

**Mahatma Education Society's
Pillai HOC College of Arts, Science and Commerce, Rasayani
(Accredited by NAAC & ISO 9001:2015 Certified)**

3.3. QnM.

Research Publication and Awards

(FROM AY 2017-18 TO AY 2021-22)




Principal
Mahatma Education Society's
Pillai's HOC College of Arts
Science and Commerce
HOC Educational Campus,
Rasayani, Tal. Khalapur,
Dist. Raigad, PIN - 410 207

**Mahatma Education Society's
Pillai HOC College of Arts, Science and Commerce, Rasayani
(Accredited by NAAC & ISO 9001:2015 Certified)**

3.3.2 QnM.

**Number of books and chapters in
edited volumes/books published and
papers published in national/
international conference proceedings
per teacher during last five years**

(FROM AY 2017-18 TO AY 2021-22)

Academic Year
2021-22
Supporting Documents

INFRASTRUCTURAL IMPEDIMENTS AND RISING COSTS OF INTERNATIONAL TRADE

A CASE STUDY OF INDIA'S EXPORT
COMPETITIVENESS IN THE GLOBAL MARKET

Dr. LATA MENON

Principal, Pillai HOC College of Applied Science & Commerce



P. Saranraj , Padmaja Santosh Kore & Narayan Dattatraya Totewad

Advances in Pharmaceutical Sciences



Volume - 1

First Edition



JPS, Scientific Publications, India

Contents

Chapter Number	Chapter Title and Authors	Page Number
1	Phytate: A Boon to Pharmaceuticals - <i>Suchita Bharambe and Swati Peshwe</i>	1 - 10
2	Pharmacologic Drugs from Natural Sources - <i>Remya Madan Gopal</i>	11 - 17
3	Actinobacteria - A Treasure House of Unexplored Biomolecules - <i>Reshma Ayswaria and Parimala Gnana Soundari Arockiam Jeyasundar</i>	18 - 37
4	Pharmaceutical Components from Marine Flora and Fauna - <i>J. Albino Wins and M. Murugan</i>	38 - 45
5	Ethanobotanical Uses of <i>Swertia chirayita</i> - <i>Abhishek Amit and Ashish Kumar Singh</i>	46 - 57
6	Pharmacological Benefits of <i>Eleutherococcus senticosus</i> - <i>B. Sindhuja, J. Rengaramanujam and J. Devakumar</i>	58 - 66
7	Pharmacological Benefits of <i>Citrus maxima</i> (Burm.) Merr (Pomelo) Fruit Juice - <i>Fathima Nadeer and Jiby Abraham</i>	67 - 76
8	General Properties and Pharmacological Benefits of <i>Syzygium jambos</i> - <i>Devakumar Joseph and Sudha Sivasubramanian</i>	77 - 83
9	Pharmacological Activities of Some Terrestrial Plants and Marine Algae - <i>T. Ramesh, A. Amuthavalli and T. Raajesh Kannan</i>	84 - 95
10	Pharmacological Benefits of <i>Cassia auriculata</i> - <i>A. S. Swarnalakshmi and Manikandan Venkatachalam</i>	96 - 106
11	Exploration of Marine Derived Compounds for Drug Discovery - <i>P. Subavathy and C. Shibana</i>	107 - 120

2

PHARMACOLOGIC DRUGS FROM NATURAL SOURCES

Remya Madan Gopal*

Pillai HOC College of Arts, Science and Commerce, Rasayani, Navi, Mumbai, Maharashtra, India

*Corresponding author: remyagopal@mes.ac.in

Abstract

Since time immemorial, humans have instinctively explored their natural surroundings to find ways to mitigate, alleviate and cure the maladies, injuries and ailments that afflict them. Rational drug design is a fairly recent development and even then, more often than not, most synthetically derived drugs trace their origin to some natural source wherein the synthetic product is a derivatised or modified form of the naturally extracted compound. Natural sources of pharmacologically important drugs can be plants, animals, microbes, minerals or the newfound and unique marine sources. Here, an attempt has been made to give a snapshot of these varied sources of drugs and each category is dealt with briefly. The compendium of naturally derived drugs is very vast and the reader is advised to search the available literature if detailed and concise information is needed on any particular source of interest.

Key words: Drugs, Pharmacology, Natural products and Plant sources.

1. Introduction

People have relied upon nature for their straightforward necessities just like the hotspots for prescriptions, covers, food stuffs, scents, clothing, flavors, manures and methods for transportation all through the ages. A drug, according to WHO definition, is any substance or product that is used or intended to be used to modify or explore the physiological system or pathological state in the benefit of the recipient. Development of a new drug usually takes one or more of these four routes:



- a) Modification of an existing approved drug.
- b) Random high throughput screening for biologic activity of large numbers of natural products, banks of previously discovered chemical entities or large libraries of organic molecules.
- c) Rational computer-aided drug design (CADD) based on an understanding of biologic mechanisms and chemical structure.
- d) Biotechnology and cloning tools

Historically, for ages, natural products have contributed to the development of new pharmaceutical drugs; approximately 50 % of all small-molecule drugs have a natural product derivation or inspiration. This would seem a pragmatic approach due to the tremendous diversity of natural products in nature.

As Frank Lovering had noted about stereo-chemical complexity in drug development, modern day rational drug design is also benefitted by the vast reservoir of complex structural information contained in many of the natural products. With the development of science of chemical analysis, the actual pharmacological active principle of the drug from the natural sources could be identified as also the mechanism of action in the body. Sertürner purified morphine from poppy in 1804 and observed that opium's analgesic and sedative effects were reproduced by morphine. This paved way for others to find the active pharmaceutical ingredient (API) of the drug, for example, quinine from cinchona, cocaine from coca etc. Further development in chemistry led to structural elucidation of the drugs and thus, we could also chemically synthesize the drug or make suitable modifications to enhance its pharmacological properties and effects.

The number of pharmacologically effective compounds that are derived from various sources is vast but the limiting factor for the actual clinical use of these compounds is more often than not, unfavorable toxic, pharmacokinetic and pharmacodynamic properties that lead to failure in clinical trials. Solubility of drugs for proper absorption is always a problem to surmount and there may be limitations to the kind of chemical modifications that may be introduced on the API for modifying it for suitable and useful properties and safety profiles. Modern drug development is a multi-billion dollar industry involving huge cost inputs but also huge returns for a successfully marketed drug. The cost and time factors involved in a new drug development and marketing has concentrated the actual drug market to a few select pharmaceutical companies referred to as Big Pharma. High Throughput Screening (HTS) and Computer Aided Drug Design (CADD) are now a routine part of rational drug design process that has contributed in bringing down drug development costs and simplified to some extent the needle in a haystack conundrum



of drug design and development. Natural sources of drugs may be broadly divided into plant sources, animal sources, microbial sources, mineral sources and semi-synthetic sources. We will look in detail into each of these sources below.

2. Plant Sources

Among the natural sources of drugs, plants are the most important and major sources of all drugs. Many plant species have known pharmacological properties. Initially, only leaves were used but now, all plant tissue parts are used and the pharmacologic properties may be due to their phyto-constituents such as glycosides, alkaloids, saponins, oils, waxes, gums and mucilages, resins, toxins, steroids, flavonoids, tannins, and terpenoids. Plant extracts may be used directly as crude drugs or further processed. Drugs from plants exert various biological activities including virucidal, bactericidal, fungicidal, anti-inflammatory, analgesic, sedative, spasmolytic, and local anesthetic activities among others. A few categories of active principle agents from above are mentioned. Alkaloids- these are mostly plant derived but some are of animal origin. They are called alkaloids because they are 'alkali-like'. Name of alkaloid-based drugs are usually suffixed with 'ine' Some important alkaloids are:

- a) Xanthine alkaloids: Caffeine
- b) Opium alkaloids: Morphine
- c) Cocaine alkaloids: Cocaine
- d) Cinchona alkaloids: Quinine
- e) Vinca alkaloids: Vinblastine
- f) Belladonna alkaloids: Atropine

Glycosides are crystalline solids composed of a sugar and non-sugar component. Pharmacological activity lies with the non-sugar component while the sugar component largely determines the pharmacokinetic properties. Digoxin, salicylic acid, digitoxin etc. are examples of glycosides.

Oils from plant sources are divided into essential oils and fixed oils. Essential oils usually contain aromatic volatile compounds that evaporate without leaving a stain like clove oil and peppermint oil. Fixed oils are higher fatty acids and triglycerides like castor oil, olive oil etc. Waxes are esters of higher fatty acids. Gums and mucilages are complex polysaccharides. Resins are bitter, insoluble polymers of volatile oils for example as a foetida. Tannins are non-nitrogenous, bitter phenolic derivatives.



3. Animal sources

Animal sources of drugs may be obtained from the whole organism, a specific part or organ or they may be glandular secretions. For example, cod liver oil is obtained from the liver of *Gadus* spp. *Hirudin* is obtained from *H. medicinalis* and heparin from *H. manillensis*.

4. Microbial sources

Many important antibiotics and life-saving drugs have been obtained from microbes. Alexander Fleming created history when he discovered penicillin, the first antimicrobial drug. Antibiotics are not the only class of drugs obtained from microbial sources. The repertoire of drugs now encompasses anticancer, lipid lowering, immunosuppressant, anti-infective, anti-atopic etc.

5. Semi-synthetic Drugs

One of the ways mentioned earlier for obtaining a new drug is to modify the properties of an existing approved drug. Many drugs obtained from natural sources can be further chemically altered and modified to impart novel and more effective pharmacological properties.

6. Marine Natural Products

An emerging and novel source of therapeutic agents is from marine sources that have distinct mechanisms of action. The limited accessibility and amount of material obtained from marine organisms had so far stunted the development of therapeutic products from these sources. Structural, chemical and biological elucidation of the compounds has made it possible to study in detail the unique therapeutic potentials of marine natural products. Scale up synthesis has now allowed drugs from marine sources to hit the market and many of them are undergoing clinical trials.

7. Antineoplastic Drugs: A special note

Cancer remains the biggest reason for morbidity and majority of drug development activities by the big pharma is directed towards finding a panacea for the same. Natural products derived anti-neoplastic drugs form a major chunk of drugs that are either obtained from natural sources or they are modified versions of some anticancer natural product. The sources are varied although maximum drugs are obtained from plants. Some of the important anti-cancer drugs obtained from natural sources are the Vinca alkaloids, taxanes, camptothecin, combretastatin, colchicine, etoposide, paclitaxel (Taxol®), docetaxel, topotecan, and irinotecan,



salvicine, ellipticine, roscovitine, maytansin, etc. Among the many microbial metabolites with anticancer properties, taxol, doxorubicin, daunorubicin, bleomycin, actinomycin D etc. are among the most important of the cancer chemotherapeutic agents.

8. Conclusion

Modern drug development depends to a great extent on high throughput screening (HTS) to find hits rummaging through chemical libraries previously synthesized or the drug design route goes through rational drug design process through computer aided drug design (CADD) in the event that structure of the druggable target protein or the ligand or both is known through X-ray crystallography or NMR. Although the methods to analyze best possible drug candidates has improved substantially, the number of drugs hitting the market is quite few. It has of course to do with safety and compatibility profile of the drugs and the stringent regulations defining safety concerns. After a dip in the past due to modern drug designing processes, the use of natural product compounds has now slowly gathered momentum recently with a better understanding of signaling pathways and also because of better lead compounds being obtained from natural sources.

9. References

- 1) Alberti F, Foster GD and Bailey AM. (2017). Natural products from filamentous fungi and production by heterologous expression. *Applied Microbiology and Biotechnology*, 101(2): 493 - 500.
- 2) Atanasov AG, Waltenberger B and Pferschy Wenzig EM. (2015). Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnology Advances*, 33(8): 1582 - 1614.
- 3) Azab A, Nassar A and Azab AN. (2016). Anti-Inflammatory Activity of Natural Products. *Molecules*, 21(10): 1321 - 1330.
- 4) Beutler JA. (2019). Natural Products as a Foundation for Drug Discovery. *Current Protocols in Pharmacology*, 86(1): e67.
- 5) Burns DC, Mazzola EP and Reynolds WF. (2019). The role of Computer-Assisted Structure Elucidation (CASE) programs in the structure elucidation of complex natural products. *Natural Products and Reproduction*, 36(6): 919 - 933.
- 6) Chikara S, Nagaprashantha LD, Singhal J, Horne D, Awasthi S and Singhal SS. (2018). Oxidative stress and dietary phytochemicals: Role in cancer chemoprevention and treatment. *Cancer Letters*, 413: 122 - 134.
- 7) Christy MP, Uekusa Y, Gerwick L and Gerwick WH. (2021). Natural Products with Potential to Treat RNA Virus Pathogens Including SARS-CoV-2. *Journal of Natural Products*, 84(1): 161 - 182.

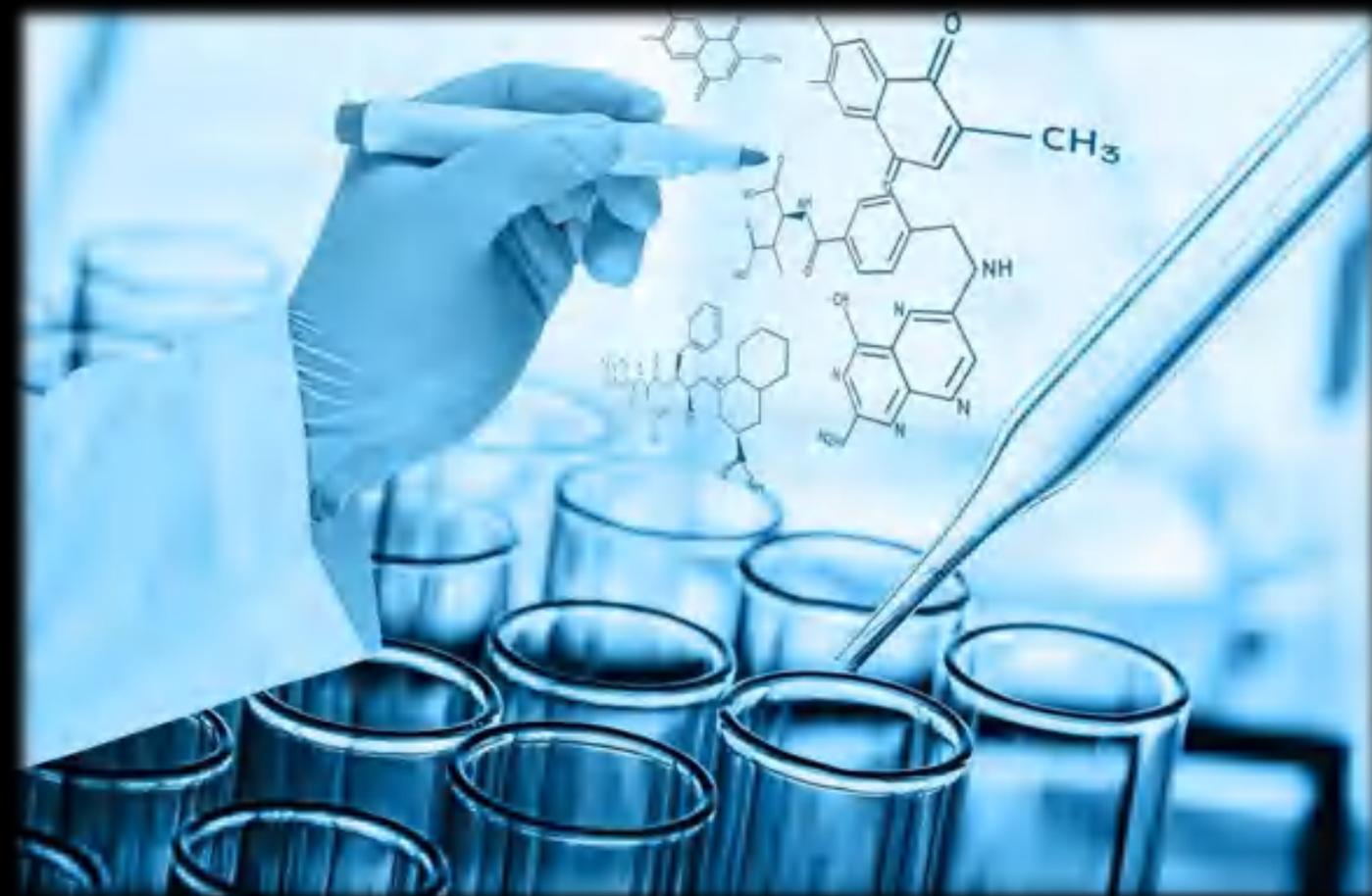


- 8) Cragg GM and Pezzuto JM. (2016). Natural Products as a Vital Source for the Discovery of Cancer Chemotherapeutic and Chemopreventive Agents. *Medical Principles and Practices*, 25(2): 41 - 59.
- 9) Da Rocha AB, Lopes RM and Schwartzmann G. (2001). Natural products in anticancer therapy. *Current Opinion in Pharmacology*, 1(4): 364 - 369.
- 10) Gezici S and Şekeroğlu N. (2019). Current Perspectives in the Application of Medicinal Plants Against Cancer: Novel Therapeutic Agents. *Anticancer Agents and Medical Chemistry*, 19(1): 101 - 111.
- 11) Hanif N, Murni A, Tanaka C and Tanaka J. (2019). Marine Natural Products from Indonesian Waters. *Marine Drugs*, 17(6): 364 - 370.
- 12) Kalkreuter E, Pan G, Cepeda AJ and Shen B. (2020). Targeting Bacterial Genomes for Natural Product Discovery. *Trends in Pharmacological Science*, 41(1): 13 - 26.
- 13) Katz L and Baltz RH. (2016). Natural product discovery: past, present, and future. *Journal of Indian Microbiology and Biotechnology*, 43(2-3): 155 - 176.
- 14) Khalifa SAM, Elias N and Farag MA. (2019). Marine Natural Products: A Source of Novel Anticancer Drugs. *Marine Drugs*, 17(9): 491 - 500.
- 15) Ma L, Zhang M, Zhao R, Wang D, Ma Y and Li A. (2021). Plant Natural Products: Promising Resources for Cancer Chemoprevention. *Molecules*, 26(4): 933 - 940.
- 16) Mondal S, Bandyopadhyay S, Ghosh MK, Mukhopadhyay S, Roy S and Mandal C. (2012). Natural products: promising resources for cancer drug discovery. *Anticancer Agents and Medical Chemistry*, 12(1): 49 - 75.
- 17) Morais MC, Souza JV and da Silva Maia Bezerra. (2020). Trypanocidal Essential Oils: A Review. *Molecules*, 25(19): 4568 - 4575.
- 18) Nugroho AE and Morita H. (2019). Computationally-assisted discovery and structure elucidation of natural products. *Journal of Natural Medicine*, 73(4): 687 - 695.
- 19) Orlikova B, Legrand N, Panning J, Dicato M and Diederich M. (2014). Anti-inflammatory and anticancer drugs from nature. *Cancer Treatment and Research*, 159: 123 - 143.
- 20) Sarmiento-Neto JF, do Nascimento LG, Felipe CF and de Sousa DP. (2015). Analgesic Potential of Essential Oils. *Molecules*, 21(1): E20 – E28.
- 21) Scuteri D, Hamamura K and Sakurada T. (2021). Efficacy of Essential Oils in Pain: A Systematic Review and Meta-Analysis of Preclinical Evidence. *Frontiers in Pharmacology*, 12: 640128.
- 22) Singh M, Sharma P, Singh PK, Singh TG and Saini B. (2020). Medicinal Potential of Heterocyclic Compounds from Diverse Natural Sources for the Management of Cancer. *Mini Reviews in Medical Chemistry*, 20(11): 942 - 957.
- 23) Sukmarini L. (2021). Recent Advances in Discovery of Lead Structures from Microbial Natural Products: Genomics and Metabolomics - Guided Acceleration. *Molecules*, 26(9): 2542 - 2550.



- 24) Thomford NE, Senthebane DA, Rowe A, Munro D, Seele P, Maroyi A and Dzobo K. (2018). Natural Products for Drug Discovery in the 21st Century: Innovations for Novel Drug Discovery. *International Journal of Molecular Science*, 19(6): 1578 - 1585.
- 25) Waltenberger B, Mocan A, Šmejkal K, Heiss EH and Atanasov AG. (2016). Natural Products to Counteract the Epidemic of Cardiovascular and Metabolic Disorders. *Molecules*, 21(6): 807 - 818.
- 26) Zhang MM, Qiao Y, Ang EL and Zhao H. (2017). Using natural products for drug discovery: the impact of the genomics era. *Expert Opinion in Drug Discovery*, 12(5): 475 - 487.





ISBN: 978-81-951323-7-9

Contents

Cemal Parlak, Berna Sümeýra Atan, Lydia Rhyman and Ponnadurai Ramasami

1 Structural and spectroscopic properties of 3-halogenobenzaldehydes: DFT and TDDFT simulations — 1

- 1.1 Introduction — 1
- 1.2 Computations — 2
- 1.3 Results and discussion — 3
 - 1.3.1 Analysis of structure — 3
 - 1.3.2 Analysis of carbonyl stretching frequency — 3
 - 1.3.3 Electronic properties — 9
- 1.4 Conclusions — 11
- Index — 13
- References — 13

Francis Opoku and Penny P. Govender

2 Atomistic insight into the significantly enhanced photovoltaic cells of monolayer GaTe₂ via two-dimensional van der Waals heterostructures engineering — 15

- 2.1 Introduction — 15
- 2.2 Computational details — 17
- 2.3 Results and discussion — 18
 - 2.3.1 Structure models and electronic properties of monolayers — 18
 - 2.3.2 The catalytic stability of the MTe₂/GaTe vdW heterostructures — 19
 - 2.3.3 Electronic properties of the MTe₂/GaTe vdW heterostructures — 20
 - 2.3.4 Optical properties — 25
 - 2.3.5 PCE of photovoltaic cells — 26
- 2.4 Conclusions — 27
- References — 28

Sulochana Bhalekar, Archana Bhagwat and Nagaiyan Sekar

3 Fluorescent styryl chromophores with rigid (pyrazole) donor and rigid (benzothienodioxide) acceptor – complete density functional theory (DFT), TDDFT and nonlinear optical study — 33

- 3.1 Introduction — 33
 - 3.1.1 Computational details — 35
 - 3.1.2 Theoretical interpretation — 35
 - 3.1.3 TDDFT analysis for absorption and emission — 36
 - 3.1.4 TDDFT emission parameters obtained from first excited state geometry — 36
 - 3.1.5 Frontier molecular orbital energies (FMOs) — 41

- 3.1.6 Electrophilicity index — 44
- 3.1.7 hyperpolarizability of 1a, 1b and 1c in all functionals — 46
- 3.1.8 Fundamental limits of hyperpolarizability — 52
- 3.1.9 Intrinsic hyperpolarizability — 53
- 3.2 Conclusions — 53
- References — 54

Suvidha Shinde, Nagaiyan Sekar

4 Comparative studies of excited state intramolecular proton transfer (ESIPT) and azo-hydrazone tautomerism in naphthalene-based fluorescent acid azo dyes by computational study — 61

- 4.1 Introduction — 61
- 4.2 Experimental section — 62
 - 4.2.1 Computational details — 62
 - 4.2.2 Synthesis — 63
- 4.3 Results and discussion — 63
 - 4.3.1 DFT study — 63
 - 4.3.2 Geometry optimization — 64
 - 4.3.3 Calculated energies of tautomeric forms — 66
 - 4.3.4 Frontier molecular orbitals — 66
 - 4.3.5 Comparative energies and FMOs of representative naphthol and its analogs — 68
 - 4.3.6 Molecular electrostatic potential analysis — 72
 - 4.3.7 Electrophilicity index — 74
 - 4.3.8 Hyper-hardness values of molecules — 75
- 4.4 Conclusions — 78
- References — 80

Ibrahim Olasegun Abdulsalami, Banjo Semire
and Isa Adewale Bello

5 Theoretical examination of efficiency of anthocyanidins as sensitizers in dye-sensitized solar cells — 83

- 5.1 Introduction — 83
- 5.2 Computational details — 86
- 5.3 Results and discussion — 87
 - 5.3.1 Molecular geometry — 87
 - 5.3.2 Frontier molecular orbitals and band gap energies of the studied molecules — 88
 - 5.3.3 Electronic properties of the molecules studied — 94
 - 5.3.4 Reactivity descriptors and electronic properties of the dye molecules in different media for the dye, semiconductor and the dye-semiconductor couples — 96

Sulochana Bhalekar, Archana Bhagwat and Nagaiyan Sekar

3 Fluorescent styryl chromophores with rigid (pyrazole) donor and rigid (benzothiophenedioxiide) acceptor – complete density functional theory (DFT), TDDFT and nonlinear optical study

Abstract: Density functional theory (DFT) and time-dependent DFT computations were employed to examine linear and nonlinear optical (NLO) characteristics of (E)-4-((1,1-dioxido-3-oxobenzo[b]thiophen-2(3H)-ylidene) ethyl)-1-phenyl-1H-pyrazol-5 (4H)-one derived styryl dyes. NLO properties were computed using the two different global hybrid functionals B3LYP, BHandHLYP and three range separated hybrid functionals CAM B3LYP, wB97, wB97X and wB97XD with the basis sets 6–311++G(d,p), cc-pVDZ and cc-pVTZ. The compounds shows higher values of dipole moment around 8–9 Debye than the other compounds. They show higher values of α_0 , β_0 and γ_0 values. The values of γ_0 are around $204\text{--}544 \times 10^{-36}$ e.s.u. with the method, B3LYP/6–311++G(d, p). We have calculated the mean absolute error (MAE) for dipole moment, α_0 , β_0 and γ_0 values. It is observed that MAE is less (0.89) for wB97/6–311++G(d,p) which indicates that wB97 is the most suited functional for all three compounds. Chemical stability and reactivity of these dyes were studied using electrophilicity index and chemical hardness and hyperhardness.

Keywords: electrophilicity index, MAE, NLO, pyrazole

3.1 Introduction

DFT is a powerful tool in the study of molecular properties [1–4]. TDDFT is used in the study of excited state properties in combination with the experimental findings [5, 6]. In the last three decades, DFT is extensively used to generate the description of chemical reactivity and it's influence in the solvent environment. Density functional approximations (DFA) are of two types generalized gradient approximations (GGA) and Local spin density approximations(LSDA), and they are used in the geometry optimization, and in obtaining accuracy of atomization energies [8–10].

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Bhalekar, S., Bhagwat, A., Sekar, N. Fluorescent styryl chromophores with rigid (pyrazole) donor and rigid (benzothiophenedioxiide) acceptor – complete density functional theory (DFT), TDDFT and nonlinear optical study Physical Sciences Reviews [Online] 2021, 1. DOI: 10.1515/psr-2019-0129

DFA employ global hybrid (GH) and range separated hybrid (RSH) functionals which have exchange correlations like short range (SR) and long range (LR) Hartree–Fock (HF) exchange SR and LR separation in these functionals are given by the equation;

$$\frac{1}{r_{12}} = \frac{\text{erf}(wr_{12})}{r_{12}} + \frac{\text{erfc}wr_{12}}{r_{12}}$$

LR SR

where, $r_{12} = r_1 - r_2$

Here, we have used different GH functionals like B3LYP (20%HF exchange), BHandHLYP(50% HF exchange) and RSH functionals like CAMB3LYP (19% SR and 65% LR) wB97 (100%LR), wB97X (15.77%SR and 100%LR) and wB97XD (22.2%SR and 100%LR) HF exchange [11]. DFT is used to illustrate and define the structure and chemical reactivity of molecules. Chemical reactivity is measured by using electronegativity (χ), chemical potential (μ), chemical hardness (η) and electrophilicity index [12, 13]. Organic Donor- π -Acceptor (D- π -A) motifs possessing large second order nonlinearity have large importance nowadays, due to their possible applications in the area of electro optic devices, [14] nanophotonics and biological imaging [15]. Compounds containing heteroaromatic ring such as thiophene [16–21] pyrrole [22], pyrazole [23, 24] show enhancement in their NLO properties.

Usually in a D- π -A system, effective charge separation leading to non-centrosymmetric charge distribution leads to a larger molecular NLO response [24–30]. Efficiency of Donor-Acceptor overlap is enhanced by rigidification of donor or acceptor. Examples of rigidified donors are julolidine derivatives [31–37], and carbazole [38–40], indanedione [41, 42]. Employing rigidified donor as well as acceptor in a single system is also used [43–45]. The present work describes computational investigation of D- π -A system having both donor and acceptor rigidified.

Pyrazole is electron rich rigid donor because the lone pair electrons on nitrogen are constrained by rigidity of the system, and therefore it is an effective rigidified donor. Five membered heterocyclic ring increases the donor strength and polarizability of the chromophoric backbone [46]. N-Aryl pyrazole derivatives have wide applications in the field of Dye Sensitized Solar Cell (DSSC) [47]. They are also extensively studied in the Organic Light Emitting Diodes (OLEDs) as blue emitting materials [48]. Pyrazole derivatives have been used in the field of electroluminescence as well as electro photography because of high whole transfer efficiency [49]. Pyrazole based derivatives show high magnitudes of first order hyperpolarizability as well as dipole moment which attributes for their NLO properties [50–52]. Pyrazoline derivatives show high fluorescent quantum yield as well as two photon absorption values [15, 53].

Benzo[b]thiophen-3 (2H)-one 1,1-dioxide (BTD) a sulfonyl group containing analog of 1,3-indanedione (1a), is a powerful rigid acceptor which generates high optical and NLO properties, and have high thermal stability along with different donors as well as spacers [54]. Styryl disperse dyes along with benzo [b]thiophen-3 (2H)-one

1,1-dioxide as acceptor show very good light fastness properties and they have applications in the dyeing of polyesters [55, 56].

Here, we have studied the NLO studies of compounds containing heterocyclic ring pyrazole as a rigid donor and BTB, a rigid acceptor. This work describes DFT, TDDFT and NLO properties of pyrazole substituted dyes. The static dipole moment (μ), static polarizability (α), first order hyperpolarizability (β) and second order hyperpolarizability (γ) of these compounds are demonstrated by using above mentioned functionals and basis sets. Also, chemical stability and chemical reactivity of dyes are illustrated by using the electrophilicity index. The compound 1b contains one phenyl group attached instead of methyl group which have effect on NLO properties. Values of α , β and γ are comparatively higher compared to other two compounds.

3.1.1 Computational details

All the computations were performed on the Gaussian 09 package. For ground state optimization, DFT methods are used and for excited state optimization time-dependent DFT was used. Global hybrid functionals used are B3LYP (Becke3-LeeYang-Parr hybrid functionals). B3LYP, (functional with 20% Hartree Fock exchange, and BHandHLYP (hybrid functional with 50% HF exchange for both short and long range separation). Range separated hybrid functionals used are CAM-B3LYP (19% short range and 65% long range HF exchange) [57], wB97 (0% HF exchange) wB97X (15% HF exchange) [8] and wB97XD (22.20% short-range and 100% long-range HF exchange) [58, 59]. Emissions were obtained by computing vertical excitations of excited state geometry at the ground state. The ground state optimized geometries were subjected to TDDFT computations for the first 10 states using B3LYP/6-311++G(d,p) methods for the calculations of the vertical excitation energies and oscillator strength.

3.1.2 Theoretical interpretation

3.1.2.1 Optimized geometries in the ground state and excited state

Molecular geometries of the dyes 1a, 1b and 1c were optimized in their ground state (S_0) and excited state (S_1) with DFT and TDDFT using B3LYP/6-311++G(d,p) for all the atoms in the gas phase and solvents DMF and EA. For 1c bond lengths observed between the bonds C8-O13, C7-C8, C17-C7, C17-C20, C20-C21, C21-N23, N23-N22, N22-C19, C19-C20, C19-C39, C19-Cl24, N22-C34, C7-S14 were 1.25676, 1.43903, 1.38462, 1.45963, 1.46875, 1.03075, 1.40386, 1.38328, 1.3769, 1.71333, 1.38304 and 1.77829, respectively. It is noted that there is a major change in dihedral angle observed in dye 1a C7-C17-C19-C38, S14-C7-C17-H18, C38-N21-C28-C23, in both the gas phase and in the solvent phase (DMF and EA) both the ground and excited states; and for 1b and 1c, there is major change observed in dihedral angle in C7-C17-C20-C21 and S14-C7-C17-H18 in gas phase and in

DMF. In 1b, there is also change in dihedral angle in N23-C21-C37-C38 and C20-C21-C37-C36 in both the gas phase and in DMF and EA. Major changes in bond angles observed are in C7-C17-C20 in 1a, 1b and 1c in both DMF and EA, and also H18-C17-C20 in 1a and 1c in DMF. Variation in bond lengths in both G.S (black color) and E.S (red color) are shown in Figures 3.1, 2 to understand the electron flow from pyrazole to benzothioophene ring. Table 3.1 represents the bond length, bond angle and dihedral angle of 1a, 1b and 1c in DMF and EA in the ground and excited states.

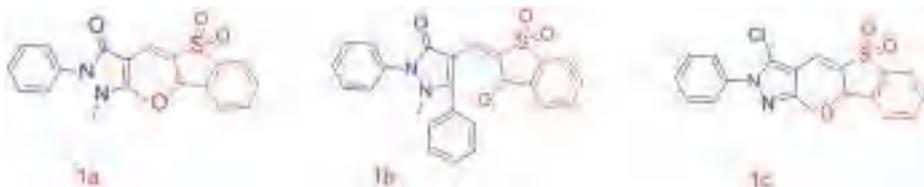


Figure 3.1: Structures of compounds.

3.1.3 TDDFT analysis for absorption and emission

3.1.3.1 Electronic vertical excitation spectra

Vertical excitations of the compounds and associated oscillator strengths along with the major orbital contribution were obtained from TDDFT using the global hybrid (GH) functionals B3LYP, BHandHLYP and RSH functionals CAM B3LYP, wB97, wB97X and wB97XD and the basis sets 6-311++G(d,p), cc-pVDZ and cc-pVTZ. The compound 1a has absorption 446 nm in DMF and 424 nm in EA) 1b 432 nm in DMF and 422 nm in EA and 1c 460 nm in DMF and 448 nm in EA. The geometries optimized in the ground state were used for TDDFT computations for the first 20 states. We have calculated mean absolute error to find out exactly which functional shows accurate results with experimental λ_{max} . It is observed that B3LYP/6-311++G(d,p) shows comparable results with the experimentally determined values with MAE 36 (Figure 3.3) Table 3.2. In both the solvents, vertical excitations are associated with the HOMO-LUMO transitions with major contribution >97.71% for compounds 1a, and 1b while for 1c, it is 56.20% in DMF and 50.61% in EA. The absorption spectra of all the dyes are mainly due to transition from the HOMO to the LUMO in both the solvents. The values of vertical excitation, oscillator strength and major orbital contribution are represented in Table S1.

3.1.4 TDDFT emission parameters obtained from first excited state geometry

The TDDFT emission values were obtained using optimized geometry of the first excited state using the method TD B3LYP/6-311++G(d,p), and the trends observed in the computed values are similar to the trends in the experimentally observed values

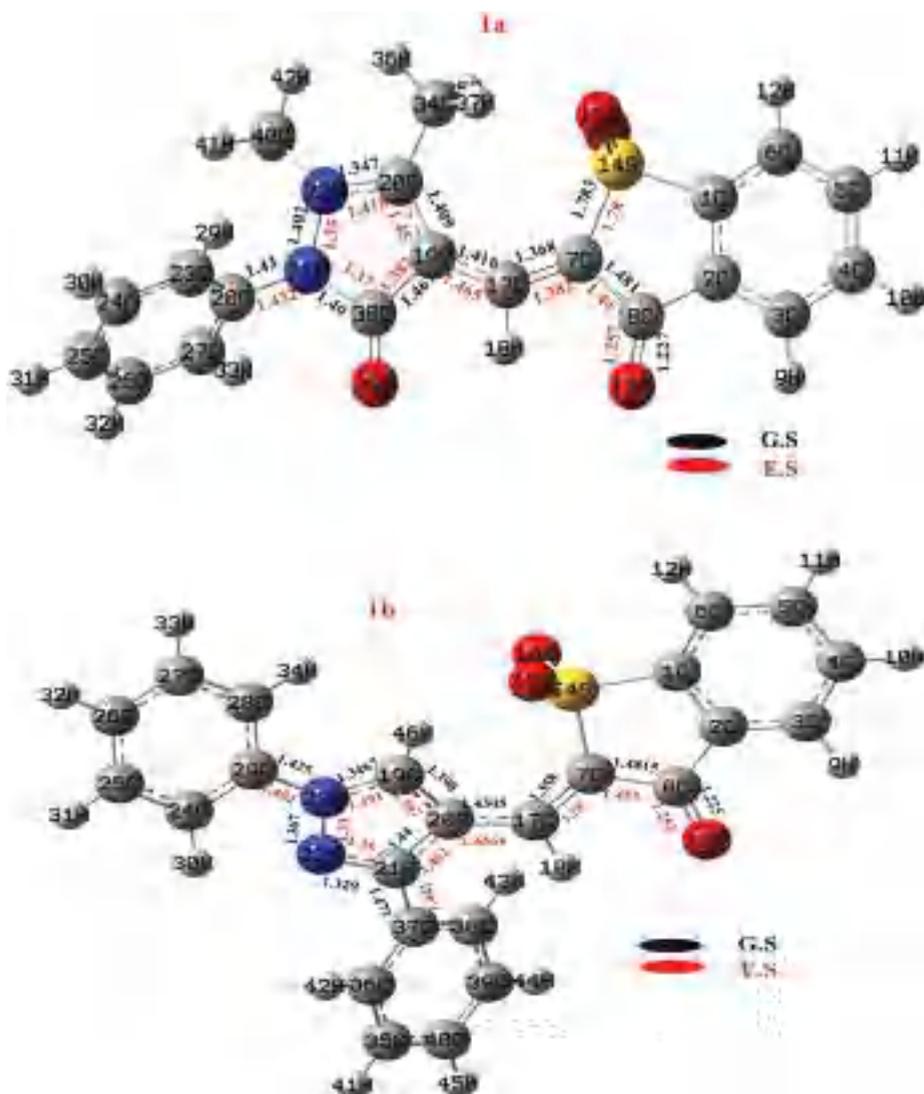


Figure 3.2: Optimized geometry of 1a and 1b.

[60]. As shown in Table 3.3, in both the solvents vertical excitations are associated with the HOMO-LUMO transitions with major contribution >98% for compounds 1a and 1b, while for 1c, it is 88 % in DMF and 77 % in EA.

Table 3.1: Bond length, bond angle and dihedral angle of 1a, 1b and 1c in DMF and EA.

Atom number	1a			1a	
	Gas	DMF		EA	
	GS	GS	ES	GS	ES
Bond length					
C8-O13	1.221	1.227	1.257	1.225	1.247
C7-C8	1.489	1.481	1.440	1.483	1.454
C17-C7	1.360	1.368	1.382	1.365	1.387
C19-C17	1.429	1.416	1.465	1.420	1.437
C19-C38	1.392	1.467	1.381	1.403	1.373
C38-N21	1.370	1.403	1.373	1.353	1.448
N21-N22	1.407	1.402	1.350	1.403	1.385
N22-C20	1.415	1.347	1.418	1.408	1.449
C20-C19	1.474	1.409	1.459	1.468	1.463
C38-O39	1.219	1.228	1.226	1.225	1.216
N21-C28	1.425	1.430	1.432	1.428	1.374
C7-S14	1.789	1.783	1.780	1.785	1.776
Bond angle					
C7-C17-C19	138.2	139.9	124.1	139.4	135.3
H18-C17-C19	110.7	110.1	118.5	110.2	113.1
N21-N22-C28	119.5	120.8	124.8	120.4	122.1
C38-N21-C38	123.2	123.9	127.2	123.3	126.8
Dihedral angle					
O13-C8-C7-C17	-5.6	-4.0	0.6	-4.6	-7.4
S14-C7-C17-C19	-1.5	-1.5	4.0	-1.7	1.0
C7-C17-C19-C38	171.2	174.1	90.1	173.1	160.6
C38-N21-C28-C27	-60.1	-76.5	-67.9	-69.2	-31.3
N21-N22-C28-C23	-25.2	-46.3	-68.3	-36.7	-19.9
C7-C17-C19-C20	-10.5	-6.9	-89.2	-8.2	-24.8
S14-C7-C17-H18	175.5	176.2	-178.4	175.8	174.6
C38-N21-C28-C23	155.7	101.7	109.7	109.1	148.4
Atom number	1b			1b	
	Gas	DMF		EA	
	GS	GS	ES	GS	ES
Bond length					
C8-O13	1.222	1.225	1.253	1.224	1.253
C7-C8	1.482	1.481	1.453	1.481	1.455
C17-C7	1.357	1.358	1.380	1.358	1.375
C17-C20	1.432	1.430	1.456	1.431	1.464
C20-C21	1.442	1.441	1.462	1.441	1.459
C21-N23	1.328	1.328	1.354	1.328	1.357
N23-N22	1.367	1.367	1.325	1.367	1.323

Table 3.1 (continued)

Atom number	1b			1b	
	Gas	DMF		EA	
	GS	GS	ES	GS	ES
N22-C19	1.348	1.346	1.401	1.347	1.402
C20-C19	1.396	1.398	1.367	1.397	1.363
C21-C37	1.473	1.473	1.440	1.473	1.439
C22-C29	1.424	1.425	1.403	1.424	1.404
N22-C29	1.424	1.425	1.402	1.424	1.403
C7-S14	1.793	1.788	1.784	1.788	1.786
Bond angle					
C7-C17-C20	130.6	131.1	128.1	131.1	127.3
H18-C17-C20	116.2	115.8	117.7	115.8	118.0
N23-N22-C28	119.9	119.9	121.3	119.9	121.4
C21-N23-C21	127.7	127.6	126.7	127.6	126.6
Dihedral angle					
O13-C8-C7-C17	-0.2	-1.1	-2.0	-0.9	-0.7
S14-C7-C17-C20	-5.9	-6.0	-5.8	-5.9	-3.1
C7-C17-C20-C19	-18.8	-18.2	-36.4	-18.2	-44.2
C19-N22-C34-C33	20.7	28.6	11.5	26.9	12.7
N23-N22-C34-C29	20.1	28.1	11.7	26.5	13.1
C7-C17-C20-C21	167.9	168.9	177.0	168.8	140.8
S14-C7-C17-H18	174.9	174.7	177.9	174.8	-179.3
C19-N22-C29-C24	-159.5	-151.2	-168.7	-152.8	-167.5
N23-C21-C37-C38	143.4	141.4	166.5	141.7	170.2
C20-C21-C37-C36	144.4	142.9	169.8	143.2	172.9
Atom number	1c			1c	
	Gas	DMF		EA	
	GS	GS	ES	GS	ES
Bond length					
C8-O13	1.222	1.224	1.257	1.224	1.256
C7-C8	1.493	1.492	1.439	1.492	1.441
C17-C7	1.357	1.358	1.384	1.358	1.382
C17-C20	1.434	1.431	1.459	1.432	1.462
C20-C21	1.437	1.439	1.468	1.439	1.468
C21-N23	1.320	1.321	1.307	1.321	1.307
N23-N22	1.372	1.373	1.403	1.373	1.403
N22-C19	1.357	1.352	1.383	1.353	1.388
C20-C19	1.402	1.404	1.376	1.403	1.373
C19-Cl24	1.709	1.712	1.713	1.711	1.713
N22-C34	1.428	1.430	1.383	1.430	1.379
C7-S14	1.800	1.794	1.778	1.796	1.779

Table 3.1 (continued)

Atom number	1c			1c	
	Gas	DMF		EA	
	GS	GS	ES	GS	ES
Bond angle					
C7-C17-C20	137.2	137.9	124.9	137.8	124.5
H18-C17-C20	112.7	112.2	118.2	112.2	118.4
N23-N22-C28	118.7	119.0	116.9	118.9	116.9
C21-N23-C21	129.9	129.6	132.7	129.6	132.8
Dihedral angle					
O13-C8-C7-C17	-10.9	-10.7	-0.8	-10.6	-0.8
S14-C7-C17-C20	-3.4	-3.6	0.5	-3.6	0.03
C7-C17-C20-C19	-20.7	-18.9	-93.5	-18.9	-92.8
C19-N22-C34-C33	47.6	53.4	19.8	51.9	17.9
N23-N22-C34-C29	43.8	49.5	19.1	48.1	17.9
C7-C17-C20-C21	160.2	162.3	89.7	162.1	89.7
S14-C7-C17-H18	172.8	173.0	179.1	173.0	179.1

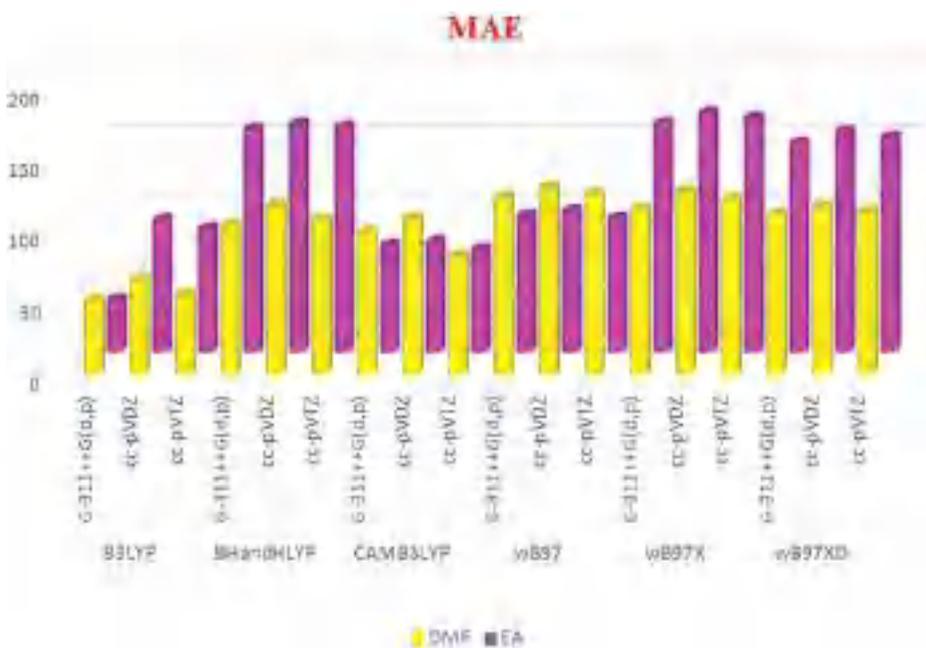


Figure 3.3: MAE of vertical excitations of all the compounds in different functionals and basis sets.

Table 3.2: MAE of vertical excitations of all the compounds in different functionals and basis sets.

Functional	Basis set	MAE	
		DMF	EA
B3LYP	6-311++G(d,p)	52.11	36.39
	cc-pVDZ	67.04	93.09
	cc-pVTZ	56.10	86.45
BHandHLYP	6-311++G(d,p)	105.46	156.79
	cc-pVDZ	119.46	160.47
	cc-pVTZ	109.99	158.41
CAMB3LYP	6-311++G(d,p)	100.85	75.076
	cc-pVDZ	109.80	77.386
	cc-pVTZ	83.703	71.823
wB97	6-311++G(d,p)	125.01	95.826
	cc-pVDZ	131.80	99.156
	cc-pVTZ	127.08	93.863
wB97X	6-311++G(d,p)	116.64	160.72
	cc-pVDZ	129.25	168.60
	cc-pVTZ	123.08	164.89
wB97XD	6-311++G(d,p)	112.69	147.05
	cc-pVDZ	117.67	155.04
	cc-pVTZ	114.79	151.08

3.1.5 Frontier molecular orbital energies (FMOs)

The FMOs (HOMO and LUMO) play a major role in determining the molecular properties like UV–visible absorption, optical and electronic properties [61]. The HOMO and LUMO energy diagram gives a quantitative idea of electronic structure and excitation properties. The electronic distribution in the frontier molecular orbitals, HOMO and LUMO in 1b is shown in Figure 3.4. In compound 1c, the HOMO is found on pyrazole ring and five-membered thiophene dioxide ring and the LUMO is placed on benzothienophene dioxide and pyrazole. In the compound 1b, HOMO is seen on diphenyl pyrazole and thiophene dione ring, while the LUMO is located on benzothienophene and pyrazole ring. In the compound 1a, the HOMO is located on phenylpyrazolidin-3 one and thiophene dioxide ring, while the LUMO is located on benzothienophene and pyrazolidine ring.

The energy gap between HOMO and LUMO as well as HOMO-1–LUMO+1 are represented in Figure 3.5. The compounds 1a and 1c show low band gap (1.97 and 1.83eV, respectively) which indicates that a maximum conjugation takes place. The compound 1b shows a band gap 2.388 eV which indicates that a minimum conjugation takes place.

Molecular electrostatic potential (MEP) is an important plot for visualizing negative and positive regions in a molecule. Electrophilic reactivity is understood in the negative (red) regions of the plot, while nucleophilic reactivity is visualized in the

Table 3.3: Vertical excitation, oscillator strength and orbital contribution in the excited state by Root TD method.

Compound	Solvent	λ_{em}^f (nm)	$\Delta\nu^p$ (cm^{-1})	λ_{em}^h (nm)	$\Delta\nu^l$ (cm^{-1})	f^l	Major contribution	%	HOMO-LUMO
1a	DMF	476	21008	411	24319	0.06	0.7010	98.28	99 → 100
	EA			407	24526	0.68	0.7025	98.69	99 → 100
1b	DMF	470	21276	414	24148	0.61	0.7018	98.51	107 → 108
	EA			411	24309	0.54	0.7017	98.49	107 → 108
1c	DMF	498	20080	373	26796	0.83	0.6660	88.72	99 → 100
	EA			370	26977	0.65	0.6223	77.45	99 → 100

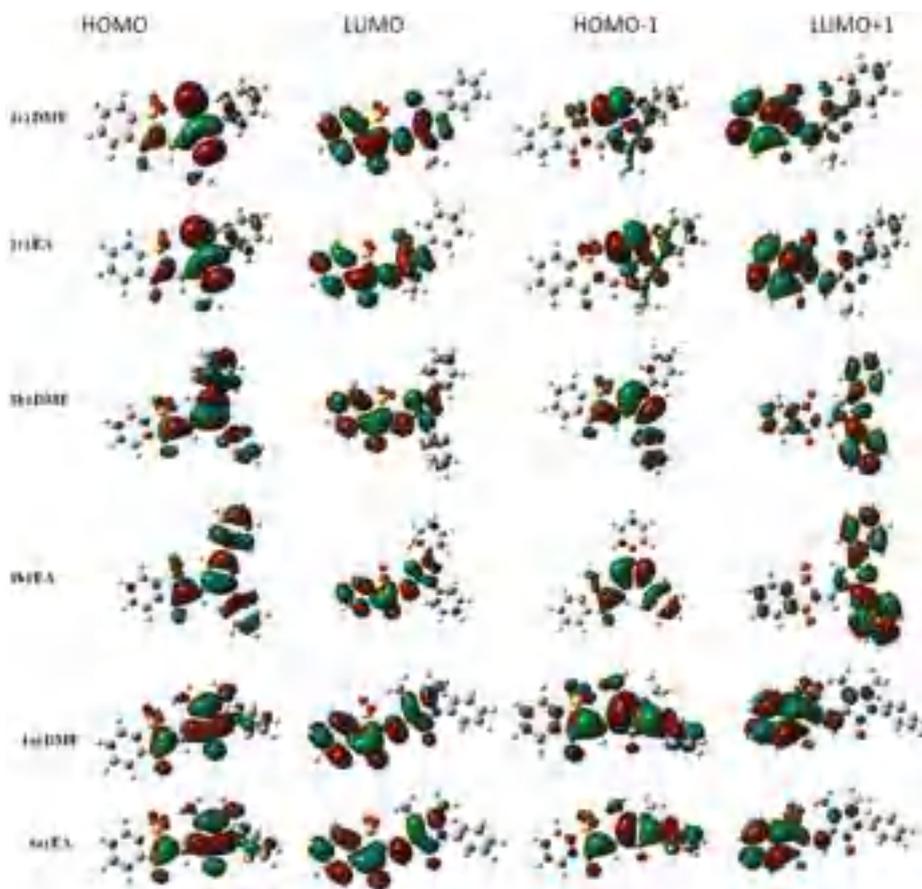


Figure 3.4: FMO diagrams of all the three compounds.

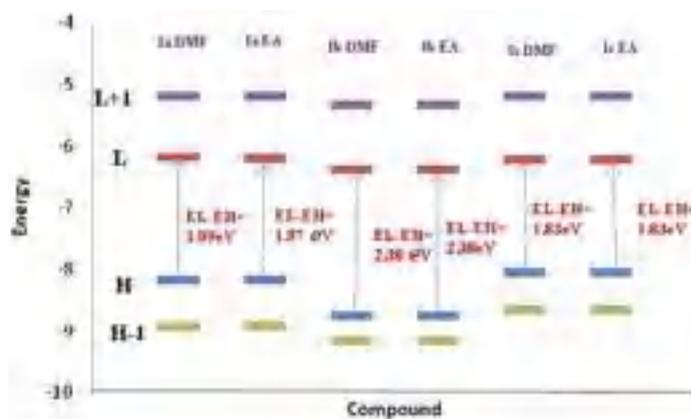


Figure 3.5: Energy band gap in HOMO, LUMO and HOMO-1 LUMO+1.

positive (blue) regions. In Figure 3.6, it is observed that the electrophilic reactivity found on phenyl rings and oxygen atoms. Nucleophilic reactivity is found on hydrogen and nitrogen.



Figure 3.6: MEP plots of the dyes 1a, 1b and 1c.

3.1.6 Electrophilicity index

Electrophilicity is a measure of the reactivity of compounds which is an intrinsic property that shows the electron-deficient and -rich species based on Lewis theory [62]. Electrophilicity index measures the stability of molecules and it depends on both chemical potential (μ) and chemical hardness (η).

Electrophilicity index can be given by the equation,

$$\omega = \frac{\mu^2}{2\eta} \quad (3.1)$$

where μ is chemical potential and η is chemical hardness. Electron affinity and electrophilicity index bear a relation with each other because electron affinity is the tendency of the electrophile to accept the electron from nucleophile and electrophilicity index is the measure of the electron flow from donor to acceptor [63, 64].

It involves the chemical potential (μ)

$$\mu = -\frac{1}{2}(I + A) \quad (3.2)$$

Chemical hardness (η) is given by equation

$$\eta = I - A \quad (3.3)$$

Electronegativity is given by equation

$$\chi = \frac{I + A}{2} \quad (3.4)$$

$$\eta = (\epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}})$$

$$I = \epsilon_{\text{HOMO}}$$

$$E = -\epsilon_{\text{LUMO}}$$

where I and A are ionization potential and electron affinity. According to Koopmans theory, I and A can be given as HOMO and LUMO. Chemical hardness indicates the stability of compounds. Along with electrophilicity index, we have calculated hyperhardness γ by using the equation [63–66].

$$\gamma = \epsilon_{\text{LUMO}} - 2 \times \epsilon_{\text{HOMO}} + \epsilon_{\text{HOMO}} - 1 \quad (3.5)$$

Table 3.4 represents the values of hyperhardness which shows negative values which indicates that the molecule is either good electrophile or good nucleophile. We have calculated chemical hardness and electrophilicity index using the above equations and correlated the chemical hardness with hyperpolarizability. The compound 1a shows comparatively higher values of electronegativity which indicates the tendency to attract electrons toward it which indicates withdrawing power of compounds which is again supported by electrophilicity index which also shows higher values in all functionals. Table 3.4 and Table S2 represent the chemical hardness, chemical potential and electrophilicity index of compounds 1a, 1b and 1c which shows higher values of ω indicating the reactivity and stability of compounds. Also, we compared chemical hardness and α , β and γ , as shown in Figures 3.1 and 7 which indicates functional wB97 shows comparatively higher values for chemical hardness than the other functionals. The observation indicates the less tendency to liberate electron which shows the chemical stability and hardness of compounds.

Table 3.4: Chemical potential, chemical hardness electrophilicity index and hyperhardness of all three compounds in DMF and EA.

1a DMF	E_{HOMO} (au)	E_{LUMO} (au)	μ (eV)	η (eV)	ω (eV)	ω^+ (eV)	ω^- (eV)	ω^* (eV)	Γ
B3LYP	-0.22006	-0.12091	-4.64	2.69	3.99	2.01	6.65	8.65	-0.90
BHandHLYP	-0.25208	-0.0862	-4.60	4.51	2.35	0.61	5.21	5.82	-0.83
CAMB3LYP	-0.26251	-0.08396	-4.71	4.86	2.29	0.54	5.25	5.79	-0.82
wB97	-0.29783	-0.04863	-4.71	6.78	1.64	0.13	4.84	4.97	-0.75
wB97X	-0.29263	-0.05375	-4.71	6.49	1.71	0.17	4.88	5.04	-0.76
wB97XD	-0.28126	-0.06304	-4.68	5.94	1.85	0.25	4.93	5.18	-0.78
1a EA	E_{HOMO} (au)	E_{LUMO} (au)	μ (eV)	η (eV)	ω (eV)	ω^+ (eV)	ω^- (eV)	ω^* (eV)	Γ
B3LYP	-0.21969	-0.11893	-4.61	2.74	3.87	1.91	6.52	8.43	-0.89
BHandHLYP	-0.25226	-0.10203	-4.82	4.09	2.84	0.94	5.76	6.71	-0.85
CAMB3LYP	-0.26269	-0.09951	-4.93	4.44	2.73	0.83	5.75	6.58	-0.84
wB97	-0.29748	-0.06582	-4.94	6.30	1.94	0.25	5.19	5.45	-0.77
wB97X	-0.29239	-0.07053	-4.94	6.04	2.02	0.31	5.24	5.55	-0.78
wB97XD	-0.2812	-0.07905	-4.90	5.50	2.18	0.42	5.32	5.74	-0.79

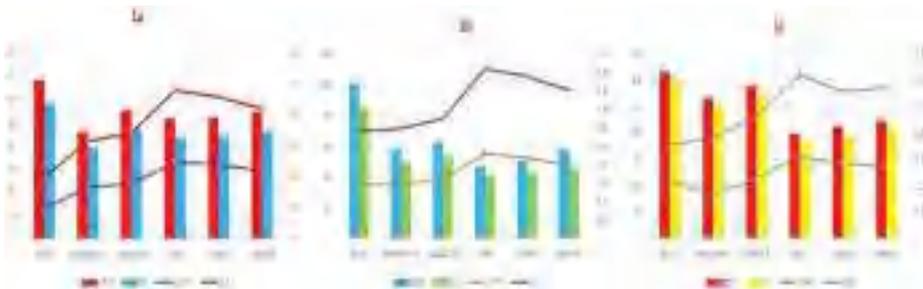


Figure 3.7: Comparison of chemical hardness and first-order hyperpolarizability of 1a, 1b and 1c.

3.1.7 hyperpolarizability of 1a, 1b and 1c in all functionals

3.1.7.1 Nonlinear optical properties

We have done the theoretical calculations using GH functionals B3LYP, BHandHLYP and RSH functionals, CAM-B3LYP, wB97, wB97X and functionals with Pople's basis set 6-311++G(d,p) and cc-pVDZ and cc-pVTZ. We have computed the values of static dipole moment (μ), mean polarizability (α_0), polarizability anisotropy ($\Delta\alpha$), static first hyperpolarizability (β_0) and second hyperpolarizability (γ) were calculated in DMF and EA solvents by using equations,

3.1.7.2 Dipole moment calculation

The total static dipole moment (μ) is expressed by using equation,

$$\mu = \left(\mu_x^2 + \mu_y^2 + \mu_z^2 \right)^{1/2} \quad (3.6)$$

The dipole moment is the response of energy with the external electric field which indicates the charge distribution of electrons in the molecules [67]. The dipole moments of all the compounds in the solvents DMF and EA were calculated using B3LYP BHandHLYP, wB97, wB97X and wB97XD functionals and basis sets and we have calculated their mean absolute error for accurate results. It is observed both the compounds show very fewer values for MAE which are given in Figure 3.8 and Tables 3.5a,b. The compound 1a shows higher values for dipole moment around 8.55–9.65 Debye while other two compounds show comparatively lower values for dipole moment 1–2 Debye. Due to the dipole-dipole interaction, there is increase in dipole moment in solvents is observed compared to gas. For compound 1a, dipole moment in gas phase is observed around four Debye and in solvents it is around 7–9 Debye. Figure 3.5a indicates the dipole moment in G.S and E.S in gas, DMF and EA.

The isotropic polarizability (α_0) can be calculated by using equation,

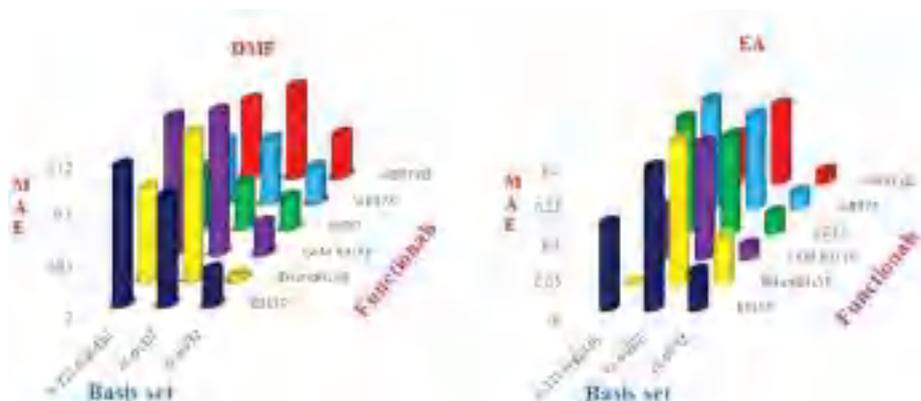


Figure 3.8: MAE of dipole moments of 1a, 1b and 1c in all functionals and basis sets.

Table 3.5a: Dipole moments of 1a, 1b and 1c in Gas, DMF and EA in G.S and E.S. Table 3.5b-MAE of dipole moments of 1a, 1b and 1c in all functionals and basis sets.

Comp	$\mu_g(\text{DMF})$				$\mu_g^a(\text{DMF}) - \mu_g^a(\text{GAS})$
	μ_g^a	μ_g^a	μ_g^b	$\mu_g^a - \mu_g^b$	
1a	4.35	9.46	9.40	0.06	5.11
1b	1.40	2.07	3.14	-1.07	0.67
1c	0.88	1.12	4.50	-3.38	0.24
Comp	$\mu_g(\text{EA})$				$\mu_g^a(\text{EA}) - \mu_g^a(\text{GAS})$
	μ_g^a	μ_g^a	μ_g^b	$\mu_g^a - \mu_g^b$	
1a	4.35	7.62	7.28	0.34	3.28
1b	1.40	1.88	3.42	-1.54	0.48
1c	0.88	1.05	4.37	-3.32	0.17

^aDipole moment in ground state, ^bDipole moment in excited state.

Table 3.5b: MAE of dipole moments of 1a, 1b and 1c in all functionals and basis sets.

Dipole moment		Comp 1c		Comp 1b		Comp 1a		MAE	
Functional	Basis set	DMF	EA	DMF	EA	DMF	EA	DMF	EA
B3LYP	6-311++G(d,p)	1.22	1.19	2.26	2.01	9.46	7.62	0.14	0.16
	cc-pVDZ	1.21	1.12	1.97	1.79	8.99	7.23	0.11	0.189
	cc-pVTZ	1.2	1.15	2.1	1.89	9.33	7.52	0.036	0.051
BHandHLYP	6-311++G(d,p)	1.19	1.1	2.53	2.26	9.08	7.34	0.09	0.0016
	cc-pVDZ	1.19	1.11	2.24	2.03	8.65	6.99	0.15	0.19
	cc-pVTZ	1.16	1.12	2.34	2.09	9.01	7.29	0.005	0.064

Table 3.5b (continued)

Dipole moment		Comp 1c		Comp 1b		Comp 1a		MAE	
Functional	Basis set	DMF	EA	DMF	EA	DMF	EA	DMF	EA
CAM B3LYP	6-311++G(d,p)	1.2	1.15	2.55	2.28	9.18	7.4	0.14	0.0678
	cc-pVDZ	1.19	1.18	2.21	2.02	8.68	6.98	0.14	0.15
	cc-pVTZ	1.19	1.25	2.37	2.13	9.07	7.31	0.03	0.018
wB97	6-311++G(d,p)	1.19	1.41	2.51	2.43	9.01	7.25	0.06	0.15
	cc-pVDZ	1.18	1.2	2.66	2.17	8.55	6.88	0.045	0.13
	cc-pVTZ	1.19	1.25	2.33	2.27	8.92	7.19	0.031	0.027
wB97X	6-311++G(d,p)	1.18	1.37	2.47	2.38	9.067	7.31	0.062	0.143
	cc-pVDZ	1.16	1.19	2.54	2.12	8.63	6.95	0.063	0.122
	cc-pVTZ	1.19	1.23	2.24	2.22	8.98	7.24	0.035	0.021
wB97XD	6-311++G(d,p)	1.19	1.3	2.37	2.1	9.19	7.42	0.076	0.063
	cc-pVDZ	1.18	1.17	2.26	2.04	8.81	7.11	0.09	0.1
	cc-pVTZ	1.18	1.19	2.37	2.12	9.11	7.35	0.042	0.015

$$\alpha_0 = \left(\frac{1}{3}\right) [\alpha_{xx}\alpha_{yy}\alpha_{zz}] \quad (3.7)$$

where α_{xx} , α_{yy} , and α_{zz} are polarizability tensor components

Anisotropy polarizability can be calculated by using equation,

$$\Delta\alpha = 2^{-1/2} [(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{yy} - \alpha_{zz})^2 + (\alpha_{zz} - \alpha_{xx})^2 + 6\alpha_{xx}^2]^{1/2} \quad (3.8)$$

The mean static first-order hyperpolarizability (β_0) can be calculated by using the equation

$$(\beta_0) = (1/2) [\beta_x^2 + \beta_y^2 + \beta_z^2] \quad (3.9)$$

$$\beta_{\text{Total}} = \frac{1}{2} (\beta_{xxx} + \beta_{xyy} + \beta_{zzz})^2 + (\beta_{yyy} + \beta_{xxy} + \beta_{yzz})^2 (\beta_{zzz} + \beta_{xxx} + \beta_{yyz})^2 \quad (3.10)$$

The mean second-order hyperpolarizability (γ) is expressed by,

$$\gamma = \frac{1}{5 \left[(\gamma_{xxxx} + \gamma_{yyyy} + \gamma_{zzzz}) + 2(\gamma_{xxyy} + \gamma_{yyzz} + \gamma_{xxzz}) \right]} \quad (3.11)$$

The values are calculated for μ , α_0 , $\Delta\alpha$, β_0 and γ for the compounds in DMF and EA using equations above for the above referred three distinct computational methods, respectively. The compound 1a shows higher dipole moment in both solvents (7.11 and 9.46 Debye) in the above GH and RSH functionals while the compounds 1b and 1c show lower values (1.03–2.34) in both solvents of dipole moment while all the other compounds show higher values of for mean polarizability as well as first-order

hyperpolarizability in B3LYP functional than CAM-B3LYP and BHandHLYP functionals. The values of computed first-order polarizability are found to be greater than that of the reported value of urea (0.38×10^{-30} e.s.u.) by several folds. Likewise, the values of computed second-order hyperpolarizability for all the compounds are found to be higher than that of the reported values of urea (0.68×10^{-36} e.s.u.) by several folds. Also, all the compounds show higher values of the mean polarizability, first as well as the second-order hyperpolarizability in DMF than in EA in all the three functionals. For α values, MAE in the method BHandHLYP/6-311++G(d,p) shows very low value (0.24 in DMF and 0.38 in EA which indicates that this functional is able to show accurate results with the average value (Table 3.6 and Figure 3.9). Also, for β_0 and γ wB97X functional and 6-311++G(d,p) show low values of deviation which indicates that these values are closer the average values (Tables 3.7, 8 and Figures 3.10, 11).

Table 3.6: Linear polarizability and MAE of all compounds in DMF and EA.

α		1a		1b		1c		MAE	
Functional	Basis set	DMF	EA	DMF	EA	DMF	EA	DMF	EA
B3LYP	6-311++G(d,p)	64.59	60.12	74.37	69.22	63.25	58.18	6.18	5.18
	cc-pVDZ	57.45	54.01	65.76	61.86	53.569	51.55	2.29	1.52
	cc-pVTZ	62.73	58.55	71.97	67.17	62.25	59.25	4.43	4.33
BHandHLYP	6-311++G(d,p)	60.09	56.02	69.16	64.55	53.2147	50.25	0.39	0.38
	cc-pVDZ	53.64	50.46	61.35	57.85	49.678	45.63	6.33	6.01
	cc-pVTZ	58.69	54.82	67.25	62.94	52.9647	49.41	1.59	1.6
CAM B3LYP	6-311++G(d,p)	61.68	57.41	70.11	65.4	58.678	54.55	3.29	3.25
	cc-pVDZ	54.81	51.51	61.95	58.4	51.648	48.25	4.06	3.15
	cc-pVTZ	60.07	56.04	68.05	63.65	56.8475	52.38	1.46	1.48
wB97	6-311++G(d,p)	60.13	55.9	69.25	63.16	57.647	53.76	2.15	1.73
	cc-pVDZ	53.39	50.13	67.16	56.44	51.468	47.62	2.86	4.48
	cc-pVTZ	58.71	54.72	68.2	61.71	55.847	52.53	0.72	0.45
wB97X	6-311++G(d,p)	60.73	56.47	68.56	63.95	57.647	54.38	2.12	2.39
	cc-pVDZ	53.91	50.63	60.57	57.12	51.649	48.16	4.82	3.9
	cc-pVTZ	59.26	55.24	66.69	62.4	57.745	53.07	1.04	1.03
wB97XD	6-311++G(d,p)	61.77	57.46	70.06	60.25	58.478	55.4	3.24	1.83
	cc-pVDZ	54.95	51.62	62.04	58.46	52.28	49.19	3.77	2.78
	cc-pVTZ	60.32	56.24	68.19	63.74	56.55	54.098	1.49	2.15

3.1.7.3 Comparison of pyrazole derivatives with reported compounds

We compared the NLO properties of pyrazole derivatives with the reported [68] compounds 5CWG and nine CWG. All these three derivatives shows higher values of first-order hyperpolarizability compared with the reported compounds.

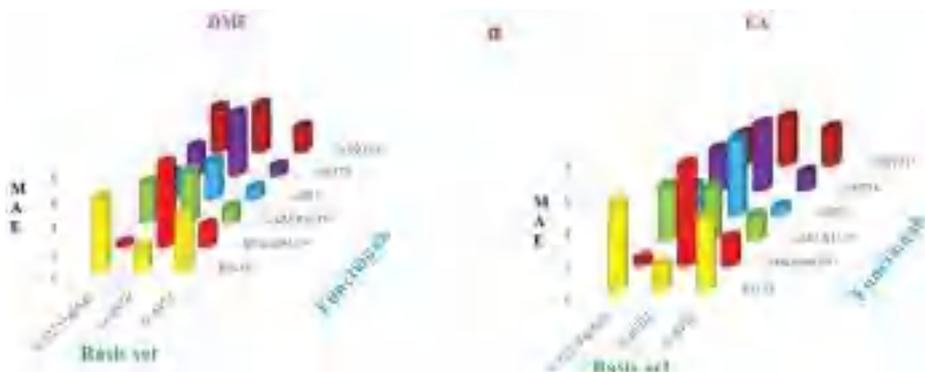


Figure 3.9: MAE of linear polarizability of compounds 1a, 1b and 1c in DMF and EA.

Table 3.7: First-order hyperpolarizability and MAE of all the compounds in DMF and EA.

β	Functional	Basis set	1a		1b		1c		MAE	
			DMF	EA	DMF	EA	DMF	EA	DMF	EA
B3LYP	6-311++G(d,p)	6-311++G(d,p)	68.61	58.31	101.13	85.73	63.45	60.53	15.66	13.29
		cc-pVDZ	63.86	54.68	86.01	73.74	55.86	53.54	6.5	5.75
		cc-pVTZ	65.83	56.29	96.36	81.77	62.84	59.25	12.94	10.87
BHandHLYP	6-311++G(d,p)	6-311++G(d,p)	46.21	39.29	58.59	50.42	53.54	50.25	9.29	8.24
		cc-pVDZ	42.35	36.56	52.16	45.37	47.289	45.63	14.81	12.38
		cc-pVTZ	44.22	37.97	57.36	49.41	51.689	49.41	10.98	9.3
CAM B3LYP	6-311++G(d,p)	6-311++G(d,p)	55.23	46.96	62.77	53.67	57.98	54.55	9.31	8.46
		cc-pVDZ	50.59	43.55	54.9	47.48	50.745	48.25	2.73	3.16
		cc-pVTZ	52.94	45.39	61.19	52.38	56.95	52.38	7.67	6.79
wB97	6-311++G(d,p)	6-311++G(d,p)	51.99	43.89	47.25	40.87	39.84	37.05	2.99	2.66
		cc-pVDZ	47.39	40.54	41.25	36.67	36.542	33.98	7.62	6.2
		cc-pVTZ	49.56	42.11	46.7	40.09	39.623	36.56	4.06	3.68
wB97X	6-311++G(d,p)	6-311++G(d,p)	52.53	44.42	50.93	43.6	42.578	38.59	0.67	1.06
		cc-pVDZ	47.82	40.99	45.3	39.29	38.514	35.49	5.48	4.67
		cc-pVTZ	50.37	42.87	50.07	42.96	40.27	38.26	2.45	1.9
wB97XD	6-311++G(d,p)	6-311++G(d,p)	54.45	46.15	58.28	45.25	44.712	41.97	3.13	1.19
		cc-pVDZ	49.58	42.57	51.87	44.87	41.568	38.52	1.68	1.28
		cc-pVTZ	52.57	44.81	57.26	48.98	44.57	41.57	2.11	1.86

Table 3.8: Second-order hyperpolarizability and MAE of all compounds in DMF and EA.

Functional	Basis set	1a		1b		1c		MAE	
		DMF	EA	DMF	EA	DMF	EA	DMF	EA
B3LYP	6-311++G(d,p)	233.72	204.75	544.86	439.01	260.37	256.11	96.82	82.13
	cc-pVDZ	173.67	157.09	412.25	337.88	198.64	191.38	12.03	10.96
	cc-pVTZ	190.89	169.87	469.03	378.74	258.964	240.25	56.79	45.13
BHandHLYP	6-311++G(d,p)	141.85	123.29	288.25	235.68	220.647	210.36	32.58	28.05
	cc-pVDZ	97.69	88.18	213.66	177.92	215.634	204.95	73.84	60.81
	cc-pVTZ	110.97	98.32	241.87	198.49	217.95	208.64	59.23	49.35
CAM B3LYP	6-311++G(d,p)	145.37	128.69	290.86	236.66	184.554	170.25	56.99	50.51
	cc-pVDZ	98.88	91.78	213.44	177.09	177.954	165.34	13.49	16.71
	cc-pVTZ	112.34	102.15	243.38	198.82	173.584	160.24	26.49	25.71
wB97	6-311++G(d,p)	112.46	100.42	175.36	164.68	129.24	116.29	10.92	0.89
	cc-pVDZ	73.57	70.01	172.67	121.08	89.847	82.22	37.91	36.93
	cc-pVTZ	84.58	78.49	165.44	135.37	101.23	92.79	32.85	25.81
wB97X	6-311++G(d,p)	121.32	107.82	222.47	180.88	142.35	125.35	12.11	9.99
	cc-pVDZ	79.35	74.92	160.2	133.14	109.25	88.45	33.67	29.19
	cc-pVTZ	91.31	84.17	181.66	148.43	115.24	99.81	20.53	17.22
wB97XD	6-311++G(d,p)	139.22	122.95	265.77	160.25	160.247	143.67	38.48	14.26
	cc-pVDZ	93.91	87.44	196.36	162.56	120.24	103.56	13.1	10.17
	cc-pVTZ	107.05	97.51	220.79	179.69	126.24	116.04	1.42	3.05

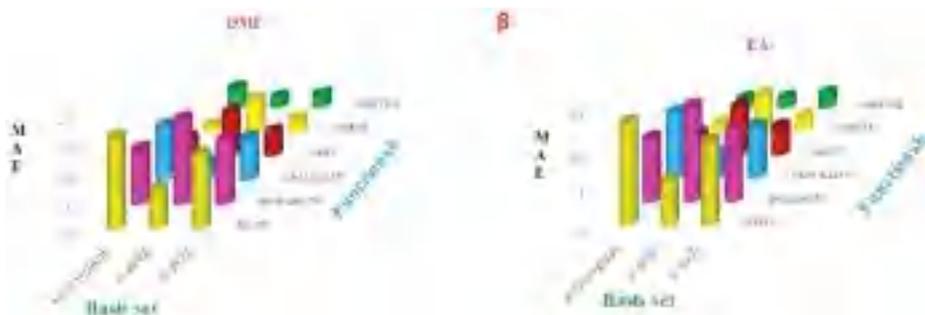


Figure 3.10: MAE of first-order hyperpolarizability of 1a, 1b and 1c in DMF and EA.

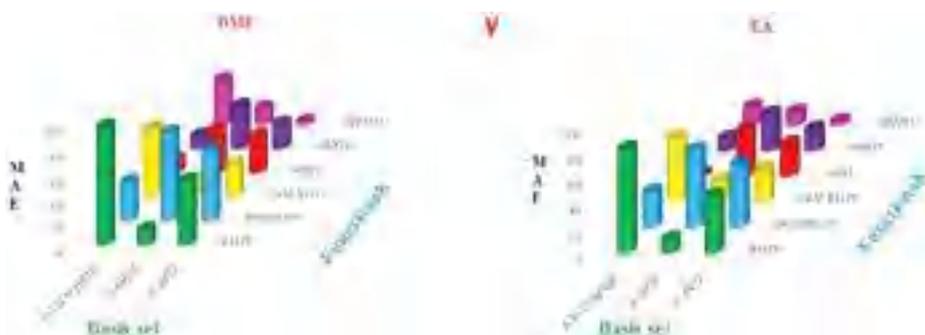


Figure 3.11: MAE of second-order hyperpolarizability of 1a, 1b and 1c in DMF and EA.

3.1.8 Fundamental limits of hyperpolarizability

The limit of first-order hyperpolarizability can be obtained from the equation [69–71]

$$\beta_{2L}^{*SR} < \sqrt{\frac{4}{3}} \left(\frac{n^2 + 2}{3} \right)^3 \left(\left[\frac{eh}{\sqrt{m}} \right] \right)^3 \left(\frac{N^3}{E_{10}^7} \right) \quad (3.12)$$

where m and e are the mass and charge of the electron, h is Plank's constant, E_{10} is the transition energy between the ground and excited states, N is the number of electrons, and n is the refractive index.

The limit for the second-order hyperpolarizability is given by the equation,

$$-\frac{e^4 h^2}{m^2} \left(\frac{N^2}{E_{10}^5} \right) < \gamma < 4 \frac{e^4 h^4}{m^2} \left(\frac{N^2}{E_{10}^5} \right) \quad (3.13)$$

By using the above two equations, we have calculated the limits of β and γ and are tabulated in Table 3.9. The values of β are less than γ .

Table 3.9: Limits of first- and second-order hyperpolarizability.

Comp	Solvent	β limit (10^{-30} esu)	γ min (10^{-36} esu)	γ max (10^{-36} esu)
1a	DMF	9596.7	31978.1	127094
	EA	7107.3	37090.7	148363
1b	DMF	11995.2	53273.6	213094
	EA	9769.8	65586.6	262346
1c	DMF	10693.1	31978.1	127912
	EA	8617.9	37090.7	148363

3.1.9 Intrinsic hyperpolarizability

Intrinsic hyperpolarizability is the ratio of first- and second-order hyperpolarizability with the fundamental limits of hyperpolarizability. NLO response of compounds can be compared with intrinsic first- and second-order hyperpolarizability values by using the equation [71–73].

$$\beta_0^{\text{int}} = \frac{\beta_0}{\beta_0^{\text{max}}} \quad (3.14)$$

where β_0^{int} is calculated using Equation 3.11

Also, the second-order intrinsic hyperpolarizability is calculated by using the equation,

$$\gamma_0^{\text{int}} = \frac{\gamma_0}{\gamma_0^{\text{max}}} \quad (3.15)$$

Table S3 indicates the values for intrinsic hyperpolarizability for both first- and second-order in the solvents, DMF and EA. Figure 3.2 represents a plot of (a) intrinsic first order hyperpolarizability with a maximum limit of β and (b) intrinsic second-order hyperpolarizability with a maximum limit of γ .

3.2 Conclusions

Here we have calculated the electronic spectral behavior of different disperse dye derived from 3-oxo-2,3 dihydrobenzothiophene 1,dioxide. A relative study of different DFT functionals different basis sets were also benchmarked to take out the best combination of functional and basis set. The results revealed that best and reliable functionals and basis sets which provided close correlation with the experimental findings. For all compounds are in good agreement with vertical excitations obtained with the method B3LYP/6–311++G(d,p). Different functionals like GH and RSH and basis sets

6-311++G(d,p), cc-pVDZ and cc-pVTZ were used for the calculations of vertical excitations, dipole moments and NLO properties. NLO properties like μ , α , β and γ were calculated by using all above functionals and basis sets and also calculated their MAE values to benchmark the functional for the NLO computations. Our findings will be helpful for the selection of basis sets and functionals for DFT and TDDFT study. The higher values for first-order hyperpolarizability and static second-order hyperpolarizability show a large nonlinear optical property and this is mainly due to the strong donor- π -acceptor conjugation. Compound 1c have low band gap indicating the better charge transfer which may be due to donating group have methyl and Cl substituents.

Acknowledgment: Author SBB and AAB are thankful for University Grants Commission (UGC) New Delhi, India for financial support as a junior and senior research fellowship.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Körzdörfer T, Brédas J-L. Organic electronic materials: recent advances in the DFT description of the ground and excited states using tuned range-separated hybrid functionals. [Internet]. *Acc Chem Res* 2014;47:3284–91. Available from: <http://pubs.acs.org/doi/10.1021/ar500021t>.
2. Barone V. *Structure, magnetic properties and reactivities of open-shell species from density functional and self-consistent hybrid methods*. [Internet]; 1995:287–334. Available from: http://www.worldscientific.com/doi/abs/10.1142/9789812830586_0008.
3. Verma D. Density Functional Theory (DFT) as a powerful tool for designing corrosion inhibitors in aqueous phase [Internet] 2018:87–104. Available from: <https://www.intechopen.com/doi.org/10.5772/intechopen.78333>.
4. Boyd RJ, Wang J, Eriksson LA. *The electron density as calculated from density functional theory*. [Internet]; 1995:369–401. Available from: http://www.worldscientific.com/doi/abs/10.1142/9789812830586_0010.
5. Zara Z, Iqbal J, Ayub K, Irfan M, Mahmood A, Khera RA, et al.. A comparative study of DFT calculated and experimental UV–vis spectra for thirty carboline and carbazole based compounds. [Internet]. *J Mol Struct* 2017;1149:282–98. <https://doi.org/10.1016/j.molstruc.2017.07.093>.
6. Casida ME. *Time-dependent density functional response theory for molecules*. [Internet]; 1995:155–92. Available from: http://www.worldscientific.com/doi/abs/10.1142/9789812830586_0005.
7. Chermette H. Chemical reactivity indexes in density functional theory. [Internet]. *J Comput Chem* 1999;20:129–54. Available from: <http://doi.wiley.com/10.1002/%28SICI%291096-987X%2819990115%2920%3A1%3C129%3A%3AAID-JCC13%3E3.0.CO%3B2-A>.

8. Chai J-D, Head-Gordon M. Systematic optimization of long-range corrected hybrid density functionals. [Internet]. *J Chem Phys* 2008;128:84106. Available from: <http://aip.scitation.org/doi/10.1063/1.2834918>.
9. Kohn W, Becke AD, Parr RG. Density functional theory of electronic structure. [Internet]. *J Phys Chem* 1996;100:12974–80. Available from: <https://pubs.acs.org/doi/10.1021/jp960669l>.
10. Tsang TY, Osbrn TA. The Spectral Properties of Time Delay Department of Physics. Brooklyn college of the city University of New York, NY 11210. *Nucl Phys* 1975;247:43–50. [https://doi.org/10.1016/0375-9474\(75\)90275-4](https://doi.org/10.1016/0375-9474(75)90275-4).
11. Mohbiya DR, Sekar N. Electronic structure and spectral properties of indole based fluorescent styryl dyes: comprehensive study on linear and non-linear optical properties by DFT/TDDFT method. [Internet]. *Comput Theor Chem* 2018;1139:90–101. <https://doi.org/10.1016/j.comptc.2018.07.015>.
12. Geerlings P, De Proft F, Langenaeker W. Conceptual density functional theory. [Internet]. *Chem Rev* 2003;103:1793–874. Available from: <https://pubs.acs.org/doi/10.1021/cr990029p>.
13. Parr RG, Szentpály LV, Liu S. Electrophilicity index. [Internet]. *J Am Chem Soc* 1999;121:1922–4. Available from: <https://pubs.acs.org/doi/10.1021/ja983494x>.
14. Rao VP, Jen AKY, Wong KY, Drost KJ. Dramatically enhanced second-order nonlinear optical susceptibilities in tricyanovinylthiophene derivatives. *J Chem Soc Chem Commun* 1993;14:1118–20. <https://doi.org/10.1039/C39930001118>.
15. Miniewicz A, Palewska K, Sznitko L, Lipinski J. Single- and two-photon excited fluorescence in organic nonlinear optical single crystal 3-(1,1-Dicyanoethenyl)-1-phenyl-4,5-dihydro-1H-pyrazole. [Internet]. *J Phys Chem* 2011;115:10689–97. Available from: <http://pubs.acs.org/doi/abs/10.1021/jp204435s>.
16. Jen AK, Rao P, Drost KJ, Michael YW. Functionalized fused thiophenes: a new class of thermally stable and efficient second-order nonlinear optical chromophores. *Chem Mater* 1994;6:2210–12. <https://doi.org/10.1021/cm00048a004>.
17. Drost KJ, Rao VP, Jen AKY. Theoretical and experimental studies of the molecular second order nonlinear optical responses of heteroaromatic compounds. *J Chem Phys* 1994;100:6818. <https://doi.org/10.1063/1.467041>.
18. Steybe F, Effenberger F, Gubler U, Bosshard C, Günter P. Highly polarizable chromophores for nonlinear optics: syntheses, structures and properties of donor-acceptor substituted thiophenes and oligothiophenes. *Tetrahedron* 1998;54:8469–80. [https://doi.org/10.1016/S0040-4020\(98\)00450-5](https://doi.org/10.1016/S0040-4020(98)00450-5).
19. Jen AK, Rao VP, Wong KY, Drost KJ. Functionalized thiophenes: second-order nonlinear optical materials 2000;90:90–2. <https://doi.org/10.1039/C39930000090>.
20. Wong KY, Drost KJ. Novel push-pull thiophenes for second order nonlinear optical applications 2000;34:1747–50. [https://doi.org/10.1016/S0040-4039\(00\)60768-2](https://doi.org/10.1016/S0040-4039(00)60768-2).
21. Wong KY, Jen AKY, Pushkara Rao V, Drost KJ. Theoretical and experimental studies of the molecular second order nonlinear optical responses of heteroaromatic compounds. *J Chem Phys* 1994;100:6818–25. <https://doi.org/10.1063/1.467041>.
22. Cidália MR, Castro R, Maurício AA, Fonseca C, Belsley M, Manuela M, et al. Highly efficient and thermally stable NLO organic materials based on pyrrole and thiophene heterocycles. Available from: https://www.researchgate.net/journal/0277-786X_Proceedings_of_SPIE-The_International_Society_for_Optical_Engineering 26 July 2011; <https://doi.org/10.1117/12.892205>.
23. Chandrakantha B, Isloor AM, Sridharan K, Philip R, Shetty P, Padaki M. Novel N-substituted-5-phenyl-1H-pyrazole-4-ethyl carboxylates as potential NLO materials. *Arab J Chem* 2013;6:97–102. <https://doi.org/10.1016/j.arabjc.2010.09.024>.

24. Makowska-Janusik M, Kajzar F, Miniewicz A, Mydlova L, Rau I. First principle calculations of the electronic and vibrational properties of the 3-(1,1-Dicyanoethenyl)-1-phenyl-4,5-dihydro-1H-pyrazole molecule. [Internet]. *J Phys Chem* 2015;119:1347–58. Available from: <https://pubs.acs.org/doi/10.1021/jp510102q>.
25. Bureš F. Fundamental aspects of property tuning in push–pull molecules. [Internet]. *RSC Adv* 2014;4:58826–51. Available from: <http://xlink.rsc.org/?DOI=C4RA11264D>.
26. Chen S, Li Y, Yang W, Chen N, Liu H, Li Y. Synthesis and tuning optical nonlinear properties of molecular crystals of benzothiadiazole. [Internet]. *J Phys Chem C* 2010;114:15109–15. Available from: <https://pubs.acs.org/doi/10.1021/jp103159b>.
27. Liyanage PS, de Silva RM, de Silva KMN. Nonlinear optical (NLO) properties of novel organometallic complexes: high accuracy density functional theory (DFT) calculations. [Internet]. *J Mol Struct THEOCHEM* 2003;639:195–201. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0166128003006377>.
28. Facchetti A, Abbotto A, Beverina L, van der Boom ME, Dutta P, Evmenenko G, et al. Azinium–(π -Bridge)–pyrrole NLO-phores: influence of heterocycle acceptors on chromophoric and self-assembled thin-film properties #. [Internet]. *Chem Mater* 2002;14:4996–5005. Available from: <https://pubs.acs.org/doi/10.1021/cm0205635>.
29. Kwon O-P, Ruiz B, Choubey A, Mutter L, Schneider A, Jazbinsek M, et al. Organic nonlinear optical crystals based on configurationally locked polyene for melt growth. [Internet]. *Chem Mater* 2006;18:4049–54. Available from: <https://pubs.acs.org/doi/10.1021/cm0610130>.
30. Oudar JL. Optical nonlinearities of conjugated molecules. Stilbene derivatives and highly polar aromatic compounds. [Internet]. *J Chem Phys* 1977;67:446–57. Available from: <http://aip.scitation.org/doi/10.1063/1.434888>.
31. Shao J, Ji S, Li X, Zhao J, Zhou F, Guo H. Thiophene-inserted aryl-dicyanovinyl compounds: the second generation of fluorescent molecular rotors with significantly redshifted emission and large Stokes shift. [Internet]. *Eur J Org Chem* 2011;2011:6100–9. Available from: <http://doi.wiley.com/10.1002/ejoc.201100891>.
32. Scholz N, Jadhav A, Shreykar M, Behnke T, Nirmalanathan N, Resch-Genger U, et al. Coumarin-rhodamine hybrids—novel probes for the optical measurement of viscosity and polarity. [Internet]. *J Fluoresc* 2017;27:1949–56. Available from: <http://link.springer.com/10.1007/s10895-017-2165-4>.
33. Haidekker MA, Brady TP, Lichlyter D, Theodorakis EA. Effects of solvent polarity and solvent viscosity on the fluorescent properties of molecular rotors and related probes. [Internet]. *Bioorg Chem* 2005;33:415–25. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0045206805000763>.
34. Paul A, Samanta A. Free volume dependence of the internal rotation of a molecular rotor probe in room temperature ionic liquids. [Internet]. *J Phys Chem B* 2008;112:16626–32. Available from: <https://pubs.acs.org/doi/10.1021/jp8060575>.
35. Shreykar MR, Sekar N. NLOphoric donor-rigidified ESIPT dyes – synthesis, pH study, solvatochromism and DFT insights. [Internet]. *J Lumin* 2017;192:343–58. <https://doi.org/10.1016/j.jlumin.2017.06.061>.
36. Haidekker MA, Theodorakis EA. Environment-sensitive behavior of fluorescent molecular rotors. [Internet]. *J Biol Eng* 2010;4:11. Available from: <http://jbioleng.biomedcentral.com/articles/10.1186/1754-1611-4-11>.
37. Jin Y-J, Park H, Ohk Y-J, Kwak G. Hydrodynamic fluorescence emission behavior of molecular rotor-based vinyl polymers used as viscosity sensors. [Internet]. *Polymer* 2017;132:79–87. <https://doi.org/10.1016/j.polymer.2017.10.054>.
38. Gao Z, Wang Z, Shan T, Liu Y, Shen F, Pan Y, et al. High-efficiency deep blue fluorescent emitters based on phenanthro[9,10-d]imidazole substituted carbazole and their applications

- in organic light emitting diodes. [Internet]. *Org Electron* 2014;15:2667–76. <https://doi.org/10.1016/j.orgel.2014.07.019>.
39. Pander P, Motyka R, Zassowski P, Etherington MK, Varsano D, da Silva TJ, et al. Thermally activated delayed fluorescence mediated through the upper triplet state manifold in non-charge-transfer star-shaped triphenylamine–carbazole molecules. [Internet]. *J Phys Chem C* 2018;122:23934–42. Available from: <http://pubs.acs.org/doi/10.1021/acs.jpcc.8b07610>.
 40. Uoyama H, Goushi K, Shizu K, Nomura H, Adachi C. Highly efficient organic light-emitting diodes from delayed fluorescence. [Internet]. *Nature* 2012;492:234–8. <https://doi.org/10.1038/nature11687>.
 41. Seniutinas G, Tomašiūnas R, Czaplicki R, Sahraoui B, Daškevičienė M, Getautis V, et al. Arylmethylene-1,3-indandione based molecular glasses: third order optical non-linearity. [Internet]. *Dye Pigment* 2012;95:33–40. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0143720812000678>.
 42. Hauck M, Stolte M, Schönhaber J, Kuball H, Müller TJJ. Synthesis, electronic, and electro-optical properties of emissive solvatochromic phenothiazinyl merocyanine dyes. [Internet]. *Chem Eur J* 2011;17:9984–98. Available from: <http://doi.wiley.com/10.1002/chem.201100592>.
 43. Haenle JC, Bruchlos K, Ludwigs S, Köhn A, Laschat S. Rigidified push-pull dyes: using chromophore size, donor, and acceptor units to tune the ground state between neutral and the cyanine limit. [Internet]. *Chempluschem* 2017;82:1197–210. Available from: <http://doi.wiley.com/10.1002/cplu.201700347>.
 44. Ruiz Delgado MC, Casado J, Hernández V, Lopez Navarrete JT, Orduna J, Villacampa B, et al. Electronic, optical, and vibrational properties of bridged dithienylethylene-based NLO chromophores. [Internet]. *J Phys Chem C* 2008;112:3109–20. Available from: <https://pubs.acs.org/doi/10.1021/jp710459c>.
 45. Lawrentz U, Grahn W, Lukaszuk K, Klein C, Wortmann R, Feldner A, et al. Donor-acceptor oligoenes with a locked all-trans conformation: synthesis and linear and nonlinear optical properties. [Internet]. *Chem Eur J* 2002;8:1573–90. Available from: [http://doi.wiley.com/10.1002/1521-3765\(200204\)8:1573::AID-CHEM1573\(3E\)3.0.CO;2B](http://doi.wiley.com/10.1002/1521-3765(200204)8:1573::AID-CHEM1573(3E)3.0.CO;2B).
 46. Kumari R, Varghese A, George L, KBA. Photophysical study of 6-amino-3-methyl-4-(4-nitrophenyl)-1,4-dihydropyrano[2,3-c]pyrazole-5-carbonitrile and estimation of ground-state and singlet excited-state dipole moments by solvatochromic approaches. [Internet]. *J Mol Liq* 2016;222:828–35. <https://doi.org/10.1016/j.molliq.2016.07.133>.
 47. Xia J Bin Li FY, Yang H, Li XH, Huang CH. A novel quasi-solid-state dye-sensitized solar cell based on monolayer capped nanoparticles framework materials. *J Mater Sci* 2007;42:6412–6. <https://doi.org/10.1007/s10853-006-1184-3>.
 48. Chen X-L, Yu R, Zhang Q-K, Zhou LJ, Wu XY, Zhang Q, et al. Rational design of strongly blue-emitting cuprous complexes with thermally activated delayed fluorescence and application in solution-processed OLEDs. *Chem Mater* [Internet] 2013 [cited 2018];25:3910–20. Available from: <https://pubs.acs.org/doi/pdf/10.1021/cm4024309>.
 49. Lanke SK, Sekar N. Pyrazole based solid state emissive NLOphores with TICT characteristics: synthesis, DFT and TDDFT studies. [Internet]. *Dye Pigment* 2016;126:62–75. <https://doi.org/10.1016/j.dyepig.2015.11.014>.
 50. Szukalski A, Sznitko L, Cypriach K, Miniewicz A, Mysliwiec J. Light amplification in derivatives of pyrazoline-based systems. [Internet]. *J Phys Chem C* 2014;118:8102–10. Available from: <http://pubs.acs.org/doi/10.1021/jp411031b>.
 51. Barberá J, Clays K, Giménez R, Houbrechts S, Persoons A, Serrano JL. Versatile optical materials: fluorescence, non-linear optical and mesogenic properties of selected 2-pyrazoline derivatives. [Internet]. *J Mater Chem* 1998;8:1725–30. <https://doi.org/10.1039/A802070A>.

52. Moylan CR, Twieg RJ, Lee VY, Swanson SA, Betterton KM, Miller RD. Nonlinear optical chromophores with large hyperpolarizabilities and enhanced thermal stabilities. [Internet]. *J Am Chem Soc* 1993;115:12599–600. Available from: <http://pubs.acs.org/doi/abs/10.1021/ja00079a055>.
53. Fang W, Shao C-S, Cheng W-D, Tang H-Y, Zheng W-C. Theoretical investigation of optical and paramagnetic resonance spectra of $[\text{NiF}_6]^{4-}$ clusters with orthorhombic symmetry. [Internet]. *Spectrochim Acta Part A Mol Biomol Spectrosc* 2013;115:483–7. <https://doi.org/10.1016/j.saa.2013.06.074>.
54. Hu ZY, Fort A, Barzoukas M, Jen AKY, Barlow S, Marder SR. Trends in optical nonlinearity and thermal stability in electrooptic chromophores based upon the 3-(dicyanomethylene)-2,3-dihydrobenzothiophene-1,1-dioxide acceptor. *J Phys Chem B* 2004;108:8626–30. <https://doi.org/10.1021/jp036728u>.
55. Singh Bhatti H, Seshadri S. Novel disperse dyes from benzo[b]thiophene-3(2H)-one-1,1-dioxide and ethyl benzo[b]thien-3(2H)-ylideneacyanoacetate,S,S-dioxide: synthesis and properties. [Internet]. *Dye Pigment* 2004;62:83–92. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0143720803002365>.
56. Shenoy VU, Patel VP, Seshadri S. Disperse dyes derived from 3-Oxo-2,3-dihydrobenzo[b]thiophene-1,1-dioxide and 3[dicyanomethylene-2,3-dihydrobenzo[b]thiophene-1,1-dioxide. *Dye Pigment* 1989;11:37–46.
57. Yanai T, Tew DP, Handy NC. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem Phys Lett* 2004;393:51–7. <https://doi.org/10.1016/j.cplett.2004.06.011>.
58. Li R, Zheng J, Truhlar DG. Density functional approximations for charge transfer excitations with intermediate spatial overlap. *Phys Chem Chem Phys* 2010;12:12697–701. <https://doi.org/10.1039/c0cp00549e>.
59. Chai J-D, Head-Gordon M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. [Internet]. *Phys Chem Chem Phys* 2008;10:6615. Available from: <http://xlink.rsc.org/?DOI=b810189b>.
60. Lanke SK, Sekar N. Coumarin push-pull NLOphores with red emission: solvatochromic and theoretical approach. *J Fluoresc* 2016;26:949–62. <https://doi.org/10.1007/s10895-016-1783-6>.
61. Carthigayan K, Xavier S, Periandy S. HOMO-LUMO UV, NLO NMR and vibrational analysis of 3-methyl-1-phenylpyrazole using FT-IR, FT-Raman FT-NMR spectra and HF-DFT computational methods. [Internet]. *Spectrochim Acta Part A Mol Biomol Spectrosc* 2015;142:350–63. <https://doi.org/10.1016/j.saa.2014.11.011>.
62. Islam N, Ghosh DC. On the electrophilic character of molecules through its relation with electronegativity and chemical hardness. [Internet]. *Int J Mol Sci* 2012;13:2160–75. Available from: <http://www.mdpi.com/1422-0067/13/2/2160>.
63. Figueredo S, Páez M, Torres F. The electrophilic descriptor. [Internet]. *Comput Theor Chem* 2019;1157:34–9. <https://doi.org/10.1016/j.comptc.2019.04.011>.
64. Gutiérrez-Oliva S, Jaque P, Toro-Labbé A. Using Sanderson's principle to estimate global electronic properties and bond energies of hydrogen-bonded complexes. [Internet]. *J Phys Chem* 2000;104:8955–64. Available from: <https://pubs.acs.org/doi/10.1021/jp000777e>.
65. Fuentealba P, Parr RG. Higher-order derivatives in density-functional theory, especially the hardness derivative $\partial\eta/\partial N$. [Internet]. *J Chem Phys* 1991;94:5559–64. Available from: <http://aip.scitation.org/doi/10.1063/1.460491>.
66. Morell C, Grand A, Toro-Labbé A, Chermette H. Is hyper-hardness more chemically relevant than expected?. [Internet]. *J Mol Model* 2013;19:2893–900. Available from: <http://link.springer.com/10.1007/s00894-013-1778-z>.

67. Sharma R, Joshi S, Bhattacharjee R, Pant DD. Solvent effect on absorption and fluorescence spectra of cinchonine and cinchonidine dications: estimation of ground and excited state dipole moments by experimental and numerical studies. [Internet]. *J Mol Liq* 2015;206:159–64. <https://doi.org/10.1016/j.molliq.2015.02.002>.
68. Ghanavatkar CW, Mishra VR, Sekar N, Methew E, Thomas E, Joe I. Benzothiazole pyrazole containing emissive azo dyes decorated with ESIPT core: linear and non linear optical properties, Z scan, optical limiting, laser damage threshold with comparative DFT studies. *J Mol Str* 2019;1203:127401. <https://doi.org/10.1016/j.saa.2014.08.111>.
69. Lytel R, Mossman S, Crowell E, Kuzyk MG. Exact fundamental limits of the first and second hyperpolarizabilities. [Internet]. *Phys Rev Lett* 2017;119:73902. Available from: <https://link.aps.org/doi/10.1103/PhysRevLett.119.073902>.
70. Kuzyk MG. Fundamental limits of all nonlinear-optical phenomena that are representable by a second-order nonlinear susceptibility. [Internet]. *J Chem Phys* 2006;125:154108. Available from: <http://aip.scitation.org/doi/10.1063/1.2358973>.
71. Kuzyk MG, Watkins DS. The effects of geometry on the hyperpolarizability. [Internet]. *J Chem Phys* 2006;124:244104. Available from: <http://aip.scitation.org/doi/10.1063/1.2205859>.
72. Pérez-Moreno J, Zhao Y, Clays K, Kuzyk MG. Modulated conjugation as a means for attaining a record high intrinsic hyperpolarizability. [Internet]. *Opt Lett* 2007;32:59. Available from: <https://www.osapublishing.org/abstract.cfm?URI=ol-32-1-59>.
73. Kuzyk MG, Pérez-Moreno J, Shafei S. Sum rules and scaling in nonlinear optics. [Internet]. *Phys Rep* 2013;529:297–398. <https://doi.org/10.1016/j.physrep.2013.04.002>.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/psr-2019-0129>).

Academic Year
2020-21
Supporting Documents

Contents	Page No
The State-of-Art Machine Learning In Bioinformatics Mujtaba A.Karim Shaikh	1
Graphical Password Authentication System Mrs.Rajeshree Mundhe, Meghna Nikam, Jagruti Rokade	6
Survey on the Usages of the Internet Ms. Fatema Shaikh, Ms. Priyanka Patil	10
Data Mining Techniques and Applications Dr. Pradhnya M Wankhade, Mrs. Shraddha Prasad Kok	12
Cybercrime and its Effects Parvati Rijal, Praseena Biju	15
Virtual Assistant: A Product of Machine Learning Shreya Chandane, Prof. Deepesh Jagdale	19
Artificial intelligence in e-commerce Shreya Chandane, Prof. Deepesh Jagdale	28
SMART BLIND STICK USING ARDUINO Neha Patil, Prof. Deepesh Jagdale	35
A Study of Robotic Process Automation Use Cases Today for Tomorrow's Business Neha Patil, Prof. Deepesh Jagdale	39
Current Automation Trends- A research on RPA development tool Rashmita Rai, Prof. Deepesh Jagdale	46
Application of RPA for Email Automation Rashmita Rai, Prof. Deepesh Jagdale	49
Data Analysis: A Common review Ajin shaji antoney, Prof. Deepesh Jagdale	54
Data Mining : Use case Review	57

Ajin shaji antoney, Prof. Deepesh Jagdale	
Voice Based Intelligent Virtual Assistance For Windows Alisha Bhoir, Prof. Deepesh Jagdale	61
Voice Based Intelligent Virtual Assistance For Windows Alisha Bhoir, Prof. Deepesh Jagdale	66
WEBSITE EVALUATION USING OPINION MINING Jitesh Kirdakude, Prof. Priyanka Sorte	71
CHEST X-RAY IN NEW CORONAVIRUS DISEASE 2019 (COVID-19) INFECTION: FINDINGS AND CORRELATION WITH CLINICAL OUTCOME Ms. Payal Jingare, Prof. Priyanka Sorte	75
AN ARTIFICIAL INTELLIGENCE SYSTEM FOR PREDICTING THE DETERIORATION OF COVID-19 PATIENTS IN THE EMERGENCY DEPARTMENT Ms. Payal Jingare, Prof. Priyanka Sorte	81
Serverless cloud computing Ms. Hemangi Mali, Ms. Priyanka Sorte	88
A Survey of Machine Learning For Big Data Processing Ms. Hemangi Mali, Ms. Priyanka Sorte	92
Implementation of Website Builder Automation Vishal Lagoo, Priyanka Sorte	96
Survey on Website Builder Automations Vishal Lagoo, Priyanka Sorte	99
FAKE PRODUCT REVIEW MONITORING SYSTEM Jitesh Kirdakude, Prof. Priyanka Sorte	101
A Study of Efficient and Secure Data Storage in Mobile Cloud Computing. Neha Mahadik, Prof. Priyanka Sorte	105
A Study of Data Warehousing Neha Mahadik, Prof. Priyanka Sorte	109
USE OF AUTOMATION (RPA) IN SHOPPING WEBSITE	112

Mukeshkumar Mishra, Prof. Priyanka Sonawane	
IMPLEMENTING TECHNIQUES OF AUTOMATION (RPA) IN SHOPPING WEBSITE Mukeshkumar Mishra, Prof. Priyanka Sonawane	119
Thyroid Disease Prediction Using Data Mining Techniques Namrata Kedari, Prof Priyanka Sonawane	127
Heart Disease Prediction System Namrata Kedari, Prof Priyanka Sonawane	131
Bypass Fraud Detection:Artificial Intelligence Approach Aftab Ibrahim Panhalkar, Priyanka Sonawane	137
Application of Artificial intelligence in Diabetic Retinopathy Aftab Ibrahim Panhalkar, Priyanka Sonawane	141
ROLE OF ARTIFICIAL INTELLIGENCE IN COVID - 19 JENCY KERLIN, Priyanka Sonawane	146
ATTENDANCE MANAGEMENT USING FACE RECOGNITION. JENCY KERLIN, Priyanka Sonawane	158
STUDENT SENTIMENT ANALYSIS Vaishnavi Nair, Prof. Priyanka Sonawane	166
CYBERSECURITY USING ARTIFICIAL INTELLIGENC Vaishnavi Nair, Prof. Priyanka Sonawane	172
Artificial Inteligence In Power Station Hari Prakash Mishra Prof, Anita Mhatre	177
Artificial Intelligence for 5G Wireless Systems:Opportunities, Challenges, and Future Research Directions MAYUR GANPAT SAKPAL, Prof. Anita Mhatre	180
Concept Artificial Intelligence, its impact And Emerging Trends MAYUR GANPAT SAKPAL, Prof. Anita Mhatre	192
Natural Language Processing Using Artificial Intelligence	200

Hari Prakash Mishra Prof, Anita Mhatre	
Transforming Digital Marketing with Artificial Intelligence Rutuja Yewale, Prof Anita Rajput	204
Transforming Digital Marketing with Artificial Intelligence Rutuja Yewale, Prof Anita Rajput	204
Image Processing Based Signature Verification Technique to Reduce Fraud in Financial Institutions Vaibhav S. Wandkar, Prof.Anita mohan Rajput	216
OFFLINE FORGERY DETECTION OF HANDWRITTEN SIGNATURE USING GAUSSIAN EMPIRICAL RULE Vaibhav S. Wandkar, Prof.Anita mohan Rajput	221
IOT Based Waste Management system Gaurav Thalkar, Darshana Wajekar	228
Cloud Based Smart Ration Card System Gaurav Thalkar, Darshana Wajekar	232
Survey paper on Robotic Process Automation Yashwin Poojary, Darshana Wajekar	236
SURVEY ON ARTIFICIAL INTELLIGENCE IN HEALTHCARE Yashwin Poojary, Darshana Wajekar	239
A Study of Mobile Cloud Computing: Applications and Security Issues Shreyash Sonawane, Darshana Wajekar	244
A Study of Data Integrity and Data Storage Security in Cloud Computing. Shreyash Sonawane, Darshana Wajekar	248
BRAIN TUMOR SEGMENTATION AND DETECTION Omkar Gujar, Prof. Darshana Wajekar	254
BRAIN TUMOR SEGMENTATION Omkar Gujar, Prof. Darshana Wajekar	258

Machine Learning : Multi-Label Classification :MPST : Movie Plot Synopses with Tags : Tags Prediction Prashant Hule, Darshana Wajekar	264
Survey On Multi-Label Classification : Movie Plot Synopses with Tags : Tags Prediction Prashant Hule, Darshana Wajekar	271
Spam Detection Using Machine Learning Algorithms Pratik Gaikwad, Prof. Swapnil Patil	276
Email Spam detection using NLP Pratik Gaikwad, Prof. Swapnil Patil	280
House Price Prediction based on Deep Learning and ARIMA Model Vivek chalke, Ashwini Khillar	283
Literature Paper on Artificial Intelligence for Social Media Chinmay Joshi, Swapnil Patil Sir	291
Literature Review Artificial Intelligence for RoboSoccer MAYUR KAMBLE, Swapnil Patil Sir	295
Review Paper on Artificial Intelligence for Social Media Chinmay Joshi, Swapnil Patil Sir	299
ROBOSOCCERS IN ARTIFICIAL INTELLIGENCE MAYUR KAMBLE, Swapnil Patil Sir	302
Application of Artificial Intelligence Systems in the Process of Crew Training Bubere Arfat , Ravindra Mhatre	309
RESTAURANT REVIEWS PREDICTION AND DEPLOY ON DJANGO Bubere Arfat , Ravindra Mhatre	313
Disease Prediction using Symptoms Pritam Borade, Prof. Ravindra Mhatre	316
Block Chain Based Proxy Re-Encryption Scheme For Secure IoT Data Sharing Sahil Shailendra Mishra, Prof. Ravindra Mhatre	319

Movie Recommender System Shadab Mohammad Saeed Mulla, Prof. Ravindra Mhatre	323
Natural Language Processing in Artificial Intelligence Aishwarya Mokal, Prof. Ravindra Mhatre	327
Personalized Healthcare : Applying Internet of Things and Machine -Learning Pritam Borade, Prof. Ravindra Mhatre	337
Advances in Natural Language Processing – A Survey of Current Research Trends, Development Tools and Industry Applications Sahil Shailendra Mishra, Prof. Ravindra Mhatre	341
A Quick Review of Machine Learning Algorithms Shadab Mohammad Saeed Mulla, Prof. Ravindra Mhatre	345
Sentiment Analysis by Natural Language Processing Aishwarya Mokal, Prof. Ravindra Mhatre	349
ARTIFICIAL INTELLIGENCE DIETICIAN Aayush Tiwari, Prof. Sulbha Nandeshwar	354
APPLICATION OF ARTIFICIAL INTELLIGENCE FOR CYBER SECURITY Meenuraj Thoppil, Prof. Sulbha Nandeshwar	359
ARTIFICIAL INTELLIGENCE IN FINANCIAL MARKETS Meenuraj Thoppil, Prof. Sulbha Nandeshwar	365
ARTIFICIAL INTELLIGENCE IN HEALTHCARE Aayush Tiwari, Prof. Sulbha Nandeshwar	372
ARTIFICIAL INTELLIGENCE IN INTERNET OF THINGS Saurav Shewale, Prof. Sulabha Nandeshwar	377
HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES NAYAN NITN SAPANE, Prof . DEEPESH JAGDALE	390
Music Genre Classification NAYAN NITN SAPANE, Prof . DEEPESH JAGDALE	397

Data Security in Cloud Computing Gargie Ambike, Lalit Patil	401
DATABASES FOR ARTIFICIAL INTELLIGENCE Tejal Deshmukh , Prof.Lalit Patil	409
Flight Price Prediction Using Machine Learning Gargie Ambike, Lalit Patil	413
Analysis on the Influence of Artificial Intelligence Development and Impact on Accounting Kunal Chavan, Lalit Patil	418
Mobile Natural Language Database Interface for Accessing Relational Data Tejal Deshmukh , Prof.Lalit Patil	421
The Artificial Intelligence in Personal Knowledge Management Kunal Chavan, Lalit Patil	423
SENTIMENT ANALYSIS FOR PRODUCT RATING Aishwarya Choudhari, Lalit Patil	426
Real-time Traffic Control in Emergency using Artificial Intelligence Aishwarya Choudhari, Lalit Patil	430

Virtual Assistant: A Product of Machine Learning

Shreya Chandane, Prof. Deepesh Jagdale

Abstract

Machine learning is a buzzword in 21 century, and it is developing very quickly day by day. We are using machine learning in our day-to-day life even without knowing it like Google Maps, Google Assistant, Alexa, etc. The products of ML works with the intention of providing easy and luxurious life to humans. One of the most commonly used and highly acceptable technologies of Machine learning is Virtual Assistant. A VA is, also called an AI assistant or digital assistant, is an application program that recognizes natural language voice controls and completes tasks for the user. It helps us in getting the information using our voice instruction. Virtual assistants use natural language processing (NLP) to meet user text or voice input to executable instructions. Virtual assistants are able to mimic human speech and acknowledge via synthesized voices. Users can ask their assistant's problems, control home automation gadgets, and media playback via voice, and manage other essential tasks such as email, to-do lists, and calendars with verbal instructions. Virtual assistants are available on most smartphones and tablets, old computers, and, now, even standalone devices just like the Amazon Echo and Google Home.



