REVIEW ON ELECTION PREDICTION USING MACHINE LEARNING TECHNIQUE

Yashashree Belkhede[1], Dr.Praveen Shetiye[2], Dr.Avinash K. Gulve[3]

[1]Department of Computer science and Engineering
Government College of Engineering Aurangabad, Maharashtra, INDIA
(yashashreebelkhede@gmail.com)

[2]Department of Computer science and Engineering
Government College of Engineering Aurangabad, Maharashtra, INDIA
(praveen.shetiye@gmail.com)

[3]Department of Computer science and Engineering
Government College of Engineering Aurangabad, Maharashtra, INDIA
(Akgulve@gmail.com)

Abstract── **Right before peoples morning get started with the newspaper in hand .But nowadays the scenario is bit different. Social media is essential because it will take us to deal with locally, nationally and also internationally. Peoples are now connected with the social networks. Peoples are using social sites for a each single step of their daily life .what social media exactly means ,which helps human in knowing that what happens in the corner of the world .the social media flaunting real as well as sometimes fake news from which the human emotions get affected. Politics is the thought provoking topic .Elections are the most awaited moment for the people .people are very much interested in, who is going to win the elections or who will be the ruling party to rule .This paper is likely to predict the election outcomes using twitter or facebook data .we will collecting tweets from the twitter and then perform pre-processing on that obtained data. Aspect and topics are obtained by using LDA.**

Keywords── **Election, LDA, MLT, predict, Social media, twitter, twitter data.**

I. INTRODUCTION

Social influence plays important part in a human life. Social influence means when a person comments on some social tweet. If that person reacts on it whether it is positive or negative which may change the situations [1]. Social media connect people all over the world with in few second. In this project by using twitter data or facebook data. We are trying to figure out the election outcome. When a people reacted on something which calculate the polarity of the situation in favor or in oppose.

In this paper we are focusing on the Indian election prediction. For example, Consider any Indian political party and the various users that is active user and inactive user[1] .when any party related news is there on the social media and the people will start reacting on it. These reactions may be positive, neutral and negative. Using sampling technique we are going to obtain various topics and aspects.
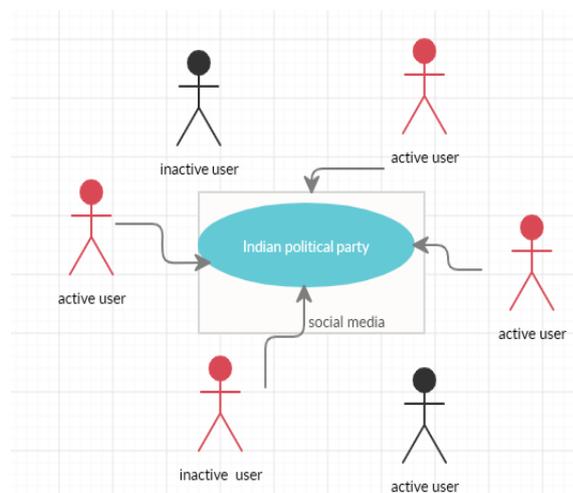

Fig. 1 Local example of social influence

II.RELATED WORK

[1]. Jiezhongqiu proposed system on Social Influence Prediction with Deep learning, here in this paper described the study of the social influence using deep learning technique and obtained the final result in the graphical format. This framework basically worked on the OAG, Digg, twitter, weibo. The main idea behind the paper is to work on the local network .Using sampling techniques and the learning techniques simultaneously

[2]. Parul Sharma and Teng-Sheng Moh research on Sentiment Analysis on Hindi Twitter data, in this paper using Twitter Archiver tool obtained the Hindi language tweets. Then completed the data mining on 42,235 tweets. By using Naïve Baye's, Dictionary Approach and Support Vector machine algorithms tried to calculate the result of the general elections. Among the three algorithms the accuracy of the Support vector Machine is higher. So from that the prediction is clear that which party had chances of winning. They have also computed the precision and recall.

[3]. Steven T. Smith studied on Influence Estimation on Social Media Networks Using Causal Inference this research is about the social media influence estimation. During French presidential elections 2017 find out the influence by using causal inference. Context are used for the graph sampling, Graph Filtering and for Graph influence

[4]. Quanzeng You Approach to Social Prediction of elections. The approach of this paper is the connection between image and the social media. Competitive Vector Auto-Regression model joins an image and the text in multimedia network. The proposed system is simply about when an election candidate post image on the social media and the viewer comments on that candidate related image. CVAR model calculate the accurate prediction of the elections.

[5].Jason Anastasopoulos studied on Political image analysis with deep neural network according to this proposed system posting text and images on social media is like daily routine though Political Candidate can interact with the voters directly. Images are the medium of communication between the candidate and the voters. Deep learning, neural network, machine learning and image analysis methodology are used in this proposed system.

In studies we found that what methodologies or techniques used early to achieve specific target. We required high power system to perform this project related work or we can use cloud instances which gives better power for processing and storage space for data. You can develop and use your own cloud and perform NLP and MTL operation on that. [9],[18].

III.PROPOSED SYSTEM

In the above model, the social media i.e. twitter or facebook is the centre of the system. There were the number of the user are handling many account on the social media, among them twitter and facebook are the trending social media accounts. As any Indian political party tweets on the twitter, user starts reacting on that topic whether they are agree or not or they have any other opinion on it. Firstly from the tweets, we are going to remove the unnecessary words. Then correctly identify the meaning of the word from the given tweet called as lemmatization process.
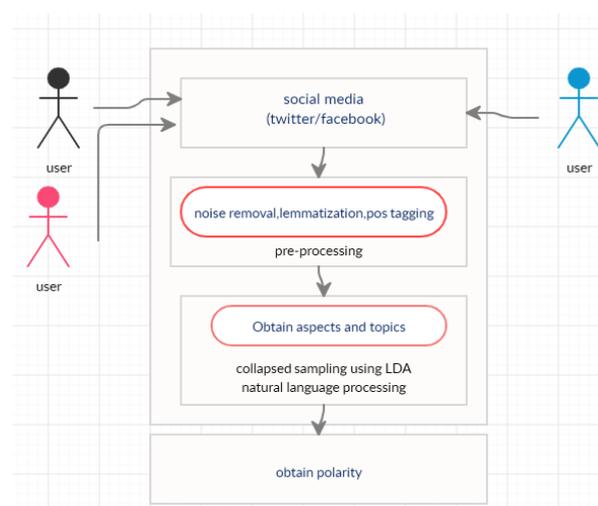

Fig. 2 Proposed System

This example of part of speech tagging shows the exact working of the process. It will separate all the part of speech as pronoun, verb, adjective, article; noun etc.[16] This is called as the part of speech tagging