1. Introduction

Although the potential to instruct devices to perform tasks via voice commands has been available since the 1950s, it's only within a previous couple of years that the proliferation of smartphones made voice interfaces accessible to users worldwide. A virtual assistant is based on the fields of Artificial Intelligence, Machine Learning, and Natural Language Processing. A virtual assistant is also referred to as an intelligent virtual assistant (IVA) or intelligent personal assistant (IPA). It enables the user to work

with the mobile device and its apps using natural voice commands.

Virtual assistant services have increased dramatically in demand and offerings. With today's technologies, a VA can work from anywhere and be an immediate contributor to your business. The accuracy, speed, and contextual abilities of VAs like Alexa, Google Assistant, and Siri are all because of Machine Learning algorithms and servers owned by their developing companies. Seldom the word "chatbot" is employed to ask virtual assistants generally or particularly accessed

by online chat (or in some cases online chat programs that are only for entertainment purposes).

Depending on the service you select, VAs can even be a representative for your company. A virtual assistant provides several services to business people or businesses from a far-off location. From digital marketing tasks, scheduling appointments, and handling events to non-public tasks. These tasks or services are supported by user input files, location recognition, and therefore the ability to access information from a spread of online sources. The virtual assistant technology works by combining mobile devices, application programming interfaces, and thus the proliferation of mobile applications. Although like almost everything, it had been created for military services, nowadays this technology is present in our daily lives, becoming a marketing trend in 2018. Not only is it set to be used with the user's voice, but you'll also interact with this software in writing or through images, which suggests that this kind of software offers endless possibilities. Also, we will find them in speakers, messaging applications, smartphones, as a part of a mobile OS, in watches, appliances, cars, etc. You can make a virtual assistant do nearly anything.

2. Literature Survey

[1] Imrie, Peter & Bednar, Peter addresses the way in which new technology could be controlled and utilized to create an Intelligent Virtual Personal Assistant (VPA) with a focus on user-based data. This paper concerns the currently available intelligent

programs with Natural Language Processing to get a better idea on the utilization of NLP for the invention of Virtual Personal Assistants holding the information within the context of the user.

[2] V.Lalitha, A.Dinesh, L.Parameswaran, and S.Dinesh Kumar presents the most widely used Virtual Personal Assistants based on Machine Learning, The authors took an online survey consisted of a mixture of open and closed questions. The goal of those questions was to supply background information regarding device usage and to differentiate between different groups supported frequency of VPA usage. The analysis of this survey revealed two major groups as 'frequent users' and 'infrequent users'. Based on usage behavior the respondents who reported using VPA at least once a week were considered as 'frequent users' and respondents who reported using VPA at least once a week were considered as 'infrequent users'.

[3] Dr. Kulhalli, K. V., Dr. Sirbi, K. & Mr. Patankar, A. J. discussed the Voice Recognition Intelligence for Personal Assistant. In this paper, the author presented a proposed architecture of PARI which is a personal assistant based on Voice Recognition Intelligence. The main feature of PARI is Voice Recognition Engine which has the capability to figure without internet connection i.e. Offline Voice Recognition. The purpose behind designing the PARI is to help native and blind persons who work on their voice commands. As mentioned in the paper, PARI is language barrier independent which actively responds to user's voice

commands faster than the Online Voice Search applications.

[4] Tulshan, Amrita & Dhage, Sudhir took a deep survey on the usability of four virtual assistants, namely, Google Assistant, Cortana, Alexa, and Siri. This survey compares the four virtual assistants where Google Assistant gains the majority of votes from users based on voice-based recognition, contextual understanding, and hand-free interaction. The survey concludes that Google Assistant's accuracy in answering the users' questions was more than that of the other three VAs.

[5] In this research paper, Arora, Shakti; Athavale, Vijay; Maggu, Himanshu & Agarwal, Abhay have studied and analyzed the working model and the efficiency of different virtual assistants available in the market. The paper shows the comparative analysis of the traffic and message communication with the length of conversation for approximately three days which is taken as input to calculate the efficiency of the designed virtual assistant. The designed intelligent virtual assistant could be integrated with Google virtual services and work with the Google Virtual Assistant interface.

[6] In this paper, V. Iyer, K. Shah, S. Sheth, and K. Devadkar explain the implementation of software that gives assistance to people with visual impairments who are less likely to access the internet. The software provides a way to interact with any websites with much ease. Instead of the traditional keyboard and mouse, with the use of voice commands, the software provides a new dimension to access and provide commands

to any website. The software uses the BERT model to add the functionality of providing a summary of the content on the website and answering the questions asked by the user based on the summary.

3. Evolution of Virtual Assistant

The history of virtual digital assistants produced nearly 100 years ago. Back in the 30s of the 20th century, specialists tried to make technologies for recognizing voices. It wasn't easy: the main difficulty was the continuous noise background and the presence of the so-called homonyms (the words, identical in sound, but with different meanings).

Today, developers use AI technology and machine learning to beat such obstacles. The goal is to show neural networks to research the context and identify the most source of the sound. However, it took a minimum of 80 years of preparatory work to form the dream into a reality.

- Within the Mid20th century, the 50s, Bell laboratory had developed a special mechanism, which could recognize numbers dictated by phone (from 1 to 9).
- In the 60s of the 20th century, IBM had created Shoebox, a speech recognition technology. Shoebox placed 16 English words, 6 arithmetic instructions, and 10 digits. Nothing just like the best virtual assistants we affect today!
- In the 1990s, the Dragon Dictate program had entered the mass market. It knew the way to recognize and write down a user-dictated text into a file.

- At the beginning of the 21st century, in 2001, Microsoft added voice input to Office XP.
- In 2002, Google launched Voice Search. Alas, the answer was inconvenient to use, and thus it had been removed to be improved (in the longer term, Voice Search became Google Assistant).
- In 2007, SRI International began started developing its speech recognition solution. A few years later, namely, in 2011, the primary personal voice assistant called Siri had seen the sunshine of day: a system was ready to look for information online, work as a voice menu, and conduct a conversation with a user.
- Within the spring of 2012, Google trapped with its competitor and launched the Google Assistant on the market. Both virtual assistants add an equivalent way and have similar functionality.
- In 2014, the Amazon Echo smart speaker with its Alexa came into existence. Other virtual digital assistants involve Bixby from Samsung and Cortana from Microsoft.

4. Working of Virtual Assistant

Virtual assistants work on the idea of speech synthesis and recognition technology. They passively read all sound signals, process them, and, if necessary, answer them. The whole operation of the Virtual assistant is given below:

The user says a passphrase and provides a voice command. After that, his request is being sent to the server for processing, which occurs in four stages:

- **Filtration.** The system filters out the signal received by removing any sound intervention.
- **Digitization.** Conversion of sound waves into a digital form.
- **Sound signal analysis.** The system selects sections containing speech in the signal and estimates its parameters that are, the probability of a correlation between two words.
- **Identify data patterns.** Working with data templates and choosing the correct pronunciation for the phrase.

Then the user receives an acknowledgment of his request: the given instruction is being executed or a search query is being formed, the results of which produce a voice response.

After processing the data on the server, if the voice-based virtual assistant doesn't understand the request or cannot find the right answer, the system asks the user to rephrase the question or provide supplementarily.

5. Applications of Virtual Assistants

Alexa: Amazon has designed the Alexa Voice Service (AVS), Amazon's cloud-based voice service, to mimic real conversations, using intuitive voice commands to get this service to perform specific tasks. "Alexa" acts as a wake-up word for the Alexa-supported devices that alert the service to start out listening to the user commands. Its Alexa Voice Service (AVS) has intelligent voice recognition and natural language understanding (NLU) services. It can play

music, provide information, deliver news and sports scores, weather reports, controls smart home services, and even allow Prime members to buy for order products from Amazon. In order to use Alexa, the user will need a tool that integrates voice technology. This typically means an Amazon device, like an Echo Dot or Fire TV Stick.

Siri: Siri is Apple's Virtual Assistant for iOS, macOS, tvOS, and watchOS devices that utilize voice recognition. It is supported by the fields of AI, Machine Learning, and natural language processing. Moreover, Siri is not all business, though. Users can have some fun and ask it more complex questions if users like. Siri can serve you when the user is out and about, with sports and entertainment information, phone calls and messages, getting organized, tips and tricks, read your last email, text your friends and family, mix-up playlist, even finding a table for 3 during a restaurant and far more stuff like this.

Google Assistant: Voice Searching, Voice Instructions, and Voice-Activated device control offered by Google's Virtual Assistant let you complete different sorts of tasks after you've said the "OK Google" or "Hey Google" wake words. It is designed to provide you conversational intercommunications. It is Google's virtual helper that permits you to urge stuff done faster. Instead of doing tasks manually by touching your finger on the screen a gazillion times, you will get the work done by utilizing voice instructions. Google Assistant is sensible and really well integrated with Android. You can use it to open apps, send messages, make calls, play a selected song on

YouTube Music, check the weather, and lots of other things.

6. Privacy policy of Virtual Assistants

Virtual assistants have a spread of privacy concerns related to them. Features like activation by voice pose a threat, intrinsically feature requires the device to always be listening.

6.1 Google Assistant

Google Assistant doesn't store your data without your permission. To collect the audio, you can go to Voice & Audio Activity (VAA) and switch on this feature. Your audio files are transferred to the cloud and used by Google to strengthen the performance of Google Assistant, but as long as you've turned on the VAA feature.

6.2 Amazon's Alexa

Amazon's Virtual Assistant Alexa only listens to your conversation once you use its wake word (like Alexa, Amazon, Echo). It starts recording the conversation after the decision of a wake word. It stops listening after 8 seconds of silence. It sends the recorded conversation to the cloud. You can erase your recording from the cloud by visiting 'Alexa Privacy' in 'Alexa'. You can stop Alexa from taking note of your discussions using the 'mute' feature of Alexa, after muting the device, it cannot hear you albeit you employ the wake words (like Alexa).

6.3 Apple's Siri

Apple doesn't record your audio to enhance Siri, it uses transcripts instead. It only sends data that is vital for analysis, as an example, if you ask Siri to read your message it won't send the message to the cloud, the machine will directly examine the message without the server's intervention. Users can opt-out anytime if they don't want Siri to send the transcripts within the cloud.

7. Uses of a virtual assistant:

Here are a couple of the ways you'll expect to use a virtual assistant on each day to day basis:

- Daily news and weather update,
- Reviewing schedule for the day,
- Booking meetings and scheduling appointments in the calendar,
- Taking notes in a meeting and sharing those notes for follow-up,
- Getting directions to enable you to go from one meeting to the next,
- Finding recipe's that you can make for dinner with the ingredients you have or what you need to get,
- Bookings for travel, for both personal or for business,
- Optimizing your week by aggregating chores together in dedicated timely blocks.

8. Types of Virtual Assistant Services

1. Social Media Virtual Assistant: "Virtual social media assistant is nothing, but a faraway worker who handles social media marketing tasks of a client". Virtual social

media assistants from a reputed agency are usually well-versed in managing all the leading social media platforms, like Facebook, Twitter, LinkedIn, Instagram, Pinterest, YouTube, etc.

2. land Virtual Assistant: land virtual assistants are remote workers who can complete an honest kind of land task, like preparing & filing paperwork, responding to email inquiries, fixing meetings, etc.

3. Virtual Research Assistant: The virtual research assistant does internet research for his or her clients.

4. Virtual Administrative Assistant: A virtual administrative assistant is someone who does a variety of administrative tasks, like managing contact lists, handling billing and accounting, handling basic HR duties, etc.

5. Virtual Marketing Assistants: A virtual marketing assistant is someone who helps their customers in several quiet marketing tasks, like Facebook marketing, Twitter marketing, email marketing, etc." A virtual marketing assistant won't only keep your marketing efforts on track, but also restrict time and money.

6. Virtual Bookkeeping Assistant: A virtual bookkeeping assistant is someone who manages the bookkeeping for the business.

7. Data Entry Virtual Assistant: Data entry virtual assistants work as a savior for entrepreneurs. they're going to prevent both time and money. Hiring a knowledge entry virtual assistant is usually cheaper than hiring a daily employee to finish data entry tasks,

especially once you don't need to affect a mountain of data fairly often.

8. eCommerce Virtual Assistant:

eCommerce virtual assistants can ease out a load of eComm businessmen. Whether you're trading products on your website otherwise you're selling them on a marketplace, an eCommerce virtual assistant can assist you to run your business smoothly.

9. Virtual Assistant Advantages

- Lower cost, fewer overheads – performing from a foreign location with no commuting.
- Majority of contact via video call, email, or telephone, at scheduled times which can be a benefit if you favor to distance yourself from employees.
- Virtual assistants are often self-employed or work for a further body, often for multiple clients simultaneously, removing needing for you to provide benefits, organize payroll or taxes.
- Can manage many equivalent jobs as a private assistant could be asked to try to, as many tasks normally require the utilization of computers and other electronic devices.
- Your virtual assistant will haven't any interaction with the other members of your team, this will be advantageous if you would like to stay things separate and personal.

10. Virtual Assistant Disadvantages

- Their role will be deeply defined without modification unless you take more time to give additional instruction.

- Generally they will be operating for many customers simultaneously, so your work may not be done at a particular time of day, they may not be able to be as responsive as you might like. Especially if you need something done quickly or if changes need to be made part way through a task.
- May is not available when you need them.
- May does not receive all their time or loyalty.
- Not available to meet you in person; you may never meet them in person if they are a vital distance away.
- Will is not able to assist with the rest of the team or help the team out in a pinch.
- Vetting a virtual assistant can be inconvenient and difficult.
- It is unlikely you will be able to teach your assistant as such.

11. The future of virtual assistants

Virtual assistants are quickly evolving to supply more capabilities and value to users. As speech recognition and tongue processing advances, so too will a virtual assistant's ability to know and perform requests. And as voice recognition technology upgrades, virtual assistant use will move deeper into business workflows. Chatbots, so popular today, still interact with consumers using text messages, but soon they'll find out how to recognize speech. Multitasking technology is being actively introduced into the software, which can soon allow one device to support several assistants. In the future, it'll be reasonable to individualize any personal voice assistant and connect it to the flavor of the buyer. Voice

assistants might arise in security cameras, thermostats, and even household devices. Tomorrow's virtual assistants are going to be built with more high-level cognitive computing technologies, which can allow a virtual assistant to know and perform multistep requests and perform more complex tasks, such as making a plane reservation.

12. Conclusion

The world of virtual assistants was born to form life easier for humans. Our tasks are often solved in a matter of seconds just by using our voice. Virtual Assistants like Alexa, Siri, Google Assistant, are going to be more and more accurate, which successively will exponentially increase the number of users who will begin to rely on this technology. Within the planet of selling, it's a poorly researched strategy, but with great potential, we could cash in. With more authentic virtual assistants, we'll see more Voice Market strategies. Today, we would like to launch a challenge: to include these virtual assistants in our email marketing strategy.

References

1. Imrie, Peter & Bednar, Peter. (2013). Virtual Personal Assistant.
2. V., A., L., & S. (2019, July 7). ML Based Virtual Personal Assistant [PDF]. Chennai: International Journal of Engineering Research in Computer Science and Engineering (IJERCSE).
3. Dr. Kulhalli, K. V., Dr. Sirbi, K. & Mr. Patankar, A. J. (2017). Personal Assistant with Voice Recognition Intelligence [PDF]. Delhi: Research India Publications.
4. Tulshan, Amrita & Dhage, Sudhir. (2019). Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa: 4th International Symposium SIRS 2018, Bangalore, India, September 19–22, 2018, Revised Selected Papers. 10.1007/978-981-13-5758-9_17.
5. Arora, Shakti & Athavale, Vijay & Maggu, Himanshu & Agarwal, Abhay. (2021). Artificial Intelligence and Virtual Assistant—Working Model. 10.1007/978-981-15-7130-5_12.
6. V. Iyer, K. Shah, S. Sheth and K. Devadkar, "Virtual assistant for the visually impaired," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 1057-1062, doi: 10.1109/ICCES48766.2020.9137874.
7. Tasca, N. (2019, September 23). Doing more to protect your privacy with the Assistant. Retrieved from <https://blog.google/products/assistant/doing-more-protect-your-privacy-assistant/>
8. GP. (2011). Retrieved from <https://www.amazon.com/gp/help/customer/display.html?nodeId=GVP69FUJ48X9DK8V>
9. Improving Siri's privacy protections. (2021, March 12). Retrieved from <https://www.apple.com/in/newsroom/2019/08/improving-siris-privacy-protections/>
10. Virtual Assistant Services. Types of Virtual Assistant, Definition, Free Trial.

(2020, May 15). Retrieved from
<https://www.acelerartech.com/va-guide/types-of-virtual-assistant-services/>

11. Botelho, B. (2017, October 31). What is virtual assistant (AI assistant)? - Definition from WhatIs.com. Retrieved from
<https://searchcustomerexperience.techtarget.com/definition/virtual-assistant-AI-assistant>

12. 12.

<https://whataftercollege.com/machine-learning/machine-learning-virtual-assistants/>

13. <https://agilie.com/en/blog/virtual-assistant-apps-benefits-and-abilities>

14. <http://timemanagementninja.com/2013/11/top-10-tasks-to-delegate-to-a-virtual-assistant/>

Artificial intelligence in e-commerce

Shreya Chandane, Prof. Deepesh Jagdale

Abstract

Artificial intelligence has the powerful ability to acquire and analyze large volumes of data and provide decisions for action. E-commerce is now adopting this technology to identify patterns based on browsing, purchase history, credit checks, account information etc. Artificial intelligence has the powerful ability to acquire and analyze large volumes of data and provide decisions for action. E-commerce is now adopting this technology to identify patterns based on browsing, purchase history, credit checks, account information etc. This data collected then form the basis of creating customized recommendations for each customer experience. This paper highlights the role of artificial intelligence in e-commerce and its application in different areas of e-commerce.

Introduction:

The term e-commerce was coined back in the 1960s, with the rise of electronic commerce – the buying and selling of goods through the transmission of data

‘E-commerce’ and ‘online shopping’ are often used interchangeably but at its core e-commerce is much broader than this – it embodies a concept for doing business online, incorporating a multitude of different services e.g. making online payments, booking flights etc.

Why Do People Buy ‘Online’?

Lower Prices: Managing an online storefront is far cheaper than an offline, brick and mortar store. Typically less staff are required to manage an online shop and warehousing is not necessarily required. so e-commerce business owners can afford to pass operational cost savings on to consumers.

Accessibility and Convenience: Unlike many offline stores, consumers can access e-commerce websites 24 hours a day. Customers can read about services, browse products and place orders whenever they wish. In that sense, online shopping is

extremely convenient and gives the consumer more control.

Wider Choice: For the past twenty years, the growth of online shopping has to a large extent customers have almost endless choice of brands and products to choose from, Items can be sourced and shipped globally.

AI is beginning to embed itself into all aspects of our lives. From the growing number of self-checkout cash registers to advanced security checks at the airport; artificial intelligence is just about everywhere.

It’s widely anticipated that AI is set to go into turbo drive in the next couple of years with giants such as Google and Microsoft already investing heavily into new AI initiatives.

If you search for AI online, you will stumble across hundreds of articles that predict a marketplace dominated by the use of artificial intelligence.

Many ecommerce businesses are already using forms of AI to better understand their customers, generate new leads and provide an enhanced customer experience. a recent study by Business Insider suggests that as

much as 85% of customer interactions will be managed without a human by as soon as 2021.

Following are powerful and practical ways that retail businesses are using AI in the world of online shopping

1. Create customer-centric search

Consumers often abandon ecommerce experiences because the product results displayed are often irrelevant. To tackle this problem, Twigggle, software model utilizes natural language processing to narrow, contextualize and ultimately improve search results for online shoppers. It enables ecommerce search engines to think the way humans do. New software platforms that drive ecommerce websites create innovative visual search capabilities. the software can successfully assist the customer in finding the product they desire. The consumer no longer needs to be shopping to see something they would like to purchase.

US-based tech start-up Clarified, enables developers to build smarter apps that 'see the world like you do', empowering businesses to develop a customer-centric experience through advanced image and video recognition. Leveraging machine learning, the AI software automatically tags, organizes and visually searches content by labeling features of the image or video.

A great example is Interest's recent update of its Chrome extension, which enables users to select an item in any photograph online, and then ask Interest to surface similar items using image recognition software. As well as finding matching products, For example, they may take a liking to a friend's new dress or a work colleague's new pair of gym Nike's. If there is a visual, then AI enables consumers to

easily find similar items through e-commerce stores.

2. Retarget potential customers

Many businesses are overloaded with unmanageable customer data that they do little or nothing with. For instance, if we take a deeper look at the retail industry, facial recognition is already being used to capture shoplifters by scanning their faces on CCTV cameras.

Some businesses are now using facial recognition to capture customer dwell times in the physical store. This means that if a customer spends a notable amount of time next to a specific product e.g. an iPod, then this information will be stored for use upon their next visit.

The face of sales is changing with businesses responding directly to the customer. It is as if businesses are reading the minds of customers and it's all thanks to the data used with AI.

3. Create a more efficient sales process

Customer's lives are now heavily influenced by a variety of different media from TV adverts to social media. In fact, even Snap chat has established itself as a viable sales and marketing tool, opening up new opportunities.

If you want to tailor your problem-solving solutions and create a strong sales message that reaches consumers at the right time on the right platform, then integrating AI into your CRM is the way to go.

Many AI systems enable natural language learning and voice input such as Siri or

Alexa. This allows a CRM system to answer customer queries, solve their problems and even identify new opportunities for the sales team. Some AI-driven CRM systems can even multitask to handle all these functions and more. They achieve this by asking the customer questions e.g. “where and when will you be using your jacket?” through voice input AI technology. IBM’s software then scans hundreds of products to find perfect matches based on real-time customer input and its own research e.g. such as weather conditions in the local area.

There is little doubt that AI is already starting to impact e-commerce and has started to evolve the sales process with new data. The changes will ensure that customers will no longer be offered products and services that are inappropriate.

4. Create a new level of personalization across multiple devices

Personalization is nothing new for ecommerce and if you frequently use Amazon then you’ll know exactly what we’re referring to. However, with the ever-increasing advances in artificial intelligence and machine learning technologies, new deep levels of personalization have started to penetrate the fast-growing e-commerce world. Whether it is a mobile application, the website, or an email campaign, the AI engine is continuously monitoring all devices and channels to create a universal customer view. This unified customer view enables ecommerce retailers to deliver a seamless customer experience across all platforms. The next time a customer is browsing iPhone cases on your website, they may receive a push notification on their mobile, informing them about your flash sale for iPhone cases. They directly make

the purchase on their phone, saving a lot of steps for both parties.

5. Provide a personal touch with chatbots

Many ecommerce retailers are already becoming more sophisticated with their AI capabilities in capturing attention, and one approach widely developing is known as ‘conversational commerce’.

If brands wish to survive then this is one of the priority business strategies that must be executed. The use of artificial intelligence through the application of ‘chatbots’ is just one way to drive the conversation in this next era of conversational commerce. A chatbot is a specific computer program that is designed to simulate conversation with human users over the Internet.

Chatbots can actively take on some of the important responsibilities that come with running an online business, particularly when it comes to executing tasks for operations and marketing.

Chatbots can automate order processes and are an effective and low-cost way of providing customer service. Customer service via social is starting to establish itself as a requirement as opposed to an option.

Often when consumers are browsing online, they are already logged into social platforms such as Facebook. With this in mind, there is a great opportunity to use messenger functionality to confirm orders or to provide instant online support.

Once the chatbot system has been integrated with one of your shopping carts, it can work with all the stores based on the platform. The more shopping carts that your chatbot

application supports, the more potential customers it has.

Chatbots provide a valuable customer support solution for ecommerce retailers.

6. Implement virtual assistants

We're all familiar with the usual suspects: Siri, Google Now and Alexa, and they have successfully introduced us to the idea of talking to a phone, laptop or even a home appliance. Virtual assistants are impacting the way customers purchase, and provide a creative opportunity for ecommerce retailers to take advantage of.

Let's take a look at Amazon's virtual assistant, Alexa.

Their virtual assistant, which has recently emerged as one of the most prominent voices in commerce, has been successfully integrated into Amazon's own products as well as products from other manufacturers. By using Alexa on Amazon's Echo device, customers can discover local gigs for the upcoming weekend through Stub Hub, arrange transport to and from the event via Uber, or even order pre-event dinner from Domino's and track the order status in real time.

7. Integrate with everyday household items

There are few more interesting examples of AI integration than the partnership between Amazon's Alexa and LG's Smart InstaView refrigerators.

LG have experimented with several previous versions of the InstaView refrigerator with enormous touchscreens built into the door. However, this time around, LG has tacked on a virtual assistant and webOS software. It's a place where a virtual assistant has real potential to be especially helpful.

In addition to providing news and weather updates, it can lend a hand with your shopping orders. You'll never have to run to the shop for milk again. Imagine the possibilities for ecommerce retailers that have direct access to the homes of consumers.

8. Improve recommendations for customers

Using AI, brands can more intelligently and efficiently scan through petabytes of data to predict customer behavior, and offer relevant and helpful recommendations to individual consumers.

For instance, Starbucks recently launched 'My Starbucks Barista', which utilizes AI to enable customers to place orders with voice command or messaging. The algorithm leverages a variety of inputs, including account information, customer preferences, purchase history, third-party data and contextual information. This allows the coffee giant to create and deliver more personalized messages and recommendations for their customers.

The process of recommendation is widely practiced by ecommerce retailers to help customers find the best solution.

For example, Amazon makes recommendations to users depending on their activities on the site and any past purchases.

Netflix makes TV and movie recommendations based on a user's interaction with categories e.g. drama, comedy and action.

9. Introduce virtual personal shoppers

There are many perks of in-store shopping that both brands and customers love. For instance, the customer has the opportunity to directly engage in conversation with a shop assistant.

They may ask the customer about a specific item in a particular color or size. These perks are limited online as customers have to go through the time-consuming (and sometimes frustrating) process of ticking boxes or entering keywords.

The impressive and friendly assistant helps to simplify mobile shopping and provides customers with the best and most relevant deals and products that are tailored to your preferences.

Sentient Technologies, the world's most funded AI Company, is enabling people to see and buy things they weren't even aware existed or even knew they wanted. The introduction of virtual personal shoppers is a true example of how AI, for the ecommerce industry, is completely disrupting traditional customer engagement techniques.

10. Bridge the gap between personalization and privacy

Whenever it comes to discussing the topic of personalization, there is often a trade off with concerns to user privacy. User privacy has been a hot topic in recent years with its importance considered stronger than ever. Brands is actively striving to take transparency, security and honesty to an entire new level. However, to achieve this, brands cannot afford to abandon user personalization, given its critical role in any successful e-commerce venture.

Users are a little more comfortable with sharing their details on the promise they are getting something truly valuable in return.

For example, if you give Google Now access to your account, it will sync your calendar, emails and search habits.

Each morning, it will greet you with a small briefing of what you currently have on your plate and will let you know if you're going to be late to the office due to a train cancellation.

Amazon's Alexa applies the same magical approach for daily life. The modern shopping assistant puts your day-to-day routine first and even helps with daily house chores. Most recently, Amazon added the required intelligence for Alexa to buy on your behalf.

11. Improve dialogue systems

Amazon have started to apply AI to widely known issues with dialogue systems, such as speech recognition, natural language understanding and question answering.

For example, by applying a class of machine learning algorithms known as 'deep learning', Amazon can effectively convert speech (spoken by customers) to text with accurate results.

Amazon are also tackling the problem of answering questions automatically using AI by leveraging content within website pages such as product descriptions and customer reviews.

12. Tackle fake reviews

Any experienced online retailer will be able tell you of at least one painful story about receiving fake reviews for their brand.

Consumers are flooded with an abundance of advertising every day, which can become

overwhelming and this will often delay their decision making. This is where word of mouth has become invaluable.

If a customer's friend has purchased your product and had a positive experience, then the customer will end up buying the product too.

Some ecommerce retailers are using artificial intelligence to fight astroturfing by putting more emphasis on verified and helpful reviews.

Amazon uses AI to combat fake product reviews and inflation of their popular star ratings.

There is also preference to those reviews that are marked as helpful by other users as well as the newer and more up-to-date critiques on its site. The business is continuously reviewing several review characteristics such as ratings to detect fake reviews. They are critical to the company as they help to build customer trust in Amazon.

13. Personalize the customer experience

Personalizing the results means a lot of overwhelming information that travelers are often presented with is removed. This allows consumers them to make faster decisions and with more confidence.

AI is expected to expand in this industry to a point where customers can type in specific adventures and it will provide them with a solution e.g. "where can I go rock climbing on my honeymoon?" The AI systems can then provide personalized recommendations for points of interest and local insights that you never knew existed.

Conclusion:

Although the term 'artificial' may imply something negative or dehumanized, artificial intelligence allows businesses to provide a more personalized experience for their customers.

AI makes it possible for ecommerce retailers to analyze millions of interactions every day and ultimately target offers down to a single customer – an experience every marketer dreams of providing.

AI technology is likely to have an enormous and beneficial impact on the ecommerce industry in the coming years. It will change, and arguably improve the way consumers find products online.

AI is already boldly walking and talking among us and in the age of Instagram and Snapchat, and the rapidly decreasing attention spans of the digital age, there is evidence to suggest that new AI-driven platforms will be essential to ecommerce success.

Reference:

International Journal of Advances in Engineering & Scientific Research, Volume 4,(Issue 4,Jun-2017), pp 62-. ISSN: 2349 – 3607 (Online) , ISSN: 2349 –4824 by Shyna K and Vishal M (2017).

<https://www.linnworks.com/blog/artificial-intelligence-in-ecommerce>

<https://feinternational.com/blog/what-is-e-commerce-an-introduction-to-the-industry/>

SMART BLIND STICK USING ARDUINO

Neha Patil, Prof. Deepesh Jagdale

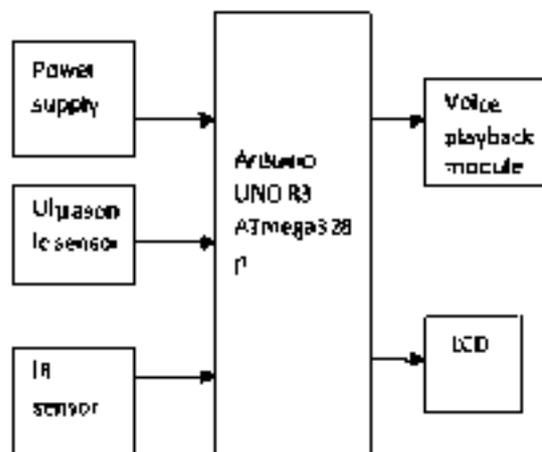
Abstract : This paper describes ultrasonic blind walking stick with the use of arduino. According to WHO, 30 million peoples are permanently blind and 285 billion peoples with vision impairment . If u notice them , you can very well know about it they can't walk without the help of other. One has to ask guidance to reach their destination. They have to face more struggles in their life daily life. Using this blind stick , a person can walk more confidently. This stick detects the object in front of the person and give response to the user either by vibrating or through command. So, the person can walk without any fear. This device will be best solution to overcome their difficulties .

Key Words: Arduino, Ultrasonic sensor, Walking Stick.

1.INTRODUCTION

This walking stick is an alternative to the traditional walking stick. Here, Arduino UNO , ultrasonic sensor, IR sensor , voice playback module , LCD display and voltage regulator are used. Arduino is a microcontroller which can do all the calculations very fastly and quickly with great accuracy. Ultrasonic sensor is used to detect the object in the front of the person by measuring the distance between the object and the stick. For left and right object detection, IR Sensor is used which is very small in range. So, it detects the object which are very close. Using more ultrasonic sensor may create calculation problem. So, IR Sensor is Preferred. The voice playback module will assist the blind person to reach the destination through the command or microphone.

2.BLOCK DIAGRAM



3. HARDWARE REQUIREMENTS

3.1 ARDUINO UNO R3

Arduino UNO is a microcontroller board based on ATmega328p. It has 20pins out of which 16 digital input and output pins and 6 analog input pins , 16MHZ Quartz crystal, power jack, ICSP header and reset button compareto PIC

microcontroller , it is very easy to perform with arduino since it is user friendly, The Operation Voltage is 5V, You can directly connect it to computer with USB cable, power it with AC-DC adapter or battery.

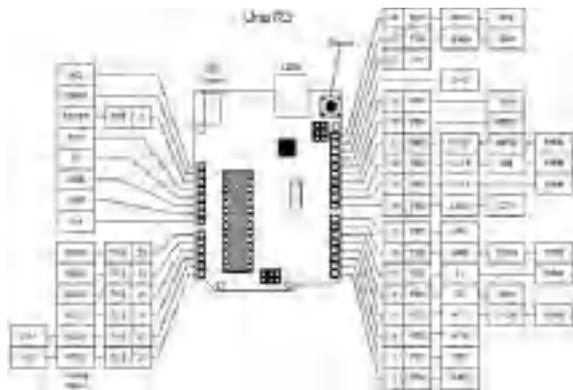


Fig 2: Pin Configuration of Arduino

3.2 ULTRASONIC SENSOR

Ultrasonic sensor is used to detect the object in front of the person . HC-SRC04 ultrasonic sensor has 4 pins-ground, Vcc, trigger and Echo . It ranging from 2cm to 400cm. Mainly it has two opening –one is transmitter which is used to transmit the signal and another one is receiver which is used to receive the signal. It sends ultrasound waves at high frequency and receive back the signal .

3.3 IR SENSOR

IR Sensor which is used as obstacle detector where it transmits the infrared waves and hits the object and reflected back the signal to sensor . it ranges from 700nm to 1mm. IR output various depending on infrared rays that have been received. Since, this variation cannot be analyses as such, output provide for

comparative circuit. If IRReceiver does not receive any signal, the output of the comparator goes low and LED does not glow whereas if it receive any signal, the output goes high and LED Starts Glowing

4. METHODOLOGY

The working behind this blind stick is that it is used for special purpose as a sensing device for the blind people. The circuit provides 5V power supply for the circuit and maintains its output of the power supply at constant level. It is used widely to detect objects using ultrasonic sensor and IR sensor. If any object is present, the ultrasonic sensor detects the object by measuring the distance between the object and the user and sends the data to the arduino UNO. To determine the distance of an object , calculate the distance between sending the signal and receiving back the signal.

$$\text{*Distance=speed*time}$$

The speed of the signal travelling through air is 341m/s. The time is calculated between the sending and receiving back the signal. Since the distance travel by the signal is double, it is divided by two i.e.,

$$\text{Distance=*Distance/2}$$

IR sensor is placed at right and left of the stick to detect the object. Since, it is very small range, it detects the closer objects. Arduino processes with this data and calculates with the command conditions. If any object is found nearer, it sends the command to the user

through the speaker or microphone. The command is already stored in the voice playback module which sends alert message to the user about the object.

The command condition is as follows:

[1] If the distance between the objects and the person is 30 inch, it will send the command as the obstacle is nearer to the person.

2] If the object is about 60-90 inch, it will send the command as the obstacle is just closer and reaching the person.

[3] If the object is about 90-120 inch, it will send the command as the object is far away from the person.

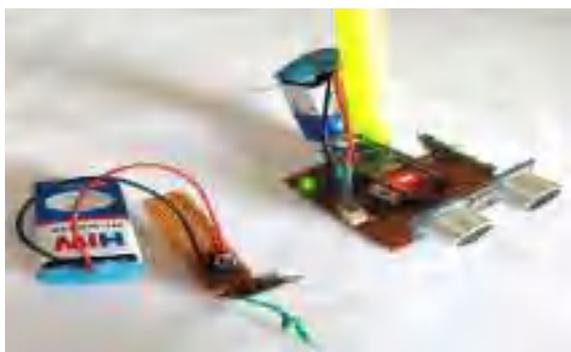


Fig 6: Output

5. FUTURE SCOPE

A variety of future scope are available that can be used of with the stick such as usage of Global positioning System can help the blind person to source to destination route information. GPS can help to find the shortest and best path as accordingly to Google (Bing map based on real time co-ordinates). GSM attachment can help in future for any immediate casualty help. It can also contain

special arrangement to connect the walking stick to aadhar card of blinds, helping the government serve the physically disable even better. Water sensor that sense any kind of water allowing the safe walk of the blind people in order to avoid slipping .

6. CONCLUSIONS

The Blind Walking Stick has been finally made into prototype which can be used to guide the blind . Its aims to solve the problems faced by the blind people in their daily life. The system also takes the measure to ensure their safety . This project will operate to help all the blind people in the world to make them easier to walk everywhere they want. It was done to help the blind to move in front very well. It is used to help the people with disabilities that are blind to facilitate the movement and increase safety.

REFERENCES

- [1] Osama Bader AL-Barrm International Journal of Latest Trends in Engineering and Technology(IJLTETJ)
- [2] Smart Cane Assisted Mobility for the Visually Impaired, World Academy of Science, Engineering and technology International Journal and Information Engineering Vol:6 No:10,2012
- [3] J.B.F. van Erp, J.A. Veltman, H.A.H.C. van Veen (2003): A Tactile Cockpit Instrument to

support Altitude Control, Proceedings Human Factors and Ergonomic Society 47th annual meeting 2003.

[4] W. Docbelle Artificial Vision for the blind by Connecting a Television Camera to the Brain, ASAIO Journal, Vol:46, page 3-9,2000.

[5] G. Prasanthi and P. Tejaswidho Sensor assisted Stick for the Blind People-Transmission on Engineering and Sciences, vol:3,number 1, pp-, 12-16,2015.

[6] I .Ulrich, J. Borenstein, The GuideCane – Applying Mobile Robot Technologies to Assist the visually impaired, IEEE Tr.SMC, Vol:31, No.2, March 2001.

[7] S. Shovel, J. Borenstein, Y. Koren, Auditory guidance with the Navbelt-a computerized travel aid for the blind, IEEE Tr. SMC, Vol.28, August 1998.

A Study of Robotic Process Automation Use Cases Today for Tomorrow's Business

Neha Patil, Prof. Deepesh Jagdale

Abstract: Historically, most of the companies across various industry domains, implemented multiple IT tools and processes (workflows), that often had a mix of automated and manual steps. These manual steps performed across these industries constantly required lot of manpower, generated human errors, reduced employee motivation due to mundane repetitive tasks, degraded productivity, increased operational expenses etc. The automated steps performed, had its own downside due to the implementation of multiple IT automation solutions adding up license/development/maintenance cost, increased integration timelines, continuous patching of code to accommodate underlying business logic changes. Businesses today and over the years had been constantly looking for IT solutions that could holistically augment their user's performance improving customer satisfaction, increase productivity and quality, reduce costs etc. Robotic Process Automation (RPA) has recently emerged as a game-changing technology outperforming multiple solutions that evolved from Business Process Management (BPM) industry over the decades. Robotic Process Automation, since its inception, due to its capability to automate rudimentary rules-based tasks that are repetitive and manual, found its way across multiple industry domains, extending the use cases and continuing to evolve with business requirements. This paper study RPA's application across various industry domains, benefits of RPA, and future of RPA. Keywords —Robotic Process Automation, RPA, Cognitive Automation.

INTRODUCTION

In today's fast paced work environment, its critical for businesses to think better ways of engaging their employees, strategizing solutions to avoid rudimentary manual repetitive human workloads, in order to generate maximum throughput. Robotic Process Automation, or RPA, is a new solution in business process management technology leveraging artificial intelligence and machine learning which helps businesses accomplish the above said goal. The IEEE (Institute of Electrical and Electronics Engineers) Standards Association defines Robotic Process Automation (RPA): "A preconfigured software instant autonomous execution of a combination of processes, activities, transactions, and tasks in one or more unrelated software systems to deliver a result or service with human exception management" (IEEE Corporate Advisory Group 2017). Robotic Process Automation is a versatile, non-invasive, scalable technology that can be readily applied to automate use cases across many different departments and industry processes. This study will review the significance of RPA across various industry domains specifically,

- Banking

- Audit
- Insurance
- Retail
- Manufacturing

Across these industries, RPA could cover application areas / use cases in various business functions not limited to,

- Sales
- Finance
- Operations
- Procurement
- HR
- Customer Support
- IT

I. WHY RPA IMPLEMENTATIONS ARE ON THE RISE?

Research studies across all the industry domains on Strategic agenda for their businesses show Continuous improvement and Automation as top priority items.

For any organization, even the simplest use of right technology can dramatically improve their business efficiency and productivity. One way to assure that your technology/system delivers desired value to your business is to align it with your strategic business priorities. Majority of organizations have already started their RPA journey / have RPA on their roadmap since RPA evolved and proved itself as a solution that can deliver significant productivity improvement and Automate many of their rudimentary business tasks. Due to its versatility, RPA is becoming a software in solving systematic automation challenges and helping organizations to grow into digital workplaces.

II. BENEFITS OF RPA

RPA can help businesses solve their systematic automation challenges that are repeated based. It is a way of creating a virtual workforce that brings in extended capacity to businesses to gain a competitive edge in customer satisfaction and enterprise agility. These Software Robots or Bots offers a magnitude of benefits, include

- Accuracy – To err is human; RPA can significantly improve the accuracy of business operations since they are impervious to human error.
- Consistency – RPA can automate any business process that is repeatable, and rule based. It can execute business processes with extreme precision at very high velocity that even smartest humans cannot.
- Reliability – RPA with its inbuilt capabilities including Monitoring and Analytics, can offer detailed Audit logs that allows users to oversee the health of their systems and processes. unexpected errors occur, could be configured to personnel and handle them appropriately with specific workflows.
- Scalability – RPA tools offer centralized management of Bots that allows businesses to Scale up / down quick. With a simplified Bot architecture and

ensuring the bots are audited in regular intervals post deployment, scaling up/down RPA infrastructure would be simple.

- Reduced Costs – In general, RPA Bots cost way less than a full By deploying bots to automate mundane business processes, RPA can significantly reduce operation cost.



Non-Invasive – RPA is designed to mimic human action; hence the technology interacts with data within the presentation layer of platforms and applications. This allows companies to implement RPA without making any significant changes to their legacy systems. This also brings in significant savings as it reduces the need for consistent IT development and technical workforce

- Improved Productivity – RPA with its ability to autonomously execute business processes 365 x 24 x 7 with high efficiency and accuracy can bring in significant business productivity gains. Bots cut down the time it takes to do mundane repetitive tasks and outperform humans as virtual workforce.
- Improved Employee Morale – RPA can relieve employees of repetitive mundane tasks, so they can focus on more engaging and challenging activities. Most employees have better morale when they invest their time and talent in jobs that are more interesting and less routine. Since better morale improves productivity, businesses can get more work done by their human team members as well.
- Improved Transparency – RPA bots can detect poor data integrity errors and enable standardization. This leads to transparency in any industry by identifying significant errors hampering both management decisions and operational performance
- Improved Compliance – RPA Bots execute processes per instructions they have been configured to follow and provide an audit trail for each step. Furthermore, if any step in a specific process need to be reviewed, bots

have ability to play back their actions. This controlled nature of bots increases transparency and eliminates fraud.

- Improved Reporting – RPA generates significant amount of data that allows organizations to analyse and identify process bottlenecks / inefficiencies. This operational insight provided by bots allows organizations to streamline existing business processes.

- Improved Quality – RPA enables businesses to streamline and standardize their processes which reduce data errors. This reduction in errors lead to high quality data, that enables more reliable analysis.

IV. RPA IN INDUSTRY In order to remain competitive in this rapidly evolving ecosystem, every industry need to make significant investment in disruptive technologies that could maximize their efficiency while keeping the costs low and offer best possible user experience to their customers. RPA is one of the disruptive technologies that companies across various industries want to implement. We will review its significance in some of the industries below,

A. RPA in Banking:

Banking industry deals with enormous amount of data every single day and its critical for banks across the globe to deal with this sensitive dataset with high level of accuracy and quick turnaround time. RPA can minimize the manual processing data errors by automating various processes including,

- Customer service – RPA can help in resolving low priority customer queries and lets human employees focus on high priority customer service issues.
- Accounts Payables – RPA can help automating the vendor invoice processing and payment to vendor account post reconciliation & validation.
- Accounts Receivables – RPA can automate processing incoming payments and clearing customer dues.
- General Ledger – RPA can help automating the collection of financial data, assets, liabilities, revenue, expenses and update GL with right information.
- Credit Underwriting – RPA can automate verification of income/expense/exposure of

credit applicant against internal and external databases.

- Compliance – RPA generates full audit trails of every process executed which significantly improves the quality of compliance process.
- Credit card processing – RPA can automate various time-consuming processes like document collection, credit checks, background check and quickly determine if the customer is eligible for credit card.
- Consumer Loan processing – Most of the consumer loans including Auto, Mortgage, personal involves credit checks, employment check and verification. RPA could accelerate this process based on predefined rules and enable quick decision making.
- Fraud detection – RPA can quickly perform if-then analysis on the customer transactions and detect anomalies that could facilitate fraud detection.

Report Generation – RPA can help banks collect information from various sources, validate them, organize it in required format with high level of accuracy and share the reports with required stakeholders.

- Account opening / closure – Account opening, and closing involves a set of predefined tasks that could be automated with RPA. All known exception scenarios could be configured as well. RPA can maintain all required compliance audit logs throughout this process.

B. RPA in Audit

- Accounting/Auditing firms also deals with massive volume of data that are analysed/audited by a variety of processes that are repetitive & time consuming and that do not require audit judgment. RPA can help in variety of Audit processes including,
- Data Collection and Cleansing – Auditing involves collection of data from various sources and validate the data for completion, duplication etc. RPA can automate this data collection and manipulation / cleansing process with high level of accuracy.
- Controls testing – RPA could be utilized in automating a variety of controls testing including auditing

segregation of duties, exception reporting, access related controls, change management controls etc.

- Risk Assessment – RPA can automate the data collection, classification of data and identify core trends as part of the annual risk assessment process.
- Reconciliation – RPA can automate data collection from various sources and reconcile data against preconfigured rules with high level of accuracy. RPA can reconcile huge volume of data consistently which is real tedious work for the human partners.
- Audit PMO – RPA can handle a variety of tasks including identifying open items, sending reminders at required intervals / criteria, tracking the progress against plan, automating reports etc.

C. RPA in Insurance

- Insurance companies across the globe is heavily inundated with manual back office processes. This necessitates insurance companies to automate various processes to meet rapid customer growth and improve their processing time. Some of the potential automation processes includes,
- Claims processing – RPA can expedite the claims processing by gathering required information from various sources and integrate all this disparate information with high level of accuracy.
- Underwriting – RPA can collect information from various internal & external data sources and perform risk assessment quickly with high level of accuracy.
- Appeals processing – RPA can automate extracting data from policy and claims decision document in-order to expedite appeals processing.
- Data collection and cleansing – RPA can automate extracting required data from internal and external data sources. RPA can cleanse this data,

format and remove duplicates, fill required information in forms ensuring high data quality.

- New policy opening / policy cancellation – RPA can follow predefined rules and checklists to automate various processes involved in policy opening and cancellation. It also maintains detailed audit trail of all processes involved for compliance purposes.
- Business process analytics – RPA allows monitoring and recording every task it executes which gives valuable insights including number of documents processed, exceptions generated etc. This insight allows further process improvements and decision making.

D. RPA in Retail

- Retail industry utilizes numerous processes that are performed manually, and retail companies are on constant look out for technologies to improve their accuracy and productivity. RPA can help automating various retail processes including
- Customer Support – RPA can efficiently track orders and provide required updates to the customers at pre-set intervals. It can also automate feedback collection from customers ensuring customer satisfaction.
- Returns processing – RPA can help automating return management, inventory adjustment and customer credit process effectively.
- Product categorization & Pricing Comparison – RPA can assist harmonizing SKU data from various data sources and allows comparing price information to facilitate better decision making.
- Accounting & Finance – RPA can assist automating numerous manual tasks in accounting including Accounts payables, Accounts

receivables, reconciliation, financial closing, financial filing etc.

- Marketing analysis – RPA can be utilized in variety of analysis functions including, Campaign analysis, Consumer behaviour analysis etc.,
- Demand planning – RPA can automate collection of planning data, cleanse the data, run simulations, identify exceptions and communicate plan data to required stakeholders.
- Logistics and Supply chain – RPA can be utilized to monitor the inventory levels, track shipment information and communicate with customers / vendors based on preconfigured events. RPA can automate a variety of use cases in Inventory management, Freight management and Vendor management.

E. RPA in Manufacturing

- Manufacturing companies constantly look for technologies that improve speed, scale and simplify their manufacturing operations. RPA can help manufacturing industry in a variety of processes including,
- Bill of Materials – RPA can help in efficiently managing information from Bill of Materials data comprising of nested multilevel raw materials, components and subcomponents for new product creation.
- Customer service - RPA can help in resolving low priority customer queries by bringing in required data from various sources and can facilitate automated email communication with users
- Logistics – RPA can automate transport management effectively monitoring goods in transit. It can also provide valuable insights in terms of carriers, insurance companies etc based on the statistical data set.
- ERP Automation – RPA can facilitate automating ERP tasks like generating a variety of reports including

payables, receivables, pricing etc. It can also automate tasks across ERP functions not limited to Inventory management, Vendor management, Customer management, Accounts payables, Accounts receivables, GL, HR etc.

- Data Migration – RPA can assist in digitizing and migrating information from legacy data sources into ERP system with high level of accuracy.
- Manufacturing Analytics – RPA can provide detailed process statistics including successful execution, process bottlenecks and exceptions triggered. This valuable insight can help refining manufacturing processes.
- Compliance – RPA system would build audit trails of every process executed which significantly improves the compliance process.

III. FUTURE OF RPA

- Robotic process Automation evolved as futuristic tool in automating rule based repetitive business tasks. This helped business automate a variety of their business processes that utilizes structured data and matured business rules. With global demand for talented business users in the rise, tremendous growth in unstructured data, evolution of new business processes, companies are looking for solutions that could automate business processes that are non-rule-based utilizing structured / unstructured / semi-structured data. RPA technology is going through a radical change being embedded with Artificial Intelligence, Machine Learning and various other cognitive technologies. RPA embedded with above-said technologies is being referred as Cognitive Automation, Intelligent Automation, Smart

Process Automation etc. This evolution would enable RPA with,

- Processing of structured / unstructured / semi-structured data using variety of Cognitive technologies including OCR (Optical character recognition), Text analytics and data mining.
- Ability to integrate NLP (Natural language processing) capability to enhance human interaction which could expand its horizon and integrate well with human workforce.
- Built in Machine Learning, it can learn to adapt to / handle extensive list of scenarios than just pre-configured list of scenarios with business rules. With the integration of Machine Learning, AI, Big Data and Cloud, businesses would be able to significantly improve quality, operational scalability and employee efficiency.

IV. CONCLUSION

Robotic Process Automation is gaining momentum across industries and evolving every day as a technology of choice for all executives who want to bring speed and agility in their business operations. Organizations who wanted to transform their business processes and maximize the outcome of RPA, would need to break down their back-office processes and analyze specific use-cases that could be automated. RPA implementation is more than just a technology implementation, it not only automates business processes; it allows business to identify business process bottlenecks and simplify / standardize business processes. As the adoption of RPA matures, businesses can look for more use-cases to automate, redesign / refine existing implementation use-cases.

REFERENCES

- [1] S. C. Lin, L. H. Shih, D. Yang, J. Lin and J. F. Kung, "Apply RPA (Robotic Process Automation) in Semiconductor Smart Manufacturing," 2018 eManufacturing & Design Collaboration Symposium (eMDC), Hsinchu, Taiwan, 2018, pp. 1-3
- [2] IEEE Guide for Terms and Concepts in Intelligent Process Automation," in IEEE Std 2755-2017 , vol., no., pp.1-16, 28 Sept. 2017 doi: 10.1109/IEEESTD.2017.8070671.
- [3] Kevin C. Moffitt, Andrea M. Rozario, and Miklos A. Vasarhelyi (2018) Robotic Process Automation for Auditing. Journal of Emerging Technologies in Accounting: Spring 2018, Vol. 15, No. 1, pp. 1-10.
- [4] R. Issac, R. Muni and K. Desai, "Delineated Analysis of Robotic Process Automation Tools," 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAIECC), Bangalore, 2018, pp. 1-5. doi: 10.1109/ICAIECC.2018.8479511.
- [5] Appelbaum, D., A. Kogan, and M. A. Vasarhelyi. 2017a. Big Data and analytics in the modern audit engagement: Research needs. Auditing: A Journal of Practice & Theory 36 (4): 1–27. <https://doi.org/10.2308/ajpt5168410.2308/ajpt-51684>
- [6] Robotic process automation in the Finance function of the future. Available at: [https://www.ey.com/Publication/vwLUAsets/EY_-_Robotic_process_automation_in_the_Finance_function_of_the_future/\\$FILE/EY-robotic-process-automationin-the-finance-function-of-the-future-2016.pdf](https://www.ey.com/Publication/vwLUAsets/EY_-_Robotic_process_automation_in_the_Finance_function_of_the_future/$FILE/EY-robotic-process-automationin-the-finance-function-of-the-future-2016.pdf). [7] Cathy Tornbohm John E. Van Decker When and How to Use Robotic Process Automation in Finance and Accounting Gartner 2017.

Available at:
<https://www.gartner.com/doc/3834757/use-roboticprocess-automation-finance> [8]
Lacity, M., L. Willcocks, and A. Craig, A. 2015. Robotic Process Automation at Telefónica 02. Available at: FINAL.pdf
[9] H. Jalonen Assessing Robotic Process Automation Potential. Master of Science Tampere University of Technology 2018. Available at:
<https://dspace.cc.tut.fi/dpub/handle/123456789/24879>
[10] C. Tornbohm "Market Guide for Robotic Process Automation Software" Gartner.com 2017 [online] Available at:
<https://www.gartner.com/doc/3506217/marketguiderobotic-process-automation>.

Current Automation Trends-

A research on RPA development tool

Rashmita Rai, Prof. Deepesh Jagdale

Abstract- RPA or Robotic Process Automation is a smarter and easier way to direct work across those machines which can be categorized easily using AI (Artificial Intelligence). RPA (Robotics Process Automation) which allows organizations to automate task just like a human being was doing them across application and systems. In any organization, Robotic process automation can perform tasks like a human being. RPA is a program in which a series of commands are executed as per the predefined rules of a company or organization. In this paper, we are going to discuss about Robotic process automation, its working and tools. Some of the popular RPA tools such as Automation Anywhere, Blue Prism and UiPath will be discussed deeply. This paper will also focus on the benefits of using RPA over traditional processes in an organization.

Index Terms- Attended automation, Life cycle, RPA, RPA market, Tools.

◆

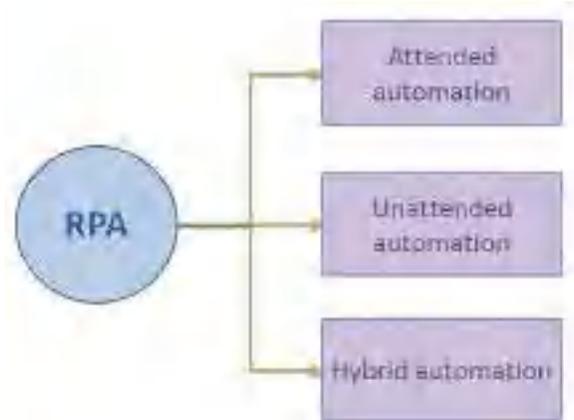
INTRODUCTION:

Robotic Process Automation (RPA) is a software technology which makes it easy for anyone to automate the tasks.[1] Using RPA software users create robot or bots which learns and mimics and then it executes rule based processes. You have to simply show the bots what is to be done and let them do their work. RPA bots can interact with applications just like humans do- except that the RPA bots can 24x7 round the clock, nonstop, with faster service and better reliability.

RPA can be used in multiple and complex tasks in various systems. Many organizations are adopting this new technology because it benefits the company in many ways. Organizations such as healthcare, HR, banking and financial services, travel and government have started using robotic process automation.

TYPES OF RPA:

- Attended automation
- Unattended automation
- Hybrid automation



1. Attended automation:

Attended Automation refers to the type of automation where the robot or the agent resides on the machine of the user and is invoked by the user at some instances. The triggering needs to happen by the user's action since the triggering points are programmatically difficult to detect. It makes the users feel good as they're still a component of the method but now not an element of the monotony. This improves productivity, not only by saving time but also by increasing their efficiency.

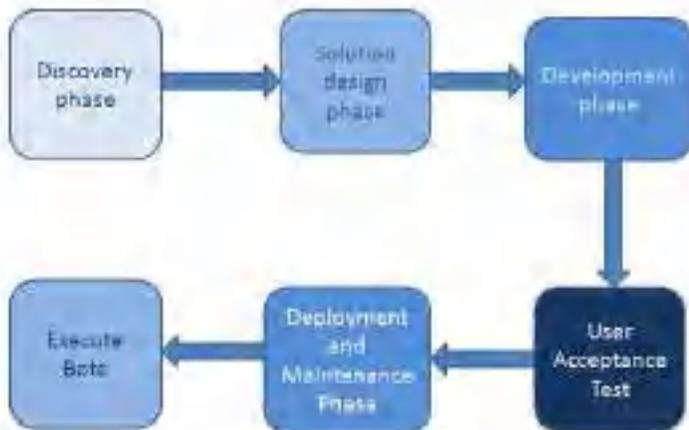
2. Unattended automation:

The nomenclature 'unattended' may not bring positive vibes about this sort of automation, but it elevates the awesomeness of RPA to a different level. Not all the tasks during a process must explicitly run – they'll run within the background, processing the essential data and giving the output. This might save plenty of your time for the back-end employees who don't need to cater to customers but more with data and processes.

3. Hybrid automation:

Another option that's emerging is Hybrid automation. This may include the utilization of unattended as a part of attended automation or when AI is employed with attended automation. The problem with most Attended automation is that they occupy the computer while the robot is executing. A far better approach is for the attended bot to perform the part that needs people's involvement and let an unattended bot perform the remainder. This is often how Hybrid automation can automate more parts of the task and more efficiently.[2]

RPA Life Cycle



a. Discovery Phase:

It is the first and important phase of the RPA implementation. In this phase, the requirements of the client are analyzed. The process can be automated or not is also decided in this phase

b. Solution Design Phase:

In this phase, according to the requirements given the steps to automate the task is designed.

c. Development Phase:

In this phase, the RPA developer creates bots which will help in automating the task.

d. User Acceptance Test:

Once the bots are created, these developed bots are tested.

e. Deployment and Maintenance Phase:

In this phase, the bots are deployed in the production environment.

f. Execute Bots:

In this phase, the bots are executed after its deployment.[3]

FUNCTIONAL COMPONENTS OF RPA:

The word of 'Robot' of Robotic Process Automation means software that runs on a physical or virtual machine. Such softwares help in configuring automation workflows which in turn automates business processes.

Following are the functional components of RPA:

- **Business process management (process):** It is used for designing and recording the automated processes.
- **Software robots (bots):** It is used to interact with apps and to execute automated processes.
- **Robot controller (orchestrator):** It is used for scheduling and monitoring the process and the bots.

RPA Tools:

RPA tools are the softwares that are used to configure different tasks to get automated and help the user to get the things done in an easier and efficient way. Most of the companies have periodic and repetitive tasks such as data updation, report generation, etc. which is manually performed by the users. Such tasks can be easily automated using these tools.

The most popular RPA tools are:

Property	UiPath	Blue Prism	Automation Anywhere
Robots	Front end and back end	Back end	Front end and back end
Learning	User friendly visual designer	Visual designer	Developer friendly
Use	BPO Automation	Desktop, Web and Citrix Automation	Achieve fair efficiency across all mediums
Cognitive Capability	Medium level of cognitive capability	Low level of cognitive capability	Medium level of cognitive capability
Technology	Based on numerous technology such as Sharepoint, cabana	Based on c-sharp	Based on Microsoft technologies
Recorder	Yes	No	Yes
Architecture	Web based orchestrator	Client server architecture	Client server architecture
Popularity	Most popular	Popular than Automation Anywhere	Least Popular
Pricing	Effective and entry level pricing	High cost of acquisition	Higher cost of deployment

Popular RPA tools [4]

RPA tools	Features
OutSystems	<ul style="list-style-type: none"> You can add your code using the language you know. Developers can build complex apps using reactive web, IoT, etc.
Eggplant	<ul style="list-style-type: none"> Developed specially for process experts and not developers and technical experts. It can test any system from mobile to mainframe using Universal fusion testing.
HelpSystems	<ul style="list-style-type: none"> IT and business teams are brought together using graphical workflow. Intelligent automation capabilities with advanced conditions and trigger scenarios.
Power Automate	<ul style="list-style-type: none"> To simplify business process automation, this application provides step by step guide. This automation tool is protected by MS security technology.
Pega	<ul style="list-style-type: none"> Open APIs that quickly connect to your existing services. It captures the insights at the desktop which helps in getting details of work gets done.
OpenConnect	<ul style="list-style-type: none"> Server Based architecture. Highly secure and encrypted data.
Contextor	<ul style="list-style-type: none"> It ensures that the tasks are properly sequenced. It can monitor the processes executed on the workstation.
Jacada	<ul style="list-style-type: none"> Reduced Average Handling Time of your organization. Helps you to increased first call resolution.

RPA Tools comparison with features

Kryon	<ul style="list-style-type: none">• Empower companies to accomplish their tasks with greater speed, accuracy, and efficiency.• Enhances productivity at every stage of your business processes.
Nice	<ul style="list-style-type: none">• Helps any organizations to take firm and accurate decisions based on advanced analytics of structured and unstructured data.• Better resource utilization.

RPA Tools with features [5]

Benefits of RPA:

1. Improved results in business:
Focusing employees on higher value added activities will lead to improvement in business metrics.
2. Reduced wage costs:
By enabling automation your personnel will be responsive to automation opportunities and keep departments with lesser number of employees.
3. Reduced risk:
RPA projects are low risk projects which can get completed quickly without any disturbance of existing systems
4. Faster services:
Employees that manually enter forms into systems or copying data between systems wait your service speed. Bots work on lightning speed without break.
5. Reduced workload
Automating reports is especially valuable for the analytics department, allowing them to focus on more sophisticated, custom analysis.

Conclusion:

In this paper we have discussed about Robotic Process Automation and its types in detail. The phases of RPA lifecycle have also been discussed. The Robot is able to process the cases in the same way a human can but it is done much faster and it also delivers consistent quality. Various RPA tools are being used to achieve faster results and better quality.

References:

- [1] <https://www.automationanywhere.com/rpa/robotic-process-automation>
- [2] <https://citrixready.citrix.com/uipath/uipath-robot.html#:~:text=Enterprise%20RPA%20platform,-.These%20digital%20workers%20execute%20the%20workflows%20you%20create%20in%20UiPath,work%20on%20higher%2Dlevel%20tasks.>
- [3] <https://medium.com/emorphis-technologies/robotic-process-automation-stages-of-rpa-life-cycle-e0f8180a11d6>
- [4] <https://www.javatpoint.com/uipath-vs-blue-prism-vs-automation-anywhere>
- [5] <https://www.guru99.com/robotics-process-automation-tools.html>

Application of RPA for E-mail Automation

Rashmita Rai, Prof. Deepesh Jagdale

Abstract-This paper presents the results of the analysis and also the methodology of RPA application. The research focuses on the methodology of robotic automation in automating email processing and segregation. RPA is a technology developed on software robots to perform certain repetitive tasks that allow them to regulate applications, collect data, feed them to other applications and manipulate data. This paper focuses on how the tedious and repetitive task of managing and segregating a college student's data can be automated using Robotic process automation. Students in a certain college are from different streams and department but the faculty receives all the students' emails within the same format. Managing and sorting out this data according to the department and emailing this data to the respective incharges can be tiring task. Also if we do this ask manually it will take more time and there is no guarantee on whether this work would be done in a proper manner. So to manage this, RPA is employed. The definition of the robotic automation of processes, the advantages of using RPA technology, stages of effective RPA implementation has also been mentioned in this paper.

Index Terms- Email Automation, Methodology, RPA.

I. INTRODUCTION

Robotic Process Automation (RPA) is one of the most innovative and trending technologies in the IT field. The purpose of RPA is to transfer the basic process from humans to bots. RPA communicates with the current IT architecture with no complicated system integration required. Robotic process automation can be used to automate labour intensive processes. It is used to perform repeating tasks without much of human interaction.[1] We can save money, time, labour and with that we can produce much more accurate result. We do not need to worry about the quality of the process since the chances of error is too less. Robotic process automation can be done by using the following tools: 1) UiPath 2) Automation Anywhere 3) Blue Prism. These are some of the trending tools used in RPA.

RPA is additionally utilized in Business purpose to increase the flexibleness, scalability, productivity increase, virtual workforce 24x7x365, Reduction of human errors, Regulatory compliance and

Government etc. RPA can save the 50-70% of the businesses cost. In banking and financial sector, RPA is used for report creation, data validation and migration between the customer management account and banking applications.

RPA is a developing methodology that utilizes programming robots to catch and decipher existing applications for preparing exchanges, controlling information and speaking with other software systems. These software robots are utilized to perform work that requires manual work and to automate redundant undertakings over numerous business applications without adjusting existing foundation and systems [2].

Healthcare systems contain many burdensome tasks that require considerable amount of resource allocation such as claim management. This leads to higher costs of operations and slow processes. Every industry has inefficiencies however only less number of industries faces the challenges of healthcare industry: strict regulations concerning patient data and fewer resources to deal with such regulations. RPA can be used in patient scheduling, claim

management, regulatory compliance, data entry, etc.

Automation is used to automate the manual processes in a single click. Robotic process can be defined as to execute in a click of a button. For Email Automation we are going to use UiPath tool.

UiPath is a Robotic Process

Automation tool that is used for Windows desktop automation. It is used to automate repetitive/redundant tasks with the assistance of drag and drop functionality and eliminates human intervention. This tool offers various editions to support different types of users and comes with a community to resolve issues. [3]

The UiPath Architecture includes the three UiPath components – Studio, Orchestrator, and Robot, along with the database servers and browser components.

1. **UiPath Studio:** It lets you build automation workflows with pre-built activities in a visual designer. It is used to create automation workflows. UiPath Studio includes: GUI Dashboard, 3 complexity levels, types of recorders, logging and exception handling, integrate with OCR technologies and reusable components.
2. **UiPath Robot:** It executes automation workflows created by the UiPath Studio. If you need to execute any of your tasks, you have to make sure that the UiPath Robot is in a running state. Single or multiple robots can run together at the same time.
3. **UiPath Orchestrator:** It is used to orchestrate the UiPath robots in continuously executing repetitive process on various platforms.



UiPath Component Architecture [4]

II. METHODOLOGY

UiPath has two types of bots those are:

- Attended bots
 - Unattended bots
1. **Attended bots:** Attended Robots help human workers speed front-office tasks. Triggered by user commands or needing human direction, Attended Robots operate in the background while users work on higher-level tasks. [5] Attended bots should not be granted to execute the tasks which users cannot. Any credentials that are required during the execution of this attended process should be known to the user and should be provided by them.
 2. **Unattended bots:** Unattended Robots run back-office tasks in a physical or virtual environment and can be scheduled to self-start. [5] These bots are used for more complex and repetitive tasks. These automations are usually performed in batches, which can be decided based upon some predefined rule. Since it is an unattended bot, we need to use it in processes which do not require human intervention.

Recorder:

It is used to create a basic framework for automation. Any action done by the user on

the screen is recorded by the recorder and which is turned into a recording sequence in the current project. In this way Robots are able to mimic human actions. There are four types of recording in UiPath Studio:

- Basic
- Desktop
- Web
- Citrix

There are two types of recording:

1. **Automatic recording:** This is used for recording multiple actions in one go. It is a good feature for automating a task. This recording can be invoked with Record icon which is available in desktop, web and basic recorders.
2. **Manual recording:** This is used for recording each step one at a time and hence has more control over the recording. Also, it can record all actions which automatic recording cannot record such as keyboard shortcuts, mouse hover, right-click, finding text from apps, and many other activities

Excel commands: These commands are one of the most powerful commands. Many processes can be done using commands such as:

- To copy data from one excel sheet to another.
- To move data from an Excel spreadsheet to another application.
- To transfer data from Microsoft Access to Excel.
- To transfer Excel data to a website form.
- To compare two cell's or two columns in an Excel spreadsheet.

- To sort data in Excel spreadsheets and deleting duplicate rows.

The Excel command contains some of the following sub-commands those are

- Open Spreadsheet
- Get Cells
- Set Cell
- Go to cell
- Delete Cells
- Activate Sheet
- Find/Replace
- Close Spreadsheet
- Save Spreadsheet

Variables in UiPath:

In UiPath, variables are used to represent fields such as numbers, text, array, e.t.c. Variables in UiPath are same as in any other programming languages. We can create, remove and manage variables. Various types of variables are:

1. **Text variables:** It is used to store text values.
2. **Number variables:** It is used to store Integer values.
3. **Boolean variables:** It is used to store boolean values.
4. **Array variables:** It is used to store an array of text or numbers.
5. **Data table variables:** It is used to store variables in the form of table.

Activities:

Activity is a unit of action performed by the bot. This activity is used to automate different applications. When all these activities come together it becomes a process. Some of the activities used in UiPath are:

- Message Box: It displays a message box with a text given that we need to show to the user.
- Assign Activity: This activity is used to assign a value to a variable.
- Write CSV Activity: It is used to save the specified DataTable to a .csv file.
- If activity: It has statement with two conditions: true or false. If the statement is true, then the first condition is executed; if not, the second condition is executed.
- The Switch activity: The Switch activity is used to make a choice. When we have various options available and we want to choose one option to execute, we frequently use the Switch activity.
- For Each Activity: It is used to iterate through each and every data so that it can be processed individually.

Screen Scraping

Screen Scraping is a method which extracts data from documents, websites, and PDFs. It is one of the powerful methods for extracting text. We can extract text using the Screen Scraper wizard. The Screen Scraper wizard has three scraping methods:

- Full Text
- Native
- OCR

Conclusion:

In this paper, we have proposed a method which will help the institution in managing the student's data using the RPA approach. This automation would help the incharge to analyze the student's data easily. This automation will also send emails to respective incharges. The main objective of

this automation is to reduce the manual repetitive tasks by using the RPA tool.

References:

[1] IRPA, "Introduction To Robotic Process Automation," tech. rep., 2015.

[2] M. C. Lacity, and L. P. Willcocks, (2016) "A new approach to automating services". MIT Sloan Management Review, Fall. ISSN 1532-9194

[3] <https://www.edureka.co/blog/what-is-uipath/>

[4] <https://dev.joget.org/community/display/DX7/Integration+with+UiPath+Robotic+Process+Automation>

[5] <https://citrixready.citrix.com/uipath/uipath-robot.html#:~:text=Enterprise%20RPA%20platform,-.These%20digital%20workers%20execute%20the%20workflows%20you%20create%20in%20UiPath,work%20on%20higher%20level%20tasks.>

Data Analysis: A Common review

Ajin shaji antoney

PG student at Pillai HOC college

Email: ajinshaji123@gmail.com

Mob:8793406170

Prof. deepesh jagdale

HOD of IT at Pillai HOC college

Email: djagdale@mes.ac.in

Mob:9028609874

Abstract— Data analysis is the process of analysing data with various methods and tools to get useful information from structured and unstructured data. Businesses to take proper decision need key information, internet as we know is creating vast amount of data, to search manually will be a big challenge and time consuming with proper data analysis method we can extract data from the source and present it in a readable format with visualisation tools in this way we can save time and take effective decision for running a smooth business. **Index Terms**— Data analysis, data collection data mining, inferential analysis, predictive analysis, diagnostic analysis

1 INTRODUCTION

As many top entrepreneurs has said data is the new oil and as common people get easy access to internet data is going to increase in a very high amount. as study shows there will be 175 billion terabytes of data generated by 2025.To analyse such big amount of data we need proper tools and method.

Data analysis is the procedure of cleaning, analysing, rendering, and visualizing data to learn valuable insights which will help business to make quick and effective decision. It helps to turn raw data into human readable format. It will help businesses to improve their services and product by analysing various types of data within structured and unstructured format. This process will give you detail insights about your customer base and will give better edge in the market. This will help you make more customer relatable product strategies and penetrate market and will also help them to focus more on return on investment. Instead of depending on on perception or understanding, analyzing data will provides you solid evidence to support conclusionsProduct teams, for example, frequently analyze client feedback to understand how clients interact with their product, what they're unsatisfied with, and what new features they would like to see. Then, they translate this insight into User experience improvements, new features, and improved functionalities

With data analysis, you can also spot the weaknesses and strengths of your opposition, detecting opportunities for enhancement. You can Analyse online reviews about your competition to answer questions like: what do clients love or hate about your competitors' products or services?

2 METHODS

The six main methods of data analysis are:

- Text Analysis
- Descriptive Analysis
- Inferential Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

2.1 Text Analysis

Text analysis, also termed text analytics or data mining, uses machine learning with natural language processing to organize unstructured data, so that it can be appropriately analyzed for valuable insights. This method is a form of qualitative analysis which is concerned with more than just statistics and mathematical values.

By converting human language into machine-readable format, text analysis tools can group text by topic, extract keywords, and read for sentiment and intent. It will tell us "*What is happening*" as specific, often independent data. It will give suggestions more in-depth and targeted views into *why* something may be happening, or *why* something happened. You can use text analysis to detect topics in client feedback, for example, and understand which aspects of your brand are important to your clients

2.2 Descriptive Analysis

Descriptive analysis delivers the "*What happened?*" scenario when analysing quantifiable data. This is the most basic and most common procedure of data analysis concerned with telling, summarizing, and identification of patterns through calculations of current data, like mean, median, mode, percentage, frequency, and range.

It is usually the baseline from which other data analysis starts. This method is very useful for producing things like revenue reports and KPI dashboards. However, as it is only concerned with statistical analysis and absolute numbers, This method will not be able to provide the reason or motivation for *why* and *how* those numbers developed.

2.3 Inferential Analysis

Inferential analysis generalizes or imagines about "*What happened?*" scenario by comparing data from groups within an entire inhabitant. it can include population of a country,

existing customer base, patients in a medical study, etc. The most common approaches for conducting inferential statistics are hypothesis tests and estimate theories. This method can be widely used in market research, to compare two variables in an attempt to reach a decision. It is also used to survey a sample set of the people in an attempt to extract information about the entire population. That's why it is necessary to properly select and calculate for representative sample of the population

2.4 Diagnostic Analysis

This method aims to answer "Why did ____ happen?" it can be also called root cause analysis it uses insights from statistical analysis to attempt to understand the cause of or reason behind these statistics. It is able to identify pattern and deviation within the data to answer why. This method can be helpful to understand client behaviour, to discover out which advertising campaigns actually surges sales, for example. There is sudden decrease in customer base you can question "why".

2.5 Predictive Analysis

This analysis uses known data to assume about future events. It deals with the scenario with "What is likely to happen." When Used in sales analysis, it often merge demographic data and purchase data with other data points to predict the activities of customers

For example, ability for certain business to exist in demographic area changes as the area changes. Or in a scenario where customers salary increases, theoretically, then the customer will be able to purchase more. The more data points you have on a certain demographic or specific customer, the more precise the prediction is likely to be.

2.6 Prescriptive Analysis

This method is the most advanced form of analysis, it work by combing all of your data and analytics, then give a output model prescription: *What action to take*. This model works to analyse multiple situations, it try to predict the outcome of each situation and decide the best action based on finding.

We can take AI as an example of prescriptive analysis which is the cutting edge of data analysis. It allow prescriptive analysis that can take and breakdown huge amount of data and efficiently teach itself how to use information and make own calculated decision

3 HOW TO ANALYSE YOUR DATA

To implement data analysis process you have to get most out of your data complexity of the process depend on the data type you have to complete the process successfully you have follow some rules and have to be disciplined which include setting goals, collection of data from the source, cleaning of the data and analysing data and at the last visualizing data with the understandable dashboards which make it easy to see pattern and trends

Now we will look in to seven steps to analyse data

1. **Define Your Goals:** Setting clear purposes is key and will help determine the type of data that you'll need to collect and analyse.

2. **Collect Your Data:** Data is all over the place, and you'll want to bring it all into one place ready for analysis. Whether you're gathering quantitative or qualitative data, Excel will be a great platform for storing your data, otherwise you could connect data sources on to your analysis tools via APIs and integrations.

3. **Clean Your Data:** It's likely that unstructured data will be got to be cleaned before analysing it to realize more accurate results. Get obviate the noise, like special characters, punctuation marks, stop words (and, too, she, they), HTML tags, duplicates, etc.

4. **Integration of Data Analysis Tools:** Most tools can easily be integrated via APIs and one-click integrations. If using an API, you would possibly need a developer's help to line it up. Once connected, your data can run freely through data analysis tools.

5. **Analyse Your Data:** Now that you've connected data analysis tools, you'll got to choose the analysis type you would like to perform (which you almost certainly defined in step one).

6. **Visualize Your Data:** Dashboards are an excellent thanks to aggregate your data, and make it easy to identify trends and patterns. Some data analysis tools have in-built dashboards otherwise you can hook up with your existing BI tools.

7. **Draw Conclusions:** Gain key insights and make databased decisions by digging into your data from every angle

4 DATA ANALYSIS TOOLS

There are many data analysis tools you'll start with, counting on your technical skills, budget, and sort of knowledge you would like to research. They're also an excellent way of speeding up manual and time-consuming data analysis, which is vital if you would like up-to-the-minute insights to form fast decisions.

Here's a fast rundown of the highest data analysis tools which will assist you perform everything from text analysis to data visualization

SQL: SQL is the perfect tool to fetch the structured data from an organisation and is the most used tool in these field

Microsoft Excel: Filter, organize, and visualize quantitative data. The perfect tool for performing simple data analysis.

Tableau: A powerful analytics and data visualization platform. Connect all of your data and make interactive dashboards that update in real-time.

Python: The preferred programming language for machine learning. Use it to create data analysis solutions for various use cases.

5 USE CASES

Some of the top applications for data analysis:

- Automatically analyze survey responses with text analysis

- Analyze customer support tickets and automatically route them This method uses diagnostic method to understand why

- Categorize potential customers
- Examine the success of marketing campaigns Predict customer churn.

5.1 Automatically analyse survey responses with text analysis

This application uses text analysis method to conduct survey at qualitative results, beyond simple Yes/No or multiplechoice questions you are also able to ask open-ended questions you have to organize them by subject or theme and automatically analyse the feelings and opinion of the responder.

Marketing and research consultancy, analysed 13,500 multiparagraph responses with text analysis solution to count public opinion around large infra-structure project. This method allowed them hundreds of employee hours and company was able to give powerful insights that help them to make optimal business decisions.

5.2 Analyse customer support tickets and automatically route them

This method allows advance data analysis tools to read automatically from customer support tickets, whether it can be phone calls, emails and chatbots. It has a feature to rate them for urgency and route them automatically to correct department or specific employee. A company which offers technology based financial services had always have urgency to answer there customer inquiry due to sensitive work nature by exercising top classification method to their accurate needs these company successively increased their initial answering time by 70% and increases employees ticket volume by 25%.

5.3 Categorize potential customers

We can use customer analysis method like survey analysis, inferential analysis and diagnostic analysis, with all this technique you are able to understand present customer base we can further categorize them into sub group and identify and analyse there needs, we can further use this for perspective building analysis model and find potential new customer. we can use key word extraction technique and can get minute information of your potential customer

5.4 Examine the success or failure of marketing campaigns

specific advertising campaigns works and others don't and we will also extract real time reaction to your campaigns on social media or reviews. What can be keywords used to specify brands a product it will help you to maximize customer engagement by monitoring social media

5.5 Predict customer churn

This method uses predictive analysis to measure customer churn by analysing demographic and purchase date. It will help you to analyse whether your customers are aging out of your brand this will help you to keep a keen eye on you customer base by using predictive analysis with these techniques you can avoid getting churn or locate your business to new area for better growth.

6 CONCLUSIONS

In this report we have explored data analysis as processoriented technology we have seen dieffernt types of methods to explore data according to bussinesss needs. we have seen different tools and technology to extract data from structured and unstructured format. We have also seen diffrent use case for data analysis.

We can conclude that data analysis as a process is very powerful technology if used properly a business can maximise their profits by understanding there customer base and act according to there needs

7 Acknowledgments

I would like to thank mr deepesh jagdale HOD of MSC IT at pillai hoc for being my guide for this paper. I would also like to thank mr Ravindra Mhatre for teaching me the subject data science which act as the foundation for this research paper

8 REFERENCES

- [1] Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses" by Michael Minelli
- [2] Big Data, Data Mining and Machine Learning (WILEY Big Data Series)" by Jared Dean
- [3] Analytics: Data Science, Data Analysis and Predictive Analytics for Business" by Daniel Covington
- [4] Elements_and_Principles_of_Data_Analysis

Data Mining: Use case Review

Ajin shaji antoney

PG student at Pillai HOC college

Email: ajinshaji123@gmail.com

Mob:8793406170

Prof. deepesh jagdale

HOD of IT at Pillai HOC college

Email: djagdale@mes.ac.in

Mob:9028609874

Abstract— Revolution in the field of data has led to an increase in the adaptation of data mining the evolving technologies of information gathering and storage have made available huge amounts of data within most application domains, such as the business world, the scientific and medical community, and public administration. The set of activities involved in the analysis of these large databases, usually with the purpose of extracting useful knowledge to support decision making, has been referred to in different ways, such as data mining, knowledge discovery, pattern recognition and machine learning. In particular, the term data mining indicates the process of exploration and analysis of a dataset, usually of large size, in order to find regular patterns, to extract relevant knowledge and to obtain meaningful recurring rules. Data mining plays an ever-growing role in both theoretical studies and applications. In this chapter we wish to describe and characterize data mining activities with respect to investigation purposes and analysis methodologies. The relevant properties of input data will also be discussed. Finally, we will describe the data mining process and its articulation in distinct phases.

Index Terms— Data analysis, data collection data mining, inferential analysis, predictive analysis, diagnostic analysis

----- ◆ -----

1 INTRODUCTION

In this world we can say that data is the building block upon which any organisation thrives. Due to advancement in technology surge of data and information has increased rapidly. Advancement in storage capacity and data collection big amount of data is easily available each second data is being created and needs to be stored and analysed to extract value. Data has also become cheaper to store so organisation tries to gather as much of data they can size and variety and change of such data require new type of data analytics method. We can also define this as the union of emerging methods, technology, or languages, including such tools as R, python, SQL, Hadoop

Data mining activities constitute an iterative process aimed at the analysis of large databases, with the purpose of extracting information and knowledge that may prove accurate and potentially useful for knowledge workers engaged in decision making and problem solving., the analysis process is iterative in nature since there are distinct phases that might imply feedback and subsequent revisions. Usually such a process represents a cooperative activity between experts in the application domain and data analysts, who use mathematical models.

The term data mining refers therefore to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired. The term mathematical learning theory is reserved for the variety of mathematical models and methods that can be found at the core of each data mining analysis and that are used to generate new knowledge.

2 MODELS AND METHODS FOR DATA MINING

Data mining, classical statistics and OLAP Data mining projects differ in many respects from both classical statistics and OLAP analyses. Such differences are shown in Table 5.1, with reference to an example. The main difference consists of the active orientation offered by inductive learning models, compared with the passive nature of statistical techniques and OLAP. Indeed, in statistical analyses decision makers formulate a hypothesis that then has to be confirmed on the basis of sample evidence. Similarly, in OLAP analyses knowledge workers express some intuition on which they base BUSINESS INTELLIGENCE between OLAP, statistics and data mining OLAP statistics data mining extraction of details and aggregate totals from data verification of hypotheses formulated by analysts identification of patterns and recurrences in data information validation knowledge distribution of incomes of home loan applicants analysis of variance of incomes of home loan applicants characterization of home loan applicants and prediction of future applicants extraction, reporting and visualization criteria. Both methods – on one hand statistical validation techniques and on the other hand information tools to navigate through data cubes – only provide elements to confirm or disprove the hypotheses formulated by the decision maker, according to a top-down analysis flow. Conversely, learning models, which represent the core of data mining projects, are capable of playing an active role by generating predictions and interpretations which actually represent new knowledge available to the users. The analysis flow in the latter case has a bottom-up structure. In particular, when faced with large amounts of data, the use of models capable of play-

ing an active role becomes a critical success factor, since it is hard for knowledge workers to formulate a priori meaningful and well-founded hypotheses.

3 REPRESENTATION OF INPUT DATA

In most cases, the input to a data mining analysis takes the form of a two dimensional table, called a dataset, irrespective of the actual logic and material representation adopted to store the information in files, databases, data warehouses and data marts used as data sources. The rows in the dataset correspond to the observations recorded in the past and are also called examples, cases, instances or records. The columns represent the information available for each observation and are termed attributes, variables, characteristics or features. The attributes contained in a dataset can be categorized as categorical or numerical, depending on the type of values they take on.

Categorical. Categorical attributes assume a finite number of distinct values, in most cases limited to less than a hundred, representing a qualitative property of an entity to which they refer. Examples of categorical attributes are the province of residence of an individual (which takes as values a series of names, which in turn may be represented by integers) or whether a customer has abandoned her service provider (expressed by the value 1) or remained loyal to it (expressed by the value 0). Arithmetic operations cannot be applied to categorical attributes even when the coding of their values is expressed by integer numbers.

Numerical. Numerical attributes assume a finite or infinite number of values and lend themselves to subtraction or division operations. For example, the amount of outgoing phone calls during a month for a generic customer represents a numerical variable. Regarding two customers A and B making phone calls in a week for 27 and 36 respectively, it makes sense to claim that the difference between the amounts spent by the two customers is equal to 9 and that A has spent three fourths of the amount spent by B. Sometimes a more refined taxonomy of attributes can prove useful.

Counts. Counts are categorical attributes in relation to which a specific property can be true or false. These attributes can therefore be represented using Boolean variables {true, false} or binary variables {0,1}. For example, a bank's customers may or may not be holders of a credit card issued by the bank.

Nominal. Nominal attributes are categorical attributes without a natural ordering, such as the province of residence.

Ordinal. Ordinal attributes, such as education level, are categorical attributes that lend themselves to a natural ordering but for which it makes no sense to calculate differences or ratios between the values.

Discrete. Discrete attributes are numerical attributes that assume a finite number or a countable infinity of values.¹

4 DATA MINING PROCESS

The definition of data mining given at the beginning of Section 1 is an iterative process, during which learning models.

4.1 Definition of objectives.

Data mining analyses are carried out in specific application domains and are intended to provide decision makers with useful knowledge. As a consequence, intuition and competence are required by the domain experts in order to formulate plausible and well-defined investigation objectives. If the problem at hand is not adequately identified and circumscribed one may run the risk of thwarting any future effort made during data mining activities. The definition of the goals will benefit from close cooperation between experts in the field of application and data mining analysts. It is possible to define the problem and the goals of the investigation as the analysis of past data and identification of a model so as to express the propensity of customers to leave the service (churn) based on their characteristics, in order to understand the reasons for such disloyalty and predict the probability of churn.

4.2 data gathering and integration.

Once the objectives of the investigation have been identified, the gathering of data begins. Data may come from different sources and therefore may require integration. Data sources may be internal, external or a combination of the two. The integration of distinct data sources may be suggested by the need to enrich the data with new descriptive dimensions, such as geomarketing variables, or with lists of names of potential customers, termed prospects, not yet existing in the company information system. In some instances, data sources are already structured in data warehouses and data marts for OLAP analyses and more generally for decision support activities. These are favorable situations where it is sufficient to select the attributes deemed relevant for the purpose of a data mining analysis. There is a risk, however, that, in order to limit memory uptake, the information stored in a data warehouse has been aggregated and consolidated to such an extent as to render useless any subsequent analysis. For example, if a company in the retail industry stores for each customer the total amount of every receipt, without keeping track of each individual purchased item, a future data mining analysis aimed at investigating the actual purchasing behavior may be compromised.

4.3 Exploratory analysis.

In the third phase of the data mining process, a preliminary analysis of the data is carried out with the purpose of getting acquainted with the available information and carrying out data cleansing. Usually, the data stored in a data warehouse are processed at loading time in such a way as to remove any syntactical inconsistencies. For example, dates of birth that fall outside admissible ranges and negative sales charges are detected and corrected. In the data mining process, data cleansing occurs at a semantic level. First of all, the distribution of the values for each attribute is studied, using histograms for categorical attributes and basic summary statis-

tics for numerical variables. In this way, any abnormal values (outliers) and missing values are also highlighted.

4.4 Attribute Selection.

In the subsequent phase, the relevance of the different attributes is evaluated in relation to the goals of the analysis. Attributes that prove to be of little use are removed, in order to cleanse irrelevant information from the dataset. Furthermore, new attributes obtained from the original variables through appropriate transformations are included into the dataset. For example, in most cases it is helpful to introduce new attributes that reflect the trends inherent in the data through the calculation of ratios and differences between original variables. Exploratory analysis and attribute selection are critical and often challenging stages of the data mining process and may influence to a great extent the level of success of the subsequent stages.

4.5 Model development and validation.

Once a high-quality dataset has been assembled and possibly enriched with newly defined attributes, pattern recognition and predictive models can be developed. Usually, the training of the models is carried out using a sample of records extracted from the original dataset. Then, the predictive accuracy of each model generated can be assessed using the rest of the data. More precisely, the available dataset is split into two subsets. The first constitutes the training set and is used to identify a specific learning model within the selected class of models. Usually, the sample size of the training set is chosen to be relatively small, although significant from a statistical standpoint – say, a few thousand observations. The second subset is the test set and is used to assess the accuracy of the alternative models generated during the training phase, in order to identify the best model for actual future predictions. The most popular classes of learning models will be discussed in detail in the following chapters.

4.6 Prediction and interpretation.

Upon conclusion of the data mining process, the model selected among those generated during the development phase should be implemented and used to achieve the goals that were originally identified. Moreover, it should be incorporated into the procedures supporting decision-making processes so that knowledge workers may be able to use it to draw predictions and acquire a more in-depth knowledge of the phenomenon of interest.

5 APPLICATIONS OF DATA MINING

Data mining methodologies can be applied to a variety of domains, from marketing and manufacturing process control to the study of risk factors in medical diagnosis, from the evaluation of the effectiveness of new drugs to fraud detection. Relational marketing. Data mining applications in

the field of relational marketing, described in Chapter 13, have significantly contributed to the increase in the popularity of these methodologies. Some relevant applications within relational marketing are:

- identification of customer segments that are most likely to respond to targeted marketing campaigns, such as cross-selling and up-selling;
- identification of target customer segments for retention campaigns;
- prediction of the rate of positive responses to marketing campaigns;

Fraud detection.

Fraud detection is another relevant field of application of data mining. Fraud may affect different industries such as telephony, insurance (false claims) and banking (illegal use of credit cards and bank checks; illegal monetary transactions). Risk evaluation. The purpose of risk analysis is to estimate the risk connected with future decisions, which often assume a dichotomous form. For example, using the past observations available, a bank may develop a predictive model to establish if it is appropriate to grant a monetary loan or a home loan, based on the characteristics of the applicant.

Text mining.

Data mining can be applied to different kinds of texts, which represent unstructured data, in order to classify articles, books, documents, emails and web pages. Examples are web search engines or the automatic classification of press releases for storing purposes. Other text mining applications include the generation of filters for email messages and newsgroups. Image recognition.

The treatment and classification of digital images, both static and dynamic, is an exciting subject both for its theoretical interest and the great number of applications it offers. It is useful to recognize written characters, compare and identify human faces, apply correction filters to photographic equipment and detect suspicious behaviors through surveillance video cameras.

Web mining.

Web mining applications, which will be briefly considered in section 13.1.9, are intended for the analysis of so-called clickstreams – the sequences of pages visited and the choices made by a web surfer. They may prove useful for the analysis of e-commerce sites, in offering flexible and customized pages to surfers, in caching the most popular pages or in evaluating the effectiveness of an e-learning training course.

Medical diagnosis.

Learning models are an invaluable tool within the medical field for the early detection of diseases using clinical test results. Image analysis for diagnostic purpose is another field of investigation that is currently burgeoning. API, you would possibly need a developer's help to line it up. Once connected, your data can run freely through data analysis tools.

6 CONCLUSIONS

In this report we have explored data mining as process-oriented technology we have seen different types of methods to explore data according to business needs. We have seen different tools and technology to extract data from structured and unstructured format. We have also seen different use cases for data mining.

We can conclude that data mining as a process is very powerful technology if used properly a business can maximise their profits by understanding their customer base and act according to their needs

7 Acknowledgments

I would like to thank Mr. Deepesh Jagdale HOD of MSC IT at Pillai HOC for being my guide for this paper. I would also like to thank Mr. Ravindra Mhatre for teaching me the subject data science which acts as the foundation for this research paper

8 REFERENCES

- [1] Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses" by Michael Minelli
- [2] Big Data, Data Mining and Machine Learning (WILEY Big Data Series)" by Jared Dean
- [3] Analytics: Data Science, Data Analysis and Predictive Analytics for Business" by Daniel Covington
- [4] Elements_and_Principles_of_Data_Analysis

Voice Based Intelligent Virtual Assistance For Windows

Alisha Bhoir, Prof. Deepesh Jagdale

Abstract- In the 21st century, everything is automated including the things we use daily washing machines, dishwashers, refrigerators, bus doors, air condition systems, turning everything in a single click, etc. In this fast-moving the present study proposes the newer concept of voice-controlled devices that recognizes one's voice; process the request and assigns the time and date of the appointment based on the request with details as name of the person; date; time and other related information. Communication and Technology has a long history but still constantly and actively growing and changing. The technology changes so fast that today everybody has an AI Personal assistant. A personal voice assistant is the software that can perform task and provide different services to the individual as per the individual's dictated commands. This is done through a synchronous process involving recognition of speech patterns and then, responding via synthetic speech. For Example, if you ask the device to change the wallpaper of your Personal Computer it will change the wallpaper by downloading wallpaper from a website and changing the wallpaper. It can also guide you the traffic between source and destination and also auto-suggest lesser traffic and time routes. The attempt has been made to develop an "Intelligent Personal Voice Assistant using Python" which helps people to control device with their voice(speech), extract information and perform tasks on their desktop.

Keywords: Personal Voice Assistants, Speech Recognition, Artificial Intelligence, automated, dishwashers, recognizes, crowd, website, destination.

◆

INTRODUCTION

A Voice-based Assistant is an artificial intelligence system that emulates human interaction to perform tasks. This project is done with speech recognition (changes text to speech or vice versa), coding (programming the activities for the device to perform and device accessing), & creativity (adding up tasks and creating healthy interaction user-friendly environment. The Voice recognition technology system was developed before the Apple's Siri was introduced in 2012. The usage of virtual assistants is expanding rapidly after 2017, more and more products are coming into the market.

In today's era almost all tasks are digitalized. We have Smartphone in hands and it is nothing less than having world at your fingertips. These days we aren't even using fingers. We just speak of the task and it is done. There exist systems where we can say Text Dad, "I'll be late today." And the text is sent. That is the task of a Virtual Assistant. It also supports specialized task such as booking a flight, or finding cheapest book online from various e-commerce sites and then providing an interface to book an order are helping automate search, discovery and online order operations. Some examples of commonly known assistants: Siri, Google Assistant, Alexa or Cortana.

Speech recognition is an alternative to keyboard typing. Simply put, you're talking to the machine and your words show on the screen. It has been developed to provide a simple way to write on a computer and can support people with a range of disabilities. It is helpful for users with hand disabilities who often find it though, impossible or painful to type. Voice-recognition apps can

-
- Alisha Bhoir is currently pursuing masters degree program in M.Sc IT in Mumbai University, India, 7378584813. E-mail: alishaanbh2019mscit@students.mes.ac.in
 - Prof. Deepesh Jagdale is currently professor masters degree program in M.Sc IT in Mumbai University, India, 9028609874. E-mail: djagdale@mes.ac.in

also support people with spelling problems, including those with dyslexia, since well-recognized words are almost always pronounced correctly. Scientists have used text generated online by people to train voice assistants to listen and respond to our requests in a more natural and meaningful way. Voice assistants will decipher questions that have been phrased in a variety of different ways and interpret what the user is most likely to want.

This project was started on the premise that there is sufficient amount of openly available data and information on the web that can be utilized to build a virtual assistant that has access to making intelligent decisions for routine user activities.

Here are some basic definitions, similarities, and differences:

Intelligent Personal Assistants (IPA):

This type of software can assist users with some basic tasks, usually using natural language. Intelligent personal assistants are also so smart that they go online and search for an answer to a user's question. It may text or voice either of them trigger an action.

Automated Personal Assistant:

Automated means the task is performed by itself. The personal assistants are using AI and deep learning according to the user's experience and behavior towards the IPA they are able to do some automated tasks.

Smart Assistant: It is usually refers to the types of physical devices that can provide various advance features and services by using smart speakers that listen for a wake up word to become active and can perform certain tasks. Amazon Echo, Google Home, and Apple HomePod are examples of smart assistant's devices.

Chatbot: Its function is similar to its name it uses text as medium to communicate and provide information and do task for user. Chatbots can imitate a conversation with a

human user. Today most of the banking companies use them in the customer service sector or managing the accounts to answer basic questions and there is not to worry it also connect with an online person if bot is not able to resolve your query.

Voice Assistant: The input key here is our voice. It is a digital assistant that uses voice recognition, speech synthesis, and natural language processing (NLP) and also AI to provide an amazing service through an application exam Siri, Ok Google. Cortana etc.

BACKGROUND

There already exist a number of desktop virtual assistants. A few examples of current virtual assistants available in market are discussed in this section along with the tasks they can provide and their drawbacks.

i) SIRI from Apple:

SIRI is personal assistant software that interfaces with the user thru voice interface, recognizes commands and acts on them. It learns to adapt to user's speech and thus improves voice recognition over time. It also tries to converse with the user when it does not identify the user request. It integrates with calendar, contacts and music library applications on the device and also integrates with GPS and camera on the device.

Supported Tasks:

- Call someone from my contacts list
- Launch an application on my iPhone
- Send a text message to someone
- Set up a meeting on my calendar for 9am tomorrow
- Set an alarm for 5am tomorrow morning
- Play a specific song in my iTunes library
- Enter a new note

Drawback:

SIRI does not maintain a knowledge database of its own and its understanding comes from the information captured in domain models and data models.

i) ReQall:

ReQall is personal assistant software that runs on smartphones running Apple iOS or Google Android operating system. It helps user to recall notes as well as tasks within a location and time context. It records user inputs and converts them into commands, and monitors current stack of user tasks to proactively suggest actions while considering any changes in the environment. It also presents information based on the context of the user, as well as filter information to the user based on its learned understanding of the priority of that information.

Supported Tasks:

- Reminders
- Email
- Calendar, Google Calendar
- Outlook
- Evernote
- Facebook, LinkedIn
- News Feeds

Drawback:

Will take some time to put all of the to-do items in you could spend more time putting the entries in than actually doing the revision.

METHODOLOGY

1. Speech recognition:

The proposed system used the google API to convert input speech into text. The speech is given as an input to google cloud

for processing, as an output, the system then receives the resulting text.

2. Backend work:

At backend the python gets the output from speech recognition and after that it identifies whether the command is a system command or a browser command. The output is sent back to python backend to give desired output to user.

3. Text to speech:

Text to speech, or TTS, is a new wave technique of for transforming voice commands into readable text. Not to mix it up with VR Systems that instead, generate speech by joining strings gathered in an exhaustive DB of pre-installed text and have been developed for different goals which form full-fledged sentences, clauses or meaningful phrases through a dialect's graphemes and phonemes. Such systems have their limits as they can only determine text on the basis of pre-determined text in their databases TTS systems, on the other hand, are practically to "read" strings of characters and dole out resulting sentences, clauses and phrases.

VOICE RECOGNITION SYSTEMS

It is the heart of a Voice application system, which has ability to understand voice input given by user, and make application work in a efficient way and generating voice feedback to the user. This system is an important component for user as a gateway to use his or her voice as an input component. In a Nutshell, for clearly understanding user voice command and to get feedback from the system, we should consider voice recognition system contains all the process by which application system directs for building speech signals to text data and few forms of important meaning of speech.

FEATURES

Accessibility (creating ability to access information or data from many sources to provide a unique solution for the user command), Availability (The quality of having the ability to be used and being available all the time for the user, Creating a tool that responds to the user depending on the command inputs), Portability (Easy to hold the device around and giving a pretty design for looks) and last but not the least Efficiency (Device which takes less time to retort to the commands and present the results). Voice assistants can perform a range of actions after hearing a wake word or voice command. They can answer questions, play music, navigate to different apps supporting the user's request, read recipes, etc. Making the program more humanized and easier to use, increase the database capacity, and add more possible keywords, responses, and data during this program, interface optimization, etc.

PROPOSED ARCHITECTURE

The system design consists of

1. Taking the input as speech patterns through microphone.
2. Audio data recognition and conversion into text.
3. Comparing the input with predefined commands
4. Giving the desired output

APPLICATIONS

There are a wide variety of services which are provided by the voice-enabled devices which range from simple commands like providing information about the weather of a place, general information from Wikipedia, movie rating from IMDB, setting an alarm or reminder, creating a to-do list and adding items to the shopping list so that we don't forget when we go shopping. It can also read books for the user or else play music from any streaming services depending on the device provider

or user preference. It can also play videos from YouTube or else from any streaming services.

RESULT

According to the overall description in the context, the purpose of the project is to develop an intelligent voice assistant with the functionalities as calling services, message transformation, mail exchange, location services, YouTube service, checking weather, searching engine (Google, Wikipedia), translator and help menu. We can have this assistant implemented either in phones or PC's or they can be in the form of speaker devices.

The common way of communication used by people in daily life is through speech. If the mobile phone or the device the voice Assistant implemented can listen to the user for the requestor handle the daily affairs, then give the right response, it will be much easier for users to communicate with their phone, and the mobile phone will be much "Smarter" as a human assistant. This project is focusing on voice control (recognition, generating and analyzing corresponding commands, intelligent responses automatically), Google products and relevant APIs (Google map, Google weather, Google search, etc.), Wikipedia API references ranging from Speech-To-Text, Text-To-Speech technology. As all those functionalities and services for the project have been explained, the main structure and construction of the project have been illustrated with its goals.

CONCLUSION

In this paper the design and implementation of an Intelligent Personal Voice Assistance is described. The project is built using available open-source software modules with visual studio code community backing which can accommodate any updates in future. The modular approach used in this project makes it more flexible and easier to integrate additional modules and features

without disturbing the current system functionalities. It not only works on human commands but also it is designed for give responses to the user on the basis of query being asked or the words spoken by the user such as opening tasks and operations. This Intelligent Voice Assistant has an enormous and limitless scope in the future. Like Siri, Google Now and Cortana most popular personal voice assistants. The project will easily able to integrate with devices near future for a Connected Home using Internet of Things, voice command system and computer vision.

REFERENCE

1. Hirschberg, Julia, and Christopher D. Manning. "Advances in Natural Language Processing." *Science* 349(6245):261–266. doi:10.1126/science.aaa8685.

2. Moore, Clayton. "The Most Useful Skills for Google Home." *Digital Trends* (May 3, 2017). <https://www.digitaltrends.com/home/google-home-most-useful-skills/>.

3. <https://www.engadget.com/2017/05/17/google-assistant-iphone-hands-on/>.

4. Webopedia, web page [Online]. Available: <https://www.webopedia.com/TERM/I/intelligent-personal-assistant.html>

5. Statista, web page [Online]. Available: <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/>

6. Smartsheet web page [Online]. Available: <https://www.smartsheet.com/voice-assistants-artificial-intelligence>

Voice Based Intelligent Virtual Assistance For Windows

Alisha Bhoir, Prof. Deepesh Jagdale

Abstract-Communication and Technology has a long history but still constantly and actively growing and changing. The technology changes so fast that today everybody has an AI Personal assistant. Most of us have it on our phones in the form of Google assistant or Siri or Bixby. The use of voice based personal assistants is increasing day by day and helping in making our life simple. This paper present intelligent voice assistant with ability to organize and maintain information it includes the management of emails, calendar events, files and to do lists. The attempt has been made to develop an "Intelligent Personal Voice Assistant using Python" which helps people to control device with their voice(speech), extract information and perform tasks on their desktop. For Example, if you ask the device to change the wallpaper of your Personal Computer it will change the wallpaper by downloading wallpaper from a website and changing the wallpaper. It can also guide you the traffic between source and destination and also auto- suggest lesser traffic and time routes.

Keywords: Personal Voice Assistants, Speech Recognition, Artificial Intelligence, voice commands, desktop.

INTRODUCTION

This system is voice-based personal assistant has always seemed a little out of place in the enterprise. It's a useful tool for search, for reminders, and to write the note just by speaking it up. Window assistant is to create voice apps for the intelligent assistant. When user need to open any other application, he/she can use the command open. E.g. Open Notepad, File explorer, goggle chrome, this will help to open the applications. When user want to write the message can use command write. And to for searching purpose search command can be used. If we want to access any document or file we can do it by voice recognition. But there are certain limitations that the user should set some voices as a security options. If the user spells out the word it automatically types in the required

field. It recognizes the speech and searches the appropriate content in the database and retrieves it. The user should select the appropriate language for the virtual assistant to understand. If any wrong or invalid communication happens it invokes some messages in dialog box. It is like a software agent which performs tasks and events based on commands. Voice-Command and speech synthesis are enhancing the level of user-interaction in applications. This trend is now approaching business-applications like ERP-Systems. Implementing an intelligent personal assistant (IPA) will empower the application not only by navigating users through the system, but it also enables the option to navigate and explain data to the users through speech synthesis. In this paper, the architecture for a new IPA was worked out and a prototype for a web-based ERP-System resting upon this architecture has been developed.

Interactions between a user and your Window assistant skill are mostly free-form, so assistant must understand language naturally and also in context.

-
- *Alisha Bhoir is currently pursuing masters degree program in M.Sc IT in Mumbai University, India, 7378584813. E-mail: alishaanbh2019mscit@students.mes.ac.in*
 - *Prof. Deepesh Jagdale is currently proffesor masters degree program in M.Sc IT in Mumbail University, India, 9028609874. E-mail: djagdale@mes.ac.in*

Window assistant determines what a user wants to do by identifying the user intent from spoken or textual input by utterance. The intent maps utterances to actions that Window assistant can take, such as invoking a dialog.

REVIEW OF LITERATURE

The proposed system was based on the voice recognition in commands which converts the speech to text. The user gets easy access to speech than getting typed in the keyboard. The user spells the commands in the voice so that the user should have good quality of microphones. The user should spell the word in correct pronunciation which should be understandable to the voice recognition system. But there exists some confusion to the system for example the words sun and son have same pronunciation so the system finds difficult to choose which word. So that it provides some powerful microphones to be audible to the system.

The voice is the convenient and efficient mode of communication. Most of the people prefer to use speech rather than using text based. The basic overview of the proposed system is it gets input signal in the form of voice. It passes on to the feature extraction and gets into the decoder. In the decoder it consists of two main models, they are acoustic model and language model. The decoder will decode the input. After all the processing going in the decoder it gives the specified output. The classification of speech recognition system are types of speech utterance, types of speaker model, types of vocabulary.

IMPLEMENTATION OF SPEECH TO TEXT COMMAND EXECUTION

1. Speech recognition:

The proposed system used the google API to convert input speech into text. The speech is given as an input to google cloud for processing, As an output, the system then receives the resulting text.

2. Backend work:

At backend the python gets the output from speech recognition and after that it identifies whether the command is a system command or a browser command. The output is send back to python backend to give desired output to user.

3. Text to speech:

Text to speech, or TTS, is a new wave technique of for transforming voice commands into readable text. Not to mix it up with VR Systems that instead, generate speech by joining strings gathered in an exhaustive DB of pre-installed text and have been developed for different goals which form full-fledged sentences, clauses or meaningful phrases through a dialect's graphemes and phonemes. Such systems have their limits as they can only determine text on the basis of pre-determined text in their databases TTS systems, on the other hand, are practically to "read" strings of characters and dole out resulting sentences, clauses and phrases.

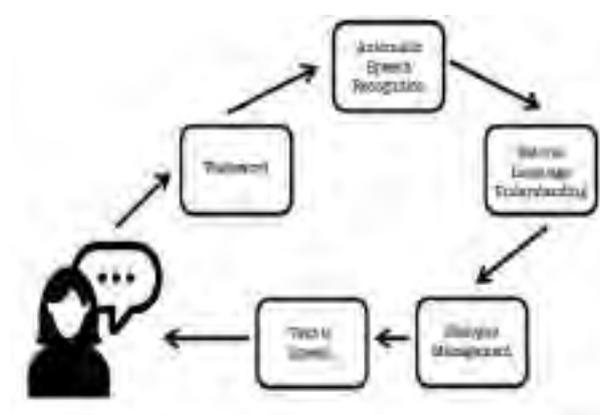


Figure: Virtual Assistant

METHODOLOGY

General Structure

Considering overall research, voice application will be used in following three ways: Firstly, command to the computer whereas secondly, to input information the computer, finally for communication with other people. In this section we will be discussing general components for voice application. As seen in Figure, voice will be divide into four different parts: front-end interface, end users, voice recognition System finally dictionary and text file database. Each section is explained as follows:

1. Front-End Interfaces

In front-end-interface, user will be having direct access to the interface and communication users by providing Input and Output with graphics designs and icon-based menu. It receives user prompt input voice and in return delivers users with a voice recognition system to detect voice inputs, and usually generates feedback of voice to users, after completion of commands by several other functions of the system.

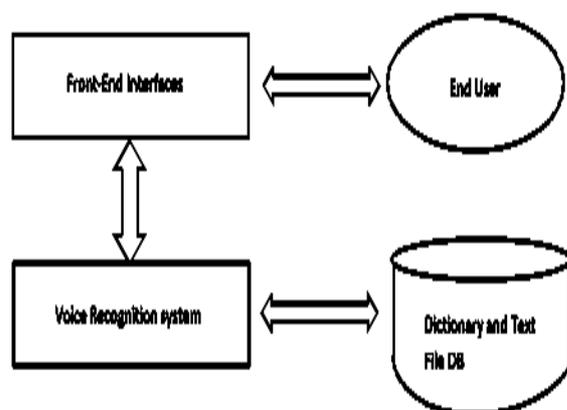


Figure: General Structure of Voice Application

2. End Users

Basically end users refers to device users. They will be using this devices for communication and feedback of voice with the use of application, and moreover end users are those who will be using this application with there personal devices like mobile and laptop users.

3. Voice Recognition Systems

It is the heart of a Voice application system, which has ability to understand voice input given by user, and make application work in a efficient way and generating voice feedback to the user. This system is an important component for user as a gateway to use his or her voice as a input component.

4. Dictionary and Test files Database

Providing the type of the device and requirements for the user, the system application needs to support few exact input types or by providing peculiar voice feedback. Whereas for language type, the system application can give additional explanations to the users or it may provide functions based on the files in the database in extension, system requires to install an additional text file database to add and update application in few different cases.

RESULT

According to the overall description in the context, the purpose of the project is to develop an intelligent voice assistant with the functionalities as calling services, message transformation, mail exchange, location services, YouTube service, checking weather, searching engine (Google, Wikipedia), translator and help menu. We can have this assistant implemented either in phones or PC's or they can be in the form of speaker devices. The common way of communication used by people in daily life is through speech. If

the mobile phone or the device the voice Assistant implemented can listen to the user for the requestor handle the daily affairs, then give the right response, it will be much easier for users to communicate with their phone, and the mobile phone will be much “Smarter” as a human assistant. This project is focusing on voice control (recognition, generating and analyzing corresponding commands, intelligent responses automatically), Google products and relevant APIs (Google map, Google weather, Google search, etc.), Wikipedia API references ranging from Speech-To-Text, Text-To-Speech technology. As all those functionalities and services for the project have been explained, the main structure and construction of the project have been illustrated with its goals.

CONCLUSION

Voice-Controlled Devices uses Natural Language Processing to process the language spoken by the human and understand the query and process the query and respond to the human with the result. The understanding of the device means Artificial Intelligence needs to be integrated with the device so that the device can work in a smart way and can also control IoT applications and devices and can also respond to query which will search the web for results and process it. It is designed to minimize the human efforts and control the device with just human Voice. The device can also be designed to interact with other intelligent voice-controlled devices like IoT applications and devices, weather reports of a city from the Internet, send an email to a client, add events on the calendar, etc. The accuracy of the devices can be increased using machine learning and categorizing the queries in particular result sets and using them in further queries. The accuracy

of the devices is increasing exponentially in the last decade. The devices can also be designed to accept commands in bilingual language and respond back in the same language queried by the user. The device can also be designed to help visually impaired people.

REFERENCES

- 1.View of Voice based Intelligent Virtual Assistance for Windows (sersc.org) journals/index.php/IJAST/article/view/11053/5872
2. Voice based Intelligent Virtual Assistance for Windows| Dotnet Project (nevonprojects.com)
3. (PDF) PROJECT REPORT ON VIRTUAL ASSISTANT SUBMITTED BY NAAZNEEN JATU | Naazneen Jatu - Academia.edu

WEBSITE EVALUATION USING OPINION MINING

Jitesh Kirdakude, Prof. Priyanka Sorte

Abstract- With the advancement in technology, online shopping has taken a new toll. Customers can access the shopping websites from anywhere with the help of the internet. The websites display all the products for the customer to choose from and the customer can post any review or their opinion on the service of the website. This system aims to perform the evaluation on the website by means of opinion mining. The opinion of the user by means of rating and feedback is taken into account for the optimal search. The main aim of this project is to provide users with a rating generated according to the shopping website's performance and standards. There are numerous shopping websites to visit but the legitimacy of all the websites is questionable. Fraud can occur on a regular basis from illegitimate websites. The users may not get the service as per their expectations. There are many renowned shopping websites which do not match their credibility and in turn disappoints the customers or users. The proposed system evaluates the shopping websites based on the reviews or opinions of the experienced customers. The performance of the website is processed down to certain parameters. According to these parameters a certain rating of that particular website is generated, which helps the customer to decide which website to opt for.

Keywords: Opinion mining, website evaluation, Feature extraction and evaluation, Web Content, Mining, Sentiment Analysis



INTRODUCTION

With the rapid advancement of Web technologies, which facilitate people to contribute rather than simply receive information, a large amount of review texts are generated and become available online. These user-generated opinion-rich contents are credible sources of knowledge that can not only help users make better judgments but assist manufacturers of products in keeping track of customer sentiments. As users, and even more as professionals, we

have all experienced that for many websites navigation is all but satisfactory, needed information is missing, transaction forms are awkward, graphical design is shabby, analytics show low popularity and success, and so on.

Every passing day the number of companies, organizations and individuals publishing their web sites is increasing. Considering all the information available on the web every individual should desire to find and access useful information. For example, users want to learn about different shopping websites and what products and services they offer using the web. By the help of this information users may learn about the websites and in turn choose a website which is suitable according to their

standards. The task of evaluating and improving the web sites can be intimidating considering the number of web sites available and the frequency of updates. As a result, automated support for web designers and web site owners become more important. Automated usability tools can help save time and money in design. User testing can improve consistency and quality of the web site. The customers wanting to buy the product would like to compare the products and the services provided by different websites before purchasing the product. The services and all the parameters related to the website can't be compared on the existing system. Therefore, it makes it difficult for the customers to decide. Although the websites would be rated but comparing the websites would make it more efficient and can provide a clear to opt for the better product. Hence it was highly important to come up with the solution for the customers which would be the evaluation of websites based on the review of the customers and rate it according to the sentiments identified in the reviews.

RAPID MINER

Opinion mining, because the name implies, is that the mining (extracting) of opinions regarding the merchandise, event, services, etc. announce by variant folks on the net. Thanks to the large enlargement of the net, folks are inspired to contribute themselves with blogs, social networking sites, etc. an outsized volume of knowledge is created thanks to these platforms that require being well-mined (extracting helpful patterns) for analysis and higher cognitive process. Opinion-mining systems analyses regarding the reviews like that half is opinion conveying, its various selection of

administrators, Rapid Miner Likewise incorporates the information mining library from the WEKA Toolkit.

LITERATURE REVIEW

The aim of Machine Learning is to develop an algorithm to increase the performance of the system using the given data or past experience. The Machine Learning provides a solution to the classification problem that involves two steps: 1. Learning the model from a corpus of training data. 2. Classifying the unseen data based on the trained model. Sentiment analysis of natural language texts is an emerging field. Converting a part of text to a feature vector is an essential step in any data driven approach to Sentiment analysis. Their proposal relies on a language modeling framework that combines a probabilistic model of opinion words and a stochastic mapping model between words to approximate a language model of products. Their work extends a preliminary approach introduced which addresses the modeling of a language of product features from customer reviews. They provide a formalized methodology for the retrieval of product features from the estimated language model of features.

METHODOLOGY

Here we propose an advanced Website Evaluation system that rates the website based on the opinion of the user. Website will be evaluated based on factors such genuineness of the website, timely delivery of the product after online transaction and support provided by the website. User will comment about the website, based on the comment system will rate the website. The system takes opinion of various users, based on the opinion; system will decide whether the website is genuine or not. The

system uses opinion mining methodology in order to achieve desired functionality. We use a database of sentiment-based keywords along with positivity or negativity weight in database and then based on these sentiment keywords mined in user comment is ranked. The system contains keywords related to fraud, genuineness, timely delivery of the product and service meters in the database. Based on these factors system will rate the website. The working of the system is follows:-

- The user logs in to the system he can view various websites posted by the admin and can comment about the website.
- User can see the comment of other user.
- System will rate the website based on the comment of various users.
- The role of the admin is to add various website to the system and to add keywords in database.
- So that system will match the comment with the keywords in database and will rate the website based on the sentiment analysis.

User can easily find out which website will deliver the product in time. And also helps to find out website which will provide good support. This application helps to find out whether the website is genuine or not that is useful for those users who do online transactions.

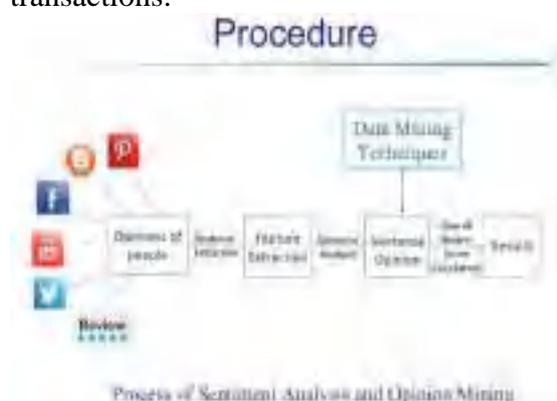


Fig: Block diagram of Website evaluation using Opinion Mining

TECHNIQUES OF OPINIONS MINING

Currently, has become a practice for websites, to facilitate the expression of opinions by guests and visitors on products marketed or on presented topics. Also, the expansion of social networking, facilitated users posting opinions online. Thus, the content of reviews has increased rapidly, making the big e-commerce sites, or recommendations of products and services sites, to contain hundreds to tens of thousands of reviews per item. The large number of reviews promotes access to useful and relevant information to visitors. They can be used, for example to compare offers from different competitors on the market and make an informed decision about buying a certain offer. It is very difficult for a visitor to read all of them and to form an opinion on the subject or product because:

- in some cases these reviews can be very long and only a few sentences may express opinions or may not contain opinions at all. Navigating only part of the may create a false impression about the topic;
- the user is not familiar with the various metrics used in comparing offers in a certain specialized field.



Fig: The overall opinion mining process

APPLICATIONS

Opinions are so important that whenever one needs to make a decision, one wants to hear others' opinions. This is true for both individuals and organizations. The technology of opinion mining thus has a tremendous scope for practical applications.

Opinion mining is useful for Individual consumers. If an individual wants to purchase a product, it is useful to see a summary of opinions of existing users so that he/she can make an informed decision. This is better than reading a large number of reviews to form a mental picture of the strengths and weaknesses of the product. He/she can also compare the summaries of opinions of competing products, which is even more useful.

RESULT

Opinion Mining has become a latest trend in the information mining industry. There is plenty of future scope for opinion mining as it requires Natural Language Processing and also Artificial Intelligence. While the development of the opinion mining tools described shows very much work in progress and initial results are promising, but still it requires a lot of refinements. Since this is a study of sentiments of a person, so it requires a lot of precision. Whenever any person talk about something, then the context in which he is talking and how the sentence is formed may change the parsing method to catch the exact opinion

said by that person. If we try to concentrate on one pattern of sentence then we may lose any other pattern of sentence from our parsing method. This is a major challenge in front of the opinion mining methods.

CONCLUSION

Our system gives the customer a better understanding about e-shopping websites. It prevents the customer from being a victim of fraud. It will give them a detailed survey about the service or customer feedback regarding those shopping websites. Our system could be a critic for the shopping sites hence driving them towards improvement on their drawbacks. .On the basis of data availability our system can be expanded to different sectors. It is also open to customer's views based on their frequent shopping experience with those websites.

REFERENCES

- <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6887540&matchBoolean%3Dtrue%26queryText%3DEvaluation+of+genuiness+of+website>
- <https://projectchampionz.com.ng/2018/11/14/website-evaluation-using-opinion-mining/>
- <https://www.sciencepubco.com/index.php/ijet/article/view/10257>
- https://www.researchgate.net/publication/324044138_Website_evaluation_using_opinion_mining

CHEST X-RAY IN NEW CORONAVIRUS DISEASE 2019 (COVID-19) INFECTION: FINDINGS AND CORRELATION WITH CLINICAL OUTCOME

Ms. Payal Jingare
PG Student,
Pillai HOC College of Arts,
Science & Commerce, Rasayani

Prof. Priyanka Sorte
Asst. Professor,
Pillai HOC College of Arts,
Science & Commerce, Rasayani

Abstract- The purpose of this study is to describe the main chest radiological features (CXR) of COVID-19 and correlate them with clinical outcome. This is a retrospective study involving patients with clinical-epidemiological suspect of COVID-19 infection, who performed CXRs at the emergency department (ED) of our University Hospital from March 1 to March 31, 2020. All patients performed RT-PCR nasopharyngeal and throat swab, CXR at the ED and clinical-epidemiological data. RT-PCR results were considered the reference standard. The final outcome was expressed as discharged or hospitalized patients into a medicine department or intensive care unit (ICU).

Keywords: chest x-ray, COVID-19, chest radiological features, intensive care unit (ICU)

INTRODUCTION

At the end of 2019 a novel virus, named SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), expanded globally from China with the first Italian cases dating back to February 2020. This new coronavirus causes a highly infectious disease, commonly called Coronavirus Disease 19 (COVID-19): Lung infection can result in severe pneumonia up to more aggressive acute respiratory distress syndrome (ARDS). Genetic sequencing of SARS-CoV-2 has permitted the rapid development of real-time reverse transcription polymerase chain reaction

(RT-PCR) of viral nucleic acid, and nowadays this is the diagnostic gold standard. However, this serologic examination has several limitations due to the high number of false-negative tests and the delayed results. Radiological evaluation of patients with clinical-epidemiological suspect of COVID-19 is mandatory, especially in the emergency department (ED) while waiting for RT-PCR results, in order to have a rapid evaluation of thoracic involvement. The recent COVID-19 radiological literature focuses primarily on computed tomography (CT) findings, which is more sensitive and specific than chest X-ray (CXR): In particular, in China CT is used as a first-line diagnostic method for COVID-19. Nonetheless, it has to be remembered that performing CT scan is not easy during this pandemic, considering not only the excessive radiation exposure especially to younger patients but also the

- *Payal Jingare is currently pursuing masters degree program in M.Sc IT in Mumbai University, India, 7057640590. Email:payalanji2019mscit@students.mes.ac.in*
- *Prof. Priyanka Sorte is currently professor masters degree program in M.Sc IT in Mumbai University, India, 96891 57442. E-mail: psorte@mes.ac.in*

mandatory scanner disinfection procedures that have to take place. The most Italian hospitals are employing CXR as the first-line method, with faster results comparing with those of RT-PCR, especially by using portable X-ray units which reduce the movement of patients and so minimizing the risk of cross-infection. Therefore, the purpose of our study is to better understand the main radiographic features of COVID-19 pneumonia, by describing the main CXR findings in a selected cohort of patients, also correlating the radiological appearance with RT-PCR examination and patients outcome (intended as discharged or hospitalized into a medicine department or intensive care unit).

METHODOLOGY

i) Patients' selection and inclusion criteria

CXR of patients with clinical-epidemiological suspect of COVID-19 infection performed at the ED of our University Hospital from March 1 to March 31, 2020, were retrospectively reviewed. Inclusion criteria were: patients' age between 18 and 99 years, RT-PCR nasopharyngeal-throat swab and CXR performed immediately at the ED access, clinical-epidemiological data suspect for COVID-19 infection and their duration at the time of ED access (fever, cough, dyspnea, respiratory impairment, diarrhea, asthenia, myalgia and dysgeusia). RT-PCR results were considered the reference standard. For the radiological assessment we selected only patients with RT-PCR-positive results. The final outcome was expressed as discharged or hospitalized patients into a medicine department or into an intensive care unit (ICU).

ii) Image acquisition and analysis

All CXRs were acquired as digital radiographs with the same portable X-ray unit (FDR Go PLUS—Fujifilm, Italia) in the isolation wards of our ED. CXRs were performed in the postero-anterior or antero-posterior projection. All images were stored in a picture archiving and communication system (PACS, Syngo-Siemens). An independent and retrospective review of each CXR was performed by two thoracic radiologists in order to define the number of radiological suspects of COVID-19 infection; after this, they defined the predominant pattern of COVID-19 pneumonia presentation in patients with a positive RT-PCR. In case of discordance, a consensual agreement was reached. Radiographic features including consolidation, ground-glass opacities (GGO), pulmonary nodules and reticular-nodular opacities were diagnosed according to the Fleischner Society glossary of terms. Moreover, CXRs were assessed for the presence of a specific distribution of the disease (mostly peripheral or perihilar predominance), monolateral (right or left lung) or bilateral disease, upper or lower or diffuse predominance. All thoracic images were also assessed for evidence of other associated pulmonary pathology (cardiomegaly, hilar vascular congestion, pleural effusion, pneumothorax). Finally, to quantify the extent of COVID-19 lung involvement, a severity score was applied (Radiographic Assessment of Lung Edema—RALE). Following RALE indications, each CXR was given a score between 0 and 48, ranging from the absence of any pathological sign (score 0) to the complete pathological involvement of lung parenchyma (score 48). The score was separately assessed by each of the two radiographers.

iii) Statistical analysis

Statistical analysis was performed with SPSS (SPSS Chicago IL, USA). Descriptive statistics of RALE score were calculated for each group of patients.

Analysis of Variance (ANOVA) was performed to detect possible differences among RALE score estimated in the following groups: discharged patients, hospitalized patients into a medicine department, hospitalized patients into an ICU (respectively, group 1, 2 and 3 in the following). Homogeneity of the variance was established between groups by means of Levene test, to adequately choose the post hoc test: Bonferroni in case of detected significant homogeneity of variance, otherwise Games-Howell. Intraclass correlation coefficient (ICC) was calculated to assess inter-reader agreement in assigning RALE score. Statistical significance threshold was set at $p = 0.05$.

RESULTS

We found 482 patients fulfilling the following selecting criteria: presence of clinical-epidemiological suspect of COVID-19 infection and RT-PCR and CXR performed at the ED admission. Patients with a RT-PCR-positive results for COVID-19 infection were 234: Of these, 153 were males (65.4%) and 81 females (34.6%), with a mean age of 66.04 years (range 18–97 years). Only 13 CXRs were negative for radiological thoracic involvement (5.6%). The others showed variable features as described in Tables 1 and 2. The following alterations were more commonly observed: 135 patients with lung consolidations (57.7%), 147 (62.8%) with GGO, 55 (23.5%) with nodules and 156 (66.6%) with reticular-nodular opacities. Patients with consolidations and GGO coexistent in the same radiography were 35.5% of total. In RT-PCR-positive patients, we found also signs nonspecific for COVID-19 pneumonia as hilar or vascular congestion (39.3%), cardiomegaly (29.9%), pleural effusion (16.6%) and pneumothorax (2.4%). Peripheral (57.7%) and lower zone distribution (58.5%) were the most

common predominance. Bilateral involvement (69.2%) was most frequent than unilateral one. Given the results, baseline CXR sensitivity in our experience is about 68.1%.

COVID-19 Radiological Features	n (%)
Normal findings (N=)	13 (5.6)
Abnormal findings (N=)	221 (94.4)
Reticular-nodular opacities	156 (66.6)
Consolidation	147 (62.8)
Cardiomegaly	72 (30.2)
Vascular congestion	72 (30.2)
Cardiomegaly	70 (29.9)
Nodules	55 (23.5)
Pleural effusion	39 (16.6)
Pneumothorax	5 (2.4)
Distribution	
Peripheral	135 (57.7)
Bilateral	161 (69.2)
Unilateral	59 (23.5)
Diffuse	39 (16.6)
Bilateral predominant	167 (73.8)
Superior predominance	71 (31.5)
Right lung	21 (9.0)
Left lung	23 (9.4)
Mixed	10 (4.0)

Table 1 Radiographic findings of our cohorts of COVID-19 patients

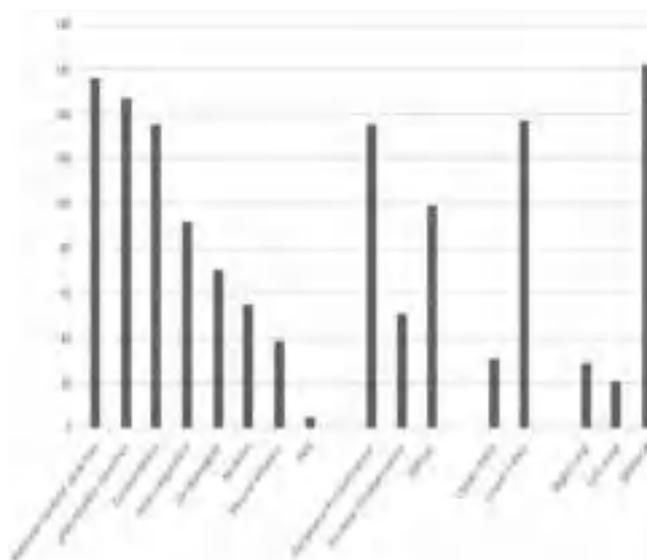
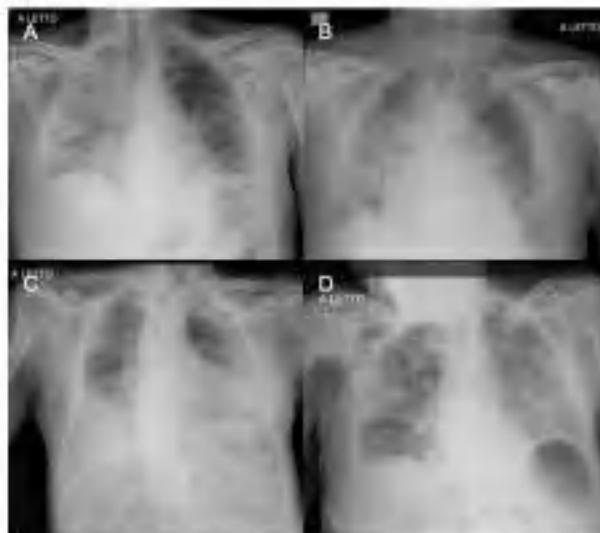


Table 2 Graphic of CXRs findings of our cohorts of COVID-19 patients

In our population the most affected patients were in the age group of 60–79 years old (43.6%, of which 71.57% males); patients older than 80 years (23.1%) often presented more advanced lung involvement (Fig. 1).

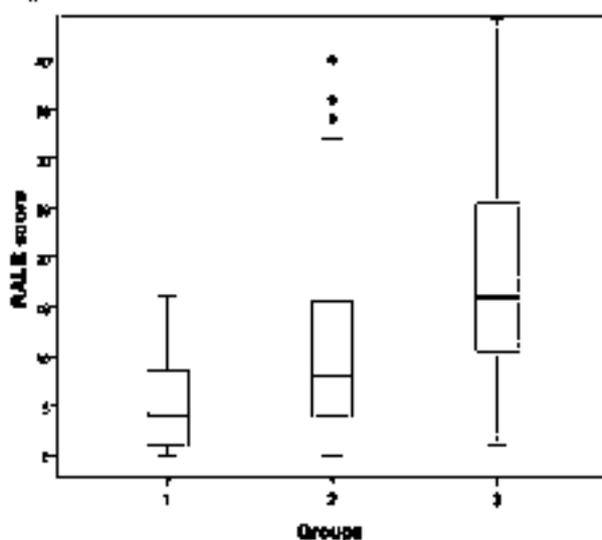
Nine patients (4%) were immediately discharged from ED, and the others were hospitalized in medicine department or ICU (Table 3). A total of 34 (15.3%) patients died in the 30 days included in this study (9 in group 2 and 25 in group 3).

Fig. 1



Diffuse lung involvement in elderly patients. Four cases of advanced lung disease with diffuse consolidations and interstitial involvement of patients older than 80 years at the emergency department

Fig. 2



Box and Whisker plot of RALE score estimated in each group defined by outcome: discharged patients (group 1), hospitalized patients into a medicine department (group 2), hospitalized patients

into an intensive care unit (group 3). RALE score showed statistically significant differences between group 1 vs 3 and 2 vs 3.

CONCLUSION

We describe the main features of COVID-19 thoracic involvement on CXR in our cohort of patients. In our experience, baseline CXR had a sensitivity of 67.1%. The RALE score can be used in the emergency setting as a quantitative method of the extent of SARS-CoV-2 pneumonia, correlating with an increased risk of ICU admission. The results of our study could help radiologists in identifying the highest risk patients, allowing for timely initiation of treatments currently available against SARS-CoV-2 infection.

REFERENCE

1. Zhu N, Zhang D, Wang W et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. <https://doi.org/10.1056/NEJMoa2001017>
2. Yang W, Sirajuddin A, Zhang X et al (2020) The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). *Eur Radiol*. <https://doi.org/10.1007/s00330-020-06827-4>
3. Hansell DM, Bankier AA, MacMahon H et al (2008) Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 246(3):697722. <https://doi.org/10.1148/radiol.2462070712>
4. Fichera G, Stramare R, de Conti G et al (2020) It's not over until it's over: the chameleonic behavior of COVID-19 over a six-day period. *Radiol Med*. <https://doi.org/10.1007/s11547-020-01203-0>

AN ARTIFICIAL INTELLIGENCE SYSTEM FOR PREDICTING THE DETERIORATION OF COVID-19 PATIENTS IN THE EMERGENCY DEPARTMENT

Ms. Payal Jingare
PG Student,
Pillai HOC College of Arts,
Science & Commerce, Rasayani.

Prof. Priyanka Sorte
Asst. Professor,
Pillai HOC College of Arts,
Science & Commerce, Rasayani.

Abstract- During the COVID-19 pandemic, rapid and accurate triage of patients at the emergency department is critical to inform decision-making. We propose a data-driven approach for automatic prediction of deterioration risk using a deep neural network that learns from chest X-ray images, and a gradient boosting model that learns from routine clinical variables. Our AI prognosis system, trained using data from 3,661 patients, achieves an AUC of 0.786 (95% CI: 0.742-0.827) when predicting deterioration within 96 hours. The deep neural network extracts informative areas of chest X-ray images to assist clinicians in interpreting the predictions, and performs comparably to two radiologists in a reader study. In order to verify performance in a real clinical setting, we silently deployed a preliminary version of the deep neural network at NYU Langone Health during the first wave of the pandemic, which produced accurate predictions in real-time. In summary, our findings demonstrate the potential of the proposed system for assisting front-line physicians in the triage of COVID-19 patients.

Keywords: COVID-19 pandemic, prediction of deterioration risk, X-ray images, neural network, image processing.

◆

INTRODUCTION

In recent months, there has been a surge in patients presenting to the emergency department (ED) with respiratory illnesses associated with SARS CoV-2 infection (COVID-19). Evaluating the risk of deterioration of these patients to perform triage is crucial for clinical decision-making and resource allocation. While ED triage is difficult under normal circumstances, during a pandemic, strained

hospital resources increase the challenge. This is compounded by our incomplete understanding of COVID-19. Data-driven risk evaluation based on artificial intelligence (AI) could, therefore, play an important role in streamlining ED triage.

As the primary complication of COVID-19 is pulmonary disease, such as pneumonia, chest X-ray imaging is a first-line triage tool for COVID-19 patients. Although other imaging modalities, such as computer tomography (CT), provide higher resolution, chest X-ray images are less costly, inflict a lower radiation dose, and are easier to obtain without incurring the risk of contaminating imaging equipment and disrupting radiologic services. In addition, abnormalities in the chest X-ray

-
- *Payal Jingare is currently pursuing masters degree program in M.Sc IT in Mumbai University, India, 7057640590. Email: payalanji2019mscit@students.mes.ac.in*
 - *Prof. Priyanka Sorte is currently professor masters degree program in M.Sc IT in Mumbai University, India, 96891 57442. E-mail: psorte@mes.ac.in*

images of COVID-19 patients have been found to mirror abnormalities in CT scans. Consequently, chest X-ray imaging is considered a key tool in assessing COVID-19 patients. Unfortunately, although the knowledge of the disease is rapidly evolving, understanding of the correlation between pulmonary parenchymal patterns visible in the chest X-ray images and clinical deterioration is limited. This motivates the use of machine learning approaches for risk stratification using chest X-ray imaging, which may be able to learn such correlations automatically from data.

The majority of related previous works using imaging data of COVID-19 patients concentrate more on diagnosis than prognosis. Prognostic models have a number of potential real-life applications, such as: consistently defining and triaging sick patients, alerting bed management teams on expected demands, providing situational awareness across teams of individual patients, and more general resource allocation. Prior methodology for prognosis of COVID-19 patients via machine learning mainly use routinely-collected clinical variables such as vital signs and laboratory tests, which have long been established as strong predictors of deterioration. Some studies have proposed scoring systems for chest X-ray images to assess the severity and progression of lung involvement using deep learning, or more commonly, through manual clinical evaluation. In general, the role of deep learning for the prognosis of COVID-19 patients using chest X-ray imaging has not yet been fully established.

Methods Outline

In this section, we first introduce our data collection and preprocessing pipeline. We then formulate the adverse event prediction task and present our multi-modal approach

which utilizes both chest X-ray images and clinical variables. Next, we formally define deterioration risk curve (DRC) and introduce our X-ray image-based approach to estimate DRC. Subsequently, we summarize the technical details of model training and implementation. Lastly, we describe the design of the reader study. Dataset collection and preparation. We extracted a dataset of 19,957 chest X-ray exams collected from 4,772 patients who tested positive for COVID-19 between March 2, 2020, and May 13, 2020. We applied inclusion and exclusion criteria that were defined in collaboration with clinical experts, as shown in Figure 2.b. Specifically, we excluded 783 exams that were not linked to any radiology report, nine exams that were not linked to any encounter information, and 5,213 exams from patients who were still hospitalised by May 13, 2020. To ensure that our system predicts deterioration prior to its occurrence, we excluded 6,260 exams that were collected after an adverse event and 187 exams of already intubated patients. The final dataset consists of 7,502 chest X-ray exams corresponding to 4,204 unique patients. We split the dataset at the patient level such that exams from the same patient exclusively appear either in the training or test set. In the training set, we included exams that were collected both in the ED and during inpatient encounters. Since the intended clinical use of our model is in the ED, the test set only includes exams collected in the ED. This resulted in 5,224 exams (5,617 images) in the training set and 770 exams (832 images) in the test set. We included both frontal and lateral images, however there were less than 50 lateral images in the entire dataset.

The data used to evaluate the models during deployment consist of 375 exams from 217 patients collected between May 22, 2020 and June 24, 2020. The exams were filtered

based on the same criteria described above. Among the 375 exams, 25 chest X-ray exams were collected from patients who were admitted to the ICU within 96 hours, and six exams were collected from patients who were intubated within 96 hours.

After extracting the images from DICOM files, we applied the following preprocessing procedure. We first thresholded and normalized pixel values, and then cropped the images to remove any zero-valued pixels surrounding the image. Then, we unified the dimensions of all images by cropping the images outside the center and rescaling. We performed data augmentation by applying random horizontal flipping ($p = 0.5$), random rotation (-45 to 45 degrees), and random translation. Supplementary Figure 1 shows the distribution of the size of the images prior to data augmentation, as well as examples of images before and after preprocessing.

Adverse event prediction:

Our main goal is to predict clinical deterioration within four time windows of 24, 48, 72, and 96 hours. We frame this as a classification task with binary labels $y = [y_{24}, y_{48}, y_{72}, y_{96}]$ indicating clinical deterioration of a patient within the four time windows. The probability of deterioration is estimated using two types of data associated with the patient: a chest X-ray image, and routine clinical variables. We use two different machine learning models for this task: COVID-GMIC to process chest X-ray images, and COVID-GBM to process clinical variables. For each time window $t \in T_a = \{24, 48, 72, 96\}$, both models produce probability estimates of clinical deterioration,

$$\hat{y}_{\text{COVID-GMIC}}^t, \hat{y}_{\text{COVID-GBM}}^t \in [0, 1].$$

In order to combine the predictions from COVID-GMIC and COVID-GBM, we

employ the technique of model ensembling [46]. Specifically, for each example, we compute a multi-modal prediction $\hat{y}_{\text{ENSEMBLE}}$ as a linear combination of $\hat{y}_{\text{COVID-GMIC}}$ and $\hat{y}_{\text{COVID-GBM}}$:

$$\hat{y}_{\text{ENSEMBLE}} = \lambda \hat{y}_{\text{COVID-GMIC}} + (1 - \lambda) \hat{y}_{\text{COVID-GBM}}, \quad (1)$$

where $\lambda \in [0, 1]$ is a hyperparameter. We selected the best λ by optimizing the average of the AUC and PR AUC on the validation set.

Clinical variables model:

The goal of the clinical variables model is to predict the risk of deterioration when the patient's vital signs are measured. Thus, each prediction was computed using a set of vital sign measurements, in addition to the patient's most recent laboratory test results, age, weight, and body mass index (BMI). The laboratory test results were represented as maximum and minimum statistics of all values collected within 12 hours prior to the time of the vital sign measurement. The feature sets of vital signs and laboratory tests were then processed using a gradient boosting model which we refer to as COVID-GBM. For the final ensemble prediction, $\hat{y}_{\text{ENSEMBLE}}$, we combined the COVID-GMIC prediction with the COVID-GBM prediction computed using the most recently collected clinical variables prior to the chest X-ray exam. In cases where there were no clinical variables collected prior to the chest X-ray, we performed a mean imputation of the predictions assigned to the validation set.

Chest X-ray image model:

We process chest X-ray images using a deep convolutional neural network model, which we call COVID-GMIC, based on the GMIC model. COVID-GMIC has two desirable properties. First, COVID-GMIC generates interpretable saliency maps that

highlight regions in the X-ray images that correlate with clinical deterioration.

Model training and selection:

In this section, we discuss the experimental setup used for COVIDGMIC, COVID-GMIC-DRC, and COVID-GBM. The chest X-ray image models were implemented in PyTorch and trained using NVIDIA Tesla V100 GPUs. The clinical variables models were implemented using the Python library LightGBM.

We applied data augmentation during training and testing, but not during validation. During testing, we augmented each image ten times and averaged the corresponding outputs to produce the final prediction.

Design of the reader study:

The reader study consists of 200 frontal chest X-ray exams from the test set. We selected one exam per patient to increase the diversity of exams. We used stratified sampling to ensure that a sufficient number of exams in the study corresponded to the least common outcome (patients with adverse outcomes in the next 24 hours). In more detail, we oversampled exams of patients who developed an adverse event by sampling the first 100 exams only from patients from the test set that had an adverse outcome within the first 96 hours. The remaining 100 exams came from the remaining patients in the test set. The radiologists were asked to assign the overall probability of deterioration to each scan across all time windows of evaluation.

Results

Dataset:

Our AI system was developed and evaluated using a dataset collected at NYU Langone Health between March 3, 2020 and June 28, 2020.¹ The dataset consists of chest X-ray images collected from patients who tested positive for COVID-19 using the polymerase chain reaction (PCR) test, along with the clinical variables recorded closest to the time of image acquisition (e.g. vital signs, laboratory test results, and patient characteristics). The training set consisting of 5,617 chest X-ray images was used for model development and hyperparameter tuning, while the test set consisting of 832 images was used to report the final results. The training and the test sets were disjoint, with no patient overlap. Table 1 summarizes the overall demographics and characteristics of the patient cohort in the training and test sets. Supplementary Table 1 summarizes the associated clinical variables included in the dataset.

Table 1: Description of the characteristics of the patient cohort included in the training and test sets used to develop and evaluate our system. The training and test sets are similar in terms of age, BMI, and proportion of females. We note that there is a higher proportion of chest X-ray images associated with deterioration across all time windows in the test set compared to the training set. This implies that there is a higher incidence of adverse events amongst ED patients than inpatients, since the test set only includes chest X-ray images collected from ED patients, while the training set also includes inpatients.

Characteristic	Training set	Test
Patients, n	2,943	711
Admissions, n	3,175	764
Females, n (%)	1,206 (41.0)	305 (41.4)
Age (years), mean (SD)	62.9 (17.2)	64.9 (17.1)
BMI (kg/m ²), mean (SD)	29.4 (7.1)	29.5 (7.1)
Survived	2,405	554
Adverse events, n	1,311	594
Intubation, n	386	97
ICU admission, n	387	111
Mortality, n	538	151
Composite outcome, n	730	221
Chest X-ray exams, n	5,224	771
Composite outcome within 24 hours, n (%)	349 (6.7%)	74 (9.6%)
Composite outcome within 48 hours, n (%)	553 (10.6%)	101 (13.1%)
Composite outcome within 72 hours, n (%)	735 (14.1%)	130 (16.9%)
Composite outcome within 96 hours, n (%)	870 (16.8%)	156 (20.2%)
Total number of images, n	3,617	831

Comparison to radiologists:

We compared the performance of COVID-GMIC with two chest radiologists from NYU Langone Health (with 3 and 17 years of experience) in a reader study with a sample of 200 frontal chest X-ray exams from the test set. We used stratified sampling to improve the representation of patients with a negative outcome in the reader study dataset. We describe the design of the reader study in more detail in the Methods section.

Explain ability of COVID-GMIC:

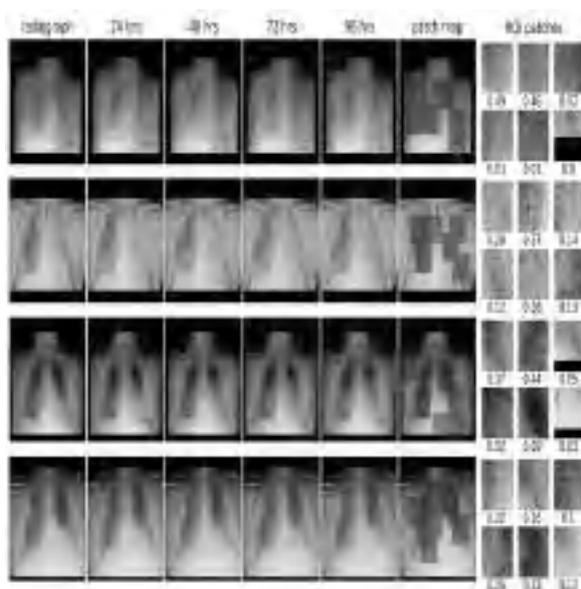


Figure 2: From left to right: the original X-ray image, saliency maps for clinical

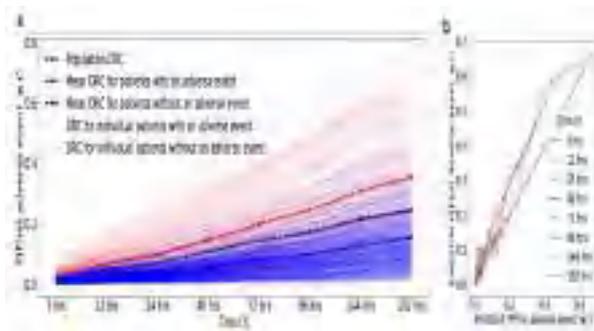
deterioration within 24, 48, 72, and 96 hours, locations of region-of-interest (ROI) patches, and ROI patches with their associated attention scores. All four patients were admitted to the intensive care unit and were intubated within 48 hours. In the first example, there are diffuse airspace opacities, though the saliency maps primarily highlight the medial right basilar and peripheral left basilar opacities. Similarly, the two ROI patches (1 and 2) on the basilar region demonstrate comparable attention values, 0.49 and 0.46 respectively. In the second example, the extensive left mid to upper-lung abnormality is highlighted, which correlates with the most extensive area of parenchymal consolidation. In the third example, the saliency maps highlight the left mid lung and right hilar/infrahilar regions which show groundglass opacities. In the last example, saliency maps highlight the right mid to lower paramediastinal and left mid lung periphery as regions predictive of clinical deterioration within 96 hours.

Deterioration risk curves:

We use a modified version of COVID-GMIC, referred to hereafter as COVID-GMIC-DRC, to generate discretized deterioration risk curves (DRCs) which predict the evaluation of the deterioration risk based on chest X-ray images. Figure 3.a shows the DRCs for all the patients in the test set. The DRC represents the probability that the first adverse event occurs before time t , where t is equal to 3, 12, 24, 48, 72, 96, 144, and 192 hours. The mean DRCs of patients who deteriorate (red bold line) is significantly higher than the mean DRCs of patients who are discharged without experiencing any adverse events (blue bold line). We evaluate the performance of the model using the concordance index, which is computed on patients in the test set who experienced

adverse events. For a fixed time t the index equals the fraction of pairs of patients in the test data for which the patient with the higher DRC value at t experiences an adverse event earlier. For t equal to 96 hours, the concordance index is 0.713 (95% CI: 0.682-0.747), which demonstrates that COVID-GMIC-DRC can effectively discriminate between patients.

Figure 3.b shows a reliability plot, which evaluates the calibration of the probabilities encoded in the DRCs. The diagram compares the values of the estimated DRCs for the patients in the test set with empirical probabilities that represent the true frequency of adverse events.



0.365 (95% CI: 0.313-0.465) PR AUC. The decrease in accuracy may indicate changes in the patient population as the pandemic progressed.

REFERENCE

- [1] Baugh, J. J. et al. Creating a COVID-19 surge clinic to offload the emergency department. *Am. J. Emerg. Med.* 38, 1535–1537 (2020).
- [2] Debnath, S. et al. Machine learning to assist clinical decision-making during the COVID-19 pandemic. *Bioelectron. Med.* 6, 1–8 (2020).
- [3] Whiteside, T., Kane, E., Aljohani, B., Alsamman, M. & Pourmand, A. Redesigning emergency department operations amidst a viral pandemic

Serverless cloud computing

Ms. Hemangi Mali

Student Of MSC.IT

Pillai HOC College of Arts, Science And Commerce
Rasayani

Ms. Priyanka Sorte

Asst. professor, Department Of IT/CS

Pillai HOC College of Arts, Science And Commerce ,
Rasayani

Abstract- Serverless cloud computing handles virtually all the system administration operations needed to make it easier for programmers to use the cloud. It provides an interface that greatly simplifies cloud programming, and represents an evolution that parallels the transition from assembly language to high-level programming languages.

Index Terms - Serverless Cloud Computing, Architecture , Attractiveness , Limitations of serverless cloud computing .

1.Introduction

The past few years have seen these advantages largely realized, but cloud users continue to bear a burden from complex operations and many workloads still do not benefit from efficient multiplexing. Cloud computing relieved users of physical infrastructure management but left them with a proliferation of virtual resources to manage. Multiplexing worked well for batch style workloads such as MapReduce or high performance computing, which could fully utilize the instances they allocated. It worked less well for stateful services, such as when porting enterprise software like a database management system to the cloud.

2. Serverless Cloud Computing

programmers from managing memory resources, whereas serverless computing relieves programmers from managing server resources.

There are three critical distinctions between serverless and serverful computing:

1. Decoupled computation and storage .The storage and computation scale separately and are provisioned and priced independently. In general, the storage is provided by a separate cloud service and the computation is stateless.

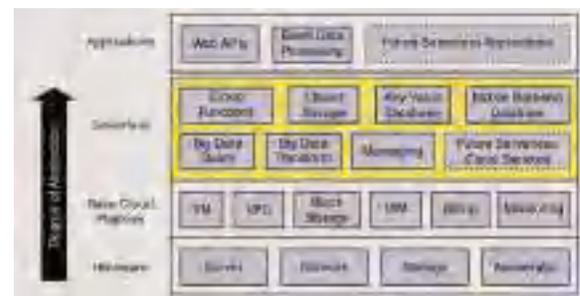
2. Executing code without managing resource allocation. Instead of requesting resources, the user provides a piece of code and the cloud automatically provisions resources to execute that code.

3. Paying in proportion to resources used instead of for resources allocated. Billing is by some dimension associated with the execution, such as execution time rather than by a dimension of the base cloud platform, such as size and number of VMs allocated.

In any serverless platform, the user just writes a cloud function in a high-level language, picks the event that should trigger the running of the function such as loading an image into cloud storage or adding an image thumbnail to a database table and lets the serverless system handle everything else: instance selection, scaling, deployment, fault tolerance, monitoring, logging, security patches, and so on.

This mirrors several of the steps of serverful cloud programming, where one first provisions resources or identifies available ones, then loads those resources with necessary code and data, performs the computation, returns or stores the results, and eventually manages resource release. The aim and opportunity in serverless computing is to give cloud programmers benefits similar to those in the transition to high-level programming languages. Other features of high-level programming environments have natural parallels in serverless computing as well. Automated memory management relieves

❖ Architecture of the serverless cloud



2.1 Attractiveness of Serverless Computing

For cloud providers serverless computing promotes business growth, as making the cloud easier to program helps draw in new customers and helps existing customers make more use of cloud offerings. For example, recent surveys found that about 24% of serverless users were new to cloud computing and 30% of existing serverful cloud customers also used serverless computing .

In addition, the short run time, small memory footprint, and stateless nature improve statistical multiplexing by making it easier for cloud providers to find unused resources on which to run these tasks. The cloud providers can also utilize less popular computers as the instance type is up to the cloud providers such as older servers that may be less attractive to serverful cloud customers.

Popularity of serverless computing use cases according to a 2018-2029 survey .

	Use Case
32%	Web and API serving
21%	Data Processing, e.g, batch ETL (database Extract Transform and Load)
17%	Integrating 3 rd party services
16%	Internal tooling
8%	Chat bots e.g., Alexa Skills (SKD for Alexa AI Assistant)
6%	Internet of things

3. Limitations of Serverless Computing Platforms

Serverless cloud functions have been successfully employed for several classes of workloads¹⁰ including API serving, event stream processing, and limited ETL¹¹. To see what obstacles prevent supporting more general workloads, we attempted to create serverless versions of applications that were of interest to us and studied examples published by others. These are not intended to be representative of the rest of information technology outside of the current serverless computing ecosystem; they are simply examples selected to uncover common weaknesses that might prevent serverless versions of many other interesting applications.

Interestingly, even this eclectic mix of applications exposed similar weaknesses, which we list after describing the applications.

3.1. ExCamera: Video encoding in real-time. ExCamera aims to provide a real-time encoding service to users uploading their videos to sites, such as YouTube. Depending on the size of the video, today's encoding solutions can take tens of minutes, even hours. To perform encoding in real time, ExCamera parallelizes the "slow" parts of the encoding, and performs the "fast" parts serially. ExCamera exposes the internal state of the video encoder and decoder, allowing encoding and decoding tasks to be executed using purely functional semantics. In particular, each task takes the internal state along with video frames as input, and emits the modified internal state as output.

3.2. MapReduce: Analytics frameworks such as MapReduce, Hadoop, and Spark, have been traditionally deployed on managed clusters. While some of these analytics workloads are now moving to serverless computing, these workloads mostly consist of Map-only jobs. The natural next step is supporting full fledged MapReduce jobs. One of the driving forces behind this effort is leveraging the flexibility of serverless computing to efficiently support jobs whose resource requirements vary significantly during their execution.

3.3. Numpywren : Linear algebra. Large scale linear algebra computations are traditionally deployed on supercomputers or high-performance computing clusters connected by high-speed, low-latency networks. Given this history, serverless computing initially seems a poor fit. Yet there are two reasons why serverless computing might still make sense for linear algebra computations. First, managing clusters is a big barrier for many non-CS scientists . Second, the amount of parallelism can vary dramatically during a computation. Provisioning a cluster with a fixed size will either slow down the job or leave the cluster underutilized.

3.4. Cirrus : Machine learning training. Machine Learning (ML) researchers have traditionally used clusters of VMs for different tasks in ML workflows such as preprocessing, model training, and hyperparameter tuning. One challenge with this approach is that different stages of this pipeline can require significantly different amounts of resources. As with linear algebra algorithms, a fixed cluster size will either lead to severe underutilization or severe slowdown. Serverless computing can address this challenge by enabling every stage to scale to meet its resource demands. Further, it frees developers from managing clusters.

3.5. Serverless SQLite : Databases. Various autoscaling database services already exist , but to better understand the limitations of serverless computing it is important to understand what makes database workloads particularly challenging to implement. In this context, we consider whether a third party could implement a serverless database directly using cloud functions, the general purpose serverless computing building block. A strawman solution would be to run common transactional databases, such as PostgreSQL, Oracle, or MySQL inside cloud functions. However, that immediately runs into a number of challenges.

4. What Serverless Computing Should Become

Now that we've explained today's serverless computing and its limitations, let's look to the future to understand how to bring its advantages to more applications. Researchers have already begun to address issues raised above and to explore how to improve serverless platforms and the performance of workloads that run on them . Additional work done by our Berkeley colleagues and some of us emphasizes data-centric, distributed systems, machine learning, and programming model challenges and opportunities for serverless computing. Here we take a broad view on increasing the types of applications and hardware that work well with serverless computing, identifying research challenges in five areas: abstractions, systems, networking, security, and architecture.

4.1 Abstraction challenges

1. Resource requirements : With today's serverless offerings the developer specifies the cloud function's memory size and execution time limit, but not other resource needs. This abstraction hinders those who want more control on specifying resources, such as the number of CPUs, GPUs, or other types of accelerators. One approach would be to enable developers to specify these resource requirements explicitly. However, this would make it harder for cloud providers to achieve high utilization through statistical multiplexing, as it puts more constraints on function scheduling. It also goes against the spirit of serverless by increasing the management overhead for cloud application developers.

2. Data dependencies : Today's cloud function platforms have no knowledge of the data

dependencies between the cloud functions, let alone the amount of data these functions might exchange. This ignorance can lead to suboptimal placement that could result in inefficient communication patterns, as illustrated in the MapReduce and numpywren examples .

4.2 System challenges

1 .Coordination/signaling service : Sharing state between functions often uses a producer-consumer design pattern, which requires consumers to know as soon as the data is available from producers. Similarly, one function might want to signal another when a condition becomes available, or multiple functions might want to coordinate, e.g., to implement data consistency mechanisms. Such signaling systems would benefit from microsecond-level latency, reliable delivery, and broadcast or group communication. We also note that since cloud function instances are not individually addressable they cannot be used to implement textbook distributed systems algorithms such as consensus or leader election .

2. Minimize startup time : There are three parts of startup time scheduling and starting resources to run the cloud function, downloading the application software environment (e.g., operating system, libraries) to run the function code, and performing application-specific startup tasks such as loading and initializing data structures and libraries. Resource scheduling and initialization can incur significant delays and overheads from creating an isolated execution environment, and from configuring customer's VPC and IAM policies. Cloud providers as well as others have recently focused on reducing the startup time by developing new lightweight isolation mechanisms.

4.3 Networking challenges

Cloud functions can impose significant overhead on popular communication primitives such as broadcast, aggregation, and shuffle. In particular, assuming that we can pack K cloud functions on a VM instance, a cloud function version would send K times more messages than an instance version, and K^2 more messages in the case of shuffle. There may be several ways to address this challenge:

- Provide cloud functions with a larger number of cores, similar to VM instances, so multiple tasks can combine and share data among them before sending over the network or after receiving it.
- Allow the developer to explicitly place the cloud functions on the same VM instance. Offer distributed communication primitives that applications can use

out-of-the-box so that cloud providers can allocate cloud functions to the same VM instance.

- Let applications provide a computation graph, enabling the cloud provider to co-locate the cloud functions to minimize communication overhead (see “Abstraction Challenges” above.)

4.4 Security challenges

1.Scheduling randomization and physical isolation

: Physical co-residency is the center of hardware-level side-channel or Rowhammer attacks inside the cloud. As a first step in these types of attacks, the adversarial tenant needs to confirm the cohabitation with the victim on the same physical host, instead of randomly attacking strangers. The ephemerality of cloud functions may limit the ability of the attacker to identify concurrently-running victims. A randomized, adversary-aware scheduling algorithm might lower the risk of co-locating the attacker and the victim, making co-residency attacks more difficult.

2.Fine-grained security contexts : Cloud functions need fine-grained configuration, including access to private keys, storage objects, and even local temporary resources. There will be requirements for translating security policies from existing serverful applications, and for offering highly-expressive security APIs for dynamic use in cloud functions. For example, a cloud function may have to delegate security privileges to another cloud function or cloud service.

4.5 Computer architecture challenges

Hardware Heterogeneity, Pricing, and Ease of Management

: Alas, the x86 microprocessors that dominate the cloud are barely improving in performance. Single program performance improvement only 3%. Assuming the trends continue, performance won't double for 20 years. Similarly, DRAM capacity per chip is approaching its limits; 16 Gbit DRAMs are for sale today, but it appears infeasible to build a 32 Gbit DRAM chip. A silver lining of this slow rate of change is letting providers replace older computers as they wear out with little disruption to the current serverless marketplace.

we see two paths for serverless computing to support the upcoming hardware heterogeneity:

1. Serverless could embrace multiple instance types, with a different price per accounting unit depending on the hardware used.
2. The cloud provider could select language-based accelerators and DSAs automatically. This automation might be done implicitly based on the

software libraries or languages used in a cloud function, say GPU hardware for CUDA code and TPU hardware for TensorFlow code. Alternatively, the cloud provider could monitor the performance of the cloud functions and migrate them to the most appropriate hardware the next time they are run.

5. Conclusions :

By providing a simplified programming environment, serverless computing makes the cloud much easier to use, thereby attracting more people who can and will use it. Serverless computing comprises FaaS and BaaS offerings, and marks an important maturation of cloud programming. It obviates the need for manual resource management and optimization that today's serverful computing imposes on application developers, a maturation akin to the move from assembly language to high-level languages more than four decades ago.

While serverful cloud computing won't disappear, the relative importance of that portion of the cloud will decline as serverless computing overcomes its current limitations. Serverless computing will become the default computing paradigm of the Cloud Era, largely replacing serverful computing and thereby bringing closure to the Client-Server Era.

6. Reference

- ▣ [http:// www.google.com](http://www.google.com)
- ▣ <https://scholar.google.com>

A Survey of Machine Learning For Big Data Processing

Ms. Hemangi Mali
Student Of MSC.IT

Pillai HOC College of Arts, Science And Commerce
Rasayani

Ms. Priyanka Sorte

Asst. professor, Department Of IT/CS
Pillai HOC College of Arts, Science And Commerce ,
Rasayani

Abstract - There is no doubt that big data are now rapidly expanding in all science and engineering domains. While the potential of these massive data is undoubtedly significant, fully making sense of them requires new ways of thinking and novel learning techniques to address the various challenges. In this paper, we present a literature survey of the latest advances in researches on machine learning for big data processing.

Index Terms- Machine learning, Advanced Learning Methods, Critical Issues, SP Techniques for Big Data .

1.Introduction

It is obvious that we are living in a data deluge era, evidenced by the phenomenon that enormous amount of data have been being continually generated at unprecedented and ever increasing scales. Large-scale data sets are collected and studied in numerous domains, from engineering sciences to social networks, commerce, biomolecular research, and security.

Over the past decade, machine learning techniques have been widely adopted in a number of massive and complex data-intensive fields such as medicine, astronomy, biology, and so on, for these techniques provide possible solutions to mine the information hidden in the data.

2.Definition of Machine Learning

Machine learning is a field of research that formally focuses on the theory, performance, and properties of learning systems and algorithms. It is a highly interdisciplinary field building upon ideas from many different kinds of fields such as artificial intelligence, optimization theory, information theory, statistics, cognitive science, optimal control, and many other disciplines of science, engineering, and mathematics. Because of its implementation in a wide range of applications, machine learning has covered almost every scientific domain, which has brought great impact on the science and society. It has been used on a variety of problems, including recommendation engines, recognition systems, informatics and data mining, and autonomous control systems .

3. Comparison of machine learning technologies

Learning Type	Data Processing task	Distinction Norm	Learning Algorithms
Supervised learning	Classification	Computational Classifiers	Support vector Machine
		Statistical Classifiers	Naive Bayes
			Hidden Markov Model
			Bayesian Network
		Connectionist Classifiers	Neural Network
Unsupervised learning	Clustering / Prediction	Parametric	K-means
			Gaussian Mixture Model
		Nonparametric	Dirichlet Process Mixture Model
			X-means
Reinforcement learning	Decision - making	Model-free	Q-learning
			R-learning
		Model-based	TD learning
			Sarsa Learning

4. Advanced Learning Methods

4.1. Representation Learning: Datasets with high dimensional features have become increasingly common nowadays, which challenge the current learning algorithms to extract and organize the discriminative information from the data .

4.2. Deep learning: Nowadays, there is no doubt that deep learning is one of the hottest research trends in machine learning field. In contrast to most traditional learning techniques, which are considered using shallow-structured learning architectures, deep learning mainly uses supervised and/or unsupervised strategies in deep architectures to automatically learn hierarchical representations .

4.3. Distributed and parallel learning: There is often exciting information hidden in the unprecedented

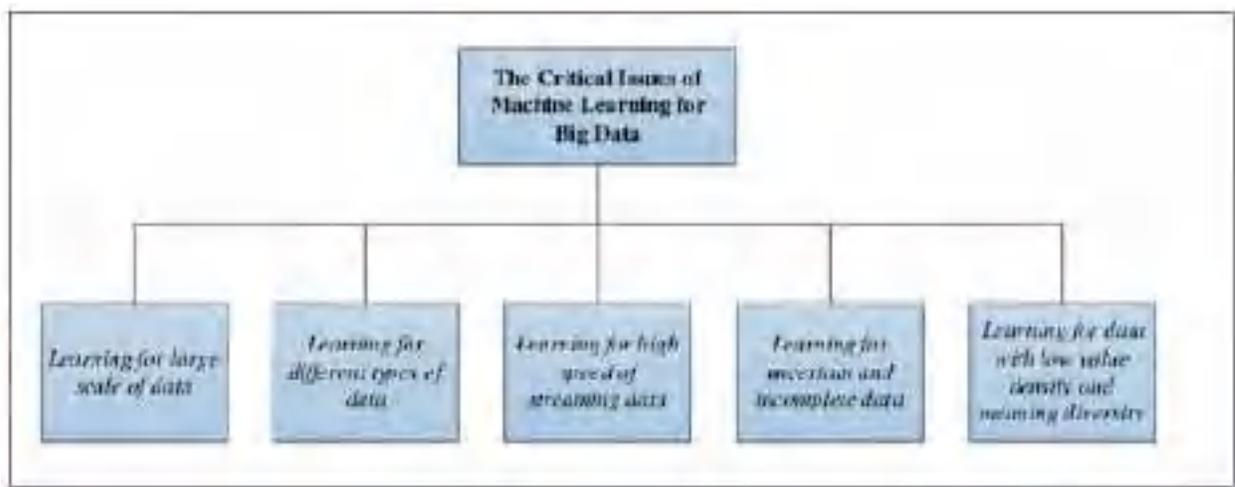
volumes of data. Learning from these massive data is expected to bring significant science and engineering advances which can facilitate the development of more intelligent systems.

4.4. Transfer learning: A major assumption in many traditional machine learning algorithms is that the training and test data are drawn from the same feature space and have the same distribution .

4.5. Active learning: In many real-world applications, we have to face such a situation: data may be abundant but labels are scarce or expensive to obtain. Frequently, learning from massive amounts of unlabeled data is difficult and time-consuming.

4.6. Kernel-based learning: Over the last decade, kernel based learning has established itself as a very powerful technique to increase the computational capability based on a breakthrough in the design of efficient nonlinear learning algorithms .

5. Critical Issues of Machine Learning for Big Data

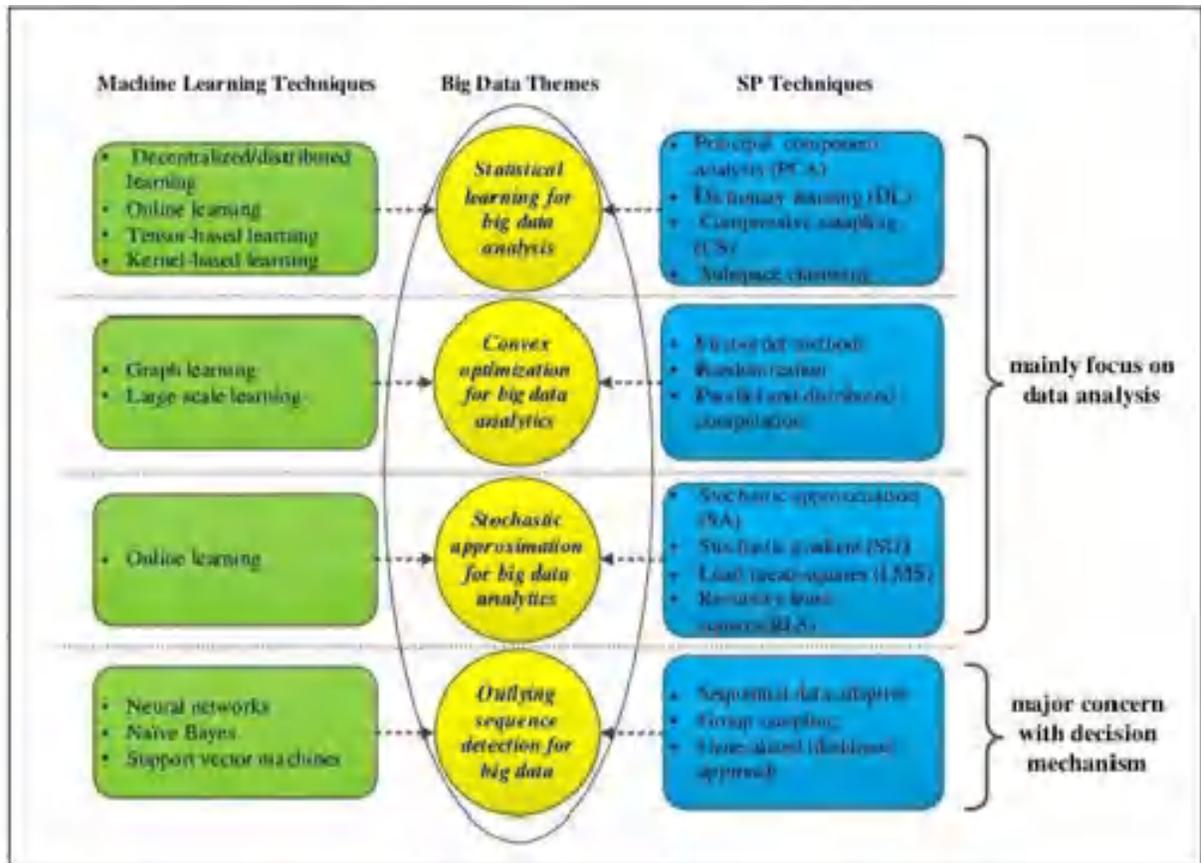


6. Connection of Machine Learning With SP Techniques for Big Data

There is no doubt that SP is of uttermost relevance to timely big data applications such as real-time medical imaging, sentiment analysis from online social media, smart cities, and so on []. The interest in big-data related research from the SP community is evident from the increasing number of papers submitted on this topic to

SP-oriented journals, workshops, and conferences. In this section, we mainly discuss the close connections of machine learning with SP techniques for big data processing. Specifically, we analyze the existing studies on SP for big data from four different perspectives. Several representative literatures are presented. we provide a review of the latest research progress which is based on these typical works.

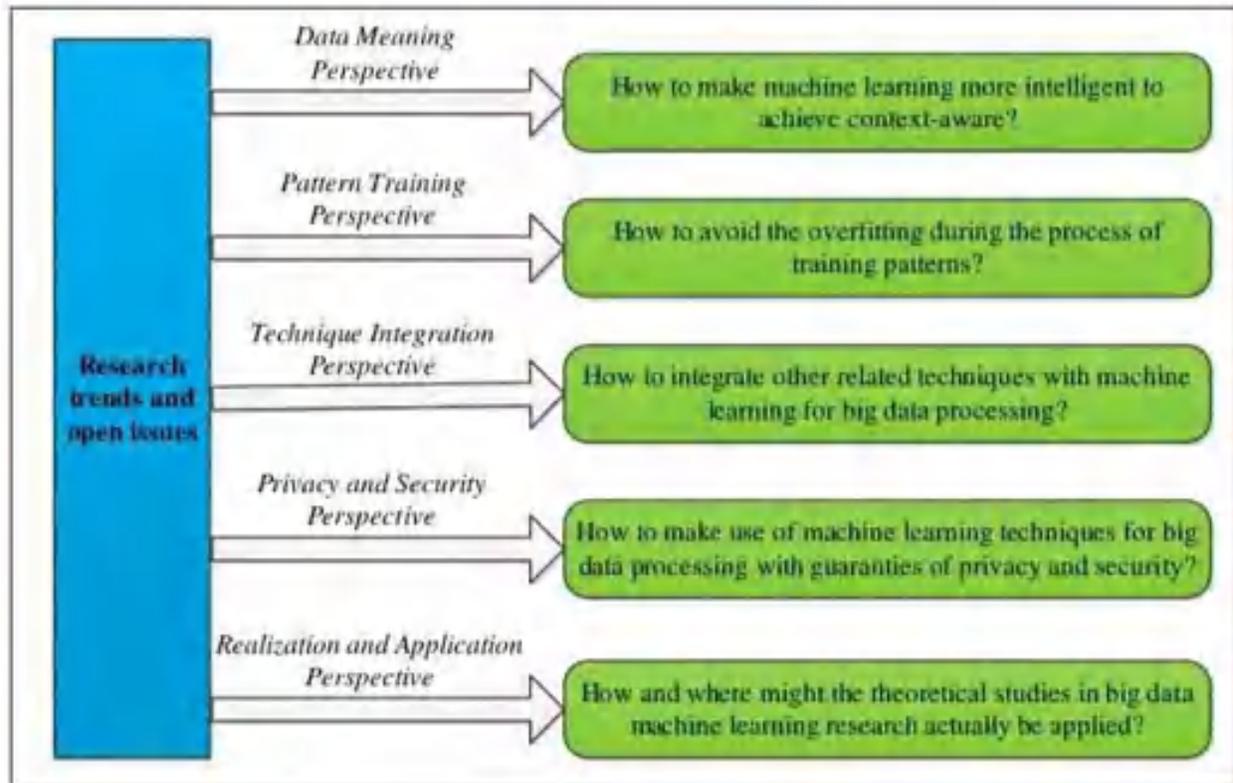
7. Connection of Machine Learning With SP Techniques for Big Data From Different Perspectives



8. Research Trends And Open Issues

While significant progress has been made in the last decade toward achieving the ultimate goal of making sense of big data by machine learning techniques, the consensus is that we are still not quite there.

The efficient preprocessing mechanisms to make the learning system capable of dealing with big data and effective learning technologies to find out the rules to describe the data are still of urgent need.



9. Conclusions

Big data are now rapidly expanding in all science and engineering domains. Learning from these massive data is expected to bring significant opportunities and transformative potential for various sectors. However, most traditional machine learning techniques are not inherently efficient or scalable enough to handle the data with the characteristics of large volume, different types, high speed, uncertainty and incompleteness, and low value density. In response, machine learning needs to reinvent itself for big data processing. This paper began with a brief review of conventional machine learning algorithms, followed by several current advanced learning methods. Then, a discussion about the challenges of learning with big data and the corresponding possible solutions in recent researches was given. In addition, the connection of machine learning with modern signal processing technologies was analyzed through studying several latest representative research papers. To stimulate more interests for the audience of the paper, at last, open issues and research trends were presented.

10. Reference

- www.google.com
- www.youtube.com
- <https://scholar.google.com>

Implementation of Website Builder Automation

Vishal Lagoo,
PG Student,

Pillai HOC College of Arts, Science & Commerce,
Rasayani

Priyanka Sorte,
Asst.Professor,

Pillai HOC College of Arts, Science & Commerce,
Rasayani

Abstract— Nowadays, every start-up, established companies have their own website to grow user engagement. Companies or businesses with a website are easily accessible. So everyone hires some developers or software company to design and manage their website. But the same can be done without prior knowledge of coding or web designing. There are many website automations / website builders available in the market which work on drag-drop basis or template basis. Our proposed AI based website builder works on template based automation which can be used to create product based websites in just a few minutes. This framework is developed in PHP CodeIgnitor using MVC architecture. MVC is used to manage backend functionality in models and load views as a website dynamically. This framework allows users to create multipage, responsive and search engine optimized websites. Users will get a separate web panel / app to manage and edit website data to make changes in website.

Index Terms— Artificial Intelligence, Website, Automations, Website Builders

1 INTRODUCTION

There are misconceptions among customers who are trying to get a website. Basically there are two types of websites, product based and service based. Service based websites are those who are developed by hiring a developer or company to develop a website. This website is all designed and developed according to users requirements and can be changed too. Product based websites are basically ready products but just a customer's data is changed. There are some pre-designed templates. User can select a design template according to his choice and page need. Product based website a content management tool which will be given to user to manage all his website content through a single panel or app. This tool doesn't require a single line of code or we can say it is totally coding free. Users have to just enter content data in respective fields and their website will be loaded using dynamic data. This proposed framework is exactly based on a product based website. Product based website offers multiple services like business email, email storage, domain, responsive designs, quick support, https security, rapid deployment, website analysis.

2 PROPOSED SYSTEM

Proposed AI based website builder works on template based automation which can be used to create product based websites in just a few minutes. This framework is developed in PHP CodeIgnitor using MVC architecture. MVC is used to manage backend functionality in models and load views as a website dynamically. This framework allows users to create multipage, responsive and search engine optimized websites. Users will get a separate web panel / app to manage and edit website data to make changes in website.

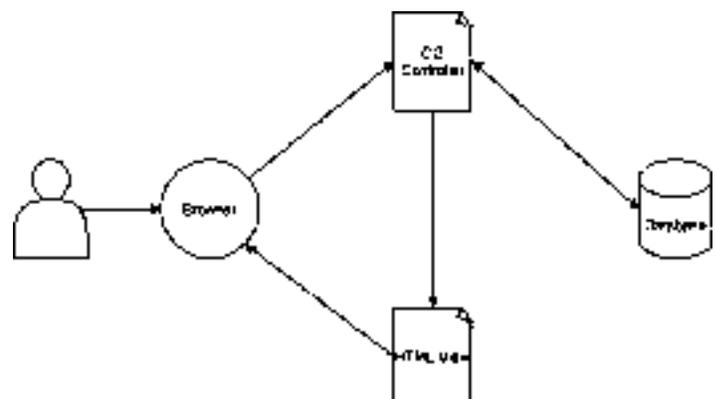
3 METHODOLOGY

3.1 Development Stack

Backend: PHP CodeIgnitor
Frontend: HTML, CSS, Bootstrap, Cordova, AngularJS
Database: MySQL

3.2 PHP CodeIgnitor

It is a framework designed specially for Application Development for people who build web apps using PHP. It provides a rich set of libraries with simple interfaces and logical structured libraries which helps to develop projects much faster if you're developing from a scratch. Also, CI provides customized URL routing. CI uses MVC (Model View Controller) Architecture to organise files. This allows developers to keep data, presentation and logical flow separate from each other.



3.3 Model

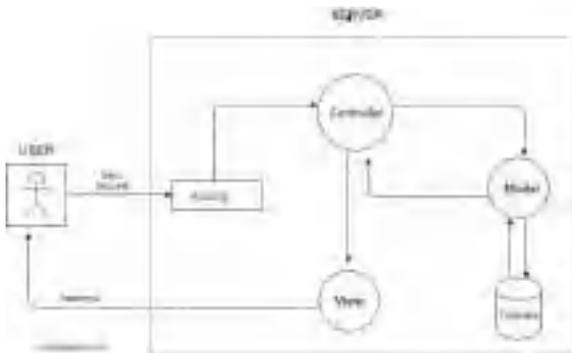
It manages the data of application and helps to implement business logic, APIs which application might need.

3.4 View

They are simple files with little or no logic that display information or logic. In our Website Builder, we are going to use view to load web pages, so it will manage all html files.

3.5 Controller

An entity which controls or manages connections between model and views.



3.6 Database

CodeIgnitor comes with featured and abstracted database classes that support both traditional structures and query builder patterns.

In this project, MySQL will be used for storing data. This data will consist of user data, website data such as pages, images, contacts and other configs.

Instead of storing each value in a separate column, we are using a json document to store key value pairs. So a whole json will be stored in a single column of MySQL database table. This will help to access the data much faster and easier and also will require much lesser programming in frontend.

4 ADVANTAGES & DISADVANTAGES OF MVC ARCHITECTURE

4.1 Advantages

1. It will separate the user interface from business logic and business logic from data access logic. So maintainability, flexibility and scalability are increased.
2. Reuse of components: CodeIgnitor MVC allows using UI components anywhere and anytime.
3. Team can work on different parts of a project at a time.
4. Helps to develop loosely coupled systems.
5. Testing of each and every component can be done as components are independent from each other.

4.2 Disadvantages

1. We need extra overhead due to layers which may cause a negative impact on the system sometimes.
2. Development of user-intensive applications can sometimes take longer if layering prevents the use of user interface components that directly interact with the database.
3. Use of layers helps to control complexity of large projects, but it can make simple projects more complex.

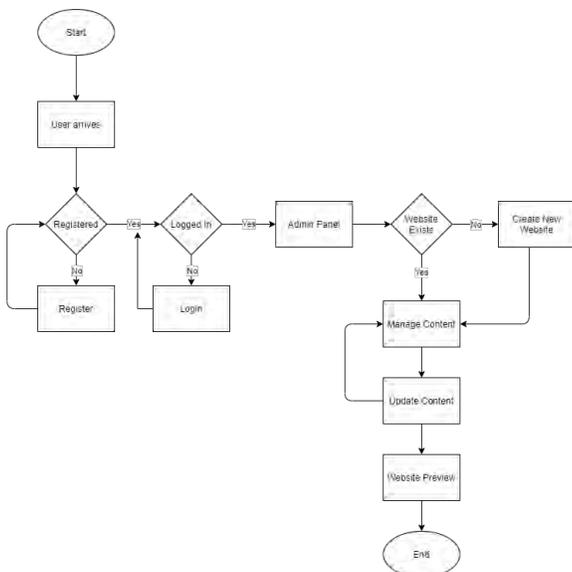
5 AUTOMATIONS

5.1 Image Color Picker

This piece of code will extract all the colors from the website logo and create an array using primary and secondary colors. Primary and Secondary colors will be ones which are most used in logos. This color array will be pushed to a json document containing website data in the database table. Also, these colors will be directly binded in website html using angularjs and thus, the website will automatically use colors from the logo so that it matches the color theme of the website logo.

5.2 Website Image Assets

All images like homepage sliders, product images, service images, gallery assets will be automated using a tag which the user will enter when creating a website. We are integrating the unsplash.com API and libraries which will fetch and make an array of all relative images from unsplash.com. For example, if a user owns a bakery shop and enters keywords / tags when creating a website as bakery, cake, muffins, this tag will be pushed to an array. An http get request will call an API and fetch all images related to these tags from the unsplash library. So these images will be binded in the website initially, and users can change them later manually.



Survey on Website Builder Automations

Vishal Lagoo,
PG Student,

Pillai HOC College of Arts, Science & Commerce,
Rasayani

Priyanka Sorte,
Asst.Professor,

Pillai HOC College of Arts, Science & Commerce,
Rasayani

Abstract— Nowadays, every start-up, established companies have their own website to grow user engagement. Companies or businesses with a website are easily accessible. So everyone hires some developers or software company to design and manage their website. But the same can be done without prior knowledge of coding or web designing. There are many website automations / website builders available in the market which work on drag-drop basis or template basis. Our proposed AI based website builder works on template based automation which can be used to create product based websites in just a few minutes. This framework is developed in PHP CodeIgnitor using MVC architecture. MVC is used to manage backend functionality in models and load views as a website dynamically. This framework allows users to create multipage, responsive and search engine optimized websites. Users will get a separate web panel / app to manage and edit website data to make changes in website.

Index Terms — Artificial Intelligence, Website, Automations, Website Builders

1 INTRODUCTION

There are misconceptions among customers who are trying to get a website. Basically there are two types of websites, product based and service based. Service based websites are those who are developed by hiring a developer or company to develop a website. This website is all designed and developed according to users requirements and can be changed too. Product based websites are basically ready products but just a customer's data is changed. There are some pre-designed templates. User can select a design template according to his choice and page need. Product based website a content management tool which will be given to user to manage all his website content through a single panel or app. This tool doesn't require a single line of code or we can say it is totally coding free. Users have to just enter content data in respective fields and their website will be loaded using dynamic data. This proposed framework is exactly based on a product based website. Product based website offers multiple services like business email, email storage, domain, responsive designs, quick support, https security, rapid deployment, website analysis.

2 EXISTING SYSTEM AND COMPARATIVE STUDY

There are multiple website builder automations available in the market. Some of the builders use drag and drop feature whereas some of them use just form control options. Following are some of the website builders available in the market.

2.1 Wordpress

It is a free, open source and most popular website builder and hosting solution in the market and nearly 39% of websites are powered by wordpress. There are thousands of

pre-designed templates / themes available that can be used for website design. Also it offers a drag & drop feature to design websites on our own. It also provides complete control on every aspect of the website and a variety of plugins that can be useful to integrate in the website.

2.2 Wix

It is another popular giant company in the website building and hosting market. It offers powerful tools and features to easily build and enhance performance of websites. Wix offers pre-build and editable templates with their drag and drop feature. Also, it offers a free plan with limited bandwidth and storage along with free ssl certificate. But moving away from wix websites is too much complicated.

2.3 BigCommerce

If someone is looking to build an online store then BigCommerce is the best choice among all builders. They have native integration with WordPress which allows you to leverage the flexibility of WordPress while taking advantage of the ecommerce power of BigCommerce. Also, they integrate payment gateways like PayPal, Apple Pay, Amazon Pay, Visa without charging transaction fees.

2.4 Weebly

Provides tons of great designs and functionality with beautifully crafted page builder which allows you to edit your website without having any coding skills. As it is a hosted platform, any kind of installation is not required.

2.5 GoDaddy

It is the largest name in website hosting and domain field in the world. They offer a simple website builder with hosting included. It also comes with a photo library using which getting professional, noncopyrighted photos for websites is made easy. It also runs on small screen devices

like mobile or tablet.

3 PROPOSED SYSTEM

Proposed AI based website builder works on template based automation which can be used to create product based websites in just a few minutes. This framework is developed in PHP CodeIgnitor using MVC architecture. MVC is used to manage backend functionality in models and load views as a website dynamically. This framework allows users to create multipage, responsive and search engine optimized websites. Users will get a separate web panel / app to manage and edit website data to make changes in website. This proposed builder will automate following things:

3.1 Website Colours

Primary and secondary colours will automatically be detected from the website logo, and websites will be designed automatically using these colours. This can be changed manually from the user's panel or admin panel

3.2 Dynamic Images

If a user enters a process or service related tag during registration process, automatically that tag related images will be automatically inserted in respective pages. For example, if it is an interior designer's website, and the user enters a tag 'interior', 'decorations', 'interior designing', images related to this will be automatically fetched and inserted in service / product or any other respective page. Also, this image data can be changed or replaced later manually using a content management tool or from the admin panel.

4 ADVANTAGES

4.1 Product based websites

This is a type in which website designs are available as direct product, instead of collecting user requirements and developing one. So these are much cheaper in rates. Product based websites cost nearly about Rs.299 per month including domain name, email storage, SSL security, content management tool with analytics app whereas typical service based websites may cost one time Rs.10000 or more with maintenance charges.

4.2 Free registration

Registration is simply free; users have to pay before making the website live.

4.3 Pricing

These websites start from Rs.299 per month.

4.4 Easy to use

Content management tool, analytical app and website creation process is simple to use, which consist of simple user

interface.

4.5 Instant deployment

When a user fills all the data, the website gets instantly ready. After payment, the website will be instantly live.

4.6 Responsive design

A website is said to be responsive when the layout / content responds or adapts based on the size of screen which they are presented on. It automatically changes design when screen width is changed. There are basically 4 screen sizes, large (big desktop / laptop screens), medium (medium desktop screen), small (tablet screen), extra small (mobile screen). So when a website is made responsive, it automatically changes its design according to screen size. But here's the catch, this doesn't need special coding for different screen sizes, single design changes automatically, thanks to Bootstrap.

4.7 Attractive Pre-designed templates

This framework will have multiple design templates available from which the user can choose one according to his page requirements, attractiveness. Users data will dynamically bind into this selected template design. Also, users can change the design template before making the website live.

4.8 Content Customization

User can customize the website whenever he wants. A tool for content management and customization will be provided to users so even after the website goes live, users can add or change images, content sequence, add / update / delete business services & products.

4.9 No coding skills required

No coding required, just enter data in respective UI fields and content will be updated at the website.

4.10 SSL Security

Websites will be SSL secured.

4.11 Support 24x7

There will be a live support option if users have some queries or concerns regarding website, payment or content management.

5 CONCLUSION

The main aim of this project is to automate the website design and content management tasks for both developer and end user. This will allow developers to quickly set up product based websites within a short period of time and end users to get a website and manage content on their own in price. This proposed project framework not only reduces cost but also saves time and hence can be used for small startups, business, institutes, etc.

FAKE PRODUCT REVIEW MONITORING SYSTEM

Jitesh Kirdakude, Prof. Priyanka Sorte

Abstract- In the current scenario, the data on the web is growing exponentially. Social media is generating a large amount of data such as reviews, comments, and customer's opinions on a daily basis. This huge amount of user generated data is worthless unless some mining operations are applied to it. As there are a number of fake reviews so opinion mining technique should incorporate Spam detection to produce a genuine opinion. Nowadays, there are a number of people using social media opinions to create their call on shopping for product or service. Opinion Spam detection is an exhausting and hard problem as there are many faux or fake reviews that have been created by organizations or by the people for various purposes. They write fake reviews to mislead readers or automated detection system by promoting or demoting target products to promote them or to degrade their reputations. The proposed technique includes Ontology, Geo location and IP address tracking, Spam words Dictionary using Naïve Bayes, Brand only review detection and tracking account used. Some companies exhibit their own review for the demand of product and company rating purpose. To resolve this problem to find out fake review in the website this "Fake Review Monitoring" system is introduced. This system includes with verification process of reviews by the reference of IP address and then separate them into spam and non-spam reviews.

Keywords: Fake, Review, Products, Online, Shopping, Genuine, Proper, Ratings, spam review detection

----- ◆ -----

INTRODUCTION

One of the very rapid growth area is ecommerce. Generally e-commerce provide facility for customers to write reviews related with its service. The existence of these reviews can be used as a source of information. For examples, companies can use it to make design decisions of their products or services but unfortunately, the importance of the review is misused by certain parties who tried to

create fake reviews, both aimed at raising the popularity or to discredit the product. They share their thoughts on internet.

Before purchasing anything, it is a normal human behaviour to do a survey on that product. Based on reviews, customers can compare different brands and can finalize a product of their interest. These online reviews can change the opinion of a customer about the product. If these reviews are true, then this can help the users to select proper product that satisfy their requirements. On the other hand, if the reviews are manipulated or not true then this can mislead user. This boosts us to develop a system which detect fake reviews for a product by using the text and rating

property from a review. The honesty value and measure of a fake review will be measured by utilizing the data mining techniques.

An algorithm could be used to track customer reviews, through mining topics and sentiment orientation from online customer reviews and will also blocked the fake reviews.

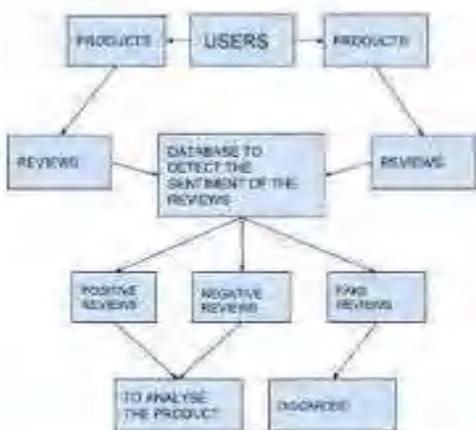


Fig: The Process of Product Review Analysis

LITERATURE SURVEY

As of late, the World Wide Web has greatly changed the technique for imparting the insights. Online audits are feedback, tweets, posts, conclusions on numerous on line tiers like survey destinations, news locales, net based totally enterprise destinations or some different lengthy range interpersonal communication destinations. Sharing audits is one of the strategies to compose a survey approximately administrations or items. Surveys are considered as a person's near domestic concept or experience about items or administrations. Client dissects reachable audits and takes preference whether to shop for the object or no longer. In this way on-line audits are critical wellspring of information approximately patron conclusions. Phony or spam audit

alludes to any spontaneous and superfluous facts about the item or administration. Spammer composes counterfeit audits approximately the contenders' item and advances possess items. The surveys composed by spammers are called phony audits or unsolicited mail audits. In this manner counterfeit surveys discovery has grown to be fundamental trouble for customers to decide higher preference on items dependable simply because the sellers to make their purchase.



Fig: Methodology to accomplish objectives of the research.

• Research and Survey:

In this one review's the most recent researches minutely related in the theory problem statement and the question. Once we analyse the thesis, the existing procedures in spam detection for non-Arabic opinion reviews, identifying the drawbacks or the flaws in the existing approaches, we prepare the strategies and solutions on how to proceed or extend in order to be overcome in our research.

• Data Acquisition:

In this step, we prepare an in-house data set of spam reviews and reviewers using human collected from online e-commerce

websites or application like Amazon, Flipkart with different characteristics and sizes. The records are chosen randomly from any of the records that are available on the website.

• **Data Integration:**

In this step, we combine the data from multiple review source data sets into a coherent form.

• **Spam Identification Labelling:**

In this step, we look for various types of the spam in the data integrated set, and labelled each record as spam and non-spam manually.

• **Pre-processing:**

In this step, we use various types of pre-processing techniques to handle the missing, noisy and inconsistent data. There are a number of pre-processing techniques such as case folding, character erase, tokenization, slang word handling, stop word removal, stemming and number handling.

• **Processing Stage:**

In this step, we will first have to implement the following steps:

1. Data mining classification.
2. Text mining classification.
3. Data-Text mining classification.

• **Evaluate the Approach:**

In this step, analyzation of the outcome and rationalize the feasibility of the approach we followed by comparing it with other previous approaches.

METHODOLOGY

Commentator Centric Approach-This method is predicated upon the behaviour of analysts. This methodology considers data about clients and all surveys that are composed with the aid of them. Highlights

utilized proper now account age, profile image, URL duration, IP address, variety of composed audits through one commentator, maximum severe rating every day and so on. Item Centric Approach-This technique for the most component facilities around the object related information. Right now, rank of object, value of object and so on are taken into consideration as highlights. There are special methods to differentiate counterfeit surveys. AI system is one of the procedures to distinguish counterfeit surveys. AI model learns and make forecast. The essential advances associated with AI are records making ready, highlight extraction, include determination, characterization model age.

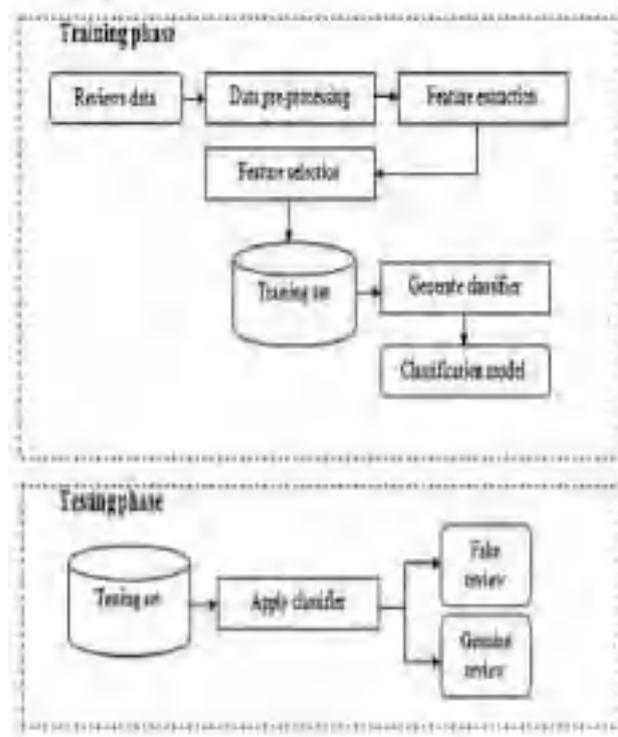


Fig: Fake review detection system

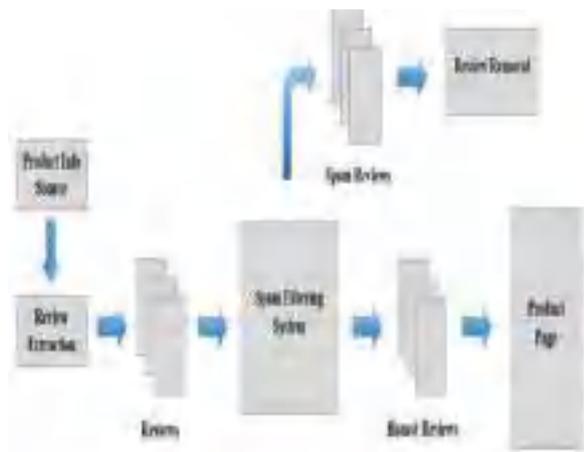


Fig: Working of system

RESULT

Now a day's technology is growing day by day and there are so many website and application are available in the online market by which they sell their product and on that products there are millions of reviews available on base of reviews user buy the product most of the time. There are some organization which posting fake reviews on fake product or on genuine product and user gets stuck. Our application which will help the user to pay for the right product without any getting into any scams. Our application will do analysis and then post the genuine reviews on genuine product. And user can be sure about the products availability on that application and reviews too.

Fake product review detection system will be useful to business organization as well as to customers. When used on e-commerce dataset the accuracy will be around 90%-96% if the data is going to be a labelled and supervised data, but, if the data is a unsupervised unlabelled data then the accuracy might vary in comparison to supervised data.

CONCLUSION

This work has investigated the user's request and has given the particular prediction and safety measure for that input. The K-Means clustering technique was used to find the clusters and fatality for the flight crash investigation. This method will also consider other factors like efficiency, weather impact and schedules of other aircraft. Results showed that the proposed algorithm obtained prediction based on user's input with efficiency. The proposed algorithm also provided improved search results for the query given by the user. Possible future work is to improve the efficiency and also increase the count of clusters used. Our software will help the user to pay for the right product. Our software will do analysis and then if any fake review is found from any IP address consistently then admin will block the IP address. In this way it monitors the fake review made on any product. And user can be sure about the products availability on the application and reviews too.

REFERENCES

1. Shirani-Mehr, Houshmand. "SMS spam detection using machine learning approach." (2013)
2. <http://www.cs.uic.edu>.
3. Fake product review monitoring survey, <https://goo.gl/forms/alqmW0q0iObLmIU23>
4. A. Rastogi, M. Mehrotra, "Opinion spam Detection in Online Reviews", Journal of information and Knowledge Management

A Study of Efficient and Secure Data Storage in Mobile Cloud Computing.

Neha Mahadik¹

Pursuing Master's degree program in IT from Pillai HOC College of Arts, Science, & Commerce, Rasayani.
E-mail:- mahadikneha19@gmail.com
PH:- 9404841656

Priyanka Sorte²

Asst. Professor, Department Of IT/ CS Pillai HOC College of Arts, Science & Commerce, Rasayani.
E-mail:- psorte@mes.ac.in
PH:- 9689157442

Abstract:- Cloud Computing is providing services like servers, database management, data storage, networking, software applications, etc. to the client on demand. Data storage is one of the most widely used service provided by cloud computing. Mobile Cloud Computing is a combination of mobile computing and cloud computing to provide computational resources to Mobile users, Network Operators and Cloud providers. Data stored by users have a great risk of altering and there's a huge risk of data integrity is lost. As data of different organizations is stored at one place on the server of cloud provider there is also risk of data sold illegally by the provider itself. In this paper we will study about security of data storage in MCC (Mobile Cloud Computing).

Index Terms:- Cloud Computing, Mobile Cloud Computing, Data Storage, Data Security.

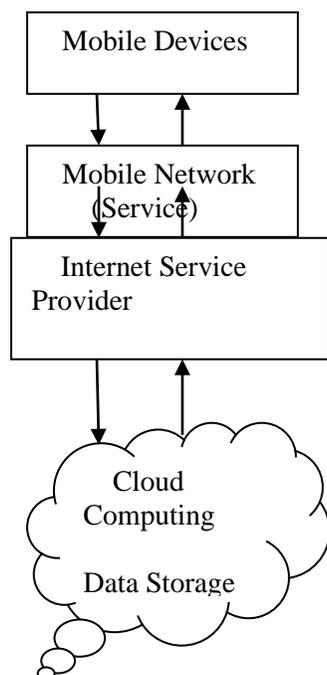
INTRODUCTION

Cloud Computing in delivery of computing services to individuals and businesses over the internet. Cloud services allow them to utilize system software and hardware that are managed by the cloud providers and allow them to access them anywhere, anytime through internet. Examples of cloud services contain online file storage, webmail, social networking sites, online business applications. Cloud computing is new an promising technology that is being accepted by various entrepreneur organizations and companies. Because of services provided by cloud computing the users don't have to install there software and hardware and data storage servers which help them saving the cost for the same. In Mobile Cloud Computing users can stored their data and enjoy high quality on demand cloud computing application without the limitation of having to purchase and maintain their own local hardware and software. However, data stored by the users is always at a risk of getting altered by external sources. That's the reason cloud computing is not adopted more widely. Data breaching is possible in cloud environment since data from various users and organizations lie together in a cloud. As the user not only sends data to the cloud but also transfers control to a third person which may raised concern about security. The Cloud Service Provider (CSP) has the responsibility to keep data secure and maintain its confidentiality and integrity but sometimes the Cloud Service Provider itself may used or corrupt data illegally. However, there is a simple solution to this the user just have to encrypt the data before uploading it to the cloud. This will help the user to keep data safe and it will not be visible to external users or the administrator itself.

Mobile Cloud Computing

- ✚ The factors included in Mobile Cloud Computing are standard Cloud Computing, wireless communication, infrastructure portable computing devices, location based services, mobile internet etc.
- ✚ Users are experiencing both unlimited online computing power and storage thank to Mobile Cloud Computing
- ✚ Due to wireless network access to cloud storage and computational resources MCC has become a prominent and powerful model.
- ✚ Unfortunately mobile devices are battery operated and has limited processing power, low capacity storage, less security feature and low energy which causes problems for applications that require high computational capabilities and demand large process capacity.
- ✚ So in order to overcome this huge computing capabilities, storage capacity and battery life these high computational capabilities and large storage capacity demanding activities should be transferred to cloud.
- ✚ However, cloud data security is still is a concern and is the main obstacle to mobile cloud computing for being widely adopted.

Mobile Cloud Computing Architecture



Architecture of Mobile Cloud Computing

- ✚ By using the base stations (base transceiver stations or BTS, access points, satellites), mobile devices can connect to the mobile networks.
- ✚ These mobile base stations establish and control the connections or air link and the functional interfaces between both the mobile devices and the networks. The location and the ID of the mobile users are sent to the central processors connected to the servers that provide mobile network services.
- ✚ With such setup, the mobile network operators can give AAA (for authentication, authorization, and accounting) services to mobile users based on the home agent (HA) and subscribers information stored in the database.
- ✚ Subscribers request are sent to a cloud via the internet where the requests are processed by the cloud controller, thus, providing the users with the corresponding cloud services

Applications of Mobile Cloud Computing

✚ Mobile Learning

Mobile learning applications provide users to learn via mobile gadgets like tablets, laptops, mobile phones etc. This applications providers m-learners flexibility as they can access the application anytime anywhere from any portable device. The cloud-based m-learning applications have more faster processing speed and more educational resources at high network transmission rate. When m-learning is combine with cloud computing the interaction between students and teachers has increased.

✚ Mobile Gaming

Mobile devices do not have high processing power which is essential for graphic rendering in mobile games. Due to which mobile cloud gaming came to existence. It is a gaming model in which the players use mobile devices to play high graphic games that run remotely on cloud server. in this technology is also known as gaming as a service.

✚ Mobile Healthcare

Mobile healthcare is a measure extension of electronic healthcare which enables the doctor or care givers to have access to patients clinical data and learn latest medical knowledge. It also allows patients with regular conditions with health issues to the main under constant observation without needing to be physically present at the clinic. In traditional medicine applications we have restrictions such as less storing capacity, minimal security and high rate of medical errors but this restrictions diminished after applying mobile cloud computing concept in health care applications.

✚ Mobile Commerce

Mobile commerce or m-commerce is accessing the internet for business transactions using mobile devices like smart phones, tablets, notebook, pc etc. As this devices provide users mobility and probability rapid growth has been taken place in m-commerce but some measure challenges faced by m-commerce are low processing power, low network bandwidth, limited storage and security. However, cloud integrated mobile applications gives lots of advantage for business industry and the user as well.

Mobile Cloud Computing Security

One of the major concerns of cloud operators is securing the mobile computing users privacy and the data or application integrity. The mobile users as well the application developer are benefiting from the large storage area the cloud they must also be aware of the danger when dealing with data and application integrity. As Mobile Cloud Computing is the combination of cloud computing and mobile networks thus issues relating to the security can be divided into Cloud security and Mobile Network Security respective.

Mobile Network Users Security

There are many types of security vulnerabilities and threats to different mobile devices such as smart phones, laptops, PDAs, etc. Some of the applications can also cause security and privacy issues for mobile users which are discussed below.

• Security For Mobile Applications

One of the simple way to detect any security threat to install and run any security software or an antivirus program on the device. However, this devices have limited processing capabilities and power it is difficult to protect this device against threat. Several methods have been developed to transfer the security and threat detection mechanism such as authentication, authorization and data encryption to the clouds verifications will be carried out to make sure that all the contents send to the mobile devices are not malicious in nature. Rather than running the antivirus and threat detection software locally mobile devices only need to perform light weight functions such as executing trace transmitted to the cloud security servers.

• Data Access and Security

Issues related to data access and data security are important for application that rely on remote data storage and need internet access to work. Sometimes data stored by a user can be accessed altered or deleted by a malicious user illegally. In order to keep data safe from such malicious users can encrypt data and then store it to the cloud.

Cloud Security

Security information on the cloud: Individual users, organizations and enterprisers know the advantage of storing large amount of data and application on cloud. But this may also raised the issues of loss of data integrity and authentication. The data stored by users should be authenticated only to the user itself. To authenticated data the user can put a pin or password

to access it. However, sometimes data stored by users can get accessed by some hackers and the integrity of data is lost. Therefore, to maintain the integrity the data stored by the user should be in encrypted format.

Secure Data Storage

The cloud data storage system, users store their data in the cloud itself while no longer have the data locally. Therefore, the correctness and availability of the data files are being stored on the distributed cloud servers as well as they are guaranteed too. Data storage security also includes the legal access to the data stored in the cloud, namely access authorization and authentication security data sharing. It also encrypt the store data to ensure data confidentiality and it design a file system which supports secure data of deletion. The secure data accordingly refers to manual and Secure data storage as collectively refers automated computing processes and technologies. It used to ensure stored data security and integrity. This can include protection of the hardware on which the data is stored as well as software security.

A Study of Data Warehousing

Neha Mahadik¹

Pursuing Master's degree program in
IT from Pillai HOC College of Arts,
Science, & Commerce, Rasayani.
E-mail:- mahadikneha19@gmail.com
PH:- 9404841656

Priyanka Sorte²

Asst. Professor, Department Of IT/ CS
Pillai HOC College of Arts, Science
& Commerce, Rasayani.
E-mail:- psorte@mes.ac.in
PH:- 9689157442

Abstract:- A data warehouse is a collection of business data which is used to help an organization and make decisions. The concept of the data warehouse has existed since the 1980 and it was developed to help transition data. The term data warehousing indicates the whole set of interrelated activities involved in designing, implementing and using a data warehouse. Here, we will learn the features of data warehouse, architecture of data warehouse, goals and functions of data warehouse.

Index Terms:- Internal, External, Personal Data, Data warehouse architecture

INTRODUCTION

The term data warehousing indicates the whole set of interrelated activities involved in designing, implementing and using a data warehouse. Data warehousing is the process of consolidating all the organizational data into one common database. A data warehouse is a database designed to enable business intelligence activities it exists to help users understand and enhance their organization's performance. It is designed for query and analysis rather than for transaction processing, and usually contains historical data derived from transaction data, but can include data from other sources. Data warehouses separate analysis workload from transaction workload and enable an organization to consolidate data from several sources. In addition to a relational database, a data warehouse environment can include and extraction, transportation, transformation, and loading solution, statistical analysis, reporting, data mining capabilities, client analysis tools, and other applications that manage the process of gathering data, transforming it into useful, actionable information, and delivering it to business users. The data warehouse works with data collected from multiple sources. It may involve transactions, production, marketing, human resources and more. In today's world of big data, the data may be many billions of individual clicks on web sites or the massive data streams from sensors built into complex machinery. Data warehouses are distinct from online transaction processing (OLTP) systems. With a data warehouse you separate analysis workload from transaction workload. Thus data warehouses are very much read-oriented systems that all users can look to; it prevents many disputes and enhances decision-making efficiency.

Internal Data

Internal data are stored for the most part in the databases, referred to as transactional systems or operational systems, that are the backbone of an enterprise information system. Internal data are gathered through transactional applications that routinely preside over the operations of a company, such as administration, accounting, production and logistics. This collection of transactional software applications is termed enterprise resource planning (ERP). The data stored in the operational systems usually deal with the main entities involved in a company processes, namely customers, products, sales, employees and suppliers. These data usually come from different components of the information system

External Data

There are several sources of external data that may be used to extend the wealth of information stored in the internal databases. For example, some agencies gather and make available data relative to sales, market share and future trend predictions for specific business industries, as well as economic and financial indicators. Other agencies provide data market surveys and consumer opinions collected through questionnaires. A further source of external data is provided by geographic information systems (GIS), which represent a set of applications for acquiring, organizing, storing and presenting territorial data

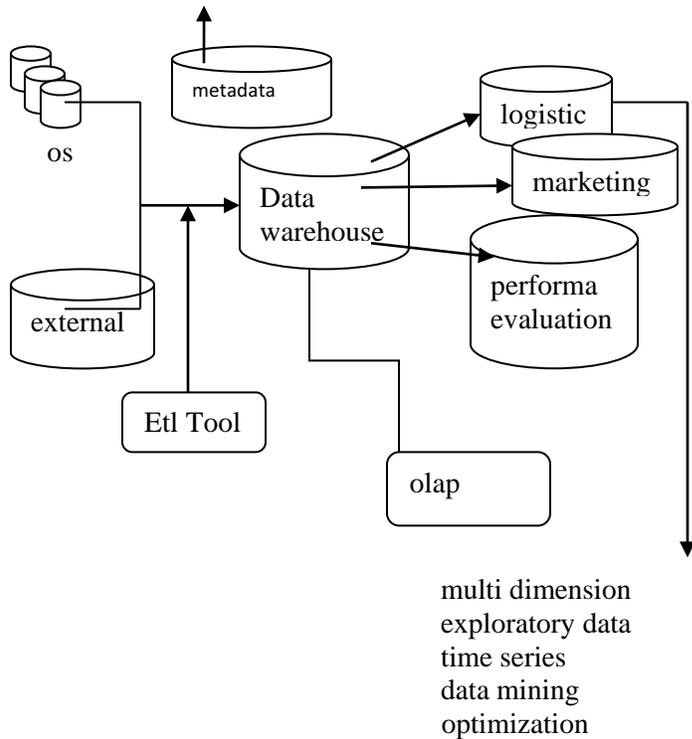
perspective

Personal Data

In most cases, decision makers performing a business intelligence analysis also rely on information and personal assessments stored inside worksheets or local databases located in their computers. The retrieval of such information and its integration with structured data from internal and external sources is one of the objectives of knowledge management systems.

Data warehouse architecture

The reference architecture of a data warehouse, shown in Figure includes the following major functional components.



The data warehouse itself, together with additional data marts, that contains the data and the functions that allow the data to be accessed, visualized and perhaps modify.

Data acquisition applications, also known as extract, transform and load (ETL) or back-end tools, which allow the data to be extracted, transformed and loaded into the data warehouse. Business intelligence and decision support applications, which represent the front-end and allow the knowledge workers to carry out the analyses and visualize the results.

The three-level distinction applies to the architecture shown in the above figure even from a technological

The level of the data sources and the related ETL tools that are usually installed on one or more servers. The level of the data warehouse and any data mart, possibly available on one or more servers as well, and separated from those containing the data sources. This second level also includes the metadata documenting the origin and meaning of the records stored in the data warehouse. The level of the analyses that increase the value of the information contained in a data warehouse through query, reporting and possibly sophisticated decision support tools. The applications for business intelligence and decision support analysis are usually found on separate servers or directly on the client PC used by analysts and knowledge workers.

ETL Tools

ETL refers to the software tools that are devoted to performing in an automatic way three main functions: extraction, transformation and loading of data into the data warehouse.

Extraction

During the first phase, data are extracted from the available internal and external sources. A logical distinction can be made between the initial extraction, where the available data relative to all past periods are fed into the empty data warehouse, and the subsequent incremental extractions that update the data warehouse using new data that become available over time.

Transformation

The goal of the cleaning and transformation phase is to improve the quality of the data extracted from the different sources, through the correction of inconsistencies, inaccuracies and missing values.

Metadata

In order to document the meaning of the data contained in a data warehouse, it is recommended to set up a specific information structure, known as metadata, i.e. data describing data. The metadata indicate for each attribute of a data warehouse the original source of the data, their meaning and the transformations to which they have been subjected. The documentation provided by metadata should be constantly kept up to date, in order to reflect any

modification in the data warehouse structure. The documentation should be directly accessible to the data warehouse users, ideally through a web browser, according to the access rights pertaining to the roles of each analyst.

External

External data is defined as any data not contained in the company's operating systems. It can be data that the company has, but is not in an operating system. ... Data from internal sources the company already has, such as forecast data, can be included in the data warehouse.

Operational System

An operational system is a term used to refer to a system that is used to process the day-to-day transactions of an organization. These systems are designed in a manner that processing of day-to-day transactions is performed efficiently and the integrity of the transactional data is preserved.

Logistics

The success of any logistics company lies in efficient inventory management and warehousing, automation of internal business processes, fast delivery and taking care of safe storage and quality of goods.

Marketing

A marketing data warehouse is a cloud based destination for storing and analyzing cross channel marketing data. Data warehouse allows marketing and analytics terms to consolidate data from multiple performs include advertising channels like facebook and google, web analytics platform like google analytics like hubspot and salesforce among others.

Performance Evaluation

In recent years, data warehouse (DW) systems have become more important for decision making processes. This was due to the fact that they are capable of enhancing the value of an organization. DW used for storing large amount of data and making it available for decision makers and business analyst to make queries, analysis and planning without considering data changes in the operational database.

Multidimensional

Multidimensional databases are used mostly for OLAP (online analytical processing) and data warehousing. They can be used to show multiple dimensions of data to users . A multidimensional database is created from multiple relational databases. The data in multidimensional databases is stored in

a data cube format.

Time Series Analysis

Time series analysis is a statistical technique that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. Cross-sectional data

Data Mining

Data mining is the process of analyzing unknown patterns of data. A data warehouse is database system which is designed for analytical instead of transactional work. Data mining is a method of comparing large amounts of data to finding right patterns.

Optimization

Data warehouse optimization (DWO), it's a strategy for identifying the "right" workloads for your data warehouse. In other words, it's making sure you're not allocating data warehouse resources on tasks that other technologies can handle more cost-effectively.

CONCLUSION

In this research paper we have discussed Internal Data, External Data, Personal Data and Data warehouse Architecture.. This paper has also evaluated Internet of things (IOT).This position paper presents the case of using artificial intelligence techniques and business as well.

REFERENCES

Hammergren, Tom, Data Warehousing Strategies, Technologies, and Techniques, New York: McGraw-Hill, 1996.
Humphries, Mark, et al., Data Warehousing: Architecture : Prentice-Hall PTR, 1999. Inmon, W. H., Building the Data Warehouse (2nd Edition), New York: Wiley, 1996.

USE OF AUTOMATION (RPA) IN SHOPPING WEBSITE

Mukeshkumar Mishra

PG Student,

Pillai Hoc College of Arts, Science & Commerce,
Rasayani

Prof. Priyanka Sonawane

Asst.Professor,

Pillai Hoc College of Arts, Science & Commerce,
Rasayani

Abstract— Nowadays, every start-up, established companies or businesses are moving toward automation to tackle the repetitive, easy tasks that plague customer service representatives, which helps to deliver a better experience to customers. In this project we are using the same concept and technology of automation in shopping website using the RPA tool i.e. UiPath. It means we are going to use automation process for repetitive and time-consuming tasks which consumes more time for same process and also requires more people work. Which will result in time saving and cost reduction, so that we will be able to give more time to complex problems. So that we can provide better customer experience. Use of automation also help into Lower operational risk by eliminating human errors such as tiredness or lack of knowledge, RPA reduces the rate of errors thereby providing a lower level of operational risk. In this paper we will see, how we can automate most of the possible operations and processes of shopping website.

Index Terms— Automation, Business, Bot, Chatbot, Robots, RPA (Robotic Process Automation), UiPath, Shopping Website.

1. INTRODUCTION

In past years we have seen many companies and businesses are shifting towards automation and found a significant growth in terms of cost and time as well, and in customer experience and engagement. Now every sector of businesses or organizations are trying to move and adapt the technology of automation in their work field.

But still the small-scale companies and businesses lacking behind in the adoption of automation in their process now might be there are some reasons such as “trust issue” or “lack of implementation”, still they believe or rely on the manual work.

In this project we are going to use RPA automation techniques in our shopping website that proves that we can implement the automation in any sector it might be either small or large, every sector with proper implementation can use automation and automate their processes which consume more time and more people for same and repetitive process.

In this project we are using the automation in most of the possible process such as auto reply to customer email and solve the customers queries on the website itself using the chatbot, and the repetitive

processes such as invoice processing, shipment validation, customer profile updating and etc.

Where we will use the famous and most used RPA tool UiPath. Where we will create and assign software robots also called as bots using various UiPath tools which will mimic the human action and will try to perform respective process automatically.

1.1. TERMINOLOGY

Before getting into the project let's explore the basic concepts and terminologies which we are going to use in this project or over which we are going to work.

- **Automation:**

The term automation is derived from the Greek word's autos meaning self, and motos, meaning moving, in simple words, is technology that deals with the application of computers and machines for the maintenance and production of goods and services. Which helps to get work done without human help or assistance and if any then less, compare to traditional systems. With the arrival of computer, there were many software's and

software systems developed to complete tasks that were previously done on the paper to handle and manage businesses or they kept as it is and not done at all because of lack of tools. Some of the examples are like inventory management, bookkeeping and communication management.

- **RPA:**

Robotic process automation (RPA) means the process where basic and complex tasks are gets automated through the software or hardware system which function with many and number of applications, like human workers used to do. RPA software is designed to perform repetitive and simple task and reduce the burden for humans. Which helps employees to give more time to complex problem rather than just performing simple and repetitive tasks. RPA performs the rule based tasks and involves the use of softwares which mimics the human actions.

- **UiPath:**

UiPath is Tool of Robotic process automation which is used for the automation of desktop automation and software process automation. UiPath is used to perform automation for simple and boring repetitive tasks. which eliminates the human interference. UiPath provides the drag and drop functionality for the most of the processes and activities. There are many tools are available for robotic process automation but UiPath is simple and most popular tool easy to use among them. Following are some of the features of UiPath that makes it attractive for RPA. UiPath provides the complete solutions to the automation with help of following three components which are UiPath Robots, UiPath Orchestrator and UiPath Studio.

1)UiPath Studio:

UiPath is efficient and advanced tools which helps to design and automate the processes in visual manner with the help of diagrams.

2)UiPath Robot:

executes the processes built in Studio, as a human would. this robot can work as unattended which run without help or supervision of humans in any environment that may be virtual or not. or as assistants which takes the help of human to trigger the process.

3)UiPath Orchestrator:

a web application that enables you to deploy, schedule,

monitor and manage Robots and processes, while business exception handling is available through centralized work queues.

- **Bot:**

Bot means “Robot Player” and “Back On Topic”. with the help of RPA softwares we create the software robot and also called as "Bot", which can learn and mimic human actions and the performs and executes the rule-based business task.

In this project we have used different types of bots for different types of processes. such we have created a software robot (Bot) which will try to automatically send reply to customer emails and problems over email using UiPath SMTP email tool, and bot for verification of shipment means bot will verify that product has been delivered or not by taking the customer feedback or acknowledgement. Another bot to take reviews and feedback of product. A bot for invoice processing, data scraping and transforming into excel or csv format to perform various types of analysis for future references. A bot for profile updating and for other activities.

1.2. MOTIVATION

In past 3 years we have seen the rapid growth in the internet usage and user as well , now almost every sector is present on the internet from online purchasing to online education , almost every businesses have their presence on the internet and still it is growing day by day ,now the situation has becomes like “resources are become less than the consumers” therefor to satisfy each and every customer with limited amount of time and employee and paperwork, it has become challenging task if we use manual work or tradition approach to tackle this issue then it will became nearly impossible in future, now the question is “how can we solve this issue without transforming IT infrastructure completely ?”

The answer to this issue is “integration of Automation in current system, business or process”. And there for nowadays most of the businesses have started integrating the automation in their process. And in this project, we have integrated the automation technique in our shopping website to tackle the challenges which can occur in the future if we practice the same tradition approach, and this was the main motivation behind this project that how can we solve this issues in our shopping website using the automation using RPA.

1.3. PROBLEM STATEMENT

There are various shopping websites available in the market which practices the same tradition approach which results in more time consuming, more paper work, more employees. Therefore, sometimes it creates the issues such as late reply or late response which makes the customer dissatisfied. which decreases the customers trust over the website which leads to the loss in business. Also, in festival seasons we get to see a huge traffic because of the discounts and increase in the demands of user for the same and rare product, which creates the load on the server which becomes hard to handle with the limited number of employee and in a limited amount of time. Therefore to resolve these problems we need to integrate the automation in our process. and to handle this type of critical situation wisely and smoothly.

2. LITERATURE REVIEW

We have seen various shopping website available in the market which uses the same and old tradition methods which leads to more time consuming, more paper work, more employees. There for sometimes it creates big problem such as late reply or late response which makes the customer unhappy and dissatisfied the customer. which decreases the customers trust over the website which leads to the loss in business. Also, in festival seasons we get to see a huge traffic because of the discounts and increase in the demands of user for the occasional product, which creates the load on the server which becomes hard to handle with the limited number of employee and in limited amount of time.

To tackle this situation, we have integrated the Automation technique in our shopping website, we have tried to practice the modern approach to solve this type of issues using RPA Automation techniques.

2.1. KEY FEATURES

This project provides the following features:

- **Email Categorization:** dividing the different emails into different folders such as order, payments, problems, etc. to prioritize the important emails.
- **Email Processing:** Automatically send reply to customer email and identify the important or critical emails and assign them to the administrative team.

- **Query Processing using Chatbot:** solving the customer issues and questions using the chatbot present on website itself.
- **Shipment Validation:** verify the shipment means product has been delivered or not by taking the user acknowledgment.
- **Taking Reviews and Feedback:** taking the reviews and feedback of customers after product delivery.
- **Invoice Processing:** generating and processing the invoices.
- **User Profile Updating:** to update the customer profile and the details regularly.
- **Web Scraping:** extracting the sales and order data into csv file and in excel format to perform various types of analysis for future references.

2.2. SOFTWARE REQUIREMENT

- UiPath Studio.
- UiPath Orchestrator.
- Windows 7/ 10.
- Web Browser: Mozilla, Google Chrome, IE8, OPERA.
- Software: XAMPP server.
- .NET Framework: Version 4.6.1 or greater.
- User Interface Design: HTML,CSS,PHP, JAVASCRIPT.

2.3. HARDWARE REQUIREMENT

- CPU: 4 x 2.4GHz 64-bit (x64)
- RAM: 8 GB
- Processor: Intel i3/i5/i7 (any one) or AMD Ryzen 3/5/7.
- Space: 200 MB minimum.
- A better internet connectivity.

3. ADVANTAGES

This project provides the following advantages to our shopping websites and our customers as well:

- ❖ **Increased speed:** use of Automation in shopping website increases the performance of website and customer relationship.
- ❖ **Time savings:** as we replace most of the manual work with the automation and assign a specific bot to that work or process which results in more work in less time.
- ❖ **Simplicity:** use of automation makes the complex process simple and more time can be given to complex problems which makes the website work normally and smoothly.

- ❖ **Scalability:** number of operations can be done within less time which provides the scalability to the website.
- ❖ **Better management:** use of automation provides better management because of automation most of the workload gets reduced which leads to better management of working process.
- ❖ **Better customer service:** because of automation repetitive task can be automated which helps to give customer response in less time which leads to better customer service, as customer gets response very immediately or in less time.
- ❖ **Increased employee satisfaction:** as most of manual work gets replaced by the automation which helps employee to give more time to important and complex problem and also reduces the workload of employees.
- ❖ **Reduced costs:** because of automaton, very less employees required to handle the website and process which leads to the cost reduction.
- ❖ **Improved analytics:** most of process gets automated and more data can be generated which helps to improve analytics for growth of website.
- ❖ **Higher quality services, greater accuracy:** automation generates very less error or we can say with proper implementation no errors get generate which provides Higher quality services, greater accuracy to the website.

4. CHALLENGES

Following are some challenges of the project:

- ❖ **Financial Expenses:** one of the big reasons because of which many organizations and businesses don't opt to implement RPA is the Financial Expenses. many organizations and businesses believe that it will cost lot more than the traditional approach that's why it is the big challenge of RPA.
- ❖ **Insufficiency in Technical Ability:** Most of the organizations and businesses and people believes that in order to implement robotic process automation, the end users and customers should know the significant technical knowledge. this misunderstanding often holds and keep them back from availing the many benefits which are available to them.
- ❖ **Major Changes:** we think adopting a new technology

requires a big change but instead, by using right tool the impact of that changes are much less disruptive and less noticeable than many will realize.

- ❖ **Fear and Redundancy:** another common challenge of RPA is the fear of peoples, that robot will replace human workers entirely but instead of this its main purpose and idea was and is to actually support humans in the business and in the workplace.

5. METHODOLOGY

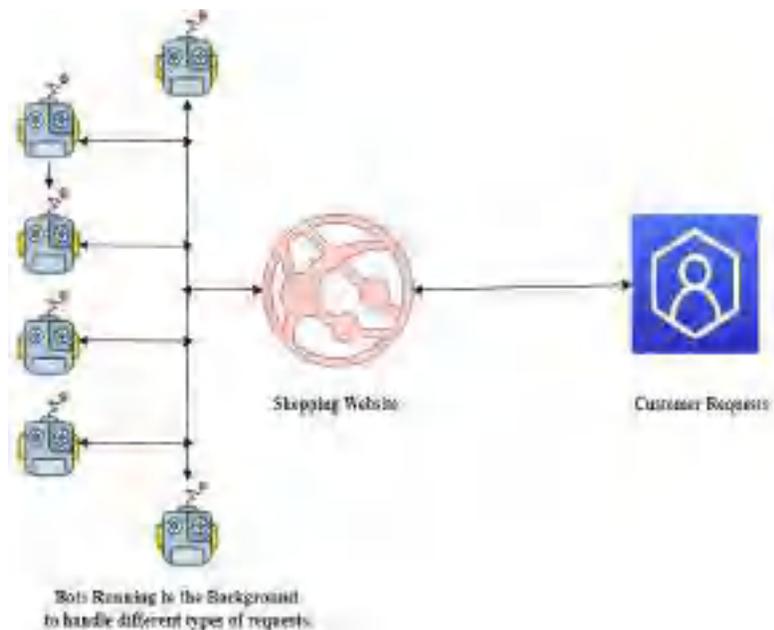


Fig.1. working of project.

In the diagram we can see the customer sends the request for specific service then it gets redirect or gets handles by the robots (bots) running in the background. For example, if customer send an email related to their order, then the bot running in the background to handle the emails, replies the customer automatically with the detail of order. Similarly, different types of Bots are assigned to handle the different types of customer request, such as a bot for chatting, another for taking feedback and shipment validation, to update profile, etc.

Following are the steps to create robots to handle different-different activities and process.

- **Email Categorization:**

In this we will create a bot which will divide the email into different types of folder such as order, delivery, payment, price, exchange, etc. Following are the steps to perform this

process:

1.First of all we have to create a sequence and then we have to add 'Get IMAP Mail Messages'. then we have to create some credentials such as email, password and mailMessages variables. Then after that we have to set our email and password for accessing our inbox. then set mailMessages type as List.

2.After that, in Properties, we have to set port no. as 903 and also server as "imap.gmail.com". then enter email and password into the login section.

3.Next step is to set the number of mails as a count which we want to filter according to their properties. Then uncheck the option of 'OnlyUnreadMessages' in order to filter emails from already read mails.

4.In Next step, we have to drag ForEach activity in sequence and iterate over mailMessages list. Then we have to add If activity to check each mail's sender email address using mail.From.Address.Contains(".....") Inside foreach. If the condition holds true, then define task to move email to respective folder. For false condition, also define a task. Where in Contains we have to pass the argument such order, price, shipment, etc., in which folder we have to move the email.

- **Email Processing:**

In this process we will create a bot which will automatically sends the email to the customer in which we will create a sequence and use SMTP protocol to send email in which we will assign the receivers email id to which we have to send email.

- **Query Processing using Chatbot:**

In this we will create a chat bot using UiPath MLPackage where in the sequence we have pass different types of argument for detecting a specific words or answer related to that word, ML package help to fetch that word and assigns the right answers to the respective question. We have to create maximum possible number of arguments and their respective solution where it detects the specific word from question and binds that word to required answer.

- **Shipment Validation:**

In this we will create a sequence in which bot will send message for verification of shipment and if answer gets in

yes then it validates the shipment as delivered successfully, else not delivered.

- **Taking Reviews and Feedback:**

In this we will create a sequence in which bot will take feedback from the customer and stores into database and displays over the website, which helps other customers to take decisions accordingly.

- **Invoice Processing:**

In this we will create a sequence where bot will create an invoice after shipment of the product.

- **User Profile Updating:**

In this process customer makes the request for making changes in the profile, then bot assigns for this task updates the details in profile. And also, bot can ask for the customer for any changes in profile regularly in order to keep customers profile UpToDate.

- **Web Scrapping:**

To extract the data from the website UiPath provides the tool such as data scrapping, where we have to create a sequence to store the extracted data and transfer into the excel or into csv format into the local database.

6. PROPOSED DESIGN

In this shopping website project, we are integrating the automation technique using RPA, but our first step is to create a shopping website.

To create shopping website, we will create user interface using the html, ajax, jQuery, and JavaScript for the background processing. And we are going to use PHP language to create a website.

After successfully creation of website, we will create different types of bots which will work in background and will replace repetitive work. Also, we have tried to replace the all-manual process by automation using the various bot.

To handle the emails, we have a different bot in which when a customer sends an email, our bot detects the type of email such as order, price, shipment, delivery, etc. after that it divides the email into the different folder. After which another bot automatically send reply to important email. If an email is very critical bot assigns the tickets to that email which helps to track the record of problem and assigns that email to the administrative team to solve the issue.

For customers questions and doubt we have a chatbot present in the

website which detect the type of question-based store arguments in the database and then fetch the correct answer related to that question and send the right answer to the customer in less time.

We have different bot to validate the shipment i.e., to verify product delivered or not and another bot to take feedback from customer related to the product in which this bot sends customer message then based on the received acknowledgment bot validates the shipment, which helps in the working process with less or no human assistance. With the help of computer many software systems were developed to solve the task that were previously done with the help of paper to manage businesses or not done at all due to the lack of tools. For example, inventory management, bookkeeping and communication management, etc. and then Bot stores the feedback of customer into the website which helps other customers to choose between the right products.

There is another bot which will use the data scraping tool of UiPath and extract the data such as product and price in order to perform the analysis of price increase and decrease. And other important analysis such as in which product customer is interested. or which type of product a respective customer buys regularly.

Another bot which handles the customer request for making changes in profile, then bot assign the new details in profile. And also, bot can ask for the customer any changes in profile regularly in order to keep customers profile UpToDate.

7. ACKNOWLEDGMENTS

Here in this project, we are implementing the Automation technique in the shopping website using RPA (Robotic process Automation) Technology, where we are trying to Automate all the possible things which we used to do manually. Such as email query processing, email categorization to prioritizing important emails, validating shipments, invoice processing, customer support email, chatting bot to solve customer problems in less time, web scrapping technique to extract order history into excel in order to perform detailed analysis for future reference, etc.

8. CONCLUSION

The main aim of this project is to replace manual work, process or task with the automation which gives the advantage from time to cost and etc.

We have tried to automate the most of the possible things and repetitive tasks which takes more time and more energy and manual interference as well.

This project will prove that we can use automation in any sector it might be either small or large.

every sector with proper implementation can use automation and automate their processes which consume more time and more people for same and repetitive process.

This project will definitely help our shopping website to operate faster and more smoothly, and will make it more responsive than others.

9. REFERENCES

1. <https://www.coursehero.com/file/31080626/Sotware-Requirements-Specification-For-Robotic-Process-Automation-For-Szabist-1doc/>
2. <https://www.uipath.com/resources/all>
3. https://en.wikipedia.org/wiki/Robotic_process_automation
4. <https://youtu.be/OJ-XuKiOoPc>
5. <https://www.uipath.com/solutions/technology>

IMPLEMENTING TECHNIQUES OF AUTOMATION (RPA) IN SHOPPING WEBSITE

Mukeshkumar Mishra

PG Student,

Pillai Hoc College of Arts, Science & Commerce,
Rasayani

Prof. Priyanka Sonawane

Asst.Professor,

Pillai Hoc College of Arts, Science & Commerce,
Rasayani

Abstract— Nowadays, every start-up, established companies or businesses are moving toward automation to tackle the repetitive, easy tasks that plague customer service representatives, which helps to deliver a better experience to customers. In this project we are using the same concept and technology of automation in shopping website using the RPA tool i.e. UiPath. It means we are going to use automation process for repetitive and time-consuming tasks which consumes more time for same process and also requires more people work. Which will result in time saving and cost reduction, so that we will be able to give more time to complex problems. So that we can provide better customer experience. Use of automation also helps in to lower operational risk by eliminating human errors such as tiredness or lack of knowledge, RPA reduces the rate of errors thereby providing a lower level of operational risk. In this paper we will see, how we can automate most of the possible operations and processes of shopping website.

Index Terms— Automation, Business, Bot, Chatbot, Robots, RPA (Robotic Process Automation), UiPath, Shopping Website.



1. INTRODUCTION

In past years we have seen many companies and businesses are shifting towards automation and found a significant growth in terms of cost and time as well, and also in customer experience and engagement. Now every sector of businesses or organizations are trying to move and adapt the technology of automation in their work field. But yet the small-scale companies and businesses lacking behind in the adoption of automation in their process now might be there are some reasons such as “trust issue” or “lack of implementation”, still they believe or rely on the manual work. In this project we are going to use RPA automation techniques in our shopping website that proves, we can implement the automation in any sector it might be either small or large, every sector with proper implementation can use automation and automate their processes which consumes more time and more people for same and repetitive process. In this project we are using the automation in most of the possible process such as auto reply to customer email and solve the customers queries on the website itself using the chatbot, and the repetitive processes such as invoice processing, shipment validation, customer profile updating and etc.

where we will use the famous and most used RPA tool UiPath. Where we will create and assign software robots also called as Bots using various UiPath tools which will mimic the human action and will try to perform respective process automatically.

1.1. TERMINOLOGY

Before getting into the project let's explore the basic concepts and terminologies which we are going to use in this project or over which we are going to work.

- **Automation:**

The term automation is derived from the Greek word's autos meaning self, and motos, meaning moving, in simple words, is technology that deals with the application of computers and machines for the maintenance and production of goods and services. Which helps to get work done without human help or assistance and if any then less, compare to traditional systems.

- **RPA:**

Robotic process automation (RPA) means the process where basic and complex tasks are gets automated through the

software or hardware system which function with many and number of applications, like human workers used to do. RPA software is designed to perform repetitive and simple task and reduce the burden for humans. Which helps employees to give more time to complex problem rather than just performing simple and repetitive tasks. RPA performs the rule based tasks and involves the use of softwares which mimics the human actions.

- **UiPath:**

UiPath is Tool of Robotic process automation which is used for the automation of desktop automation and software process automation. UiPath is used to perform automation for simple and boring repetitive tasks.

- **Bot:**

Bot means “Robot Player” and “Back On Topic”. with the help of RPA softwares we will create the software robot and also called as "Bot", which can learn and mimic human actions and performs and executes the rule-based business task.

In this project we have used different types of bots for different types of processes. such as we have created a software robot (Bot) which will try to automatically send reply to customer emails and problems over email using UiPath SMTP email tool, and bot for verification of shipment means bot will verify that product has been delivered or not by taking the customer feedback or acknowledgement. Another bot to take reviews and feedback of product. A bot for invoice processing, data scraping and transforming into excel or csv format to perform various types of analysis for future references. A bot for profile updating and for other activities.

2. OBJECTIVE

In past 3 years we have seen the rapid growth in the internet usage and user as well , now almost every sector is present on the internet from online purchasing to online education , almost every businesses have their presence on the internet and still it is growing day by day ,now the situation has becomes like “resources are become less than the consumers” therefor to satisfy each and every customer with limited amount of time and employee and paperwork, it has become challenging task if we use manual work or tradition approach to tackle this issue then it will became nearly impossible in future, now the question is “how can we solve this issue without transforming IT

infrastructure completely?”

The answer to this issue is “integration of Automation in current system, business or process”. And therefore, nowadays most of the businesses have started integrating the automation in their process. And in this project, we have integrated the automation technique in our shopping website to tackle the challenges which can occur in the future if we practice the same tradition approach, and this was the main motivation behind this project that how can we solve this issues in our shopping website using the automation using RPA.

There are various shopping websites available in the market which practices the same tradition approach which results in more time consuming, more paper work, more employees. Therefore, sometimes it creates the issues such as late reply or late response which makes the customer dissatisfy. which decreases the customers trust over the website which leads to the loss in business. Also, in festival seasons we get to see a huge traffic because of the discounts and increase in the demands of user for the same and rare product, which creates the load on the server which becomes hard to handle with the limited number of employee and in a limited amount of time.

Therefor to resolve these problems we need to integrate the automation in our process. and to handle this type of critical situation wisely and smoothly.

3. TECHNOLOGY AND SOFTWARE SPECIFICATION

To Implement this idea, we first need a shopping website, and to create a shopping website we are going to use HTML and CSS as frontside processing and to make the interface user friendly and easy to use and operate, and for background processing we are using the JavaScript.

Then we integrate this website with the UiPath in order to use automation and have a different bots for different activities.

Now we have used many websites thus we know the basic idea and concept of websites and its elements. But what’s new in this project is the integration of UiPath which is the idea behind this project, so let’s first understand what is UiPath and what are the elements and features of UiPath which we are going to use in this project.

- **UiPath:** UiPath is a Robotic process automation tool used for Windows desktop automation. It is founded in the year 2005 by the Romanian entrepreneurs Daniel Dines and Marius Tirca. It is used to automate boring repetitive tasks. It also eliminates human

intervention. With drag and drop functionality for all activities, it is the simplest RPA tool. There are lots of tools for robotic process automation, however UiPath is the most popular among them. Following are some of the features of UiPath that makes it attractive for RPA. UiPath offers a complete solution with the help of three components namely UiPath Studio, UiPath Orchestrator and UiPath Robots.

- 1) **UiPath Studio** - an advanced tool that enables you to design automation processes in a visual manner, through diagrams.
- 2) **UiPath Robot** - executes the processes built in Studio, as a human would. Robots can work unattended (run without human supervision in any environment, be it virtual or not) or as assistants (a human triggers the process).
- 3) **UiPath Orchestrator** - a web application that enables you to deploy, schedule, monitor and manage Robots and processes, while business exception handling is available through centralized work queues.

Now to summarize the specification following are the software and hardware requirement for this project

3.1. SOFTWARE REQUIREMENT

- UiPath Studio.
- UiPath Orchestrator.
- Windows 7/ 10.
- Web Browser: Mozilla, Google Chrome, IE8, OPERA.
- Software: XAMPP server.
- .NET Framework: Version 4.6.1 or greater.
- User Interface Design: HTML,CSS,PHP, JAVASCRIPT.

3.2. HARDWARE REQUIREMENT

- CPU: 4 x 2.4GHz 64-bit (x64)
- RAM: 8 GB
- Processor: Intel i3/i5/i7 (any one) or AMD Ryzen 3/5/7.
- Space: 200 MB minimum.
- A better internet connectivity.

4. IMPLEMENTATION

In the following flowchart diagram, we can see how the customer send the request for specific service then it gets redirects or gets handles by the robots (bots) running in the background.

For example, if customer send an email related to their order, then the bot running in the background to handle the emails, replies the customer automatically with the detail of order. Similarly, different types of are assign to handle the different types of customer request, such as a bot for chatting, another for taking feedback and shipment validation, to update profile, etc.

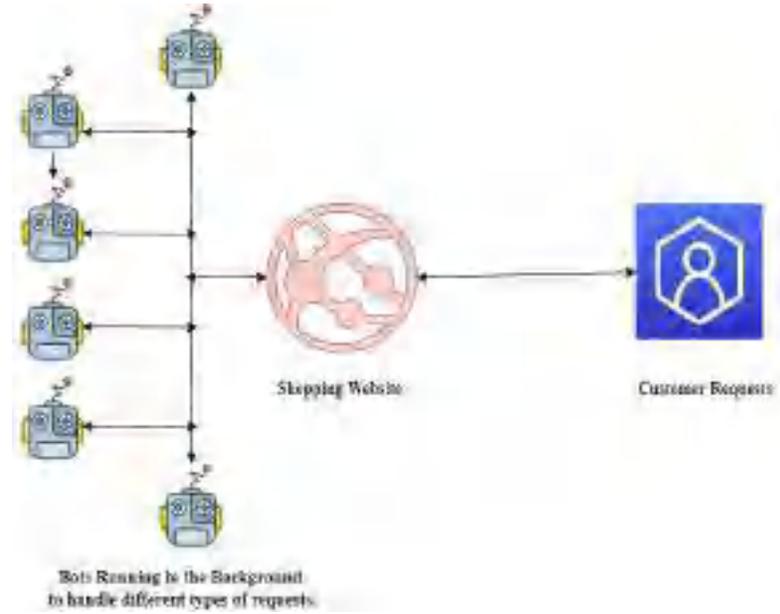


Fig.1. Working of Project.

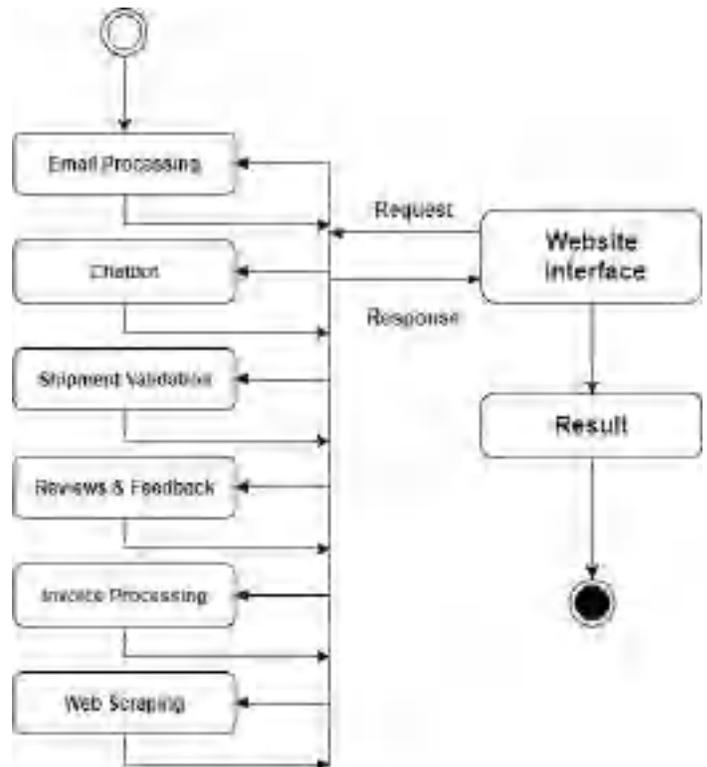


Fig.2. Activity Diagram of working processes.

above Activity diagram shows that how and when different activities are gets performed and when user makes the request. and Diagram shows how activities are perform from one operation to another and from start to end.

Following diagram shows the sequence of process such as when a specific action gets performed by user then, how it gets directs to UiPath and how specific bot gets assigned, and how bot takes the proper action and gives the response back to the user via website.

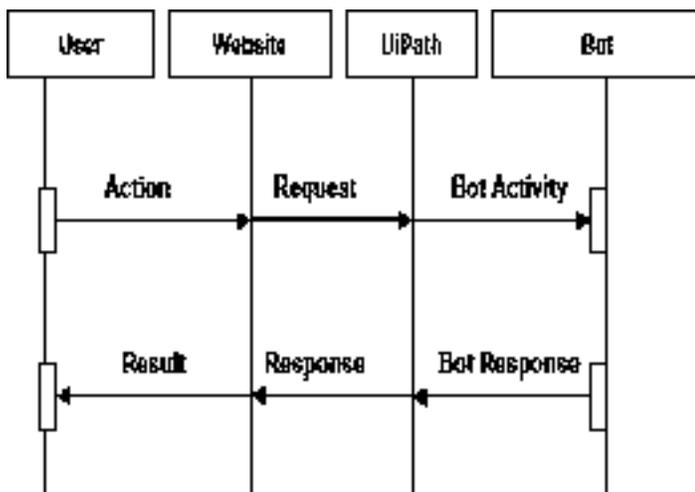


Fig.3. Sequence Diagram of working processes.

Now to implement the project and to use UiPath robots for different-different activities we have to create a bot for different-different activities.

Now, following are the steps to create robots to handle different-different activities and process.

- **Email Categorization:**

In this we will create a bot which will divide the email into different types of folder such as order, delivery, payment, price, exchange, etc. Following are the steps to perform this process:

1.First of all we have to create a sequence and then we have to add 'Get IMAP Mail Messages'. then we have to create some credentials such as email, password and mailMessages variables. Then after that we have to set our email and password for accessing our inbox. then set mailMessages type as List.

2.After that, in Properties, we have to set port no. as 903 and also server as "imap.gmail.com". then enter email and

password into the login section.

3.Next step is to set the number of mails as a count which we want to filter according to their properties. Then uncheck the option of 'OnlyUnreadMessages' in order to filter emails from already read mails.

4.In Next step, we have to drag ForEach activity in sequence and iterate over mailMessages list. Then we have to add If activity to check each mail's sender email address using mail.From.Address.Contains(".....") Inside foreach. If the condition holds true, then define task to move email to respective folder. For false condition, also define a task. Where in Contains we have to pass the argument such order, price, shipment, etc., in which folder we have to move the email.

- **Email Processing:**

In this process we bot will automatically sends the email to the customer in which we will create a sequence and use SMTP protocol to send email in which we will assign the receivers email id to which we have to send email.

- **Query Processing using Chatbot:**

In this we will create a chat bot using UiPath MLPackage where in the sequence we have pass different types of argument for detecting a specific words or answer related to that word, ML package help to fetch that word and assigns the right answers to the respective question. We have to crate maximum possible number of arguments and their respective solution where it detects the specific word from question and binds that word to required answer.

- **Shipment Validation:**

In this we will create a sequence in which bot will send message for verification of shipment and if answer gets in yes then it validates the shipment as delivered successfully, else not delivered.

- **Taking Reviews and Feedback:**

In this we will create a sequence in which bot will take feedback from the customer and stores into database and displays over the website, which helps other customers to take decisions accordingly.

- **Invoice Processing:**

In this we will create a sequence where bot will create an invoice after shipment of the product.

- **User Profile Updating:**

In this process customer makes the request for making changes in the profile, then bot assigns for this task updates the details in profile. And also, bot can ask for the customer for any changes in profile regularly in order to keep customers profile UpToDate.

- **Web Scrapping:**

To extract the data from the website UiPath provides the tool such as data scrapping, where we have to create a sequence to store the extracted data and transfer into the excel or into csv format into the local database.

5. OUTCOMES

In this shopping website project, we are integrating the automation technique using RPA, but our first step is to create a shopping website. To create shopping website have used following steps, first we have created user interface using the html, CSS and JavaScript for the background processing. After successfully creation of website, we have made different types of bots which will work in background and will replace repetitive work. Also, we have tried to replace the all-manual process by automation using the various bot.

So similarly, this project gives the following Outcomes and provides the following Key features:

- **Email Categorization:**

Working of this Bot is, dividing the different emails into different folders such as order, payments, problems, etc. to prioritize the important emails. As shown in following diagram.

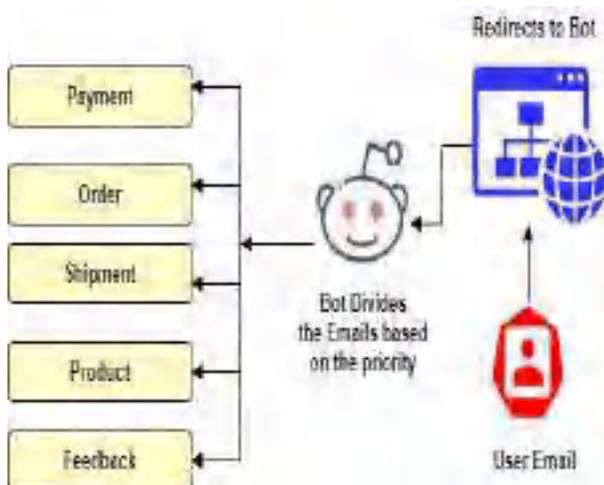


Fig.4. Email Categorization.

- **Email Processing:**

This Bot Automatically send reply to customer email and identify the important or critical emails and assign them to the administrative team. As shown in following diagram.

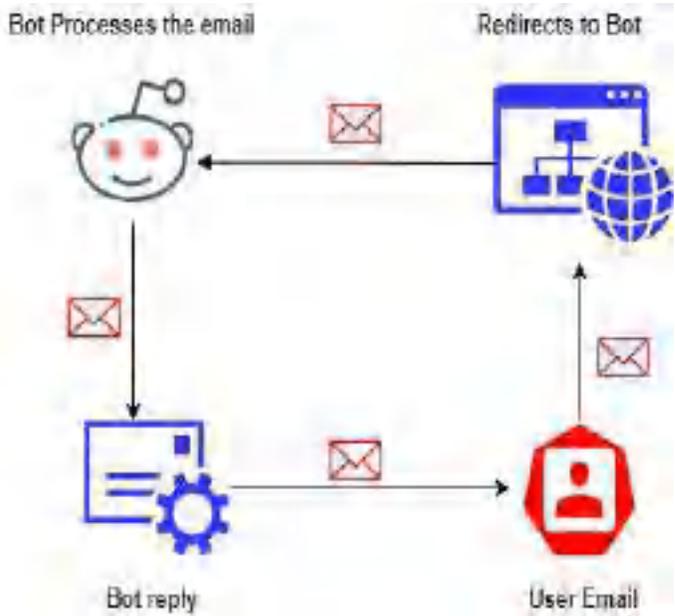


Fig.5. Email Processing.

- **Query Processing using Chatbot:**

solving the customer issues and questions using the chatbot present on website itself. As shown in following diagram.

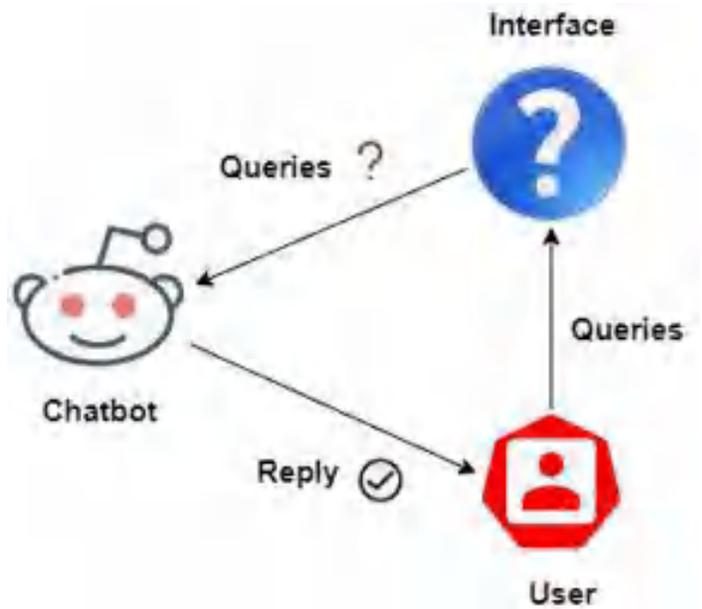


Fig.6. Query Processing using Chatbot.

• Shipment Validation:

In this process Bot verifies the shipment means product has been delivered or not by taking the user acknowledgment. As shown in following diagram.

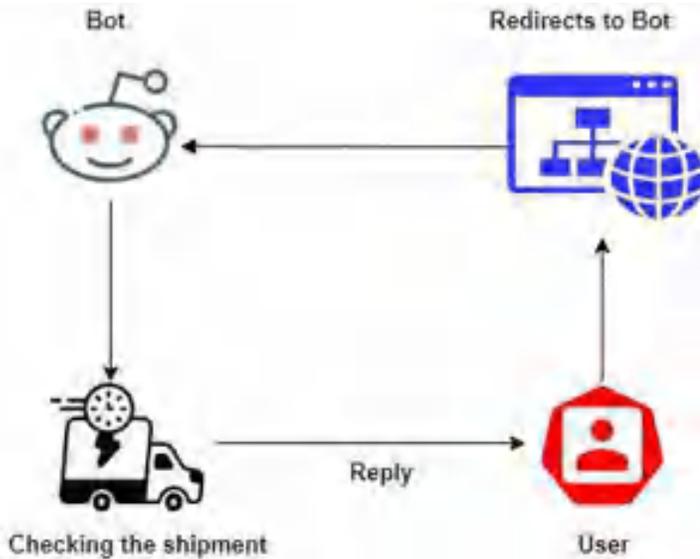


Fig.7. Shipment Validation.

• Taking Reviews and Feedback:

In this process Bot takes the reviews and feedback of customers after product delivery. As shown in following diagram.

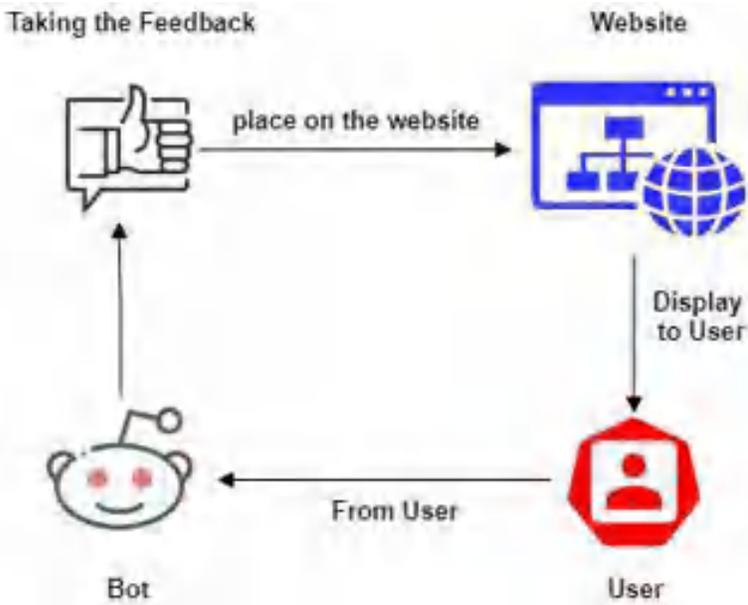


Fig.8. Taking Reviews and Feedback.

• Invoice Processing:

In this process task of Bot is to generate and process the invoices.

As shown in following diagram.

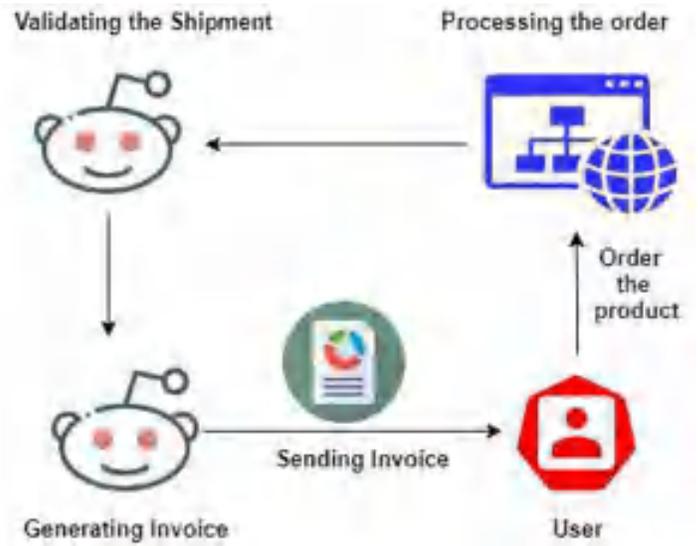


Fig.9. Invoice Processing.

• User Profile Updating:

In this process Bot works to update the customer profile and the details regularly. As shown in following diagram.

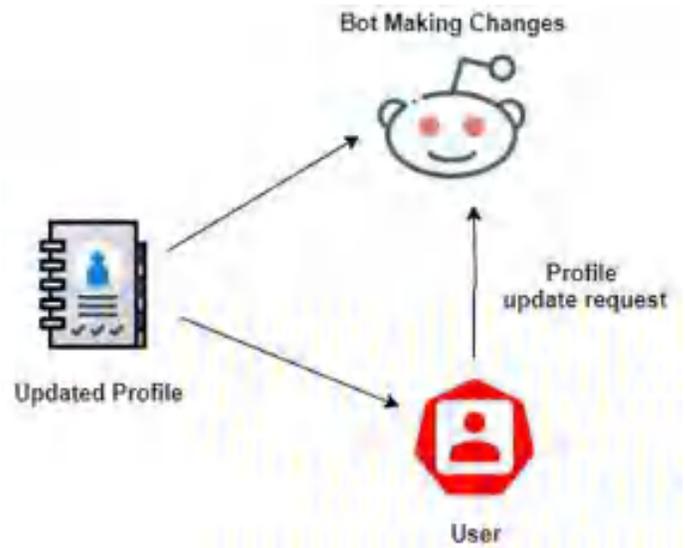


Fig.10. User Profile Updating.

• Web Scraping:

In this process, task of this Bot is extracting the sales and order data into csv file and in excel format to perform various types of analysis

for future references. As shown in following diagram.

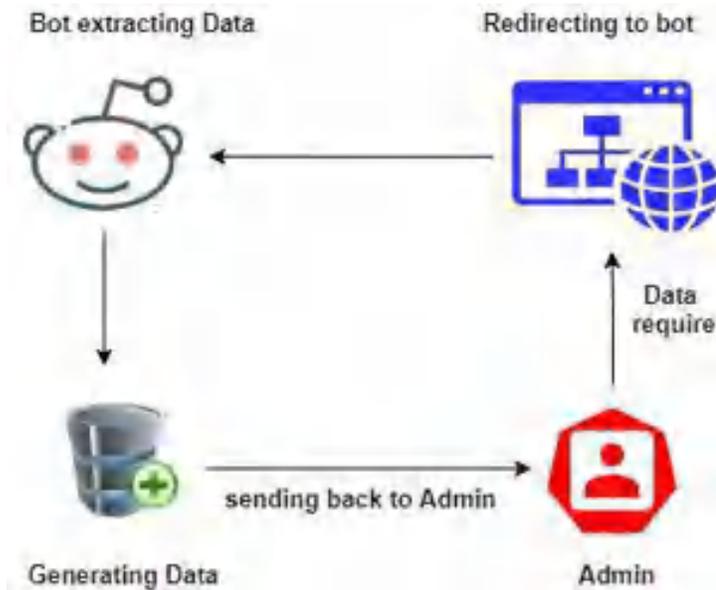


Fig.11. Web Scraping.

6. ADVANTAGES

This project provides the following advantages to our shopping websites and our customers as well:

- ❖ **Increased speed:** use of Automation in shopping website increases the performance of website and customer relationship.
- ❖ **Time savings:** as we replace most of the manual work with the automation and assign a specific bot to that work or process which results in more work in less time.
- ❖ **Simplicity:** use of automation makes the complex process simple and more time can be given to complex problems which makes the website work normally and smoothly.
- ❖ **Scalability:** number of operations can be done within less time which provides the scalability to the website.
- ❖ **Better management:** use of automation provides better management because of automation most of the workload gets reduced which leads to better management of working process.
- ❖ **Better customer service:** because of automation repetitive task can be automated which helps to give customer response in less time which leads to better customer service,

as customer gets response very immediately or in less time.

- ❖ **Increased employee satisfaction:** as most of manual work gets replaced by the automation which helps employee to give more time to important and complex problem and also reduces the workload of employees.
- ❖ **Reduced costs:** because of automaton, very less employees required to handle the website and process which leads to the cost reduction.
- ❖ **Improved analytics:** most of process gets automated and more data can be generated which helps to improve analytics for growth of website.
- ❖ **Higher quality services, greater accuracy:** automation generates very less error or we can say with proper implementation no errors get generate which provides Higher quality services, greater accuracy to the website.

7. CHALLENGES

This project also has some challenges as well, as we know many times while creating and implementing new idea and its working project which is not yet done by others, we face many challenges, Now Following are some challenges we have faced while implementing this project:

- **Financial Expenses:** one of the big reasons because of which many organizations and businesses don't opt to implement RPA is the Financial Expenses. many organizations and businesses believe that it will cost lot more than the traditional approach that's why it is the big challenge of RPA.
- **Insufficiency in Technical Ability:** Most of the organizations and businesses and people believes that in order to implement robotic process automation, the end users and customers should know the significant technical knowledge. this misunderstanding often holds and keep them back from availing the many benefits which are available to them.
- **Major Changes:** we think adopting a new technology requires a big change but instead, by using right tool the impact of that changes are much less disruptive and less noticeable than many will realize.
- **Fear and Redundancy:** another common challenge of RPA is the fear of people that robot will replace human workers entirely but instead of this its main purpose and idea was and is to actually

support humans in the business and in the workplace.

- **Inability to automate end-to-end processes:** For the more complex processes, RPA tools may be insufficient for directly automating all the process steps. “Divide and conquer” is way to go about this. Redesigning these sophisticated tasks, breaking them into simpler parts, and starting automation. Additionally, with the joint work of RPA and other digital technologies like machine learning or optical character recognition. Keeping in mind though the extra costs involved by this, so we cannot strive for end-to-end intelligent automation when cost-efficiency becomes questionable.
- **Insufficient assistance from the business department:** Relying solely on the IT department is among the common RPA challenges that should be actively avoided throughout the automation project. According to RPA expert Nicole Schultz, “finance cannot depend on IT for RPA; it needs to be owned by the business side.” Business processes require a Process Design Document for the pilot phase, including workflow diagrams, data-specific business rules (for various types of data), a comprehensive list of technical exceptions that the operations unit may face during manual processing, etc. It is more likely that the pilot paves the way for successful long-term development if the business team gives feedback for bots’ performance.

8. FUTURE SCOPE

In future in this project, we will try to replace all the task by automation.

following are the some of the ideas on which we would like to work in the future.

- **Stock details and verification:** A Bot which will Automatically updates the stock details and request for new stock.
- **Product suggestion:** A Bot which will Automatically provide the better suggestion to customer related to the product they like very much, or they buy very often.
- **Short Stories:** Provide story related to interested product of respective customers within the same website. with the help of bot will try to create short stories related to the products.
- **Complete Automation:** Replace all small task by different

types of bot.

- **Automatic maintenance:** A Bot which will Automatically keep the maintenance of website using automation.

9. CONCLUSION

The main aim of this project is to replace manual work, process or task with the automation which gives the advantage from time to cost and etc.

We have tried to automate most of the possible things and repetitive tasks which takes more time and more energy and manual interference as well.

This project will prove, that we can use automation in any sector it might be either small or large.

every sector with proper implementation can use automation and automate their processes which consume more time and more people for same and repetitive process.

This project will definitely help our shopping website to operate faster and more smoothly and will make it more responsive than others.

10. REFERENCES

1. <https://www.coursehero.com/file/31080626/Sotware-Requirements-Specification-For-Robotic-Process-Automation-For-Szabist-1doc/>
2. <https://www.uipath.com/resources/all>
3. https://en.wikipedia.org/wiki/Robotic_process_automation
4. <https://youtu.be/OJ-XuKiOoPc>
5. <https://www.uipath.com/solutions/technology>

Thyroid Disease Prediction Using Data Mining Techniques

Namrata Kedari, Prof Priyanka Sonawane

Abstract— Classification based data processing plays important role in various healthcare services. In healthcare field ,the important and challenging task is to diagnose health conditions and proper treatment of disease at the first stage. There are various diseases which will be diagnosed early and can be treated at the first stage. As as an example, Thyroid diseases. the normal ways of diagnosing thyroid diseases depends on clinical examination and plenty of blood tests.The Main task is to detect disease diagnosis at the first stages with higher accuracy.Data mining techniques plays an important role in healthcare field for creating decision,disease diagnosis and providing better treatment for the patients at low cost. Thyroid disease Classification is an important task.The purpose of this study is predication of thyroid disease using different classification techniques and also to seek out the TSH, T3,T4 correlation towards hyperthyroidism and hypothyroidism and also to finding the TSH, T3,T4 correlation with gender towards hyperthyroidism and hypothyroidism.

Index Terms Data mining techniques,classification model, thyroid diseases,Decision tree,knn,svm,Naive Bayes,ID3 classification

1 INTRODUCTION

Diagnosis of Thyroid Disease is incredibly tedious and difficult tasks. The diagnosis thyroid disease within the traditional way includes clinical examination and also the many blood tests. But then the most task is to diagnosis the disease at early stages with high accurate percentage. In medical field ,Data mining plays an important role for diagnosis of disease. data processing provides many classification techniques for the prediction of disease accuracy. The gathered patient data collected from many health care organization is helpful for the chance factors analysis for many diseases. In healthcare and life science ,the applications based on data processing are very beneficial and important. The large amount of knowledge gathered from health care organization has no organizational value unless transformed into most useful information and knowledge,which might be helpful in cost controlling, increasing the profits, and top quality maintainance of patient healthcare.Classification Algorithms is one in every of the most important applications within the field of knowledge mining, which can be usefull for deciding in many real world problems

2 OVERVIEW OF THYROID

The thyroid is an endocrine gland.The function of thyroid gland is to produce of thyroid hormones.It

reaches to all other organs through the bloodstream and control metabolism and growth development.The important functions of thyroid gland includes respiration,blood circulation,gut movements,temperature control,muscle functioning ,digestion and functioning of brain.Any dys function in the thyroid gland may affects the normal physiological functioning human body.The thyroid hormone affects the growth and development depending on the amount of secretion.When the production of thyroid hormone is very less and is referred as hypo-thyroidism.When the production of thyroid hormone is very high then this type of thyroid disease is referred as hyperthyroidism.

A. Thyroid Harmones

The ductless gland produces are tri-iodothyronine (T3)and Lthyroxine (T4) and .The thyroid hormones regulates various metabolic activities like generation of warmth, the consumption of carbohydrates, protein and fats. The pituitary gland controls production of tri- iodothyronine and L-thyroxine hormones.The Thyrotropin-Stimulating Hormone from pituitary gland is released when hormone is required and circulates through the bloodstream to succeed in thyroid. TSH then stimulates the thyroid glands for the

assembly of T4 and T3 hormones .The production of internal secretion are controlled by the feedback system of pituitary gland.The TSH production is a smaller amount when T3,T4 are more within the circulation and TSH production is more when T3,T4 are less

B. Thyroid and its Health Effects

Thyroid disorder are most common endocrine disease,across the worldwide. In an Indian survey 42 million people are suffering from this diseases.The Thyroid diseases is different from others in terms of their the relative visibility, predication ease, medical treatment modeaccessibility .The inappropriate production of thyroid hormone affects health conditions.

1) Hyperthyroidism

Increase production in the thyroid hormones causes hyperthyroidism.Graves' disease is one of the autoimmune disorder that causes hyperthyroidism.The symptoms[are dry skin , increase sensitivity to temperature,thinning of hair,weight loss, increase heart rate, high blood pressure, excess sweating, neck enlargement, nervousness, menstrual periods shorten, frequent gut movements and hands trembling .

2) Hypothyroidism

Decrease production in the thyroid hormones causes Hypothyroidism.The medical term hypo means deficient or less. The causes for hypothyroidism are inflammation and thyroid gland damage.The Symptoms includes obesity, low heart rate, increase in cold sensitiveness, neck swelling, dry skin , hands numbness,hair problem, heavy menstrual periods and digestive problems.And these Symptoms may worsen over period if not treated .

3 LITERATURE SURVEY

Deepika Koundal et al.[6] have studied the existing the earlier automatic tools for diagnosis of disease at the easier stage in an efficient way.Also the metrics study about the different evaluation of performance and also investigations on the trends and future developments are studied.

Nikita Sigh and Alka Jindal [7] have compared Support Vector Machine with K –Nearest Neighbor and Bayesian and concluded Support Vector Machine better then KNN and Bayesian with an accuracy about 84.62%.KNN found the nearest neighborhood automatically.The results is represented by graph with object as each vertices.The probability classification is done using Bayesian which indicates the sample data belongs to a class.

Edgar Gabriel et al.[8] have proposed a texture-based segmentation i,e two parallel versions of a code for Fine Needle Aspiration Cytology thyroid images is the most important first step in identifying a fully automated Computer Aided Design solution.The code is developed in MPI version to exploit computer resources such as PC clusters.

Preeti Aggawal et al.[9] listed the method for an automatic segmentation.The study shows the summary obtained by applying specific algorithm(automatic) segmentation and automatic tools on both thyroid US as well on lung CT [7]. For segmentation of thyroid US images they have used Analyze 10.0 and Mazda .

Eystraints G[10]have provided system TND(Thyroid Nodule Detector) using a technique called computeraided diagnosis(CAD).During thyroid UltraSound examinations ,a nodular tissue detection is used in ultrasound(US) and thyroid images videos acquired

Alfonso Bastias et al.[4]have aimed at developing an machine learning classifier using AIS for diagnosis of health condition and of the proposed classifier for capability investigation.The proposed classifier successfully improved the thyroid gland disease identification process .

Ali keles et al. [7] proposed an expert system for predicting of thyroid that is known as Expert System for Thyroid Disease Diagnosis(ESTDD).This expert system diagnose thyroid diseases through neuro fuzzy rules with 95.33% of accuracy.

4 DATASET DESCRIPTION

Dataset is taken from UCI machine learning repository Database consists of patients thyroid records. Each thyroid patients record is consists of 15 attributes listes below.

Attribute can be Boolean or continuous valued are in given below table I

SN	Attribute Name	Value Type
1	Age	continuous
2	Sex	m,f
3	On_thyroxine	f,t
4	Query_on_thyroxine	f,t
5	Thyroid_surgery	f,t
6	Query_hypothyroid	f,t
7	Query_hyperthyroid	f,t
8	Pregnant	f,t
9	Goitre	f,t
10	TSH value	continuous
11	T3 value	continuous
12	TT4 value	continuous
13	T4U value	continuous
14	FTT value	continuous
15	TBG value	continuous

5 CLASSIFICATION TECHNIQUES IN DATA MINING

a) Decision Tree

A decision tree has 3 styles of node like internal node that represents test attribute, the classes or class attribute are denoted by the leaf node, the highest most is denoted by the basis node of. the advantages of using decision tree is to spot and eliminate the redundant data called "tree pruning" to improve the accuracy of the classification. the decisions are made on attribute with the very best normalized data also it can applied to both continuous and discrete values. on the opposite hand the disadvantages includes

b) Backpropagation Neural Network

Backpropagation could even be a neural network algorithm. It consists of three different layers, input layer - the inputs are given here, hidden layer - the input to hidden layer can the outputs with weights [5], number of hidden layer's arbitrary, output layer - the input to the output layer is from hidden layers, which eliminates prediction of the network's. Thus the benefits includes high accuracy, Very flexible for noisy and when the info is inconsistency, easy update of weights. The Disadvantages of Backpropagation Neural Network are representation of knowledge, it is difficult for humans to interpret, correctness. Decreases the accuracy of the network by the removal of weighted links. Selection of coaching dataset is difficult.

c) Support Vector Machine

The hyper plane separates positive training sample with the uncooperative training data sample in an plan. The advantages includes a simple extend, used for pattern reorganization, quadratic optimization problem can be formulated. The another disadvantages are suitable just for real valued space. It allow only 2 classes for classification using binary method and several other strategies for multiple class classification. For user its very hard to understand Hyper plane. One

of the kind of learning system algorithm is Support Vector Machine[8], which is employed to perform classification in an exceedingly better accurately and uses 2 class classifier, referred as hyper plane as “decision boundary or decision surface”.

d) Density-based clustering

In Density based clustering algorithm, an unassigned object is chosen from the given data set classification method like Hierarchical multiple classifier is employed classify the given dataset. Thus it’s an efficient thanks to classify an data with accurate information in reduced time and cost[13]The density based clustering algorithm falls under data clustering algorithm: an area is taken into account with given set of points its grouping together points that are closely packed together i.e., points with many closely neighbors. The most common clustering algorithms and also most cited for scientific literature is density based algorithm. It is opposite to k-means, using an R*tree.

6. PROPOSED WORK

In healthcare services data mining technique is mainly used for making conclusion, disease diagnosing and giving better treatment to the patients at comparately low cost. Classification of thyroid disease plays is an main task in the prediction of disease. capacity reduction may be done as a future work so that number of blood test the thyroid will be reduced and also time required to diagnose disease.

The thyroid Dataset is taken from UCI data depository site. In the Database involve of thyroid patient records. The Patients record is having different attributes described in the data set

description and different data mining techniques are applied to get the predication of thyroid disease and then Linear regression is performed to obtain the which hormone among

TSH, T3, TT4 affect the male and female. And also which among the TSH, T3, TT4 influence the hypothyroidism and hyperthyroidism

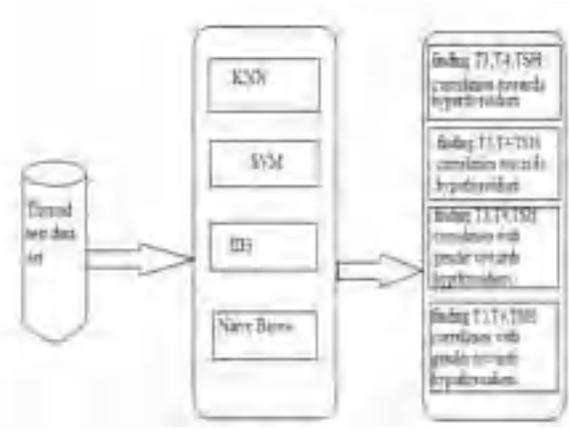


Fig.1 Proposed Classification model for thyroid disease prediction

7 CONCLUSION

Thus this survey is needfull to spot a way to predict the thyroid disorder at earlier stage using data processing techniques. Data Mining classification algorithms are used to diagonise the thyroid problems and provides different level of accuracy for every techniques. These techniques help to minimize the noisy data of the patient’s data from the information bases. Data mining Algorithms like KNN, Naïve bayes, Support vector machine, ID3 are considered for the study. These various algorithm results are supported speed, accuracy and performance of the model and value for the treatment. Also these classification of effective data are helps to search out the treatment to the thyroid patients with better cost and facilitates the management.

8 REFERANCE

[1] Roshan Banu D, and K.C.Sharmili “A Study of Data Mining Techniques to Detect Thyroid Disease” In-

ternational Journal of Innovative Research in Science, Engineering and Technology (Vol. 6, Special Issue 11, September 2017

[2]Irina IoniÑă and Liviu IoniÑă” Prediction of Thyroid Disease Using Data Mining Techniques”The Classification Technique for Talent Management using SVM, (ICCEET), 978-1- 4673-0210-4/12, pp. 959-963, 2017

[3] Khushboo Taneja, Parveen Sehgal, Prerana “Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network” International Journal of Research in Management, Science & Technology (E-ISSN: 2321- 3264) Vol. 3, No. 2, April 2016

[4] Hanung Adi Nugroho, Noor Akhmad Setiawan, Md. Dendi Maysanjaya,” A Comparison of Classification Methods on Diagnosis of Thyroid Diseases”IEEE International Seminar on Intelligent Technology and Its Applications,2017

[5] M.N Das ,Brojo Kishore Mishra ,Shreela Dash,” Implementation of an Optimized Classification Model for Prediction of Hypothyroid Disease Risks” 2017 IEEE Conference on Big Data and Analytics (ICBDA)

[6] Rasitha Banu,G.— “Predicting Thyroid Disease using Linear Discriminant Analysis (LDA) Data MiningTechnique”, IJTRA

[7] Izdihar Al-muwaffaq and Zeki Bozkus” MLTDD : USE OF MACHINE LEARNING TECHNIQUES FOR DIAGNOSIS OF THYROID GLAND DISORDER” Proceeding of Annual International Conference of the IEEE Engineering in Medicine and Biology Society,2017

[8] [8]Umadevi S , Dr .JeenMarseline K.S,”Applying Classification Algorithms to Predict Thyroid Disease” International Journal of Innovative Research in Science, Engineering and Technology (Vol. 6, Special Issue 10, September 2017

[9] Dr. Srinivasan B , Pavya K “ Diagnosis of Thyroid Disease: A Study” International Research Journal of Engineering and Technology Volume: 03 Issue: 11 | Nov - 2016

[10] K. Rajam and R. Jemina Priyadarsini” A Survey on Diagnosis of Thyroid Disease Using Data Mining Techniques” ,IJCSMC, Vol. 5, Issue. 5, May 2016, pg.354– 358.

[11] Rouhani Modjtaba and Mansouri Kamran “Multi-classification approach for detecting Thyroid attacks”, International Journals Pharm Bio Sci, pp. - 4(3),July 2017.

[12] Gurmeet Kaur, Er.Brahmaleen Kaur Sidhu, — Comparison of anfis neural network with several other anns and svm for diagnosing hepatitis and thyroid diseases ,IRJET Volume1, Issue 11, July-

[13] Saeed Shariati and Mahdi Motavalli Haghighi, Proposing Neural Network Efficient Training Model for Thyroid Disease Diagnosis In Proceedings of IEEE IACSIT 2017, pages 526-528.

[14] Aytürk Keleş and Keleş, Ali. "ESTDD: Expert system for thyroid diseases diagnosis." International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 11 | Nov -2017 34.1 (2017): 242- 246

[15]<http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid>

Heart Disease Prediction System

Namrata Kedari, Prof. Priyanka Sonawane,

Abstract— The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease. In this paper, we develop a heart disease predict system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal. Secondly, we develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a user-friendly heart disease predict system (HDPS). The HDPS system will be consisted of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result). Our approaches are effective in predicting the heart disease of a patient. The HDPS system developed in this study is a novel approach that can be used in the classification of heart disease.

Index Terms- Logistic regression, Native Bayes, Random forest, K nearest neighbor, Decision tree, Support vector classifier, Neural network, Extreme gradient boost

1 INTRODUCTION

Heart is one of the most important part of our body. It affects a lot of other biochemical reactions in our body. Diseases affecting heart include coronary artery disease, cardiovascular failure, arrhythmias and myocardial infarction. Cardiovascular diseases are basically heart related diseases that may also involve narrowed or blocked blood vessels that may lead to myocardial infarction, angina or stroke. Other conditions may affect heart's valves – auricles and ventricles. Early detection of heart diseases can prevent a lot of problems. Due to lack of knowledge in the medical field, it is imperative to diagnose the condition at an earlier stage.

Heart disease is one of the biggest causes of death among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of clinical data analysis. The amount of data in the healthcare industry is huge. Data mining turns the large collection of raw healthcare data into information that can help to make informed decisions and predictions. In this research, we discuss Machine Learning approaches (and eventually comparing them) for classifying whether a person is suffering from heart disease or not, using one of the most used dataset — Cleveland Heart Disease Dataset (CHDD) from the UCI Repository. It consists of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (precision, accuracy, recall, f1-score and support). We have used the models — Logistic Regression, Naïve Bayes, Random Forest, Extreme Gradient Boost, K-Nearest Neighbour, Decision Tree, Support Vector Classifier and Neural Network.

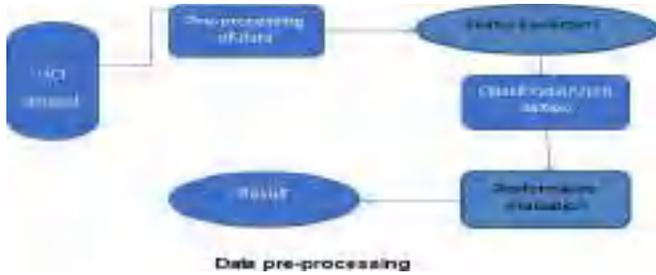
Inference:

This paper used two neuron model to classify the UCI dataset of heart disease data and managed to get an accuracy of 80%

2 Methodology

2.1 Data pre-processing

Heart disease data is pre-processed after collection of various records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing. The multi class variable and binary classification are introduced for the attributes of the given dataset. The multi-class variable is used to check the presence or absence of heart disease. In the instance of the patient having heart disease, the value is set to 1, else the value is set to 0 indicating the absence of heart disease in the patient. The pre-processing of data is carried out by converting medical records into diagnosis values. The results of data pre-processing for 303 patient records indicate that the records of both suffering from heart disease and not suffering are same.



Information:

Only 14 attributes used:

1. #3 (age)
2. #4 (sex)
3. #9 (cp)
4. #10 (trestbps)
5. #12 (chol)
6. #16 (fbs)
7. #19 (restecg)
8. #32 (thalach)
9. #38 (exang)
10. #40 (oldpeak)
11. #41 (slope)
12. #44 (ca)
13. #51 (thal)
14. #58 (num) (the predicted attribute)

3 Feature selection

From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease. As previously mentioned in this experiment, several (ML) techniques are used namely Logistic Regression, Naïve Bayes, Random Forest, Extreme Gradient Boost, K-Nearest Neighbour, Decision Tree, Support Vector Classifier and Neural Network.

4 Model selection and training

A) Logistic Regression

Logistic regression is a type of regression analysis in statistics that uses a set of predictor or independent variables to predict the outcome of a categorical dependent variable (a dependent variable with a limited number of values). The dependent variable in logistic regression is always binary (with two categories). The major characteristics of logistic regression includes prediction and calculating the probability of success. Fitting an equation of the form to the data is just what Logistic Regression should be all about:

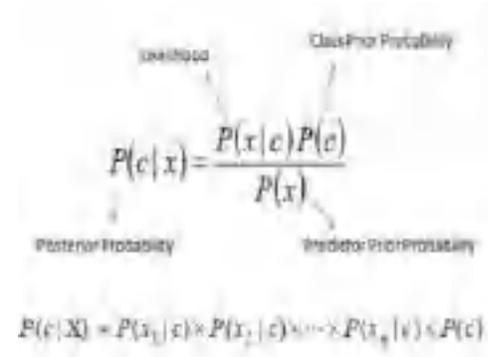
$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n - eq. 1$$

Maximum likelihood estimation is frequently used to measure

regression coefficients. The maximum likelihood ratio is a statistical tool that would be used to calculate the statistical relevance of independent variables on dependent variables. The likelihood-ratio test determines how important individual predictors are (independent variables). Then, using the odds ratio, $P/(1-P) = e^Y - eq$, the likelihood (p) of each case is computed. 2 The p-value is calculated using this information. This estimates the probability of the individual developing coronary heart disease.

B) Naïve Bayes

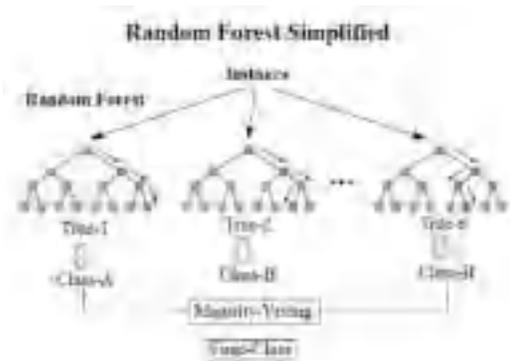
Based on the Bayes Theorem, Naive Bayes is a simple but efficient classification method. It assumes predictor independence, which means that the characteristics or features should not be correlated with one another or linked in any way. Even though there is a dependence, all of these characteristics or qualities contribute to the likelihood independently, which is why it is called Nave. When all 13 features of the CHDD are used, Naive Bayes achieves an accuracy of 83.49 percent



C) Random Forest

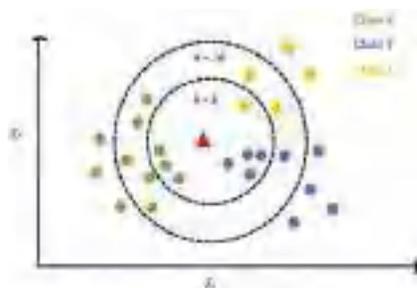
Random Forest is another supervised machine learning algorithm that is widely used. This approach can be used for both regression and classification tasks, but it performs best in the latter. Before producing an output, the Random Forest method considers various decision trees, as the name implies. As a result, it's essentially a collection of decision trees. This method is based on the idea that a larger number of trees would eventually lead to the correct decision. It uses a voting system for classification and then determines the class, while it uses the mean of all

the decision tree outputs for regression. It works well with big datasets that have a lot of dimensions.



D) K-Nearest Neighbour

KNN is a nonparametric, lazy, and straightforward classifier. When all features are constant, KNN is the best option. KNN, also known as case-based reasoning, has been used in a variety of applications, including pattern recognition and statistical estimation. To ascertain the class of an unknown sample, find its closest neighbour. Because of its fast convergence and ease of use, KNN is favoured over other classification algorithms. The closest neighbour classification is shown in the diagram below. There are two stages to KNN classification:



1. In the dataset, find the k number of instances that are nearest to instance S.
2. A vote is taken among the k instances to determine the class of instance S.

KNN accuracy is determined by the distance metric and K value. Cosine and Euclidian distance are two methods of determining the distance between two examples. KNN computes its K closest neighbours to analyse the new unknown sample and assigns a class by majority voting.

E) Decision Tree

A decision tree is a tool that predicts a response to data using

classification or regression. When the features are grouped, classification is used, and when the data is continuous, regression is used. One of the most common data mining techniques is the decision tree. A root node, branches, and leaf nodes make up a decision tree. Follow the path from the root node to a leaf node to analyse the data. As mentioned in the training section, decision trees must be constructed using a purity index that splits the nodes. Each of the 303 tuples is assessed down the decision tree for the CHDD, resulting in a positive or negative assessment of heart disease. These are contrasted with the initial decisions.

F) Support Vector Classifier

In classification and regression analysis, a support vector machine is a type of model that is used to interpret data and find patterns. When your data has exactly two classes, the support vector machine (SVM) is used. The best hyper plane that separates all data points of one class from those of the other class is found by an SVM. The better the model, the wider the margin between the two classes. There must be no points in the interior area of a margin. The data points on the margin's edge are known as support vectors. SVM is a mathematical function-based model that can be used to model complex, real-world problems. SVM works well with data sets with a lot of features, like the CHDD.

G) Neural Network

Neural Network with Multilayer Perceptrons (MLPNN) Multilayer Perceptron is one of the most fundamental models in Artificial Neural Networks (MLP). The system is implemented using a Multilayer Perceptron Neural Network architecture (MLPNN). One input layer, one output layer, and one or more hidden layers make up the MLPNN. Each layer has one or more nodes, which are depicted as small circles. The lines connecting nodes represent the flow of data from one node to the next. Signals from external nodes are received by the input layer. Through weighted connection links, the output of the input layer is passed on to the hidden layer. It runs computations and sends the results to the output layer via the output layer.

Steps:

1. Input data is provided to the input layer for processing, which results in a predetermined output.
2. The predicted output is subtracted from actual output, and the error value is calculated.
3. The network then adjusts the weights using a Backpropagation algorithm.
4. Weights adjustment begins weights between output layer nodes and last hidden layer nodes, and works backwards through network.
5. The forwarding process resumes once the back propagation is completed.
6. The procedure is repeated until the difference between the predicted and actual output is minimised.

Backpropagation network

Backpropagation is the most widely used training algorithm for multilayer and feed forward networks. Back propagation gets its name because the difference between real and predicted values is calculated and propagated backwards from output nodes to nodes in the previous layer. This is done to increase processing weights. The following are the steps that make up the Backpropagation algorithm:

Steps to take:

1. Provide the network with training data.
2. Make a comparison between the real and expected results.
3. Determine each neuron's error.
4. Determine what each neuron's output should be and how much lower or higher output must be adjusted to achieve the desired output.

Adjust the weights after that.

H) Extreme Gradient Boost

Extreme Gradient Boosting is abbreviated as XGBoost. It's a learning algorithm that's supervised. It's a library for creating high-performance gradient boosting tree models in a short amount of time. XgBoost's main benefit is parallel computing. It has recently gained popularity due to its adaptability, scalability, and effectiveness. It promotes gradient boosting, stochastic gradient boosting, and regularised gradient boosting, all three types of gradient boosting. XGBoost runs these gradient boosting methods in parallel to find the most effective model.

Gradient boosting can be explained as follows:

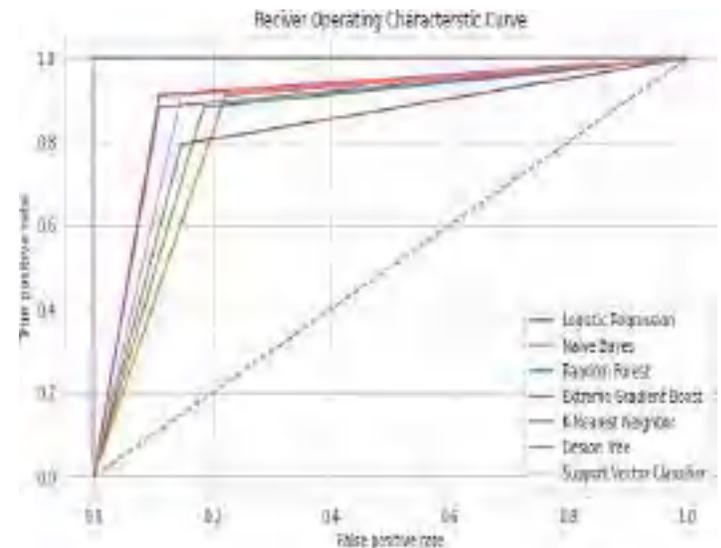
1. Consider a binary classification problem. Small decision trees are referred to as weak learners. Each data point is given a weight (all of which are initially equal), and the error is the sum of the weights of misclassified examples.
2. We have a weak predictor that classifies the data points here.
3. Now, we'll range the weights of the points misclassified by the previous predictor and learn another predictor on data points. As a result, the residual is reduced.
4. This procedure is repeated for the specified number of iterations until the residual is minimal.

During training, the K-fold cross-validation method is used to improve the model's accuracy. The training set is divided into k subsets using this technique. One subset is considered the testing fold for each iteration, while the remaining K 1 subsets are considered the training folds. The model is used to train the training folds, and the testing fold is used to assess the model. During various iterations, all K subsets are treated as testing folds. This improves the overall accuracy of the model before it is tested on the testing set.

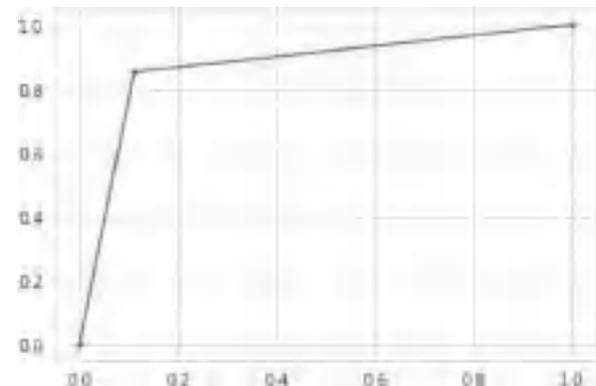
5 Result

	Model	Accuracy
0	Logistic Regression	85.245902
1	Naive Bayes	85.245902
2	Random Forest	85.245902
3	Extreme Gradient Boost	90.163934
4	K-Nearest Neighbour	88.504590
5	Decision Tree	81.967213
6	Support Vector Machine	68.624590
7	Neural Network	95.041323

(Accuracy table for all the algorithms applied)



(ROC for ML models)



(ROC for ANN)

6 Conclusion

The prediction models are developed using 13 features and the accuracy is calculated for modeling techniques. The highest accuracy is achieved by neural network method in comparison with existing methods. For future work, we propose a DNN for deeper network and more accuracy.

7 Reference

- [1] A. S. Abdullah and R. R. Rajalaxmi, "A data mining model for predicting the coronary heart disease using random forest classifier," in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.
- [2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease," in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306–311.
- [3] N. Al-milli, "Backpropagation neural network for prediction of heart disease," J. Theor. Appl. Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.
- [4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, "Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.

Bypass Fraud Detection: Artificial Intelligence Approach

Aftab Ibrahim Panhalkar, Priyanka Sonawane

Abstract— Telecom organizations are significantly broken via way of means of pass fraud (SIMboxing). However, There is a scarcity of posted studies to tackle this problem. The conventional approach of Test Call Generators (TCG) is easily conquer via way of means of fraudsters and the want for greater state-of-the-art approaches is inevitable. In this work, we're growing shrewd algorithms that mine a big amount of cell operator's records and hit upon the SIMs that used to pass international calls. This approach will make it tough for fraudsters to generate sales and prevent their network. Also via way of means of lowering fraudulent activities, pleasant of provider may be extended in addition to purchaser satisfaction. Our method has been examined and evaluated on actual global cell operator records, and it proved to be very efficient.

Keywords—Artificial Intelligence, Bypass Fraud, Fraud Detection, SIMboxing

I. INTRODUCTION

We have studied the hassle of skip fraud (SIMboxing) and located that even though skip fraud has been adverse telecom corporations severely, there was a scarcity of posted studies to take a look at it and clear up it. So our paintings has been fold, one is to growth attention about this massive hassle and to reveal that traditional strategies couldn't be used to clear up it. We applied a few shrewd strategies to correctly detect SIMboxing fraud and save you it from affecting telecomm corporations now no longer best when it involves sales however additionally in relation to denial of service, best of service, and communications network congestion.

In this work, a detection system is designed and tested in cooperation with the Tier 1 mobile operator in India (Top 10 Mobile Phone Company). The system is utilizing artificial intelligence techniques to detect whether a SIM card is used by a normal customer or by a fraudster.

II. IMPACT OF FRAUD

Telecommunications, which have become a necessity worldwide, became a target for fraudsters who are making a lot of money out of illegally accessing communication networks and using it to make huge profits, by selling services at much lower prices than their original prices. According to a survey by the Communication Fraud Control Association (CFCA) [1], the mobile telecom industry lost more than 38 billion dollars in 2015 alone due to telecom fraud. Besides those big losses, telecom fraud causes other indirect losses to mobile operators, like: decrease in quality of service, deny of service and network congestion.

Bypass fraud costs telecom companies 6 billion annually and ranked the 2nd most costly fraud worldwide. Fig. 1 shows the top 3 fraud types with their annual losses. The numbers are huge, since major mobile operators in Libya are owned by the state, the revenue obtained by these major companies could help in Libya's economic growth, and anything that would affect it would degrade the country's GDP.



Fig. 1. CFCA 2015 survey, Top 3 fraud losses globally

III. TYPES OF FRAUD

Phone systems fraud is not just due to phone theft or hacking, It is much more than that. Yelland [2] lists seven of them. The top three types that cause a significant loss are:

International Revenue Share Fraud (IRSF)

IRSF is the largest contributor to the overall fraud losses according to CFCA. It is when a Fraudster makes an agreement with a local carrier in high cost destination to share profit for increasing traffic, then the fraudster hacks into any organization's public branch exchange (PBX) and gets illegal access to generate calls. After that, the fraudster generates high traffic calls to high cost destinations and gets revenue from the sharing agreements. [4].

Premium Rate Service Fraud

It must have happened to you; that is receiving a text message saying that you won a big prize and all what you need to do is to call a certain number. If you called that number you have been tricked to use some premium rate service. Premium rate service is an agreement between some service provider and telecom companies to share revenues generated by traffic to the premium service number, it is used in TV shows and contests and entertainment services. The fraudsters try to stimulate the costumers by giving them a missed call or a message, then make money from share of the call back revenues [4].

Bypass fraud

To explain how bypass fraud is committed, firstly we describe the legitimate way for international calls. Let's assume that caller A and caller B live in different countries. Caller A makes a call to caller B over the mobile operator. The mobile operator of country A takes the call and send it through his international gate to a transient operator. The transient operator then routes the call through voice over IP (VoIP) to the country B mobile operator and pays a toll. After that, the mobile operator of country B terminates the call through his network to caller B. Fig. 2 shows the legitimate route.

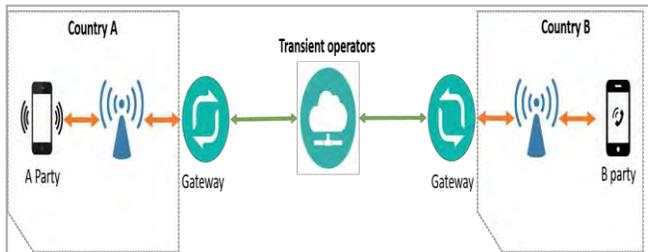


Fig. 2. The legitimate route of international call

In bypass fraud, the transient operator route the call through a SIMbox placed in country B using VoIP, the SIMbox then re-route the call through country B mobile operator and pay for just the local call. Fig. 3 shows the bypass fraud route.

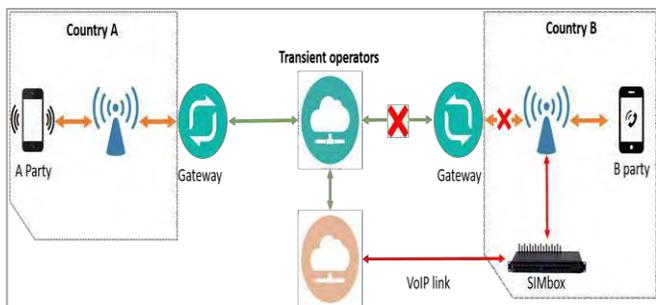


Fig. 3. The bypass fraud route of international call

The incentive here is the toll charge by country B mobile operator is much higher than the local call fee, so the bypass will be financially viable.

Bypass fraud is committed when fraudsters install SIMboxes with multiple low-cost, prepaid SIM cards. SIMbox equipments includes SIM slots, antennas. The SIMbox is connected to the internet through Ethernet port. Fig. 4 shows a SIMbox and its components.



Fig. 4. A SIMbox and its components

IV. BATTLING BYPASS FRAUD

The major methods used today for battling fraudsters are:

Test Call Generation (TCGs)

Test call generation is used as an active method to detect bypass fraud, where operators test different international routes to their network and see whether calls go through legitimate routes or SIMbox routes. This method detects fraud with no false positive (that is when a normal user assumed to be a SIMbox). However, this method is probabilistic in nature and costly in terms of the need to test huge number of international routes. Also, fraudsters use tricks to avoid test call detection as we will see in the anti-spam method.

Fraud Management Systems (FMS)

Fraud management systems use measures to detect the abnormal usage of SIM cards. FMS analyze Call Details Record (CDR) data to make usage profiling that distinguishes normal users from SIMboxes.

SIM Card Distribution Control (SDC)

SIM cards are vital in the bypass cycle, and fraudsters must maintain an adequate supply of SIM cards to be in business. However, SIM card distribution control will make this process difficult. Requiring government IDs and limiting the number of SIM cards per ID will prevent fraudsters from obtaining a large number of SIM cards to install in their SIMboxes.

V. HOW FRAUDSTERS AVOID DETECTION

Fraudsters and anti-fraud are in eternal battle, every time detection technology improves, fraudsters are developing their methods to avoid detection and increase profit. This section describes different methods used by fraudsters to avoid SIM blocking.

Anti-Spam (Test Call Detection)

One of the effective methods to detect SIMs used in SIMboxes is generating test calls (TCG) using different routes to known local network numbers. The incoming call will appear weather it is coming from a local number or from an international number; if it was coming from a local number then it must be associated with some SIM card used in a SIMbox and easily processed by the fraud department. However, the fraudsters analyze the voice call traffic coming toward their SIMboxes and based on usage and other patterns they could determine whether the calls were real subscriber calls or they were originated from a TCG system. They could then either block the test calls and prevent them from reaching the SIM box, to begin with, or reroute the calls to a legitimate route so as to avoid detection.

Human Behavior Simulation (HBS)

According to the literature [3], some features can be used to identi-

fy SIMbox fraud, for example: 1. The SIMbox is not moving. 2. Most calls are outgoing calls. 3. No usage of network services like SMS, GPRS. and others. However, Smart SIMboxes are designed to mimic the behavior of normal customers by using Human Behavior Simulation (HBS). This technique makes detection of fraudsters very difficult if no advanced detection algorithms were used. HBS encompasses the following:

SIM Migration (Movability)

Fraudsters are deploying many gateways in different locations, for example, one in the city center and another in a shopping mall or some other crowded place and once in a while they swap the SIM cards between the gateways, so it would look like that the user is moving. The swapping operation could be done manually or automatically using software.

SIM Rotation

SIMboxes can be detected easily if fraudsters operate their SIMs around the hour excessively, so they limit their usage by rotation of the SIMs as workers shifts. This will make SIMs operate in limited hours a day, which simulates the behavior of ordinary customers.

Usage of Other Network Services

Most of the SIMboxes are using just voice services and that makes them vulnerable to detection. In order to mitigate this issue smart SIMboxes are making calls and sending SMS to each other. Also, sometimes they use some internet services provided by the network operator.

Traditional SIMboxes just reroute the call from VoIP to the GSM network, so they make calls to large numbers of different network customers. A smart way to avoid this is by using family lists, where each SIM is assigned to reroute calls to a specific list of numbers. This leads to escaping the trap of large different numbers detection.

To summarize, Human Behavior Simulation (HBS) makes dealing with bypass fraud harder and harder and time-consuming. Advanced measures must be taken to tackle this problem. In this work, intelligent machine learning algorithms were used to detect the bypass fraud by analyzing huge CDR data

VI. LITERATURE REVIEW

Even though the telecommunication industry suffers major losses due to fraud, there is no comprehensive published research on this area mainly due to lack of publicly available data to perform experiments on. On the other hand, any broad research published publicly about fraud detection methods will be utilized by fraudsters to evade detection.

This section presents the literature of the available research published. Most research investigated CDR analysis combined with machine learning algorithms to detect fraudulent SIMboxes [4] [5] [3], and few others used Audio analysis [6].

In [4], 234,324 calls made by 6415 subscribers from one Cell-ID during two months were analyzed. The dataset consisted of 2126 fraud subscribers and 4289 normal subscribers which are equivalent to two thirds of legitimate subscribers and one third of fraudulent SIMboxes. The researchers extracted 9 features, like Total Calls, Total Number Called, Total Minutes and Average Minutes, etc. Then they used

the extracted features to train an Artificial Neural Network (ANN) classifier, where three architectures of neural networks were considered and three hidden layers; 5, 9 and 18 hidden nodes in each layer. They found that the best architecture was when two hidden layers were used, each having five hidden neurons, with a learning rate of 0.6 and a momentum term of 0.3. The accuracy reached 98.7% with just 20 accounts were wrongly classified as false positive.

In [5], researchers extended their previous work in [4] to design and compare two classifiers; Support Vector Machine (SVM) classifier and ANN classifier. Same dataset and features were used. They found that SVM has better accuracy compared to ANN. SVM gave 99.06% accuracy while ANN model gave 98.69% accuracy. In addition, the SVM training time was found to be three times less than the time consumed by the ANN training time.

Another more broad research was conducted by [3]. In contrast with [4] and [5], where data was constrained to only one Cell-ID, a larger dataset was used with accounts distributed nationwide. They analyzed CDRs from main cellular operator in the United States (AT&T). The dataset contains CDRs of 93000 legitimate accounts and 500 of

fraudulent accounts. For training the classifier they split the dataset to two-thirds for training and one third for testing. Using International Mobile Equipment Identity (IMEI) as a device identifier other than the subscriber identifier they computed 48 features characterizing patterns of legitimate and fraudulent

IMEIs.

They observed that fraudulent SIMboxes have common patterns as the following:

High number of Int. mobile subscriber identity number (IMSI) per IMEI.

Static physical location.

Large number of international phone calls.

Large volume of outgoing calls generated compared to incoming calls.

For classification, they used a combination of decision tree classifiers (alternating decision tree, functional tree, and random forest). The classification rule was the linear combination of their results. The accuracy of the classification rule was 99.95% with lowest false positive achieved by the random forest.

On the other hand, instead of using CDR analysis, [6] used real-time call audio analysis to detect fraudulent calls. They designed a system that relies on the raw voice data received by the tower during a call to distinguish errors in GSM transmission from the distinct audio distortions caused by delivering the call over a VoIP. They used fast signal processing techniques to identify whether individual calls are likely made by a SIMbox and then to develop profiles of SIM cards. Their resultant system was able to detect 87% of real SIMbox calls in only 30 seconds of audio with no false positives. Their system promises a real-time detection capability if it were deployed on network towers and embedded on the operators Base Transceiver Station (BTS).

VII. DESIGNING THE DETECTIN SYSTEM

Designing the detection system was done in four stages:

Data Collection and Cleaning: In cooperation with Almadar Aljadeed Mobile Phone Company, fully anonymized Call Detailed Records (CDR) were obtained and utilized. The CDR's sensitive data like user numbers or identities were obfuscated and hashed by the mobile operator. CDR data was processed in order to remove missing values, duplicate information and useless fields (here we used Pathon filters).

Feature Extraction and Engineering: To prepare the data for the machine learning model, informative features were extracted. In this stage features for each SIM were extracted. Also, features had been engineered in order to get the best features for training the model. Feature engineering includes feature selection and dimensionality reduction in generating more useful features.

Model Training: The features extracted in the previous stage were used to train machine learning models. In this stage, different models were trained to detect SIMs that were used in SIMboxes. SVM and Decision trees (random forest algorithm) had been used as supervised learning algorithms, and since the labeled data is scarce, we used unsupervised learning algorithms to cluster the SIMs in order to get insights of how we could improve the designed algorithm.

Model Evaluation and Testing: The final stage was testing and comparing the performance of designed algorithms in terms of accuracy and precision and that showed that our system is very efficient. We tested our system on real time data on site, where we took our system to our sponsor (Almadar Aljadeed Mobile Phone Company) headquarters and our tests showed the high efficiency of our bypass fraud detection system.

Fearing that fraudsters might use information published here to their advantage, affected how much details we could reveal, but, our data and detailed work techniques would be available for all of those concerned in our presentation.

VIII. CONCLUSIONS

We have studied the problem of bypass fraud (SIMboxing) and found that although bypass fraud has been damaging telecom companies severely, there has been a shortage of published research to study it and solve it. So our work has been two fold, one is to increase awareness about this big problem and to show that conventional techniques could not be used to solve it. We utilized some intelligent techniques to effectively detect SIMboxing fraud and prevent it from affecting telecomm companies not only when it comes to revenues but also when it comes to denial of service, quality of service, and communications network congestion.

REFERENCES

- [1] Communication Fraud Control Association, 2015 Global Fraud Loss Survey [Internet]. 2015 [cited 2016 Dec 3]. Available from: <https://goo.gl/HT9IER>
- [2] M. Yelland, "Fraud in mobile networks," *Comput. Fraud Secur.*, vol. 2013, no. 3, pp. 5-9, 2013.
- [3] I. Murynets, M. Zabaranin, R. P. Jover, and A. Panagia, "Analysis and detection of SIMbox fraud in mobility networks," *Proc. - IEEE INFO-COM*, pp. 1519-1526, 2014.
- [4] R. S. A. H. Elmi, S. Ibrahim, "Detecting SIM Box Fraud Using Neural Network," *IT Converg. Secur.* 2012, vol. 215, pp. 575-582, 2013.
- [5] R. Sallehuddin, S. Ibrahim, A. Mohd Zain, and A. Hussein Elmi, "Classification of SIM Box Fraud Detection Using Support Vector Machine and Artificial Neural Network," *Int. J. Innov. Comput.*, vol. 4, no. 2, pp. 19-27, 2014.
- [6] B. Reaves, E. Sherman, A. Bates, H. Carter, and P. Traynor, "Boxed Out : Blocking Cellular Interconnect Bypass Fraud at the Network Edge," *USENIX Secur. Symp.* 2015, pp. 833-848, 2015.

Application of Artificial intelligence in Diabetic Retinopathy

Aftab Ibrahim Panhalkar, Priyanka Sonawne

Abstract— In the field of Medicine Artificial intelligence is an evolving technology which holds much promise for diagnosis and also accurate screening. The big step in complex computing is the ability to perform pattern recognition based on input data and comparing it with performance standards. An ever increasing problem is Diabetic retinopathy. Early diagnosis and screening may help in reducing sight threatening retinopathy. A tool that helps patients and ophthalmologists in quick screening without more help from human trainee would be a boon. In this review we discuss the current status of use of artificial intelligence in diabetic retinopathy and few other common retinal disorders.

Index Terms— Artificial intelligence, IDx-DR, fundus image, screening, CNN, Diabetic Retinopathy

1 INTRODUCTION

Artificial intelligence (AI) has risen to prominence as a major research area of computer science. This technology will improve healthcare affordability, efficiency, and accessibility. In simple terms, AI refers to the completion of a job primarily by a machine or a robot with limited human involvement. In other words, AI is the software/machine simulation of human intelligence. It is essentially the ability of a computerized system to show cognitive abilities.^[2] Just like learning in humans, the AI systems need to be exposed to a database which allow them to first “learn” simple targets with regards to a dedicated finding or disease. However, AI is much more than simply a humongous database. Following initial steps of learning, the system or machine is then taught to “improve”, i.e., evolve upon its initial learning to become more accurate and efficient.^[2] This learning is further compounded by use of complex mathematical equations for the system to understand nonlinear relationships between different variables through an information flow referred to as “neural networks.” In essence, this form of “higher training” allows AI to judge and weigh possibilities of different outcomes, much like an ideal physician! Many of these technological advances are in part due to software that have now been made available by resources related to information and technology. An example of commercially available machine learning software is Scikit-learn which require computer languages like Python as a platform for working. Scikit learn is a machine learning library for use in python programming language and it has been used for diabetic retinopathy detection. Several commercial software currently integrates artificial Intelligence and machine learning for fundus image processing and evaluation.

Fundus evaluation software are being made as student projects in several engineering colleges for the past few years. Open Indirect Ophthalmoscope by LVPEI and MIT has an inbuilt option for DR detection by Machine Learning. Also, Kavya Kopparapu’s Eyagnosis app, which along with a 3D printed smartphone fundus camera, were developed in 2016 and tested in several leading eye hospitals.

AI-assisted medical screening and diagnosis based on images are currently evolving.^[3,4] Application of this technology in ophthalmology is currently focused mainly on the diseases with a high incidence, such as diabetic retinopathy (DR), age-related macular degeneration (ARMD), glaucoma, retinopathy of prematurity (ROP), age-related or congenital cataract, and retinal vein occlusion (RVO).

DR is a form of eye disease that causes mild to extreme vision loss and is the leading cause of blindness in working-age people who have had diabetes for a long time. (5) The health burden is accentuated by the huge per capita cost. This has further increased since the introduction of anti VEGF agents. Very often the disease does not show overt symptoms until it reaches an advanced stage; however, if detected early on, vision impairment can be averted by early intervention which is also the most cost-effective option. In view of the alarming increase in the number of people with diabetes and dearth of trained retinal specialists and ophthalmologists, a computer-based analysis of the fundus images by an automated approach would lessen the burden of the health systems in screening for DR and offer a near ideal system for its management.^[6-8] Therefore, screening will be valuable at any stage of the disease and will also be helpful in avoiding blindness among 90% patients.^[9]

We searched the PubMed and Springer databases for all ophthalmology studies conducted in English. The articles published in last 10 years that we deemed relevant were summarised. The keywords used for the PubMed search were artificial intelligence, ophthalmology, deep learning, machine learning, and diabetic retinopathy.

2 THE PROBLEM

2.1 Review Stage

Regardless of the form of diabetes, all people diagnosed with DM need annual retinal screening in order to detect diabetic retinopathy early and effectively treat it (DR). 10 and 11 Traditionally, fundus examinations by ophthalmolo-

gists or colour fundus photography with traditional fundus cameras (mydriatic or nonmydriatic) are used to test for retinopathy by trained eye technicians or optometrists.^[12] The primary issue is the grading of the retinal images by ophthalmologists (retinal specialists) or trained persons, whose numbers are very scarce compared to the load of patients requiring screening. Second, some of these patients are based in rural areas and can't visit an eye care provider. Thirdly, as such follow ups are required for years together, the attitude, and/or behavioral aspects negatively impact the patients practice despite knowledge of consequences. These issues can be solved with provision of an automated imaging system within easy reach of the patient. Hence, there has been an increasing interest in the development of automated analysis software using computer machine learning/artificial intelligence (AI) for analysis of retinal images in people with diabetes thus solving at least some part of the problem.^[13]

2.2 The Solution: Principle Behind Artificial Intelligence

It's simply a way of training a machine to recognise patterns like this. It has previously been used for a variety of technical tasks, including the accurate classification of high-resolution photographs. The following major categories are used to classify AI system techniques. [2] machine learning techniques,^[14] natural language processing processes, voice, vision, expert systems, and so on. So far, the machine learning techniques are more utilized in ophthalmology.^[15] The training and validation sets are the two key parts of the machine learning process. This process begins with a large amount of training data, such as thousands of retinal images with varying degrees of DR, being provided to the machine/system as the training collection. [number 16]. Most of the data are labeled as per features in advance by the authoritative professionals. After being exposed to numerous annotated retinal images the machine learns to grade DR by itself by building a model of complex relationships between input data and generalizing a performance standard. In addition, some other data are used to verify the established algorithm i.e., validation set [Flow Chart 1].

Two main deep learning models have been in papers in described: the convolutional neural network (CNN) and the massive-training artificial neural network (MTANN).^[17]

2.3 Diabetic Retinopathy and Artificial Intelligence

In patients with uncontrolled diabetes, at least yearly fundus screening is recommended for early diagnosis and treatment to reduce the disease burden in the population. Just half of these patients, however, are screened. [nineteen] Currently, screening involves a referral to an ophthalmologist, and patients may be unable to see the specialist due to a lack of an eye specialist in their region logistical barriers, or cost of the visit. One way of addressing such problems is by obtaining color fundus images and sending those to eye specialist or optometrists for reading.^[19] Although these programs increase screening rates, still there

are logistical barriers, cost issues, and time delays.^[20] Such limitations created interest in assessment of images using fully automated AI-based grading systems. In real time the system would decide whether a patient requires referral and could potentially be much cheaper than having ophthalmologists conduct screening. In April 2018, the US Food and Drug Administration (FDA) approved an AI algorithm, developed by IDx, used with Topcon Fundus camera (Topcon Medical) for DR identification.^[21]

The enthusiasm in the field of artificial intelligence has led to several studies using retinal images to test the performance of AI grading systems for detecting DR. Few of the studies are worth mentioning. [Table 1] Historically the Wisconsin Fundus Photograph Reading Centre (FPRC) has been the gold standard for trials that require grading of the severity of DR, including the Diabetes Control and Complications Trial (DCCT), Diabetic Retinopathy Clinical Research Network (DRCR.net) studies. Wong *et al.*^[22] proposed a model to classify the DR stages based on microaneurysms and hemorrhages. Imani *et al.*^[23] shaped a different technique in which, by using morphological component analysis, they spotted the exudation and blood vessel. The vessel map is obtained by using adaptive thresholding. Yazid *et al.*^[24] published identification of hard exudation and optic disc based using inverse surface thresholding. The main objective of this paper was to detect both hard and soft exudates. However, since the color of the exudates is similar to that of optic disc, they are usually detected together. Basically, fuzzy c-means clustering, edge detection and Otsu thresholding were utilized to separate edge pixels of the exudates from the background. Akyol *et al.*^[25] by using key point detection, texture analysis, and visual dictionary techniques detected the optic disc of fundus images. Studies have reported sensitivity of automatic DR screening ranging from 75% to 94.7% and the specificity accuracy is comparable.^[26] A study using smartphone-based fundus photography system to evaluate the usefulness of an automated AI-based interpretation of screening at a physician clinic has also shown positive results.^[27] The study reported a "high sensitivity for detection of DR above 95% using the EyeArt software when used on retinal images taken with Fundus on Phone (FOP)". A recent study done by EyeNuk with retinal images taken with traditional desktop fundus cameras showed that EyeArt's sensitivity for DR screening was 91.7% and 91.5% specificity.^[28] Another very recent major study on validation of deep learning (AI) by Ting *et al.*^[29] done in Singapore with multiple retinal images taken with conventional fundus cameras showed a high sensitivity and specificity for identifying DR and other eye diseases like ARMD. An automated tele-retinal DR screening program, IRIS (intelligent retinal imaging system), compared non-mydriatic retinal images with a standard data set images from Early Treatment Diabetic Retinopathy Study (ETDRS), and proposed recommendations for referral. Any patient with severe NPDR or more advanced disease was considered suitable for the referral.^[8] This screening program reported good sensitivity and a low false-negative rate.

Use of AI to evaluate retinal images is alluring as it fits in with the present trend of tele-ophthalmology and telemedicine. Selected patients who have sight-threatening

DR would need to meet the eye specialist. Urgent referral of these patients is critical, since DR affects people during their prime productive years of life.^[30]

of security about the retinopathy status. For the present, a comprehensive dilated eye examination remains the gold standard of screening and cannot be replaced with this device till appropriately proved otherwise. Diabetes has numerous ocular manifestations other than DR, which

2.4 IDx-DR

If you IDx DR is the first FDA-approved AI algorithm for detecting DR in non-ophthalmic offices. healthcare practitioners.^[21] The device is paired with a non-mydratric retinal camera (TRC-NW400, Topcon) and the captured images are sent to a cloud-based server. The server then utilizes IDx-DR software and a “deep-learning” algorithm to detect retinal findings consistent with DR based on autonomous comparison with a large dataset of representative fundus images. The software provides one of the two results: (1) If more than mild DR detected, refer to an eyecare professional (ECP); (2) If the results are negative for more than mild DR, re-screen in 12 months.^[31]

The FDA approval of IDx-DR device was based on a study on 900-subjects in a primary-care setting (10 primary care sites) with automated image analysis. Two 45-degree digital images per eye (one centered on the macula, one centered on the optic nerve) were obtained and analyzed. These images were compared with the stereo, widefield fundus imaging interpreted by the Wisconsin Fundus Photograph Reading Centre (FPRC). After procurement of retinal images, the artificial intelligence system is able to make a diagnosis in just 20 seconds.

Based on the analysis a new entity called more than minimal DR (mtmDR) was defined. It in nothing but the presence of ETDRS level 35 or higher (microaneurysms plus hard exudates, cotton wool spots, and/or mild retinal hemorrhages) and/or DME in at least one eye.^[32] 96 percent of acquired images were of sufficient quality for algorithmic assessment, which was really high numbers in primary care settings. Sensitivity and specificity of the technology was 87.4% and 89.5% respectively for detecting more than mild DR. It's worth mentioning that 100% of subjects with ETDRS levels of 43 or higher DR were correctly identified by the algorithm. As the device delivers a screening decision without necessitating an eye specialist, it can also be used by non-ophthalmic healthcare professionals.

3 DOWNSIDES OF AI

Since the device's sensitivity and accuracy are both below 90%,^[21] 1 in 10 patients may potentially have a false positive or false negative test. As a consequence, it isn't absolutely failsafe. Thus, it is crucial to educate patients and doctors that the present generation devices are not 100% reliable. A false negative result may provide a pseudo sense

Includes glaucoma, age-related macular degeneration (ARMD), cataract, dry eye. A comprehensive examination is obligatory for proper diagnosis and management in these patients.^[32]

Diabetic macular oedema is the leading cause of vision loss in patients with diabetes. Stereoscopic macular examination coupled with optical coherence tomography remains the gold standard for diagnosing this condition. Though all subjects with ETDRS level 43 or higher DR were detected via IDx-DR, but many cases of subtle DME were missed because of non-addressal. Legal accountability in cases of misdiagnosis with artificial intelligence is another subject that is yet to be fixed.

4 FUTURE OUTLOOK

AI's application in medical diagnostics, especially in ophthalmology, ushers in a new age. If this technology is shown to be responsive and precise enough, it has the potential to completely transform the way we think about screening services and community-based ophthalmology programmes. Most of the present systems use conventional of 30–50° fundus images. Perhaps applications based on wide field imaging and OCT angiography based vascular analysis might yield even more consistent results. However, the high cost of wide field imaging and OCT angiography may be a limiting factor for this at present. A lot of work is also being done on identifying serum biomarkers for early detection and monitoring of diseases like diabetic retinopathy. Thus, a comprehensive analysis of ocular imaging, systemic parameter profile and other serum biomarkers using AI might provide better insights, perhaps even better conclusions than what human intelligence is capable of deriving.

5 CONCLUSION

The AI DR method can provide the clinician with fundus image review, which can help inform the next steps in the patient's care more quickly. Also, without mydriasis, doctors will attend to more patients that need treatment. Emerging healthcare technologies emphasize on reducing visits to eye specialists, curtailing the overall cost of treatment and optimizing the number of patients seen by each doctor. AI can help the health care professional in achieving the goal. Though it assists in health care sector but should not substitute a clinician at its current level. Novel developments in the sector of artificial intelligence are opening up new promises for running DR detection and grading algorithms.

6 REFERENCES

1. Hamet P, Tremblay J. Artificial intelligence in medi-

- cine. *Metabolism* 2017;69S: S36-40.
- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, *et al.* Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc Neurol* 2017;2;2:230-43.
 - Doi K. Computer-aided diagnosis in medical imaging historical review, current status and future potential. *Comput Med Imaging Graph* 2007;31:198-211.
 - Doi K. Diagnostic imaging over the last 50 years research and development in medical imaging science and technology. *Phys Med Biol* 2006;51:R5-27.
 - Kocur I, Resnikoff S. Visual impairment and blindness in Europe and their prevention. *Br J Ophthalmol* 2002;86:716-22.
 - Abramoff MD, Niemeijer M, Suttorp-Schulten MS, Viergever MA, Russell SR, van Ginneken B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care* 2008;31:193-8.
 - Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-10.
 - Walton OB, Garoon RB, Weng CY, Gross J, Young AK, Camero KA, *et al.* Evaluation of automated tele-retinal screening program for diabetic retinopathy. *JAMA Ophthalmol* 2016;134:204-9.
 - Vashist P, Singh S, Gupta N, Saxena R. Role of early screening for diabetic retinopathy in patients with diabetes mellitus: An overview. *Indian J Community Med* 2011;36:247-52.
 - Fong DS, Aiello LP, Ferris FL, Klein R. Diabetic retinopathy. *Diabetes Care* 2004;27:2540-3.
 - Namperumalswamy P, Nirmalan PK, Ramaswamy KM. Developing a screening program to detect sight threatening retinopathy in south India. *Diabetes Care* 2003;26:1831-5.
 - Ryan ME, Rajalakshmi R, Prathiba V, Anjana RM, Ranjani H, VenkatNarayan KM, *et al.* Comparison among methods of retinopathy assessment (CAMRA) study: Smartphone, nonmydriatic, and mydriatic photography. *Ophthalmology* 2015;122:2038-43.
 - Bhaskaranand M, Cuadros J, Ramachandra C, Bhat S, Nittala MG, Sadda SR, *et al.* Automated diabetic retinopathy screening and monitoring using retinal fundus image analysis. *J Diabetes Sci Technol* 2016;10:254-61.
 - Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. *JAMA* 2016;315:551-2.
 - Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, *et al.* Automated identification of post-operative complications within an electronic medical record using natural language processing. *JAMA* 2011;306:848-55.
 - Lee A, Taylor P, Kalpathy-Cramer J, Tufail A. Machine learning has arrived. *Ophthalmology* 2017;124:1726-8.
 - Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol* 2017;10:257-73.
 - Murchison AP, Hark L, Pizzi LT, Dai Y, Mayro EL, Storey PP, *et al.* Non-adherence to eye care in people with diabetes. *BMJ Open Diabetes Res Care* 2017;5:e000333.
 - Rosenberg JB, Tsui I. Screening for diabetic retinopathy. *N Engl J Med* 2017;376:1587-8.
 - Taylor CR, Merin LM, Salunga AM, Hepworth JT, Crutcher TD, O'Day DM, *et al.* Improving diabetic retinopathy screening ratios using telemedicine-based digital retinal imaging technology: The Vine Hill study. *Diabetes Care* 2007;30:574-578.
 - US Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect cer-

- tain diabetes-related eye problems. Available from: <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm>. Published April 11, 2018. [Last accessed on 2018 Aug 12].
22. Wong LY, Acharya R, Venkatesh YV, Chee C, Min LC. Identification of different stages of diabetic retinopathy using retinal optical images. *Inf Sci* 2008;178:106-21.
 23. Imani E, Pourreza HR, Banaee T. Fully automated diabetic retinopathy screening using morphological component analysis. *Comput Med Imaging Graph* 2015;43:78-88.
 24. Yazid H, Arof H, Isa HM. Automated identification of exudates and optic disc based on inverse surface thresholding. *J Med Syst* 2012;36:1997-2004.
 25. Akyol K, Şen B, Bayır Ş. Automatic detection of optic disc in retinal image by using keypoint detection, texture analysis, and visual dictionary techniques. *Comput Math Methods Med* 2016;2016:6814791.
 26. Niemeijer M, Abramoff MD, van Ginneken B. Fast detection of the optic disc and fovea in color fundus photographs. *Med Image Anal* 2009;13:859-70.
 27. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye* 2018;32:1138.
 28. Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda S, *et al.* Automated Diabetic Retinopathy Screening and Monitoring Using Retinal Fundus Image Analysis. *J Diabetes Sci Technol* 2016;10:254-61.
 29. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, *et al.* Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318:2211-23.
 30. Rajalakshmi R, Amutha A, Ranjani H, Ali MK, Unnikrishnan R, Anjana RM, *et al.* Prevalence and risk factors for diabetic retinopathy in Asian Indians with young on set type 1 and type 2 diabetes. *J Diabetes Complications* 2014;28:291-7.
 31. Abramoff M. Artificial intelligence for automated detection of diabetic retinopathy in primary care. Paper presented at: Macular Society; February 22, 2018; Beverly Hills, CA. Available from: http://webeye.ophth.uiowa.edu/abramoff/MDA_MacSocAbst_2018_02_22.Pdf [Internet]. [Last accessed on 2019 Mar 26].
 32. Pros and Cons of Using an AI-Based Diagnosis for Diabetic Retinopathy: Page 4 of 5 N.d. *Optometry Times*. Available from: <http://www.optometrytimes.com/article/pros-and-cons-using-ai-based-diagnosis-diabetic-retinopathy>. [Last accessed on 2018 Oct 29].
 33. Ferris FL, Wilkinson CP, Bird A, Chakravarthy U, Chew E, Csaky K, *et al.* Clinical classification of age-related macular degeneration. *Ophthalmology* 2013;120:844-51.
 34. Chou CF, Cotch MF, Vitale S, Zhang X, Klein R, Friedman DS, *et al.* Age-related eye diseases and visual impairment among U.S. adults. *Am J Prev Med* 2013;45:29-35.
 35. Mookiah MR, Acharya UR, Fujita H, Koh JE, Tan JH, Nonha K, *et al.* Local configuration pattern features for age-related macular degeneration characterization and classification. *Comput Biol Med* 2015;63:208-18.
 - Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med* 2017;82:80-6.
 36. Bogunovic H, Waldstein SM, Schlegl T, Lang G, Sadeqhipour A, Liu X, *et al.* Prediction of Anti-VEGF treatment requirements in neovascular amd using a machine learning approach. *Invest Ophthalmol Vis Sci* 2017;58:3240-8.
 37. Zhao RQ, Chen ZH, Chi ZR. Convolutional neural networks for branch retinal vein. *Conf Proc IEEE* 2015. Available from: <https://ieeexplore.ieee.org/document/7279547>.
 38. Gilbert C, Fielder A, Gordillo L, Quinn G, Semiqlia R, Vintin P, *et al.* International NO-ROP Group Characteristics of infants with severe ROP in countries with low, moderate, and high levels of development implications for screening programs. *Pediatrics* 2005;115:e518-25.
 39. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020-the right to sight. *Bull World Health Organ* 2001;79:227.
 40. Ataer-Cansizoglu E, Kalpathy-Cramer J, You S, Keck K, Erdogmus D, Chiang MF. Analysis of underlying causes of inter-expert disagreement in retinopathy of prematurity diagnosis. Application of machine learning principles. *Methods Inf Med* 2015;54:93-102.
 41. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, Bozkurt A, Erdogmus D, Kalpathy-Cramer J, *et al.* i-ROP Research Consortium Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity performance of the "i-ROP" system and image features associated with expert diagnosis. *Transl Vis Sci Technol* 2015;4:5.
 42. Bolón-Canedo V, Ataer-Cansizoglu E, Erdogmus D, Kalpathy-Cramer J, Fontenla-Romero O, Alonso-Betanzos A, *et al.* Dealing with inter-expert variability in retinopathy of prematurity: A machine learning approach. *Comput Methods Programs Biomed* 2015;122:1-15.
 43. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Eng* 2015;62:2693-2701.
 44. Musch DC, Gillespie BW, Niziol LM, Lichter PR, Varma R. Intraocular pressure control and long-term visual field loss in the collaborative initial glaucoma treatment study. *Ophthalmology* 2011;118:1766-73.
 45. Oddone F, Lucenteforte E, Michelessi M, Rizzo S, Donati S, Parravan M, *et al.* Macular versus retinal nerve fiber layer parameters for diagnosing manifest glaucoma a systematic review of diagnostic accuracy studies. *Ophthalmology* 2016;123:939-49.

ROLE OF ARTIFICIAL INTELLIGENCE IN COVID - 19

JENCY KERLIN

ROLL NO 15

GUIDE – PRIYANKA SONAWANE

ABSTRACT—

“ARTIFICIAL INTELLIGENCE can be harnessed for forecasting the extends and out spreads of virus and growth of its pre warning systems by extracting data from social media stages , calls and daily news sites and provide ear opening information about the vulnerable regions and for prediction of morbidity and mortality”

◆

I INTRODUCTION

The CORONA VIRUS pandemic is taking a colossal toll in every creatures suffering and lives. A significant amount of new scientific research and data sharing is underway due to the pandemic which is still rapidly spreading. There is now a growing amount of corona-virus related datasets as well as published papers that must be leveraged along with artificial intelligence (AI) to fight this pandemic by driving news approaches to drug discovery, vaccine development, and public awareness. AI can be used to mine this avalanche of new data and papers to extract new insights by cross-referencing papers and searching for patterns that AI algorithms could help discover new possible treatments or help in vaccine development. Drug discovery is not a trivial task and AI technologies like deep learning can help accelerate this process by helping predict which existing drugs, or brand-new drug-like molecules could treat COVID-19. AI techniques can also help disseminate vital information across the globe and reduce the spread of false information about COVID-19. The positive power and potential of AI

must be harnessed in the fight to slow the spread of COVID-19 in order to save lives and limit the economic havoc due to this horrific disease

II BACKGROUND

In this worldwide health crisis, the medical industry is looking for new technologies to monitor and control the spread of COVID-19 (Coronavirus) pandemic. AI is one of such technology which can easily track the spread of this virus, identifies the high-risk patients, and is useful in controlling this infection in real-time. It can also predict mortality risk by adequately analyzing the previous data of the patients. AI can help us to fight this virus by population screening, medical help, notification, and suggestions about the infection control. This technology has the potential to improve the planning, treatment and reported outcomes of the COVID-19 patient, being an evidence-based medical tool. This shows the general procedure of AI and non-AI based applications that help general physicians to identify the COVID-19 symptoms.

2. Main applications of AI in COVID-19 pandemic

I)

Early detection and diagnosis of the infection

AI can quickly analyze irregular symptom and other 'red flags' and thus alarm the patients and the healthcare authorities. It helps to provide faster decision making, which is cost-effective. It helps to develop a new diagnosis and management system for the COVID 19 cases, through useful algorithms. AI is helpful in the diagnosis of the infected cases with the help of medical imaging technologies like Computed

tomography (CT), Magnetic resonance imaging (MRI) scan of human body parts.

II)

Monitoring the treatment

AI can build an intelligent platform for automatic monitoring and prediction of the spread of this virus. A neural network can also be developed to extract the visual features of this disease, and this would help in proper monitoring and treatment of the affected individuals. It has the capability of providing day-to-day updates of the patients and also to provide solutions to be followed in COVID-19 pandemic.

III)

Contact tracing of the individuals

AI can help analyze the level of infection by this virus identifying the clusters and 'hot spots' and can successfully do the contact tracing of the individuals and also to monitor them. It can predict the future course of this disease and likely reappearance.

IV)

Projection of cases and mortality

This technology can track and forecast the nature of the virus from the available data, social media and media platforms, about the risks of the infection and its likely spread. Further, it can predict the number of positive cases and death in any region. AI can help identify the most vulnerable regions, people and countries and take measures accordingly.

V)

Development of drugs and vaccines:

AI is used for drug research by analyzing the available data on COVID-19. It is useful for drug delivery design and development. This technology is used in speeding up drug testing in real-time, where standard testing takes plenty of time and hence helps to accelerate this process significantly, which may not be pos-

sible by a human. It can help to identify useful drugs for the treatment of COVID-19 patients. It has become a powerful tool for diagnostic test designs and vaccination development. AI helps in developing vaccines and treatments at much of faster rate than usual and is also helpful for clinical trials during the development of the vaccine.

VI)

Reducing the workload of healthcare workers

Due to a sudden and massive increase in the numbers of patients during COVID-19 pandemic, healthcare professionals have a very high workload. Here, AI is used to reduce the workload of healthcare workers. It helps in early diagnosis and providing treatment at an early stage using digital approaches and decision science, offers the best training to students and doctors regarding this new disease. AI can impact future patient care and address more potential challenges which reduce the workload of the doctors.

VII)

Prevention of the disease

With the help of real-time data analysis, AI can provide updated information which is helpful in the prevention of this disease. It can be used to predict the probable sites of infection, the influx of the virus, need for beds and healthcare professionals during this crisis. AI is helpful for the future virus and diseases prevention, with the help of previous mentored data over data prevalent at different time. It identifies traits, causes and reasons for the spread of infection. In future, this will become an important technology to fight against the other epidemics and pandemics. It can provide a preventive measure and fight against many other diseases. In future, AI will play a vital role in providing more predictive and preventive healthcare.

III LITERATURE SURVEY

es on management of covid-19 can be found in ^[27].

For working on this paper, many of the related works were studied.

RELATED WORKS:

To the date of writing this paper, a number of research have tried to perform a review modern approaches in tackling the pandemic. In this section, we perform an overview on the existing works in the area. In ^[8], a review on the role of IoT, Drones, AI, Block-chain, and 5G in managing the pandemic is performed. In ^[9], a review on the current automatic CT scan image processing approaches is performed. A review on the modeling techniques for predicting the pandemic including mathematical and AI approaches is performed in ^[10]. In another work ^[11], a review of modern approaches in tackling covid-19 is presented. Another review is performed in ^[12], where different areas in which AI has been used are discussed. A review on Deep Transfer Learning techniques in managing the pandemic is proposed in ^[13]. In ^[14], an overview of audio, signal and speech and language processing has been performed. A review of machine learning and AI algorithms for managing the pandemic is performed in ^[15]. In ^[16] the limitations, constraints and pitfalls for application of AI in battling the disease has been over-viewed. A survey on the state-of-the-arts of application of AI and big data for the pandemic is offered in ^[17]. In ^[18], ^[19], an early review on the application of AI in processing chest X-Ray images is presented.

A short review of AI application for covid-19 is presented in ^[20], ^[21]. In ^[22], ^[23], a review on the potential of using AI in developing countries is performed. A review on automatic detection and forecasting of covid-19 using DNN algorithms is performed in ^[24]. In ^[25], re survey on AI-based algorithms for combating the pandemic is performed. A review on machine learning algorithms in processing medical images regarding the disease can be found in ^[26]. A review on AI approach-

In [28], a review on data-driven methods for monitoring, modeling and forecasting the pandemic is presented. In [29], a survey on epidemic models for the disease is presented. A discussion on how big data can help better manage the pandemic is presented in [30].

In [31], a review on the data science approaches to combat the disease is presented. An overview of recent studies using machine learning in tackling the disease is presented in [32]. A review on the research on using machine learning algorithms in predicting the number of cases is presented in [33]. A review on the application of AI in discovering drugs can be found in [34]. A review is performed in [35] that covers the research on application of AI in managing critical covid-19 patients.

A review on the application of imaging characteristics and computing models applied to covid-19 related images is presented in [36]. In this work, CT positron emission tomography (PET/CT), lung ultrasound and magnetic resonance imaging (MRI) applied for detection, treatment and follow-up are studied.

In [37] some of many considerations for managing the development of AI applications including planning, unpredictable, unexpected or biased results, repurposing, the importance of data and diversity in AI team membership is addressed. The author provides implications for research and for practice according to each of the considerations. In [38], the role of AI for detection of the patients, finding the current pandemic pattern and possibility of future relapses are discussed. In [39], it is argued that there has been a great enthusiasm in diagnosing covid-19 AI approaches. So the authors examine 14 of the studies to discover the weakness of the solutions. The authors argue that “scientific community should be careful in interpreting statements, results and conclusions regarding AI use in imaging”. In [40],

five of the most important challenges in responding to covid-19 are presented and it is discussed how each of them can be managed via machine learning and artificial intelligence. In [41], overviews the challenges in fighting covid-19 and presents an

overview of ways in which machine learning can help in managing the disease. A review on potential technological strategies to control the pandemic is presented in [42]. In [43], a review on AI techniques in data acquisition, segmentation and diagnosis is presented. In another work [44], a review on machine and deep learning models for detecting and predicting the disease is presented. A review of Biological data mining and machine learning techniques in detecting and diagnosing the virus is presented in [45]. In [46], a review on AI approaches for covid-19 prognosis is presented.

IV COMPARATIVE STUDY OF TECHNOLOGIES:

Artificial Intelligence algorithms play a tremendous role in rapid detection, classification, identification, screening, and quantitation of patients with COVID-19 as shown in Table 2. These AI algorithms have been used in machine learning, deep learning and computer vision to discover insightful patterns in datasets. Javaid et al. (2020) stated that there are limited uses of AI technologies due to lack of data. Also, Wim (2020) further stated that AI has not been fully explored on tracking and prediction of COVID-19 cases in affected continents such as Europe, South and North America, and Africa. This might be attributed to the lack of a vast amount of historical data to train the AI models, which results in developing AI forecasting models that rely on noisy data and social media data. This severely affects the performance and accuracy of the forecasting model because of different data formats, lack of data standardization and interoperability, and missing values which is often inaccuracy and unreliable (Agbehadji, Bankole, Alfred, & Richard, 2020; Elliot, Fanwell, & Kinsley, 2018). The current literature, depicted in Table 2, shows that China is the leading pack in implementing AI technologies in fighting COVID-19 pandemic. Countries such as the United States of America (USA), South Africa, Brazil,

and India have recorded high COVID-19 cases of 5,595,835; 589,886; 3,343,925; 2,701,604, respectively as of August 17, 2020; have not completely and successfully implemented AI techniques in combating COVID-19 (Worldometers, 2020). These countries with high infection rate can utilize AI to detect, diagnose, identify and predict COVID-19 new cases. Majority countries diagnose COVID-19 using transcrip-tase-polymerase chain reaction (RT-PCR) test which takes up to 2 days to complete and there is currently a shortage of RT-PCR test kits (Xueyan et al., 2020). Health systems are overwhelmed with increasing demand for RT-PCR test kits which led some countries to focus only on contact tracing rather than testing the affected populace. Therefore, there is a need for AI models for early detection and diagnosis of COVID-19 using chest computed tomography (CT) images and can save radiologists' time. For example, Wang et al. (2020) developed a COVID-Net deep learning model (with 98.9% accuracy) to diagnose COVID-19 using chest CT images. Also, AI models can be used to develop COVID-19 vaccine development and drug discovery. For instance, Abhimanyu, Vineet, and Oge (2020) state that Flinders University applied AI-based program called Search Algorithm for Ligands (SAM) which generates trillions of synthetic compounds and determine the best trial candidates as vaccine adjuvants, thus reducing COVID-19 vaccine development process. This could benefit health policymakers, health care professionals to effectively allocate resources to high-risk zones and facilitate research (Raju et al., 2020). It is undoubtedly that AI technologies are conceivably reducing the burden of COVID-19; however, these technologies face the following challenges such as: (1) limited access to a large COVID-19 dataset for training and testing AI models; (2) The reliability and accuracy of AI models are also threatened with the availability of unstructured, noisy, and outlier COVID-19 data; (3) Failing to detect asymptomatic COVID-19 suspected individuals (Sera et al., 2020). In this paper, the hybrid approach of sentiment analysis is used which combines two approach-

ble on the web are also compared with the presented approach. This section shows the comparative details of lexicon-based approaches and other API-based service analysis. There APIs namely Test Analytics API by Microsoft, Alchemy language API and Aylien Text API were evaluated on the student feedback dataset.

Each of the Technologies are explained below:



V RESEARCH METHODOLOGY:-

We based our review on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [16], and extensively searched two databases (PubMed and EMBASE) for all English-language papers published from December 1, 2019, to March 31, 2020, using the search terms “novel coronavirus,” “2019 novel coronavirus,” “2019-nCoV,” “coronavirus disease 2019,” “COVID-19,” and “SARS-CoV-2.” The database search was supplemented by reference list checks. Papers reporting new data on AI applications for COVID-19 were included. Review papers and commentaries without new data were excluded.

The quality of included studies was assessed using a modified TRIPOD (transparent reporting of a multi-variable prediction model for individual prognosis or diagnosis) statement for adherence to reporting standards [17] and PROBAST (prediction model risk of bias assessment tool) for risk of bias [18]. A thematic analysis and narrative review of AI applications for COVID-19 was then conducted.

Accelerating research Open data projects and distributed computing to find AI-driven solutions to the pandemic, e.g. drug and vaccine development.	Detection	Early warning Detecting anomalies and digital “smoke signals”, e.g. BlueDot	Diagnosis Pattern recognition using medical imagery and symptom data, e.g. CT scans
	Prevention	Prediction Calculating a person’s probability of infection, e.g. EpiRisk	Surveillance To monitor and track contagion in real time, e.g. contact tracing
	Response	Delivery Drones for materials transport; robots for high-exposure tasks at hospitals, e.g. CRUIZ robot	Service automation Deploying triaging virtual assistants and chatbots, e.g. Canada’s COVID-19 chatbot
	Recovery	Monitor Track economic recovery through satellite, GPS and social media data, e.g. WeWork	

Before the world was even aware of the threat posed by the coronavirus (COVID-19), artificial intelligence (AI) systems had detected the outbreak of an unknown type of pneumonia in the People’s Republic of China (hereafter “China”). As the outbreak has now become a global pandemic, AI tools and technologies can be employed to support efforts of policy makers, the medical community, and society at large to manage every stage of the crisis and its aftermath: detection, prevention, response, recovery and to accelerate research .

VI RESULTS

Camus was prescient as the world has seen the ravages of the novel coronavirus and COVID-19 through most of this year with no immediate relief in sight. The Canadian company BlueDot was purported to be the first organization to disclose the possibility of an outbreak back in December by using machine learning and natural language processing. Artificial intelligence since then, however, has been somewhat under-leveraged as a valuable resource to mitigate the stress and burden of frontline healthcare workers in the tenacious fight against the virus ([1]), and it is important to under-

fective. First and foremost, it is essential to have a full understanding of this virus and the pandemic in the context of global health and epidemiological tenets. This foundation can then be applied to our appreciation for how artificial intelligence has helped to combat this virus currently and how it will be more impactful in the future.

VII CONCLUSIONS

Adopting a three-pronged approach based on testing, isolation and contact tracing is warranted to combat COVID19. It is necessary to exploit the available knowledge base to develop effective chemotherapeutic agents against COVID-19, taking cues from lessons learnt in the past during other such outbreaks. As there is no silver bullet available to cure the disease, we need to hasten progress on all fronts ranging from surveillance and monitoring to prevention and treatment. As this is the third outbreak of a coronavirus in recent times and many coronaviruses are circulating in animal reservoirs, we must focus on deciphering the molecular mechanism of SARS-CoV-2 and other coronaviruses and increasing our preparedness by capacity building for preventing future outbreaks [67]. As the current scenario warrants the need for immediate delivery of solutions, response to this outbreak was hugely augmented by various digital technologies and AI [68]. AI was found to be on par with and even more accurate than human experts in COVID-19 diagnosis and drug discovery. We need bigger datasets for training AI models and a legal framework and ethical considerations for sharing data before AI takes the forefront in diagnosis and other areas. Several bottlenecks in harnessing AI to its full potential in the current scenario are availability and sharing of clinical and epidemiological data, computational resources, scalability, privacy and ethical concerns.

Adopting a three-pronged approach based on testing, isolation and contact tracing is warranted to combat COVID19. It is necessary to exploit the available knowledge base to develop effective chemotherapeutic agents against COVID-19, taking cues from lessons

learnt in the past during other such outbreaks. As there is no silver bullet available to cure the disease, we need to hasten progress on all fronts ranging from surveillance and monitoring to prevention and treatment. As this is the third outbreak of a coronavirus in recent times and many coronaviruses are circulating in animal reservoirs, we must focus on deciphering the molecular mechanism of SARS-CoV-2 and other coronaviruses and increasing our preparedness by capacity building for preventing future outbreaks [67]. As the current scenario warrants the need for immediate delivery of solutions, response to this outbreak was hugely augmented by various digital technologies and AI [68]. AI was found to be on par with and even more accurate than human experts in COVID-19 diagnosis and drug discovery. We need bigger datasets for training AI models and a legal framework and ethical considerations for sharing data before AI takes the forefront in diagnosis and other areas. Several bottlenecks in harnessing AI to its full potential in the current scenario are availability and sharing of clinical and epidemiological data

VIII ACKNOWLEDGMENT

This research was supported by Pillai HOC College of Arts, Science & Commerce, Rasayani. It gives me great pleasure to present this research report on Role of artificial intelligence in treatin covid 19. It's a great pleasure and moment of immense satisfaction for me to express my profound gratitude to my Project Guide, Prof. Priyanka Sonawane.

REFERENCES

1. Cascella M., Rajnik M., Cuomo A., Dulebohn S.C., Di Napoli R. *Statpearls [internet]* StatPearls Publishing; 2020. Features, evaluation and treatment coronavirus (covid-19) [[Google Scholar](#)]
2. Razai M.S., Doerholt K., Ladhani S., Oakeshott P. Coronavirus disease 2019 (covid-19): a guide for UK GPs. *BMJ*. 2020;368 [[PubMed](#)] [[Google Scholar](#)]
3. Tayarani-N. M., Yao X., Xu H. Meta-heuristic algo-

Trans Evol Comput. 2015;19(5):609–629. [[Google Scholar](#)]

4. Browning L., Colling R., Rakha E., Rajpoot N., Rittscher J., James J.A. Digital pathology and artificial intelligence will be key to supporting clinical and academic cellular pathology through covid-19 and future crises: the pathlake consortium perspective. *J Clin Pathol.* 2020 [[PubMed](#)] [[Google Scholar](#)]

5. Ruiz Estrada M.A. The uses of drones in case of massive epidemics contagious diseases relief humanitarian aid: wuhan-covid-19 crisis. Available at SSRN 3546547. 2020 [[Google Scholar](#)]

6. Liang T. Handbook of covid-19 prevention and treatment. *The First Affiliated Hospital, Zhejiang University School of Medicine Compiled According to Clinical Experience.* 2020;68 [[Google Scholar](#)]

7. Wang C.J., Ng C.Y., Brook R.H. Response to covid-19 in taiwan: big data analytics, new technology, and proactive testing. *JAMA.* 2020;323(14):1341–1342. [[PubMed](#)] [[Google Scholar](#)]

8. Chamola V., Hassija V., Gupta V., Guizani M. A comprehensive review of the covid-19 pandemic and the role of IoT, drones, ai, blockchain, and 5g in managing its impact. *IEEE Access.* 2020;8:90225–90265. [[Google Scholar](#)]

ATTENDANCE MANAGEMENT USING FACE RECOGNITION.

JENCY KERLIN NADAR

ROLL NO 15

GUIDE : PRIYANKA SONAWANE

ABSTRACT—

In this digital era, face recognition system plays a vital role in almost every sector. Face recognition is one of the mostly used biometrics. It can be used for security, authentication, identification, and has got many more advantages. Despite of having low accuracy when compared to iris recognition and fingerprint recognition, it is being widely used due to its contactless and non-invasive process. Furthermore, face recognition system can also be used for attendance marking in schools, colleges, offices, etc. This system aims to build a class attendance system which uses the concept of face recognition as existing manual attendance system is time consuming and cumbersome to maintain. And there may be chances of proxy attendance. Thus, the need for this system increases. This system consists of four phases- database creation, face detection, face recognition, attendance updating. Database is created by the images of the students in class. Face detection and recognition is performed using Haar-Cascade classifier and Local Binary Pattern Histogram algorithm respectively. Faces are detected and recognized from live streaming video of the classroom. Attendance will be mailed to the respective faculty at the end of the session. Keywords—Face Recognition, Face Detection, Haar-Cascade classifier, Local Binary Pattern Histogram (LBPH), attendance system

◆

I INTRODUCTION

Attendance maintenance is a significant function in all the institutions to monitor the performance of the students. Every institute does this in its own way. Some of these institutes use the old paper or file-based systems and some have adopted strategies of automatic attendance using some biometric techniques. A facial recognition system is a computerized biometric software which is suited for determining or validating a person by performing comparison on patterns based on their facial appearances. Face recognition systems have upgraded appreciably in their management over the recent years and this technology is now vastly used for various objectives like security and in commercial operations. Face recognition is a powerful field of re-

search which is a computer based digital technology. Face recognition for the intent of marking attendance is a resourceful application of attendance system. It is widely used in security systems and it can be com-

pared with other biometrics such as fingerprint or eye iris recognition systems. As the number of students in an educational institute or employees at an organization increases, the needs for lecturers or to the organization also increase the complication of attendance control. This project may be helpful for the explanation of these types of problems. The number of students present in a lecture hall is observed, each person is identified and then the information about the number of students who are present is maintained.

II BACKGROUND

2.1 Counterpart Approach to Attendance and Feed-back System using Machine Learning Techniques:

In this paper, the idea of two technologies namely Student Attendance and Feedback system has been implemented with a machine learning approach.

This system automatically detects the student per-

formance and maintains the student's records like attendance and their feedback on the subjects like Science, English, etc. Therefore the attendance of the student can be made available by recognizing the face. On recognizing, the attendance details and details about the marks of the student is obtained as feedback.

2.2 Automated Attendance:-

System Using Face Recognition: Automated Attendance System using Face Recognition proposes that the system is based on face detection and recognition algorithms, which is used to automatically detects the student face when he/she enters the class and the system is capable to marks the attendance by recognizing him. Viola-Jones Algorithm has been used for face detection which detect human face using cascade classifier and PCA algorithm for feature selection and SVM for classification. When it is compared to traditional attendance marking this system saves the time and also helps to monitor the students.

2.3 Student Attendance:-

System Using Iris Detection: In this proposed system the student is requested to stand

of making attendance system easy and secure, in [6] the author proposed a system with the help of artificial neural networks, they used PCA to extract face images and testing and training were achieved by neural networks, their system performs in various orientation. A 3D face recognition approach for attendance management system was proposed by MuthuKalyani.K, VeeraMuthu.A has proposed, they marked attendance with monthly progress of each student. There is need for an alternative algorithm which can enhance the recognition on oriented faces. Efficient Attendance Management system is designed with the help of PCA algorithm the have achieved accuracy up to 83% but their system performance decreases due to slightly changes in light condition. An eigen face approach along with PCA algorithm for marking face recognition attendance system have introduced by author in , they mention comparison of different face recognition algorithm in their paper. Overall it was good approach to maintain record of attendance. he system to mark attendance for the student. Some algorithms like Gray Scale Conversion, Six Segment Rectangular Filter, Skin Pixel Detection is being used to detect the iris. It helps in preventing the proxy issues and it maintains the attendance of the student in an effective manner, but in one of the time-consuming process for a student or a staff to wait until the completion of the previous members.

2.4 Face Recognition-based Lecture Attendance System:

This paper proposes that the system takes the attendance automatically recognition obtained by continuous observation. Continuous observation helps in estimating and improving the performance of the attendance. To obtain the attendance, positions and face images of the students present in the class room are captured. Through continuous observation and recording the system estimates seating position and location of each student for attendance marking. The work is focused on the method to obtain the different weights of each focused seat according to its location. The effectiveness of the picture is also being discussed to enable the faster recognition of the image.

IV COMPARATIVE STUDY OF TECHNOLOGIES:

While humans can recognize faces without much effort, facial recognition is a challenging pattern recog-

III LITERATURE SURVEY

For working on this paper, many of the related works were studied.

RELATED WORKS:

In recent years, a number of face recognition based attendance management system have introduced in order to improve the performance of students in different organization. In Jomon Joseph, K. P. Zacharia proposed a system using image processing, PCA, Eigen faces, Microcontroller, based on Matlab. Their system works only with front face images and there is need of a suitable method which works with the orientation of the system. Ajinkya Patil with their fellows in proposed a face recognition approach for attendance marking using Viola jones algorithm, Haar cascades are used to detect faces in images and recognition performs through Eigen face method. Another approach

nition problem in computing. Facial recognition systems attempt to identify a human face, which is three-dimensional and changes in appearance with lighting and facial expression, based on its two-dimensional image. To accomplish this computational task, facial recognition systems perform four steps. First face detection is used to segment the face from the image background. In the second step the segmented face image is aligned to account for face pose, image size and photographic properties, such as illumination and grayscale. The purpose of the alignment process is to enable the accurate localization of facial features in the third step, the facial feature extraction. Features such as eyes, nose and mouth are pinpointed and measured in the image to represent the face. The so established feature vector of the face is then, in the fourth step, matched against a database of faces

Traditional



Some eigenfaces from AT&T Laboratories Cambridge

Some face recognition algorithms identify facial features by extracting landmarks, or features, from an image of the subject's face. For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. These features are then used to search for other images with matching features.

Human identification at a distance

To enable human identification at a distance (HID) low-resolution images of faces are enhanced using face hallucination. In CCTV imagery faces are often very small. But because facial recognition algorithms that identify and plot facial features require

high resolution images, resolution enhancement techniques have been developed to enable facial recognition systems to work with imagery that has been captured in environments with a high signal-to-noise ratio.

3-dimensional recognition

3D model of a human face.

Three-dimensional face recognition technique uses 3D sensors to capture information about the shape of a face. This information is then used to identify distinctive features on the surface of a face, such as the contour of the eye sockets, nose, and chin. One advantage of 3D face recognition is that it is not affected by changes in lighting like other techniques. It can also identify a face from a range of viewing angles, including a profile view

Thermal cameras

A pseudocolor image of two people taken in long-wavelength infrared (body-temperature thermal) light.

A different form of taking input data for face recognition is by using thermal cameras, by this procedure the cameras will only detect the shape of the head and it will ignore the subject accessories such as glasses, hats, or makeup. Unlike conventional cameras, thermal cameras can capture facial imagery even in low-light and nighttime conditions without using a flash and exposing the position of the camera. However, the databases for face recognition are limited. Efforts to build databases of thermal face images date back to 2004. By 2016 several databases existed, including the IIITD-PSE and the Notre Dame thermal face database. Current thermal face recognition systems are not able to reliably detect a face in a thermal image that has been taken of an outdoor environment.

V RESEARCH METHODOLOGY

Image acquisition: Image is acquire using a high definition :- camera which is placed in the classroom.

This image is given as an input to the system. Dataset Creation:

Dataset of students is created

before the recognition process. Dataset was created only to train this system. We have created a dataset of 5 students which involves their name, roll number, department and images of student in different poses and variations. For better accuracy minimum 15 images of each students should be captured. Whenever we register student's data and images in our system to create dataset,

deep learning applies to each face to compute 128-d facial features and store in student face data file to recall that

face in recognition process. This process is applies to each image taken during registration.

Storing: We have used JSON to store the student's data. JavaScript Object Notation (JSON): To represent a structured data based on JavaScript object syntax, a standard text based format is introduced. JSON is used for transmitting data in web application. It is a perfect solution for storing temporary data that's consumed by the entity that's creates the data. JSON can store data in String, Number, Object, Array, Boolean, Null form which means it has limitation of storing data in functions, dates Image acquisition: Image is acquire using a high definition camera which is placed in the classroom. This image is given as an input to the system.

Dataset Creation: Dataset of students is created before the recognition process. Dataset was created only to train this system. We have created a dataset of 5 students which involves their name, roll number, department and images of student in different poses and variations. For better accuracy minimum 15 images of each students should be captured. Whenever we register

student's data and images in our system to create dataset,deep learning applies to each face to compute 128-d facial

features and store in student face data file to recall that face in recognition process. This process is applies to each

image taken during registration.

Storing: We have used JSON to store the student's data. JavaScript Object Notation (JSON): To represent a structured data based on JavaScript object syntax, a standard text based format is introduced.

JSON is used for transmitting data in web application. It is a perfect

solution for storing temporary data that's consumed by The entity that's creates the data. JSON can store data in

String, Number, Object, Array, Boolean, Null form which means it has limitation of storing data in functions, dates

Image acquisition: Image is acquire using a high definition camera which is placed in the classroom. This image is given as an input to the system.

Dataset Creation: Dataset of students is created before the recognition process. Dataset was created only to train this system. We have created a dataset of 5 students which involves their name, roll number, department and images of student in different poses and variations. For better accuracy minimum 15 images of each students should be captured. Whenever we register

student's data and images in our system to create dataset, deep learning applies to each face to compute 128-d facial

features and store in student face data file to recall that face in recognition process. This process is applies to each

image taken during registration.

Storing: We have used JSON to store the student's data.

JavaScript Object Notation (JSON): To represent a structured data based on JavaScript object syntax, a standard text based format is introduced. JSON is used for transmitting data in web application. It is a perfect solution for storing temporary data that's consumed by the entity that's creates the data. JSON can store data in String, Number, Object, Array, Boolean, Null form which means it has limitation of storing data in functions, dates Image acquisition: Image is acquire using a high definition camera which is placed in the classroom. This image is given as an input to the system.

Dataset Creation: Dataset of students is created before the recognition process. Dataset was created only to train this system. We have created a dataset of 5 students which involves their name, roll number,

department and images of student in different poses and variations. For better accuracy minimum 15 images of each students should be captured. Whenever we register

student's data and images in our system to create dataset, deep learning applies to each face to compute 128-d facial

features and store in student face data file to recall that face in recognition process. This process is applies to each image taken during registration.

Storing:

We have used JSON to store the student's data. JavaScript Object Notation (JSON): To represent a structured data based on JavaScript object syntax, a

standard text based format is introduced. JSON is used for transmitting data in web application. It is a perfect

solution for storing temporary data that's consumed by the entity that's creates the data. JSON can store data in

String, Number, Object, Array, Boolean, Null form which means it has limitation of storing data in functions, dates The proposed system is designed for automating the attendance of the different organization and reduces the flaws of existing manual system. The

system calculate the attendance subject wise, that is the data of students and subjects are added manually by administrator, and whenever time for corresponding subject arrives the system automatically starts taking snaps and find whether human faces are appear in the given image or not. We have used Histogram of

Oriented Gradient for face detection and deep learning techniques to calculate and compare 128-d face features for face recognition. Once faces are detected and recognize with the existing database, system calculate attendance for the recognize students with the respective subject id in real time. And an excel sheet generated and saved by the system automatically. Our system splits into two parts, First the front end side which consist of GUI which is based on Electron JS that is

JavaScript stack which is serving as a client and the second is the backend side which consist of logic and based on Python which is serving as a server. And we know that both the languages cannot communicate with each other directly so we have used IPC (Inter Personal Communication) techniques with zero library as a bridge to communicate these two languages. The

Electron JS call the python functions and interchange

data via TCP with help of Zero PC Library. Data Acquisition Image acquisition: Image is acquire using a high definition camera which is placed in the class-room. This image is given as an input to the system.

Dataset Creation: Dataset of students is created before the recognition process. Dataset was created only to train this system. We have created a dataset of 5 students which involves their name, roll number, department and images of student in different poses and variations. For better accuracy minimum 15 images of

each students should be captured. Whenever we register student's data and images in our system to create dataset, deep learning applies to each face to compute

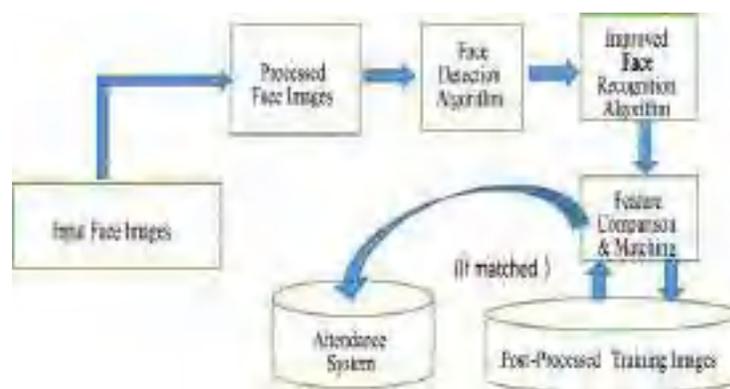
128-d facial features and store in student face data file to recall that face in recognition process. This process is applies to each image taken during registration.

Storing: We have used JSON to store the student's data. JavaScript Object Notation (JSON): To represent a structured data based on JavaScript object syntax, a standard text based format is introduced. JSON is used for transmitting data in web application. It is a perfect

solution for storing temporary data that's consumed by the entity that's creates the data. JSON can store data in String, Number, Object, Array, Boolean, Null form which means it has limitation of storing data in functions, dates

VI RESULTS

The main working principle of the project is that, the video captured data is converted into image to detect and recognize it. Further the recognized image of the student is provided with attendance, else the system marks the database as absent.



VII CONCLUSIONS

Smart attendance management system is designed to solve the issues of existing manual systems. We have used face recognition concept to mark the attendance of student and make the system better. The system performs satisfactory in different poses and variations. In future this system need be improved because these system sometimes fails to recognize students from some distance, also we have some processing limitation, working with a system of high processing may result even better performance of this system.

VIII ACKNOWLEDGMENT

This research was supported by Pillai HOC College of Arts, Science & Commerce, Rasayani. It gives me great pleasure to present this research report on "Attendance management using face recognition". It's a great pleasure and moment of immense satisfaction for me to express my profound gratitude to my Project Guide, Prof. Priyanka Sonawane.

REFERENCES

- [1] Kar, Nirmalya, et al. "Study of implementing automated attendance system using face recognition technique." *International Journal of computer and communication engineering* 1.2 (2012): 100. [2] RoshanTharanga, J. G., et al. "Smart attendance using real time face recognition (smart-fr)." *Department of Electronic and Computer Engineering, Sri Lanka Institute of Information Technology (SLIIT), Malabe, Sri Lanka* (2013) [3] Selvi, K. Senthamil, P. Chitrakala, and A. Antony Jenitha. "Face recognition based attendance marking system." *Corresponding Author: S. Rajkumar*, Email: rajkumarsrajkumar@ gmail. com* (2014). [4] Joseph, Jomon, and K. P. Zacharia. "Automatic attendance management system using face recognition." *International Journal of Science and Research (IJSR)* 2.11 (2013): 327- 330. [5] Patil, Ajinkya, and Mrudang Shukla. "Implementation of classroom attendance system based on face recognition in class." *International Journal of Advances in Engineering & Technology* 7.3 (2014): 974. [6] Kanti, Jyotshana, and Shubha Sharm. "Automated Attendance using Face Recognition based on PCA with Artificial Neural Network." *International journal of science and research IJSR*(2012). [7] MuthuKalyani, K., and A. VeeraMuthu. "Smart application for AMS using face recognition." *Computer Science & Engineering* 3.5 (2013): 13. [8] Deshmukh, Badal J., and Sudhir M. Kharad. "Efficient Attendance Management: A Face Recognition Approach." (2014). [9] Wagh, Priyanka, et al. "Attendance system based on face recognition using eigen face and PCA algorithms." *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*. IEEE, 2015. [10] Bhattacharya, Shubhobrata, et al. "Smart Attendance Monitoring System (SAMS): A Face Recognition Based Attendance System for Classroom Environment." *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 2018. [11] Samet, Refik, and Muhammed Tanriverdi. "Face recognition-based mobile automatic classroom attendance management system." *2017 International Conference on Cyberworlds (CW)*. IEEE, 2017. [12] Li, Xiang-Yu, and Zhen-Xian Lin. "Face recognition based on HOG and fast PCA algorithm." *The Euro-China Conference on Intelligent Data Analysis and Applications*. Springer, Cham, 2017. [13] Arsenovic, Marko, et al. "FaceTime—Deep learning based face recognition attendance system." *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, 2017. [14] Rekha, N., and M. Z. Kurian. "Face detection in real time based on HOG." *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 3.4 (2014): 1345-1352. [15] Kwolek, Bogdan. "Face detection using convolutional neural networks and Gabor filters." *International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2005. [16] Ashwini, C., et al. "An Efficient Attendance System Using Local Binary Pattern and Local Directional Pattern." *Journal of Network Communications and Emerging Technologies (JNCET)* www.

jncet. org 8.4 (2018). [17] Karnalim, Oscar, et al. "Face-face at classroom environment: Dataset and exploration." 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2018. [18] Mian, Ajmal. "Realtime face detection and tracking using a single pan, tilt, zoom camera." 2008 23rd International Conference Image and Vision Computing New Zealand. IEEE, 2008. [19] Mehta, Preeti, and Pankaj Tomar. "An Efficient Attendance Management Sytem based on Face Recognition using Matlab and Raspberry Pi 2." International Journal of Engineering Technology Science and Research IJETSR 3.5 (2016): 71-78. Smart Attendance Management System Using Face Recognition 5

STUDENT SENTIMENT ANALYSIS

Vaishnavi Nair, Prof. Priyanka Sonawane

ABSTRACT— In an Educational institution, overall development of the student should be a primary concern. When it comes to the overall Development, Academic achievements and Curricular Activities are not the only parameters focused. Along with them, we need to analyze the Positive Psychology and Mental well-being of the Student. Moreover, nowadays it has become important for the Teachers teaching in an educational institution to analyze the Sentiments or Emotional Quotient of the students studying in the institution. So, in this paper, an attempt is made to analyze the sentiments of the students and give a result about their attitude towards a situation. In Student Sentiment Analysis, the identification of the sentiments or emotions behind the textual feedback given will be done in the terms of Positivity and Negativity. For this, the Hybrid approach is proposed, which is the combination of Sentiment Lexicons and Machine-learning methods, to determine the Semantic Orientation behind the textual feedback. In short, the paper aims to present a new perspective for studying sentiment analysis of the students and the nature of language used.

Keywords — Sentiment Analysis, Emotional Quotient, Hybrid approach, Sentiment Lexicons, Semantic Orientation.

◆

I INTRODUCTION

Sentiment analysis means to analyze the emotion or attitude of a person. It is also called as opinion of a person. It is also called as opining mining or emotional AI. The objective of the sentiment analysis is to determine the attitude of a speaker, writer or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event.

In recent years, there has been an increase in popularity of data science and analytics. In addition, the appearance of Internet everywhere offers the opportunity to access huge volumes of data, most often in written text form. For example, on social networks, many users could post comments on different educational services, or fill in questionnaires that measure their satisfaction, items that would require free answers, etc. All this data is subject to interpretation.

Sentiment analysis is a relative new research area. It is also known as opinion mining. It involves application of data mining, NLP and text analysis techniques in order to identify and classify subjective opinions from different materials such as written documents.

In this paper, we will analyze the student's attitude or response towards some Questions based on Emotional Quotient.

II BACKGROUND

In the past few years, Sentiment Analysis has been extensively studied. There are three main approaches towards the reported work:

- A. Machine Learning-based
- B. Lexicon-based Method
- C. Hybrid

A. Machine-Learning based:

In machine learning-Based approach (ML), several machine learning algorithms are used. Here, two things are needed to be done: - learn a predictive model using the provided training dataset and Evaluate the performance of the learned

model on the test dataset. There are two classifications in ML-based approach.

1. Supervised learning methods.
2. Unsupervised learning methods.

1. Supervised machine learning approach:

Here, Training classifiers are involved, and linguistic features are extracted from the text. For this, we need a labeled dataset of test documents to train classifiers. Linguistic features that have been widely used for sentiment analysis includes n-grams, word representation, part of speech (POS) tags, punctuations and emotions.

A supervised learning approach to predict sentiments from student feedback. The reported models were trained using n-gram features from the feedback text. Naïve Bayes, Maximum entropy and support vector machine (SVM) algorithms are used to train models.

2. Unsupervised learning methods:

Here, dataset is not required to be annotated with true sentiment labels. An unsupervised approach of sentiment analysis was proposed where Aggregation of the polarities of phrases containing adjectives and adverbs determined the polarity of the textual content. Polarity of a phrase was identified using the employment of point wise mutual information (PMI) based method. Then, a method of computing semantic orientation of unstructured text based on dependency parsing technique was presented. In that proposal, sentiment lexicons were used which were created using semi-automatic polarity expansion algorithm.

B. Lexicon-Based approach:

In Lexicon-Based approach of sentiment analysis sentiment lexicon is used to determine the polarity of a given textual content. A lexicon can also be called as a dictionary. It shows a list of words with related sentiment polarity. Manual and automatic creation of lexicon is possible. Another lexicon-based approach proposes that determines the polarity of a word by using the dic-

tionaries construct. Several general purpose and domain-specific lexicons have been constructed such as MPQA subjectivity lexicon, Harvard General Inquires, Linguistic Inquiry and Word counts Database and many others. One limitation of lexicon-based approach is that the contextual and domain-specific semantic orientation of a word is generally ignored.

The use of lexicon was to determine the sentiment polarity of the feedback given by a student. The general-purpose sentiment dictionary was modified by them so that the polarity of an opinion can be determined in the context of the academic domain. They concluded that the domain-specific sentiment lexicon was better in comparison to any general-purpose sentiment lexicon.

C. Hybrid approach:

Hybrid is the combination of sentiment lexicon and machine learning methods. An example is the hybrid approach for sentiment analysis of twitter data. For labelling training dataset with sentiment polarities, an opinion lexicon was used. To train a binary classifier to predict sentiment polarity on the evaluation dataset, they used labelled dataset. Sentiment analysis is performed at the sentence level using a hybrid approach. The basis of their approach was sentiment lexicon extended using SentiWordNet and fuzzy sets and sentiment polarity of a sentence was determined

This paper also proposes Hybrid approach that combines the use of sentiment dictionary and machine learning methods to determine the semantic orientation of a textual feed provided by students.

III LITERATURE SURVEY

For working on this paper, many of the related works were studied.

RELATED WORKS:

According to Gareth Huges, Mehr Panjwani, Priya Tulcidas and Dr. Nicola Byrom, Participants reported large numbers of students experiencing mental health difficulties [1]. Utilization of both text mining and qualitative coding was included in a computer-assisted literature review presented by

Mika V. Mäntylä, Daniel Graziotin, Miikka Kuutila [2], with an analysis of 6,996 papers from Scopus. They found that the roots of sentiment analysis were in the research on public opinion analysis at the beginning of 20th century and. In the text subjectivity analysis was performed by the computational linguistics community in 1990's. However, the availability of subjective texts on the Web was the only reason which resulted into the outbreak of computer-based sentiment analysis.

Francis F. Balahadia; Ma. Corazon G. Fernando; Irish C. Juanatas [3], developed the teacher's performance evaluation tool using opinion mining with sentiment analysis. They collected the feedback from the students and identified the strength and weakness of the teacher. Evaluation of the qualitative and quantitative data was done by them and they provided sentiment score of the teacher in a school. Also, a study on the combination of machine learning and lexi-con-based approaches for sentiment analysis by collecting feedback from students was done by Zarmeen Nasim, Quratulain Rajput, Sajjad Haider [4]. The textual feedback, typically collected towards the end of a semester, provides useful insights into the overall teaching quality and suggests valuable ways for improving teaching methodology.

Saida Ulfa, Rex Bringula, Citra Kurniawan, Muhibuddin Fadhli [5] explored the implementation of sentiment analysis techniques on student feedback in an online learning environment.

D Selvapandian, Thamba Meshach W., K.S.Suresh Babu, R. Dhanapal, Jebakumar Immanuel D. [6] examined the feedback concept, where opinion examination helps to distinguish how students are communicated in writings and whether the articulations demonstrate positive (ideal) or negative (troublesome) and conclusions toward the subject. In this research work efficient fusion based neural network (EF-NN) classifier was introduced to predict the frequent context patterns used in the student feedback dataset.

An analysis of those feedbacks were done by using machine learning techniques like Support Vector Machines (SVM), Maximum Entropy (ME), Naive Bayes (NB), and Complement Naive Bayes and Complement Naive Bayes (CNB) and applying neutral class by Mohammad Aman Ullah [7]. And, found SVM as the highest performer with an accuracy of 97% by applying different preprocessing

and feature extraction techniques and avoiding neutral class, which outperform state-of-art work by 2%.

IV COMPARATIVE STUDY OF TECHNOLOGIES:

In this paper, the hybrid approach of sentiment analysis is used which combines two approaches. However, the other sentiment analysis tools available on the web are also compared with the presented approach. This section shows the comparative details of lexicon-based approaches and other API-based service analysis. There APIs namely Test Analytics API by Microsoft, Alchemy language API and Aylien Text API were evaluated on the student feedback dataset.

Each of the Technologies are explained below:

2.1 Text-Analysis API:

In this, a numeric-sentiment score between 0-1 is returned. Score near to 1 means that the textual feedback is positive whereas the score near to 0 means the textual feedback is negative. For evaluating the present dataset, a trial version of API was used. API was provided with the student's feedback. A numeric score was returned instead of discrete sentiment labels. Hence, for comparison with discrete sentiment labels, a decision tree classification algorithm was used.

2.2 Alchemy Language API:

Here discrete sentiment labels are returned as {positive, negative, neutral}. Here also, free subscription plan of API was used. The student's comments were passed to API. The result set were compared to manually annotated sentiment labels.

2.3 Aylin Text API:

In this, text input is required. It is analysed and return the polarity labels as {positive, negative, neutral}. Here also for evaluation purpose, free subscription plan of API is used. Student feedback were passed to API and returned labels were compared with manually annotated sentiment labels.

2.4 Lexicon-based approach:

Here, sentiment lexicons are used to for evaluating sentiment scores of each comment. Sentiment scorers were assigned with sentiment labels. Positive was the feedback if the score was greater than 0. Neutral was the feedback if the score of the comments was 0, and the feedback was negative if the score is negative.

2.5 Proposed Hybrid approach:

By comparing the above discussed approaches and the proposed approach. the result got was hybrid approach proved to be better than any other approach. A comparative analysis is shown in the below Table I . The presented hybrid approach had given better results than Aylien Text API, Alchemy language API and Text Analytics API. Domain-specific features were included in this approach. This was better for the academic domain than the general-domain models used by the three APIs 2% improvement were seen in the results of presented approach as compared with Lexicon-based approach

2.6 Comparative analysis between proposed approach and other sentiment analysis approaches.

	Accuracy	F-measure
Aylien Test API	0.67	0.76
Alchemy Language API	0.49	0.57
Test Analytics API	0.79	0.74
Lexicon-based approach	0.91	0.90
Proposed Hybrid approach (TF-IDF with domain specific lexicon)	0.93	0.92

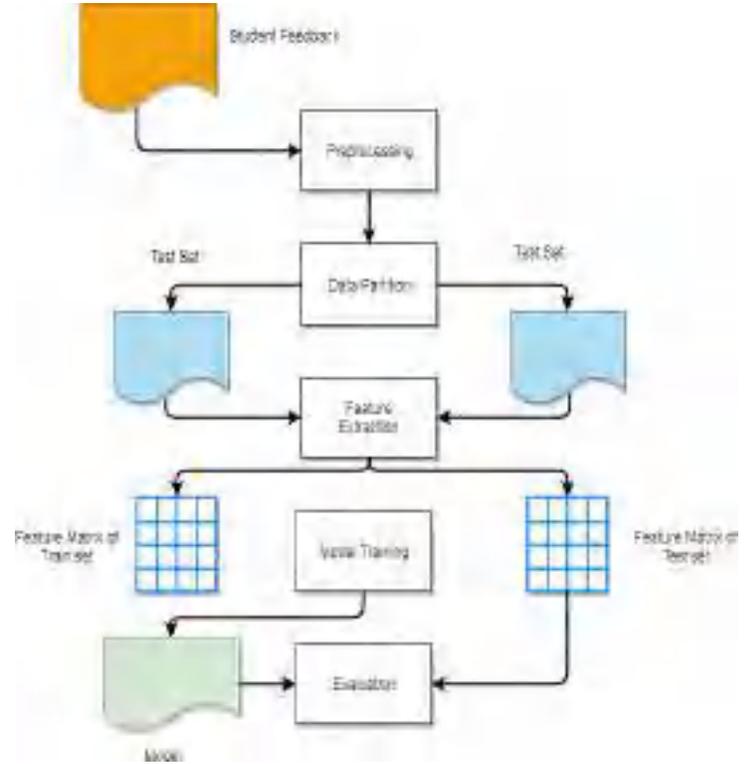
Table 1.1. Comparative analysis between proposed approach and other sentiment analysis approaches.

V RESEARCH METHODOLOGY

PROPOSED RESEARCH APPROACH

Implementation Approaches:

To implement the concept, the methodology given in the fig. is used. The methodology classifies the Sentiment Polarity as Positive, Negative and Neutral.



Fig, 5.1. Methodology of Student Sentiment Analysis System

The process workflow is further divided into following sub-sections:

1. **Dataset Description:** Here, a Dataset used in this system comprises of 2000+ positive words and 4000+ negative words. This was classified and labelled as Positive and Negative.
2. **Pre-processing:** Pre-processing include the following:
 - i. **Punctuations:** Removal of Punctuations, numbers and other special characters as these characters do not carry useful information related to sentiment analysis.
 - ii. **Tokenization:** Tokenization is the process of splitting text stream into a list of words.

- iii. **Case Conversion:** After tokenization, words were transformed into lower case.
- iv. **Stop words:** In natural language processing, words that are frequently used such as helping verbs, prepositions, articles are termed as stop-words. Stop-words generally do not provide any useful information and therefore were removed from the feedback text.

- 3. **Data Partition:** For training and evaluation purposes, the manually labelled dataset of students' feedback was randomly split into train set and test set. 70% of the dataset was used for training and the remaining dataset was used for the evaluation purpose.
- 4. **Feature Extraction:** After data splitting, feature extraction was applied on both training and testing datasets. During the feature extraction stage, the pre-processed text was converted into a numerical feature vector.
- 5. **Model Training:** After the extraction of features from the train and test dataset, learning algorithms were applied for training model.

VI RESULTS

Here, an attempt was made to create a web application which will be used to analysis the Sentiments of a student with the help of few Emotional-Quotient questions. Sentiment Analysis is also called as Opinion Mining. This refers to the techniques and processes that help to retrieve information about how a person react to a instance.

This system is significant for all the educational institutions where the overall development of the personality of a student is the primary concern. Here, the teachers can analyse the overall Sentiment of a Student and can help them to inculcate positive attitude towards the life along with their academic progress.

In educational institutions, Positive Psychology of students is very important. There we can use the Student Sentiment Analysis System where the attitude of the students is analysed.

VII CONCLUSIONS

This Student Sentiment Analysis system has made a way to analyse the sentiments of the students at any educational institution. Here, Hybrid approach was used for performing sentiment analysis on student feedbacks. Employment of Machine learning methods along with the sentiment lexicons was described in this presented approach. Investigation of other APIs available for Sentiment Analysis were also done and the results were compared with the presented Hybrid approach. The proposed approach, where the Sentiment Lexicons were combined with the use of machine learning techniques, was capable of predict the Sentiment of the Textual content even in the absence of opinion words in the Lexicon. Semantic Orientation behind the textual feedback is determined.

In this paper, an attempt was made to the analyse Sentiments of a student with the help of few Emotional-Quotient questions. Sentiment Analysis is also called as Opinion Mining. This refers to the techniques and processes that help to retrieve information about how a person react to a instance. This system is significant for all the educational institutions where the overall development of the personality of a student is the primary concern. Here, the teachers can analyse the overall Sentiment of a Student and can help them to inculcate positive attitude towards the life along with their academic progress. In educational institutions, Positive Psychology of students is very important. There we can use the Student Sentiment Analysis System where the attitude of the students is analysed.

VIII ACKNOWLEDGMENT

This research was supported by Pillai HOC College of Arts, Science & Commerce, Rasayani. It gives me great pleasure to present this research report on "Student Sentiment Analysis". It's a great pleasure and moment of immense satisfaction for me to express my profound gratitude to my Project Guide, Prof. Priyanka Sonawane.

REFERENCES

1. Student Mental Health: The Role and Experiences of Academics by Gareth Huges, Mehr Panjwani, Priya Tulcidas and Dr. Nicola Byrom._
<http://hdl.handle.net/10545/622114>
2. The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers by Mika V. Mäntylä, Daniel Graziotin, Miikka Kuutila, M3S, ITEE, University of Oulu -
https://www.researchgate.net/publication/311458740_The_Evolution_of_Sentiment_Analysis
3. Teacher's performance evaluation tool using opinion mining with sentiment analysis_
<https://ieeexplore.ieee.org/document/7519384>
4. Sentiment Analysis of Student Feedback Using Machine Learning and Lexicon Based Approaches by Zarmeen Nasim, Quratulain Rajput and Sajjad Haider, Faculty of Computer Science_
<https://ieeexplore.ieee.org/document/8002475>
5. Student Feedback on Online Learning by Using Sentiment Analysis: A Literature Review -
<https://ieeexplore.ieee.org/document/9276578/>
6. An Efficient Sentiment Analysis on Feedback Assessment from Student to Provide Better Education -
<https://ieeexplore.ieee.org/document/9243594/>
7. Sentiment analysis of students feedback: A study towards optimal tools -
<https://ieeexplore.ieee.org/document/7860361/>

CYBERSECURITY USING ARTIFICIAL INTELLIGENCE

Vaishnavi Nair, Prof. Priyanka Sonawane

ABSTRACT – Artificial intelligence is ability of a computer to perform tasks with human intelligence. A intelligence works by learning iteratively again and again, like humans. With the emergence of technologies, threats are also increasing. Financial rewards are focused on by the cyber attackers. But the department has found a new way to prevent attacks before they occur, as the old traditional way can no longer be relied upon. This paper presents a study on the requirement for the evolution of skills of cybersecurity and how artificial intelligence will help to improve skills using machine learning algorithms and artificial neural networks.

Keywords – Artificial Intelligence, Cybercrime, Cyber- attacks, Cyber Security.

◆

I INTRODUCTION

Artificial Intelligence (AI) is now emerging in the field of cybersecurity industry. AI refers to a few techniques that will support security. Machine learning algorithms are used to identify the threats and respond to it. These steps can predict the malicious threats as they are coming. These attacks are not new in today's life as we can see the cyber-attacks in some places frequently. Hence, a plan is being proposed where we can create a database where all the information is collected related to the threats and stored to machine algorithms giving a way to prevent the threat before it occurs.

Moving into classifications, first classification will look at the characteristics of the data and decide whether it is malicious or not. Second technique includes the detection of anomalies while the data processing fast in real time. Here, the intrusion is detected by the deviation from the standard of behavior which are tracked by an algorithm and considered as normal. The last technique used in this field is probabilistic programming, where a set of computer languages is used to distribute probabilities which track the history of the malware and informs future responses.

A large data can be quickly processed using Neural Networks and Deep Learning Algorithms. These techniques recognize the features of the data on the internet and use them to determine that the data is malicious or not. Data can be of any kind, like

images or texts. Learning from the past data will help in reducing the mistakes. The result cannot be concrete, whereas it is complex for the interpretation by humans. This is leading to the decline of traditional solution for security the and helping in the emergence of AI in Cybersecurity which increases speed and scalability.

II BACKGROUND

A. ARTIFICIAL INTELLIGENCE FOR CYBERSECURITY

Information Security is the primary concern in Information & Communication Technology field. Many studies were attempted to address this issue. Some of the techniques include malware detectors, intrusion detection and prevention systems (IDPS), sophisticated firewall setups and data encryption algorithms. There were some debates addressing that monitoring human behavior is sufficient or not sufficient for Information Security. A balance between humans, technology and policy management is necessary. Cyber Security uses algorithms and physical devices like sensors and actuators.

In the first generation of antivirus systems, viruses were identified by scanning the bit signature. Here, it was

assumed that all the viruses have same pattern and structure every time. Hence, there was a fixed signature and algorithm. Sometimes, the signature catalog was updated whenever the internet was connected to the device. But the approach was ineffective due to the release of the vast malware. Then came the approach of signature-less techniques. Here, behavioral detections and AIs were used to detect and mitigate malware attacks. AI applications had enhanced later giving possibility to design a system that automatically identify and previous malicious activities. Prevention of cyber-attacks were achieved by supporting existing technological methods.

Identification of the most efficient technique and its impact were always a matter of research. Some researchers say that AI has improved Organization's Information Security. Selection of existing methods or surveyed a sample of systems and assess their performance in comparison to theirs was demonstrated by Most existing studies. The chances of level of selection is high. But, an aggregated literature that provide summaries on issues, is needed to understand the challenges and future research directions within the domain.

III RELATED WORKS

Many works were done in these field. Some of them are mentioned here. A system was proposed by Harini and Dharani [1] which uses the techniques Neural Nets, Expert Systems, and Intelligent Agents. Two parts are included in these Expert Systems: Knowledge and engine that uses inference. The technique of Neural Nets works like neurons in the brain. An identification of malicious codes is done in this technique where these code causes malware installation in the user's system. Sensors and Actuators are used by Intelligent Agents which stops DDOS attack.

Another system was proposed by Swapnil Ramesh Kumbar [2] which uses fuzzy system techniques, pattern recognition, image processing, and data mining techniques. Data mining prevents Phishing and Fake Auctioning. Fingerprinting, facial recognition, voice recognition is used in Pattern Recognition as security. In the field of defense and military, image processing techniques come into use. Fuzzy logic is prominently used at the time of

malware invasion.

Mohana K.V et al. [3] have given a proposal of a system that gives Data security with the Chaotic Neural Network and genetic algorithms. Generic Algorithm generates random numbers and passed as a parameter used in the processing of the neural network. An encryption key is generated after processing the neural network. For decryption, the above process is reversed. Decryption is not easy.

Artificial Immune Systems, Ontology, General intelligence are used by the system which was proposed by Selma Dilek et al. [4]. For encountering spam, the air-based email system was used. Strong AIs are used to represent knowledge and General Intelligence in the Ontology technique.

In the system by Alberto Perez Veiga [5], Machine Learning technology is used. Small parts of the complex problems are solved by Machine learning.

Ranjeev Mittu et al. have proposed a system that is used in the detection of advanced persistent threats. The Cyber Threats are addressed by the system addresses.

Enn Tygu has proposed a system in which Applications based on neural networks are used in cyber defense. This application can be used as a defense method where the neural network is not efficient.

Detection of disruption and intelligence agents are used in the system proposed by L.N Wijesinghe et al. Detection of cybercrimes is done by IDS Technology. Cyber-attacks and alert users are done by Intrusion Prevention Systems.

Arockia Panimalar.S et al. have proposed a that detects DDOS attacks.

Anna L. Buczak has proposed a system that provides a review of the DM and ML methods used for cybersecurity. However, there are few problems that make DM and ML hard to use.

IV COMPARISON OF DIFFERENT CYBERSECURITY IN AI

Table. 1 Comparison of different techniques used in cybersecurity for AI

SL. NO	Approach	Methods	Result	Advantages
1.	Survey and analyzing different methods for improving cybersecurity with the use of AI	Expert Systems, Neural Networks Intelligent Agents	Improving security with the use of Artificial Intelligence	Increase accuracy and efficiency of Intrusion Detection
2.	Overview of different techniques in AI that can be implemented in cybersecurity	Artificial neural networks, Image processing, data mining, fuzzy system, expert system, pattern recognition	Increasing the performance of security systems using different AI techniques	Can adapt to different situations, help with decision making, accurate and quick
3	Using a genetic algorithm for optimization problems	An artificial neural network, Chaotic neural networks, encryption, decryption, security.	Encryption and Decryption of original data by using chaotic neural networks and genetic algorithms	Ensures optimal security when compared to open channels
4	Study and demonstrate different AI methods for combating cyber	The artificial immune system, Intelligent Agents, Fuzzy sets, Genetic	Techniques used in applications to combat cybercrime	Non- linearity, mobility, Dynamic structure, resilience, versatility
5	Provides usage of AI in network security and informs about the evolution of cybersecurity	Artificial Intelligence with Machine learning	Usage of ML in cybersecurity provides reactive real-time security	Real-time effective protection against cybercrime
6.	Engineering approaches to mitigate cyber threats using AI	Artificial Intelligence	Advance persistence threats are difficult to detect so using of AI becomes necessary	AI application are present which can detect and continue future threats
7.	Provides a survey of Artificial Intelligence methods used in cyber defense	Neural nets, Expert system, Intelligent agents	Application based on neural networks are used in cyber defense	Development in AI applications will help decrease cyber threats in the world
8.	Introducing in an advance cybercrime defense system involving Artificial Intelligence agents	Intrusion detection and prevention system, Artificial agent, Intelligence agent	Provides information about the cybercrimes and advances made using AI with a security system	Tracing the attacker and responding to the source can be done in the most efficient way using Artificial Intelligence
9.	Usage of Artificial Intelligence implemented in cybersecurity	Expert systems, Neural nets, Intelligent agents	Artificial intelligence systems with human assistance in the form of training to provide optimal performance in detecting and controlling cyber attacks	Better decision support, better DOS detection, and good mobility
10.	Machine learning in data mining methods for better security	Cyber analytics, machine learning, data mining	Provides a brief survey of usage of machine learning and data mining techniques in cybersecurity.	Algorithms for improving performance, complexity, and accuracy is provided.

V DISCUSSION

AI is a field of study where models are designed to simulate human intelligence and behavior. Several technologies are used such as expert systems, intelligent agents, etc. to secure our networks more efficiently and accurately. The advantages of using these technologies are better decision support, good mobility, and better accuracy, which gives better performance.

VI CONCLUSION

The main objective of data security is safeguarding the data. Attackers who use AI for cyber-attacks are now a threat to society. AI is used to alert people before hackers attack their private data. Algorithms are used for user data using certain algorithms. AI plays an important role in data security. Some disadvantages are also there. We need to update the technologies according to the latest requirements. It could become a boon or a curse in the field of cybersecurity as the AI is still developing. Using Expert systems, neural networks and intelligent agents as techniques, we can achieve Cybersecurity. Since AI uses machine learning, deep learning etc. continuous human interactions and training is needed.

VII FUTURE SCOPE

We can develop Social engineering using human interaction to get access to a system by manipulating human behavior. Therefore, in a security environment, the weakest link is the human mind, which an attacker can exploit easily. Making use of artificial intelligence to make our systems more secure and efficient is the solution. We can make use of expert systems, neural networks etc. to increase security in the systems. The nature of future attacks can be predicted by a good AI system using the existing attack logs. It is in the future scope that it is possible to reach the level of security which hackers to intrude.

VII ACKNOWLEDGMENT

This research was supported by Pillai HOC College of Arts, Science & Commerce, Rasayani. It gives me great pleasure to present this research report on "Student Sentiment Analysis". It's a great pleasure and moment of immense satisfaction for me

to express my profound gratitude to my Project Guide, Prof. Priyanka Sonawane.

REFERENCES

1. Artificial Intelligence in Cyber Security-An Investigation”
https://www.academia.edu/34850503/ARTIFICIAL_INTELLIGENCE_IN_CYBER_SECURITY_A_N_INVESTIGATION
2. An Overview on Use of Artificial Intelligence Techniques in Effective Security Management
https://www.researchgate.net/publication/338223306_Artificial_intelligence_in_the_cyber_security_environment_Artificial_intelligence_in_the_cyber_security_environment
3. Data Security using Genetic Algorithm and Artificial Neural Network
<https://www.semanticscholar.org/paper/Data-Security-using-Genetic-Algorithm-and-Neural-Mohana-Venugopal/72fe885d09b9b631a7cf90b0c2b35f63e9d7b11e>
4. Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review
<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9152956>
5. Application of Artificial Intelligence (AI) to Network Security
https://www.appliedaicollege.com/?gclid=Cj0KCCQjw16KFBhCgARIsALB0g8LtX13QMbzCw1ZICyUw9Op18A4gUY8AxuUGRftKVWeFZmijGLQpCb8aAIGTEALw_wcB

Artificial Intelligence In Power Station

Hari Prakash Mishra Prof, Anita Mhatre

Abstract— Artificial intelligence is that the science of automating intelligent behavior presently doable by humans. installation has big tremendous over a number of decades. because the size and complex of the facility system consisting of generators, transmission lines, power transformers, distribution transformers etc. will increase the likelihood of inviting faults. The acquisition of knowledge, the process of these knowledge to be used by the operator, and management of remote devices are the elemental building blocks of all fashionable utility management systems. Manual calculations, technical analysis and conclusions at first adopted the facility system style, operation and management. because the installation grew it become additional advanced because of the technical advancements, selection and dynamic needs

Index Terms— Artificial intelligence, Expert system, Artificial neural network, Fuzzy logic, Power station.

1 INTRODUCTION

There are three unit varieties of country plants best-known for the large electricity generation : i) Thermal power plants, ii) Hydal power plants, iii) atomic energy plants. One might expect that the mobile sensing can play Associate in Nursing more and more vital role within the observance of facility. computing is understood to be the intelligence exhibited by machines and software system, for instance, robots and laptop programs. Associate in Nursing professional system obtains the data of somebody's professional in a very slender such that domain into a machine implementable kind. professional systems area unit unable to be told or adopt to new issues or things

Advantages of artificial neural networks : ☺

- Speed of process. ☺
- They are doing not would like any acceptable information of the system model. ☺
- They need the power to handle things of incomplete knowledge and data, corrupt knowledge.
- They're fault tolerant. ☺
- Artificial neural networks square measure quick and sturdy. Disadvantages of artificial neural networks : ☺
- giant spatiality. ☺ Results square measure invariably generated even though the computer file square measure unreasonable. ☺ they're not climbable i.e. once a man-made neural network is trained to try and do bound task, it's troublesome to increase for alternative tasks while not preparation the neural network. Fuzzy knowledgeable systems also are referred to as as information primarily {based} systems or rule based systems. knowledgeable

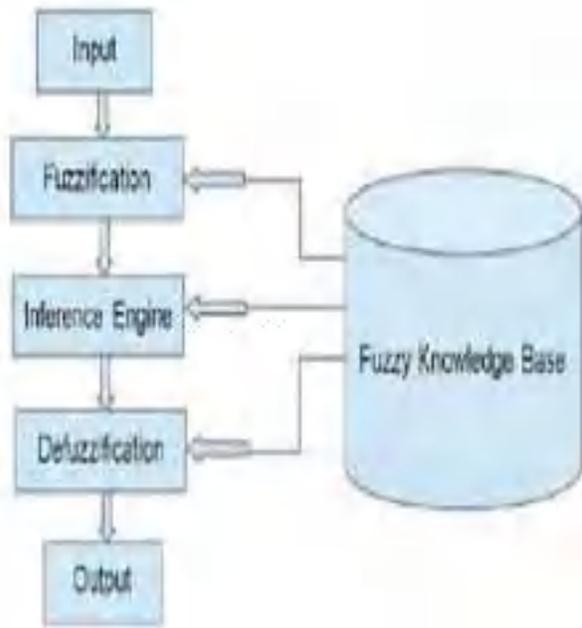
systems square measure laptop programs that have proficiency and competency in an exceedingly explicit field. Artificial neural networks square measure biologically impressed systems that convert a group of inputs into a group of outputs by a network of neurons, wherever every vegetative cell produces one output as a perform of inputs. A basic vegetative cell may be thought-about as a processor that makes an easy non linear operation of it's inputs manufacturing one output. they're classified by their design : variety of layers and topology : property pattern, feed forward or repeated logic or fuzzy systems square measure logical systems for standardisation and rationalization of approximate reasoning. it's the same as human higher cognitive process with a capability to provide precise and correct solutions from bound or maybe approximate data and knowledge. symbolic logic is that the method like that human brain works, and that we will use this technology in machines in order that they will perform somewhat like humans.

METHODOLOGY

There are mainly three techniques :

- i)Expert system techniques,
- ii)Artificial neural networks,
- iii)Fuzzy logic systems.

- Since professional systems are essentially laptop programs, the method of writing codes for these programs is less complicated than really scheming and estimating the worth of parameters utilised in generation, transmission and distribution.
- Any modifications even when style is simply done as a result of they're laptop programs. As artificial neural networks operate biological institutes and perform biological analysis of globe issues, the issues in generation, transmission and distribution of electricity is fed to the factitious neural networks in order that an appropriate answer is obtain net of Things, Cyber Physical Systems and net of Everything.
- Given the constraints of a sensible transmission and distribution system, the precise values of parameters is determined.
 - ⊗ For example, the worth of inductance, capacitance and resistance during a conductor is numerically calculated by artificial neural networks taking in varied factors like environmental factors, unbalancing conditions, and different attainable issues.
 - ⊗ Fuzzy logic is used for coming up with the physical elements of power systems.



RESULT

- replacement human staff for dangerous and extremely specialised operations, like live maintenance of high voltage transmission lines, has been a protracted standing result within the power community.
- Operation in dangerous environments, like hot locations in nuclear plants, access to tight areas, like cable viaducts and cooling pipes, and precise positioning of measuring instrumentality.
- professional systems use the interface mechanism and information to resolve issues that can not be or tough to be resolved by human talent and intellect. ⊗ Results area unit permanent and consistent. ⊗ are often simply documented.
- Results are often simply transferred and reproduced. ⊗ The understanding of the operating of neurons and also the pattern of their interconnection are often accustomed construct computers for finding

planet issues of classification of patterns and pattern recognition.

- Fuzzification provides superior communicative power, higher generality and an improved capability to model advanced issues at low or moderate resolution price
- Stability analysis and improvement. ☉ power grid management. ☉ Fault identification. ☉ Load prognostication. ☉ Reactive power coming up with and its management
- Operation of power grid like unit commitment, hydro-thermal coordination, economic dispatch, congestion management, maintenance programming, state estimation, load and power flow.
- coming up with of power grid like generation growth coming up with, power grid dependability, transmission growth coming up with, reactive power coming up with
- management of power grid like voltage management, stability management, power flow management, load frequency management of stations like fuel cells power plant management, thermal station management. ☉ Automation of power grid like restoration, management, fault identification, network security.
- will be utilized in something from tiny circuits to massive mainframes. ☉ will be accustomed increase the potency of the elements utilized in power systems.
- most of the info utilized in power grid analysis are approximate values and assumptions, formal logic will be of nice use to derive a stable, precise and ambiguity free output.

CONCLUSION :

A reliable, continuous offer of current is important for the functioning of today's fashionable complicated

and advanced society. Electricity is one in all the most important factors for the expansion and valuable for our society. So, implementation of computing is extremely necessary in power station

REFERENCES [

- [1] Stuart Russell, Peter Norvig, Artificial Intelligence : A Modern Approach, Pearson. [2] Jeff Heaton, Artificial Intelligence for Humans, Volume 1 : Fundamental Algorithms, Create Space Independent Publishing Platform.
- [3] Philip C. Jackson Jr., Introduction to Artificial Intelligence, Dover Publications. [4] Keith Frankish, William M. Ramsey, The Cambridge Handbook of Artificial Intelligence, Cambridge University Press.
- [5] Kevin Warwick, Artificial Intelligence : The Basics, Routledge.
- [6] Patrick Henry Winston, Artificial Intelligence, Pearson.
- [7] George F. Luger, Artificial Intelligence : Structures and Strategies for Complex Problem Solving, Pearson.
- [8] Stephen Lucci, Danny Kopec, Artificial Intelligence in the 21st Century, Mercury Learning & Information.
- [9] David L. Poole, Alan K. Mackworth, Artificial Intelligence : Foundations of Computational Agents, Cambridge University Press.
- [10] Mohamad Hassoun, Fundamentals of Artificial Neural Networks, A Bradford Book.
- [11] Sandhya Samarasinghe, Neural Networks for Applied Sciences and Engineering, Auerbach Publications.
- [12] Martin T Hagan, Howard B Demuth, Mark H Beale, Neural Network Design, Martin Hagan.
- [13] Laurene V. Fausett, Fundamentals of Neural Networks : Architectures, Algorithms and Applications, Pearson.

Artificial Intelligence for 5G Wireless Systems: Opportunities, Challenges, and Future Research Directions

MAYUR GANPAT SAKPAL, Prof:- *Anita Mhatre*

Abstract:

The appearance of the wireless communications systems augurs new cutting-edge technologies, including self-driving vehicles, unmanned aerial systems, autonomous robots, the net of-Things, and game. These technologies should high data rates, ultra-low latency, and high Credibility, all of which are promised by the fifth generation of wireless communication systems (5G). Many research groups state that 5G cannot meet its appeal to without artificial intelligence (AI) combination as 5G wireless networks are expected to urge unprecedented traffic giving wireless research designers access to big data which will help in predicting the strain and adjust cell designs to fulfill the users' requirements. later, many researchers applied AI in many aspects of 5G wireless communication design including radio resource allocation, network management, and cyber-security. during this paper, we provide an in-depth review of AI for 5G wireless communication systems. during this respect, the aim of this paper is to survey AI in 5G wireless communication systems by discuss many case studies and also the associated challenges, and shedding new light on future research directions for leveraging AI in 5G wireless communications.

Index Terms: 5G Wireless Communication,

Machine Learning, Deep Learning, Energy Efficiency, Channel Coding, Scheduling, Cybersecurity.

1.INTRODUCTION

5G Wireless communication and mobile networks face many challenges to satisfy the unprecedented growing demands for access to wireless services with ultra-low latency and high data rates.5G network today is that the core technology of the various cutting-edge technologies just like the internet-of-things (IoT), smart grid, unmanned aerial systems, and self-driving vehicles.5G wireless networks are required to be permits a versatile 5G system which is ready to regulate itself in real-time to optimize resource allocation while enhancing

characterized by high flexibility in design and resource management and allocation to fulfill the increasing demands of those heterogeneous networks and users. The 5G specifications, released in 2017 by 3GPP, consider the pliability of design collectively but a fundamental pillar of 5G New Radio, which might be achieved through the mixture of the software-defined network (SDN) and virtual network function (NMF) functions. Such flexibility the standard of experience of users, which needs accurate prediction of the network

behaviour, the traffic demands, and user's mobility. Many Wireless research leading groups predict that AI (AI) is that the following big "game-changing" technology, poised to provide 5G with the pliability and also the intelligence needed. for this reason, many researchers have investigated the efficiency of this theory in many aspects of 5G wireless communications including modulation, channel coding, interference management, and scheduling, 5G slicing, caching, energy efficiency, and cyber-security. Many survey and tutorial papers provided an summary of wireless discipline. as an example, the authors of [1] provided a comprehensive tutorial on how deep learning can empower several applications in wireless systems. Specifically, the authors focused on some sorts of neural networks, like recurrent, spiking, and deep neural networks, and also the way it can empower some wireless communication problems. Nevertheless, most of these survey papers focus on AI theory, which is well-established, over they concentrate on how this technology can solve practical problems in wireless communications. during this paper, we review wireless computing with a specific focus on how AI methodologies can solve complex problems in 5G wireless networks considering many aspects of wireless communication and networking. for every aspect, we illustrate how machine learning could also be applied using well-selected examples; which we pinpoint the benefits and downsides of using such machine learning/deep learning to unravel each particular problem. We also provide some future research directions to beat the

challenges facing AI driven wireless communication and networking. the rest of this paper is printed as follows. Section II presents several applications of AI to unravel issues in 5G wireless communication and networking. For reasons of space, the thought of machine learning and deep learning isn't considered during this paper. Section III gives an summary on the challenges facing the blending of AI in 5G wireless networks additionally as some future research direction to completely exploit AI during this context. Conclusions and future research directions are drawn in section IV.

2.MACHINE LEARNING AND DEEP LEARNING=

AI technology includes machine learning and deep learning. For the remainder of this paper, we target deep learning due to the success deep learning have achieved. the speculation of deep learning is comprehensively and well established. Nevertheless, for completeness reasons, we briefly give an outline of machine learning and deep learning. Machine learning techniques mainly can be classified into three main categories supervised learning, unsupervised learning, and reinforcement learning. In the first category, there is a mapping between the input and output. The machine learning models are given the labels of the dataset at the output, and it's to optimize the weights of the price function so it can best learn the representations of the input file and therefore the rules that map these inputs and their outputs. samples of techniques under this category include logistic regression, support vector machine, decision tree, and random

forest. In contrast, within the second category, the output's labels don't seem to be specified to the machine learning models, which itself has to underline any hidden patterns within the input and cluster the weather of the input dataset. Thus, it is said that the first function of unsupervised learning is underlying patterns instead of mapping the input and its labels. Examples of techniques under this category are clustering techniques like K-means and self-organizing maps. In both supervised and unsupervised learning, there is no reward function, which is present in reinforcement learning that defines reward mechanisms to grant feedback to the model. The last type is reinforcement learning, which is made upon establishing a bequest mechanism. just like supervised learning, in reinforcement learning, there's a mapping between the input and also the output. Over the past decade, a category of techniques called deep learning, which may be either supervised, unsupervised or reinforced, has been employed in many technologies. Deep learning may be defined as a model which involves many hidden layers between the input layer and therefore the output layer. Deep learning reveals unknown correlations in large data sets by using the feed-forward and back-propagation algorithms. One popular class of deep learning is convolutional neural network. A neural network could be a network of neurons that are interconnected, and every neuron consists of a weighted sum of the inputs and one activation function, such as sigmoid function, rectified linear measure (RELU), threshold, and softmax. the most foundations on which neural networks are built are feed-forward

propagation and backward-propagation algorithms. the primary calculates the output as a function of the inputs. The latter computes the weights to reduce the error between the output predicted and therefore the real one.

III. WIRELESS AI= during this section, we selected several use cases in 5G wireless communication and networking empowered by machine learning and deep learning. for every example, we show the concepts, the advantages, and disadvantages of AI-enabled methodologies.

A. Massive MIMO and Beamforming=
Massive MIMO is one feature of 5G.

Through the utilization of an unlimited number of antennas, 5G can focus the transmission and reception of signal power into ever-smaller regions of space. However, several issues are related to this technology. Machine learning/deep learning has been applied in Massive MIMO to beat these issues. For instance, an accurate estimate of the channel with simple estimation methods and an affordable number of pilots is challenging in massive MIMO: the low complexity least-squares (LS) estimator does not achieve satisfactory performance, while minimum mean square error (MMSE) channel estimation is extremely complex.

B. Automatic Modulation Classification
Automatic modulation classification (AMC) may be a core technique in non-cooperative communication systems. Modulation recognition is one task which will help in classifying the modulation kind of a received

signal, which can be a necessary step towards understanding and sensing the wireless environment. High-quality sensing and adaptation improve spectral efficiency and interference mitigation. Deep learning-based AMC systems incorporates three main parts: the first part is signal processing to spice up the quality of the received samples, a frequency offset correction, gain control, amplifiers, and filtering. The second part involves the extraction of features like the amplitude, phase, and frequency of the received signal. The last part may be a signal classifier: classification of the modulation types. Deep learning can do high accuracy of modulation classification.

C. Channel Coding=

an understandable feature of the air interface of the 5G is that the utilization of recent channel coding techniques: Data channels use low-density parity-check (LDPC) codes, and control channels use polar codes [18]. However, the employment of these techniques have some limitations. as an example, polar codes are able to do excellent performance, but it takes several iterations to achieve this performance, and there is no because of predict how briskly polar codes can reach this desired performance. additionally, LDPC codes suffer from high complexity of decoding when either it's used with large block or the channel is under colored noise.

D. Intelligent Radio Resource and Network Management=

Radio resources are scarce, and there's an increasing demand of wireless traffic. Intelligent wireless network management is

that the way forward to fulfill these increasing demands. Machine learning/deep learning are a promising feature for resource allocation in 5G wireless communication networks. Deep learning will be a decent alternative for interference management, spectrum management, multi-path usage, link adaptation, multi-channel access, and holdup. as an example, the authors of [23] proposed an AI scheduler to infer the free slots during a multiple frequencies time division multiple access to avoid congestion and high packet loss. Four last frames state are fed to a neural network, which consists of two fully connected hidden layers. The proposed AI scheduler was tested in an exceedingly very wireless sensor network of 5 nodes and should reduce the collisions with other networks with 50%

E. Energy Efficiency Maximization=

Information and Communication Technology (ICT) is in command of 2% to 10% of the earth energy consumption in 2007, and it's expected to still grow [32]. Also, quite 80% of ICT is from radio access network (RAN), which is deployed to satisfy the peak traffic load and stays thereon even that the load is light. Motivated by saving energy for green communication, 5G specifications require that energy use should decrease to 100% of the normal 4G/LTE networks. This objective will be achieved by reducing the ability consumption of the underside stations and mobile

F. 5G Slicing and Caching=

Two prominent features of 5G are the network slicing and caching. the first allows operators to deliver different service types over the one

network infrastructure. The latter predicts the content that users may request for efficient usage of the storage of the bottom station. Thus, the 5G requires accurate predictions of the needed resources during a slice and also the longer term content of the users.

G. 5G Cybersecurity Deep learning has also been investigated in cybersecurity of 5G wireless communications. for instance, the authors of [39] proposed an unmanned aerial vehicle (UAVs) aided 5G wireless communications with deep reinforcement learning against jamming attacks. The relay UAVs are accustomed establish the communication of legitimate nodes. to see the optimal policy of the relay UAV, the authors addressed proposed a deep reinforcement learning. The methodology can restore the communication between the underside station and thus the legitimate users, but several issues must be addressed to enable these anti-jamming methods.

IV= DISCUSSION AND FUTURE RESEARCH DIRECTIONS=

In this section, we revisit the benefits of deep learning and machine learning in building intelligent 5G wireless communication and networking. We discuss also the challenges facing the mixing of AI in wireless communication further as some future research directions to hurry up this integration.

A. Advantages

the use of deep learning to form intelligent 5G systems has many advantages. for example, in signal processing, deep learning is capable of

performing automatic feature extraction, which could be a hard task in wireless network engineering that usually requires human expertise. Deep learning can perform this task with high accuracy. Another advantage is that deep learning models, in some cases, can do high-performance accuracy and outperform traditional techniques. 5G wireless networks are expected to induce a large amount of knowledge at high data rates. Deep learning can enable 5G systems to profit of this to engineer optimized wireless networks.

B. Challenges =

the mixture of AI in 5G wireless communication systems faces many challenges. variety of those challenges are listed as follows:

1) The reliability and speed trade-off: The reliability of these techniques is much but traditional techniques in wireless communications in solving some problems. as an example, deep learning can compete with LS and MMSE in wireless channel estimation in massive MIMO, but slow feedback characterizes these techniques. Deep learning inference may elongate the system reaction time. this can be because not most wireless devices have access to cloud computing, and whether or not it's the case, communication with cloud servers goes to introduce extra delays.

2) The complexity:

Deep learning algorithms, in due course, must be implemented in wireless devices. However, many wireless devices have limited memory and computing capabilities, which isn't suitable for complex algorithms. the gathering of

enormous samples and training deep learning models takes considerable time, which can be a big impediment to deploy them on some wireless devices having limited power and storage. Also, some applications require processing, and on-fly sampling and training often can't be performed easily. In some cases, the upper the quantity of samples and therefore the more significant the training time are, the higher the accuracy of recognition of the signal and network features is. Acquiring more samples and training the models for extended times incur slow feedback. Therefore, the deep learning models should be designed to attain the most effective accuracy with fewer samples and within a quick time.

3) Data Collection and Cleansing:

it's necessary to collect data and build large comprehensive datasets to educate AI models, and this task isn't often easy to accumulate because mobile service providers, as an example, cannot release these datasets, which contains counselling about the users and may risk the violation of the privacy of their consumers. Also, even with transfer learning, which refers to use models trained on the previous dataset, it's a necessity to adapt these models for specific networks and scenarios which require re-training of the models.

4) Privacy: Preserving the privacy of the users is the first concern of mobile and repair providers. One in all the most challenges in wireless AI is how one can enable the training on a dataset belonging to users without sharing the computer file and putting the personal

information of users in danger. It's necessary to own a security approach to spice up the combination of deep learning in wireless communications.

5) Security: the protection of deep learning models itself is another challenge, as neural networks are prone to adversarial attacks. Attackers can affect the training process by injecting fake training datasets; such injection can lower the accuracy of the models and yield wrong design, which can affect the network performance. Research within the security of deep learning or machine learning, in general, remains shallow.

C. Future Research Directions=

In order to ease the combination of deep learning, research efforts are needed in several directions. For instance, the acceleration of deep neural network alongside advanced parallel computing, faster algorithm, and cloud computing, distributed deep learning systems present a chance for 5G to build the intelligence in its systems to deliver high throughput and ultra-low latency. There have been some recent efforts in deep neural network acceleration [43]. The acceleration of deep neural network, is at three levels: architecture level, computation level, and implementation level. At the architecture level methods will be used, including layer decomposition [44], pruning [45], projection [46], and knowledge distillation [47]. At the implementation level, several characteristics is explored like advanced GPU [48] and FPGA designs [49]. Using deep learning acceleration methods can achieve lower the complexity of deep learning with small loss in the accuracy of

those models

V. CONCLUSION=

In this paper, we presented AI for 5G wireless communication systems. We studied several case studies including modulation classification, channel coding, massive MIMO, caching, energy efficiency, and cybersecurity. As a conclusion of this in-depth study, AI enabled 5G wireless communication and networking may be a promising solution that may provide wireless networks with the intelligence, efficiency, and flexibility required to manage the scarce radio resource well and deliver prime quality of service to the users. However, some efforts are still needed to scale back the complexity of deep learning so it will be implemented in time-sensitive networks and low power devices and test the models in more realistic scenarios.

REFERENCE=

- [1] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials*, 2019.
- [2] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652-655, 2019.
- [3] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning based channel estimation for massive mimo with mixed-resolution adcs," *IEEE Communications Letters*, vol. 23, no. 11, pp. 1989-1993, 2019.
- [4] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114-117, 2017.
- [5] Y. Wang, M. Narasimha, and R. W. Heath, "Mmwave beam prediction with situational awareness: A machine learning approach," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1-5.
- [6] E. Balevi and J. G. Andrews, "Deep learning-based channel estimation for high-dimensional signals," *arXiv preprint arXiv:1904.09346*, 2019.
- [7] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmwave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852-855, 2018.
- [8] M. S. Safari and V. Pourahmadi, "Deep ul2dl: Channel knowledge transfer from uplink to downlink," *arXiv preprint arXiv:1812.07518*, 2018.
- [9] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive mimo csi feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748-751, 2018. [

- [10]** M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive mimo," arXiv preprint arXiv:1906.04610, 2019.
- [11]** G. Gao, C. Dong, and K. Niu, "Sparsely connected neural network for massive mimo detection," EasyChair, Tech. Rep., 2018.
- [12]** M. Alrabeiah and A. Alkhateeb, "Deep learning for tdd and fdd massive mimo: Mapping channels in space and frequency," arXiv preprint arXiv:1905.03761, 2019.
- [13]** L. Sanguinetti, A. Zappone, and M. Debbah, "Deep learning power allocation in massive mimo," in 2018 52nd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2018, pp. 1257-1261.
- [14]** T. Maksymyuk, J. Gazda, O. Yaremko, and D. Nevinskiy, "Deep learning based massive mimo beamforming for 5g mobile network," in 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS). IEEE, 2018, pp. 241-244.
- [15]** J. Jagannath, N. Polosky, D. O'Connor, L. N. Theagarajan, B. Sheaffer, S. Foulke, and P. K. Varshney, "Artificial neural network based automatic modulation classification over a software defined radio testbed," in 2018 IEEE International Conference on Communications (ICC). IEEE, 2018, pp. 1-6.

Concept Artificial Intelligence, its impact And Emerging Trends

MAYUR GANPAT SAKPAL, Prof :-ANITA MHATRE.

Abstract: Whilst there is plenty of new hype around computer science (AI) as we move into the third decade of the 21st century, the history books would tell you that it's been around for several years. it's best to start with a short history of AI before digging deeper into the current principles. Application of machine learning came into existence for the primary time during Second World War when a Computer Scientist named Alan Turing tried to crack a bit of code called referred to as the "Enigma code" which was employed by the German forces for secure communication. According to Turing, a machine which might communicate with humans without the humans knowing it could be defined as an "intelligent" machine.

Keywords – Artificial Intelligence, Machine Learning, Deep Learning

INTRODUCTION

Whilst the term "artificial intelligence" wasn't coined until 1956, Turing prompted an accelerated amount of research into this branch of knowledge with a vision on transforming the world. Despite some early promise like the primary neurocomputer built by Marvin Minsky (known as Ferranti Mark 1) and exploration in robotics just like the Unimate robot from GM, research of computer science (AI) in 1970s headed into an "AI Winter." Scientists found it hard to create intelligence without more data and computing power. In turn, this led to a slump in funding and investment across all areas of AI. It wasn't until the dawn of the new Millennium that there was a real upturn as improvements in component meant companies had more data resulting in opportunities for generating machine learning propositions. because the Big 4 (Google, Apple, Facebook and Amazon) began to get heavily involved in AI

technology, investment grew exponentially from the beginning of autonomous cars in 2005, IBM Watson in 2006 (a famous case where it defeated a Jeopardy champion), Netflix streaming and Google street view in 2010 and Alexa in 2015. We now have countless other applications founded on evolving technologies like augmented reality (AR), video game (VR), computer vision, language processing (NLP) et al. Terms like machine learning and deep learning became synonymous with the world but there tends to be clouds on exactly what they mean.

2. WHAT IS AI?

AI is also a field of computing that studies how machines can imitate the intelligence of their human counter parts. Over the last decade, definitions of the term became quite loose and check with almost any computerized or automated function. However, the difference between an AI system and traditional software packages is that the flexibility to make informed judgments and decisions by responding to

patterns in data. Many folks use movies as a guide to the current state of computing. In November 2019, the foremost recent release within the "Terminator" movie series sparked some debate on the subject as did previous films within the series. the explanation is that movies tend to supply up AI on the thought that it'll cause doomsday type scenarios instead of

2.1 Artificial Narrow Intelligence (ANI)

Commonly called machine learning, ANI solutions specialise in one area and one problem at a time. this is often that the shape of AI that we see within the market today and complete tasks like recommending a product or predicting the forecast. ANI comes very near replicating and sometimes surpassing how humans perform tasks and is that the only variety of AI that really exists today (some reports might challenge that, but it's certainly the sole monetized form).

2.2 Artificial General Intelligence (AGI)

AGI is that the subsequent level up from ANI and refers to AI which includes a "human level of cognitive function." to realize success, an AGI system would want to attach potentially thousands of ANI systems together to imitate human behavior. to place it into context, the market leading IBM Watson system took 40 minutes to simulate only one second of neuro-activity. Big companies are striving to realize AGI which we'll get there but just nearly yet.

2.3 Artificial Super Intelligence (ASI)

This is the aim where we start thinking of phantasy. An ASI system is one which is able to completely surpass any variety of human intelligence. It may be creative, make rational decisions, build relationships and choose

whether it wanted to be good or evil. it's thought that the progression from AGI to ASI wouldn't necessarily be that vast. If machines can start springing up with their own concepts with AGI, an excellent intelligent system would be the next logical step. to higher understand its current state, the 2 main applications that be the umbrella of ANI are machine learning and deep learning. Before staring at how AI impacts society, we are visiting take a deeper take a glance at both techniques.

3.WHAT IS MACHINE LEARNING?

Machine learning is the most common form of AI in the globe today and how most use cases for ANI are realized. the rationale we've been able to develop AI solutions at a rapid rate within the last few years is due to vast volumes of information being generated within the world. Machine learning is that the process where computer systems become capable of gaining intelligence through data. Devices and Systems which are built with machine learning algorithms can learn from experience within the kind of historical data. When we talk about algorithms, these are programming codes, a small amount like how a developer would build a website or some other online functionality. within the data science, the two codes that tend to be used are Python and R. once you hear these names banded about, it refers to the language in which an algorithm has been created certain it to function. Both Python and R have pros and cons. See the references section of this paper for details on where you'll be able to find more information on Python and R. A recent example of machine learning would be

Amazon Alexa. Everyone are going to be acquainted with the voice activated device which sits within the front room and responds to commands but might not have considered the AI system which sits as its foundation. to place this into a machine learning context. A user will make an invitation by reproval Alexa. Alexa will recognize what the user is asking employing a text to speech algorithm. this is often essentially the simplest way of turning unstructured information like spoken words into data. The converted speech is distributed to the cloud and matched against a limiteless pool of existing data to determine the simplest possible response. Alexa then sends the reply and converts it to audio for the user to pay attention to. the strategy Alexa goes through is machine learning (ANI), a system trained for a selected task that it does incredibly well. There are many other similar examples that we'll discuss later during this paper. There are four common styles of machine learning which are summarized below.

3.1 Supervised Learning=

This method takes existing data and trains a model to figure out a way to classify a replacement piece of information. for instance, it could hold data on the symptoms of diabetes and when it receives biopsy results of a new patient, it's able to form a diagnosis prediction. Initially somebody's would train the machine how to classify symptoms into "Has Diabetes" or "Does Not Have Diabetes." Over time, with enough data, an AI system are going to be ready to take a replacement set of knowledge

and build its own prediction as to which classification the new patient falls into.

3.2 Unsupervised Learning =

Unlike, supervised learning, these models will try and classify data with none prior knowledge. The algorithms look to search out patterns themselves and put data into groups. a customary example are some things like customer purchasing behaviors. The algorithm won't have existing labels and might choose its own the way to classify the info, often called clustering. Imagine visiting a celebration where everybody may be a stranger. Your mind will probably classify people supported age, gender or clothing. You don't know them but have still discovered the classifications.

3.3Semi-Supervised Learning =

is a mix between supervised and unsupervised learning. in an exceedingly large volume of data, it's normal that some items are labelled, and a few aren't. A semi-supervised model would have some labelled data to grasp that classification does exist. it's then trained on unsupervised data to define the boundaries of what it's viewing and potentially specify new classifications that the human failed to specify when labelling. as an example, machine learning is being employed to detect fraud in banking by identifying patterns within the info. However, initially you will be able to only classify the fraudulent activity that you simply just understand. the complete criminal mind set of fraud is about undertaking activity that no-one can detect meaning classifying it's impossible. A semi-supervised algorithm will

take new data and retrain this a mix between supervised and unsupervised learning. in an exceedingly large volume of data, it's normal that some items are labelled, and a few aren't. A semi-supervised model would have some labelled data to grasp that classification does exist. it's then trained on unsupervised data to define the boundaries of what it's viewing and potentially specify new classifications that the human failed to specify when labelling. as an example, machine learning is being employed to detect fraud in banking by identifying patterns within the info. However, initially you will be able to only classify the fraudulent activity that you simply just understand. the complete criminal mind set of fraud is about undertaking activity that no-one can detect meaning classifying it's impossible. A semi-supervised algorithm will take new data and retrain the model each time to add to its classification methods. Whilst a computer being left to do this on its own might not be 100% accurate, it is better than having no labels at all.

4. WHAT IS DEEP LEARNING?

Deep learning is another subset of AI and also the term is usually used interchangeably with machine learning, but the 2 applications are different. within the simplest form, deep learning algorithms have numerous layers, each providing a separate interpretation of the data it's supported. This multi-layer approach is usually spoken as a synthetic neural network as their function is intended to (at least attempt) replicate that of somebody's brain. one in every of the key differentiators of machine and deep learning lies in classification. Earlier during this paper,

we spoke about how supervised learning techniques depend on labelling images to classify predictions accurately. a regular cited example is that of distinguishing between pictures of cats and dogs. In machine learning, there would be batches of labelled images for the machine to find out from and remove to come to a decision if future pictures will be classified as a dog or cat. Solving the identical problem with deep learning wouldn't use labelled cat and dog data. Instead, a brand new picture is distributed through the multiple layers of our neural network to define the assorted features and are available up with a reasoned decision. The human brain works within the identical due to find the suitable identifiers. Even children play the parlor game "Guess Who" which is all about feature recognition to assist the brain on its way towards logical predictions Deep learning is being employed in places where there's just too much data for traditional machine learning algorithms to derive conclusions from or problems that are highly complex. Autonomous vehicles are one technology that are reliant on deep learning to succeed. For them to think variety of a personality's, there must be a network of models working simultaneously. as an example, one has to understand the way to drive, another has to perceive the environment, one must understand road signs then on. Until the AI can do of these tasks at the identical time, we won't see full commercialized samples of completely driverless cars. Advancements in deep learning will ultimately be the key to AGI.

5. IMPACT ON SOCIETY

Machine learning and deep learning are already having an impression on society across many industries and applications. variety of the key cases are outlined below. The AI sector is by no means limited to those, but they represent the primary uses within the market.

5.1 Customer Service one all told the foremost common styles of AI is that the conversational chatbot. These are messaging apps, speech-based assistants or voice activated devices that are wont to automate communication and make a really personalized customer experience. These applications (often known Machine learning and deep learning are already having an impression on society across many industries and applications. variety of the key cases are outlined below. The AI sector is by no means limited to those, but they represent the primary uses within the market. 5.1 Customer Service one all told the foremost common styles of AI is that the conversational chatbot. These are messaging apps, speech-based assistants or voice activated devices that are wont to automate communication and make a really personalized customer experience. These applications (often known the Internet of Things or IoT) can process vast amounts of information instantly meaning they can make faster and more accurate responses than a person's would ever be ready to. Similar personalization that produces best use of information can be used in marketing. This is where we get emails that are relevant to us and social media ads that just happen to be something, we are interested in. In some

cases, each customer can even see different website homepages depending on their likely preferences and what will interest them the foremost. Utilizing AI in these ways is a great way to ensure customer loyalty through a personalised experience.

5.2 Data Security and Fraud AI =

AI are often wont to help identify fraudulent transactions and forestall unauthorized access to data. in a very rapidly growing digital world, Artificial Intelligence play a crucial role in defending cyber-attacks. Powerful algorithms can find malware and combat spam as an example. Machine learning will detect irregular patterns in the data and inform businesses when there's a possible threat. As well as this we are seeing the increased utilization of identity checks other than passwords such as identity verification and fingerprint technology. These unique identifiers based on unstructured data are far harder to hack and offer a great layer of protection for businesses. AI are often wont to help identify fraudulent transactions and forestall unauthorized access to data. in a very rapidly growing digital world, Artificial Intelligence play a crucial role in defending cyber-attacks. Powerful algorithms can find malware and combat spam as an example. Machine learning will detect irregular patterns in the data and inform businesses when there's a possible threat. As well as this we are seeing the increased utilization of identity checks other than passwords such as identity verification and fingerprint technology. These unique identifiers based on unstructured data are far harder to hack and offer a great layer of protection for

businesses.

5.3 Business Process Automation

Businesses that are established for an extended time tend to have several manual processes. AI could be a natural partner to optimize these efforts given its efficiency at handling routine tasks, improving interfaces, willingness and speed to try and do monotonous tasks and skill to handle massive amounts of data. There are some obvious processes like using robotics in factories, managing conditions in product storage, processing payments and registering customer requests but these only touch the surface of the chances. Doctors can use AI devices to dictate clinical notes which automatically fills within the relevant forms and orders a prescription. Lawyers will use AI to process contracts and agreements in an exceedingly blink of an eye which will have taken them days or weeks.

5.4 Predictive Analytics

Machine learning is getting used in prediction-based systems. as an example, consider an individual applying for a loan. As they enter their data, machine learning algorithms can predict in real-time whether or not they are likely to be an honest or bad future risk i.e. will they ultimately neglect payments. The model could then choose the interest rates or term of the loan instantly. Predictive applications are becoming quite common across several industries now. One such way is via recommender systems. for example, Netflix predict what show we wish to stream, Spotify tell us the music we'd prefer to hear, and Amazon know the

products we wish to buy for. Even a Google search is predicting what we'd wish to grasp after typing or saying some words.

5.5 Staff Training=

AI is being employed in businesses to form personalized training plans. Some companies could have huge knowledge bases that take staff weeks or perhaps months to be told. AI has been shown to chop this in half by presenting content to the learner within the way that the majority accurately fits them. this might include the order they learn items in, the length of it slow between when learners are presented with repeat information or the sort of fabric like written, visual and audio. Training is both more useful and enjoyable.

6. EMERGING TRENDS6

Many emerging AI trends are focused on machine and deep learning techniques.a transparent priority getting into 2020 is autonomous AI. we have already spoken about vehicles but whilst they are not ready to be commercialized yet, innovations like drones and robots that progress their own are likely to own a wonderful bigger impact. Google and Amazon are already within the testing phase of drone technology but other applications like in agriculture, construction and logistics and making accelerated headway. In each of these, drones are removing immense amounts of toil and driving better efficiency and productivity. Beyond that, cybersecurity needs to be at the top of AI strategy lists. Gartner say that as much as 30% of AI cyberattacks leverage data poisoning or AI model theft to

compromise systems. Organizations have to start doing all they can to prevent innovative systems from being penetrated. Whilst machine learning and deep learning are critical to innovation, they're also a keen way for hackers to develop techniques that can carry out new kinds of cyberattacks. the planet of search and conversation is additionally changing. we've got already spoken about chatbots in this paper, but further Gartner research has said it expects as much as 70% of workers will work with conversational platforms every day. This could be in the form of mobile, voice activated or service devices but in whichever form, conversational AI is becoming business. 5G networks will provide super-fast download and upload speeds which leads to greater access to data. These networks are available now but are still expensive and confined to quite specific regions Whilst it isn't an emerging trend as such, connectivity enhancements like 5G will likely bring technologies like augmented and computer game back to the forefront of AI systems. This technology has already shown to have practical applications in healthcare, construction and education but continues to be yet to achieve full potential. of these emerging trends incrementally move us from ANI towards an AGI world. Slow and steady wins the AI race.

7. CONCLUSION

AI has been around us for several years but as we move towards 2020, more is anticipated from the technology than ever before. With the way within which it's changed way of life via machine and deep

learning, AI has become embedded as a part of what we do. In fact, much of the time we don't even recognize something as AI because it is so familiar, rather like how we take using the online as a right. AI has always been considered as futuristic if we blow over films and television but because the examples within the paper show, it I a good deal within the here and now.

REFERENCES

- [1] Emerging AI And Machine Learning Trends to Watch In 2020, November 2019, <https://www.crn.com/slide-shows/internet-of-things/6-emerging-ai-and-machinelearning-trends-to-watch-in-2020>
- [2] Sudweeks, F. (1989). Artificial Intelligence in design. J. S. Gero (Ed.). Springer.
- [3] Chrisley, R., & Begeer, S. (Eds.). (2000). Artificial intelligence: critical concepts (Vol. 1). Taylor & Francis.
- [4] Li, D., & Du, Y. (2017). Artificial intelligence with uncertainty. CRC press.
- [5] Pannu, A. (2015). Artificial intelligence and its application in different areas. *Artificial Intelligence*, 4(10), 79-84.
- [6] Self, J. (Ed.). (1988). Artificial intelligence and human learning: intelligent computer-aided instruction. London: Chapman and Hall.
- [7] Shi, Y. (2019, February). The Impact of Artificial Intelligence on the Accounting Industry. In *The International Conference on Cyber Security Intelligence and Analytics* (pp. 971-978). Springer, Cham.
- [8] Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business &*

Industrial Marketing.

[9] Jackson, P. C. (2019). Introduction to artificial intelligence. Courier Dover Publications.

[10] Davis, L. K., & Uppu, R. M. (2019). Fundamental Theory of the Evolution Force (FTEF): Gene Engineering utilizing Synthetic Evolution Artificial Intelligence (SYN-AI). bioRxiv, 585042.

[11] Engelbrecht, A.P. (2002). Computational Intelligence: An Introduction.

[12] Hemmert, F., Becker, P., Görts, A., Hrlic, D., Netzer, D.V., & Weld, C.J. (2019). aicracy: Everyday Objects from a Future Society Governed by an Artificial Intelligence. Mensch & Computer.

[13] Aslan, Oner Deniz et al. "Using Artificial Intelligence for Automating Pavement Condition Assessment." (2019).

[14] Hurley, J.D., Johnson, C., Dunham, J., & Simmons, J. (2019). Nonlinear Algorithms for Combining Conflicting Identification Information in Multisensor Fusion. 2019 IEEE Aerospace Conference,1-7

Natural Language Processing Using Artificial Intelligence

Hari Prakash Mishra, Prof Anita Mhatre

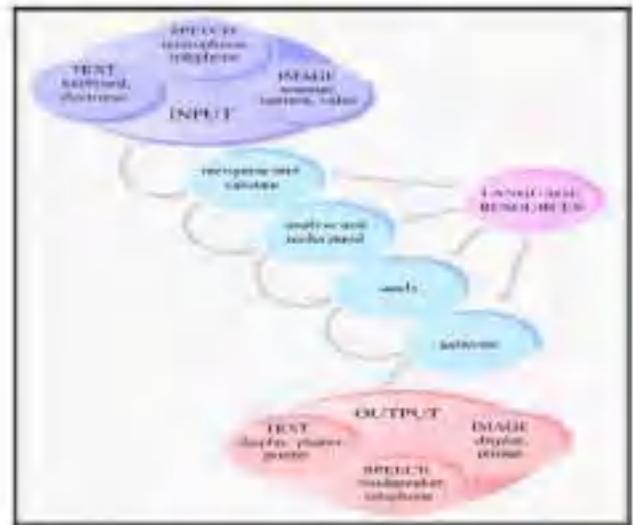
Abstract— Artificial Intelligence (AI) is that the study of a way to create computers (machines) do things that, at the instant, folks do higher. There ar several applications of the factitious intelligence. language process (NLP) is one among the coming applications of AI. The goal of the language process is to style and build computer code that may analyze, understand, and generate languages that humans use naturally, in order that eventually you'll be ready to address your laptop like you were addressing another person. language process is that the use of computers to method written and language for a few sensible, useful, purpose: to translate languages, to induce data from the net on text information banks thus on answer queries, to hold on conversations with machines, and to induce recommendation regarding, say, pensions and then on. sensible applications of language process ar MT, info access, data retrieval, text categorization, extracting information from text etc. Still no such utterly operating system has developed nonetheless however analysis goes on. And it's going to be done shortly. Some basic systems ar already developed like ELIZA, INTELLISHRINK, and AMALGAM etc.

Index Terms — Artificial Intelligence, Natural Language Processing.

1 INTRODUCTION

Artificial Intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. As a theory in the philosophy of mind, artificial intelligence (or AI) is the view that human cognitive mental states can be duplicated in computing machinery. Accordingly, an intelligent system is nothing but an information processing system. Natural Language process (NLP) is one in every of the future applications of AI. The goal of the tongue process (NLP) is to style and build package which will analyze, understand, and generate languages that humans use naturally, so eventually you. are able to address your pc as if you were addressing another person. This goal isn't simple to succeed in. "Understanding" language means that, among alternative things, knowing what ideas a word or phrase stands for and knowing method to|a way to} link those ideas along in an exceedingly purposeful way. It's ironic that tongue, the image system that's best for humans to be told and use, is hardest for a pc to master. Long when machines have tested capable of inverting massive matrices with speed and beauty, they still fail to master the fundamentals of our spoken and written languages.

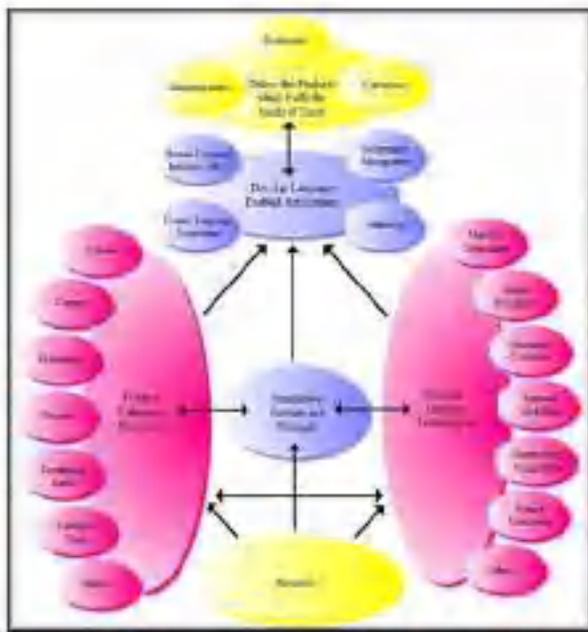
COMPONENTS OF NLP SYSTEM



- Shows a model of a Language Enabled System. inside this general model there ar, of course, many alternative configurations. looking on the appliance of the technology, not of these elements are required. the essential processes of linguistic communication process ar shown within the diagram on top of. These are loosely involved with:
- Getting into material into the PC, mistreatment speech, written text or handwriting, or text either keyed in or introduced electronically.
- Recognizing the language of the fabric, distinctive separate words, for instance, recording it in symbolic kind and corroboratory it.

- Building Associate in Nursing understanding of the that means of the fabric, to the suitable level for the actual application
- mistreatment this understanding in Associate in Nursing application like transformation (e.g. speech to text), info retrieval, or human language translation.
- Generating the medium for presenting the results of the appliance finally, presenting the results to human users via a show of some kind: a printer or a plotter; a loud speaker or the phone.

ARCHITECTURE OF NLP SYSTEM



By far the most important a part of human language happens as speech. written communication could be a fairly recent invention and still plays a less central role than speech in most activities. however process written communication is simpler, in some ways in which, than process speech. as an example to create a program that understands auditory communication, we want all the facilities of a written communication perceive further as enough extra data to handle all noise and ambiguities of the audio signal. therefore it's helpful to divide the complete language process downside into 2 tasks:

- 1) process written communication, victimization lexical, syntactic, and linguistics data of the language further because the needed universe data.
- 2) process auditory communication, victimization all {the data|the data|the data} required higher than and extra knowledge regarding descriptive linguistics further as enough additional information to handle the more ambiguities that arise in speech.

The first level there square measure variety of generic categories of application, such as:

- language translation
- data management (multi-lingual)
- authoring (multi-lingual)
- human/machine interface (multi-lingual voice and text)
- At the second level, these facultative applications square measure applied to universe issues across the social and economic spectrum. So, for example:
 - Infomation management are often employed in associate degree info service, because the basis for analyzing requests for info and matching the require Artificial Intelligence (AI) is that the study of the way to build computers (machines) do things that, at the instant, folks do higher. There square measure several applications of the substitute intelligence. language process (NLP) is one amongst the approaching applications of AI. The goal of the language process is to style and build code that may analyze, understand, and generate languages that humans use naturally, so eventually you may be able to address your laptop like you were addressing another person. language process is that the use of computers to method written and voice communication for a few sensible, useful, purpose: to translate languages, to urge info from the online on text knowledge banks therefore on answer queries, to hold on conversations with machines, and to urge recommendation concerning, say, pensions so on. sensible applications of language process square measure computational linguistics, information access, info

- retrieval, text categorization, extracting knowledge from text etc. Still no such fully operating system has developed nevertheless however analysis goes on. And it should be done shortly. Some basic systems square measure already developed like ELIZA, INTELLISHRINK , and AMALGAM etc.est against a information of text or pictures, to pick out the data accurately.
- Authoring tools square measure usually employed in data processing systems however also can be accustomed generate text, like business letters in foreign languages, still as in conjunction with info management, to produce document management facilities
- Human language translation is presently accustomed offer translator workbenches and automatic translation in restricted domains
- Most applications will usefully be given language user interfaces, as well as speech, to enhance their usability

MAIN STEPS IN THE PROCESS

1. Morphological Analysis: Individual words are analysed into their elements, and non word tokens, like punctuation, are separated from the words.

2. Syntactic analysis: Linear sequences of words are reworked into structures that show however the words relate to every alternative. Some word sequences could also be rejected if they violate the language's rules for the way words could also be combined. for instance, AN English grammar analyser would reject the sentence "Boy the go the shop."

3. Semantic Analysis: The structures created by the grammar analyser are assigned which means. In alternative words, a mapping is created between the grammar structures and therefore the objects within the task domain. Structures that no such mapping is feasible could also be rejected. for instance, in most universes, the sentence "Colorless inexperienced ideas sleep furiously" would be rejected as semantically abnormal.

4. Discourse Integration: The which means of a private sentence could depend upon the sentences that precede it and should influence the which means of the sentences that follow it. for instance, the word "it" within the sentence, "John wished it", depends on the previous discourse context, whereas the word "John" could influence the which means of later sentences.

5. Pragmatics Analysis: The structure representing what was aforesaid is reinterpreted to work out what was truly meant. for instance, the sentence "Do you recognize what time it is?" ought to be understood as missive of invitation to be told the time.

COMPARISION OF NLP SYSTEM

S. N.	SYSTEM NAME	DOMAIN	LANGUAGE	APPROACH	YEAR
1	GENLIDS	NATURAL	NATURAL	LEXICAL ANALYSIS	2009
2	PNLIDS	ACADEMIC	PORTUGAL-SPANISH	SHALLOW PARSER	2010
3	HNLIDS	EMPLOYEE	HINDI-ENGLISH	SHALLOW PARSER	2011
4	PORTABLE NLP	NATURAL	CFG FRAMEWORK	EMBEDDED SEMANTICS	2012

CONCLUSION

The complete method of the natural language processing system, at several places had created the work easier. simply by our natural language(any language) we will direct the automaton, which may do oral communication with pc and there's no want of someone to figure as a translator for the oral communication between 2 persons UN agency don't recognize any common language. Still current program haven't reached this level however they will do thus terribly presently. Language technologies may be applied to a large vary of issues in business and administration to supply higher, simpler solutions. they'll even be utilized in education, to assist the disabled, and to bring new services each to organizations and to shoppers. There square measure variety of areas wherever the impact is important like competitory in a very international market, giving services directly through telebusiness, supporting electronic commerce, enhancing diversion, leisure and ability.

REFERENCES

[1]. Natural Language Processing (Special Issues of Artificial Intelligence) Paperback – Import, 11 May 1994 by Fernando C N Pereira (Author).

[2]. Eugene Charniak and Drew McDermott, Introduction to Artificial Intelligence, Pearson, 1998, Chapter4. [3]. Enhanced Text Retrieval Using Natural Language Processing Elizabeth D. Liddy President1, 2 Article first published online: 31 JAN 2005 DOI: 10.1002/bult.91

[4]. Ambient intelligence—the next step for artificial intelligence C Ramos, JC Augusto, D Shapiro - Intelligent Systems, IEEE, 2008ieeexplore.ieee.org References 1. IST Advisory Group, Scenarios for Ambient Intelligence in 2010.

[5]. Natural language processing: an introduction
PrakashMNadkarni, 1. LucilaOhnoMachado,2 and Wendy W
Chapman2 in volume 18 on page 540.J Am Med Inform Assoc.
2011 Sep-Oct; 18(5): 544–551.

[6]. K.R. Chowdhary Professor & Head CSE Dept. M.B.M.
Engineering College, Jodhpur,India.April29,
2012.NaturalLanguageProcessing.

[7]. Rada Mihalce, Hugo Liu, and Henry
Lieberman Computer Science Department, University of North
Texas rada@cs.unt.edu Media Arts and Sciences, Massachusetts
Institute of Technology. NLP (Natural Language Processing)
for NLP (Natural Language Programming)

Transforming Digital Marketing with Artificial Intelligence

Rutuja Yewale, Prof Anita Rajput

Abstract -Artificial Intelligence Marketing is a technique of optimally utilizing technology to enhance customer's experience. In recent times it becomes really important for the companies to recognize and understand customer needs and their expectations in terms of products as well as services. With the help of AI, Marketers can process the huge amount of data, carry out individualized sales and fulfil customer expectations. Besides this, the perception they get about the customer and their requirements in a shorter time frame help them to boost performance and Return On Investment (ROI) rapidly.

In this paper we have provided whole concept of using Artificial Intelligence in marketing. It also shows benefits of implementing AI in marketing, and how AI is reshaping marketing, need of AI in marketing which is illustrated with real world examples from different businesses.

This paper entails a section dedicated to implementation of different AI technology throughout the customer life cycle. It also well-inform you about the modern marketing methods. By the end of the paper you would have a complete overview of the present-day Artificial Intelligence Marketing strategies and would have a holistic view about AIM (Artificial Intelligence Marketing) sector.

Index Terms - Artificial Intelligence, Marketing, Marketing Strategy, Personalization.



1 INTRODUCTION

Marketing is a technique for creating, communicating, exchanging and delivering offerings that have value for clients, partners, customers or consumers. With the introduction of AI in the field marketing is already transformed the face of marketing as we know. Initially, it seems really difficult to relate the field of Artificial Intelligence with various marketing strategies as one has to think out of the box to study or explore this field. It is even harder to imagine any field that has not been impacted by AI. The development of Artificial Intelligence has altered the dynamics of marketing across the world.

1.1 WHAT IS ARTIFICIAL INTELLIGENCE MARKETING?

AI marketing uses artificial intelligence AI marketing uses artificial intelligence technologies to make automated decisions based on data analysis, data collection, and additional observations of customers or economic trends that can impact marketing efforts. AI is oftenly used in marketing efforts where speed is an essential factor. Artificial Intelligence tools uses data and customers profile to learn how to best communicate with customers, then serve them at the right time without intervention from marketing team members, ensuring maximum efficiency.

1.2 Benefits of Artificial Intelligence

Evaluation of some ways that

can help marketers to shape marketing strategies (Refer Fig 1)



fig no 1

- **Profound Hunting** : Because of rapid growth in the technology customers can find whatever they are looking for at any time and at any place with help of rapid fire search engines (Google, Yahoo etc.) . What AI does is help marketers in analyzing customers' search data and determining the key factors to which they must focus their efforts.
- **Effective Advertisements**: With the availability of huge data, marketers can create smarter and more effective online advertisements. Artificial Intelligence solutions can deeply evaluate a customer's keyword searches and social profiles and aid in creating personalized advertisements.
- **Filtered Content**: Audience analytics can help marketers to understand customers on an individual basis. AI can be used to filter the potential buyers and create customized content which would be ideal and relevant to the potential customers groups.
- **AI Bots**: Customer retention is most important as the generation of new customers. Artificial Intelligence serves as the driver for customer retention. AI Bots are used to run chat functions direct-to-consumer . Through this marketer can cut their extra expenditures and also save their time.
- **Progressive Learning**: AI can not only use access to hidden insights but also can be taught to integrate formerly hidden insights into latest campaigns. With passage of time these AI solutions will become

even smarter and more effective in eliminating trash data and promoting real-time decision-making.

1.3 How is Artificial Intelligence (AI) transforming marketing?

It has become customary to perform large- scale data analysis in the marketing sector. Elements like Artificial intelligence, Big data analytics , machine learning are making their mark on corporate operations. After increased use of these trailblazing techniques marketing teams have wide scope to convey massive value to AI's potential data. AI is the ability of machines to perform intellectual tasks that one can relate with human minds such as problem solving, learning, exercising creativity, reasoning, perceiving and interacting with the environment.

Emergence of AI-powered marketing solutions is to reducing the responsibilities of marketer and assigning these tasks to machines.

Graph 1, displays statistics related to end results achieved through the implementation of AI Marketing.

Graph 2 highlights AI Adoption in different sectors and fields.

Reaping Rewards: Companies have to focus on three major areas to exploit the benefits of AI, which are Marketing and Commercial Talent Management, Data Strategy, Data Expertise. Over time the processing power and data harvesting advances and AI-powered systems become smarter. But the level of effectiveness of these systems depends on how well the organizations can interpret answers and work on them to construct their future commercial strategies.

Customer Segmentation: The job of marketers is to furnish the Right information to the Right individual at the Right point of time, to complete this marketer prefer customer segmentation. Most organizations prefer clustering their customers on certain factors which are similar at that time. AI helps marketers to distinguish their customers and discover what motivates them. This information helps marketers in creating a long-lasting relationship with their customers.

Emerging Trends: Artificial Intelligence is enterprising three megatrends in marketing: personalization, forecasting and automation. Marketing leaders mainly

need to focus on these three areas to develop beneficial capabilities. Automation is a process of making decisions by the use of machines or helping human-decision makers by providing recommendations. Product recommendation and dynamic pricing are examples of automation. Forecasting is the development of models for future scenarios by finding some patterns in historical (previous) recorded data. These forecasting systems keep on learning and inculcating from new data, this implies they are not static. Personalization refers to usage of results or outcomes developed, which are tailored according to different customer segments. With the usage of comprehensive customer information, these systems analyze every potential customer and cluster them according to some common factor

2. CORE ELEMENTS OF ARTIFICIAL INTELLIGENCE

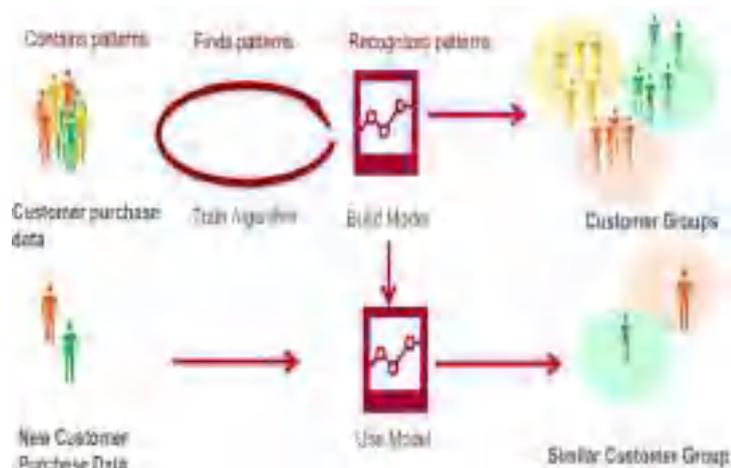


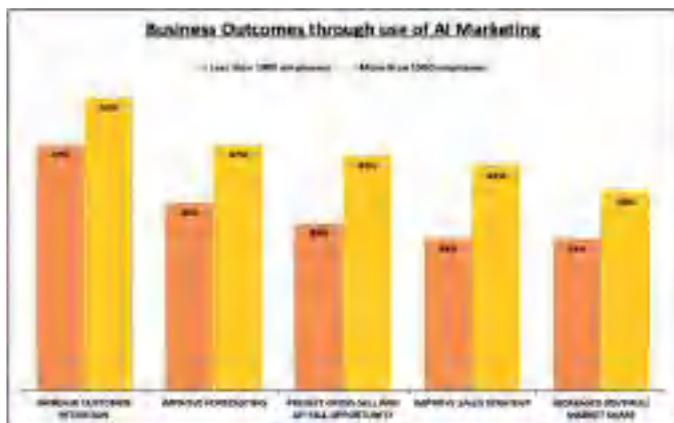
Fig 2

Figure 2 explains the sequence for execution of core elements from gathering data to producing powerful solutions. AI Marketing include few key elements which make it as powerful as it is today. These elements function in series as they are explained below:

Big Data can be considered the basic concept of collecting a huge amount of data of customers' purchase patterns as well as new customers' purchase preferences. It can also be referred as marketer's competence to aggregate and segregate extensive data sets with minimal manual work. Such altered data can be used by the marketing teams for ensuring that the right message is being delivered to the right person at the appropriate time via a channel of choice.

Machine Learning in general can be referred to as building and utilizing models based on recognized patterns. Machine learning platform comes into the picture when marketers try to retrieve significant information from huge data repositories. This can help in recognizing and understanding trends or common instances and successfully anticipate the insights and reactions, helping marketers in understanding the major factor and probability of certain actions repeating.

Powerful Solutions the end result, provided by artificial intelligence marketing truly conceive the world in a similar manner a human would. The platforms provided by AI can recognize the significant concept and themes across huge data sets astonishingly fast. While considering the technical aspects AI solutions also interpret communication, feelings and desires like a human which make this platform widely accepted.



Graph 1

	Service operation	Product and /or Service Development	Marketing and Sales	Supply Chain Management	Manufacturing	Risk	Human Resources	Storage and Copyright Finance
Telecom	75	41	30	26	11	20	17	10
High Tech	40	39	34	29	20	27	18	17
Financial Services	40	34	33	7	4	40	19	14
Professional Services	30	42	46	16	11	26	16	14
Electric Power and Natural Gas	30	41	30	14	19	24	16	10
Healthcare Systems	34	37	37	12	9	28	18	11
Automotive and Assembly	27	39	37	11	49	2	8	4
Travel, Transport and Logistics	31	34	37	18	4	4	2	4
Retail	25	11	37	16	7	9	8	6
Pharma and Medical Products	31	41	37	13	13	3	6	4

3. REAL WORLD EXAMPLES OF ARTIFICIAL INTELLIGENCE MARKETING

Possibilities with AI are limitless for marketers. The following are some examples of how brands are implementing AI as their marketing strategy:

- **Norwegian Air**, an airline company tried figuring out key markets based on various criteria like flight location. They developed a custom ML model to target users within their designated market areas (DMAs), who accounted for customers which are most likely to engage with their airline and complete booking. The campaign overreached Norwegian Air's targets, delivering a "cost per booking" CPA that was 170% lower than the CPA goal. The company's secondary goal is to aware key markets about "World's Best Long-Haul, Low Cost Airline." was also accomplished by this strategy.
- **Starbucks** used predictive analytics by using loyalty cards and user mobile applications to gather and analyze consumer data. This strategic plan of application of AI and big data was presented in 2016. Starbucks delivered personalized marketing messages to customers including recommendations. Their mobile application allowed users to place an order directly from phone via speech commands. This implementation of AI helped the company in increasing annual revenue by 11% from year 2018 in contrast to 2017.
- **Wowcher**, a leading e-commerce company, ramped up its social media in the year 2018 with the usage of AI- powered copy- writing technology. This tool was having the advantage of studying and analyzing emoticons as a part of speech and writing patterns. After this analysis AI engine developed an emoji based social writing which resonated with most users. Wowcher achieved a 31% reduction in cost per lead, as well as a higher Relevance Score as a consequence of this technology.
- **Nestle or Unilever** used AI technology to harmonize insights from a pool of references

like songs and popular media content in public domains for trends in food consumption. Unilever discovered a relationship between breakfast and ice cream. Additional research displayed that brands such as Dunkin Donuts are so far serving ice cream in their breakfast menu and sweet products were becoming popular in breakfast in the US. They considered this as an opportunity and they came up with a variety of cereal flavored ice cream or "Breakfast for Desserts" which became an industry standard.

- **Alibaba**, a retail giant, launched "FashionAI" outfitter in Hong Kong. Its aim was to streamline the customers' fashion retail experience with implementation of AI and using smart garment labels. These labels recognize when the product is grabbed and smart mirrors recommends correlated items along with the garment's description. The subsequent idea of Alibaba is to show a list or virtual wardrobe of all the garments tried or touched by a user during his/her store visit. The employment of new technology by Alibaba is a need of hour in times of consumers' ever-changing expectation. As per a survey conducted 46% of respondents were of the opinion that their encounter with technology instills additional credence in a specific brand showing a trust and positive response.
- **Nike** launched a program wherein customers were provided with the flexibility to design their sneakers in

their own way at the store, in 2017. The technology permits users to design the pattern to be printed on plain Nike shoes, according to their own choice. This is a great gimmick to increase sales of the company and furthermore it helped in collecting huge amounts of information relating to customers' predilection about sneakers. With this information and machine learning algorithms Nike developed designs for future products and also delivered personal product recommendations or messages to users.

- **Amazon** pioneered personal recommendation in early stages of AI Marketing. They developed new algorithms for dynamic pricing of products and in recent years they have launched checkout-free physical stores in San Francisco, Chicago, Seattle which works on AI-powered sensors and detect automatically which item a customer has picked up and charges them automatically of amazon go application.
- **Netflix** provides personalized recommendations to every user with implementation of AI based on what the user likes and what user watches. It analyses customer reactions to different shows, documentaries, films and then after looking into billions of records suggests the best one. Most of the shows watched by users are discovered with this AI based recommendation system.

REFERENCE

- [1] Weber Shandwick Homepage, <https://www.webershandwick.com/wp-content/uploads/2018/04/AI-Ready-or-Not-report-Oct12-FINAL.pdf>, last accessed 2019/12/24.
- [2] Thiraviyam T., Artificial Intelligence Marketing, In: Proceedings of International Journal of Recent Research Aspects, pp. 449–52, April (2018)
- [3] HBR Homepage, <https://hbr.org/resources/pdfs/comm/people.ai/accelerating-sales.pdf> last accessed 2020/02/12
- [4] Mckinsey Homepage, <https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain>, last accessed 2020/02/07.

CONCLUSION

Technological advancements have always helped businesses by creating new opportunities for reaching customers. One of the greatest technologies of our time is Artificial Intelligence (AI)

which is creating quite the buzz in the digital space. Given its potential for storytelling and marketing, Artificial Intelligence in B2B sales and marketing is here to transform the way people interact with brands, information and services.

Artificial intelligence is continually becoming an empowering tool for digital marketers and works on the tricks to get the information from an individual or the group of people to and make your brand a huge one. It is safe to say that Artificial intelligence is a secured investment, which is sure to get dividends.

Transforming Digital Marketing with Artificial Intelligence

Rutuja Yewale, Prof Anita Rajput

Abstract -Artificial Intelligence Marketing is a technique of optimally utilizing technology to enhance customer's experience. In recent times it becomes really important for the companies to recognize and understand customer needs and their expectations in terms of products as well as services. With the help of AI, Marketers can process the huge amount of data, carry out individualized sales and fulfil customer expectations. Besides this, the perception they get about the customer and their requirements in a shorter time frame help them to boost performance and Return On Investment (ROI) rapidly.

In this paper we have provided concept of using Artificial Intelligence in marketing. It also shows benefits of implementing AI in marketing ,and Changing scenario of marketing with introduction AI. Literature Review for the same and Adoption rate.

Index terms : Artificial Intelligence , Machine Learning ,Customer Lifecycle, Marketing

◆

1. INTRODUCTION

Artificial Intelligence is an increasingly popular term that lacks a unified, concrete definition. It is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment. Technically speaking, Artificial intelligence is an integration process in between cloud computing, network devices, robots, computer, and digital content production and in various business processes, systems, and daily life operations. Artificial intelligence computing was in past, today and will be in future. Embracing the increase and development of Artificial Intelligence is crucial to future marketing efforts. Every day, companies are using artificial intelligence software to optimize their own processes, reduce overhead, decrease turn around time, and improve output. Technology is evolving at an

unprecedented rate, and teams already making the move to marketing AI software are at a distinct advantage to jump on the next innovation.

2. CHANGING SCENARIO OF MARKETING

Change is the dominant fact of life in every business today. And the ability to master and exploit change has become one of the most sought-after management skills. Current global and competitive business environment constantly asks for innovation, existing knowledge base is getting obsolete, continuously thriving for advancement in process improvement. The learning curve is always put to test, and every company is striving to remain ahead of the curve. Due to this shift in the way business is getting conducted has thrown out new reality of ever shortening product and service life cycle. More and more companies are coming out with customized products and finding ways to differentiate from competition.

In marketing, the very tempo of change is constantly quickening. The major change in marketing environment is the emergence of electronic data-processing equipment as a major tool of scientific marketing not only for reporting data but also, more importantly, for planning and control by management. Most companies are taking advantage of electronic data-processing analyses, online communications, and information-retrieval systems as tools to help make marketing more efficient. This has led to the emergence of real – time marketing.

Real-time marketing involves creating a marketing strategy focused on up-to-date events. Marketers are making use of current trends and customer feedback, while working to connect consumers with products and services they could use. Real-time marketing is growing in popularity with the rise of social media because it provides businesses with access to real-time information on target audiences; marketers were able to tap into that information and transform it into messages and products that they thought the customer might like. Real-time marketing now makes use of all kinds of customer data to help companies understand exactly how customers behave. Done correctly, real time personalization ensures that you are always serving relevant and timely content to your website visitors, and it can make your website visitors feel like your entire website experience has been crafted specifically for them. That level of personalized attention makes any prospect feel valued, and valued consumers are

far more likely to develop brand loyalty than those who feel like they're simply one more lead.

Today, there's real-time, always-available access to the data and tools that enable rapid analysis. This has propelled Artificial Intelligence and machine learning and allowed the transition to a data-first approach. In the past, AI's growth was stunted due to limited data sets, representative samples of data rather than real-time, real-life data and the inability to analyze massive amounts of data in seconds.

3. LITERATURE REVIEW

Thomas Davenport in his research paper proposed a structure for comprehending the effect of AI specifically on the influence of AI on marketing strategies as well as customers' behavior. He concluded: AI's short- and moderate-term impact may be more limited and if AI augments (rather than replace) human managers it will be more effective.

Neha Soni elucidates the impact of AI on business – from innovation, research, market deployment to future shift in business models. She explains the two major factors responsible for making AI the core technology for extreme automation. Furthermore, she describes the concept of "AI Divide" or "The dark side of AI".

Muhammad Zafeer Shahid did a qualitative research by interviewing marketing professionals from different firms to analyze the parameters that contribute in integration of AI with marketing. His findings highlighted benefits of amalgamating AI in marketing whereas technical compatibility was most challenging.

Gijs Overgoor elaborated how Cross – Industry Standard Process for Data Mining (CRISP – DM) structure can be used for developing AI SEolutions to Marketing problems. He illustrated this idea with an interesting case study of Automatic Scoring Images for Digital Marketing.

N. Ramya in her research reviewed different factors influencing consumer buying behavior. Brands have an opportunity to develop strategy and personalized marketing message by recognizing and comprehending the factor that

that have impacted on their customers.

Nausherwan Raunaque study focuses on the factors which online buyers take into consideration while shopping online. His findings include and how their security and privacy concerns about online marketing influences their online buying behavior.

Dr. Md. Tabrez Quasim presents different types of forecasting and AI techniques that are useful in business forecasting. In his research he examined some present approaches of AI that seemed to be beneficial and promising for business forecasting.

Albert Annor-Antwi states that AI when integrated with other technologies ML, Big data, Data Analytics would yield more accurate results than any other forecasting method. He elaborated shortcomings of AI, being costly and cybercrime threats to AI- powered forecasting.

Naresh K. Malhotra in his study amalgamates both academic as well as practitioner outlook in order to examine the issues and emerging trends that will shape the role of marketing research.

The objective of the research by Dhanushanthini Ajanthan is to examine and identify the effect on brand equity who's major dimensions are: Brand loyalty, brand awareness, brand image and perceived quality of social media marketing.

Neil A. Morgan in his research not only unveils the important challenges to marketing strategy research but also uncovers numerous opportunities for generating highly pertinent modern and creative theories.

Anyuan Shen did an investigative research of customers' lived experiences of commercialized recommendation. According to him, researchers may discover a new standpoint – the customer's outlook – with a view to analyse theories of personalized marketing.

Jerry Vesanen through his research paper makes the meaning of personalization clear to maximum extent. Findings reveal that personalization is not yet widely applied as it is not clearly understood by majority of marketers as it has different meanings for different businesses.

Ki Youn Kim did a study with aim to qualitatively recognize topologies characteristics of the big data marketing strategy. She focused impacts of Big Data Analytics in business perspective, with execution of Q methodology.

Alexandra Amado evaluated the implementation of Big Data in Marketing, with the goal to spot trends in these fields. His findings exhibited that theirs is and increasing interest for Big Data in marketing over the years. Hence it is necessary to expand efforts in regard with the business so that Big Data can flourish in the Marketing sphere.

Stefan Lessman research focuses on empirical targeting models. Author argue adduce about standard practices to thrive such models do not gauge sufficiently for business objectives.

Ayse Bengi Ozelik elaborates about psychological interpretation or behavior of recipients in the effectiveness of online ads customized using behavioral targeting.

Aman Abid tested the impact of different content characteristics and cues on followers' online expression and understood the moderating consequences of content curation. He evaluated role of marketer generated content in building online relationships.

Samira Khodabandehlou developed a framework for customer churn prediction, which included six stages for precise prediction and averting customer churn in business. Author found that the discount, reception of returned items, prize, number of items and distribution time came best predictor variables apart from frequency and monetary (RFM) variables.

Jean Paul Simon provided an outline of the key trends in the domain of Artificial Intelligence and a global overview about regions along with companies. Author says that legal, ethical, socio-economic aspects (ELSE) can turn into barriers to deployment of AI technologies and explains how demand seems to be uncertain for AI on both sides i.e. business and consumer.

Thanos Skouras lay out a collation as well as assessment of reception of pricing by disciplines of economics and marketing. The major reasons for differences in both approaches are dissimilar historical origin, key concerns and doctrinal evolution

APPLYING ARTIFICIAL INTELLIGENCE ACROSS CUSTOMER LIFECYCLE

Analyzing and predicting consumer journey is troublesome for marketers. Consumers convey their opinion in the form of blogs, Tweets, “likes,” videos, search, comments and conversations and through various channels. In figure 4 customer interaction and value are displayed over a period of time along with the role of machine learning, propensity modelling, AI application during different stages. Customer experience is a competitive driver of growth when successful and the greatest source of risk when failing.

Machine Learning comprehends the analysis of historical data from various business interactions with audiences and their responses. This data helps in identification of the success factors of a person’s communications, including targeting, offers, copy and frequency. Algorithms for Machine Learning generate insights via predictive analytics, and it depends on marketers to take actions according to those insights or set certain directives for AI to act on them.

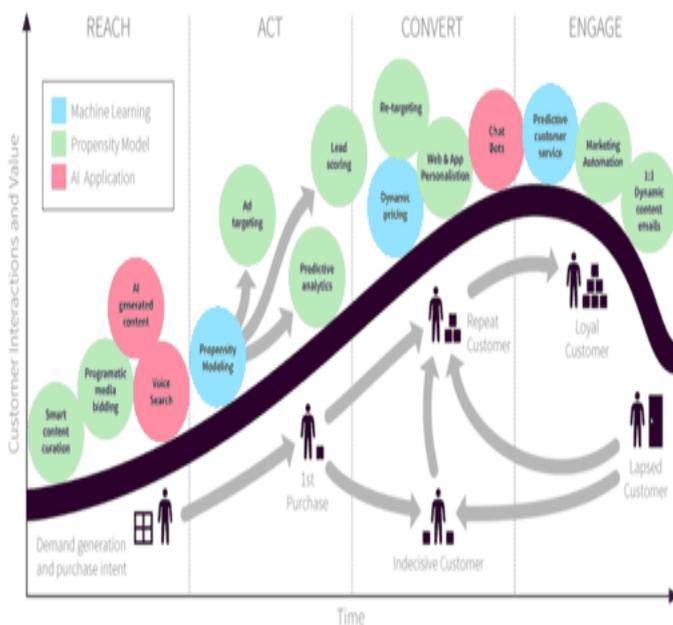


Fig 1

Estimation of the possibility of subjects executing specific types of behavior (e.g. purchasing of a product) by taking into account independent covariates and confounding variables that affect

such behavior is known as propensity modelling.

The four major stages of customer lifecycle are explained as follows:

1. Reach

For marketers, sometimes it is challenging to track when a need is triggered among customers as it eventuates at category level rather than the brand level. A key marketing scheme is when consumers initiate deliberate possible offerings to match up to their requirements, to inculcate the brand into customers' consideration set. Marketers job comprises of expanding visibility of the brand and intensifying crucial grounds for consideration.

2. Act

After customers pen down their brand preferences, marketers pitch to implant certainty in the offerings as well as to coax them that they are making best choices. AI help in accomplishing these objectives by predictive analysis, lead scoring, learning from consumers behavior and providing real time insights. Marketers can also employ emotional AI to perceive consumers' feelings about their brand publicly.

3. Convert

After consumers deduce the worth of their preferred brand and the amount, they are willing to spend, marketers shift their focus from decision process to action process. They achieve this task by strengthening brand’s value juxtaposed with its competitors. The purchase process of the consumers can be completely transformed with help of AI. Marketers implement intelligent purchasing systems for their consumers and can also regulate the “sweet spot” for pricing known as dynamic pricing.

4. Engage

During this stage, consumers are in a position to evaluate their interest and satisfaction with a particular brand and can scrutinize whether to repurchase from that brand i.e. engaging with name of brand. The post purchase services are one of deciding factors and marketers using AI enable chatbots for better customer service. Marketers can segment their most valuable customers or loyal customers, and they can

further focus in customer- relationship management campaign.

5. REGIONAL ANALYSIS (ADOPTION RATE)

AI is populated on such a big scale that the total revenue for this technology is expected to increase from 7.4 billion US Dollars in 2018 to 89.85 billion in 2025 . Across the globe marketers have discovered two major uses of AI Marketing which are task automation and data analysis. With increase in essence of customer data for marketers and marketing strategies, most of global marketing professionals are shifting or planning to shift for use of AI

Below mentioned are data of AI Marketing used across different regions across the globe. 51% of marketers are already using AI, while 27% of them are planning to incorporate it within their digital marketing strategy .The global market for AI Marketing is estimated to grow at a significant rate between the forecast period of 2018 - 2023. Table 1 and graph 5 shows the execution of AI Marketing for various purposes in different regions across the globe.



FIG 2

6 .CONCLUSION

AI Marketing is new playbook for Marketers which is making them shift from marketing automation to marketing personalization. The need for customization, reactive design and dynamic

engagement have been long talked about and introduction of AI has served as catalyst to initiate this required marketing transition. There has been a growth in scope of AI with regard to its application in marketing. This marketing aeon can be considered both interesting as well as challenging and frightening. With the advent of artificial intelligence marketing, automated and traditional marketing techniques took a backseat and things like personalization, speech and image recognitions, chatbots, churn predictions, dynamic pricing and customer insights came into the vision. Availability of wide range of data has made it possible for the marketers to carry out individualized sales and marketing and fulfil customer expectations to the maximum extent.

Currently, AIM is in its infant stage and with its fast pace, it is expected to reshape marketing strategies and business models. Some market research topics may become insignificant as these roles will be taken over by machines and advanced jobs will be created, which will require high potential and knowledge. In near future AIM is likely to substantially change both marketing strategies as well as customer behavior. AIM will surely explode up with a lot more surprises in the coming future.

7. REFERENCES:

- <https://intotomorrow.com/how-ai-is-changing-digital-marketing>
- [.https://digitalmarketinginstitute.com/blog/2018-01-07-how-ai-is-changing-digital-marketing](https://digitalmarketinginstitute.com/blog/2018-01-07-how-ai-is-changing-digital-marketing)
- Ajanthan D. (2017), The Impact of a Social Media Marketing on Customer Based Brand Equity- A Special Reference to Travel & Tourism Industry in Sri Lanka, TRANS Asian Journal of Mar.

Image Processing Based Signature Verification Technique to Reduce Fraud in Financial Institutions

Vaibhav S. Wandkar
Currently pursuing masters degree
Program in M.scIT in Mumbai university, India
E-mail:vaibhavsw20mscit@student.mes.ac.in

Guide by Prof.Anita mohan Rajput
E-mail:anitamohanrajput@mes.ac.in

Abstract. Handwritten signature is broadly utilized as personal verification in financial institutions ensures the necessity for a robust automatic signature verification tool. This tool aims to reduce fraud in all related financial transactions' sectors. This paper proposes an online, robust, and automatic signature verification technique using the recent advances in image processing and machine learning. Once the image of a handwritten signature for a customer is captured, several pre-processing steps are performed on it including filtration and detection of the signature edges. Afterwards, a feature extraction process is applied on the image to extract Speeded up Robust Features (SURF) and Scale-Invariant Feature Transform (SIFT) features. Finally, a verification process is developed and applied to compare the extracted image features with those stored in the database for the specified customer. Results indicate high accuracy, simplicity, and rapidity of the developed technique, which are the main criteria to judge a signature verification tool in banking and other financial institutions.

1 Introduction

A signature is a gained behavioural biometric of a user to declare his/her unique identity on printed documents. The demand of authorization based on signature is increased including credit card validation, security systems, banking system, checks, contracts, etc., as shown in figure 1. It is widely used as proof of identity and a socially accepted authentication method in daily life. The system stakeholders are person, organization or banks that need to verify signatures [1, 2]. The stakeholders are Bank's customers who must write their signature, and bank's employees have to verify if the sample signature is the original signature in database, to complete any transaction required on that account. Another customers are organizations' employees: any organization that still depend on paper works, employees must take supervisor's signature [3, 4].

Automatic signature verification system compete the current visual verification that depends mainly on the experience, mood and working environment of the verifier. Moreover, it is difficult for the eyes of any experts to precisely verify the ratios between lines and angles of a genuine signature to a fraud signature [5]. One reason is that signature is just a special way of handwriting that contains complex geometric patterns and often unreadable plots [6]. A signature forgery is replicating the genuine signature by the forger after careful practice. This type of forgeries is called the skilled forgery which harden the signature verification task. The other two types are random forgery where the

forger does not know the shape of the original signature, and the simple forgery where the forger knows the shape of the original signature but does not practice enough to increase the similarity value between the fraud and the genuine signatures.



Fig. 1. An example of a handwritten signature to prove user identity on a banking check. Authentication of this signature is the key process to provide the user with an access to his/her bank account.

Automatic verification systems that authenticates the person's signature can be categorized as two types, an online (dynamic) and an offline (static) system [6, 7]. In online systems, dynamic data can be obtained from an online user display suchlike electronic tablet with an instructed pen and in this case, the input is a sequence of dynamic features about the user writing activity such as

the applied pressure, speed of writing, etc. On the other hand, in the offline systems, signatures are written in a paper which is processed as two dimensional image and has been converted to the system with the aid of scanner or camera [8, 9]. The signature verification architecture usually starts by extracting the features of the genuine signatures followed by classification of a set of genuine and skilled test signatures. The offline features can be categorized as Global and local features, the Global features describe the image as whole the image size, while local features are commonly extracted by partitioning the image into a grid and extract the features in each of its parts [10, 11]. Lately, Interest points are picked using SIFT (Scale-Invariant Feature transform) [SIFT] and/or SURF (Speed-up Robust Features) [SIFT] to perform the signature verification task [12, 13]. These models extract the interest points in each image, then extract the features/descriptors for each interest point to verify the matching between signatures. The accuracy of the solution of these methods depends on the number of matches of the genuine signatures to other genuine and forged signatures [14, 15].

The work proposed in this paper uses a different approach in utilizing the SIFT/SURF extractors, where the matching depends on the summing up of the Euclidian distance between the interest points in the two images. The work applied here is based on the database of offline genuine and skilled forged signatures extracted in the work in [16] and in [17]. The results shows 95% classification accuracy which is higher than that of current research.

The rest of this paper is organized as follows: Section 2 presents the previous work of applying machine learning tools to perform the signature verification task. Section 3 describes the proposed approach, while the experimental work and discussion appear in section 4 and finally the conclusion is in section 5.

2 Materials and Methods

The signature image is initially pre-processed to facilitate the job of the feature extractor technique. The pre-processing includes the cropping of the signature area, removal of the noise, banalization of coloured image to grayscale one and finally edge detection or smoothing of the signature lines. Features are categorized into two types, global and grid features. Global features define the entire structure of the signature like the height, height-to-width ratio, and the image projection. Maximum horizontal and vertical projections represent the row and column that contain the maximum number of black pixels respectively. The second type of features is the grid features where a virtual grid is created of 12x8 segments for describing the detailed parts of the image. Pixels density, pixels distribution, and predominant axial slant are examples of the extracted grid features.

The training is applied on a set of genuine signatures using either a simple distance calculation method or a complex machine learning techniques. This approach is considered a unique method that multiple features are extracted which have global features such as statistical

features, image gradient came from distribution of pixels in a signature image and descriptors, the classification contains variation between signature of the same user and done a distribution in distance space. For any tested signature the method gains a distribution which is compared with the saved distributions and a similarity between them is obtained. This method does not utilize set of forgery signature in the training. The approach utilizes the geometric center for feature extraction. This center gains both horizontal and vertical of the signature. The classification and description is done by Euclidean classifier which defines vectors between two signatures. This method tested by database of 20 writer, 10 signatures for genuine and 10 for forgeries. The method achieved 11.4% AER. The use of complex machine learning techniques like Hidden Markov model, support vector machine and neural network show better results in the domain of Hand-written signature verification. Lately, SIFT and SURF detectors and extractors show better results in the field of offline signature verification.

2.1 SIFT/SURF algorithms

Scale Invariant Feature Transform (SIFT) is powerful and successful approach to do feature detection. Speeded Up Robust Features (SURF) is based on the same steps and principle but it uses different schema and provides better and faster result. The SURF algorithm is divided into main two steps: firstly, interest points are detected. Secondly, interest point description is performed. Both of these steps depend on a scale space representation.

SIFT algorithm: SIFT Generates for the image, another images of 1/2, 1/4 and 1/8 of the original sizes. Then for each image apply the Gaussian blurring operator on the image of intensity function $I(x, y)$ as shown in equation 1.

$$L(x, y, kn \blacklozenge a) = G(x, y, kn \blacklozenge a) \blacklozenge I(x, y) \quad (1)$$

Equation 1 is applied for different scales for all image sizes (octave), such that $o = kn \blacklozenge o$ values where $k = 1.2$ and $n = 0, 1, 2, 3, 4$. Then apply the lablacian of Gaussian approximation, DOG, for each two successive layer in each octave as shown in equation 2.

$$DoG(x,y) = (L(x,y, kn \blacklozenge a) - L(x,y, kn-l \blacklozenge a)) \quad (2)$$

Then extract the potential interest points, by detecting the local extrema for each point in 3x3x3 neighbourhood window. Remove all points smaller than a specific threshold, and accept only points that match with the Harris corners. Finally, for each point of interest, 128 descriptor is extracted. It shows the orientation and magnitude of the 4x4 grids surrounding the interest point in the eight 45-angled regions.

SURF algorithm: Meanwhile, SURF excludes the step of creating different images in its processing. It uses Hessian matrix that expresses the local changes in area of each point in x and y direction as shown in equation 3.

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{bmatrix} \quad (3)$$

Where, $L_{xx}(p, \sigma)$ is the convolution of the image with the second derivative of the Gaussian as displayed in equation 4.

$$L_{xx}(p, \sigma) = I(p) = \frac{\partial^2 g(\sigma)}{\partial x^2} \quad (4)$$

A variable sized filter of a corresponding scale σ is used to form a scale space images. Afterwards, based on the sum of Haar wavelet responses, construct a square window centered around the interest point. This is to select dominant orientation of interest points (x,y) weighted Haar-wavelet responses (dx,dy) . This allows the selection of only a single $\pi/3$ region along each of the selected orientation, out of the 8 regions in SIFT. The 4 values of orientation dx,dy and magnitude $|dx|,|dy|$ of each point (x,y) is considered for the 4×4 grids in the 16×16 window around the interest point, which will form, $4 \times 4 \times 4$, 64 descriptors vector for each feature.

3 The proposed signature verification model

The first step in the proposed verification system is the pre-processing of the image. In image pre-processing step, training and testing signature images are passed into this step. The main purpose of this phase is to make signature image ready for extracting features. The pre-processing phase includes grey scale conversion and binarization.

The image is converted into grey scale: Binary image is converted into grey scale to make process easier and feature extraction more accurate.

Binarizing image: In this step, grey scale image is converted to black and white image using a reasonable threshold. This will make signature clearer as it will be in black and the background will be in white. Threshold filtering is considered the simplest way to binarize an image. Threshold filtering does image binarization by using particular threshold value and intensities of image pixel. All pixels in the signature image with intensities equal all higher than specified threshold value will be converted to white otherwise, all pixels with intensities lower than threshold value will be converted to black which will make signature image black and white. The filter accepts grey scale images for processing, and the optimal threshold values differ from image to image and are obtained after deep investigation of the trained image. The next step is applying the SURF / SIFT feature extraction method as described above. The final step is the verification/matching method. This is done by calculating the Euclidean distance between the descriptors. For every key point in the descriptor in the original image, two closed neighbours are computed and the distance will be taken to match between them and find similarity. The matching of each two signatures is

applied based on the number of key points in each one as shown in figure 2.

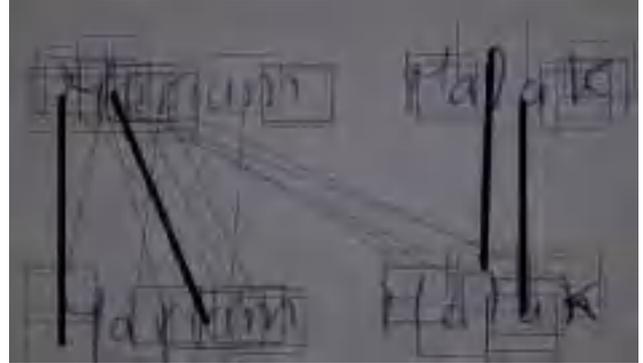


Fig. 2. Matching of features between two genuine signature images

4 Experimental Results

4.1 Accuracy of the proposed model

The proposed model is applied on two benchmark datasets available in [16] and [17]. For the first dataset [16], the results in figure 3 and 4 show that, the first 4 points are genuine-genuine matchings, while the rest of the points are genuine-forged matchings. In this test, 30 forged signatures are tested for verification of signatures of user one and user two. The defined threshold here is applied as the two points out of four which have the lowest Euclidian distance. Accordingly, the results shows nearly 85% classification accuracy of proving the forgery of user one and 92% classification accuracy for user two, as displayed in figure 3 and figure 4; respectively. Note that these results are based on the Euclidian distance values, not based on the number of matchings.

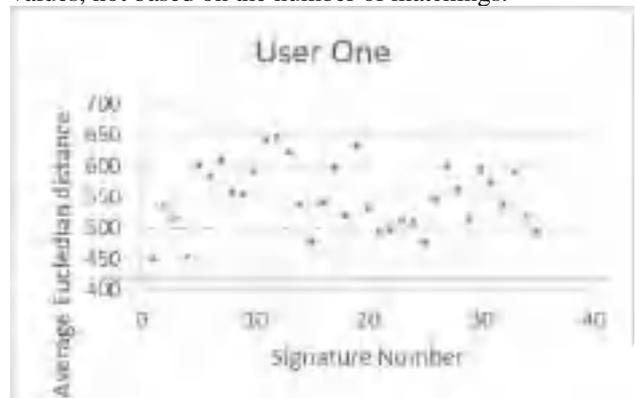


Fig. 3. Genuine-Genuine vs Genuine-Forged signatures for user one.

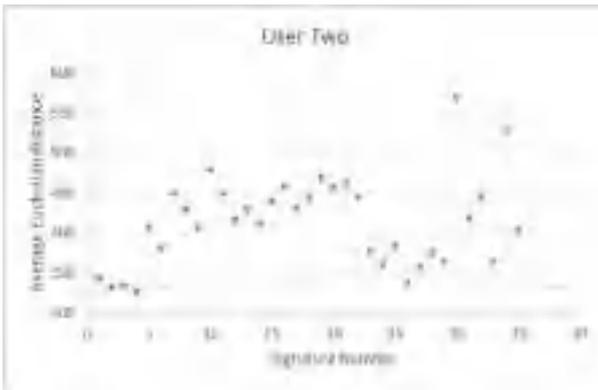


Fig. 4. Genuine-Genuine vs Genuine-Forged signatures for user two.

The same test is applied on the second dataset [17], but is applied by comparing between the first two genuine signatures and between the first genuine signature and another 3 forged signatures, the results are 100% verification accuracy of the forged signatures.

4.2 Parametric investigation of the model

A parametric study is applied in this work by changing threshold value in SURF detector. SURF detector depends on the Hessian matrix to find all interest points in a particular image. SURF divides the signature image into using second order Gaussian kernel and computes these kernels with box filter. The main function in EmguCV library is to detect interest points using SURF with Hessian threshold value. Experimental results show that best value for hessian Threshold could be from 300 to 500.

The SURF detector considers those features in the signature image whose hessian is larger than a specified hessian threshold. Therefore, as high specified threshold value, as less key points in the image will be taken by detector but with more accuracy. A low specified threshold, high key point will be taken.

When the Hessian threshold is 500, key points are 329 and when it is 100, key points are 1004. 1004 key points means many unnecessary key points is extracted from the feature, many calculation will be done which will increase the computational time of the algorithm and feature extracting will be not accurate. Therefore, after testing multiple signatures with different hessian threshold values best value for hessian threshold could be from 300 to 500.

The chart below in figure 5 shows the relation between hessian threshold value with the accuracy and the number of the detected points. Therefore, the relation of hessian threshold value and number of interest point is inversely proportional. as shown in fig.5, higher hessian threshold value will result in less key points and lower hessian threshold will result more key points.

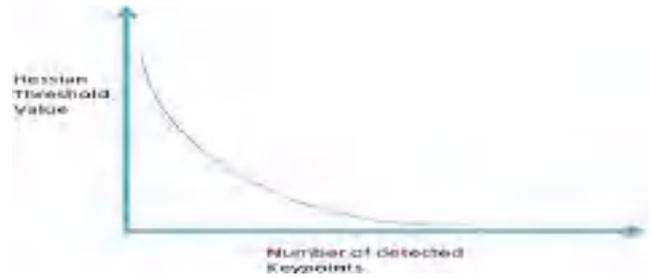


Fig. 5. Relation between hessian threshold value and Number of interest points.

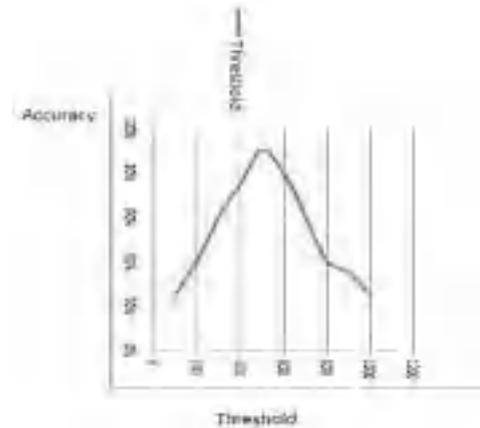


Fig. 6. Relation between hessian threshold value and Accuracy of SURF detector.

As shown in fig.6, after testing multiple signatures with different hessian threshold value shows that the best accuracy of the software is when hessian threshold value between 300 and 500. As shown in the figure 5 above, higher hessian threshold value high accuracy and lower threshold value low accuracy until 500 after hessian threshold; value bigger than 500, gives less detected points in the signature image which will make accuracy of the software less so the best value of hessian threshold is 500.

5 Conclusion

Although the existence of an automatic signature verification tool is necessary, it is not yet applied in most of the financial institutions. The reason is that most of the currently available tools work with a highest accuracy of ca. 80%, which makes them not reliable in the verification task. For many years, researchers are trying to develop more robust signature verification tools using the advances in image processing algorithms.

The main objective of this work is to offer an economically and efficient offline handwritten signature verification system, in order to achieve the objective multiple methods have been reviewed and surf features algorithm is used in this paper as strong image descriptor. Databases of signatures were collected and saved containing known writer's signatures. The proposed model was successfully verified signatures of users with a

lowest accuracy of 85%, indicating its promising implementation and making a room for more improvements to be researched and investigated.

The future work of the current study is to enhance the feature extraction step of the algorithm by adopting features related to cross correlation, and signature energy and skewness. Finally, an automatic feature extraction tool may be developed to predict the relevant features, which define each signature and reduce the verification effort.

References

1. J. Chambers, W. Yan, A. Garhwal, and M. Kankanhalli (2015) Currency security and forensics: a survey, *Multimed Tools Appl* vol. 74, pp. 4013-4043.
2. Bhanu Priya Taneja, Navdeep Kaur, (2015) Biometric System Based on Off-Line Signatures, *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4(5) , May 2015.
3. Maykin Warasart and Pramote Kuacharoen (2012) Paper-based Document Authentication using Digital Signature and QR Code, 2012 4TH International Conference on Computer Engineering and Technology (ICCET 2012).
4. Vijaypal Singh Dhaka, Mukta Rao, Manu Pratap Singh (2009) Signature Verification on Bank Checks Using Hopfield Neural Network, *KARPAGAM Journal of Computer Science*, vol 3(4).
5. Ilkhan CüceloXlu, Hasan Ogul (2014) Detecting handwritten signatures in scanned documents, 19th Computer Vision Winter Workshop, Krtiny, Czech Republic, February 3–5, 2014.
6. Biswajit Halder, Rajkumar Darbar, Utpal Garain, Abhoy Ch. Mondal (2014) Analysis of Fluorescent Paper Pulps for Detecting Counterfeit Indian Paper Money, *Information Systems Security*, series Lecture Notes in Computer Science, vol. 8880, pp 411-424.
7. Jain, U.A.; Patil, N.N. (2014) A comparative study of various methods for offline signature verification, in *Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014 International Conference, pp.760-764, 7-8 Feb. 2014.
8. Anu Rathi, Divya Rathi, Parmanand Astya (2012) Offline Handwritten Signature Verification By Using Pixel Based Method, *International Journal of Engineering Research & Technology*, vol. 1(7), September - 2012.
9. Saroj Ramadas, Geethu P.C (2015) Comparative Study On Offline Handwritten Signature Verification Schemes, *International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)*, vol. 2(10), March 2015.
10. Karrar Neamah, Dzulkifli Mohamad, Tanzila Saba, Amjad Rehman (2014) Discriminative Features Mining for Offline Handwritten Signature Verification, *3D Res* (2014) 5:2
11. Roy, Sayantan, Maheshkar Sushila (2014) Offline Signature Verification using Grid based and Centroid based Approach, *International Journal of Computer Applications*, vol. 86(8), pp. 0975–8887, January 2014.
12. Javier Ruiz-del-Solar, Christ Devia, Patricio Loncomilla, Felipe Concha (2012) Offline Signature Verification Using Local Interest Points and Descriptors, *Progress in Pattern Recognition, Image Analysis and Applications*, of the series Lecture Notes in Computer Science, vol. 5197, pp. 22-29.
13. Malik, M.I.; Liwicki, M.; Dengel, A.; Uchida, S.; Frinken, V. (2014) Automatic Signature Stability Analysis and Verification Using Local Features, in *Frontiers in Handwriting Recognition (ICFHR)*, 2014 14th International Conference, pp. 621-626, 1-4 Sept. 2014.
14. Rajpal Kaur, Pooja Choudhary (2015) Offline Signature Verification in Punjabi based on SURF Features and Critical Point Matching using HMM, *International Journal of Computer Applications* , vol. 111(16), pp. 0975–8887, February 2015.
15. J. Ruiz-Shulcloper and W.G. Kropatsch (2008) Signature Verification Using Local Interest Points and Descriptors, *CIARP 2008, LNCS 5197*, pp. 22–29, 2008.
16. Javier Galbally, Moises Diaz-Cabrera, Miguel A. Ferrer, Marta Gomez-Barrero, Aythami Morales, Julian Fierrez (2015) On-line signature recognition through the combination of real dynamic data and synthetically generated static data, *Pattern Recognition*, vol. 48(9), September 2015, pp. 2921–2934.
17. Marcus Liwicki, Michael Blumenstein, Elisa van den Heuvel, Charles E.H. Berger, Reinoud D. Stoel, Bryan Found, Xiaohong Chen, Muhammad Imran Malik (2011) SigComp11: Signature Verification Competition for On- and Offline Skilled Forgeries, in *proceedings of 11th Int. Conference on Document Analysis and Recognition*.

OFFLINE FORGERY DETECTION OF HANDWRITTEN SIGNATURE USING GAUSSIAN EMPIRICAL RULE

Vaibhav S. Wandkar
Currently pursuing masters degree
Program in M.scIT in Mumbai university, India
E-mail: vaibhavsw20mscit@student.mes.ac.in

Guide by Prof. Anita mohan Rajput
E-mail: anitamohanrajput@mes.ac.in

ABSTRACT - Signature authentication is most widely used in verifying a person's identity. In this paper Global and Geometric features are discussed. Before extracting the features, preprocessing of a scanned image is necessary to isolate the region of interest part of a signature and to remove any spurious noise present. The system is trained initially with the data-set of signatures obtained from those individuals whose signatures are to be authenticated by the system. All the features are computed for training samples of signature. There are some variation itself in features of genuine set of signatures. If testing signature sample satisfies the Gaussian empirical rule it is authenticated as original signature otherwise a forged one.

KEYWORDS: Global and Geometric features, Gaussian empirical.

1 INTRODUCTION

A signature is a signatory or signer of a person identity and it may be confused with an autograph, which is chiefly an artistic signature. This can lead to confusion when people have both an autograph and signature. Verification can be achieved either Offline or Online.



Figure 1 A sample signature image

Signature Verification system can be generally divided into two categories: A static method (called as Offline) that extracts shape related information as in Figure 1. It is generally used for lower authentication needs. Off-line data is a 2-D image of the signature. Processing Off-line signature is complex due to the absence of stable dynamic characteristics. Difficulty also lies in the fact that it is hard to segment signature strokes due to highly stylish and unconventional writing styles. A dynamic method (called as Online) with time related information that extracts dynamic features such as speed, pressure, time information which cannot be imitated. It is used for higher authentication. In present system a signature sample is taken and its global and geometric features are calculated and are verified using Euclidean distance to authenticate whether the sample considered is original or a forged one.

1.1 FORGERY AND ITS TYPES

The action of forging a copy or imitation of a document, signature, banknote, or work of art. Forgery is one the techniques of fraud including identity theft with a produced altered objects of another's rights illegally.

1.1.1 Types of forgery

The main task of any signature verification system is to detect whether the signature is genuine or not. The forgeries signature is difficult to obtain the instrument and the results of verification depend on the type of the forgery as in Figure 2.

Basically there are three types:

1. Random forgery: A signature sample that belongs to a different writer
2. Simple forgery: A signature with the same shape or the genuine writer's name
3. Skilled forgery: A Signature is written by a person who had access to a genuine signature for practice.

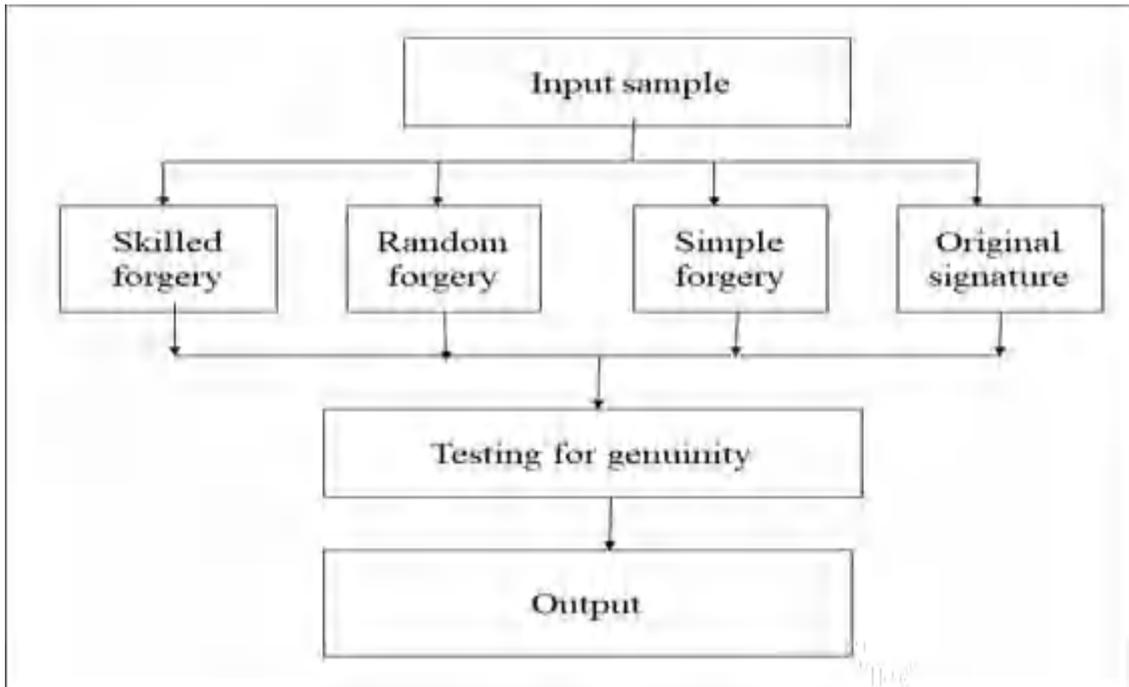


Figure 2 Forgery types

2. BACKGROUND WORK

2.1 Steps Involved In Preprocessing

The preprocessing steps are very much necessary for the signature authentication as without the region of Interest is very difficult to cross verify the steps involved in preprocessing are shown in the below section.

2.1.1 RGB to Gray Scale conversion:

The first pre-processing step is to convert an RGB image in Figure 3 to a Gray Scale image in Figure 4 that means conversion of a true color image RGB to the grayscale intensity image.



Figure 3 RGB Image Figure



Figure 4 Gray Scale Image

2.1.2 Binarization:

Binarization means converting a gray scale image into a binary image. A binary image as in Figure 6 is a digital image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white though any two colors can be used. Binary images are also called bi-level or two-level. Statistical analysis of gray-level images may include determination of mean, variance, standard deviation, contrast stretch, histogram etc. or it can be a combination of any of these Determination of a threshold value is very much important and perhaps the most sensitive part of any image binarization scheme because a wrong value of threshold may result in losing some image information (an object can be considered as part of background and vice versa).

$$Q=S/(X*Y) \tag{1}$$

where

- Q=Threshold Value
- S=Sum of all the pixel value in the image
- X=height of the image
- Y=width of the image

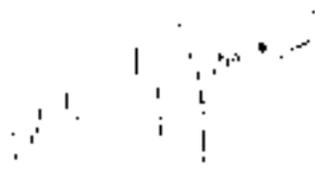


Figure 5 Original Image

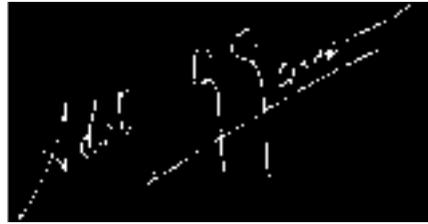


Figure 6 Binarized Image

2.1.3 Edge Detection:

Edge detection is a method of identifying edges in a digital image at which the image brightness changes sharply. In this paper we are using the canny edge detector is an edge detection operator that uses to detect a wide range of edges in images.

2.1.4 Noise removal:

Image noise is a method of removing noise from image. In this for removing noise from an image as in Figure 7 after applying Median filter as in Figure 8 that removes noise efficiently.



Figure 7 Image with noise

Figure 8 after applying Median Filter

2.1.5 Region of Interest:

Cropping is done with respect to bounding box of image by calculation the first foreground row, first foreground column, last foreground row, last foreground column to obtain region of interest.



Figure 9 Image showing region of interest

2.2 FEATURE EXTRACTION

The objective feature of this phase is to extract the features of the test image that will be compared to the features of training image for verification process. There are two types of features ie. Global and Geometric feature [3]. The Global features are Height, Width, Aspect ratio, Area of black pixels and Normalized Area. The Height is the distance between two points in the vertical projection and must contain more three pixels of the image for a binary image. For a binary signature image, width is the distance between two points in the horizontal projection and must contain more than 3 pixels of the binary image. Aspect Ratio is defined as ratio of width to height of a signature. Area of black pixels can be

defined as the number of black pixels in the image. The normalized area of black pixels can be defined as ratio of number of black pixels to the total number of pixels in the image. The Geometric features [8] are Mean, Variance and Standard deviation. The mean of a data set is simply the arithmetic average of the values in the set, obtained by summing the values and dividing by the number of values.

$$\mu = \sum X/N \tag{2}$$

In a data set is the arithmetic average of the squared differences between the values and the mean

$$\sigma^2 = (\sum (X-\mu)^2)/N \tag{3}$$

The standard deviation is simply the (positive) square root of the variance.

$$\sigma = \sqrt{((\sum (X-\mu)^2)/N)} \tag{4}$$

3 PROPOSED WORK

Gaussian distribution can accommodate around 99.7% of features or observations which fall within three standard deviation of the mean and which is between $\mu-3\sigma$ and $\mu+3\sigma$. Feature points are considered as the features extracted from query signature images while mean and standard deviation are found from signature database. Each feature point is tested within the range of three standard deviation of the mean. The equation below is used to select some important feature points from a given signature.

$$|\mu - x| \leq k * \sigma \tag{5}$$

where

μ and σ are mean and standard deviation

x is the value of some feature extracted from signatures

k can be 1, 2, 3 and derived from Gaussian empirical properties to test the closeness of the current query sample to the distribution mean of database.

The purpose of forgery detection is to design a program through which we can detect a forged signature from that of an original one. The existing algorithms are used to detect forged signatures but the false acceptance rate and false rejection rate are high. In order to reduce these para

meters we use Gaussian Empirical Rule [5].

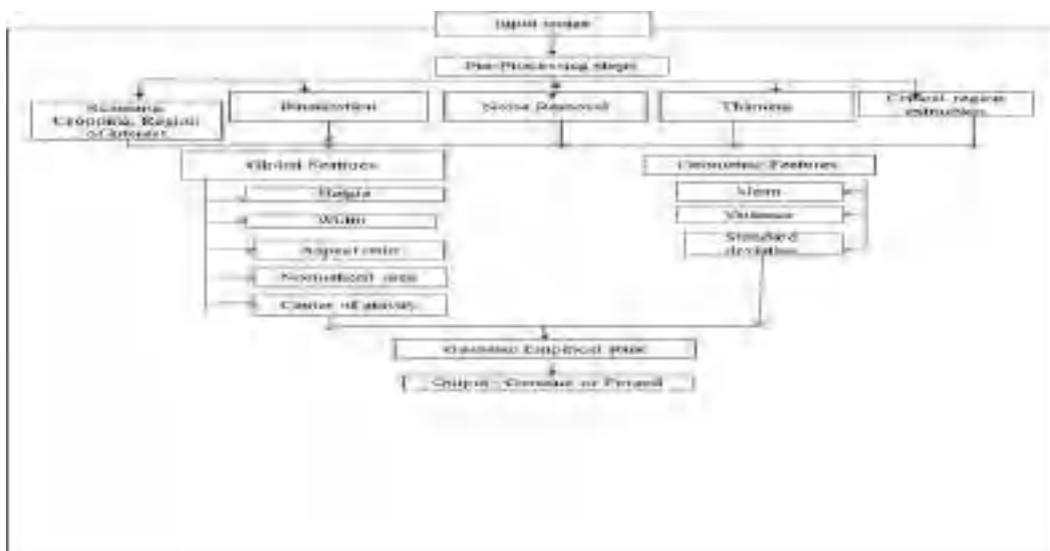


Figure 10 System Architecture

3.1 ALGORITHM STEPS FOR GAUSSIAN EMPIRICAL RULE

The Gaussian rule[5] is applied to the query signature sample for which all the global and geometric features are calculated. The geometric features are considered from that of the corresponding original signature of query signature that is already available in the trained set as shown in Figure10. Trained set of Signature originals should be collected

from various persons. Based on these trained signatures the query signatures are authenticated whether they are genuine or forged one. In Gaussian Empirical rule implementation as explained in equation 5. The geometric features mean and standard deviation are considered from trained set itself.

Pre- processing steps are

Consider an input query signature.

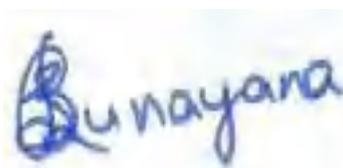


Figure 11 Query Image

In Figure 11 query is now to undergo the pre-processing steps like RGB to Gray scale image, binarization, edge detection, noise removal and region of interest.

Step 1: The Gray image is obtained as follows

Step 2: Binarized Image and Edge detection (using canny edge detection algorithm).

Step 3: Noise free (Using median filter to remove salt and pepper noise from the image).

Step 4: Region of interest by cropping.

Step 5: Extract the global features like height of the image, width of the image, aspect ratio, area of black pixels in the image and its normalized area.

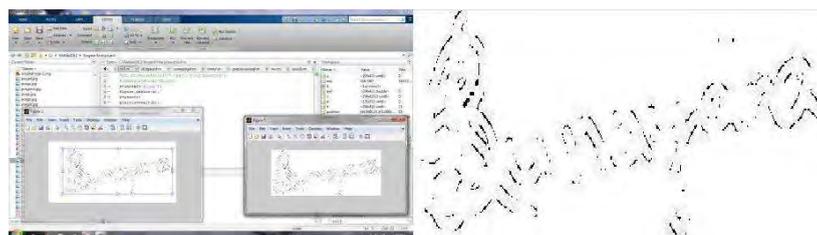


Figure 12 Preprocessing Steps

4 EXPERIMENT RESULTS

4.1 INPUT:

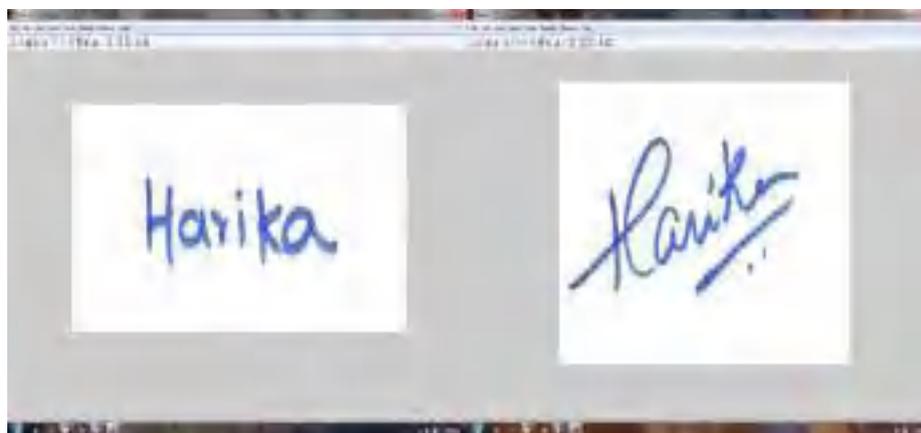


Figure 13 A query sample

Figure 14 The trained sample

OUTPUT:

The query sample when authenticated using Gaussian empirical rule is resulted as a forged one.

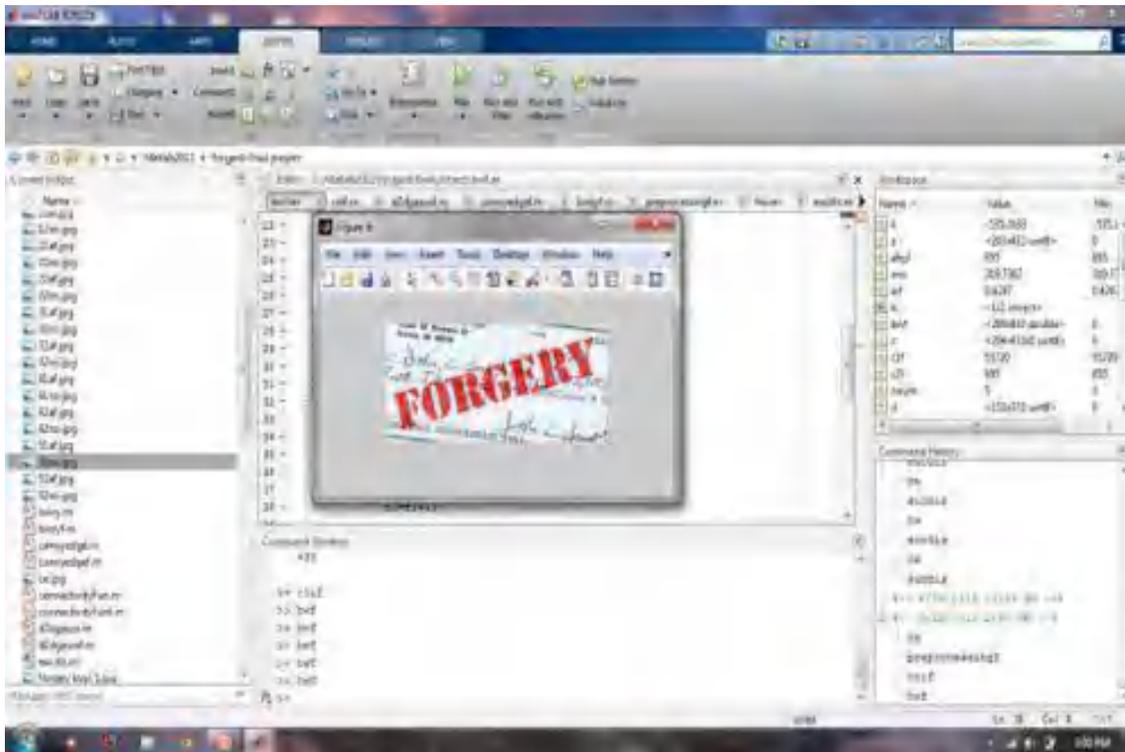


Figure 15 Output obtained after Gaussian rule implementation

4.2 COMPARISON OF TEST RESULTS:

The comparison of outputs is done using proposed system characteristics and existing system characteristics. The characteristics involved in this process are False Acceptance Rate and False Rejection Rate. False acceptance rate can be defined as ratio of the number of forged signatures authenticated as original to the total number of forged signatures. False rejection rate can be defined as ratio of the number of original signatures authenticated as forged to the total number of original signatures.

Table 1 Proposed System Characteristics

Types of Signatures	False Acceptance Rate (FAR)	False Rejection Rate (FRR)
Original	-----	0
Forged	0.13	-----

Table 2 Existing System Characteristics

Types of Signatures	False Acceptance Rate (FAR)	False Rejection Rate (FRR)
Original	-----	0.6
Forged	0.6	-----

5 CONCLUSION

The results of Gaussian Empirical Rule are observed and the implementation is better than that of Euclidean distance in authenticating the signatures. In Gaussian empirical rule, each feature is compared with the multiple of standard deviation and if all the features satisfies then it is authenticated as original otherwise forged but where as in Euclidean distance, an acceptance range is considered and authentication of signatures is decided which varied the FAR and FRR.

By comparing the two characteristics FAR and FRR results of proposed system are shows the efficient. The scope of improvement to the proposed system can be developed in various algorithms such as graph theory, SVM etc.

REFERENCES

- [1] Armand, S., Blumenstein, M. and Muthukkumarasamy V, “*Offline signature verification using the enhanced modified direction feature and neural-based classification*”, Proceedings of International Joint Conference on Neural Networks, Vancouver, pp.684–691, (2006).
- [2] Anu Rathi, Divya Rathi, Parmanand Astya , “*Offline handwritten Signature Verification by using Pixel based Method*” International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Volume 1 Issue 7, September (2012).
- [3] K. Huang, H. Yan, “*Off-line signature verification based on geometric feature extraction and neural network classification*”, Pattern Recognition, volume 30, issue 1, pp. 9-17, (1997).
- [4] M. Hanmandlu, K.R. Murali Mohan, S. Chakraborty, G. Garg, “*Fuzzy modeling based signature verification system*”, in: Proceedings of the sixth International Conference on Document Analysis and Recognition, USA, pp.110- 114, (2001)
- [5] D. R. Kisku, A. Rattani, P. Gupta, “*A Novel Approach to Offline Signature Verification using Gaussian Empirical Rule*”, 6th European Conference on Information Warfare and Security (ECIW 2007), pp. 139-150, (2007).
- [6] R. Sabourin, R. Plamondon, G. Lorette, “*Offline identification with handwritten signature images: survey and perspectives*”, Structured Image Analysis, Springer, New York, 1992, pp 219-234. Advanced in Information Sciences and Service Sciences Volume 2, issue 3, September (2010)
- [7] R. Plamondon, S.N SriHari, “*Online and Offline handwriting recognition: a comprehensive survey*”, IEEE Trans. Pattern Anal. Mach. Intell. Volume 22 issue 1, pp 63-84, (2000).
- [8] V. Nguyen, M. Blumenstein, G. Leedham, “*Global features for the offline signature verification problem*”, 10th International Conference on Document Analysis and Recognition, ICDAR.,(2009)

IOT Based Waste Management system

Gaurav Thalkar, Darshana Wajekar

Abstract— Waste management is one of the primary problem that the earth faces no matter the case of developed or developing country. The key issue within the waste management is that the rubbish bin at public places gets overflowed well before the commencement of subsequent cleaning process. It successively results in various hazards like bad odor & ugliness there to place which can be the basis cause for spread of varied diseases. To avoid all such hazardous scenario and maintain public cleanliness and health this work is mounted on a wise garbage system. the most theme of the work is to develop a sensible intelligent garbage alert system for a correct waste management. during this paper we've developed a coffee cost, low power waste management system which may be applicable in regions which are not economically sound. this technique enables us to collect the trash as and when the can is full or when the trash inside is decomposed compared to daily collection. This has been designed using an Arduino Uno board incorporating additional modules kind of a GSM module to send messages.

Index Terms— Waste Management System, IOT, Toxicity level, GSM, Ultrasonic Sensor, Gas Sensor, Arduino, Smart Bin.

1 INTRODUCTION

Waste management is becoming a worldwide problem because of increase in population, metropolitan government disorganization, lack of understanding among public, limited support for garbage management program. because of the shortage of care and a spotlight of the authorities of the rubbish cans, it always seems to be crowded. This disorganization creates an enormous untidiness within the society.

India is one among the foremost populated country during which one sixth population of the planet resides. As the population is high, an outsized amount of waste are produced from houses and industries. India approximately generates 1, 33,760 tons of Municipal solid waste (MSW) management per day and approximately 91,152 tones of which are being collected, and on this process of collection a huge amount of money is used. And now by 2050 waste production of world is expected to be approx. 27 billion tons per year, and Asia will contribute one third of this waste, because of China and India as major contributor [9]. A huge amount of waste is produced from smart cities. Mumbai produce a huge waste amongst 50 other cities. It generates 11,000 tons of waste per day [7]. To manage this waste, dustbins are placed in different areas of a city but mainly they overflow and remain unattended. The municipal corporation doesn't get real time information about overflow of dustbins [6] [8]

We use dustbin to throw the waste but still there are some problem. Such as: -

1. Sometimes waste is thrown outside the dustbin. It is not monitored.
2. Harmful gases emerged from dustbin but there is no record of it.
3. Dustbin overflows and are not cleaned timely.
4. Generation of garbage can't be stopped but we can maintain and monitor it. It is necessary to manage the solid waste with proper database.

In India many of the streets are given dustbins. People throw their waste in these dustbins which they're cleaned by municipal corporation (MC). But sometimes dustbin overflows and are not cleaned timely. a wise system should be used to get the important time status of dustbin [10]. We are surrounded by technology in several forms so we should al-

ways always use it to urge the important time monitoring of all the dustbins in any particular area. When dustbin gets full a message should be sent to the MC who is responsible to wash the dustbin. All the info of dustbin is stored for future use. These solutions will definitely make the task of garbage collection more efficient and easier.

2. LITERATURE REVIEW:

Waste management is generally defined as any activity and actions required to manage waste from its inception to its final disposal with the intention of reducing adverse effects of waste on health, the environment or aesthetic. The common waste management system used by the government leads to various problems like social, environmental and health issues. Various reasons for this are non-proper use of dustbin and improper dustbin placements. The garbage management systems should be efficient and also implemented efficiently.

Need for managing large amount of garbage has led to lot of research. Review of papers is carried to know the background, current working procedure and existing system flaws, where we can work on unsolved problems.

A variety of related papers have been reviewed and summarized as follows:

They focus mainly on collection. Current management activity of system and Waste characterization was the main objective to study and to determine[1]. It was demonstrated at toubal muncipalpty, which highlights municipal solid waste management (MSWM) system. And also leaves with suggestions, present management system to be improved which will be useful for authorities to work further also. Smart bin is not used here.

In paper[2] Author introduces technologies like Global Position System, Radio Frequency Identification, General Packet Radio Service, Geographic Information System and web camera are integrated devices. RFID reader device in truck will read both customer and also bin information. Efficient waste collection is achieved. Truck management is also done but is not in optimized way.

Writer of this paper introduces two models of routing they are dynamic and semi static[3]. Further, accumulation of waste in city is done by facilitating IoT a routing model dy-

namically is designed and in case of tragedy it's robust. Research related to waste collection concentrates on routing and scheduling dynamic design. Though lesser amount of research indicates gathering waste for smarter cities assistance.

Chalke and Bhalerao[4] proposed a method that uses various technologies are GPRS, GSM, RFID, IoT etc which can make a positive impact in the society. This can reduce the cost for fuel and transport system and human effort as in this level of dustbin is continuously tracked by the sensors and informed to the concerned authorities when full.

Another method by Wagh and Shahare [5], using Ultrasonic sensors, Wi-Fi Module, Buzzer, LCD display, Power supply it gives an alert when the dustbin is full but it is a non-realistic approach as LCD display is not feasible in practical use of dustbin in public places.

3 BACKGROUND:

3.1 Figures and Tables

Waste management System using internet of things contains of three basic components:

- Smart bin
 - monitoring and control system
 - Truck trash
- 1) Smart Bin:



Fig. 1 Arduino UNO Board

Arduino Uno as shown in the Fig. 1 is a microcontroller board based on the ATmega328P. It has 6 analog inputs, 14 digital input/output pins (of which 6 can be used as PWM outputs), a USB connection, a power jack, a 16 MHz quartz crystal, a reset button and an ICSP header. It consists of everything you need to support a microcontroller, all you need to do is power it using an AC-to-DC battery adapter or connect it with a personal computer with the USB cable to get it started [11].

B.Ultrasonic sensor:



Fig. 2 Ultrasonic Sensor HC – SR04

Fig. 2 is the HC-SR04 ultrasonic ranging sensor. This economically sensor provides non-contact measurement functionality from 2 centimetres to 400 centimetres with a ranging accuracy that can easily reach up to 3mm. Every HCSR04 module includes a receiver, an ultrasonic transmitter, and a control circuit. There are only four pins that you simply got to worry about on the HC- SR04: Trig (Trigger), VCC (Power), GND (Ground) and Echo (Receive) [12].

C.Gas Sensor:



Fig. 3 MQ3 – Gas Sensor

Alcohol Gas Sensor (MQ3) module comes of great use to detect any leakage of gas [industries as well as at home]. It is suitable for detecting LPG, H2, CH4, Alcohol, CO, Propane or smoke. Due to its fast response time & high sensitivity, measurement can be taken as quickly as possible. The sensitivity of the gas sensor are often adjusted by employing a potentiometer.. The sensor value only shows the approximated value of gas concentrations in a error range that is permissible, it does not in any way represent the exact gas concentrations. The detection of certain components in the air usually requires a more costly instrument which is quite precise, and this cannot be done using a single gas sensor [13].

D.GSM Module



Fig. 4 SIM900A GSM Module

SIM 900A module showed in Fig. 4 is used in this work to communicate the information to the user. The board itself contains pins /provisions to attach mic and speaker. These types of provisions vary with different modules [14].

- 2) monitoring and control system:
 Monitoring and system display the bin status instantane-

ously, the sensed data is stored during a database to be shown in graph and support the distribution of the containers decision. The monitoring and system send the bin status and thus the situation message to the rubbish truck to empty the bin if it's full.

3) Garbage Truck:

This part disposes the waste when received monitoring and control system's message in case of full status of the bin.

4. ARCHITECTURE:

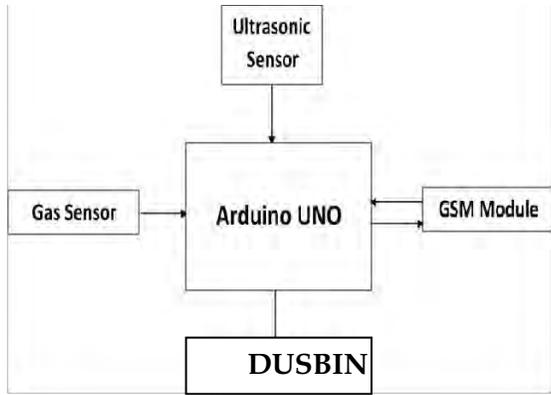


Fig.5: Block diagram of Smart Bin

Fig. 5 depicts the smart bin. The prototype consists of an Arduino (Uno) board, a GSM module, a gas sensor, and an ultrasonic sensor interfaced. The system continuously monitors the level of the can and the level of decomposition of that trash inside. On either the extent or the decomposition has reached its predefined limit, the message are sent to tell the authority so the ashcan collected.

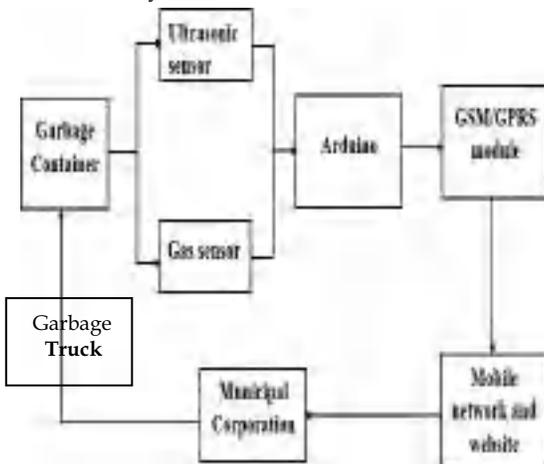


Fig. 6 Waste management system architecture

The following fig.7 illustrates the process flow of waste management system

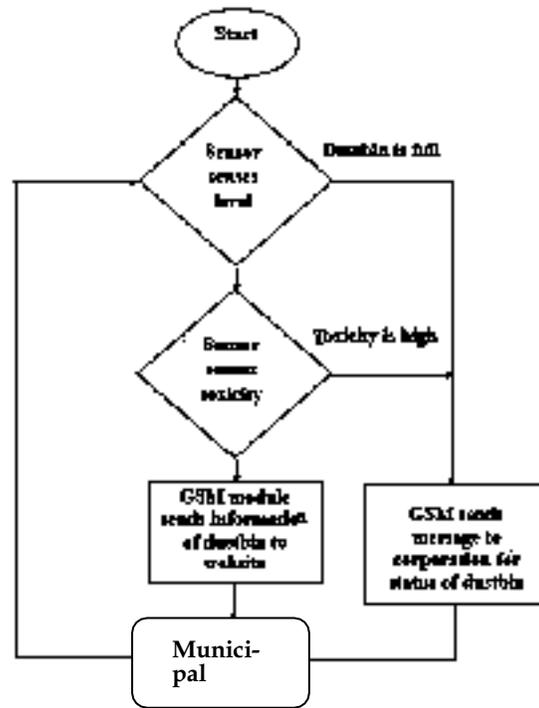


Fig.7: Process flow of waste management system using IoT

Algorithm:

1. Ultrasonic sensor senses the filling level of dustbin.
2. Gas sensor measures the toxicity level of dustbin. Due to presence of waste, dustbin produce hazardous gases which increases toxicity of dustbin.
3. If level is a smaller amount than ten cm or toxicity of gases is high then message is distributed to megahertz through GSM module.
4. the info of dustbin is additionally sent to the web site after a hard and fast interval of dustbin so this information remains store on the web site .
5. If MC get to understand about filling of dustbin then it'll send teamster to wash it. during this manner the dustbins area unit clean timely.

5 CONCLUSION

The waste management system is one among the foremost important systems and it helps within the process of cleanliness of the environment and reduce the difficulties in cleaning operations, where many cities are seeking to use the concept of smart cities and supply simpler services. this technique is far helpful for citizens and Municipal Corporation to manage waste and monitor the dustbin time to time. Smart system

provides the filling status of dustbin using message and it'll save time, fuel and money of Municipal Corporation. As there was a drag of checking real time status of dustbin so it'll be cleaned timely. So, during this project this drawback is solved and proper info is managed on-line. during this suggests the municipal corporation shall work efficiently

REFERENCES

- [1] Ch. Raghmani Singh and Mithra Deyh, "Solid Waste Management of Thoubal Municipality, Manipur- A Case Study", Department of Ecology & Environmental Science, Assam University, Silchar-788011, India environ Jaghu@yahoo.co.in, mithradey@gmail.com, 978-1-4673-0178-7/11, IEEE 2011.
- [2] Md. Shafiqul Islam, Maher Arebey, M.A. Hannan, Hasan Basri "Overview for Solid Waste Bin Monitoring and Collection System" 2012 International Conference on Innovation, Management and
- [3] Theodoros Vasileio, Anagnostopoulos and Arkady Zaslavsky, " Effective Waste Collection with Shortest Path Semi-Static and Dynamic Routing", Springer International Publishing Switzerland, LNCS 8638, pp. 95-105, 2014.
- [4] S. Bhuravane, M. Panindre, S. Patole, and P. Therade, "Survey of Wi-Fi Trash Bin," IOSR Journal of Computer Engineering (IOSR-JCE), vol. 4, Oct. 2017
- [5] H. Shinde, M. Diwakar, N. Salvi, R. Sarde, and S. Shaikh, "Survey on Smart Garbage Collection System Using IOT," International Journal of Innovative Research in Computer And Communication Engineering, vol. 5, Issue 10, Oct. 2017.
- [6] Haribabu P, Kassa SR, Nagaraju J, Karthik R, Shirisha N, Anila M, "Implementation of a smart waste management system using IoT" in International Conference on Intelligent Sustainable System (ICISS) (pp. 1155-1156), IEEE.
- [7] The Hindustan times available at:https://www.google.com/amp/S/m.hindustantimes.com/Mumbai-news/105-rise-in-waste-generated-in-mumbai-from-1999-to-2016-study/story-QEI4qWehSae9ZYtImNdBaM_amp.html (Accessed on 10 Jan 2019).
- [8] Ravale U, Khade A, Patel N, Chaurse S. "Smart Trash: An Efficient Way for Monitoring Solid Waste Management" in 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) 2017 Sep 8 (pp. 1135-1137). IEEE.
- [9] V. Praveen, T. Arunprasad, R. Naveen, and Dr. P. Gomathi, "A Survey on Garbage Monitoring System using IOT," International Journal of Trend in Scientific Research and Development, International Open Access Journal, vol. 2, 2017.
- [10] Joshi J, Reddy P, Reddy P, Agarwal A, Agarwal R, Bagga A, Bhargava, "Cloud computing based smart garbage monitoring system" in 3rd International Conference on Electronic Design (ICED), 2016 Aug 11 (pp. 70-75), IEEE.
- [11] <https://www.arduino.cc/en/main/arduinoBoardUno>
- [12] <https://www.sparkfun.com/products/13959>
- [13] http://wiki.seeed.cc/Alcohol-Gas_Sensor-MQ3/
- [14] <http://www.circuitstoday.com/interface-gsm-module-with-arduino>

Cloud Based Smart Ration Card System

Gaurav Thalkar, Darshana Wajekar

Abstract— Ration Card is one among the important documents for each Indian family. Ration card is a crucial document that provides data associated with a shopper. The aim of card is to produce ration to poor families as allotted by the government at low value. The ration merchandise embrace rice, wheat, sugar, kerosene, etc. It also can be used to apply for gas affiliation. During this paper, RFID tags square measure used to replace the present card. A singular RFID tag is allotted to each shopper; it'll facilitate in authentication of consumer data. The data of the buyer is kept in a very info that is connected to Amazon net Services (AWS). To attain a safer authentication, a One Time Parole (OTP) is employed. Once an acquisition request is initiated on the system, AN OTP is received by the buyer on their registered mobile range. When shopper enters the received OTP, the system sends it to the info for authentication. If the OTP entered by the buyer is correct, the system accepts the request and so ration content is allotted to the buyer. When the allocation, the ration the system updates the info. The buyer will receive AN SMS relating to the ration purchase.

Index Terms— Ration card, Cloud, RFID, SRC, AWS, OTP, Arduino, WiFi Module.

1 INTRODUCTION

Ration card plays a crucial role for each subject of Republic of India. It includes personal data regarding the buyer and it acts as address and identity proof. Card relies on the condition of the family and that they are available varied colours. Food security system developed by the govt of Republic of India subsidizes food and fuel for the poor. It includes distribution of food grains like wheat, rice, salt, sugar, oil and fuel from the ration search. The ration is allotted to the buyer on a monthly basis. The present card has several disadvantages and downsides. A market keeper will sell the product on the black market to achieve profits because the overnment doesn't have any record of the product. The market keeper will even hike the worth of the merchandise. The market keeper will even create pretend entries so as to deceive the govt. Typically the customers might not bear in mind of the product offered at the ration search. So as to eliminate these errors we tend to area unit implementing Cloud primarily based good card System victimisation RFID. It'll facilitate U.S.A. to scale back the protection problems gift in current manual card system. It ensures higher management of public distribution system.

2. LITERATURE REVIEW:

Ration distribution is one of the widely controversial issue that involves wrong entries in stock register of shop containing wrong stock information of the products that is supplied to the public, Rajesh C. Pingle et.al. Suggested a ration system using RFID and GSM Module to Prevent Irregularities, in this automated system conventional ration card is replaced by smartcard in which all the details about users are provided including their AADHAR number which is used for user authentication [2].

Now day's ration card is very important for every home and used for various field such as family members details, to get gas connection, it act as address proof for various purposes etc. But this system is having two draw backs, first one is weight of the material may be inaccurate due to human mistakes and secondly, if consumers do not buy the materials at the end of the month, they will sale to others without any in-

timisation to the government and customers So Ms. T. Sheela et.al. proposed a new system. To get the materials in ration shops need to show the RFID tag into the RFID reader, then controller check the customer codes and details of amounts in the card. After verification, these systems show the amount details. Then customer need to enter they required materials by using keyboard, after receiving materials controller send the information to government office and customer through GSM technology. In this system provides the materials automatically without help of humans [3].

Huge amount of Govt. money get wasted due to corruption in the conventional Ration Distribution System. So the system proposed in this paper implements a simple PDA device (personal data assistant) with RFID tag used as an eration card in place of a conventional ration card. This PDA device is similar to the ticketing machine used by bus conductor or bank pigmy agent and the e-ration card is similar to swipe card. The Subscriber has to use this card instead of a traditional ration card to get ration from the dealer [4].

The current PDS involves corruption and illegal smuggling of goods because of manual work. So the system proposed in this paper includes the automation of the Ration system using Arm7. The distribution is done automatically using mechanical devices. The Arm7 work as the central processing unit for the system. So, this paper provides a way to automatically distribute the contents using microcontroller, but it does not provide any way to monitor the PDS efficiently [5].

3. Background:

Ration card was first time introduced in India during war II to provide food grains, sugar and kerosene oil at a comparatively cheaper rate. Card acts because the address/identity proof of a personal. It includes the identity of the person in conjunction with his members of the family, their names, ages, gender.

1) Arduino:



Fig. 1. Arduino Mega 2560R3

Arduino is a hardware prototyping platform. It is a microcontroller board which processes inputs and send outputs. Arduino Mega2560R3 as shown in fig.1 is based on ATmega 2560 microcontroller. It has 54 digital input/output pins (of which 15 are often used as PWM outputs), 16 analog inputs, 4 UARTs (hardware serial ports), a 16 MHz quartz oscillator , a USB connection, an influence jack, an ICSP header, and a push button [1]. It works on 12V DC power supply. It can be programmed using arduino IDE.

2) EM-18 RFID Reader:



Fig. 2 EM-18 RFID Reader

The EM-18 RFID reader module is based on radio frequency identification technique. It is shown in fig.2. It operates on 125 KHz frequency band. It can scan a passive RFID tag from a distance of upto 10cm. It are often connected to any microcontroller using either UART or RS232 port. It requires 12V DC power supply for operation.

3) ESP8266 Wi-Fi Module:



Fig.3. ESP8266 Wi-FiModule

Fig3. Shows ESP8266 Wi-Fi module. It's a micro chip that makes use of Wi-Fi 802.11 standards. It's wont to connect the microcontroller boards to web victimisation Wi- Fi technology.. It requires 3.3V DC power supply.

4) Software Tools:

a. AWS:

Amazon Web Services (AWS) is a subsidiary of Amazon.com. A paid subscription of AWS permits you to access numerous cloud computing platforms. AWS offers large vary of virtual computers and process power. AWS supports numerous in operation systems and networking platforms. Subscribers can connect their system to AWS using any modern browser.This virtual window allow subscribers to configure their system, change settings, etc. AWS offers a good vary of cloud computing services.

b. MongoDB:

Mongoddb is a NoSQL database program. It is free and open source. Mongoddb supports many operating systems such as windows, linux, Mac, etc. It is document oriented database program. It saves database in a document format. JSON (Java Script Object Notation) is used to make the mongoddb database. Mongoddb makes the task of adding new entries in previous databases terribly easy.

c. Robomongo:

The visual tool used to produce and manage Mongoddb information is termed robomongo.it's free and open source software.

d. PuTTY:

Putty is a Secure Shell Access (SSH) client. To firmly connect with a foreign server or remote pc putty is employed. It permits you to perform an exact command or set of commands on your pc and have another pc execute the action.

SIM 900A module showed in Fig. 4 is used in this work to communicate the information to the user. The board itself contains pins /provisions to attach mic and speaker. These types of provisions vary with different modules [14].

2) monitoring and control system:

Monitoring and system display the bin status instantaneously, the sensed data is stored during a database to be shown in graph and support the distribution of the containers decision. The monitoring and system send the bin status and thus the situation message to the rubbish truck to empty the bin if it's full.

3) Garbage Truck:

This part disposes the waste when received monitoring and control system's message in case of full status of the bin.

4) BLOCK DIAGRAM:

Block diagram of smart ration card system is shown in fig. 4. The block diagram of smart ration card is shown in figure. In this system, we use an arduino mega microcontroller, Power supply, RFID reader, RFID tag, Wi-Fi module (ESP8266), 20*4 LCD display, 4*4 alphanumeric keypad and Database.

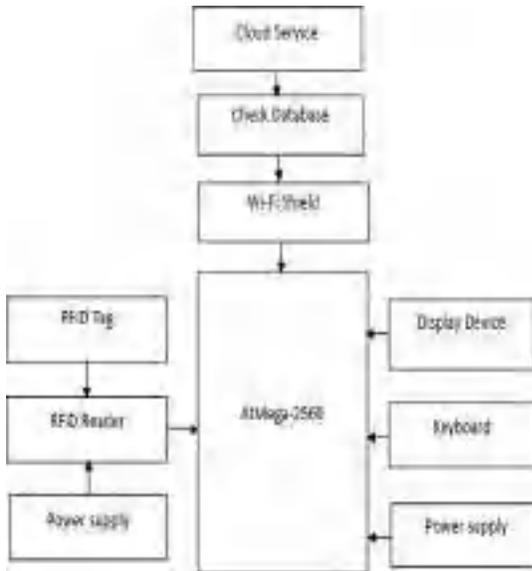


Fig 4. Block diagram of Cloud based Ration Card System

The following fig.7 illustrates the process flow of wastemanagement system.

6) FLOW CHART:

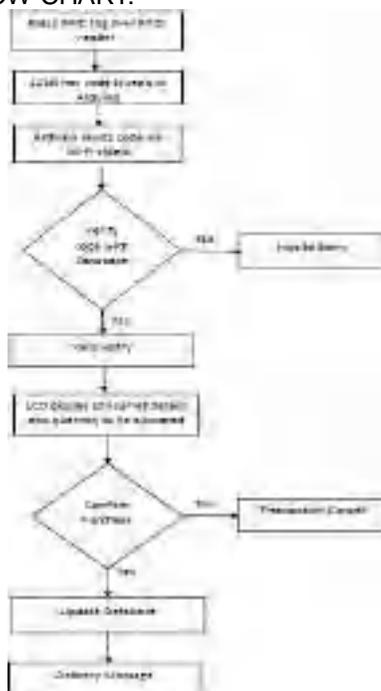


Fig. 5. Flow Chart of Smart Ration Card System

IV. ALGORITHM:

A. Database:

A consumer database is created using the mongodb and robomongo software's. The database objects are written in json format. The database consists of all the information of the consumer such as consumer ID, mobile number, quantity of ration allocated, etc.

B. Cloud Services:

An account is created on AWS with paid subscription to cloud services. Using required configurations and putty, the database created mongodb is uploaded on AWS cloud. So it can be accessed anywhere, anytime.

C. Input:

First the consumers need to swipe the RFID tag provided to them over the RFID reader module. The input to the system is in form of 12bit hex code which is scanned using EM-18 RFID reader module. This hex code is then sent to the arduino microcontroller board.

D. Processing:

Atmel atmega2560 microcontroller present on the arduino takes care of processing work. The hex code sent by the rfid reader module is matched with the database. The arduino is connected to the web via a Wi-Fi hotspot using esp8266 Wi-Fi module. Arduino also fetches the buyer info from the AWS cloud servers.

E. Displaying Information:

For displaying the information to the consumer a 20*4 LCD module is used. The ration items and quantity available to the consumer with their rate is displayed on the LCD display if the consumer is found to be authentic.

F. Data Input:

For proving data input to the system a 4*4 keypad is used .The consumer have to enter the OTP received on their registered mobile number using this keypad. In the final step the purchase is confirmed using the keypad.

V. ADVANTAGES:

- Easy to use
- Data Integrity
- Security/Safety
- Flexibility

VI. CONCLUSION:

The traditional ration distribution system faces several issues. additionally it's terribly tough for the govt. to manage such a huge system. therefore our system aims at mistreatment the technology accessible these days to create the ration system sensible. The Cloud primarily based Ration System mistreatment RFID can revolutionize the ration system.

it'll be straightforward and secure for the user to urge the abundant required ration content. Government are able to keep track of all the transactions and malpractices and corruption are reduced.

VII. FUTURE WORK:

Although our system is smart, it can be further modified. We can add extra security by linking AADHAR number of consumer to the ration card .By doing so we can enable fingerprint verification using fingerprint readers. It will become more secure than our current system.

REFERENCES

- [1] Y.A. Badamasi, "The working principle of an Arduino", in Electronics, Computer and Computation (ICECCO), 11th International Conference on Information Communication Embedded Systems, pages 1-4, 2014.
- [2] Rajesh C. Pingle and P. B. Borole, "Automatic Rationing for Public Distribution System (PDS) using RFID and GSM Module to Prevent Irregularities", HCTL Open International Journal of Technology Innovations and Research, vol 2, pp. 102- 111, Mar 2013.
- [3] Ms T.Sheela, Dr. PM Murali, Dr. T. Muthumanickam and Mr.D.Padmarajan, "RFID Based Automatic Ration Selling System", International Journal of Engineering, Business and Enterprise Applications, 12(2), March-May 2015.
- [4] KashinathWakade, PankajChidrawar, Dinesh Aitwade, BirudevTarate and Prof. M. M. Pawar, "e-Ration System Using RFID and GSM Technology", SVERIAN Scientific, Mar 2014.
- [5] A.N.Madur, Sham Nayse, "Automation in Rationing System Using Arm 7," International journal of innovative research in electrical, electronics, Instrumentation and control engineering vol.1, Issue 4, Jul 2013.
- [6] Yerlan Berdaliyev, Alex Pappachen James, " RFID-Cloud Smart Cart System", IEEE Intl. Conference on Advances in Computing A Communications and Informatics (ICACCI), Sept. 21-24, 2016.
- [7] Bichlien Hoang, Ashley Caudill, "RFID", IEEE Emerging Technology Portal, 2006 - 2012.
- [8] Anshu Prasad, Aparna Ghenge, Sonali Zende, Sashikala Mishra, Prashant Godakh, "Smart Ration Card Using RFID, Biometrics and SMS Gateway", IEEE International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017.

Survey paper on Robotic Process Automation

Yashwin Poojary, Darshana Wajekar

Abstract - The automation of robotic processes has been experiencing an increasing trend of interest in recent times. However, most of literature describes only theoretical foundations on RPA or industrial results after implementing RPA in specific scenarios, especially in finance and outsourcing. This paper presents a systematic mapping study with the aim of analyzing the current state-of-the-art of RPA and identifying existing gaps in both, scientific and industrial literature. Firstly, this study presents an in-depth analysis of the 54 primary studies which formally describe the current state of the art of RPA. These primary studies were selected as a result of the conducting phase of the systematic review. Secondly, considering the RPA study performed by Forrester, these paper reviews 14 of the main commercial tools of RPA, based on a classification framework defined by 48 functionalities and evaluating the coverage of each of them. The result of the study concludes that there are certain phases of the RPA lifecycle that are already solved in the market. However, the Analysis phase is not covered in most tools. The lack of automation in such a phase is mainly reflected by the absence of technological solutions to look for the best candidate processes of an organization to be automated. Finally, some future directions and challenges are presented.

Index Terms - Robotic process automation, RPA, Systematic mapping study.

1 INTRODUCTION

Although the term “Robotic Process Automation” (RPA) encourages thinking about robots doing human tasks, really, it is a software solution. In the context of RPA, a “robot” corresponds to a software program. For business processes, the term RPA means the technological extrapolation of a human worker, whose objective is to tackle structured and repetitive tasks (very common in ERP systems or productivity tools), quickly and profitably. It is possible to say that “RPA aims to replace people by automation done in an outside-in manner. This differs from the classical inside-out approach to improve information systems”. Adopting RPA implies a low level of intrusiveness since, according to the Institute for Robotic Process Automation and Artificial Intelligence (IRPA-AI), this technology is not part of the information technology infrastructure of a company, but rather sits on top of that.

With relation to cost, Capgemini suggests that an RPA software license may cost between 1/3 and 1/5 of the price of a full-time employee. In addition, Lacity and Willcocks argue that a robot can perform structured tasks equivalent to two or five humans. Anyway, the use of RPA by companies provides the following advantages

- RPA is easy to configure, so developers do not need programming skills.
- The RPA software is not invasive, it is based on existing systems, without the need to create, replace or develop expensive platforms.
- RPA is secure for the company, RPA is a robust platform that is designed to meet the IT requirements of the company in terms of security, scalability, and auditability and change management.

Although the benefits in cost savings are significant, not all business processes are suitable for their use. Fung suggests that the business processes tasks where RPA may be

applied should meet the following criteria: those with a low level of knowledge, those which are high-frequency executed, query different systems and applications, those which are standardized with a low level of exceptions to control and those susceptible to end in error caused by human errors. Considering these criteria, the best candidates for implementing a RPA are the companies which business is based on back-office areas.

As mentioned above, there are several scientific proposals in which an implementation of RPA is presented for a specific domain. However, to the best of our knowledge, it appears that RPA is being more used in industrial than scientific contexts. In this sense, opening a discussion about the disparities and coincidences between RPA and similar technologies, and formally classifying what is being investigated relative to this technology, is of vital importance for the community to grow and open new research lines.

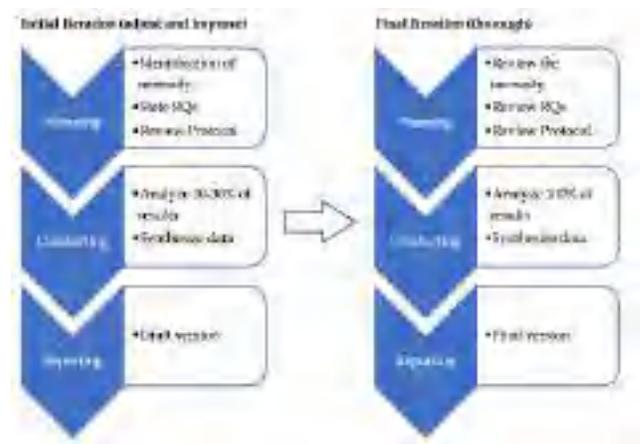


Figure 1:- Two iteration SMS Proposal

The method has been executed in two iterations (c.f. Figure 1). In the first one, a review with a percentage of the

articles found with the aim of strengthening the research line and is carried out, in the second one, after reviewing and refining the process, it is executed for the rest of the Articles and the new ones if any. Considering the extension of the document and to improve its readability, the results of this SMS will be grouped into the three main blocks that SLR method defines: planning, conducting and reporting for both, scientific and industrial scopes. The rest of the paper is organized as follows. Section II introduces the background and context of RPA. Section III describes the entire process carried out in both scientific and A industrial scope. Section IV details possible threats to the validity of the method executed. Section V summarizes the closest related work found in the literature. Section VI shows the more outstanding conclusions of this study and finally, Section VII draws up strategic views regarding future research lines in the field of RPA.

2 METHODS

There are several methods to initiate the Robotic Process Automation. Below is given a detailed description of the same.

- Pick The Right Technology

Just because it's in the Gartner Magic Quadrant doesn't mean it's the right solution to your business problem.

It's also likely to also be the most costly solution to your business problem.

There is an almost overwhelming range of software on the market to suit every business need so selecting a vendor should be more about fit with needs rather than who is winning the most deals and who is less risky.

Big software companies often have the same painful process issues you are trying to fix, so why should you try to fix your problems with the same kind of thinking that created them?

- Start with The Customer

Too many organisations have a myopic view which focuses firstly on internal processes prior to looking at the customer.

Focusing on what the customer needs and looking at end-to-end process from a customer experience perspective stops wasted time and effort improving processes and tasks which shouldn't actually exist in the first place.

Start with the customer, define what they need and align your organisational processes to that need and you can't go far wrong.

- Implement Change Management (Properly!)

Change Management. That's that warm and fuzzy feelgood stuff, puffed up with words like "engagement" and "actualisation", right? Wrong.

It's time everyone started to take change management more seriously – studies have shown that projects are six times more likely to succeed if they have good change management.

Get a change manager on board and on-board early. A good change manager is worth their weight in gold – it's a specialised skill that requires experience, expertise and a structured approach.

- Bring The Business into the Project (Fully!)

Part-time business commitment is the first sure-fire sign of a delivery disconnect.

If the business are unwilling to commit at least one fulltime resource to a process improvement project then stop the bus and let everyone off. Not only is the business's expertise essential to progressing the project, but a lack of commitment on a project that is designed to improve the business itself shows that they either do not adequately understand or appreciate the benefits that the initiative will bring or have no faith in your ability to deliver!

Either way, business commitment is another critical step on the path to success.

- Prove The Concept

Start small. Never go "big bang". Prove the concept.

It's far too risky in this day and age to start with a huge transformational implementation of software if you have never implemented the software before. Start with a small project that can prove the benefits of your business case and ramp it up from there.

3 TOOLS

There are various Tools that can be used in the Robotic Process Automation. As RPA is slowly becoming a new dominant technology in the industry, more and more companies are getting inclined toward inculcating RPA in their back-office tasks. This has led to an exponential growth in the demand for RPA tools as well. There are various tools available in the market to implement RPA. These tools are used by organizations to define the workflow of the task that needs to be automated using RPA. In this RPA tutorial on RPA tools, we will learn about various you will see a long RPA tools list, how they are different and much more.

Categories of RPA Tools

Here are the three main categories of RPA tools in the market.



Figure 2:- Categories of RPA Tools

Programmable RPA Bots: They are the first generation of RPA tools. It included programming, and the programmers needed to code and understand its working to make the bots perform actions to complete tasks.

Self-learning Tools: After vigorous employee activity, the developer/employee understood the process, took over the platform, and started performing the task.

Cognitive or Intelligent Automation Bots: Cognitive automation bots are augmenting self-learning bots with advanced functionalities such as Machine Learning, Natural Language Processing, and image recognition. They can deal with both structured and unstructured data.

4 USE CASE

RPA can be used to automate repetitive tasks both in the back office and front office that require human intervention. Some common RPA examples and use cases we encounter are automation of data entry, data extraction, and invoice processing. There are additional examples of RPA use cases automating tasks in different business departments (Sales, HR, operations, etc.) and industries (banking, retail, manufacturing, etc.).

There are various fields in which RPA is used widely and successfully. Here are some of the examples of the same. Common business processes and activities, Activities in commercial functions, Customer Relationship Management, Activities in support functions, Technology, Finance, HR, Operations, Procurement, Financial Services, Telecom, Retail, Manufacturing, Healthcare & Government

1- Quote-to-cash: - Every business needs to sell to survive. Issues in the operations side of selling are costly, they can result in customer complaints or selling at reduced prices due to clerical errors.

2- Procure-to-pay (p2p) / source-to-pay (s2p):-Since procure-to-pay process involves extracting invoice and payment data from multiple systems like supplier emails, enterprise resource planning (ERP), customer relationship management (CRM), banks, vendors, logistics companies.

3- Customer onboarding:-Most B2C businesses have a customer onboarding process that is critical to reduce churn and get customers to start using the product.

4- Data migration and entry :-Legacy systems still perform critical functions at companies.

5- Data updates: - Most departments including HR, customer service and marketing routinely need to update cus-

tomers/personnel data which is constantly changing.

6- Data validation:-Most data validation controls can be embedded in databases.

7- Extracting data from PDFs, scanned documents and other formats: - Screen scraping, OCR (Optical Character Recognition) and basic pattern recognition technologies enable data extraction from almost any format, reducing the need for keying in data.

8- Periodic report preparation and dissemination: - Every business requires regular reports to inform managers and ensure teams are aware of their progress. Preparing such reports and sending them over every week or month is not labor intensive but it distracts employees.

9- Generating mass email: - Mass emails relying on data from multiple systems are painful to produce manually. Especially if you are sending them frequently, consider automating the process.

5 REAL LIFE EXAMPLE

There are various examples where the RPA is extensively used for the development and smooth running of any process. In this document we will take an example of uses of RPA in **Financial Services**.

RPA helps in bridging the gap between different applications present in the legacy system. Even though there are many system implementation software, such as SAP and Oracle but they fall short in processing and integrating micro-level tasks.

Real-Life Use Case Example:

A lot of Canadian banks, for example, have been eagerly adopting RPA technology with success. Royal Bank of Canada, the largest financial institution in Canada is using smart chatbots for more than a year to improve customer service. Also, the Guardian Group has also embraced RPA technology in their processes and have seen early success in their journey. The technology has enabled them to calculate, copy, paste, access, or use existing business rules to handle, use, and feed data into the core enterprise application.

4 Ways of Implementing RPA in Finance Industry

Other Processes of RPA Use Case in Financial Services Can Include:

1. Accounts Payable (AP)

RPA bots can transfer inbound invoices from PDFs, internal spreadsheets, and web applications. Also, RPA bot can place a duplicate PDF on an internal server and this makes it useful in order to ensure regulatory compliance, and can reduce up to 60% of vendor invoice processing cycle times.

2. Accounts Receivable (AR)

Bots are capable of faster and more accurate maintenance of customer master files and credit approvals. The

same goes for AR cash receipts and order, and their processing. Also, late notices can be sent by email more quickly with the minimum hassle that naturally comes with last-minute updates.

3. Financial Close and Reporting

This is the core of inter-departmental, multisystem processing. Drafting tax entry data from various business units is a boring and monotonous task. This is why it is better to have a robot mitigate it with the system for error-proof entries in the processes.

4. Accounting Reconciliation

Accounting data calls for reconciliation taken from a variety of sources, such as Excel sheets, or customer invoices. With automation, you can process and download the data into the desired format and validate the data for any exceptions.

Benefits of RPA Implementation in Finance:

- Non-invasive application
- Customizable workflow
- Major lifting
- Speedy installations
- Reduce errors in work
- Nonstop performance

6 CONCLUSION

RPA is a powerful automation solution that offers a variety of opportunities to improve quality, increase control, add flexibility, and unlocks a wide scope of automation possibilities. However, a clear RPA vision and strategy, feeding into the implementation approach and operating model are prerequisites for success.

RPA adoption will continue to grow in the years to come. Meanwhile more advanced robotics solutions will continue to mature and be adopted across industries. Those organizations that are successful in scaling RPA are best positioned to advance to and reap the benefits of enhanced and cognitive automation solutions.

7 REFERENCE

[1] A. Lintukangas, "Improving indirect procurement process by utilizing robotic process automation," M.S. thesis, School Bus. Manage., Lappeenranta Univ. Technol., Lappeenranta, Finland, 2017.

[2] ActiveBatch. (2019). ActiveBatch (Advanced Systems Concepts). Accessed: Sep. 2019. [Online]. Available: <https://www.advsyscon.com>

[3] S. Agostinelli, "Synthesis of strategies for robotic process automation," in Proc. 27th Italian Symp. Adv. Database Syst. (CEUR), Grosseto, Italy, vol. 2400, M. Mecella, G. Amato, and C. Gennaro, Eds., Jun. 2019.

[4] S. Aguirre and A. Rodriguez, "Automation of a business process

using robotic process automation (RPA): A case study," in Proc. Appl. Comput. Sci. Eng., 4th Workshop Eng. Appl. (WEA), 2017, pp. 65–71.

[5] W. A. Ansari, P. Diya, S. Patil, and S. Patil, "A review on robotic process automation-the future of business organizations," Available at SSRN 3372171, 2019.

[6] A. Asatiani and E. Penttinen, "Turning robotic process automation into commercial success – case OpusCapita," J. Inf. Technol. Teach. Cases, vol. 6, no. 2, pp. 67–74, Nov. 2016.

SURVEY ON ARTIFICIAL INTELLIGENCE IN HEALTHCARE

Yashwin Poojary, Darshana Wajekar

Abstract— The complexity and rise of data in healthcare means that artificial intelligence (AI) will increasingly be applied within the field. Several types of AI are already being employed by payers and providers of care, and life sciences companies. The key categories of applications involve diagnosis and treatment recommendations, patient engagement and adherence, and administrative activities. Although there are many instances in which AI can perform healthcare tasks as well or better than humans, implementation factors will prevent large-scale automation of healthcare professional jobs for a considerable period. Ethical issues in the application of AI to healthcare are also discussed.

Index Terms— Artificial intelligence, clinical decision support, electronic health record systems

1 INTRODUCTION

Artificial intelligence (AI) and related technologies are increasingly prevalent in business and society, and are beginning to be applied to healthcare. These technologies have the potential to transform many aspects of patient care, as well as administrative processes within provider, payer and pharmaceutical organisations.

There are already a number of research studies suggesting that AI can perform as well as or better than humans at key healthcare tasks, such as diagnosing disease. Today, algorithms are already outperforming radiologists at spotting malignant tumours, and guiding researchers in how to construct cohorts for costly clinical trials. However, for a variety of reasons, we believe that it will be many years before AI replaces humans for broad medical process domains. In this article, we describe both the potential that AI offers to automate aspects of care and some of the barriers to rapid implementation of AI in healthcare.

2 METHODS

Artificial intelligence is not one technology, but rather a collection of them. Most of these technologies have immediate relevance to the healthcare field, but the specific processes and tasks they support vary widely. Some particular AI technologies of high importance to healthcare are defined and described below.

Machine learning - neural networks and deep learning: - Machine learning is a statistical technique for fitting models to data and to 'learn' by training models with data. Machine learning is one of the most common forms of AI; in a 2018 Deloitte survey of 1,100 US managers whose organisations were already pursuing AI, 63% of companies surveyed were employing machine learning in their businesses. It is a broad technique at the core of many approaches to AI and there are many versions of it.

In healthcare, the most common application of

traditional machine learning is precision medicine - predicting what treatment protocols are likely to succeed on a patient based on various patient attributes and the treatment context. The great majority of machine learning and precision medicine applications require a training dataset for which the outcome variable (eg onset of disease) is known; this is called supervised learning.

Natural language processing: - Making sense of human language has been a goal of AI researchers since the 1950s. This field, NLP, includes applications such as speech recognition, text analysis, translation and other goals related to language. There are two basic approaches to it: statistical and semantic NLP. Statistical NLP is based on machine learning (deep learning neural networks in particular) and has contributed to a recent increase in accuracy of recognition. It requires a large 'corpus' or body of language from which to learn.

In healthcare, the dominant applications of NLP involve the creation, understanding and classification of clinical documentation and published research. NLP systems can analyse unstructured clinical notes on patients, prepare reports (eg on radiology examinations), transcribe patient interactions and conduct conversational AI.

Rule-based expert systems:- Expert systems based on collections of 'if-then' rules were the dominant technology for AI in the 1980s and were widely used commercially in that and later periods. In healthcare, they were widely employed for 'clinical decision support' purposes over the last couple of decades and are still in wide use today. Many electronic health record (EHR) providers furnish a set of rules with their systems today.

Expert systems require human experts and knowledge engineers to construct a series of rules in a particular knowledge domain. They work well up to a point and are easy to understand. However, when the number of rules is large (usually over several thousand) and the rules begin to conflict with each other, they tend to break down.

Diagnosis and treatment applications: - Diagnosis and treatment of disease has been a focus of AI since at least the

1970s, when MYCIN was developed at Stanford for diagnosing blood-borne bacterial infections. This and other early rule-based systems showed promise for accurately diagnosing and treating disease, but were not adopted for clinical practice. They were not substantially better than human diagnosticians, and they were poorly integrated with clinician workflows and medical record systems.

Implementation issues with AI many healthcare organisations. Although rule-based systems incorporated within EHR systems are widely used, including at the NHS, they lack the precision of more algorithmic systems based on machine learning. These rule-based clinical decision support systems are difficult to maintain as medical knowledge changes and are often not able to handle the explosion of data and knowledge based on genomic, proteomic, metabolic and other 'omic-based' approaches to care.

3 TOOLS

Artificial Intelligence has facilitated the processing of a large amount of data and its use in the industry. The number of tools and frameworks available to data scientists and developers has increased with the growth of AI and ML. This article on Artificial Intelligence Tools & Frameworks will list out some of these in the following sequence:

List of Artificial Intelligence Tools & Frameworks

1. Scikit Learn
2. TensorFlow
3. Theano
5. Caffe
5. MxNet
6. Keras
7. PyTorch
8. CNTK
9. Auto ML
10. OpenNN
11. H2O: Open Source AI Platform
12. Google ML Kit

Scikit Learn :-Scikit-learn is one of the most well-known ML libraries. It underpins many administered and unsupervised learning calculations. Precedents incorporate direct and calculated relapses, choice trees, bunching, k-implies, etc.

Tensorflow On the off chance that you are in the realm of Artificial Intelligence, you have most likely found out about, attempted or executed some type of profound learning calculation. The fascinating thing about Tensorflow is that when you compose a program in Python, you can arrange and keep running on either your CPU or GPU. So you don't need to compose at the C++ or CUDA level to keep running on GPUs. **Theano**:-Theano is wonderfully folded over Keras, an abnormal state neural systems library that runs nearly in parallel with the

Theano library. Keras' fundamental favorable position is that it is a moderate Python library for profound discovering that can keep running over Theano or TensorFlow.

Caffe:-'Caffe' is a profound learning structure made with articulation, speed, and measured quality as a top priority. It is created by the Berkeley Vision and Learning Center (BVLC) and by network donors. Google's DeepDream depends on Caffe Framework.

MxNet:-It allows for trading computation time for memory via 'forgetful backprop' which can be very useful for recurrent nets on very long sequences. Built with scalability in mind (fairly easy-to-use support for multi-GPU and multi-machine training).

Keras :- If you like the Python-way of doing things, Keras is for you. It is a high-level library for neural networks, using TensorFlow or Theano as its backend.

PyTorch:-PyTorch is an AI system created by Facebook. Its code is accessible on GitHub and at the present time has more than 22k stars. It has been picking up a great deal of energy since 2017 and is in a relentless reception development.

CNTK:-CNTK allows users to easily realize and combine popular model types such as feed-forward DNNs, convolutional nets (CNNs), and recurrent networks (RNNs/LSTMs).

Auto ML:-Out of all the tools and libraries listed above, Auto ML is probably one of the strongest and a fairly recent addition to the arsenal of tools available at the disposal of a machine learning engineer.

OpenNN :- Jumping from something that is completely beginner friendly to something meant for experienced developers, OpenNN offers an arsenal of advanced analytics.

H2O: Open Source AI Platform:-H2O is an open-source deep learning platform. It is an artificial intelligence tool which is business oriented and helps them to make a decision from data and enables the user to draw insights.

Google ML Kit :- Google ML Kit, Google's machine learning beta SDK for mobile developers, is designed to enable developers to build personalised features on Android and IOS phones.

4. ADMINISTRATIVE APPLICATIONS

There are also a great many administrative applications in healthcare. The use of AI is somewhat less potentially revolutionary in this domain as compared to patient care, but it can provide substantial efficiencies. These are needed in healthcare because, for example, the average US

nurse spends 25% of work time on regulatory and administrative activities. The technology that is most likely to be relevant to this objective is RPA. It can be used for a variety of applications in healthcare, including claims processing, clinical documentation, revenue cycle management and medical records management.

Some healthcare organisations have also experimented with chatbots for patient interaction, mental health and wellness, and telehealth. These NLP-based applications may be useful for simple transactions like refilling prescriptions or making appointments. However, in a survey of 500 US users of the top five chatbots used in healthcare, patients expressed concern about revealing confidential information, discussing complex health conditions and poor usability.

Another AI technology with relevance to claims and payment administration is machine learning, which can be used for probabilistic matching of data across different databases. Insurers have a duty to verify whether the millions of claims are correct. Reliably identifying, analysing and correcting coding issues and incorrect claims saves all stakeholders – health insurers, governments and providers alike – a great deal of time, money and effort. Incorrect claims that slip through the cracks constitute significant financial potential waiting to be unlocked through data-matching and claims audits.

There has been considerable attention to the concern that AI will lead to automation of jobs and substantial displacement of the workforce. Deloitte collaboration with the Oxford Martin Institute suggested that 35% of UK jobs could be automated out of existence by AI over the next 10 to 20 years. Other studies have suggested that while some automation of jobs is possible, a variety of external factors other than technology could limit job loss, including the cost of automation technologies, labour market growth and cost, benefits of automation beyond simple labour substitution, and regulatory and social acceptance. These factors might restrict actual job loss to 5% or less.

To our knowledge thus far there have been no jobs eliminated by AI in health care. The limited incursions of AI into the industry thus far, and the difficulty of integrating AI into clinical work-flows and EHR systems, have been somewhat responsible for the lack of job impact. It seems likely that the healthcare jobs most likely to be automated would be those that involve dealing with digital information, radiology and pathology for example, rather than those with direct patient contact.

But even in jobs like radiologist and pathologist, the penetration of AI into these fields is likely to be slow. Even though, as we have argued, technologies like deep learning are making inroads into the capability to diagnose and categorise images, there are several reasons why radiology jobs, for example, will not disappear soon.

Finally, there are also a variety of ethical implications around the use of AI in healthcare. Healthcare decisions have been made almost exclusively by humans in the past, and the

use of smart machines to make or assist with them raises issues of accountability, transparency, permission and privacy. Perhaps the most difficult issue to address given today's technologies is transparency. Many AI algorithms – particularly deep learning algorithms used for image analysis – are virtually impossible to interpret or explain. If a patient is informed that an image has led to a diagnosis of cancer, he or she will likely want to know why. Deep learning algorithms, and even physicians who are generally familiar with their operation, may be unable to provide an explanation.

5 Conclusion

We believe that AI has an important role to play in the healthcare offerings of the future. In the form of machine learning, it is the primary capability behind the development of precision medicine, widely agreed to be a sorely needed advance in care. Although early efforts at providing diagnosis and treatment recommendations have proven challenging, we expect that AI will ultimately master that domain as well. Given the rapid advances in AI for imaging analysis, it seems likely that most radiology and pathology images will be examined at some point by a machine. Speech and text recognition are already employed for tasks like patient communication and capture of clinical notes, and their usage will increase.

The greatest challenge to AI in these healthcare domains is not whether the technologies will be capable enough to be useful, but rather ensuring their adoption in daily clinical practice. For wide-spread adoption to take place, AI systems must be approved by regulators, integrated with EHR systems, standardised to a sufficient degree that similar products work in a similar fashion, taught to clinicians, paid for by public or private payer organisations and updated over time in the field. These challenges will ultimately be overcome, but they will take much longer to do so than it will take for the technologies themselves to mature. As a result, we expect to see limited use of AI in clinical practice within 5 years and more extensive use within 10.

It also seems increasingly clear that AI systems will not replace human clinicians on a large scale, but rather will augment their efforts to care for patients. Over time, human clinicians may move toward tasks and job designs that draw on uniquely human skills like empathy, persuasion and big-picture integration. Perhaps the only healthcare providers who will lose their jobs over time may be those who refuse to work alongside artificial intelligence.

5 CONCLUSION

Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. Authors are strongly

encouraged not to call out multiple figures or tables in the conclusion—these should be referenced in the body of the paper.

REFERENCES

- [1] Deloitte Insights State of AI in the enterprise. Deloitte, 2018.
- [2] W Lee SI, Celik S, Logsdon BA, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid -leukemia. *Nat Commun* 2018;9:42. [PMC free article] [PubMed] [Google Scholar]
- [3] Sordo M. Introduction to neural networks in healthcare. *Open-Clinical*, 2002.
- [4] Fakoor R, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. A conference -presentation The 30th International Conference on Machine Learning, 2013. [Google Scholar]
- [5] Vial A, Stirling D, Field M, et al. The role of deep learning and -radiomic feature extraction in cancer-specific predictive modelling: a review. *Transl Cancer Res* 2018;7:803–16. [Google Scholar]

A Study of Mobile Cloud Computing: Applications and Security Issues

Shreyash Sonawane, Darshana Wajekar

Abstract - The mobile technology becomes robust in the present scenario because now days, mobile devices become capable to support huge variety of applications. On the second thought, the mobile devices are experiencing many challenges as they have narrow pool of resources like battery life, bandwidth and capacity for storing data etc. But after the emergence of mobile cloud computing technologies, the mobile services become more prominent. Mobile Cloud Computing (MCC) refers to cluster of cloud computing and mobile networks where mobile applications transfer the computation power and storage from the mobile devices to the cloud to enhance the capabilities of mobile devices. In MCC, computation offloading is a technique that helps in transferring the complex application modules that require intense computation from a mobile device to the resource-rich cloud. The benefits of cloud based computation offloading are that it helps in improving the performance of integrated application, also enhances the overall execution time of the application and improves battery life of mobile devices. This paper focuses on offloading benefits, its techniques, its related issues and challenges.

Index Terms - Cloud Computing, Mobile Cloud Computing, Computation Offloading.

1 INTRODUCTION

Before the concept of MCC, the technology of Cloud Computing was introduced in 1996. Cloud Computing is an emerging computing technology that preserves data and applications by using the internet and central remote servers. It helps in configuring and customizing the applications efficiently. With Cloud Computing users can access the resources through internet from anywhere, at any time with pre- defined charges which is known as “pay-as-you-use principle.”

In order to take the advantage of Cloud Computing, an abstract idea of MCC was introduced. MCC is the concept that combines the Cloud Computing and the mobile networks. In MCC, the data computation as well as data storage takes place in remote cloud servers, outside of the mobile device. This is because mobile devices are resource-constrained devices; they have fewer resources (i.e. battery life, network bandwidth and storage capacity). MCC provides mobile device users with data storage and processing services in clouds, as all the resource intensive computing can be carried out in cloud. Therefore, there is less consumption of mobile device resources and hence making mobile devices more efficient

1.1 The Architecture of MCC

MCC architecture is as follows: Multiple mobile devices are linked to the mobile networks through base stations such as BTS, satellite or access points

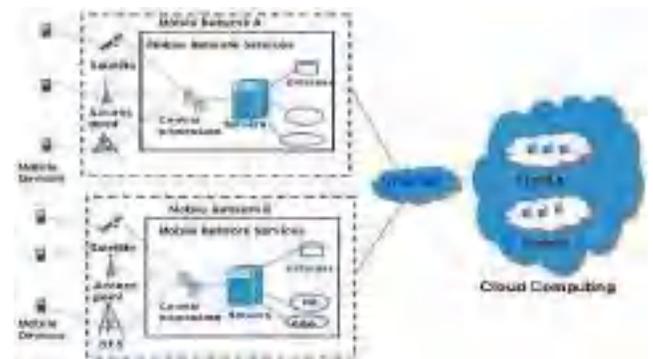


Figure 1. MCC Architecture [1]

These are capable to create and control the connections and functional interfaces between the networks and mobile devices. Mobile user's information and requests are transmitted to the central processors that are connected to servers, providing mobile network services. Now the users are given the services (like authentication, authorization, accounting based on data stored in database) by the mobile network operator. After authorizing user, the cloud server receives the users request with the help of internet. Basically, internet plays the role of intermediary that sends the user requests from mobile networks to the cloud. Afterwards, cloud controllers present in the cloud are responsible to process the user requests and provide the corresponding cloud services.

1.2. Applications In MCC:

Mobile applications have been used on large scale and have a great share in a worldwide mobile market. Various mobile applications like mobile commerce, mobile learning, mobile sensing, mobile healthcare, mobile gaming, multimedia sharing, mobile social networking and many more; have taken advantage of mobile cloud computing. In this section, some of the typical MCC applications are discussed briefly.

1.2.1. MobileGaming

Mobile games expected to have small scope as mobile devices lack high processing power, which is required for the graphic rendering. As a result, it generally depends on simple play rather than graphics. Due to this, mobile gaming use to transfer the game engine to the cloud as it is computation intensive and local device is not able to handle such application modules that needs high computing power and cloud in return sends the output after computing which the game users receive on their mobile devices. This standard brings many benefits such as power saving, enhancing game playing speed due to cloud's high computation speed. MAUI (memory arithmetic unit and interface), is a system that emphasizes more on energy saving, so at the run time, it divides the application codes based on the cost of the network communication and CPU processing speed on the local mobile device. The results show it helps in energy reduction remarkably for the mobile device. Moreover, the performance of mobile applications is also improved. The diagnosed statistics shown by MAUI are: energy saved for video game is 27% and 45% for chess while the refresh rate of game maximizes from 6 to 13 frames per second.

1.2.2. Mobile Learning

Mobile learning applications enable users to have distance learning via mobile gadgets such as tablets, notebooks, mobile phones etc. These m-learning applications are basically the electronic learning with mobility. It provides m-learners flexibility, as they can easily access the applications from anywhere, at any time, from any portable device and in any format. In comparison to traditional m-learning applications, Cloud-based m-learning applications provide users distance education at faster processing speed with more educational resources at high network transmission rate. The interaction status between the students and teachers has increased with the evolution of m-learning when combined with cloud computing. Through a web site built on Google Apps Engine, students [9] can interact with their teachers at any time. Also, the information regarding the student's understanding level of the course can be drawn by the teachers and can provide solutions to student's queries in a timely manner

1.2.3. Mobile Healthcare

Mobile healthcare applications are medical applications in the mobile environment used for diagnosing the health issues, helps in providing medical treatment etc. In tradition medical applications have the limitations such as less storing capacity, minimal security, and high rate of medical errors. But, these limitations are diminished after applying the MCC concept in health applications. Mobile healthcare applications enable users to access the resources at faster pace at any time, from any place. With the use of cloud, these applications provide variety of on-demand services on clouds. Using such mobile healthcare applications, there arise challenges to safeguard the user's health information, so there have to be proposed solutions to increase the privacy of the users, as have to be done in the traditional applications.

Therefore, a solution for protecting participant's health information is proposed by [10]. P2P model has been used to enhance the privacy of the services provided to mobile users. This model associates the clouds to report security related issues and safeguards the data.

1.2.4. Mobile Commerce:

Mobile commerce is the capability that provides commerce using a mobile device. The m-commerce applications handle multiple tasks that need mobility

functions such as mobile banking, mobile money transfer, mobile marketing, mobile browsing, mobile purchasing etc. M-commerce related services are increasing rapidly, due to this there arouse various challenges such as low network bandwidth, and security related problems.

Therefore, 4G E-commerce platform based on cloud computing [14] mingles the benefits of both cloud computing and 4G network to enhance the speed at which the data is being processed and security level [15] based on PKI (public key infrastructure), which operates on an encryption-based access mechanism.

It helps in safeguarding the subscriber's privacy and it also provides secure access to data stored in cloud.

2. CONCEPT OF OFFLOADING IN MCC

Offloading is a method of migrating the application modules, that require complex processing and heavy computations, from the local mobile devices (i.e. resource-constrained devices) to remote cloud servers (i.e. resource rich devices).

The concept of MCC involves the offloading of the task that is to be executed by the remote server. The part of application that need to be offloaded from mobile phone to cloud could be done in two fashions that is partial offloading or full offload-

ing [27,28]. In the full offloading architecture the full application along with all the data associated to it has been offloaded to the cloud where the entire computation take place and the final results have been sent back to the mobile device as shown in figure 2.

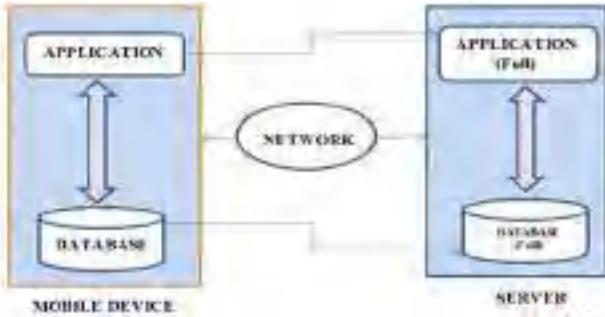


Figure 2. Full Offloading

In the partial offloading architecture, only that part of the application which consumes more energy or have high complexity in terms of computation has been offloaded to the cloud. In this, both mobile phone and the cloud are responsible for the computation and final results comes after merging the individual results of both the computations that is in mobile device and at the cloud as shown in fig.3.

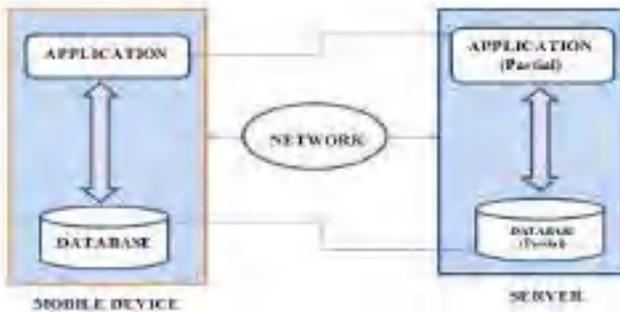


Figure 3. Partial Offloading

Offloading decision helps to: (i) improve performance, (ii) save energy, (iii) Increase storage capacity, (iv) Increasing reliability

Conditions for improving performance and saving energy through offloading are presented [19]. These conditions are as follows

1	Amount of computational data
2	Speed of the mobile system
3	Sending input data
4	Total Time to execute the application in mobile device

5	Server Speed
6	Total Time to offload and execute application in cloud
7	Bandwidth
8	Power on the mobile system
9	Power required to send data from the mobile system over the network
10	Power consumption during execution of data in server

2.1 Techniques for Offloading of an Application:

2.1.1. Static Offloading is easier to implement as the application is partitioned into modules during the early phase of development. In static environment, criterions like size of the data, energy consumption, and execution time are known beforehand, however it may not yield the promised offloading benefits. The related work on static offloading has been discussed further.

On the basis of estimation of energy consumption, program partitioning takes place before the program execution. Through calculating trade-off between the communication and computation costs, the optimal program partitioning for offloading [3] is calculated. But, offloading scheme [4] based on describing information about computation time and data sharing at level of procedure calls is described, where the cost graph is constructed and if we apply the branch and bound algorithm on it, then this result in minimization of total energy consumption.

Before the program implementation, on basis of estimation of the energy expenditure, program splitting is done. On the basis of trade-off between the calculation and transmission rates, the optimum program splitting for offloading is computed. Some explanations are supported to find the optimal result for partitioning functions before offloading.

For some offloading methods, the prior estimation of execution time is not required. For such methods online statics of computation time [2] are used in order to calculate the optimal timeout, and if the computation is not completed within time limit, then it is migrated to remote servers. It results in saving up to 17% energy as compared to the other methods. An automatic distributed partitioning system (ADPS) called Coign, [26] demonstrates the optimal distribution of program that is automatically converted to the distributed application. In this, conversion takes place without using the source code of a program and a graph model is constructed to observe the distri-

bution.

2.1.2. Dynamic Offloading is harder to implement as in dynamic environment the modules of an application may be transferred to cloud when the application is running, however it may result in higher performance and may yield maximum benefits since offloading is performed only when necessary.

Offloading systems in wireless environments [5] is analysed, this examines three circumstances in order to evaluate the fact that how efficient the offloading is when we execute the application on cloud servers. Firstly, the whole application is executed locally on mobile device. In this, no offloading takes place. Secondly, the offloading is performed but without considering any failure. In the third aspect, the offloading takes place but with the failure recoveries.

The sub-modules that failed to offload, they are re-offloaded after failure.

A system follows three steps for partitioning [6] an application. The first step is the application structuring. In this, the application runs dynamically between the mobile and cloud. At the run time, the application decides which module should run at the cloud server and which part of the application to be executed locally. Second step involves the partitioning choice. In order to minimize the energy consumption, the system applies the optimal strategy. Third step is the security. Rather than offloading the application module containing sensitive data, it is executed locally on mobile device.

The decision between the local and remote executions can be made by dynamically evaluating [22] the trade-offs between the computational cost and communication cost. This evaluation is based on input parameters and channel conditions.

3. ISSUES AND CHALLENGES IN OFFLOADING

3.1. Security: Security is the foremost issue of the users in mobile environment. The major concern is the protection of private data of users. For instance, the sensitive data of the user stored in the cloud can be misused by the cloud service provider, without the awareness of the user. For example, when visualizing the integrated global positioning system (GPS) devices [12], the privacy concern arises. The GPS devices are considered to be more security intensive because it can cause subscribers to be tracked. Hence, it is a major concern to provide security and privacy to users.

3.2. Authentication: As huge amount of data/applications are stored on a cloud both the cloud providers

and the users should be careful while tackling with the sensitive data or applications. Such sensitive data need to be authenticated to avoid any misuse. Users should have authorised keys [13] while using external resources, therefore secure authentication mechanisms should be implemented.

3.3. Low Bandwidth: Bandwidth is the critical issue in mobile cloud environment. Bandwidth utilisation [16] relies upon the sub modules of the application that are offloaded from local mobile device to the remote cloud. If the offloaded data is huge then it results in the delay between the transferring the data on cloud and the final result to be sent back to local mobile device which in turn results in less efficiency and high energy consumption.

3.4. Heterogeneity: In mobile environment, WCDMA, GPRS, WiMAX, and WLAN are various network types that are used simultaneously for mobile communication. As a result, it becomes very difficult to handle heterogeneous network [13] connectivity while

fulfilling mobile cloud computing requirements (i.e. scalability, improved mobility, on-demand availability etc.). So, heterogeneity is a major problem in computational offloading.

4. CONCLUSION

MCC is one of the emerging mobile technologies as it collaborate the benefits of both mobile computing and cloud computing, thereby delivering best services for mobile users. We discussed few mobile applications that are taking advantage of MCC because mobile devices do not encourage some complex applications to be computed locally as these devices face deficiency of enough hardware, software and battery lifetime. Offloading is technique of MCC that migrates the heavy computations to cloud servers which are more resourceful and the results are received by the mobile systems. This method enables the mobile device users to access all those applications which require large memory storage, high computational power, long battery lifetime and large network bandwidth. In this paper, we analyse the situation in which performance can be improved and energy can be saved by offloading. We listed down some of the research areas related to offloading. Further this paper presented some major issues in offloading such as security, authentication, low bandwidth, heterogeneity. These issues eventually devastate the efficiency of the process. So, these issues should be taken into consideration for the future work.

A Study of Data Integrity and Data Storage Security in Cloud Computing.

Shreyash Sonawane, Darshana Wajekar

Abstract - Cloud Computing is providing service like servers, database management, data storage, networking, software applications, etc. to the client on demand. Data storage is one of the most widely used service provided by cloud computing. Cloud service provider hosts the data of the owner on their server, and it can be accessed by the owner from these servers whenever they want. As owners and servers are different identities, the paradigm of data storage brings up many security challenges. An independent mechanism is required to make sure that data is correctly hosted in to the cloud storage server. In this paper we will discuss about the integrity of data stored, challenges faced to secure the data and techniques used to secure data storage in cloud.

Index Terms:- Cloud Computing, Data Integrity, Data Storage, Data Security.

◆

1. INTRODUCTION

Cloud Computing in delivery of computing services to individuals and businesses over the internet. Cloud services allow them to utilize system software and hardware that are managed by the cloud providers and allow them to access them anywhere, anytime through internet. Examples of cloud services contain online file storage, webmail, social networking sites, online business applications. Many companies, organizations as well small entrepreneurs are moving into cloud because it provides great services in cheap rate and it is reliable as well. Cloud providers provide only that amount of services that user has asked and charges only for that. Because of services provided by cloud the user do not have to install their own hardware and software and servers as well. It saves energy and money of the users. However big responsibility comes with bigger risks, Data breaching is possible in cloud environment since data from various users and organizations lie together in cloud. As the user not only sends data to the cloud but also transfers controls to a third person which may raise concern about its security. The major task of CSP (Cloud Service Provider) is to preserve the confidentiality, integrity, and availability of data to the users. But sometimes the CSP itself may use or corrupt the data illegally. However, to overcome this the user simply can encrypt the data before uploading it onto the cloud. This will make sure that the data uploaded is not visible to external users or the cloud administrator. However, it has limitations too that plain text-based searching algorithm are not applicable. In this paper, we discuss the security algorithms that can be used to keep the data secure and

maintain its integrity.

1.1 USE OF CLOUD COMPUTING IN VARIOUS SECTORS

- **Education**:- In education, it helps both teachers as well as students. Teachers can upload their learning materials whenever they want, and students can access them whenever needed via internet. Students can store their notes as well share with their classmates on that platform.
- **Healthcare**:- Cloud Computing in Healthcare makes the industry more efficient and reduces cost as well. It makes sharing record of a patient convenient and makes accessing it easier. Once a patient is diagnosed and its record uploaded to the cloud it becomes easier to maintain it an update it as well. It also helps in creation and maintenance of healthcare applications.
- **Business**:- The Cloud Computing model offers Infrastructure as a Service and Platform as a Service as well. With IaaS you can offer a client to use our infrastructure to host their cloud devices or create website to promote clients products and services. Alternatively, you can offer your platform to clients using cloud computing. Companies are accepting cloud in order to eliminate their computer hardware reducing the cost.

1.2 CHARACTERISTICS OF CLOUD COMPUTING

- **Large-scale:-**In large-scale computing, the scale of cloud is large convergence. The cloud of Google has owned more than one million servers access. For example, IBM, Microsoft, Yahoo, Rediff, Amazon they have more than hundreds of thousands of servers. There are a lot of servers in an enterprise control access.
- **Virtualization:-** With the help of virtualization users share the Infrastructure hence, Virtualization has a very important role in the cloud computing technology. Virtualization allows users to share the data present in the clouds like application etc. Just by using smartphone, tablet, laptop, or notebook pc users can complete the process through an internet service. Users can get the task done or share it securely through a simple way, anytime, anywhere. Users can complete a task that cannot be completed in a single computer.
- **High reliability:-** Reliability in Cloud computing is measured by the application or service provided by the provider is running smoothly and the user can access it from anywhere anytime. There should be no downtime or interruptions for the services and the connection should be secure. Proper planning and risk detection and solving makes the cloud more reliable.
- **Service on Demand:-** Cloud is an enormous resource pool, which you will pay according to your requirement. Cloud is just like that running water, electric, and gas that can be charged by the amount that you used.
- **Inexpensive:-** Cloud computing reduces the IT expenditures of Organizations and companies those invest in them as they don't need to install their own hardware or carry specific software requirements. The users only need to pay the space they are using to the cloud administrator, there's no need to pay any extra charges moreover, some space is allotted for free. As the user don't need to install their own hardware, servers there's a lot of energy is being saved which also makes it more economical.

2. SECURITY IN CLOUD COMPUTING

The main concern is to protect security against unauthorized access of data. Data relocation on high level has negative implications for protecting the data safety and data security as well as data accessibility. Therefore, the main apprehension with reference to safety of data residing within the Cloud is: at the remainder how to safe security and avoid unauthorized. Even though, customers understand the situation and no data mobility access, question with reference to security and confidentiality of data. The Cloud Computing area has no confusion become larger as a result of its accessibility and wide network access. However, we can also believe in terms of a secure and safe atmosphere for the personal data and information of the user is being needed.

a. CLOUD COMPUTING SECURITY ELEMENTS

The cloud computing of the security elements is detailed about Data Integrity, Data confidentiality, Data availability, and Data privacy. In this paper explain the detail explanation about security elements of cloud computing as given below.

Data Integrity:- Data integrity is considered one of the most critical security elements in several information systems. In general, data integrity means that protecting data from unauthorized modification process, fabrication or deletion. Managing entity's rights and access to specific enterprise resources ensures that important data and services are not abused, illegal access, or stolen. Data integrity is definitely achieved in a standalone system with a single database. Data integrity in the standalone system is maintained through database transactions and constraints, which is frequently finished by a database management system (DBMS). Transactions ought to follow ACID (atomicity, consistency, isolation, and durability) properties to ensure data integrity. The majorities of databases are to support ACID transactions and can protect data integrity. Data integrity in the cloud system means that protecting information integrity. The data should not be modified or lost or by unauthorized users access. Data integrity is the basis to give cloud computing service like SaaS, PaaS, and IaaS. Moreover, data storage of large scaled data, cloud computing environment typically provides data processing service. Data integrity can be obtained by using these techniques like digital signature and RAID-like strategies.

Data Confidentiality:- Data confidentiality is very important for users to store their confidential data or private information within the cloud services. In data confidentiality is used to ensure authentication and access control strategies. The data confidentiality, authentication and access control problems are mainly to protect in cloud computing might be self-addressed by improving the cloud reliability and trustiness [8]. As a result of the users don't trust the cloud providers and cloud storage service providers. These are virtually not possible to eliminate potential corporate executive threat; it is terribly dangerous for users to store and protect their sensitive data in cloud storage directly. The simple encryption process is faced with the key type management drawback and cannot support the advanced requirements like parallel modification, query, and fine-grained authorization. The cloud computing have several techniques for enhancing and developing data confidentiality.

1. Homomorphic encryption: encryption is typically used to make sure the data confidentiality. Homomorphic encryption is a kind of encryption system to authority of the data.
2. Encrypted Search and Database: Because the homomorphic encryption algorithm is ineffective. The homomorphic encryption algorithm is the study of the applications of limited within the cloud environment researchers. Encrypted search is a general operation to protect data from unauthorized resources.
3. Distributed Storage: In the Distributive storage of data is also a promising approach in the cloud environment.
4. Hybrid Technique. A hybrid technique is projected for ensuring data integrity and data confidentiality, which uses both key sharing technique and authentication technique.
5. Data Concealment. Data concealment is mainly used to maintain the data confidentiality in the cloud.

Data Availability:-Data availability mean that the following: when accidents occur like hard disk damage, IDC fire, electronic circuit failure, and network failures, the coverage that user's data are often recovered or utilized and how the users verify their data by techniques rather than depending on the credit guarantee by the cloud service provider alone. The transborder servers under the main issue of storing data is detail about a serious concern of clients for the reason that the cloud vendors are governed by the local laws and, as a result, the cloud clients must be cognizant of these laws. In addition, the cloud service provider should make sure the data security, notably data confidentiality and data integrity. The cloud provider should share and

protect all such concerns with the client and establish trust relationship with this connection establishment. The cloud marketer ought to give guarantees of data safety and build a case for jurisdiction of native lawsto the client's process management. Establishing data connection can support users to extend their trust relationship on the cloud. Cloud storage affords the transparent storage service for performing with users, which can reduce the cloud complexity, however, it also reduces the control ability of data storage of users.

3. SECURITY AND PRIVACY ISSUES IN DATA STORAGE:-

Cloud Computing allows the users to store their data on the storage location maintained by a third party. Once the data is uploaded into the cloud the user loses its control over the data and the data can be tampered by the attackers. The attacker may be an internal(CSP) or external. Unauthorized access is also a common practice due to weak access control. The protection of information arises the following challenges: The security and privacy issues related to data storage are confidentiality, integrity and availability.

A. Confidentiality:- The major dispute in cloud computing is confidentiality. Data confidentiality means accessing the data only by authorized users and is strongly related to authentication. In another way confidentiality means keeping users data secret in the cloud systems. As we are storing the data on a remote server and transferring the control over the data to the provider here arises the questions such as: For ensuring confidentiality, cryptographic encryption algorithms and strong authentication mechanisms can be used. Encryption is the process of converting the data into a form called cipher text that can be understood only by the authorized users. Encryption is an efficient technique for protecting the data but have the obstacle that data will be lost once the encryption key is stealed. algorithms. Blowfish is a fast and simple encryption algorithm.

B. Integrity:- Another serious problem faced by cloud computing is integrity. Integrity of data means to make sure that the data has not been changed by an unauthorized person or in an unauthorized way. It is a method for ensuring that the data is real, accurate and safeguarded from unauthorized users. As cloud computing supports resource sharing, there is a possibility of data being corrupted by unauthorized users. Digital Signatures can be used for preserving the integrity of data. The simple way for providing

integrity is using Message AuthenticationCode(MAC).

C. Availability:- Availability refers to being available and accessible to authorized users on demand. The aim of availability in cloud computing systems is to ensure that its users can use them at any place and at any time

4. CONCLUSION:-

Cloud computing proves an extremely successful application for each and every organization's performance. For the reason that organizations have large amount of data to store and cloud provides that space given to user and also enables its user to access their data from anyplace anytime in a simple manner. Improved use of cloud computing for storing data is definitely increasing the trend of improving the ways of storing data in the cloud. As peoples are saving their personal information and important data to clouds, therefore it becomes a major issue to store that data safely. Data available in the cloud can be at risk if not protected in a trustful manner. In the cloud there are a number of existing techniques used to implement security prevention. The study provided an overview and discuss about the cloud computing and security issues and how to improve the security algorithms for cloud computing.

5. REFERENCES:-

- [1] V.Nirmala, R.K.Sivanandhan, Dr.R.Shanmuga Lakshmi, "Data Confidentiality and Integrity Verification using User Authenticator scheme in cloud", Proceedings of 2013 International Conference on Green High Performance Computing (ICGHPC2013). March 14-15, 2013, India.
- [2] Arjun Kumar, Byung Gook Lee, HoonJae Lee, Anu Kumari, "Secure Storage and Access of Data in Cloud Computing", 2012 International Conference on ICT Convergence (ICTC), 15-17 Oct. 2012.
- [3] M.R.Tribhuwan, V.A.Bhuyar, Shabana Pirzade, "Ensuring Data Storage Security in Cloud Computing through Two-way Handshake based on Token Management", 2010 International Conference on Advances in Recent Technologies in Communication and Computing.
- [4] Mr. Prashant Rewagad, Ms.Yogita Pawar, "Use of Digital Signature with Diffie Hellman Key Exchange and AES Encryption Algorithm to Enhance Data Security in Cloud Computing", 2013 International Conference on Communication Systems and Network Technologies.
- [5] Uma Somani, Kanika Lakhani, Manish Mundra, "Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing", 1st International Conference on Parallel, Volume 3,

BRAIN TUMOR SEGMENTATION AND DETECTION

Omkar Gujar, Prof. Darshana Wajekar

Abstract: For study of tumour detection and segmentation the magnetic resonance imaging pictures is extremely helpful in recent years. thanks to magnetic resonance imaging pictures we will sight the tumour. For detection of surprising growth of tissues and blocks of blood in systema nervosum may be seen in an magnetic resonance imaging pictures. the primary step of detection of tumour is to visualize the isosceles and uneven form of brain which can outline the abnormality. once this step ensuing step is segmentation that relies on 2 techniques 1) F-Transform (Fuzzy Transform) 2) Morphological operation. These 2 techniques area unit accustomed style the image in magnetic resonance imaging. currently by this facilitate of style we will sight the boundaries of neoplasm|brain tumour|tumour|tumour|neoplasm} and calculate the particular space of tumor. during this the f-transform is employed to present the bound data like restored of missing edges and extracting the silent edges. Accuracy ANd clarity in an magnetic resonance imaging pictures depends on one another.

Keywords: Tumor, MRI Images, Fuzzy re-model, Morphological operation.



I. INTRODUCTION :

Now days the adult male pictures area unit terribly helpful in a very Medical field like Medical image process. The tumour defines the weird growth of tissues and uncontrolled cells proliferation thus thanks to this the natural pattern of cell growth and death is failing. The tumour is of 2 stages :- 1) Primary stage 2) Secondary stage.

When growth unfold in any a part of brain then it's called tumour. currently once tumour will known variety of symptoms as well as seizures, mood ever-changing, problem in walking and hearing, vision, and muscular movement etc. tumour is assessed into Gliomas, medulloblastoma, epeldymomas, system malignant neoplastic disease and oligodendrogloma.

In primary stage the growth may be removed however in secondary stage ,the growth malady unfold, thanks to this once removal of growth the rarely remains and grow back once more thus this can be the largest downside within the secondary stage of growth . Why this downside happens? It occurs thanks to inaccurately location of space of growth. ensuing step is

detection techniques. during this the any segmentation and detection area unit to live detection techniques the imaging of tumour may be done by- 1) MRI scanning that's magnetic resonant image 2) CT scanning i.e. pc picturing 3) Ultra sound etc.

There area unit many technique to sight an brain neoplasm by that the tumor technique we will diagnose and sight additional simply .some edges area unit nuclear network algorithmic program watershed and edge detection, fuzzy c mean algorithmic program, spatiality of brain is employed to sight an abnormality . The downside of edge detection is that the one amongst the foremost engaging problem for the image process thanks to this it's numerous applications. Candy-edge detection is that the one amongst the foremost helpful feature in image segmentation. during this candy-edge detection is employed for extraction of edges. F-transform is an intelligent technique to handle unsure data. this can be helpful for detection of growth boundaries. it's terribly simple technique for detection may be a promising and economical technique for future and edge extraction progress.

II. BASIC METHODOLOGY :

Fig one shows the fundamental diagram of tumour detection and segmentation. magnetic resonance imaging pictures of brain area unit taken for process.

Image Acquisition: 1st thought-about that the magnetic resonance imaging scan pictures of a given patient area unit either color, Gray-scale or intensity pictures herein area unit displayed with a default size of 220×220. If it's color image, a Gray-scale reborn image is outlined by employing a massive matrix whose entries area unit numerical values between zero and 255, wherever zero corresponds to black and 255 white as an example. Then the tumour detection of a given patient encompass 2 main stages specifically, image segmentation and edge detection.

Pre-processing stage : Pre-processing stage consists of Noise removal this could be done by exploitation numerous abstraction filters linear or nonlinear filters (Median filter). different artifacts like text removed by some morphological operations. RGB to gray conversion and reshaping conjointly takes place here. It includes median filter for noise removal. The possibilities of arrival of noise in trendy magnetic resonance imaging scan area unit terribly less. it should arrived thanks to thermal impact. Image Smoothing: it's the action of simplifying a picture whereas conserving vital data. The goal is to cut back noise or useless details while not introducing an excessive amount of distortion thus on modify subsequent analysis.

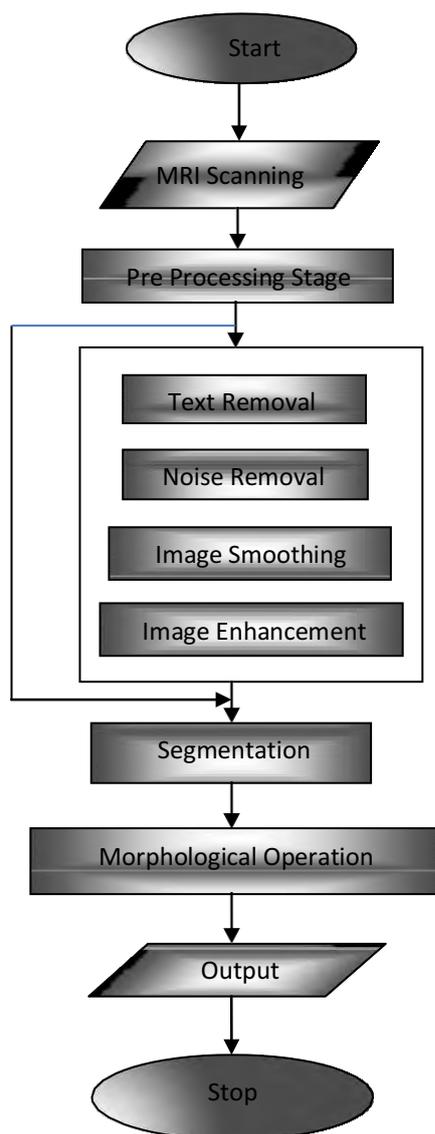
Image Registration: Image registration is that the method of transfer 2 or additional pictures into abstraction correspondence (aligning them). within the context of medical imaging, image registration permits for the synchronal use of pictures smitten completely different modalities (e.g. magnetic resonance imaging and CT), at

{different|totally completely different|completely different} times or with different patient positions. In surgery, for instance, pictures area unit nonheritable before (preoperative), still as throughout (intra-operative) surgery. attributable to time constraints, the time period intraoperative pictures have a lower resolution than the pre-operative pictures obtained before surgery. Moreover, deformations that occur naturally throughout surgery build it tough to relate the highresolution pre-operative image to the lowresolution intra-operative anatomy of the patient. Image registration makes an attempt to assist the Dr. relate the 2 sets of pictures [8].

Image Segmentation: The segmentation is that the most vital stage for analysing image properly since it affects the accuracy of the next steps. However, correct segmentation is tough attributable to the good verities of the lesion shapes, sizes, and colours at the side of completely different skin sorts and textures. additionally, some lesions have irregular boundaries and in some cases there's sleek transition between the lesion and also the skin. to handle this downside, many algorithms are projected. they will be generally classified as thresholding, edge-based or region-based, supervised and unsupervised classification techniques • Threshold segmentation • Water shed segmentation • Gradient Vector Flow (GVF) • K-mean cluster • Fuzzy C-means cluster.

Morphological Operations : Once segmentation morphological process is applied to get rid of unwanted half. It consists of image gap, image closing, dilation, erosion operations. At the top the choice has taken weather that magnetic resonance imaging image consists of any growth or not and weather it traditional or abnormal.

Fig. 1. Basic diagram of tumour detection and segmentation



III. REVIEW OF THE VARIOUS PAPERS:

The 2016 WHO i.e. world health organization on classification of tumour of central systema nervosum is Associate in Nursing conceptual furthermore as pertain summary of precursor .the WHO classification system tumour that is employed molecular parameters for its identification structure. more than 2016 system WHO presence the new diffuse glomas and alternative tumour and defines the new feature like each microscopic anatomy furthermore as molecule [1]. The fourth edition of the WHO i.e. world health organization classification of

tumour of central systema nervosum revealed in 2007.there square measure many new titles and knowledge list as well as brain tumor, papillary, glioneuronal tumour etc. the microscopic anatomy variants square measure capable of various edge distribution, location, symptoms and also the behaviours or clinical [2]. Fuzzy clump is technique that wide used medicine to discover the image. The effective fuzzy clump algorithmic rule is employed in abnormal Mr brain image segmentation. By mistreatment clump in tumor segmentation we will diagnose accurately the region of cancer.to provides higher identification of tumor magnetic resonant pictures is applied [3]. Now a day's tumor is one in every of the foremost hazards diseases thus its detection ought to be quick and correct. It will be achieved by machine-controlled tumour detection techniques on medical pictures and one in every of the machine-controlled tumour detection techniques is imaging pictures .which defines the tumour growth region and also the edges detection. As compare to alternative techniques with this is often provides a lot of correct furthermore as clear and blessings of machine-controlled tumour detection techniques is employed for removal of tumour if required [4]. The neural networks could be a new technology has been discovered .the neural network square measure Associate in Nursing "HOT" analysis space, sort of a medical specialty, radiology, medicine etc.to solve extremely advanced drawback 3 is combination of neurons into layers permits for artificial neural network. In Associate in Nursing medical applications the neural network square measure like ANNs etc. and also the medical application the neural network square measure accustomed map Associate in Nursing input into a desired output [5]. It is a brand new technique of detection of tumor and for excellent result and accuracy. The watershed technique is combined with edge detection operation. The colour brain imaging pictures will be obtained by this algorithmic rule. during this the RGB image is converts into on HSV color image so the image is separated in three regions that square measure

called hue, saturation and intensity. The clever edge detector is applied is applied to Associate in Nursing output image for remodeled method of edge happens during this .at last combining the 3 pictures and also the final resultant tumor divided image is obtained. This algorithmic rule is applied on twenty brain imaging pictures for glorious result [6]. In Associate in Nursing imaging image the extremely irregular boundaries of tumour tissues is seen. For a segmentation of medical image, the deformable modes and region base strategies square measure used. the most issues square measure there in imaging pictures like undefinable location of tumour square measure unseen boundaries or knowledge loss at boundaries and a silent edge not extended. By mistreatment this algorithmic rule the silent edge is extended and located boundary of tumour location or space and once the boundary or location of tumour is seen clearly. Then removal of tumour will be manifest itself [7].

IV. CONCLUSION :

In this paper, we've got planned completely different techniques to find and phase tumour from magnetic resonance imaging pictures. To extract and phase the tumour we have a tendency to used completely different techniques like Kyrgyzstani monetary unit bunch, k-mean bunch, Fuzzy C-mean technique, curvelet remodel. It will be seen that detection of tumour from magnetic resonance imaging pictures is finished by varied ways, conjointly in future work completely different automatic ways deliver the goods additional accuracy and additional economical.

REFERENCES

[1] D.N.Louis, et al,"The 2007 WHO classification of tumor of central nervous system," *Acta neuropathologica*, vol114, pp97-109/2007.
[2] P.Kleihues, et al."The new WHO classification of brain tumor," *brain pathology*, vot3, pp255-268, 1993.
[3] D.J.Hemanth, et al,"effective fuzzy clustering algorithm for abnormal MR brain image segmentation," in *advance computing conference 2009.IACC 2009.IEEE international*, 2009, pp-609-614.

[4] S.chrutha and M.J.Jayashree, "An efficient brain tumor detection by integrating modified texture based region growing and cellular automata edge detection, " in *Control Instrumentation, Communication and Computational Technology (ICCICCT)*, 2014 International Conference On, 2014, pp.1193-1199.
[5] A. Abdullah, et al., "Implementation of an improved cellular neural network algorithm for brain tumor detection," in *Biomedical Engineering(Isobel)*, 2012 International Conference on, 2012, pp. 611- 615.
[6] I. Maiti and M. Chakra borty, "A new method for brain tumor segmentation based on watershed and edge detection algorithms in HSV color model," in *Computing and Communication Systems (NCCCS)*, 2012 National Conference on, 2012, pp. 1-5.
[7] R. Preetha and G. R. Suresh, "Performance Analysis of Fuzzy C Means Algorithm in Automated Detection of Brain Tumor," in *Computing and Communication Technologies (WCCCT)*, 2014 World Congress on, 2014, pp. 30-33.
[8] Bandana Sharma et al. "Review Paper on Brain Tumor Detection Using Pattern Recognition Techniques" published in *International Journal of Recent Research Aspects* ISSN: 2349-7688, Vol. 3, Issue 2, June 2016, pp. 151-156.

BRAIN TUMOR SEGMENTATION

Omkar Gujar, Prof. Darshana Wajekar

Abstract- Brain tumor is one of the leading causes of cancer death. Accurate segmentation and quantitative analysis of brain tumor are critical for diagnosis and treatment planning. Since manual segmentation is time-consuming, tedious and error-prone, a fully automatic method for brain tumor segmentation is needed. Recently, state-of-the-art approaches for brain tumor segmentation are built on fully convolutional neural networks (FCNs) using either 2D or 3D convolutions. However, 2D convolutions cannot make full use of the spatial information of volumetric medical image data, while 3D convolutions suffer from high expensive computational cost and memory demand. To address these problems, we propose a novel Separable 3D ResU-Net architecture using separable 3D convolutions and the BRATS dataset.

Index Terms- Fully Convolutional Neural Networks (FCNs), 3D ResU-Net, BRATS dataset, 3D convolutions , Tversky.



1. INTRODUCTION:

Image segmentation, especially semantic segmentation, is a fundamental and classic problem in computer vision. It refers to partitioning an image into several disjoint semantically meaningful parts and classifying each part into a pre-determined class. In the application of brain tumor segmentation, the task includes the division of several sub-regions, such as GD-enhancing tumor, peritumoral edema, and the necrotic and non-enhancing tumor core. Accurate segmentation and quantitative analysis of brain tumor are critical for diagnosis and treatment planning. Generally, manual segmentation of brain tumor is known to be time-consuming, tedious and error-prone. Therefore, there is a strong need for a fully automatic method for brain tumor segmentation. However, brain tumor segmentation is a challenging task because MR images are typically acquired using various protocols and magnet strengths, which results in the non-standard range of MR images. Brain tumor segmentation methods can be divided into different categories according to different principles. Broadly, these methods can be divided into two major categories: generative methods and discriminative methods. Generative methods typically rely on the prior information about the appearance of both healthy tissues and tumors. The proposed models often regard the task of segmentation

as a problem of a posteriori distribution estimation. On the contrary, discriminative methods use very little prior information and typically rely on a large number of low-level image features to learn the distribution from the annotated training images. Inspired by S3D architecture for video classification and the state-of-the-art U-Net architecture for medical image segmentation, we propose a novel framework named S3D-UNet for brain tumor segmentation. To make full use of 3D volumes, we design a new separable 3D convolution by dividing each 3D convolution into three branches in a parallel fashion, each with a different orthogonal view, namely axial, sagittal and coronal.

DESCRIPTION OF PROBLEM :

A brain tumour is a collection of irregular cells in the brain that may develop into benign or malignant tumours. The cause and location of brain tumours are also different. Primary brain tumours are cancerous or non-cancerous tumours that grow directly in the brain. Gliomas and medulloblastomas are two of the most common malignant tumours. Hemangioblastomas, meningiomas, pituitary adenomas, osteomas, pinealomas, and other benign primary tumours are rarer than

malignant primary tumours. Secondary brain tumours arise from tumour cells that have spread to the brain and multiplied there from other organs that have already been affected by cancer. Breast cancer, prostate cancer, skin cancer (melanoma), kidney cancer, colon cancer, and other cancers are among the most common tumours that cause brain metastases. Since the causes are still unknown, when symptoms appear, a radiological test with a biopsy will be performed to decide the best course of action. A tumour, whether cancerous or not, creates prisms in some parts of the brain as it develops. Treatment varies depending on the type of tumour and its location; typically, surgery is used to remove the tumour if it is available, and otherwise, radiation therapy is used. Treatment varies depending on the type of tumour and its location; typically, surgery is used to remove the tumour if it is available, and radiosurgery is used if the tumour is malignant.

We use segmentation to effectively identify and segment brain tumours for effective surgery. Brain tumours can be segmented in two ways. The first is manual segmentation, which is a discretionary decision that does not provide the desired results since completely extracting brain tumours without destroying surrounding brain tissue, i.e. healthy tissue, is difficult. As a result, automatic segmentation, the second method, is required for clinical planning and quantitative assessment. It identifies brain tumours rapidly and accurately.

In addition, there are far too many methods for segmenting medical images. Classical approaches, machine learning approaches, and those focused on deep learning are the three groups. Regional methods, as used in the past, seek to identify homogeneous areas of images, indicating the existence of regions. Contour approaches, on the other hand, search for picture discontinuities that indicate the existence of area boundaries. Brain tumours are

often detected and segmented using machine learning, or more specifically clustering. Because of their ease of execution, K-means and GMM are the most well-known and widely used in unsupervised learning. We built a completely convoluted 2D segmentation network using the ResUNet architecture in our project. We used a Tversky loss to adjust to unbalanced samples. The Cancer Imaging Archive - TCIA datasets were used to train and test the proposed system. Our fully automated system resulted in superior tumour region segmentation as compared to the reality of the manually delineated ground.

2. LITERATURE SURVEY :

Brain tumour segmentation has been a widely studied area of research since the early days of the Internet. The authors of paper [1] proposed a technique where a set of features extracted by the discrete wavelet transform (DWT) feature extraction technique from the segmented brain MRI images, to train the DNN classifier for brain tumors classification. In paper [2], the authors present a novel CNN architecture which differs from those traditionally used in computer vision. Their CNN exploits both local features as well as more global contextual features simultaneously. Also, different from most traditional uses of CNNs, our networks use a final layer that is a convolutional implementation of a fully connected layer which allows a 40 fold speed up. They also describe a 2-phase training procedure that allows us to tackle difficulties related to the imbalance of tumor labels. Finally, they explore a cascade architecture in which the output of a basic CNN is treated as an additional source of information for a subsequent CNN which is 30 times faster than state of art representation.

Status of research/knowledge in the field and Literature Review :

Paper 1

Automated brain tumor segmentation on multi-modal MR image using SegNet

Date: 22 July 2017

Abstract:

The potential of improving disease detection and treatment planning comes with accurate and fully automatic algorithms for brain tumor segmentation. Glioma, a type of brain tumor, can appear at different locations with different shapes and sizes. Manual segmentation of brain tumor regions is not only time-consuming but also prone to human error, and its performance depends on pathologists' experience. In this paper, we tackle this problem by applying a fully convolutional neural network SegNet to 3D data sets for four MRI modalities (Flair, T1, T1ce, and T2) for automated segmentation of brain tumor and sub-tumor parts, including necrosis, edema, and enhancing tumor. To further improve tumor segmentation, the four separately trained SegNet models are integrated by post-processing to produce four maximum feature maps by fusing the machine-learned feature maps from the fully convolutional layers of each trained model. The maximum feature maps and the pixel intensity values of the original MRI modalities are combined to encode interesting information into a feature representation. Taking the combined feature as input, a decision tree (DT) is used to classify the MRI voxels into different tumor parts and healthy brain tissue. Evaluating the proposed algorithm on the dataset provided by the Brain Tumor Segmentation 2017 (BraTS 2017) challenge, we achieved *F*-measure scores of 0.85, 0.81, and 0.79 for whole tumor, tumor core, and enhancing tumor, respectively.

Experimental results demonstrate that using SegNet models with 3D MRI datasets and integrating the four maximum feature maps with pixel intensity values of the original MRI modalities has potential to perform well on brain tumor segmentation.

Inference:

The complete *F*-measure obtained in this research paper was 0.85 and it was implemented on BRATS 2017 dataset.

Paper 2

Evaluation of Three Methods for MRI Brain Tumor Segmentation

Date: 09 March 2012

Abstract:

Imaging plays a central role in the diagnosis and treatment planning of brain tumor. An accurate segmentation is critical, especially when the tumor morphological changes remain subtle, irregular and difficult to assess by clinical examination. Traditionally, segmentation is performed manually in clinical environment that is operator dependent and very tedious and time consuming labor intensive work. However, automated tumor segmentation in MRI brain tumor poses many challenges with regard to characteristics of an image. A comparison of three different semiautomated methods, viz., modified gradient magnitude region growing technique (MGRRGT), level set and a marker controlled watershed method is undertaken here for evaluating their relative performance in the segmentation of tumor. A study on 9 samples using MGRRGT reveals that all the errors are within 6 to 23% in comparison to other two methods.

Inference:

This was a comparative study of three methods, and on a limited sample showed an error of 6-23%.

Block diagram:

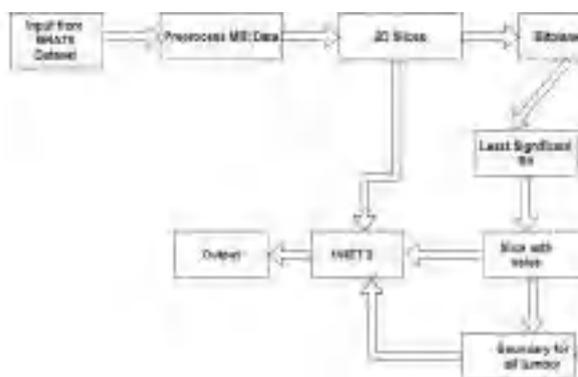


Fig 1: Block Diagram

3. METHODOLOGY:

We want to use ResUNet, which is based on U-Net, to solve the problem of brain tumour segmentation in this project. As seen in, the network architecture is an encoder decoder architecture, which means it has a down sampling path and an up sampling path. The contracting section, the bottleneck part, and the expansion part are the three parts of this architecture. Four convolutional blocks make up the contracting direction. Each block is made up of three convolutional layers, each with a 3 x 3 filter, one-by-one stride, and a batch normalised rectifier (ReLU) activation feature. In this contracting path with a stride of 2 x 2, maxpooling is applied at the end of each block. The number of function maps increases from 1 to 1024 at each contracting tab as a result of the pooling. There is no maximum pooling in the bottleneck part, which is made up of three simple convolutional layers with batch normalisation. Four blocks are included in the expansion route. Unlike the original ResUNet architecture, we use three four extra convolution layers per block to achieve more efficient brain tumor segmentation, and zero paddings to maintain the output dimension for all convolutional layers of the encoding and decoding path. Finally, one-by-one convolution on the final layer used to map each element of the feature vector according o the desired number of classes. We did not apply L1, L2 or dropout to the network because there is no significant improvement in performance.

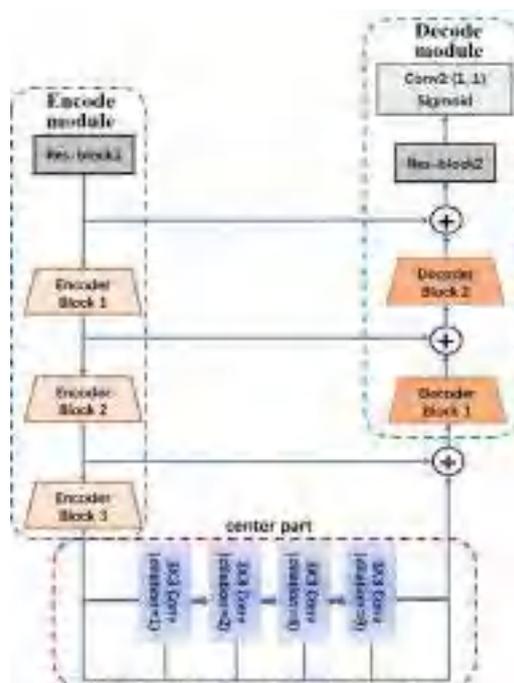


Fig 2: Seg Model

Model: "seg_model"

Total params: 1,210,513
 Trainable params: 1,206,129
 Non-trainable params: 4,384

Results (analysis and associated uncertainties)

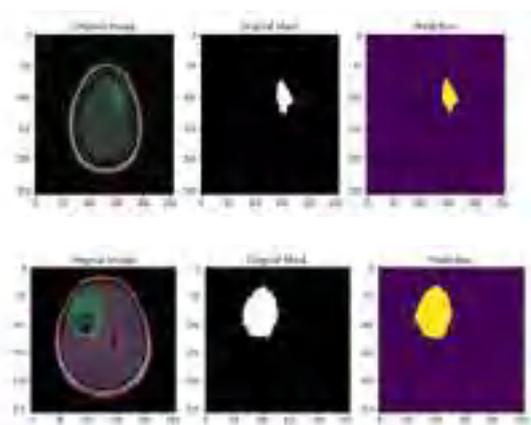


Fig 3: Result Analysis

FEATURES:

- 1. Collection of the Dataset:** In this we collect the dataset using available resources. We've used BRATS dataset.

2. **Data preprocessing:** In this we perform operations on the dataset so that it gives more generic results while training the model.
3. **Creating UNET Architecture:** We settle on one architecture after an iterative process of training and evaluation for designing our model.
4. **Compiling the model:** We compile the deep learning model using an appropriate loss function and optimizer and learning rate. We also set the hyperparameters and the parameters for the model training.
5. **Training the model and evaluation:** We train the model for a certain number of epochs. We also set checkpoints at the end of every epoch so that the best model is saved. After that we draw the accuracy loss curves to evaluate the performance.
6. **Saving the model and development of Flask application:** After this, we save the trained model and use it in our flask application so that the users also can use the model for marking the tumour affected area of the patient's brain.

4. RESULT:

Since we know, brain tumors are caused by abnormal cells growing in human brain. The current incidence of malignant brain tumors is relatively high, and this occurs a huge influence to humans and society. U-Net has achieved great success in the field of medical image segmentation, and it is also the mainstream of current MRI brain tumor segmentation methods. However, during the process of downsampling, U-Net constantly reduces the dimension of the image, which results in the poor segmentation accuracy for the small-scale tumors. Considering that attention mechanism can enhance local feature expression, to solve the insufficient segmentation accuracy of small-scale tumors by

U-Net, we aim to explore the effectiveness of attention gate, an efficient attention module for image segmentation task, for brain tumor segmentation problem, and a novel Attention Gate Residual U-Net model, i.e., AGResU-Net, is also put forward for this task. Experimental results on three brain tumor segmentation benchmarks illuminate that attention gate is benefit for the brain tumor segmentation task. The main contributions of this paper are summarized in three folds: We propose an end-to-end ResU-Net model for MRI brain tumor segmentation task, and its working is the block diagram. AGResU-Net not only extracts more abundant semantic information, but also pays more attention to the information of small-scale brain tumors, which improves the segmentation effect of brain tumors. ResU-Net integrates residual modules and attention gates with a primeval and single U-Net architecture, in which a series of attention gate units are added into the skip connection for highlighting salient feature information while disambiguating irrelevant and noisy feature responses. On the one hand, residual modules enhance the ability of feature extraction and expression, and contribute to the classification in the process of downsampling. On the other hand, attention gates pays more attention to small-scale tumors and obtain more information about the location of small-scale tumors, so that the upsampling process is helpful to restore the location information of small-scale tumors.

5. CONCLUSION:

We created a fully automated brain tumour segmentation system to enhance patients' quality of life. We used a deep learning approach in this project, using deep convolution networks based on the ResUNet model. The Cancer Imaging Archive (TCIA) datasets, which contain LGG patients, were used to test our system. As compared to the manually delineated ground reality, this approach provided an efficient and robust segmentation, with a Tversky metric of 0.8293 for the test dataset.

This Intelligent Voice Assistant has an enormous and limitless scope in the future. Like Siri, Google Now and Cortana most popular personal voice assistants. The project will easily able to integrate with devices near future for a Connected Home using Internet of Things, voice command system and computer vision.

REFERENCES:

- [1] Işın, C. Direkoğlu and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods", *Procedia Comput. Sci.*, vol. 102, pp. 317-324, 2016.
- [2] S. Bauer, R. Wiest, L.-P. Nolte and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies", *Phys. Med. Biol.*, vol. 58, pp. R97-R129, Jul. 2013.
- [3] H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., "The multimodal brain tumor image segmentation benchmark (BRATS)", *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993-2024, Dec. 2014.
- [5] S. Cui, L. Mao, J. Jiang, C. Liu and S. Xiong, "Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded"

Machine Learning : Multi-Label Classification : MPST : Movie Plot Synopses with Tags : Tags Prediction

Prashant Hule, Darshana Wajekar

ABSTRACT — In today's world social tagging of movies reveals a wide range of heterogeneous information about movies, such as plot structure, genre soundtracks, emotional experiences, metadata & visual. Such information can be valuable in building automatic systems for creating tags of movies. Automatic tagging systems can help recommendation engines to improve the retrieval of similar movies as well as help viewers to know what to expect from a movie in advance. In this paper, we set out to the task of collecting a corpus of movie plot synopses and tags. We describe a methodology that enabled us to build a fine grained set of around 70 tags exposing heterogeneous characteristics of movie plots & the multi-label associations of these tags with around 14K movie plot synopses. We investigate how these tags correlate with movies & the flow of emotions throughout different types of movies. Finally, we use this corpus to explore the feasibility of inferring tags from plot synopses. We expect the corpus will be useful in other tasks where analysis of narratives is relevant.

PROBLEM STATEMENT

Suggest the tags based on the content that was there in the Movie in its Title and Synopsis.

REAL WORLD / BUSINESS OBJECTIVES AND CONSTRAINTS

1. Predict as many tags as possible with high precision and recall.
2. Incorrect tags could impact customer experience.
3. No strict latency constraints.

REVIEW OF LITERATURE

Machine Learning problem

**Data Overview : All of the data is in 1 file:
mpst_full_data.csv**

Data is divided into 3 Sections Train, Test and Val data within the file

Train Consist of 64% of the data.

Test Consist of 20% of the data.

Val Consist of 16% of the data.

Number of rows in mpst_full_data.csv = 14828

Data Field Explanation : Dataset contains 14,828 rows. The columns in the table are:

title — Title of the movie.

imdb_id — IMDB id of the movie.

plot_synopsis — Plot Synopsis of the movie.

tags — Tags assigned to the movie separated by “,”

synopsis_source — From where the plot synopsis was collected

split — Position of the movie in the standard data split, like Train, Test or Val

Example Data point :

imdb_id: tt0113862

title : Mr. Holland's Opus

plot_synopsis:

Glenn Holland, not a morning person by anyone's standards, is woken up by his wife Iris early one bright September morning in 1964. Glenn has taken a job as a music teacher at the newly re-named John

F. Kennedy High School.....

tags : inspiring, romantic, stupid, feel-good

split : train

synopsis_source: imdb

MAPPING THE REAL-WORLD PROBLEM TO A MACHINE LEARNING PROBLEM :

Type of Machine Learning Problem :

It is a multi-label classification problem

Multi-label Classification:

Multi-label classification assigns to each sample a set of target labels. This can be thought as predicting properties of a data-point that are not mutually exclusive, such as topics that are relevant for a document. A Movie might be about any Genre like romantic, action, thriller, horror at the same time or none of these.

Performance metric :

Micro-Averaged F1-Score (Mean F Score) : The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is:

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

In the multi-class and multi-label case, this is the weighted average of the F1 score of each class.

'Micro f1 score':

Calculate metrics globally by counting the total true positives, false negatives and false positives. This is a better metric when we have class imbalance.

'Macro f1 score':

Calculate metrics for each label, and find their unweighted mean. This does not take label imbalance into account.

Hamming loss :

The Hamming loss is the fraction of labels that are incorrectly predicted.

Data Analysis :

Let us see some meta data, i.e. data about our dataset.

Total Number of Rows : 14828

Total Number of Columns : 6

Columns : imdb_id, title, plot_synopsis, tags, split, synopsis_source, tag_count

tag_count : is a newly added column that specifies how many tags our movie consist of.

Snapshot of the Raw Data

Duplicate Movies : 0
 Movies with 0 tags : 0

TAGS ANALYSIS :

We can see that there are 5516 movies with 1 tag, similarly 2 movies with 3124 tags, all the way to we have 1 movie with 25 tags in it. This gives a lot of information of the distribution of tags among the movies.

71 unique tags we have in our data-set

As we can see there are 71 unique tags we have in our dataset, we will use

these tags to predict the test data.

Some of the popular tags

We can see that "murder", "violence", "romance", "flashback" and "cult" as some of the popular tags we have in the dataset.

Avg. number of tags per movie: 2.794578 ~ 3 tags per movie

Let's draw our Observations over the analysis :

1. There are total 20 tags which are used more than 500 times.
2. 9 tags are used more than 1000 times.

1	5516
2	3124
3	1959
4	1238
5	918
6	686
7	478
8	318
9	223
10	195
11	187
13	52
12	51
14	33
15	28
16	18
18	11
17	9
21	3
19	2
20	2
25	1

Name: tag_count, dtype: int64

3. Most frequent tag (i.e. murder) is used 5782 times.
4. Since some tags occur much more frequently than others, Micro-averaged F1-score is the appropriate metric for this problem.

DATA PRE-PROCESSING :

Data Cleaning :

	Tags	Counts
43	murder	5782
88	violence	4428
28	flashback	2937
57	romantic	2906
20	cult	2647

We clean the data, as in we have raw form of textual data, we remove the following :

1. punctuation
2. extra spaces
3. stop words
4. we change words like "won't" to "will not", "can't" to "can not" etc.
5. we change words like "I've" to "I have", "I'll" to "I will"

After Cleaning the data it looks like :



MACHINE LEARNING MODELS :

Converting tags for multi-label problems :

X	y1	y2	y3	y4
x1	0	1	1	0
x1	1	0	0	0
x1	0	1	0	0

We convert all the 71 tags into a binary bow, where for each movie we will put 1 against those tags which are present for the movie. After converting all the tags into a bow feature, we have tags in the form of (14828, 71) data set.

Split the data into test and train :

We split the data in the manner of (80:20) split. After splitting we have data points as :

- Number of data points in train data X : (11862, 1)
- Number of data points in train data Y : (11862, 71)
- Number of data points in test data X : (2966, 1)
- Number of data points in test data Y : (2966, 71)

FEATURIZING DATA : THE MOST IMPORTANT PART OF ANY MODEL

Let us talk about the most creative and most toughest part about any Machine Learning model, i.e. creating the features out of the raw data..

Lets us look into features which i came up with while solving the problem.

Lexical Features :

1. n-Gram : 1,2,3
2. Char n-Gram : 2,3
3. k-Skip-n-Gram

Will Explore the Lexical Features one-by- one :

1. n-Gram : 1,2,3

We will understand this with an example : Let's take this sentence into consideration "The quick brown fox jumps over the lazy dog"

Now for a feature of 1-gram it will consider each word to be a vector :The, quick, brown, fox, jumps, over, the, lazy, dog

Now for a feature of 2-gram it will consider combination of 2 Words to be a vector : The quick, quick brown, brown fox.... and so on.

Similarly for 3-gram it will take : The quick brown, quick brown fox,... and so on.

They are basically a set of co-occurring words within a given window.

2.Char n-gram : 2,3

Character n-Gram implies the same concept as n- gram the only difference is, it works on a character level.

For Example :

Consider the word "Machine Learning"

Char-2-Gram : "ma", "ac", "ch", "hi" ... and so on Char-3-Gram : "mac", "ach", "chi" ... and so on

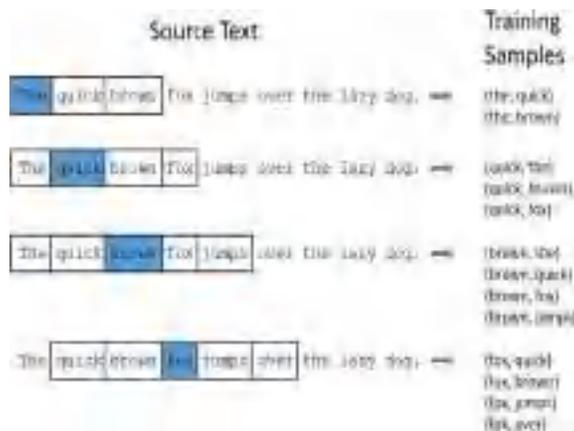
3.k-Skip-n-Gram :

In skip gram architecture of word2vec, the input is the center word and the predictions are the context words. Consider an array of words W, if W(i) is the input (center word), then W(i-2), W(i-1), W(i+1), and W(i+2) are the context words, if the sliding window size is 2.

Source Google

As we say, picture speaks more than words, well that's absolutely true. The above image just saved me from a lots of explanation. It is clear how k-Skip- n-Gram works.

BAG OF CONCEPTS : PARTS OF SPEECH



TAGGING

It is a process of converting a sentence to forms — list of words, list of tuples (where each tuple is having a form (word, tag)). The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on.

POS tagging Word Wise

Basically POS tagging is word wise, so we have to very careful before using them that how are we suppose to use them for our machine, in my case i counted the number of POS i have in my sentence and made a bow model and updated the count, this way

```

POS TAGS
=====
NN: NOUN
VB: VERB
JJ: ADJECTIVE
IN: PREPOSITION
DT: DETERMINER
RP: ADVERB
MD: MODAL
CC: COORDINATOR
CD: CARDINAL
PUNCT: PUNCTUATION
SYM: SYMBOL
UH: INTERJECTION
EX: EXCLAMATION
TO: TO
FW: FOREIGN_WORD
XX: UNKNOWN
=====
The quick brown fox jumps over the lazy dog.
=====
The: DT, quick: JJ, brown: JJ, fox: NN, jumps: VB, over: IN, the: DT, lazy: JJ, dog: NN
=====

```

there will be no dimension mismatch in our data.

Sentiments and Emotions :

I used "SentimentIntensityAnalyzer" to find out the sentiment values of a particular sentence, this function returns us with 4 dimension feature along with their value for a particular sentence, the values are basically :

neg, neu, pos, compound : these are scores based on the analysis of the text given.

WORD EMBEDDING :

I have used 4 word Embedding, namely :

- 1.BOW
- 2.TF-IDF
- 3.Avg W2V
- 4.TFIDF weighted AvgW2V

Lets go through each one of them and understand their significance :

BOW :

The bag-of-words model is a way of representing text data when modeling text with machine learning algorithms. Machine learning algorithms cannot work with raw text directly; the text must be converted into numbers. Specifically, vectors of numbers. A bag-of-words model, or BoW for short, is a way of extracting features from text for use in modeling. A bag-of-words is a representation of text that describes the occurrence of words within a document.

For example : Let's consider 2 sentences :

1. "This is a good place to stay"
2. "This is a good place to eat and drink"

Now, we need to have 2 things to make up our BoW model,

- i. A vocabulary of known words.
- ii. A measure of the presence of known words.

So, lets us consider a something called a Binary BoW, which specifies whether that sentence contains the word or not.

Here our BoW vector will be like [This, is, a, good, place, to, stay, eat, and, drink]. Hence the binary BoW for 1st and 2nd sentence will be as follows :[1,1,1,1,1,1,0,0,0] and [1,1,1,1,1,1,0,1,1] respectively.

TF-IDF :

Tf-idf stands for term frequency-inverse document frequency, and the tf-idf weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus.

TF:

Term Frequency, which measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear much more times in long documents than shorter ones. Thus, the term frequency is often divided by the document length (aka. the total number of terms in the document) as a way of normalization:

$$TF(t) = (\text{Number of times term } t \text{ appears in a document}) / (\text{Total number of terms in the document}).$$

IDF:

Inverse Document Frequency, which measures how important a term is. While computing TF, all terms are considered equally important. However it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by computing the following:

$$IDF(t) = \log_e(\text{Total number of documents} / \text{Number of documents with term } t \text{ in it}).$$

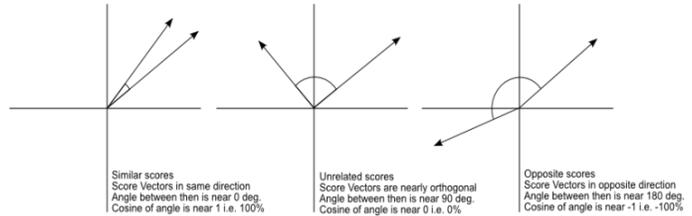
Avg W2V :

Word2vec basically place the word in the feature space in such a way that their location is determined by their meaning i.e. words having similar meaning are clustered together and the distance between two words also have the same meaning.

Let's first understand what is cosine similarity because word2vec uses cosine similarity for finding out the most similar word. Cosine similarity is not only telling the similarity between two vectors but it also tests the orthogonality of a vector. Cosine similarity is represented by formula:



$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

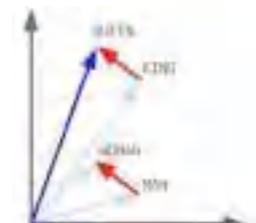


If angle is close to zero then we can say that vectors are very similar to each other and if theta is 90 then we can say vectors are orthogonal to each other (orthogonal vector not related to each other) and if theta is 180 we can say that both the vectors are opposite to each other.

We need to give large text corpus where for every word it creates a vector. It tries to learn the relationship between vector automatically from raw text. Larger the dimension it has, larger it is rich in information the vector is.

PROPERTIES:

1. If word w1 and w2 are similar than vector v1 and v2 will be closer.
2. Automatic learn the relationship between words/vector.



we are looking into Male-Female graph we are observing that distance between man and woman is same as distance between king (male) and queen (woman) Not only different gender but if we look into same-gender we observe that distance between queen and woman and distance between king and man are same (king and man, queen and woman represent same-gender comparison hence they must be equal distance)

how to convert each document to vector?

suppose you have w1, w2, ...wn word in one document(row). in order to convert into vector.



each word has one vector, we will convert average word2vec than divide by the number of word in a document.

TFIDF weighted WORD2Vec :

In this method first, we will calculate tfidf value of each word. than follow the same approach as above section by multiplying tfidf value with the corresponding word and then divided the sum by sum tfidf value.

Hence, these 4 techniques were used to add as a feature to my model's training.

Numerical Feature for Text :

1.Len of each Plot Synopsis : Basically, i took the length of each plot synopsis and made that as an additional feature.

2.Len of Unique words in Plot Synopsis : Counted each unique word for every plot synopsis and included that as a feature.

Combining All the Hand-Made Features :

Finally after combining all the features, we have all together of 12 features to be used for Training of my model.

METHODOLOGY :

1.Multi-Label Classification : Machine Learning

1.1: One Vs Rest : Logistic Regression :

One-vs-the-rest (OvR) multi-class/multi-label strategy. Also known as one-vs-all, this strategy consists in fitting one classifier per class. For each classifier, the class is fitted against all the other classes. In addition to its computational efficiency (only n_classes classifiers are needed), one advantage of this approach is its interpretability. Since each class is represented by one and one classifier only, it is possible to gain knowledge about the class by inspecting its corresponding classifier. This is the most commonly used strategy for multi-class classification and is a fair default choice.

This strategy can also be used for multi-label learning, where a classifier is used to predict multiple labels for instance, by fitting on a 2-d matrix in which cell [i, j] is 1 if sample i has label j and 0 otherwise. In the

multi-label learning literature, OvR is also known as the binary relevance method. After using OvR : LR and hypertuning it, we got the below results :

```

Test F1 Score with prob > 0.25 for each tags : 0.6533833574983932
Train Accuracy : 0.8777423059017767
Train Training Loss : 0.028553719891477472
Train Micro-average quality numbers
Precision: 0.9365, Recall: 0.3390, F1-measure: 0.4956
Train Macro-average quality numbers
Precision: 0.7307, Recall: 0.1275, F1-measure: 0.2021
    
```

```

Test Accuracy : 0.6533833574983932
Test Training Loss : 0.028553719891477472
Test Micro-average quality numbers
Precision: 0.9365, Recall: 0.3390, F1-measure: 0.4956
Test Macro-average quality numbers
Precision: 0.7307, Recall: 0.1275, F1-measure: 0.2021
Time taken to run this cell : 0:02:44.161760
    
```

Here, all the tags are given a probability for being in the plot synopsis, hence i considered only those which had a probability of more than 0.25

1.2: One Vs Rest : MultinomialNB

I used, Multinomial Naive Bayes, as a classifier along with OVR, lets visualize the results :

```

Train Accuracy : 0.8777423059017767
Train Training Loss : 0.028553719891477472
Train Micro-average quality numbers
Precision: 0.9365, Recall: 0.3390, F1-measure: 0.4956
Train Macro-average quality numbers
Precision: 0.7307, Recall: 0.1275, F1-measure: 0.2021
    
```

Hyper parameter tuning.

```
Test Accuracy : 0.83230682400539447  
Test Hamming Loss : 0.82890595541678934  
Test Micro-average quality numbers  
Precision: 0.2500, Recall: 0.9828, F1-measure: 0.3936  
Test Macro-average quality numbers  
Precision: 0.9046, Recall: 0.6667, F1-measure: 0.7894
```

Test F1 Score with prob > 0.495000 for each tags : 0.076542

Very less, cannot be considered.

MODEL SUMMARY :

Machine Learning :

Model	Precision	Recall	Macro F1-Score
Logistic Regression	0.6804	0.3381	0.4855
SVM	0.25	0.9828	0.4936

CONCLUSION :

1. In today's era, we are more used to see scores above 90%. But given a very limited data size sample of 14K datapoints, we have actually managed to get a decent micro averaged F1 score.
2. The maximum micro averaged F1 score is 0.4015 & the maximum value of recall is 0.4827.
3. Using featurization like bow, avg word2VEC, tfidf word2Vec and combination of TF-IDF and Word2Vec features, our models behaved surprisingly better than the previous implementation.
4. Char N-gram features proved to be significantly powerful than word N-gram features. Skip Grams were also useful.

FURTHER IMPROVEMENTS :

We can use more features into Deep Learning models, some more fine tuned architecture, more hyper tuning can be done.

We can achieve a good f1 score when we add more features, and consider a deeper network.

REFERENCES :

- [1] **Research Paper:** : <https://www.aclweb.org/anthology/L18-1274>
Code References : <https://www.appliedaicourse.com/>
- [2] **Ideas :** https://en.wikipedia.org/wiki/Multi-label_classification
- [3] **Binary Relevance:**

http://scikit.ml/api/skmultilearn.problem_transform.br.html
[4] **Classifiers:** <https://scikit-learn.org/stable/>

[5]Strategies:

<https://www.analyticsvidhya.com/blog/2017/08/introduction-to-multi-label-classification/>

[6] **Useful Links Source :**
<https://www.kaggle.com/cryptexcode/mpst-movie-plot-synopses-with-tags>

[7] Additional Info :

<http://ritual.uh.edu/mpst-2018/>

[8] <https://www.kaggle.com/wiki/MeanFScore>

[9] http://scikitlearn.org/stable/modules/generated/sklearn.metrics.F1_score.html

[10] <https://www.kaggle.com/wiki/HammingLoss>

[11] <http://scikit-learn.org/stable/modules/multiclass.html>

Survey On Multi-Label Classification : Movie Plot Synopses with Tags : Tags Prediction

Prashant Hule, Darshana Wajekar

Abstract— In today's world social tagging of movies reveals a wide range of heterogeneous information about movies, such as plot structure, genre soundtracks, emotional experiences, metadata & visual. Such information can be valuable in building automatic systems for creating tags of movies. Automatic tagging systems can help recommendation engines to improve the retrieval of similar movies as well as help viewers to know what to expect from a movie in advance. In this paper, we set out to the task of collecting a corpus of movie plot synopses and tags. We describe a methodology that enabled us to build a fine grained set of around 70 tags exposing heterogeneous characteristics of movie plots & the multi-label associations of these tags with around 14K movie plot synopses. We investigate how these tags correlate with movies & the flow of emotions throughout different types of movies. Finally, we use this corpus to explore the feasibility of inferring tags from plot synopses. We expect the corpus will be useful in other tasks where analysis of narratives is relevant.

Index Terms— Tag generation for movies, Movie plot analysis, Multi-label data set, Supervised Learning



PROBLEM STATEMENT

Suggest the tags based on the content that was there in the Movie in its Title and Synopsis.

REAL WORLD / BUSINESS OBJECTIVES AND CONSTRAINTS

1. Predict as many tags as possible with high precision and recall.
2. Incorrect tags could impact customer experience.
3. No strict latency constraints.

MACHINE LEARNING PROBLEM

Data Overview : All of the data is in 1 file: mpst_full_data.csv

Data is divided into 3 Sections Train, Test and Val data within the file

Train Consist of 64% of the data.

Test Consist of 20% of the data.

Val Consist of 16% of the data.

Number of rows in mpst_full_data.csv = 14828

Data Field Explanation : Dataset contains 14,828 rows. The columns in the table are:

title — Title of the movie.

imdb_id — IMDB id of the movie.

plot_synopsis — Plot Synopsis of the movie.

tags — Tags assigned to the movie separated by “,”

synopsis_source — From where the plot synopsis was collected

split — Position of the movie in the standard data split, like Train, Test or Val

Example Data point :

imdb_id: tt0113862

title : Mr. Holland's Opus

plot_synopsis:

Glenn Holland, not a morning person by anyone's standards, is woken up by his wife Iris early one bright September morning in 1964. Glenn has taken a job as a music teacher at the newly re-named John

F. Kennedy High School.....

tags : inspiring, romantic, stupid, feel-good

split : train

synopsis_source: imdb

MAPPING THE REAL-WORLD PROBLEM TO A MACHINE LEARNING PROBLEM :

Type of Machine Learning Problem :

It is a multi-label classification problem

Multi-label Classification:

Multi-label classification assigns to each sample a set of target labels. This can be thought as predicting properties of a data-point that are not mutually exclusive, such as topics that are relevant for a document. A Movie might be about any Genre like romantic, action, thriller, horror at the same time or none of these.

Performance metric :

Micro-Averaged F1-Score (Mean F Score) : The F1 score can be

interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is:

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

In the multi-class and multi-label case, this is the weighted average of the F1 score of each class.

'Micro f1 score':

Calculate metrics globally by counting the total true positives, false negatives and false positives. This is a better metric when we have class imbalance.

'Macro f1 score':

Calculate metrics for each label, and find their unweighted mean. This does not take label imbalance into account.

Hamming loss :

The Hamming loss is the fraction of labels that are incorrectly predicted.

MACHINE LEARNING MODELS :

Converting tags for multi-label problems :

X	y1	y2	y3	y4
x1	0	1	1	0
x1	1	0	0	0
x1	0	1	0	0

We convert all the 71 tags into a binary bow, where for each movie we will put 1 against those tags which are present for the movie. After converting all the tags into a bow feature, we have tags in the form of (14828, 71) data set.

Split the data into test and train :

We split the data in the manner of (80:20) split. After splitting we have data points as :

- Number of data points in train data X : (11862, 1)
- Number of data points in train data Y : (11862, 71)
- Number of data points in test data X : (2966, 1)
- Number of data points in test data Y : (2966, 71)

FEATURIZING DATA : THE MOST IMPORTANT PART OF ANY MODEL

Let us talk about the most creative and most toughest part about any Machine Learning model, i.e. creating the features out of the raw data..

Lets us look into features which i came up with while solving the problem.

Lexical Features :

1. n-Gram : 1,2,3
2. Char n-Gram : 2,3
3. k-Skip-n-Gram

Will Explore the Lexical Features one-by- one :

1. n-Gram : 1,2,3

We will understand this with an example : Let's take this sentence into consideration "The quick brown fox jumps over the lazy dog"

Now for a feature of 1-gram it will consider each word to be a vector :The, quick, brown, fox, jumps, over, the, lazy, dog

Now for a feature of 2-gram it will consider combination of 2 Words to be a vector : The quick, quick brown, brown fox.... and so on.

Similarly for 3-gram it will take : The quick brown, quick brown fox,... and so on.

They are basically a set of co-occurring words within a given window.

2.Char n-gram : 2,3

Character n-Gram implies the same concept as n- gram the only difference is, it works on a character level.

For Example :

Consider the word "Machine Learning"

Char-2-Gram : "ma", "ac", "ch", "hi" ... and so on Char-3-Gram : "mac", "ach", "chi" ... and so on

3.k-Skip-n-Gram :

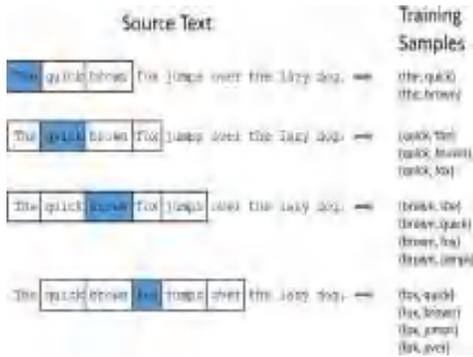
In skip gram architecture of word2vec, the input is the center word and the predictions are the context words. Consider an array of words W, if W(i) is the input (center word), then W(i-2), W(i-1), W(i+1), and W(i+2) are the context words, if the sliding window size is 2.

Source Google

As we say, picture speaks more than words, well that's absolutely true. The above image just saved me from a lots of explanation. It is clear how k-Skip- n-Gram works.

BAG OF CONCEPTS : PARTS OF SPEECH TAGGING

It is a process of converting a sentence to forms — list of words, list of tuples (where each tuple is having a form (word, tag)). The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on.



POS tagging Word Wise

Basically POS tagging is word wise, so we have to be very careful before using them that how are we supposed to use them for our machine, in my case I counted the number of POS I have in my sentence and made a bow model and updated the count, this way there will be no dimension mismatch in our data.

Sentiments and Emotions :

I used "SentimentIntensityAnalyzer" to find out the sentiment values of a particular sentence, this function returns us with 4 dimension feature along with their value for a particular sentence, the values are basically :

```

neg: 0.0
neu: 0.0
pos: 0.0
compound: 0.0
    
```

neg, neu, pos, compound : these are scores based on the analysis of the text given.

WORD EMBEDDING :

I have used 4 word Embedding, namely :

1. BOW
2. TF-IDF
3. Avg W2V
4. TFIDF weighted AvgW2V

Lets go through each one of them and understand their significance :

BOW :

The bag-of-words model is a way of representing text data when modeling text with machine learning algorithms. Machine learning algorithms cannot work with raw text directly; the text must be converted into numbers. Specifically, vectors of numbers. A bag-of-words model, or BoW for short, is a way of extracting features from text for use in modeling. A bag-of-words is a representation of text that describes the occurrence of words within a document.

For example : Let's consider 2 sentences :

"This is a good place to stay"

"This is a good place to eat and drink"

Now, we need to have 2 things to make up our BoW model,

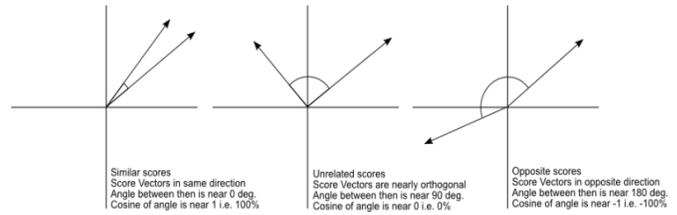
- i. A vocabulary of known words.
- ii. A measure of the presence of known words.

So, lets us consider a something called a Binary BoW, which specifies whether that sentence contains the word or not.

Here our BoW vector will be like [This, is, a, good, place, to, stay, eat, and, drink]. Hence the binary BoW for 1st and 2nd sentence will be as follows : [1,1,1,1,1,1,0,0,0] and [1,1,1,1,1,1,1,0,1,1] respectively.

TF-IDF :

Tf-idf stands for term frequency-inverse document frequency, and the tf-idf weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus.



TF:

Term Frequency, which measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear much more times in long documents than shorter ones. Thus, the term frequency is often divided by the document length (aka. the total number of terms in the document) as a way of normalization:

$$TF(t) = (\text{Number of times term } t \text{ appears in a document}) / (\text{Total number of terms in the document}).$$

IDF:

Inverse Document Frequency, which measures how important a term is. While computing TF, all terms are considered equally important. However it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by computing the following:

$$IDF(t) = \log_e(\text{Total number of documents} / \text{Number of documents with term } t \text{ in it}).$$

Avg W2V :

Word2vec basically place the word in the feature space is such a way that their location is determined by their meaning i.e. words having similar meaning are clustered together and the distance between two words also have the same meaning.



lets first understand what is cosine similarity because word2vec uses cosine similarity for finding out the most similar word. Cosine similarity is not only telling the similarity between two vectors but it also test for orthogonality of vector. Cosine similarity is represented by formula:

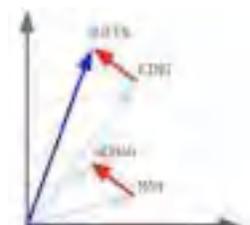
$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

if angle are close to zero than we can say that vectors are very similar to each other and if theta is 90 than we can say vectors are orthogonal to each other (orthogonal vector not related to each other) and if theta is 180 we can say that both the vector are opposite to each other.

We need to give large text corpus where for every word it creates a vector. it tries to learn the relationship between vector automatically from raw text. larger the dimension it has, larger it is rich in information the vector is.

PROPERTIES:

1. If word w1 and W2 are similar than vector v1 and V2 Will be closer.
2. Automatic learn the relationship between words/vector.



we are looking into Male-Female graph we are observing that distance between man and woman is same as distance between king (male) and queen (woman) Not only different gender but if we look into same- gender we observe that distance between queen and woman and distance between king and man are same(king and man, queen and woman represent same-gender comparison hence they must be equal distance)

how to convert each document to vector?

suppose you have w1, w2, ...wn word in one document(row). in order to convert into vector.



each word has one vector, we will convert average word2vec than divide by the number of word in a document.

TFIDF weighted WORD2Vec :

In this method first, we will calculate tfidf value of each word. then follow the same approach as above section by multiplying tfidf value with the corresponding word and then divided the sum by sum tfidf value.

Hence, these 4 techniques were used to add as a feature to my model's training.

Numerical Feature for Text :

1.Len of each Plot Synopsis : Basically, i took the length of each plot synopsis and made that as an additional feature.

2.Len of Unique words in Plot Synopsis : Counted each unique word for every plot synopsis and included that as a feature.

Combining All the Hand-Made Features :

Finally after combining all the features, we have all together of 12 features to be used for Training of my model.

References:

[1] **Research Paper :** <https://www.aclweb.org/anthology/L18-1274>

[2] **Code: References :** <https://www.appliedaicourse.com/Ideas> : https://en.wikipedia.org/wiki/Multi-label_classification Binary

[3] **Relevance :** http://scikit.ml/api/skmultilearn.problem_transform.br.html

[4] **Classifiers:** <https://scikit-learn.org/stable/Strategies>: <https://www.analyticsvidhya.com/blog/2017/08/introduction-to-multi-label-classification/>

[5] **Others :** <https://towardsdatascience.com>
<https://www.analyticsvidhya.com> <https://www.quora.com>
<https://deepsense.ai> <https://datascience.stackexchange.com/>

[6] Useful Links :

[7] **Source :** <https://www.kaggle.com/criptexcode/mps-t-movie-plot-synopses-with-tags>

Email Spam Detection Using Machine Learning Algorithms

Pratik Gaikwad, Prof. Swapnil Patil

Abstract— Receiving spam email messages are quite obvious. Most of the time such emails are commercial. But repeatedly, such emails may contain some phishing links that have malware. This arises the need for proposing careful mechanism to detect or identify such spam emails so as that time and memory space of the system are often concentrated to an exquisite extent. during this paper, we presented the identical mechanism which could filter spam and non-spam emails. an proposed algorithm generates dictionary and fea- tures and trains them through machine learning for effective results.

Index Terms— spam; machine learning; naive-bayes; data set

1 INTRODUCTION

Receiving spam email decrease the efficiency of the user and it's quite annoying to the user [5]. Usually, such emails are forwarded for profit-making purposes or even through fraud agencies which can snatch money through phishing. Spam detection identification has the foremost objective to inform users about fake-emails and/or relevant e-mails. Spam could be a web terminology that refers to either unsolicited commercial email (UCE) or unsolicited bulk email (UBE). Internet surfers usually call such quite emails as junk email which are forwarded for any commercial or business promotional purposes. Here unsolicited simply means the recipient of such junk emails has not granted permission for the messages to the sender [1].

Spam and Ham: in step with Wikipedia “the use of electronic messaging systems to send unsolicited bulk messages, especially mass advertisement, malicious links etc.” are called as spam. “Unsolicited implies that those things which you didn’t asked for messages from the sources. So, if you do not comprehend the sender the mail may be spam. People generally don’t realize they only signed certain those mailers when they download any free services, software or while updating the software. “Ham” this term was given by Spam Bayes around 2001 and it's defined as “Emails that aren't generally desired and isn't considered spam”.

Machine learning approaches are more efficient, a set of training data is employed, these samples are the set of email which are pre classified. Machine learning approach-

es have plenty of algorithms that may be used for email filtering. These algorithms include “Naïve Bayes, support vector machines, Neural Networks, etc.”

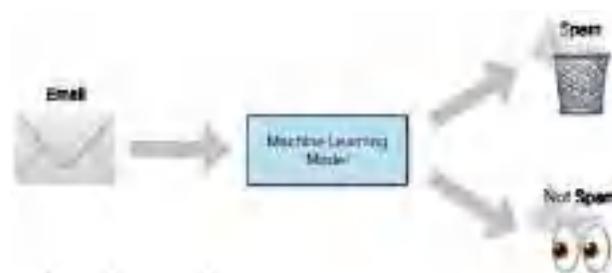


Fig.1. Classification into Spam and non-spam

2 Need of spam detection

Now our proposed algorithm will run to check whether the entered email address is s pam or not. By applying this mechanism, a user can work efficiently as comparatively fewer spam emails will be received. This mechanism also saves the time and memory of the system. So there is a need [4] for some mechanism that can reduce or even provide some sort of panacea to from these spam emails. Keep this situation in mind, in this paper; we are presenting a machine learning-based spam detection mechanism that uses a dataset of approximately 6000 valid and invalid collection of emails. Our proposed model will first make a dictionary that remove helping verbs form the contents of the email.

Consider a case in which someone over the internet is sending bulk emails regarding the promotion of their products for purchasing it, or someone sending a link to click or activate to win some lucrative prizes; such emails are generally considered as negative marketing strategies or fraud activities. As a

receiver, you are helpless in this scenario [6]. These unwanted emails may consume a lot of memory of your system & also waste your precise time. It is also observed that one can be distributed by receiving such bogus emails again and again.

3 Proposed model

In the next phase, features are generating to train the dataset. After that our algorithm will be executed to check the possibility of an email to be spam or not. Finally, the machine learning model will be tested on a real-world emailing environment. In this model, we are firstly creating a dictionary that includes a library named "stopwords" [7] [8] to remove all possible helping verbs from the content mentioned in the email.

A. Generate Dictionary

To segregate junk or non-junk emails, first of all, we have got collected some specific words and insert them into the dictionary which may be utilized within the proposed model. within the literature, numerous word extracting methods exists. So there is a desire to choose the foremost suitable model precisely. Firstly we have got to rearrange the information for generating a dictionary for our algorithm. This dictionary could also be further utilized to extract the required features which determine our algorithm. These are parameters through which a user can segregate spam emails and general or simply non-spam emails.

B. Generating features

Once the features are generated, it is now time to work on the prediction model for the algorithm. Through this prediction model, different words will be inserted as inputs. The model will calculate and produce the result as total occurrences of each and every word throughout the document inserted. The specific features can further be extracted. These extracted features must be thoroughly tested. When all the test cases are successful then this data are often passed to the Naive Bayes algorithm within the sort of inputs. After generating a dictionary, now it's time to come up with features. For this, we are going to utilize the information set which has been trained by applying machine learning processes.

C. Generating Machine Learning model

A data structure has been prepared that may run through the Naive Bayes algorithm. As an outcome, we are able to achieve a prediction model. In literature, the Naive Bayes algorithm is additionally termed as a "probabilistic classifier". the explanation behind it's that it's based an excellent extent on performing the calculation the probability of an email which can be further classified into numerous domains.

D. Testing of Machine Learning Model

This is a very crucial step in our model. In this phase, we test our proposed model as providing data set in the form of input

4. Working of model

A Sending of data

This application provides one of the main features of the system, which is content spam filter. The feature of attaching the file is disabled in the application as the file may contain images. This is feature is intentionally disabled because it is quite possible that spam got failed to detect hidden objects of the image(s). For every client, there exists an option of composing an email through which a text email can be composed. This data can be further exchanged between different clients whenever required

B Receiving the data

These features may include inbox, writing, sending, replying, forwarding, filtering and even deleting the emails. The observation that should be made here is that if the content in the email represents spam, subsequently the message is blocked by the tool before it is delivered to the target client machine. This way only the non-spam emails are exchanged between clients. When a textual email is sent from one client then it reaches the inbox of other clients. It is quite obvious that any email service system user has the basic features of the email provided by the application being used.

C Spam report module

The server system provides messages which can be categorized into spam and non-spam categories. This data can be further analysis for the determination of compromised and non-compromised machines based on the well-defined degree of fault-tolerance. Working with our model is crucially dependent on the spam report module. In this section, the server system is activated and the spam report module is deployed in the active supervision of the server.

These spams are detected when the filtering algorithm is applied to the system. To maintain the privacy issues of various clients; it is also mandatory to encrypt the messages before entering it to the network. Discrete results of spam and non-spam machines are also recorded. It also keeps track of the client's system details as the client's name, timestamping of emails, spam details, etc for analyzation purposes.

D Spam filtering algorithm

This spam filtering approach identifies and controls spam emails. With the help of this algorithm, we can effectively differentiate between legitimate and illegitimate spam emails. It has been also noticed that the chances of getting spam messages are higher if the client receives such messages in the past as well. It means spam messages are growing successively by the time. The content of the messages must be verified by any source of the algorithm. So that internet surfer can work over the internet without any disturbance. In today's digital era; the main challenge is to filter and detect the spam emails which are being forwarded uninterruptedly over the internet. The Bayesian spam filter can be applied to control such junk messages to a great extent. Most of the spammers are adapting to new technologies and are becoming more effective

There can be some words which are spam for one organization but not in other organization. For these types of words, the algorithm verifies the frequency of the particular word the

number of times it has occurred with respect to some organization and recognizes the spam based on this probability ratio. The most important part of this algorithm is, it detects the spam based on the patterns. It identifies the section of words against its spam city rather than individual words for more effective detection of spam content in an email.

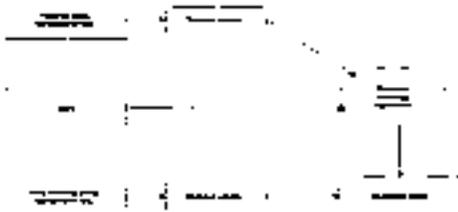


Fig. 2 Proposed Process architecture

Working of our proposed process model includes following basic steps:

- 1) Training texts and document: First we will obtain the data and then we will train the dataset according to the algorithm.
- 2) Feature Vector: Feature vector x composed of n words coming from spam emails. This vector comprises of n words coming from spam emails. The Naïve Bayes classifier makes the probability that the word is not dependent on each other. The main assumption of the classifier is each word is independent of each other.
- 3) Machine Learning algorithm: The Naive Bayes algorithm relies on Bayes Rule. This algorithm will look of the object will be n to each object by identifying all features separately. The algorithm shows us how we can calculate the posterior probability for one feature. The posterior probability will be calculated for a feature and these will be multiplied to get the final probability. This will be calculated on other classes.
- 4) Expected Label: This will give the outcome of the algorithm. We will let u know that the mail that is detected is spam or not.

E Naive Baye' s classifier

According to computer science, Naive Bayes models are also known as simple Bayes and independence Bayes models. The algorithm uses the references of Naïve Bayes; decision rule, but we know that naive Bayes is not a prominent method. It is a prominent method of technique for developing classifiers; models that assign class labels to problem instances represented as vectors of feature values, where the class labels are drawn from some finite set).

The technique is highly scalable as it can be easily updated on the arrival of new data and adapts well to the future spam techniques. It is a less expensive approach as maximum likelihood training is done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation. It is the best approach to identify emails whether they are spam or not. This algorithm gives very low false positive spam detection rates and can change according to the

need of the user. It is an intelligent approach that considers the complete message as the examining criterion rather than the single word thus saving a lot of time.

Naive Bayes algorithm is a capability of refining filtering that uses features to recognize spam mails. It is an analytical technique used to classify text. The algorithm analyzes data and then apply naive Bayes to check whether an email is a spam or not. [2] [3].

5 Advantages Of Prosoed Algorithm

- The algorithm has the feature of implementing according to the requirements of the user. Thus can be implemented in any organization.
- It is an intelligent approach as it considers the complete message as the examining criterion rather than a single word.
- It is highly scalable as it provides scalability of the algorithm that the user can utilize for better efficiency.
- The algorithm is easy to update on the arrival of new data. It is best suited for text classification.

6 Spam Identification Methods

Numerous methods are presented in the literature to determine incoming emails that might be spam. The most prominent methods are Mail Header Analysis, Keyword Checking, URL Checking, etc.

1) Mail Header Checking

Mail header checking consists of well-defined protocols that, if a mail header matches, triggers the mail server to revert back the messages that have a blank "From" field. It also prepares a list that has numerous email addresses within the "To" from the identical source, that has too many digits in email addresses.

2)Keyword Checking

This is also a widely used method for spam detection. Unlike mail header checking algorithm it scans both the body and therefore the subject a part of the e-mail. Using rules like combinations of keywords may be a good solution to boost filtering efficiency combinations of words is specified and the list that has to appear within the spam email will be updated further the messages with these words are going to be blocked [4].

3) URL Checking

The URL checking method is also well-known spam detection method. In URL checking emails are classified on the basis of their URL. A list of addresses is recorded from which we never want to receive emails, any email coming from these addresses will be blocked.

7 Conclusion

As of now recent days spam emails are increasing day by day and it's creating problem to the user so by spam detector, we'll identify which mail is spam or not, by this efficiency of users are visiting be increased. We are using the Naïve Bayes classifier which is ready to provide u the probabilistic index of that and can identify whether the mail is spam or not.

8 References

[1] Sara Radicati, "Email Statistics Report, 2014-2018", The Radicati Group, Inc., Palo Alto, CA, USA, 2014.

[2] Ramachandran, D. Dagon, and N. Feamster, "Can DNS-based blacklists keep up with bots?" in CEAS 2006. The Second Conference on Email and Anti-Spam, 2006

[3] G. Cormack, "Email spam filtering: A systematic review," Foundations and Trends in Information Retrieval, vol. 1, no. 4, pp. 335–455, 2008. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[4] <http://www.securelist.com/en/threats/spam?chapter=97> last accessed on 23 February 2020.

[5] A. Dalmia, M. Gupta, et al. Towards interpretation of node embeddings. In Companion of the The Web Conference 2018 on The Web Conference 2018, pages 945–952. International World Wide Web Conferences Steering Committee, 2018.

[6] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada. Survey of review spam detection using machine learning techniques. J. Big Data, 2(1):1, 2015.

[7] <https://pypi.org/project/stop-words/> last accessed on 03 March 2020.

[8] <https://www.geeksforgeeks.org/removing-stop-words-nltk-python/> last accessed on 03 March 2020.

Email Spam detection using NLP

Pratik Gaikwad, Prof. Swapnil Patil

Abstract —A upsurge within the number of unwanted emails called spam has created an intense need for the event of more dependable and robust anti-spam filters. Machine learning methods of recent have gotten accustomed successfully detect and filter spam emails. We present a scientific review of form of the favored machine learning based email spam filtering approaches. Our review covers survey of the important concepts, attempts, efficiency, and then the research trend in spam filtering. The preliminary discussion within the study background examines the applications of machine learning techniques to the e-mail spam filtering process of the leading internet service providers (ISPs) like Gmail, Yahoo and Outlook emails spam filters. Discussion on general email spam filtering process, and also the numerous efforts by different researchers in combating spam through the utilization machine learning techniques was done. an review compares the strengths and disadvantages of existing machine learning approaches then the open research problems in spam filtering. We recommended deep learning and deep adversarial learning because the long term techniques which may effectively handle the menace of spam emails.

Index Terms— Email; Spam; Non-Spam; Spam detector; naive-bayes; Logistic Regression

1 INTRODUCTION

A lot of corporates often get a lot of mails in flux, and it is imperative to find whether the sent mail is a spam or not. Emails form an important way of communicating in life. It is impossible to think people can communicate without electronic mails. Despite the reaping benefits of electronic mails, it is important to note that there are often incoming of unsolicited and fraudulent emails along with the ones which are important. Such mails can cause harm to the companies or corporates on both financial level and psychological level. To solve the issue of the harm such mails can bring, it is important to nullify the effect such mails can bring, it is important to bring in a system which detects spam. Email spam detection facilitates the segregation of spam mail and non-spam mail. Thus, it can be inferred that spam detection is one of the first step which is of utmost importance in email filtering process which ensures that spam mail doesn't enter the user's inbox, particularly in this era of huge spam mails due to bulk mailing tools that has pushed up the limit of spam mails [1].

There have been various techniques employed in email spam detection. There has been Artificial Neural Networks employed which only succeeded in giving 86% accuracy which is still considered far from perfect. A low accuracy model means that the corporates can face huge trouble in such scenario as important mails classified as spam mail will be ludicrous to the existence of the whole system. The proposed system uses Natural Language Processing Techniques like tokenization, vectorization and POS tagging for processing the first. And on top of that, we use Machine Learning Algorithms used for Classification – for example, Logistic Regression and Naïve Bayes. The accuracy achieved through logistic regression in our research is 98.85% and Naïve Bayes is 99.180. We have used the dataset created by Spam Assassin which is a public corpus available for ham and spam emails. The spam directory contains 501 files and ham directory contains 2551 mails [2].

Description of Problem

Currently, email is an enlightening tool which is crucial and very necessary for facilitating quick and cheap communication. It is an extensive medium and method, and plays an essential part in our lives. However, spam or unsolicited email is an nuisance in this form of communication. They can be in the form of advertisements or similar explicit content which may contain malicious code embedded in them. A study has found out that 70 percent of the total business emails are spam. It has also been found that speedy growth of spam email has ramifications like overflowing of users' mailboxes, consumption of bandwidth and storage space, compromising of important emails and problems to the user regarding consuming their time to sort through all emails. Presently, there is growing interest in the field of spam classification because of the complexity that has been introduced by the spammers making it difficult to distinguish between spam mails and ham mails. The complexity can be due to attacks like Tokenisation and Obfuscation that is hiding certain features by adding HTML or some other codes such as 'free' coded as FR3E or frexe, which are used to alter the information on particular features. For spam classification, various spam filtering methods are used. The function of a spam filter is to spot spam email and prevents it from visiting the mailbox. With the help of filters, the adverse impact of spam email is mitigated and operates sort of a predictable and reliable tool to eliminate unwanted emails [3]. However, there exists a little risk of misclassification or removal of legitimate emails. In the research, various spam filtering methods are tested but none were found to be perfect. However, filters are beneficial for an email recipient who needs to undergo the burden of detecting spam. the prices which are incurred in reading every email for identifying spam involves over just time consumption. it's now becoming difficult for a user to tell apart between emails merely by read-

ing the subject or the e-mail content, thereby increasing the requirement of spam filter. Rarely but sometimes, filters also make mistakes but are utilized in conjunction with users to reduce these errors.

Status of research/knowledge in the field and Literature review

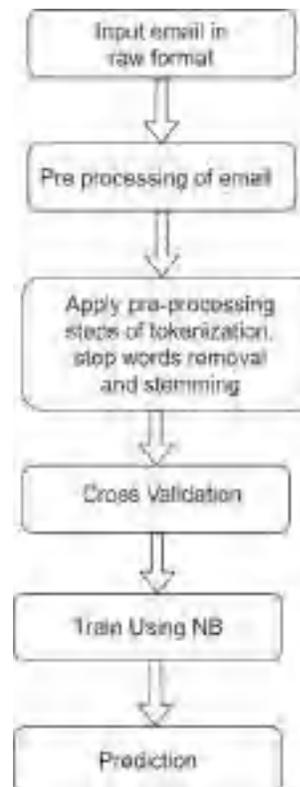
The adaptive nature of unsolicited email by the use of huge mailing tools prompts the need for spam detection. Implementation of different spam detection methods based on machine learning techniques was proposed to solve the problem of numerous email spam ravaging the system. Previous algorithm used in email spam detection compares each email message with spam and non-spam data before generating detectors while our proposed system inspired by the artificial immune system model with the adaptive nature of negative selection algorithm uses special features to generate detectors to cover the spam space. To cope with the trend of email spam, a novel model that improves the random generation of a detector in negative selection algorithm (NSA) with the use of stochastic distribution to model the data point using particle swarm optimization (PSO) was implemented.

Local outlier factor is introduced as the fitness function to determine the local best (Pbest) of the candidate detector that gives the optimum solution. Distance measure is employed to enhance the distinctiveness between the non-spam and spam candidate detector. The detector generation process was terminated when the expected spam coverage is reached. The theoretical analysis and the experimental result show that the detection rate of NSA-PSO is higher than the standard negative selection algorithm. Accuracy for 2000 generated detectors with threshold value of 0.4 was compared. Negative selection algorithm is 68.86% and the proposed hybrid negative selection algorithm with particle swarm optimization is 91.22%.

The current generation spams are growing in volume and also got varieties of it. Therefore the spam filters have to be moved away from central servers to the local servers, to be efficient. From monolithic repositories toward dynamic knowledge bases. But there are chances where a legitimate email may be classified as spam. Because the art of spam filtering is tending towards collaborative filters from the state of content-based filtering and users have to feed the information about the false positives and false negatives. This feedback helps in tracking the concept drift in spam and how to retain it in the case of false positives. In this way these filters achieve impressive accurate rates. But in order to rectify the problem of erroneous classification of a needed email as spam.

We are making three assumptions to locate the presence of false positives. So we are proposing the concept of collaborative, personalized agent based spam filtering using Bayesian filtering as an agent function and email spam filtering and we are discussing how the spam spread through the mails. To achieve this, an Intelligent Agent is used. An Intelligent agent is an entity that observes and acts upon an environment to achieve its goal.

Description of Methodology/Approach



1 Input Email in raw format

The input email is present in the raw format in the dataset. It consists of the following fields: return-path, delivered-to, received array, messageId, to, from, subject, from, date, content-type, x-spam status, x-spam level, url, date

Return-Path: <rsshfeeds@example.com>

Delivered-To: yyyy@localhost.example.com

Received: from localhost (jalapeno [127.0.0.1])
by jmason.org (Postfix) with ESMTP id 4BDC816F6D
for <jm@localhost>; Tue, 1 Oct 2002 10:36:14 +0100 (IST)

Received: from jalapeno [127.0.0.1]

by localhost with IMAP (fetchmail-5.9.0)

for jm@localhost (single-drop); Tue, 01 Oct 2002 10:36:14 +0100 (IST)

Received: from dogma.slashnull.org (localhost [127.0.0.1]) by dogma.slashnull.org (8.11.6/8.11.6) with ESMTP id g91825K15748 for <jm@jmason.org>; Tue, 1 Oct 2002 09:02:06 +0100

Message-Id: <200210010802.g91825K15748@dogma.slashnull.org>
 To: yyyy@example.com
 From: newscientist <rssfeeds@example.com>
 Subject: Drug halves genital herpes transmission
 Date: Tue, 01 Oct 2002 08:02:05 -0000
 Content-Type: text/plain; encoding=utf-8
 X-Spam-Status: No, hits=-580.8 required=5.0 tests=AWL
 version=2.50-cvs
 X-Spam-Level:
 URL: <http://www.newsisfree.com/click/-1,8397801,1440/>
 Date: Not supplied

The antiviral drug significantly reduces the spread of the incurable infection through sex

2 Pre-Processing Of Email

In this step, we first extract the subject and the body of the email for parsing. Then the HTML parser gets the content of the mail which is then converted into plain text. We then drop the null values, and prepare training and testing dataset. CountVectorizer helps to tokenize the text and to create a vocabulary. We also use TFID vectorizer to highlight interesting words

3 MODEL SELECTION AND TRAINING

In this step, we use logistic regression and Naïve Bayes to train the vocabulary.

A) Logistic Regression

Logistic regression is a type of regression analysis in statistics used for prediction of outcome of a categorical dependent variable (a dependent variable that can take a limited number of values) from a set of predictor or independent variables. In logistic regression the dependent variable is always binary (with two categories). Logistic regression is mainly used to for prediction and also calculating the probability of success. Logistic Regression involves fitting an equation of the form to the data:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n - \text{eq. 1}$$

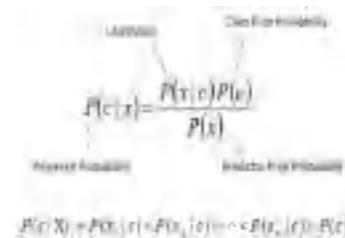
The regression coefficients are usually estimated using maximum likelihood estimation. The maximum likelihood ratio helps to determine the statistical significance of independent variables on the dependent variables. The likelihood-ratio test assesses the contribution of individual predictors (independent variables). Then the probability (p) of each case is calculated using odds ratio, $P/(1-P) = e^Y$ - eq. 2 From this p-value is found out. This gives the probability or chance for the individual to have coronary heart disease.

B) Naïve Bayes

Naïve Bayes is a simple but an effective classification technique which is based on the Bayes Theorem. It assumes inde-

pendence among predictors, i.e., the attributes or features should be not correlated to one another or should not, in any-way [4], be related to each other.

Even if there is dependency, still all these features or attributes independently contribute to the probability and that is why it is called Naïve. Naive Bayes has achieved an accuracy of 83.49% when all 13 attributes of the CHDD are used.



Results (analysis and associated uncertainty)

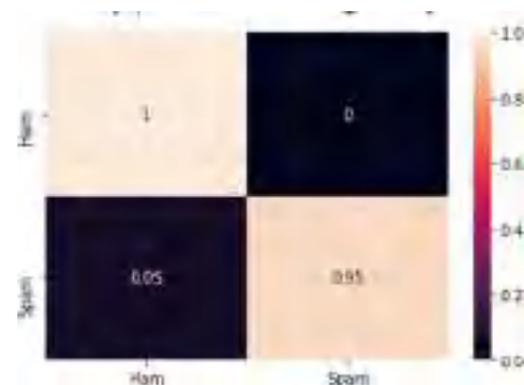
Multinomial Naïve Bayes

Precision: 100.0
 Recall: 95.0
 Accuracy: 99.18032786885246
 AUC Score: 97.5

Logistic Regression

Precision: 100.0
 Recall: 93.0
 Accuracy: 98.8524590163944
 AUC Score: 96.5000000000001

(Confusion Matrix)



App



Conclusions and Proposals for Future Work

Email spam has become one of the most demanding research topics due to increasing cyber crime and increasing spammers. The prediction models are developed using NLP and classification algorithms and vectorizers which achieved 99.180% accuracy and AUC score of 97.5 on the dataset used. The vectorizer used was TFIDF and classifier used was Multinomial Naïve Bayes Classifier. Naïve Bayes having probability distribution property determines the possible class for the email content from the spam or non-spam class on the basis of keywords present in the email content. The future work would include using more complicated classification algorithms coupled with NLP for the same purpose and use a bigger corpus for training. Naïve Bayes approach can also be used with another swarm optimization-based concept like any colony optimization, artificial bee colony optimization, firefly algorithm etc. for better results

References

- [1] Akram, A.U., Khan, H.U., Iqbal, S., Iqbal, T., Munir, E.U. and Shafi, M. (2018), "Finding rotten eggs: a review spam detection model using diverse feature sets", *KSII Transactions on Internet and Information Systems*, Vol. 12 No. 10, pp. 5120-5142
- [2] Al-Shboul, B.A., Hakh, H., Faris, H., Aljarah, I. and Al-sawalqah, H. (2016), "Voting-based classification for e-mail spam detection", *Journal of ICT Research and Applications*, Vol. 10 No. 1, pp. 29-42.
- [3] Balakumar, M. and Vaidehi, V. (2008), "Ontology based classification and categorization of email", *International Conference on Signal Processing, Communications and Networking (ICSCN '08)*, IEEE, pp. 199-202.
- [4] Feng, W., Sun, J., Zhang, L., Cao, C. and Yang, Q. (2016), "A support vector machine based naïve Bayes algorithm for spam filtering", *IEEE 35th International Performance Computing and Communications Conference (IPCCC '16)*, IEEE, pp. 1-8.

House Price Prediction based on Deep Learning and ARIMA Model

Vivek Chalke, PG Student, Pillai HOC
of Arts, Science and
Commerce, Rasayani

Ashwini Khillari, Asst. Professor, Pillai
HOC of Arts, Science and
Commerce, Rasayani

Abstract: The nonlinear relationship between influential factors and house price and inadequate number of sample size might be the explanation for the poor performance of the normal models. Meanwhile, the daily data of the important estate market is extremely huge and it's increasing rapidly. The normal house price prediction approaches are lack of capacity for enormous data analysis, causing low utilization of knowledge. To deal with these concerns, a house price prediction model supported deep learning is proposed during this paper, implemented on the TensorFlow framework. Adam optimizer is employed to coach the model, where Relu function is adopted to be the activation function. Then house price trend is predicted supported the ARIMA model. By using Scrapy, housing data are obtained from Internet to be the experimental dataset. Comparative experiments were conducted between the proposed approach and SVR method. The experimental results show that individual house price predicted by the proposed approach is best than that of SVR method. and therefore the predicted house price trend is especially agreement with the important situation

Inference: In this paper, a comparative study was made between SVR and ARIMA model and conclusion was that ARIMA model was better.



Introduction (Technical and Social Motivation)

Predictive models for calculating the selling price of houses in large countries such as the United States are still a difficult and time-consuming process. The value of properties in them is determined by a variety of interconnected variables. The size of the house, its location, and its amenities are all important factors that can influence the price. An empirical analysis was carried out in

this research work by considering the data set that is still accessible to the public by explaining the available housing properties in Kaggle, a machine learning hackathon platform. There are sixteen features in the data set. The aim of this research is to develop a predictive model for evaluating price based on the factors that influence price. Multiple linear regression (Least Squares), Logistic, Random Forest and Linear regression models, help vector regression, and boosting algorithms such as Extreme

Gradient Boost (XG Boost) Regression are used in modelling explorations and Light Gradient Boosting Machine (LGBM). These models are used to build a predictive model and to choose the best performing model by comparing the predictive errors obtained by these models. The goal here is to develop a predictive model for evaluating price based on price-influencing factors.

Description of Problem

Machine learning algorithms are used in modelling, where the machine learns from the data and applies what it has learned to predict new data. Regression is the most commonly used model for predictive analysis. The proposed model for accurately predicting future outcomes has applications in economics, finance, banking, healthcare, e-commerce, entertainment, sports, and other fields. Multiple variables are used in one approach for forecasting house prices. Multiple linear regression is a statistical technique for evaluating the relationship between several independent variables and the (dependent) target variable. To predict price, regression techniques are commonly used to construct a model based on several factors. In this report, we attempted to create a house price prediction regression model for a data set that is still open to the public on Kaggle's Machine hackathon platform. We looked at five different prediction models, including linear regression, logistic regression, and random forest

regression, as well as two different modelling techniques, extreme gradient boost (XG Boost) and light gradient boost (LG Boost).

Description of Methodology/Approach

- DATA UNDERSTANDING AND PRE-PROCESSING:

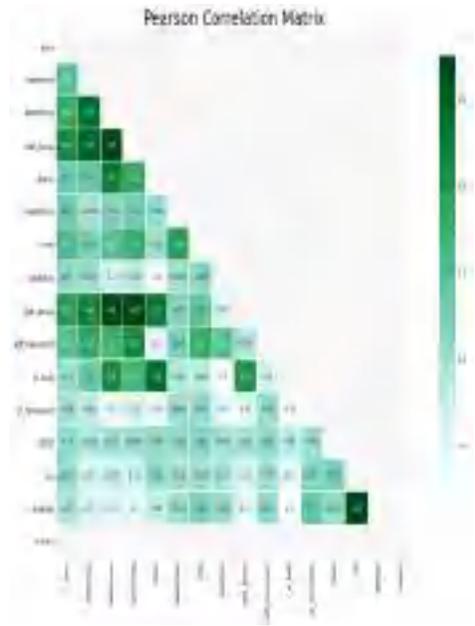
The two data sets used in the project, the train set and test data, were obtained from the Machine Hackathon platform. It is made up of characteristics that characterise house-property in the United States. In both data sets, there are 16 elements. The characteristics are as follows:

```
In[14]: pandas.core.frame.DataFrame
Int64Index: 4688 entries, 0 to 4687
Data columns (total 16 columns):
 #   Column          Non-Null Count  Dtype
---  ---
 0   price           4688 non-null   int64
 1   bedrooms       4688 non-null   int64
 2   bathrooms      4688 non-null   float64
 3   sqft_living    4688 non-null   int64
 4   floors         4688 non-null   int64
 5   waterfront     4688 non-null   int64
 6   view           4688 non-null   int64
 7   condition     4688 non-null   int64
 8   soft_sabote    4688 non-null   int64
 9   soft_basement  4688 non-null   int64
10  yr_built       4688 non-null   int64
11  yr_renovated   4688 non-null   int64
12  street        4688 non-null   int64
13  city          4688 non-null   int64
14  statezip     4688 non-null   int64
15  country       4688 non-null   int64
dtype: float64(1), int64(13)
```

With 16 features available, we try to build regression models to predict house price. We predicted the price of test data set with the regression models built on train data set.

- Data understanding and basic EDA

The aim is to build a model that can predict housing costs. The data is divided into functions and goal variables. In this section, we will attempt to comprehend an overview of the original data set, including its original features, and then conduct an exploratory review of the data set in order to obtain useful observations. There are 4600 records in the train data collection, with 16 explanatory variables. There were approximately 893 records in the test data collection, each with 16 variables. We often need to convert categorical (text) features to numeric representations when constructing regression models. The two most popular methods are to use a label encoder or a single hot encoder. The sklearn library can be used to encode labels in Python.



(Pearson correlation Matrix)

- Data Pre-processing

The general steps in data pre-processing are:

- To suit a linear regression model, categorical features are transformed into numerical variables.
- Null records are filled in with acceptable values.
- Data scaling
- Split into train test sets.

- **REGRESSION MODELS AND EVALUATION METRICS USED**

One of the most well-known statistical methods is linear regression. In statistics and machine learning, algorithms are used. The goal of a linear regression model is to discover

a relationship between two variables or more characteristics (independent, explanatory, or predictive variables) as well as a continuous target variable (dependent/response). The model is simple linear if there is only one function. If there are several attributes, the model is linear regression with multiple variables.

- **Linear regression**

The formulation for multiple regression model is

$$\hat{Y} = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n$$

The assumptions in the model are:

- The error terms are normally distributed.
- The error terms have constant variance.
- The model carries out a linear relationship between the target variable and the functions.

The multiple regression models are created by the researchers in this case. Ordinary Least Squares (OLS) is a least squares solution. It's difficult to assess the model's accuracy without using it. Using both train and test data sets to evaluate its performance. This is a possibility that can be achieved by the use of some form of performance

metric. It may be a result of measuring some kind of mistake, the goodness of a fit, or some calculation that is very useful.

- **Logistic Regression**

Logistic regression is named for the function used at the heart of the method, the logistic function. The logistic function, also known as the sigmoid function, was created by statisticians to explain the properties of population growth in ecology, such as how it grows rapidly and gradually exceeds the environment's carrying capacity. It's an S-shaped curve that can map any real-valued number to a value between 0 and 1, but never exactly between those two points.

$$1 / (1 + e^{-\text{value}})$$

Where e is the natural logarithms' base and value is the numerical value to be transformed.

- **Random Forest Regression**

- A Random Forest is an ensemble technique that uses several decision trees and a technique called Bootstrap and Aggregation, also known as bagging, to perform both regression and classification tasks. Instead of depending on individual decision trees, the basic concept is to combine several decision trees to

determine the final production. As a base learning model, Random Forest uses multiple decision trees. Row and function sampling are done at random from the dataset, resulting in sample datasets for each model. Bootstrap is the name of this section. We need to approach the Random Forest regression technique like all other machine learning technique

- Design a selected question or data and obtain the source to work out the specified data.
- Make sure the info is in an accessible format else convert it to the specified format.
- Specify all noticeable anomalies and missing data points which will be required to realize the specified data.
- Create a machine learning model
- Set the baseline model that you simply want to realize
- Train the info machine learning model.
- Provide an insight into the model with test data
- Now compare the performance metrics of both the test data and therefore the predicted data from the model.
- If it doesn't satisfy your expectations, you'll try improving your model accordingly or dating your data or use another data modeling technique.
- At this stage you interpret the info you've got gained and report accordingly

• **Modelling Techniques**

- XGBoost Features
- Extreme Gradient Boosting is laser focused on computational speed and model performance, intrinsically there are few frills. Nevertheless, it does offer variety of advanced features.
- Model Features
- The implementation of the model supports the features of the scikit-learn and R implementations, with new additions like regularization. Three main sorts of gradient boosting are supported:
 - Gradient Boosting algorithm also called gradient boosting machine including the training rate.
 - Stochastic Gradient Boosting with sub-sampling at the row, column and column per split levels
 - Regularized Gradient Boosting with both L1 and L2 regularization.
- LightGBM
- It is a gradient boosting framework supported decision trees to increases the efficiency of the model and reduces memory usage.
- It uses two novel techniques: Gradient-based One Side Sampling and Exclusive Feature

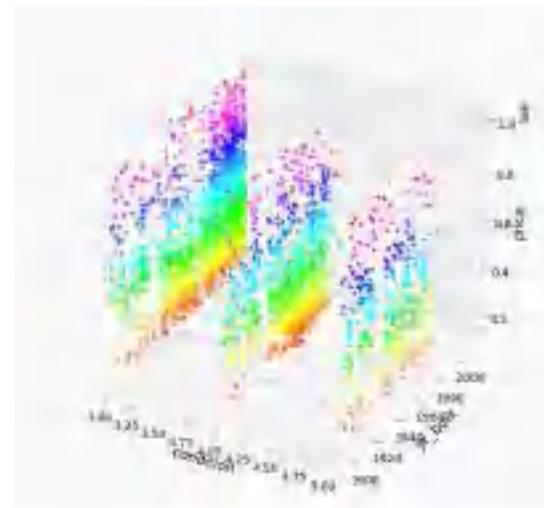
Bundling (EFB) which fulfills the restrictions of histogram-based algorithm that's primarily utilized altogether GBDT (Gradient Boosting Decision Tree) frameworks. the two techniques of GOSS and EFB described below form the characteristics of LightGBM Algorithm. They comprise together to make the model work efficiently and provide it a number one edge over other GBDT frameworks

- Gradient-based One Side Sampling Technique for LightGBM:
- Different data instances have varied roles within the computation of knowledge gain. The instances with larger gradients(i.e., under-trained instances) will contribute more to the knowledge gain
- . GOSS keeps those instances with large gradients (e.g., larger than a predefined threshold, or among the highest percentiles), and only randomly drop those instances with small gradients to retain the accuracy of data gain estimation. This treatment can cause a more accurate gain estimation than

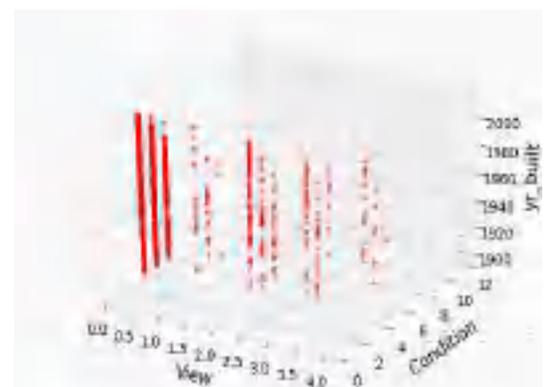
uniformly sampling , with an equivalent target rate , especially when the worth of data gain features a large range.

Results (analysis and associated uncertainties)

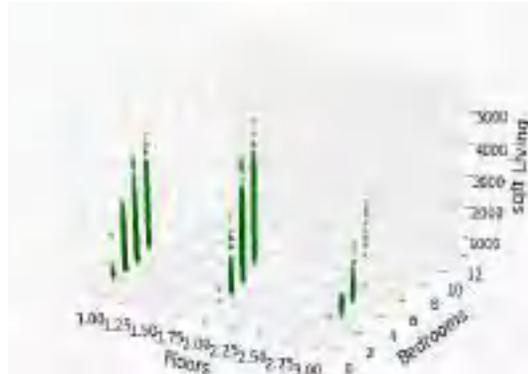
After training the model, we pickled the best model for our website. The following results were achieved:



(condition vs year built 3d satter plot)



(condition vs view 3d satter plot)



(floors vs bedrooms 3d scatterplot)

Accuracy achieved:

	Model	r2_train	r2_test	MAE_train	MAE_test	RMSE_train	RMSE_test
2	LR	0.14	0.04	7.79	7.63	602.828131	490.458603
1	GB	0.11	0.02	5.07	4.18	348.21142	312.81119
3	Random Forest	0.58	0.10	3.13	3.01	196.38597	188.88631

We created a portal which gives the user the prediction for real estate after some features are inputted. Through our method, a 95.58% accuracy was achieved. The model was also deployed on a Flask based server application.

Bibliography

[1] Limsombunchai, Visit. 2004. House price prediction: hedonic price model vs. artificial neural network New Zealand Agricultural and Resource Economics Society Conference.

[2] Wu, Jiao Yang(2017). Housing Price prediction Using Support Vector Regression.

[3] Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm

Conclusions and Proposals for Future Work

Literature Paper on Artificial Intelligence for Social Media

Author:-Chinmay Joshi
PG Student, Pillai HOC College of
Arts, Science, Commerce, Rasayani

Guide Name:-Swapnil Patil Sir
Assistant Professor, Pillai HOC College of
Arts, Science, Commerce, Rasayani

ABSTRACT-Research in Artificial Intelligence in last few years has greatly improved performance both Artificial Intelligence and Social Media. Currently there is needed to show article on this topic showing importance of Artificial Intelligence in Social Media and Social Media Marketing/Businesses. This paper shows the role of Artificial Intelligence in Social Media. It also glorifies the impact of AI in social Media. It also gives the idea of Present situation of AI in Social media and marketing, also gives the idea of how would it be in future.

1. INTRODUCTION

In 21st century Artificial Intelligence has become an important part of research in almost all fields engineering, robotics, medical, education, accounts, and finance as well as also in social media the presence of social media has already changed the statistics and with addition of Artificial Intelligence it has taken game a much higher. The use of social media is part of everyday life in both private and professional environments. Social media is used for communication, data exchange and the distribution of news and advertisements.

2. ARTIFICIAL INTELLIGENCE AND SOCIAL MEDIA

Social media dominances characteristics our era, where digital media can easily be created, communicated and read globally. As we know that

most of the people worldwide uses social media. We can't imagine our life without social media. They constantly update their profile their updates there post tweeting their views uploading new photos videos etc. This Information can be used by Social Media Marketers to buy and sell products by using Artificial Intelligence. These Figures obviously are expected to grow as mobile device users and social media users will grow among social media platforms. Face Book tops at list with no of users while what's App, Messenger ranks second, Instagram, Snap Chat, tumblr are also included in top ten lists. The potential of Artificial Intelligence have extended its platforms across several social media platforms as they use to make sense of pool of human data. Over all Artificial Intelligence is slowly integrating with humans.

3. ARTIFIAL INTELLIGENCE: THE PRESENT AND FUTURE OF SOCIAL MEDIA.

Integration of Artificial Intelligence Technology with social media for marketing and to gain competitive stage is driving the market. Rather than AI, there is no other field spreading in digital space. It's computing to Machine learning, marketing via chat bots, building industrial robots, gaming with Natural Language Processing and so on. Let's go some back to see what is Artificial Intelligence and what is Social Media? Artificial

intelligence is a way of making a computer, computer controlled robot. It is based on the study of how human brain thinks, how learns, how humans decide, and work while trying to solve problems and improve their intelligence by learning from past. Let's look at social media now, as we know social media is all about connecting people with each other; sharing information and communicating; making their presence in social and

impacting each other in good or bad way. This technological age is shifting at rapid speed some believe that that Web4.0 has arisen but not fully emerged. This age is signified with AI integrating and forming a relationship with humans. It is straight forward that AI which is based on, learns

As there is rapid development in Artificial Intelligence Field from few decades, one of those is demanding Artificial Intelligence. Present Artificial Intelligence is ruling over the world and in future all almost all fields will be depending on Artificial Intelligence for:-e.g. medicinal service, hospital ,also in service sectors etc ,it has also developed impact on social media as well as also in Social Media and helped in the growth of social media business. In Present functionality of social media applications has increased due to AI as well as it also playing vital

4. IMPACT OF ARTIFICIAL INTELLIGENCE ON SOCIAL MEDIA.

Artificial Intelligence has greatly influenced the lives of the people.AI helps to a difficult task easier way and it is been less costly than traditional mode. Social media and social media marketers saves lot of time and results in to reducing the cost to great extent.AI helps to understand the importance of need to growth for the organization.AI plays important role in improving customer services to great extent by fulfilling the needs of customers that would help them to stay ahead in their business, due to which it becomes one of the prime reason in increase in sales in different retail and product companies. Artificial Intelligence technologies are being used by companies to improve an evaluate efficiency of their different activities. As we share a vast array of data through social media platforms which include photographs, thoughts, etc. as per record the total number of social median users are up to 3billion and is expanding at a steady rate. But extracting the relevant data from this vast ocean of information and data is an extremely challenging task for any business. This is where Artificial Intelligence provides solution for analyzing the huge storehouse of data effectively and efficiently. There are no of companies which provide social listening services, and the social analyst makes use

human tactics will be an important part of social media evolving in to social ecosystems that shows excellent behavior, demonstrating, and explaining, giving advice to the users. We will be witnessing the huge change in social platforms, methods of sharing, offering of service

role in social media marketing as well. Artificial Intelligence has been an interesting topic from few years especially when looking at the future. As ha great effect on business in future Ai will be widely used by businessmen. Also AI will help to solve smallest errors threat for free environment for social media businessmen. At last it is found that Artificial Intelligence has made positive effects on social media application of social media and also on different sector in world.

of Artificial Intelligence to help them for sorting of data. Social media is an important platform for social media business to interact with their customers to enhance their brands/products. With some specific algorithms, AI can scan through user's comments and posts to help business monitor.

As we know Artificial Intelligence is used in various platforms ,Social Media is one only source for companies to grow their business, like Ai based Chat bots are helping business companies to stay connected with their audience .From small scale businesses to large scale business are adopting AI in their social media business. The messenger apps like face book, what's app have started playing a vital role in the field of social media marketing by improving customer interaction, it is predicted that, In future 80% of all customer service interaction will be driven by Artificial Intelligence. The chat bots we use helps us for faster responses which allow quicker interactions with customers and improve customer experiences. The tweets and comments of a customer provide lot of information regarding their choice patterns. The Impact of Artificial Intelligence in Social Media Marketing has major advantages:-

4.1. Cost Reduction

AI uses a complex network of computers capable of functioning independently, When algorithms and points are defined, AI functions with minimum human interaction. This will reduce marketing cost and can work with smaller team as well.

4.2. Improve Customer Service

Customers have power than before which shows customer experience is become increasingly important.80% of customers are busy in using social media to interact with the brands, so companies must communicate with their customers n a way suitable to them. But unless and until companies can't afford to hire social media staff it is hard to this task. Luckily these companies can use AI tools to answer the questions of customers and keep communication with them.

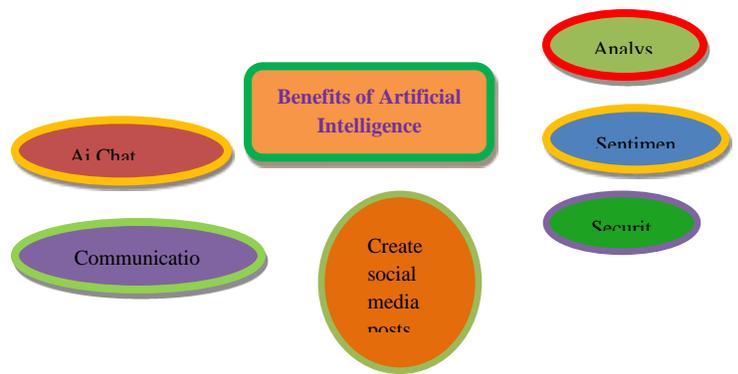
4.3. Increment in Revenue

AI can enhance the efficiency of Social media posts or content, which translates into a better result from investments in social media.AI helps us to determine which posts will be more effective, which audience is more relevant and on which channel works best. It also forces to boost revenue by making processes more effective and efficient.AI also provide deep understanding to improve engagement and make brand more visible to audience.

4.4. Customizing Social Media Content

The evolution of Artificial Intelligence has led to the generation of more personalized content. It may be benefit for social media business and business strategies there is need for customizing data when it comes to social media. Automated bots and AI assistants made it possible through the use of Artificial Intelligence. Social media marketers make use of these assistants to improve marketing strategies.ai uses text mining to answer emails daily. There are companies providing AI aided customer service. In near future al the service oriented jobs will be based on Artificial Intelligence.

5. BENEFITS OF USING ARTIFICIAL INTELLIGENCE FOR SOCIAL MEDIA



5.1. Recognize images

AI powered image recognition software and tools help in recognition of images to understand the change in user's behavior. Through algorithms it can go millions of such images to collect information.

5.2. Communication with the Audience

Now a day's social media is used by companies to promote and share their products, post an update about their brands. All social media platforms have messaging and commenting feature which allow their followers to communicate with companies about their products. Sometimes customers might have queries. So it is not possible to neglect their queries as it can make on company's products reputations, It helps these to solve this issue by automatically giving reply to these messages by developing user interactive AI chat bots.

5.3. Ai Powered Chat bots

Business running over social media applications can use AI to reduce workload and provide a faster response to customers, some groups of organizations use chat bots to answer the query or doubts of customers in no time.AI enabled chat bots can efficiently conduct conversations between customer and service provide them the required answers by understanding the intent of their queries. Chat bots if possible can be programmed by set of replies for immediate reply for frequently asked questions. Chat bots are not only cost effective for hiring team members but also they can do it in more efficient way and sometimes also in human manner.

5.4. Predictive Analysis

Predictive Analytics makes use of machine learning and statistics to analyze behavior of user. Since humans are always predictable and we have a



general awareness and machines have come to make these predictions more specific

5.5. Analyzing Sentiments

As AI can analyze the nature of intent of query or posts posted by user, it can help brands to identify sentiments. For this task AI uses another subset known as Natural Language Processing as it helps in finding out positive and negative words in post posted by users and comments of users.

5.6. Increased Security

AI can help social media platforms to protect the user data and increase the privacy of information. Through user authentication and fraud prevention and other features this technology helps users to improve the security of their social media.

5.7. Create, Analyze Social Media Posts

Regularly posting an update can help to grow followers as well as audience and build a strong social media presence. Having multiple accounts in social media, it becomes difficult for company to handle each of these platforms .AI helps to develop platform specific content by providing information. It identifies patterns and user behavior to help understand brands. AI helps allows brands to schedule post automatically applying specific time interval. There are specific time intervals when posts tends to get more engagements, Artificial Intelligence tools can identify the best time for brands content to post them automatically.

CONCLUSION

We can say that Artificial Intelligence is dominating the world in present and will dominate in future too in more advanced way. AI is helping social media business to create a strong bond between brands and customers, It has also help to develop strong bond between two parties. If AI works in full form in social media marketing it would help to extend the growth of business up to height and productivity .Artificial Intelligence is face of a society in future and proper use of AI will transform the world in better position. To conclude there are several advantages of using AI in Social Media. Technology is still evolving and developing which will help for benefit for social media business.

ACKNOWLEDGEMENT

I want to give a vote of thanks to our guide Swapnil Patil Sir (Assistant Professor, Pillai HOC College of Arts, Science, Commerce) for their guidance. Because of their guidance I was able to complete this task. Thank you Once Again!

REFERENCE

- 1) Role of Artificial Intelligence in Social Media, Quytch Site:-<https://www.quytch.com/blog/role-of-artificial-intelligence-in-social-media/>
- 2) Effects of Social Media: Papers Owl Site:-<https://papersowl.com/examples/negative-effects-of-social-media/>
- 3) Impact of AI on Social Media: -
<https://www.rtu.lv/en/studentsservice/about-us-ssd/news-11/open/the-impact-of-ai-on-social-media>
- 4) Impact of Artificial Intelligence on Social Media, Social Media Explorer Site:-
<https://socialmediaexplorer.com/content-sections/news-and-noise/the-impact-of-artificial-intelligence-on-social-media/#:~:text=Create%2C%20Schedule%2C%20and%20Analyze%20Social%20Media%20Posts&text=AI%20can%20help%20to%20create,content%20is%20getting%20more%20engagement>
- 5) International Journal of Technology and Engineering: - <https://www.ijrte.org/wp-content/uploads/papers/v8i4S3/D10011284S319.pdf>
- 6) Future of Artificial Intelligence on Social Media, Warren Knight Site:-<https://warren-knight.com/2017/06/02/future-artificial-intelligence-social-media/>

Literature Review Artificial Intelligence for RoboSoccer

MAYUR KAMBLE

PG Student, Pillai HOC College of
Arts, Science, Commerce, Rasayani.

Guide Name:-Swapnil Patil Sir
Assistant Professor, Pillai HOC College of
Arts, Science, Commerce, Rasayani

ABSTRACT:- Robotic football is a sport organised by the Federation of International Robot-soccer Association. It aims to create a team of robots capable of beating a human side at football by 2050.

INTRODUCTION

Multi-agent system is a subfield of Artificial Intelligence (AI) that aims to provide both principles for construction of complex systems involving multiple agents and mechanism for coordination of agents behaviors. An agent can be considered to be an entity, such as robot, with goals, actions and domain knowledge, situated in an environment. The way it acts is called its behavior. Although the ability to consider coordinating behavior of autonomous agents is a new research direction, the field is advancing quickly by building upon preexisting work in the field of Distributed Artificial Intelligence (DAI).

Increasingly, Multi-agent systems are being designed for a variety of complex, dynamic domains including autonomous vehicles and even some human agents.

Effective agent interactions in such domains raise some of the most fundamental research challenges for agentbased systems. An agent in such domains must model

other agents behaviors, learn/adapt from its interactions, form teams and act effectively in a team, negotiate with other agents, and so on. For each of these research problems, the uncertainty and the presence of multiple cooperative and non-cooperative agents, conspires to exacerbate the difficulty.

Consider the challenge of designing multi-agent teamwork, which has become a critical requirement across a wide range of multi-agent domains. In this case, an agent team must address the challenge of designing roles for individuals, doing so with fairness, and reorganizing roles based on new information. Furthermore, agents must also be flexibly coordinate and communicate, so as to be robust despite individual members incomplete and inconsistent view of the environment, and despite unexpected individual failures. Learning in a team context also remains a difficult challenge.

Robotic Soccer was proposed to be a general test-bed for studying multi-agent system techniques. Originated by Mackworth, it has been gaining popularity in recent years, with several international competitions in several different leagues. The

simulation league is of particular interest and attracts the largest number of participants. The stated research goals of the simulation league are to investigate the areas of multi-agent teamwork, agent modeling, and multi-agent learning. It can be used to evaluate different multi-agent system techniques in a direct manner: teams implemented with different techniques can play against each other.

ROBOTIC SOCCER SYSTEM OVERVIEW

The first robotic soccer system was the Dynamo System. Barman et al. Built a 1 vs. 1 version of the game. Now there are mainly three kinds of competitions in Robotic Soccer:

- Real Robot League: Using physical robots to play soccer games.
- Software Agent League: Using software or synthetic agents to play soccer games on an official soccer server over the network.
- Expert Skill Competition: Competition of robots which have special skills, but are not able to play a game.

As discussed above, robotic soccer can be played either with real robots or in a simulator. Of particular interest in this review is the simulation robotic soccer. A particularly good simulator is the soccer server developed by Noda. The simulator was first used for a competition among twenty-nine teams from around the world in 1997 and contains to be used for this purpose currently.

This simulator server simulates the players bodies, the ball and the environment. Software agents provide the brains for the simulated bodies. Thus, 22 agents, who do not share memory, are needed for a full game. Visual and audio

information as the sensory perception sensed by the player body is sent to the player agent, which can then send action commands as actuators to control the simulated body. The server constrains the actions an agent can take and the sensory information it receives. For instance, with the server used in the 1997 competition, a player could only send one action every 100 milliseconds and receive perceptual updates every 300 milliseconds. The server also simulates stamina: If a player has been running too hard, it gets tired, and can no longer dash as effectively. Both actions and sensors contain a noise factor, and so are not perfectly reliable. The quality of perceptual information depends on several factors, such as distance, view angle and view mode. All communications between players are done via the server, and are subject to limitations such as bandwidth, range and latencies.

TECHNICAL CHALLENGES IN ROBOTIC SOCCER

Robotic Soccer is an attempt to promote AI and robotics research by providing a common task, Soccer, for evaluation of various theories, algorithms and agent architectures. Starting with the first competition held in 1996 and continuing since then, there has been a great deal of robotic soccer-related research. It has been presented both at dedicated robotic soccer workshops held in conjunction with the competition and in other scientific forums.

For an agent to play soccer reasonably well, a wide range of technologies need to be integrated. The range of technologies spans both AI and robotics research, such as design principles of autonomous agents, multi-agent collaboration, strategy acquisition, real-time reasoning and planning, intelligent robotics, sensor fusion and so forth.

Robotic Soccer challenges can be organized into their major classes:

1. Synthetic Agent Challenge

2. Physical Agent Challenge
3. Infrastructure Challenge

The Synthetic agent challenge deals with technologies involving multi-agent control and robotic soccer strategies, which can be developed using software simulator. The research issues in this aspect involve; learning challenge; teamwork challenge; opponent challenge.

The physical agent challenge intends to promote research using real robots. Details of this challenge were described in. The infrastructure challenge is to facilitate research to establish infrastructure aspect of Robot Soccer, AI and robotics in general. Such challenge includes education programs, common robot platforms and components, standard automatic commentary systems and intelligent studio systems for robot soccer games.

LAYERED LEARNING METHOD

Layered learning is a general- purpose machine learning paradigm for complex domains in which learning a mapping directly from agents sensors to their actuators is intractable. Given a hierarchical task decomposition layered learning allows for learning at each level of the hierarchy, with learning at each level directly affecting learning at the next higher level.

- Neural networks are used by individual players to learn how to intercept a moving ball.
- With the receivers and opponents using the first learned behavior to try to receive or intercept passes, a decision tree(C4.5) is used to learn the likelihood that a given pass would succeed.
- TPOT-RL(Team-Partitioned, Opaque-Transition Reinforcement Learning) algorithm [21], a new machine

learning method used to train collaborative and adversarial team behavior, is used to learn pass selection, taking advantage of the learned pass-evaluation capability to construct the input representation for learning. However, due to the system complexity, the results haven't been applied directly.

CONCLUSION

Robotic Soccer is a useful domain for the study of multi-agent systems. The review describes control strategies in the field of robotic soccer from various multi-agent viewpoints. It provide an introduction to people unfamiliar with the field and gives an organized overview of the research in this area.

The goal of our project is to develop and evaluate the performance of strategies for Distributed Cooperation and Control for Autonomous Air Vehicles. Some of the common technical challenges between robotic soccer and control of autonomous air vehicles include the following:

1. Machine learning in a multi-agent, collaborative and adversarial environment;
2. Multi-agent architectures, enabling real-time multi-agent planning and plan execution;
3. Opponent modeling.

In general, understanding the progress in research for robotic soccer will also be helpful in designing distributed cooperation and control strategies for multiple autonomous air vehicles.

ACKNOWLEDGEMENT

I want to give a vote of thanks to our guide Swapnil Patil Sir (Assistant Professor, Pillai Hoc College of Arts, Science, Commerce) for their guidance. Because of their guidance I was able to complete this task. Thank you Once Again!

programming. Robotics and Autonomous Systems, 21:161-176, 1997.

REFERENCES

- [1] D. Andre and A. Teller. Evolving team Darwin United. In M. Asada and H. Kitano, editors, RoboCup-98; Robot Soccer World Cup II, pages 377-388. Springer Verlag: Berlin, 1999.
- [2] M. Asada and H. Kitano. Lectures notes in artificial intelligence. In M. Asada and H. Kitano editors, Robocup98: Robot Soccer World Cup II, Pages 373-388. Springer Verlag; Berlin, 1999
- [3] M. Asada, M. Kuniyoshi, A. Drogoul, M. Mataric H. Asama, D. Duhout, P. Stone, and H. Kitano. The robocup physical agent challenge: Phase-I, Applied Artificial Intelligence(AAI), 12:251263, 1998
- [4] R. A. Banman, S. J. Kingdom, J. J. Mackworth, A. K. Pai, D. K. Sahota, M. Wilkinson, and Y. Zhang. Dyanmo: Realtime experiments with multiple mobile robots. In Intelligent Vehicles Symposium, Pages 261- 266, Tokyo, 1993
- [5] H. Kitano, Robocup: A challenge problem for AI. AI Magazine, 18:340-347, 1997
- [6] J. H. Kim. Path Planning for multimobile robots in the dynamic environment. In Proceedings of the Micro-Robot World Cup Soccer Tournament Taejon, Korea, 1996.
- [7] J. Li. De La Rosa. A. Oller, J. Vehi and J. Puyol. Soccer team based on agentoriented

Review Paper on Artificial Intelligence for Social Media

Author-Chinmay Joshi
PG Student, Pillai HOC College of
Arts, Science, Commerce, Rasayani

Guide Name:-Swapnil Patil Sir
Assistant Professor, Pillai HOC College of
Arts, Science, Commerce, Rasayani

ABSTRACT-As we know social media platforms like Face book, Twitter, What's app, Instagram plays a vital role for us. It enables people over worldwide to get connected and communicate with each other. We also know social media systems helps humanity to aware about harmful diseases spreading, to find a way of problem faced by humans. On the other side social media can also do harm like sharing false and fake news spreading wrong information in society in order to violate the privacy of individual. By using Artificial intelligence with help of machine learning algorithms has changed the way of using social media by humans. This paper explores the use of artificial intelligence for protecting social media from attacks. It also specifies the role of artificial intelligence in social media and social media marketing.

1. INTRODUCTION

Artificial Intelligence is a new trend in the world of advertisement and recommendations. It is effective at making an effect on social media. Artificial Intelligence is the power of thinking and imagination allocated to machines which are not capable of it For example: Robotics. Artificial Intelligence is like a gift for development of progress of technologies. Artificial intelligence has always been holding the individuals potential and possibilities. As Social Media is vast platform, it is providing us with lot of content. Social Media provide the support for emergency need and allows sharing of information so that people are informed.

1.1. What is Artificial Intelligence?

Artificial Intelligence is the development of computer systems that are able to perform tasks that requires human intelligence. Artificial Intelligence can be used to personalize and create relevant data. For e.g. Netflix uses this technology to identify the perfect show preference of the user and recommend the shows in future.

Also we can say that Artificial Intelligence is wide ranging branch of computer science that refers to ability of a computer or, machine to mimic the activities of human mind like responding to language, recognizing objects, solve problems like human intelligence.

1.2. What is Social Media?

Social media are digitally created technologies for sharing exchanging of information, ideas, and

their forms of expression. Users usually access social media platforms from their own mobile forms computers or laptops.

As users are engaged in social media they can create individual platforms, they can create interactive problems due to which individual, communities, organizations can share, discuss, and participate user generated content posted online.

2. RESEARCH METHODOLOGY

The aim is to understand the interaction of Artificial Intelligence with Social Media. This paper will explore the ideas related to AI.

- To understand the functional Artificial Intelligence social media
- To understand the importance of Artificial Intelligence in Social Media.
- To explore the potential of AI in Social Media
- To evolve the role of Artificial Intelligence for Security and Privacy for social media

3. ARTIFICIAL INTELLIGENCE FOR SOCIAL MEDIA

Keeping in mind the reality that almost a third of world's population is using social media. It will not wrong to say that we live in social media's era. AI used machine learning techniques to have a role in Social Media. Machine learning techniques are applied for various social media platforms because they predict location of user and analyze it. On the other side social media platforms are subjected to cyber attacks any unauthorized attack may lead to

changes in the content of messages shared. Some malicious software helps attackers to create fake profiles and spread false information. Artificial intelligence uses machine learning algorithms to find source of all these fake things spreading and stop it. Artificial Intelligence can assist Social Media by developing and evaluating AI based tools and framework to collect, analyze, summarize, visualize data, NLP i.e. Natural Language Processing is one of the AI based tool which is commonly used for understanding words. This too is helpful to Social Media where a group of languages and emojis are used by users. This tool involves the task of information extracting, categorization, summarization and translation of data to new data. Now a day's Industries use this technology for analyzing opinions and reviews of customers on social media.

4. USE OF ARTIFICIAL INTELLIGENCE IN SOCIAL MEDIA

First of all there should be healthy balance in human intelligence and Artificial Intelligence. Artificial Intelligence is a key component of the popular Social Media. The data gathered by social networks is enabling to analyze for humans and to exploit them. Let us see how various social media platforms use Artificial Intelligence.

4.1. How Face book uses Artificial Intelligence?

Face book uses advanced machine learning techniques to do everything from serve the content to you to recognize your photos to target users with advertising. Face book puts large amount of focus on AI and have developed face book recognition tool that helps to find a perfect person for tagging in a picture and also uses machine learning techniques for facial recognition of users.

We all have seen that face book is growing day by day; today this platform is becoming more robust and upcoming with new features over social networking. As face book is building its business high velocity by learning about its users for benefit of advertisers, the company function with the aim to connect the people on the planet through face book owned products like Whatsapp, Instagram. Face book has evolved a platform making able conversation and communication between people as a source of knowing their interests, hobby, and lifestyle. And that where Artificial Intelligence

comes, AI enables machines to learn to clarify data .simplest example of it, it is that analysis of image.

4.2. How Twitter uses Artificial Intelligence?

Twitter is social media platform with almost 300 million users. Twitter offers entertainment option not only to users even also to businessman uses twitter to promote their events, products as well as spread awareness about their brands. Twitter has become ninth largest network in the world. Twitter uses AI to improve experience of the users let's see how it is utilizing.

Major challenges that are faced by twitter is showing suggestion to all twitter users, but some of those suggestions are not of use to other users rather than who have interest. AI helped to solve this issue. AI algorithms scans and rank them according to individuals preference. By using AI twitter is successful in detecting and eliminating twitter accounts related to terrorists. ai algorithms helps to identify such accounts.

4.3. How Instagram uses Artificial Intelligence?

Instagram has millions of users around the world, additional it is owned by face book. Every day millions of photos, videos etc are uploaded on this platform by users to share with their followers. Meanwhile people simply busy with these posts by liking posts, commenting on it, watching videos, reels, etc. All of these activities requires large amount of data. Use of AI in Instagram is as follows:-

4.3.1. Spam Filter

The spam filter is capable of detecting fake messages and deletes them from instagram accounts. To avoid spam instagram uses Artificial Intelligence. Once these platforms detect fake messages it will permanently remove them. Apart from messages instagram is also able to identify fake accounts and fake followers. So if you wishing to buy followers so make it sure it is real and authentic. Still no needs to worry because most of the sellers are trustworthy.

4.3.2. To understand human interaction and conditions

Through analysis Instagram uses Artificial Intelligence to understand human behavior. For instance AI uses to analyze how people use language patterns; it can be used to study peoples dressing patterns and much more. The use of AI and AI related

technologies allow for the extraction of insights that can be analyzed to detect human behavior and conditions.

4.4. How LinkedIn uses Artificial Intelligence?

LinkedIn is social media network that has been using Ai in everyday processes. LinkedIn is Microsoft owned site having millions of subscribers recently approached to handling accounts that has inappropriate content and etc. LinkedIn uses artificial intelligence to provide a better recommendation or to offer a better job-candidate match, creating a just environment for both the parties. LinkedIn uses this to rate candidates for companies based on prior hiring patterns, location, past work experience and job descriptions.

Its technology provides connection suggestions and targets posts to users and offers job recommendations for better candidates for employers. LinkedIn uses Data analytics to improve their member's experience. Members view it every day by getting job recommendations and suggestions to connect with people to reading this useful content. The AI system had a huge impact on both sides of platform creating a lot of value. There was up to 30% increase in job application because of improvement of personalized tool. The email responses to cold emails from recruiters have also increased. AI has also improved its recommendations in feed by 10%. With all these amazing features LinkedIn platform has reached almost 600 millions of users and engagement has increased over 50% year. LinkedIn is also adding tremendous value by adding AI to make recruiting candidates a no brainer.

4.5. How Chat Bots use Artificial Intelligence?

Ai is a powerful way for chat bots to connect with prospects. Chat bots provide faster customer service resolution, as well as provide right kind of support. Chat Bots is also an Artificial Intelligence tool which is able to maintain a conversation with a user using natural language on various platforms like email, websites or mobile applications. Chat bots provide faster customer service resolution, as well as optimum support, based on previous experiences with customers and also the immediate response.

Seeing earliest versions of chat bots it seems to be frustrating and time consuming processes. These bots were equipped to respond every input and could not any information outside of parameters. Due

to this communicating with a bot was far less an option for most of the customers than speaking with human. Artificial Intelligence has changed this.

Ai powered NLP i.e. Natural Processing Language has enabled chat bots to mimic human conversation, they can identify intent behind the text a real person type, then deliver a response that matches the intent. Additionally chat bots with NLP can now learn from past conversations and improve their ability to provide appropriate responses and solutions. And with this advanced level of processing, they are becoming extremely helpful in customer service situations.

4.6. How Snap Chat use Artificial Intelligence?

This social media company started acquiring two AI companies in order to improve their animated lenses and to start enhancing their video capabilities. Use of Artificial Intelligence in Snap Chat is:-

4.6.1. Recognizing text in videos

Snap Chat uses Artificial Intelligence Companies to recognize text in the video which will then add content for e.g. text "beautiful" automatically create comic icon.

4.6.2. Snap chat cameo feature

Artificial Intelligence use snap chat to edit someone's face in a video. Once user has picture of themselves or any other persons then they can insert into a short video through cameo feature.

4.7. How YouTube uses Artificial Intelligence?

YouTube is the major video platform that uses Ai throughout the different parts of their website.

4.7.1. Skimming the videos

AI application is used to skim the videos that are being posted repeatedly in order to find opposable content. Surely this kind of AI must be powerful enough to check thousands of videos every day. According to research almost 8 millions of videos were removed in first quarter of 2019. More than half of these videos were identified by AI. This shows how much AI is important to remove videos that are against company's guidelines.

4.7.2. YouTube's up next Feature

Another feature where AI is used is Up Next Feature. This algorithm is working in real time to suggest best video for you. This initial algorithm is based on how many people clicked the video.

4.7.3. Google Brain

As Google is the owner of this video platform, Google brain uses unsupervised learning. The main characteristic of this algorithm is how easy is for the algorithm to find relationships between different inputs. For e.g. Brain Algorithm started recommending short videos for mobile users and long term videos for YouTube TV apps.

5. ARTIFICIAL INTELLIGENCE FOR SECURITY AND PRIVACY FOR SOCIAL MEDIA

Artificial Intelligence can also define as machine to do smart intelligence without human support. The Social Media systems can get attacked by malicious attacks, but to avoid these, machine learning techniques are explored to find such attackers. The tough challenge is to find out fake users on social media, fake users may have false email id and accounts, due to which they can easily post fake info. Second problem is that to secure the privacy of individuals. For Securing Social Media. AI Security tools are used to identify good versus bad by comparing behavior of users. Also they work to discover, analyze, learn security threats without human interruption common security tools work on following capabilities Fast Detection, Advanced Techniques, Learning, at last presenting multiple solutions for best remediation.

6. CONCLUSION

Artificial intelligence is everywhere, it works silently in the background every time we open our computer or we interact with applications, it gathers data and learns from it. This paper has discussed the role of Artificial Intelligence in Social Media. The role Artificial Intelligence in Social Media is just beginning .As mentioned in article above AI can be a powerful tool for social media companies, it can compete thousands of queries in less than a second. In order to conclude we can say that Artificial Intelligence is emergence for social media and social media business. Based on the above contents it can be concluded that social media and artificial intelligence together are ending up being effectively for organizations.

ACKNOWLEDGEMENT

I want to give a vote of thanks to our guide Swapnil Patil Sir (Assistant Professor, Pillai HOC College of Arts, Science, Commerce) for their guidance. Because of their guidance I was able to complete this task. Thank you Once Again!

REFERENCE

- 1) Artificial Intelligence for Social Media Safety and Security:-
https://www.researchgate.net/publication/343583656_ARTIFICIAL_INTELLIGENCE_FOR_SOCIAL_MEDIA_SAFETY_AND_SECURITY
- 2) The Role of Artificial Intelligence and Cyber Security for Social Media
<https://ieeexplore.ieee.org/abstract/document/9150376>
- 3) Ai in Social Media, Klint Marketing Site:-
<https://klintmarketing.com/ai-social-media/>
- 4) Anja Bechmann and Geoffrey C Bowker(2014):-
https://Www.Researchgate.Net/Publication/330879115_Unsupervised_By_Any_Other_Name_Hidden_Layers_Of_Knowledge_Production_In_Artificial_Intelligence_On_Social_Media
- 5) Emerging Role of Artificial Intelligence in Modern Society:-
https://www.researchgate.net/publication/343179789_THE_EMERGING_ROLE_OF_ARTIFICIAL_INTELLIGENCE_IN_MODERN_SOCIETY

ROBOSOCCERS IN ARTIFICIAL INTELLIGENCE

MAYUR KAMBLE

PG Student, Pillai HOC College of Arts, Science and
Commerce, Rasayani

SWAPNIL PATIL

Asst. Professor, Pillai HOC College of Arts, Science and
Commerce, Rasayani

Abstract:- Football is one of the best sport and the most popular sport which played by all the people in the world. Football has a high element of co-operation, contains tactics, strategy, state, and has goal to target and also plenty of complexity. Nowadays, football orsoccer played by all kind of things not only human but also animals and Robots.

I. Introduction

Robosoccer was organized at the Moscow institute of physics and technology, Russia, in the autumn of 2018 by the honorary graduate of MIPT Azer Babaev. It was organized by the team Starkit.

Team Starkit has two subteams in Robocup – Humanoid League team and Standard Platform League team. Team basically consists of students. Seeking to emulate the behavior’s and actions of living beings, artificial intelligence and robotic systems has the aim to emulate human abilities.

Our external goal is to participate in the competitions and win the prizes. Our internal goal is to create a scientific school in modern humanoid robotics with artificial intelligence and to teach new generations of students.

II. Main Features of Robosoccer

Robosoccer is a competition which is played by Robots autonomously. Now it is little bite different from virtual environment for playing „Robots“ on real lawn.

In such competitions there are several difficulties that player has to meet. Firstly, the environment is changing rapidly, so in that robot must have to work autonomously. Secondly, you have to find a place by yourself on the field, find the ball, analyze your opponents, and goal. Thirdly, Robots participate in the game they should know how to kick the ball, when to pass the ball to your teammate, run on two legs and get up quickly after a fall. Team place robots against each other and compete it for a period. The rules of game are same as similar to the FIFA.

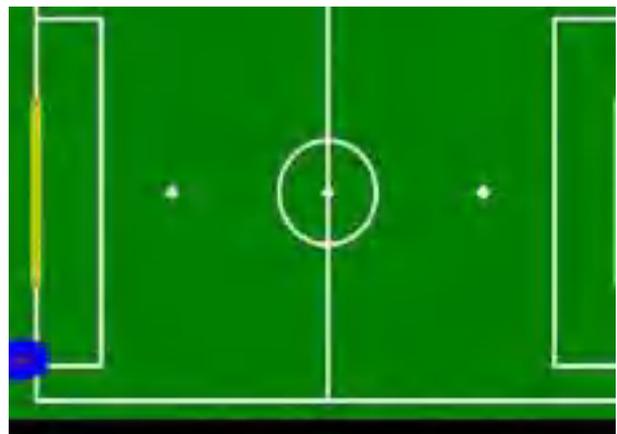
It uses the classical methods of computer vision and deep learning using convolutional neural networks(CNN). It is use for recognizing the ball, regions of interest are highlighted on the camera image, and then the bounding box is fed to the CNN.

What is CNN.? So CNN is a Convolutional **Neural Network** (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The movement is the most important task. The walking engine is simple closed cycle with a reaction only to pressure sensors located in the legs. It allows you to

understand when the step is finished and you can start the next. Trajectories of each node of leg are described by polynomials of the fifth degree. This allows you to get more human-like walking.

Soccer is a team game. Robots use onboard artificial intelligence for this. They have need to interact themselves and with the enemy. Receiving messages from the teammate's position of the opposite team player from each other, they have to make a decision who will go to the ball and pass the ball or make a goal. Also, they share the estimated positions of the ball after a hit. The AI decides what the robot will do on the field.



III. Computer Vision

Development of new systems of computer vision and its modification is one of the most important task in making of autonomous system. It requires correct applications of received information and calculations of the particular program to be made for playing the robots automatically. In this, requirements for complete autonomy for the robot on a computer impose strict

restrictions on the number of operations and the complexity of algorithms. In this computer vision, the programmer has to make a program that should detect the line, goal, ball and the opponent player. They should have to make the perfect algorithm for the line detection for the robot. The system allows the students to implement their C++ and python in real life.

A. Object Classification:-

Object classification means classifying and determining what objects are at image and classify them. It has two stages, Region of interest in which regions are highlighted on the camera and these regions are sent to CNN input for binary classification. In these stage, it has made the program for robot to identify the ball size and able to predict the size depending on its position on the image. This make it possible to accurately determine the distance from the ball and process of detecting the goalposts. It consist of convolutional layer with a 5x5 core, MaxPooling layer with a 8 core and fully-connected layer.

B. Obstacle Detection:-

Obstacle detection is the process of using sensors, data structures, and algorithms to detect objects or terrain types that impede motion. Playing robosoccers, there are limitaions which are imposed on the environment. In this robot acts in smooth part of environment in which certain colours and size has

been program for them by the coder. In this the obstacle has to find non-green and non-white colour in the field in which the robot has to follow and detect that given instructions. Obstacle detection is the main feature for playing robosoccer, if it is no there then robot will not able detect and analyze the line, ball or goalpost.



C. Object Approach

Object approach is the interaction of the robots with the object in the field, localization of the position and making a path to the object. During competitions it was mandatory to approach different objects and perform some actions, the only difference is to stopping the opponent robot. A function has been developed that receives the input from the parameters of object and stop the opponent player. For example, if a robot is ruining the ball towards the goalpost or he has to pass the ball to other teammate and someone is approaching from towards to take the ball, in this the robot has to analyze that someone is approaching from front and he has to pass the ball or make way from taking the ball from him. Now, the algorithm

IV. Strategy

Development of strategy for autonomous systems requires many tests. As a robot has a finite resources it is more comfortable and profitable to test algorithms in simulation. Moreover simulation allows integrate Reinforcement Learning algorithms for tasks solving. For these purposes program of robot's behavior simulation is used. It has a graphic interface and is able to reflect both decisions taken by a real robot and simulation. This helps students to develop complex logic systems and tune them up in their laptops.

A. Dribbling

In robosoccer competitions it is customary to kick the ball whenever possible. To make the game more dynamic and extend the time of ball control there was an attempt to introduce dribbling. First of all a movement to keep the ball in front of a robot and direct it towards goalposts was developed. Then it was required to integrate this movement in the game logics. Regions to dribble the ball and ones to keep the ball were chosen expertly.

B. Kick Strategy

In kick strategy the robot has to determine a kick direction before kicking a ball. Decisions depends upon many factors, it uses the knowledge of ball position in which direction at the field and direction of grass leaning on the field. In this leaning factor is most important factor. It can lean from one

goalpost to another goalpost. There are 2 types of kick strategy in robosoccer.



Kick-off strategy

In this, all the robots need staying in their own half-court. Before the ball was out, the robots not in the kick-off team must be three meters or more to the ball. One robot in the Kick-off team should maintain next to the ball, the robots not in the kick-off team must be two meters or more to the ball until the robots in the kick-off team go into the playing state. The ball stops in the center marks the point, all the robots are not allowed to touch the ball before the ball in the state of playing except the kick-off one. After the referee gives the "start" signal, a robot in kick-off side kicks the ball off. The kick-off robot can use the kicking machine to kick the ball and let it move more than 0.5 m freely, but it cannot drag the ball or dribble ball. Only the attacker, which is nearest to the ball, goes for the ball. When the ball was

kicked out of sports after the game was started immediately kick the other side only when the robot touches the ball after the goal to be effective.



Corner Kick Attacking Tactics

Corner Kick is also a good opportunity to score, when we are the offensive players and defensive players or goalkeeper move the ball out of the goal line, the referee will sentence us with a corner Kick. Assists players leave immediately after touch the ball to assort with the Offensive or block defensive robot off the line. Forward attack the ball quickly after kick-off. Obviously, for the attacking of free kick in the opposing half and corner kick, players need to make use of their attacking tactics to create temporary space and make use of it efficiently so as to attack actively.



V. Conclusion

Studying all the necessary disciplines for playing robosoccer its is a complex disciplinary game/task that has to been played with tactics and knowledge which requires interaction in several modern areas of science and technology.

This is a high-tech game and it has been creation for all educational field to learn in robotic field for primarily for universities. It has been sufficiently developed in the world for competition which is first represented by the Robocup competitions, Robosoccer should been studied more intensively in order to to bring robotics based on Artificial intelligence into real life faster.

It includes various types of systems that the robots has to been understand and work autonomously while playing the game.

VI. Acknowledgements

We wish to thank our colleagues at Pillai HOC college of engineering and technology for giving suggestions and stimulating the discussions. Thanks to the Robocup team Starkit team for giving lots of information and supporting us.

VII. References

1. Starkit team official site: starkit.ru
2. Robocup official site: robocup.org
3. Rhoban team official site: rhoban.com
4. SPQR team official site: spqr.dis.uniroma1.it
5. R.Emery, K.Siorski and T.Balch, Protocols for Collaboration,Coordination and Dynamic Role Assignment in a Robot Team. Proceedings of IEEE International Conference on Robotics&Automation; 2002, p.3008-3015.
6. Liu Zuo-shi; Chen Fei; Li Qiang. Two kinds of kinematic model of soccer robot in RoboCup mid-size league. Modern Manufacturing Engineering 2008; 1: 124-126.

Application of Artificial Intelligence Systems in the Process of Crew Training

Bubere Arfat , Ravindra Mhatre

Abstract— The article considers the problems of managing the method of integrated training of orbital space platform crews within the context of conversion to advanced digital smart technologies, computer-assisted training and computer science. The proposed approach relies on the employment of automated information systems to support the design and management of crew training on integrated and special-purpose simulators using a synthetic intelligence technology.

Keywords: *artificial intelligence, cosmonaut training, automatic information system, information support.*

INTRODUCTION

The application of advanced digital, smart technologies, robotic systems, new materials and style techniques, creation of huge processing systems, computer-aided learning and AI (AI) are relevant for various branches of science and technology, including manned space programs. Some technology concepts and pilot systems supported the AI (3-D computer vision, automated systems for planning and evaluating the activities of cosmonauts, inquiry and communications system) were developed within the industry over several decades [1, 2, 3].

Cosmonaut training is one among the foremost important problems with the manned flights program. Its last may be a comprehensive crew training, aimed toward forming necessary profession. The article considers the issues of managing the tactic of integrated training of orbital space station crews within the context of conversion to advanced digital smart technologies, computer-assisted training and applied science. The proposed approach relies on the utilization of automated information systems to support the planning and management of crew training on integrated and special-purpose simulators employing a synthetic intelligence technology. al qualifications of crew members for performing IVA operations in accordance with the objectives of this sort of coaching [4]. the excellent crew training because of the difference in technological features of the training process is currently divided into:

- comprehensive training of spacecraft crews (the Soyuz spacecraft);

- comprehensive training of orbital space laboratory crews (the ISS crews).

The training of the TMV crews is applied on integrated and special-purpose simulators of the spacecraft. The training of the TMV crews in Russia has traditionally been formed on the premise of the methodological, technological and organizational principles of the "Voskhod" and "Soyuz" spacecraft programs. The training of the TMV "Soyuz" crews relies on the practice of several basic flight operation modes (orbital injection, rendezvous, docking, descent and landing). the identical crew training technique was successfully accustomed train cosmonauts on the "Salyut" program, when an orbital piloted complex comprised one inhabited module with a limited number of onboard systems. (Fig.1).

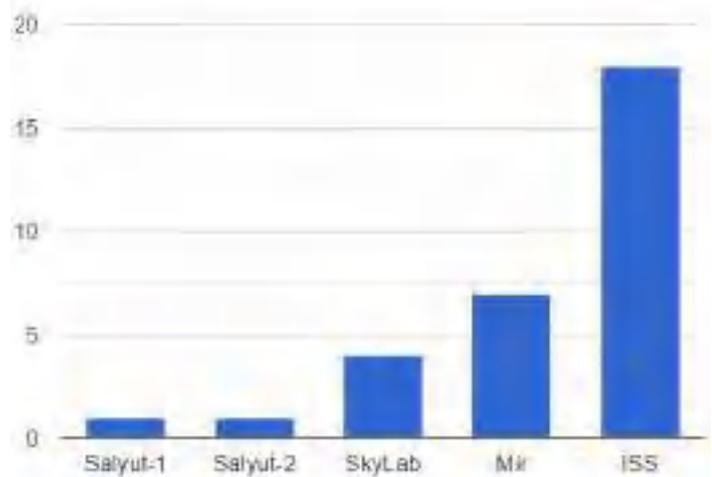


Fig. 1. Number of modules of orbital stations

- Author Arfat Aziz Bubere is currently pursuing masters in Information Tecchnology from University of Mumbai, India, PH-9773020605. E-mail: arfatbubere.ab@gmail.com

- Co-Author Ravindra Mhatre is currently Professor in University Of Mumbai, India, PH-9967719262.

The training experience of the “Mir” crews showed the requirement to boost the technique of the simulator training of space crews. This was because of the complex configuration and increased volume of the orbital station. the utilization of recent technologies and control systems during creation of the International artificial satellite led to the very fact that the amount of flight operations, applied aboard the ISS has increased 100- fold by 2010 as compared with the quantity of these, which were performed aboard the “Mir” orbital station.(Fig. 2) [5].

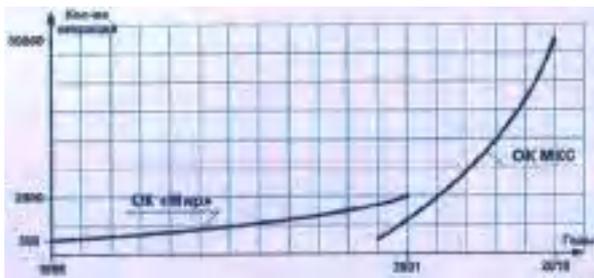


Fig. 2. Dynamics of the flight operations performed aboard the orbital stations

The current problems with the ISS crew training using the integrated and special-purpose simulators are characterized by:

- 1.increasing number of flight operations performed aboard the ISS; increase of complexity of the cosmonaut training; sophistication of fabric and technical facilities and communication system.
- 2.a significant increase of knowledge flows employed in the training process; the data rate exceeds the capacity of its processing by a choice maker (training instructor).
- 3.increasing number of varieties of training, training bases (Russia, USA, Europe, Japan, Canada) and as a consequence, rigid requirements to the training periods.
- 4.lack of your time for decision-making, and as a result, failure to require into consideration a chunk of useful information. This results in erroneous decisions within the process of coaching. Additional resources could also be needed to correct the results of erroneous decisions.

The process of the crew training includes planning, activity arrangement and performance control. so as to realize the goals of crew training, it's necessary to optimize the employment of appropriate resources.

This demonstrates the structure of a training session with the utilization of full task simulators.

The stepping up requirements for the fast decision-making has led to the necessity to develop a computer-assisted control within the study of cosmonaut training. thanks to the complexity of this sort of coaching, as a rule, the decision- making

process can not be formalized on the premise of one mathematical model. this stage of development of computing facilities allows to resolve this problem through the utilization of AI systems. during this particular case, the synthetic intelligence is known as a capability of automatic or automated systems to perform the functions of human intelligence, that is, to create optimal decisions on the premise of the gained experience and analysis of the external influences.

The artificial intelligence can be used primarily in planning and evaluating the quality of cosmonaut training .

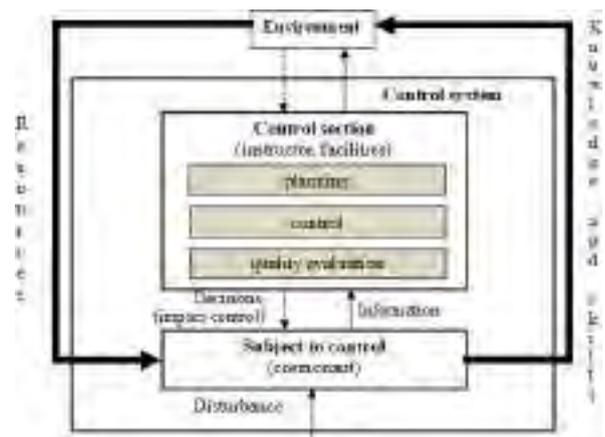


Fig. 3 Structure of the training process

In order to create the artificial intelligence to provide a simulator training of cosmonauts, it is necessary to:

- understand the goals, objectives, functions of the entire system of crew training with the use of integrated and special- purpose simulators;
- provide and use the necessary data at various levels of training, presented in the form of information flow diagrams;
- apply the classification and coding systems for flight operations and emergency situations;
- master the methodology of creating conceptual and information-logical models, which reflect the interconnection of information;
- create and store information arrays on data medium, taking into account that this will require modern software [3].

The process of the simulator training of the ISS crews can be divided into four interrelated components:

1. formation of a simulator training program, which represents a sequence of training sessions and exercises to be worked out by a crew.
2. compilation of a training sequence diagram.
3. control of crew training process

4. evaluation of cosmonauts' performance during a training session.

Automated control of the ISS crew training has resulted from the need to record large amounts of information (the number of flight operations aboard the ISS reaches several thousand) and to take into account all factors affecting the planning and control of training sessions.

An automated control of the training process can be based on the use of integrated automated training system (KAOSPK), which is the interaction of the simulator system itself and the automated information system intended for planning and conducting the training sessions (AISPTT). The integrated automated training system combines the simulator system (including instructors' terminals and crews' interfaces) with the terminals designed to train crews and instructors for each training session (Fig. 4). The use of such a system allows to reduce dataflow processing time by applying appropriate

software, interactive systems and databases. In this case, both the logical-analytical process of planning and control, and the preparation of documents necessary for conducting the training, become automated. Besides, all operations can be performed by an instructor from the same workstation.

The automated information systems involve the participation of humans and the use of technical means for data processing, but the major functions are carried out by the computer.

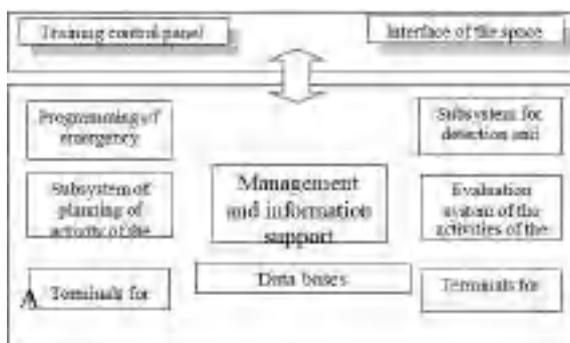


Fig 4. Scheme of the integrated automated training system

The integrated automated training systems should provide:

- descriptions of flight operations and scenarios of emergency situations;
- proximate order of flight operations and emergency situations;
- information support of cosmonauts and instructors upon request;
- reference information for the instructor on the

ongoing flight operations and emergency situations;

- automated compilation of training sequence diagrams;
- software to download the scenarios of emergency situations for the upcoming training sessions;
- emergency scenario training;
- detection of crew errors in the process of training;
- detection of simulator failures in the course of training, their correlation with the planned emergency situations;
- regulatory requirements for the crew's actions in nominal and off-nominal situations;
- options for changing the course of the training (training sequence diagram).

The indicated tasks can be combined into five groups (subsystem) according to their functional characteristics:

- management and information support subsystem;
- data input subsystem;
- detection and recognition subsystem;
- crew activity planning subsystem;
- crew performance rating subsystem.

The mode of interaction between the system's components while operational depends on the sort of the matter being solved.

Considering the widespread use of computing systems, they will be classified by: the realm of use; applicability of information; automated functions; degree of automation functions performed; degree of structuredness of this tasks. In cosmonaut training, AI-based automated systems will be classified, for instance, by purpose and naturally of the information presented

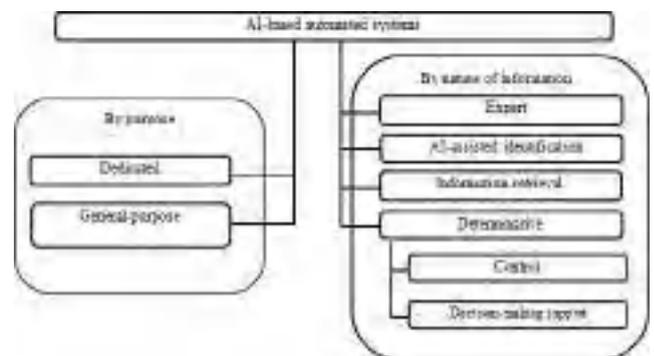


Fig. 5. Classification of automated training systems using artificial intelligence to train cosmonauts

By the nature of the information use AISPTT refers to a class of information-determinative systems, which carry out all information processing operations

according to a specific algorithm. Such systems can be classified according to the degree of impact of the resulting information on the decision-making process and divided into two classes of systems: control systems and decision-making support systems. AISPPT can be referred to control information systems. These systems produce information on the basis of which a person makes decisions. These systems are characterized by computationally intensive tasks and large data volumes. The development of AISPPT elements is carried out by using systems of visual object-oriented programming and object-oriented databases.

Thus, the automated control of comprehensive simulator training of orbital space station crews requires the synthesis of software and information support of modular AI systems in order to create integrated automated training systems. The proposed structure of AISPPT system requires the development of basic mathematical models, focused on solving issues within each subsystem. The goals of using the automated information systems are: improvement of quality of crew training; flight safety; reduction of instructors' labor contribution in planning and conducting simulator training. Further development of the integrated automated training system involves the use of artificial intelligence methods. This will make it possible to introduce a cosmonaut's individual training, taking into account his personality, to compile his "passport", which will contain data on his successes in mastering the course content, check test results, etc. Based on these data, the system will allow to generate additional tests and tasks for more successful mastering of the training documentation.

The introduction of AI technology will also contribute to the development of joint teamwork of cosmonauts, through the implementation of various scenarios on simulators, which can only be solved by a coordinated work of each cosmonaut in his workstation. This will require the development of dedicated software to create a user-friendly interface. But there's no doubt that the quality of training, and the volume of educational material will recoup the expenses connected with the development of AI technology as a part of the crew training system.

REFERENCES

- [1] Approaches to the Improvement of Information Support for the Comprehensive Cosmonaut Training of cosmonauts / A.A. Kuritsyn [and others] // Proceedings of VIII International Scientific and Practical Conference: Book of abstracts. – Star City, 2009. – pp.57–59.
- [2] Shcherbinin D.Yu. The use of information technology in the cosmonaut training: experience and prospects // International Conference VEonPC'2006 and III Scientific Conference SIM'2006. – Moscow.: IFTI.2006. – pp.155– 161.
- [3] Theoretical foundations of automated control / B.Ya. Sovetov., V.V. Tsekhanovskiy, V.D. Chertovskiy. – Moscow: Vysshaya Shkola., 2006. – 463 c. – ISBN 5-06- 005496-9.
- [4] Information Technologies of Training Cosmonauts for a Space Flight. A. Kuritsyn, Yu.V. Lonchakov, V.I. Yaropolov. 2017 International Workshop Engineering Technologies and Computer Science – 2017 IEEE DOI 10.1109/EnT.2017.24. Moscow, Russia. 978-1-5386- 3537-7/17.
- [5] Application of Information Technologies in the Course of Cosmonaut Training/ Kharlamov M.M., Kuritsyn A.A., Kovrigin S.N. // J. "Manned Space Flights", 2013, No. 1(6). ISSN 2226-7298.
- [6] New approaches to cosmonaut training of the program of scientific – applied research and experiments aboard the ISS Russian segment. Yuriy V.Lonchakov, Boris I.Kryuchkov, Andrey A.Kuritsyn, Valeriy A. Sivolap, Petr A.Saburov, Igor G.Sokhin, IAC Paper, IAC-15, B3,5,7x28425, 5 p. – ISSN 1995-625

RESTAURANT REVIEWS PREDICTION AND DEPLOY ON DJANGO

Bubere Arfat , Ravindra Mhatre

Abstract -Nowadays with the proliferation of the placement aware technologies and smart phones people tend to convey reviews for all kinds of products services and place them online . it's vital to extract knowledge or information of the occupiers within the vast amount of obtainable text reviews. For these, user's sentiment is additionally monumental and important . If any business owners want to require decision on future planning, they need to consider their clients sentiment and work thereon. during this research, we proposed a noble strategy to predict user sentiments from their online reviews given for a selected business by using supervised machine learning techniques. Our proposed machine learning model will provides a hand to restaurant owners to spot their customer's feedback and check positions. Online reviews can be helpful to utilized and make sustainable marketing strategies within the restaurant industry, which contributes to national sustainable economic development. This study, the most aspects (including food/taste, experience, location, and value) from 294,034 reviews on Yelp.com were extracted empirically using the Latent Dirichlet Allocation (LDA) and positive and negative sentiment were assigned to every extracted aspects. Positive sentiments were related to the food/taste, while negative sentiments were related to value. This study further proves a sturdy classification algorithm supported Support Vector Machine (SVM) with a Fuzzy Domain Ontology (FDO) algorithm outperforms other traditional classification algorithms like Naïve Bayes (NB) and SVM ontology in predicting the helpfulness of online reviews. This study enriches the literature on managerial aspects of sustainability by analyzing an outsized amount of plain text data that customers generated. The results of this study may be used as sustainable marketing strategy for review website developers to style sophisticated, intelligence review systems by enabling customers to sort and filter helpful reviews supported their preferences. The extracted aspects and their assigned sentiment could also help restaurateurs better understand a way to meet diverse customers' needs and maintain sustainable competitive advantages.

Key Words: MongoDB; Django;

INTRODUCTION

Many people holds the misunderstanding that sustainability relates only to the natural environment and overlook the importance of promoting sustainable economic development and creating sustainable business strategy. Metrics of sustainable economic development include, but don't seem to be limited to: Local economic process, local and tiny business growth, and value of living [1]. The restaurant industry play a crucial role in promoting sustainable economic development within the U.S. in line with National Restaurant Association [2], as of 2019, revenues generated by the restaurant industry within the U.S. was estimated at \$863 billion, accounting for 4% of the U.S. gross domestic product (GDP). Moreover, the restaurant industry employs approximately \$15.3 million people, which accounts for 10% of the general U.S. workforce. However, it's noticed that restaurants are struggling to survive because of variety of things like rising food prices, intense competitions, and high labor costs [3]. An earlier research study has demonstrated that around 60% of restaurants fail within three years [4]. Forbes [5] further reported that the restaurant failure rate is 30 percent within the primary year, and 30% of these that survive shutter within the following two operation years. As such, the

way to achieve sustained business performance becomes a critical issue for the restaurant industry [6]. on condition that the event of restaurant industry could reduce percent and promote local economic process, it's important to develop sustainable marketing strategies to drive restaurant performance. In today's digital era, eWOM (electronic Word-of-Mouth) have outweighed the normal marketing strategies regarding influence on customers' purchase decisions. in line with 2017 Restaurant Marketing Survey, it's reported that 94% of diners choose a restaurant supported online reviews. Restaurateurs agreed that online listing service is one amongst the foremost effective marketing channels for driving more businesses [7]. the number of online reviews is growing at an unprecedented rate. in line with Statistic, the quantity of reviews submitted to Yelp has reached \$148.3 million, which is over twice the number from 2014 [8]. Customers often feel overwhelmed while confronting the abundance of messages online [9]. it's [10] observed that customers tend to depend on a really limited number of reviews in making purchase decisions. Therefore, they'll resort to helpful reviews to realize a general idea of products or services. Ahelpful review is defined as "a peer-generated product evaluation that facilitates the consumer's purchase decision process" [11]. Presenting helpful reviews could help customers reduce the time and energy of searching relevant information from an outsized volume of online reviews [12]. it's also valuable for marketers to realize customers feedback to enhance their products or services. Thus, it's important for restaurant owners and marketers to know the way to make use of helpful online reviews to form their businesses stand out from competitors listed online. Numerous prior studies have identified factors influencing review helpfulness for both search goods and knowledge goods (e.g., restaurant, hotel) [11,13,14]. However, these factors are dominantly measured using Likert scales or numerical metrics

- Author Arfat Aziz Bubere is currently pursuing masters in Information Tecchnology from University of Mumbai, India, PH-9773020605. E-mail: arfatbubere.ab@gmail.com
- Co-Author Ravindra Mhatre is currently Professor in University Of Mumbai, India, PH-9967719262.

(e.g., review volume, star rating; sentence length) [11,15], neglecting more hidden semantic structures, like emotions and linguistic styles conveyed through a textual content. With digital texts increasing in size, variety of studies have applied machine learning (ML), which studies a computer's ability to find out from data without being explicitly programmed [16], to predict review helpfulness. However, prior scholars tend to predict review helpfulness within the hotel and e-commerce industry (e.g., Amazon.com), there's a comparatively small number of studies predicting review helpfulness within the restaurant industry. Also, few studies have made effort to propose an appropriate ML-based text mining technique to predict restaurant review helpfulness using both important dining aspects and emotional contents. to handle these literature gaps, this study aims to extract both emotions and most often mentioned dining aspects, then predict review helpfulness by comparing different ML-based text mining techniques. This study is guided by the subsequent three questions: (1) What dining aspects are most significant to customers? (2) What attitudes (positive or negative) are expressed regarding each dining aspect? (3) Considering restaurant aspects and their sentiment, which machine learning method performs better in predicting review helpfulness? it's expected that these study results could function a guidance for website developers to style better review systems to stay and achieve a sustainable competitive advantage over numerous online review websites. The extracted restaurant aspects and results from sentiment analysis could also help restaurateurs better understand the way to constantly improve customers' dining experiences. the remainder of this paper proceeds as follows. Section 2 illustrates related work on eWOM and review helpfulness prediction. The methods, including data collection and data analysis process are explained in Section 3. the most results are presented in Section 4. within the final section, the conclusion, a discussion on theoretical contributions and practical implications, limitations and future research are presented.

Django is a free open source python web framework which is employed to develop websites faster and easier. Django follows MTV(Model Template View) architecture pattern which makes it fast and straightforward for development and deployment. Now lets understand briefly what these three terms Model,Template and consider means:-

1. Model- Model represents shape of the info and business logic. It maintains the info of the applying. Model objects retrieve and store model state in an exceedingly database.
- 2.Template- The Template could be a presentation layer which handles computer program part completely.
3. View- View may be a programme. View display data using model to the user and also enables them to switch the info.

II. USING NATURAL LANGUAGE PROCESS

Any sentiment analysis workflow begins with loading data. But what does one do once the data's been loaded? you wish to process it through a natural language processing pipeline before you'll be able to do anything interesting with it. The necessary steps include (but aren't limited to) the following:

- 1.Tokenizing sentences to break text down into sentences, words, or other units
- 2.Removing stop words like "if," "but," "or," so on
- 3.Normalizing words by condensing all varieties of a word into one form
- 4.Vectorizing text by turning the text into a numerical

representation for consumption by your classifier All these steps serve to scale back the noise inherent in any human-readable text and improve the accuracy of classifier's results. There are plenty of great tools to assist with this, like Natural Language Toolkit, TextBlob, and spaCy. For this tutorial, you'll use spaCy.

III. USING MACHINE LEARNING CLASSIFIERS TO PREDICT SENTIMENT

There are variety of tools available in Python for solving classification problems. Here are a number of the more popular ones:

TensorFlow
PyTorch
scikit-learn

This list isn't all-inclusive, but these are the more widely used machine learning frameworks available in Python. They're large, powerful frameworks that take lots of your time to actually master and understand.

TensorFlow is developed by Google and is one in every of the foremost popular machine learning frameworks. you utilize it primarily to implement your own machine learning algorithms as critical using existing algorithms. It's fairly low-level, which provides the user plenty of power, but it comes with a steep learning curve.

PyTorch is Facebook's answer to TensorFlow and accomplishes many of the identical goals. However, it's built to be more familiar to Python programmers and has become a really popular framework in its claim. Because they need similar use cases, comparing TensorFlow and PyTorch is a useful exercise if you're considering learning a framework.

scikit-learn stands in contrast to TensorFlow and PyTorch. It's higher-level and allows you to use off-the-shelf machine learning algorithms instead of building your own. What it lacks in customizability, it quite makes up for in easy use, allowing you to quickly train classifiers in barely some lines of code. Luckily, spaCy provides a reasonably straightforward built-in text classifier that you'll study a touch later. First, however, it's important to know the overall workflow for any variety of classification problem.

IV.EVALUATING THE PROGRESS OF MODEL TRAINING

Since you'll be doing variety of evaluations, with many calculations for every one, it is sensible to put in writing a separate evaluate_model() function. during this function, you'll run the documents in your test set against the unfinished model to induce your model's predictions so compare them to the right labels of that data.

Using that information, you'll calculate the subsequent values:

- **True positives** are documents that your model correctly predicted as positive. For this project,this maps to the positive sentiment but generalizes in binary classification tasks to the category you're trying to spot.
- **False positives** are documents that your model incorrectly predicted as positive but were of course negative.
- **True negatives** are documents that your model correctly predicted as negative.
- **False negatives** are documents that your model incorrectly predicted as negative but were after all positive.

Because your model will return a score between 0 and 1 for every label, you'll determine a positive or negative result supported that score. From the four statistics described above, you'll calculate precision and recall, which are common measures of classification model performance:

1. **Precision** is the ratio of true positives to any or all items your model marked as positive (true and false positives). A precision of 1.0 implies that every review that your model marked as positive belongs to the positive class.
2. **Recall** is the ratio of true positives to all or any reviews that are actually positive, or the amount of true positives divided by the full number of true positives and false negatives.

V. Studies on Review Helpfulness Prediction

A number of online platforms have offered the mechanisms for other users to judge online reviews [32]. As an example, Amazon.com and TripAdvisor allow customers to vote for the reviews that are perceived as helpful in their decision-making process. The quantity of helpful votes could signal the standard of message contents [33]. Retail website developers could also increase their website traffic by presenting helpful reviews as a differentiation strategy [11]. Numerous previous scholars have identified the factors influencing "helpfulness" of online customer reviews. Supported a heuristic systematic model (HSM) explored factors influencing review helpfulness can be divided into two types: (1) Central route cues, which are related to review content features, like review content quality, review length, review readability, review types and review extremity [13,34,35]; and (2) peripheral cues, which are related to information source features, like reviewer expertise, reviewers' gender, reviewer reputation [36–38]. However, Hong et al. [39] noticed that the predictors of review helpfulness yield inconstant conclusion. They conducted a meta-analysis and located that review readability and review ratings failed to significantly influence review helpfulness. However, it's observed that the aforementioned determinants of review helpfulness are dominantly featured by numerical features [15]. In recent years, research on review helpfulness has focused on semantic features and linguistic style in online reviews [15,40]. Table 1 summarizes the studies on predicting review helpfulness on two popular business review sites (Yelp and TripAdvisor) during the amount from 2015 to 2019. Racherla and Friske [37] examined the impact of reviewer factors and review factors on perceived review usefulness across three kinds of categories. Among which, restaurants were chosen as experiential-based services. The results indicated that experiential-based services are a function of individual taste. Liu and Park [38] later extended the work of Racherla and Friske [37] by focusing exclusively on restaurant reviews. They added review readability and customer perceived enjoyment as two qualitative antecedents of review usefulness and identified these two variables because the most influential factors of review helpfulness. Ngo-Ye and Sinha [41] further conducted recency (the recency of purchase), frequency (frequency or total number of purchases), and also the average amount spent per transaction (monetary value) (RFM) analysis in predicting review helpfulness. Additionally to widely examined review and reviewer-specific characteristics, other factors influencing review helpfulness include review orders [42], emotions, and linguistic styles [15], temporal, exploratory and sensory cues [43]. Restaurants in the big apple City, Las Vegas, and LA are presumably to be selected as study samples [15,37,42]. Although recent studies have begun to examine the deeper structure and patterns of textual data, few studies have taken into

consideration both emotions and restaurant experience aspects in predicting restaurant review helpfulness

VI. REFERENCES

- 1) https://www.researchgate.net/publication/336061601_Predicting_the_Helpfulness_of_Online_Restaurant_Reviews_Using_Different_Machine_Learning_Algorithms_A_Case_Study_of_Yelp
- 2) https://www.researchgate.net/publication/322208456_Machine_Learning_Based_Class_Level_Prediction_of_Restaurant_Reviews
- 3) <https://realpython.com/sentiment-analysis-python/>
- 4) <https://github.com/rishabh706/McDonalds-Restaurants-Reviews-Sentimental-Analysis>
- 5) International Economic Development Council. Green Metrics: Common Measures of Sustainable Economic Development. 2017. Available online: https://www.iedonline.org/clientuploads/Downloads/edrp/IED_DC_Greenmetrics.pdf (accessed on 29 August 2019)
- 6) National Restaurant Association. Restaurant Industry Factbook. 2019. Available online: https://www.restaurant.org/Downloads/PDFs/Research/SOI/restaurant_industry_fact_sheet_2019.pdf (accessed on 26 July 2019)
- 7) Rafi, M.; Hassan, S.; Shaikh, M.S. Content-Based Text Categorization Using Wikitology. 2012. Available online: <https://arxiv.org/pdf/1208.3623.pdf> (accessed on 6 August 2019).
- 8) Qazi, A.; Syed KB, S.; Raj, R.G.; Cambria, E.; Tahir, M.; Alghazzawi, D. A concept-level approach to the analysis of online review helpfulness. *Comput. Hum. Behav.* 2016, 58, 75–81
- 9) Ali, F.; Kwak, K.S.; Kim, Y.G. Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification. *Appl. Soft Comput.* 2016, 47, 235–250.
- 10) Rampton, J. How Online Reviews Can Help Grow Your Small Business. *Forbes*. 31 May 2017. Available online: <https://www.forbes.com/sites/johnrampton/2017/05/31/how-online-reviews-can-help-grow-yoursmall-business/#7ecbc990737b> (accessed on 13 July 2019).
- 11) Mudambi, S.M.; Schuff, D. What makes a helpful online review? A study of customer reviews on amazon.com. *Society for Information Management and The Management Information Systems Research Center. MIS Q.* 2010, 34, 185–200.
- 12) Malhotra, N.K. Reflections on the information overload paradigm in consumer decision making. *J. Consum. Res.* 1984, 10, 436–440.

Disease Prediction using Symptoms

Pritam Borade, Prof. Ravindra Mhatre

Abstract- people face various diseases thanks to the environmental condition and their living habits. that the prediction of disease at earlier stage becomes important task. But the accurate prediction on the idea of symptoms becomes too difficult for doctor. the right prediction of disease is that the most challenging task. To overcome this problem data processing plays an important role to predict the disease. life science has great amount of knowledge growth annually. thanks to increase amount of knowledge growth in medical and healthcare field the accurate analysis on medical data which has been benefits from early patient care. With the assistance of disease data, data processing finds hidden pattern information within the huge amount of medical data. We proposed general disease prediction supported symptoms of the patient. For the disease prediction, we use K-Nearest Neighbor and Convolutional neural network ML algorithm for accurate prediction of disease. For disease prediction required disease symptoms dataset. during this general disease estimate the living ways of person and check up data consider for the exact prediction.

The accuracy of general disease prediction by using CNN is 84.5% which is more than KNN algorithm. And the time and the memory requirement is also more in KNN than CNN. After general disease prediction, this system able to give the risk associated with general disease which is lower risk of general disease or higher

Index Terms – Machine learning – KNN ,CNN , healthcare

1 INTRODUCTION

Artificial Intelligence made computer more intelligent and can enable the computer to things. Artificial intelligent study consider machine learning as subfield in numerous research work. Different analysts feel that without learning, insight can be created. There are numerous kinds of Machine Learning Techniques like Unsupervised, Semi Supervised, Supervised, Reinforcement, Evolutionary Learning and Deep Learning. These learnings are used to classify huge data very fastly. So we use K-Nearest Neighbor and Convolutional neural network ML algorithm for fast classification of big data and correct prediction of disease. Because medical data is cumulative day by day so usage of that for expecting correct disease is crucial job but processing big data is actual crucial in overall so data mining plays very significant role and organization of large dataset using ML becomes so easy. It is critical to comprehend the accurate diagnosis of patients by clinical examination and evaluation. For compelling determination decision support systems that depend on computer may assume an indispensable job. Health care field creates enormous information about clinical evaluation, report in regards to patient, cure, subsequent meet-ups, medicine and so forth. It is intricate to orchestrate appropriately.

Quality of the information suggestion has been influenced appreciations to unsuitable management of the knowledge .

Upgrade within the measure of data needs some legitimate thanks to focus and process information viably and professionally. One of the numerous machine learning applications is employed to construct such classifier that may separate the info supported their features. Data set is separated into two or over two classes. Such classifiers are used for medical data study and disease prediction.

A. Proposed System:

In our project the user dialogue are going to be a linear design that proceeds from symptom extraction, to symptom mapping, where it identifies the corresponding symptom, then diagnosis the patient with its disease and finds an appropriate doctor are remarked the patient by using google map, the user are identified by the login details which is stored within the database. In system dialogue design is represented using finite state graph. so as to realize an accurate diagnosis, the logic for state transitions are made, linguistic communication generation templates were used, and system initiative to the user and obtain responses from the user. Our agent has three main conversational phases: acquisition of basic information, symptom extraction, and diagnosis. Our website starts off by asking about the users userid and password for login then enters a loop of

symptom extraction states until it acquires sufficient information for a diagnosis. Once the disease is identified the web site checks the disease.and open then google maps that the user can find an appropriate doctor for him/her.

B. Blockdigram:

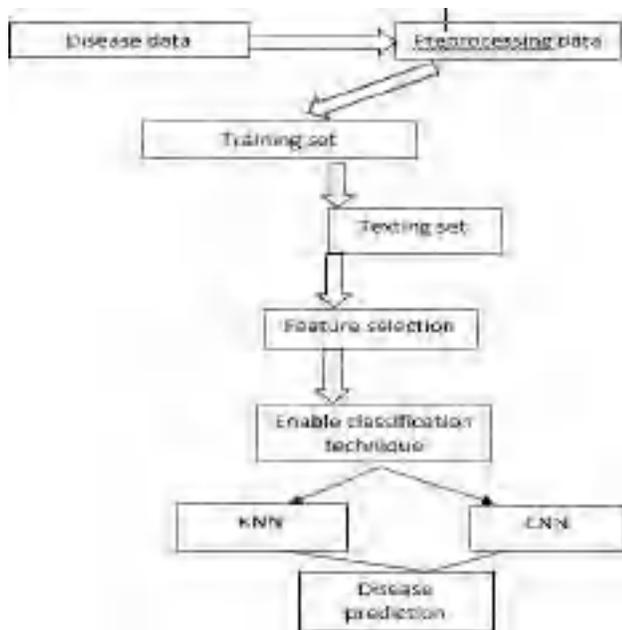


Fig .1

C. Methodology :

1. **Collection of the Dataset:** In this we collect the dataset using available resources. We've symptoms and disease dataset from github
2. **Data preprocessing:** In this we perform operations on the dataset so that it gives more generic results while training the model.
 1. **Creating KNN and CNN Architecture:** We settle on one architecture after an iterative process of training and evaluation for designing our model.
 2. **Compiling the model:** We compile the deep learning model using an appropriate loss function and optimizer and learning rate. We also set the hyperparameters and the parameters for the model training.

3. **Training the model and evaluation:** We train the model for a certain number of epochs. We also set check.points at the end of every epoch so that the best model is saved. After that we draw the accuracy loss curves to evaluate the performance.

4. **Saving the model and development of Flask application:** After this, we save the trained model and use it in our flask application so that the users also can use the model for predicting disease the patient suffers from..

D. Technologies used

- Machine Learning

ML is an application of artificial intelligence that provides system the ability to mechanically learn and improve from knowledges without being explicitly programmed. ML focuses on the development of computer programs that can access information and use it learn for themselves. The development of learning begins with observation or data, such as example, direct experiences, or teaching, in order to look for designs in data and make better decision in the future based on the examples that we provide. The main aim is to allow the computers learn mechanically without human or assistance and adjust action accordingly. The process of training and prediction involves the use of specialized algorithms. We feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. One such algorithm is KNN classification. It takes a test data, and finds k nearest data values to this data from test data set. Then it selects the neighbor of extreme frequency and gives its properties as the estimate result ML can be roughly divided into three categories:

- Supervised learning

The ML program is both given the input data and the consistent labelling. This means that the learn data has to be labelled by a humanoid being beforehand.

- Unsupervised learning

No labels are providing to the learning algorithm. The algorithm has to fig. out the a clustering of the input data.

- Reinforcement learning

A computer program animatedly interacts with its environment. This means that the program accepts positive and

negative response to improve its performance.

- Flask

Flask is a web application framework inscribed in Python. Armin Ronacher, who leads a global group of Python enthusiasts named Pocco, develops it. Flask is based on Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

Flask supports delays that can add application structures as if they were implemented in Flask itself. Delays exist for object-relational mappers, form validation, upload handling, various open authentication knowledges and several common basic related tools. Delays are updated far more than the core Flask program.

2. Summary:

We proposed a general disease prediction system based on machine learning algorithms. We utilized KNN and CNN algorithms to classify patient data because today medical data is growing very rapidly and that needs to process existing data for predicting exact diseases based on symptoms. We propose to compare the results between KNN and CNN algorithms in terms of accuracy and time. The analysis will be done using Diseases and Symptoms datasets available on Github.

3. Conclusion

The implementation of a healthcare system is based on machine learning algorithms. We have built a system that is designed to support patients and health professionals alike. A healthcare system will provide personalized diagnosis based on symptoms. The system will predict disease by accepting symptoms from the user. The system will also find a doctor for a particular disease. The usage of the system will be user-friendly and can be used by any person.

4. Future scope

The health care system project can be further improved in the future. Along with disease prediction, the system will also predict whether the patient has a major disease or a minor disease. If a patient has a minor disease, the system will give suggestions about medicine. And if the disease is major, then a Google Map will be open to search for nearby hospitals. Also, we will add information about

doctors so that if a patient has any doubts, they can directly contact with doctors.

5. References:

[1] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities", IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.

[2] B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," Springer Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

[3] IM. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, "Wearable 2.0: Enable human-cloud integration in next generation healthcare system," IEEE Commun., vol. 55, no. 1, pp. 54–61, Jan. 2017.

[4] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "HealthCPS: Healthcare cyberphysical system assisted by cloud and big data," IEEE Syst. J., vol. 11, no. 1, pp. 88–95, Mar. 2017.

[5] L. Qiu, K. Gai, and M. Qiu, "Optimal big data sharing approach for telehealth in cloud computing," in Proc. IEEE Int. Conf. Abstract

Block Chain Based Proxy Re-Encryption Scheme For Secure IoT Data Sharing

Name – Sahil Shailendra Mishra - MSC IT (Part 1)

**Paper Guide – Mr. Ravindra Mhatre
Asst. Professor, Department of IT/CS
Pillai HOC College of Arts , Science and
Commerce , Rasayani**

Abstract— Data is central to the web of Things (IoT) system. Most of the present IoT systems are unit victimisation centralized cloud-based knowledge sharing systems. Involvement of such third-party service supplier needs additionally trust from each device owner and device knowledge user. Moreover, the fees got to be procured their services. To tackle each the quantifiability and trust problems and to automatize the payments, this paper presents a blockchain based mostly proxy re-encryption theme. The system stores the IoT knowledge in a distributed cloud when cryptography. To share the collected IoT knowledge, the system establishes runtime dynamic good contracts between the device and also the knowledge user while not the involvement of a trusty third party. It additionally uses AN economical proxy re-encryption theme that permits that the info is just visible by the owner and also the person gift within the good contract. The projected system is enforced in AN Ethereum based mostly testbed to research the performance and security properties.

Index Terms—Proxy Re-Encryption, Blockchain, Smart Contracts, IoT knowledge Sharing, Security, Ethereum

I. INTRODUCTION

The Internet of Things (IoT) is AN rising technology that has nice technical, social, and economic significance. Current predictions for the impact of IoT are unit terribly spectacular. With the event of 5G, it's anticipating that a hundred billion connected IoT devices are employed by 2025 [1], [2]. It also will have a world economic impact of a lot of than \$11 trillion [3] [4].

Data is central to the IoT paradigm. IoT knowledge is collected to serve many various sorts of applications like good home, smart city, wearable, healthcare, smart grid, autonomous vehicles, good farms, industries and producing, and retail sector [4]—[6]. Therefore,

various heterogeneous sensors exist to live a spread of parameters. The collected knowledge from these IoT sensors may be helpful for various stakeholders. For example, air quality measurements are unit of interest to governmental organizations, application developers and inhabitants of the relevant areas. However, several challenges arise once organizing this knowledge sharing as these IoT devices, that are unit generally resource-constrained, need economical mechanisms to ensure the info integrity and to alter correct process and security [7]. Thanks to the massive variety of IoT devices, ascensible preparation, and maintenance prices [5] ought to even be taken under consideration. Currently, the majority of the device systems transfer the info to a centralized cloud and share the device data with totally different stakeholders, WHO prove access to the cloud storage [8]. The sensors get services from the third-party cloud service supplier, like access management additionally to the info storage. In this case, each device and device knowledge user ought to trust the third-party service supplier and additionally got to pay some fee for his or her services. Additionally, it's required to ascertain AN agreement between the third-party service supplier and device knowledge user. Most of those agreements are unit static and take various time and administration to be established [9]. It'll result during a important increase of your time before the particular knowledge sharing may be accomplished [10]. Thus, the present centralized design model in IoT systems can struggle to rescale to satisfy the stress of future IoT systems.

Our Contribution: to resolve these problems, we tend to propose a completely unique blockchain primarily based theme together with a proxy re-encryption mechanism to confirm the confidentiality of the information. Here, the advantage of mistreatment blockchain mechanisms to sell the detector measurements with completely different users is that the corresponding money transactions square measure mechanically managed through the in agreement good

contract, keep at the blockchain. Moreover, the supply and alternative quality of service needs from the legal contract between each parties are often mechanically applied. Consequently, compared to the business state of affairs wherever the information is keep in an exceedingly cloud- primarily based infrastructure, there's no want for manual verification of the payments and therefore the predefined needs. Also, disputes on these aspects square measure utterly avoided. The remainder of this paper has the subsequent structure. Section II provides an outline of connected work. The planned design and proxy re-encryption theme is explained in Section III and IV. Section V discusses the implementation of the planned theme. The performance analysis results square measure conferred in Section VI. Finally, Section VII presents our conclusions

II. connected WORK

There exist completely different studies on the protection and privacy of the IoT [11]—[16] and therefore the overwhelming majority of this analysis work is on understanding and distinctive these threats [17]—[21]. what is more use of blockchain to secure numerous IoT Platforms were mentioned in [22]—[26]. The IoT devices sense, gather, and share an outsized quantity of information, therefore gap up important security and privacy considerations. Khan and worship [27] in their paper have reviewed totally different security challenges to IoT and known insecure transferring of IoT information as a high-level security risk. Authors in [28] incontestable the shortage of basic security by hacking off-the-rack good home IoT devices.

In 1998, Blaze, Bleumer, and Strauss [29] at the start introduced the construct of proxy re-encryption and made the primary biface proxy re-encryption application. Authors in [30], [31] conjointly propose the same theme however it's not dynamic, therefore creating it unsuitable for cloud information sharing. In [32], a awfully economical answer for information storage within the cloud is planned employing a pairing free proxy re-encryption theme. However, the theme isn't enforced in follow. though the underlying structure of our planned theme relies on that, some necessary modification just like the inclusion of data is enclosed to confirm a sensible usage of the theme. Most of the previous work part addresses the matter of securely sharing the IoT information. it's nearly not possible to come back up with device-embedded security to unravel all the protection threats to the IoT devices. restricted computing and power resources of IoT conjointly build the execution of advanced security algorithms more durable on the device. we tend to

propose exploitation the mix of a blockchain and a pairing free proxy re-encryption theme to supply a commerce platform and to confirm secure transfer of the device information to the user

III. planned design

In this section, we tend to gift our new design supported the mechanisms of blockchain and re-encryption for secure storing and sharing of the device information. we tend to take into account four entities within the system: IoT sensors, information requester, cloud supplier, and also the blockchain, as shown in Figure.1

- 1) The sensors' owner activates the sensors, and registers them on the blockchain via a wise contract operate.
- 2) After triple-crown registration, the sensing elements' owner provides the sensor with the desired key material specified the measured knowledge will be sent encrypted to the cloud storage server.
- 3) A user requests access to 1 (or a gaggle off sensor(s) of the owner via the good contract perform.
- 4) After receiving the request, the sensors' owner associate degreed re- someone come back to an agreement, a wise contract is generated and deep-mined on the blockchain. The requester interacts with the blockchain to share the general public scientific discipline key and manages all the money associated transactions.
- 5) On receiving the user request, cloud storage is notified by the blockchain. The computer code then filters the information in keeping with the request.
- 6) Re-encryption scientific discipline key from the detector owner is updated on the good contract once the user request is received.
- 7) Cloud server decrypts and re-encrypts filtered knowledge, before storing it once more on a short lived location onto the cloud server.
- 8) The encrypted knowledge is temporally hold on on the server and a dealing containing the address of the hold on knowledge is deep-mined on the blockchain.
- 9) When the information is prepared, the requester is notified of the temporary location by the blockchain. The requester will decipher the information victimization its personal scientific discipline key.

We configured and connected all the devices t the internet we used the auto- discovery protocols of geth to connect the miners and the full nodes, and configured google firebase.

A. Miners

The proposed system consists of three miners that generate a block of transactions on average every 13 seconds. These miners are running on a virtual machine

with the same hardware capabilities. All the mining devices were configured to use one Ethereum wallet that collects the mining reward. These miners are running on Geth v1.80 [34] with four mining threads each.

B. Smart Contracts

We developed two smart contracts' on truffle [35] and compiled them with Solidity 0.4.24 [36]. The first smart contract consists of the functions to register the sensor, request data, and financial functions. The second smart contract is dynamically created in the runtime when the user requests for the data.

C. IOT Sensors

Each sensor TI Sensortag CC2650 connects to a Raspberry Pi 3 Model B (RSP) through Bluetooth Low Energy, as shown in Figure 2. This RSP manages the sensor and the Ethereum account to perform transactions on the blockchain on behalf of sensors. A sensor application is developed in Python.

C. User Application

A customized application is designed as the user interface in Python 2.7.12, running on a Raspberry Pi 3 attached to a touchscreen. This application uses JSON-RPC to get the sensors' information from the blockchain. After selecting the required sensor, the user enters details for specifying the data requirements. We deploy a new smart contract on the blockchain in run-time based on the user-selected options for the requested data (e.g. Sensor selection, Price). This application keeps track of the Ethereum wallet along with ECC [37] secret key of the data requester. The application downloads the data from the cloud server, checks for the signature and integrity, and decrypts the requested data.

D. Cloud Storage Server

The cloud storage server consists of the RSP and the Google Firebase. RSP acts as ethereum full node and connects to the blockchain, while Google Firebase is used for the storage of the data. The authentication and integrity of the data are performed on the RSP and encrypted sensor data along with the meta-data is upload to the Google Firebase in JSON format. This cloud also performs proxy re-encryption and updates the smart contract variable for data address sharing.

V. I. PERFORMANCE ANALYSIS

In this section, we describe the experiments to evaluate the proof of concept implementation. Experiments were designed to study the performance of the framework. We have performed multiple experiments to test the impact of proxy re-encryption on the overall system and performed some scalability tests.

. CONCLUSIONS

In this paper, we have proposed a blockchain based trading platform with the combination of a pairing free proxy re- encryption scheme to ensure secure transfer of the sensor data to the user. We have also validated the proof of concept model on a private Ethereum testbed and demonstrated the practicality of the system design using off-the-shelf laptops and raspberry pis.

In the future, we plan to extend the proposed system with an implementation on a different blockchain platform e.g. Hyperledger. We also plan to extend our architecture by adding a distributed cloud storage to make the system more scalable.

REFERENCES

1. Taylor, D. Baron, and D. Schmidt, "The world in 2025-predictions for the next ten years," in *Micro.s y.stem.s, Packagiri g, A.s.sembl y arid Circuit.s Techriolog y Conference (IMPACT), 2015 10th International*. IEEE, 2015, pp. 192-195.
2. Liyanage, I. Ahmad, A. B. Abro, A. Gurtov, and M. Ylianttila, *A Comfpreheri.five Guide to SG Securit y*. John Wiley & Sons, 2018.
3. Suryadevara and S. Mukhopadhyay, "Internet of things: A review and future perspective," *Reliance*, 2018.
4. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on multi-access edge computing for internet of things realization," *IEEE Communication.s Survey.s & Tutorial.s*, vol. 20, no. 4, pp. 2961-2991, 2018.
5. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and
6. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Commwticaciori.s Survey.s & Tutorial.s*, vol. 17, no. 4, pp. 2347-2376, 2015.
7. Rajakaruna, P. Porambage, A. Manzoor, M. Liyanage, A. Gurtov, and
8. Ylianttila, "Enabling end-to-end secure connectivity for low-power iot devices with uavs," in *rat 2nd Work.shof> ari Intelligent*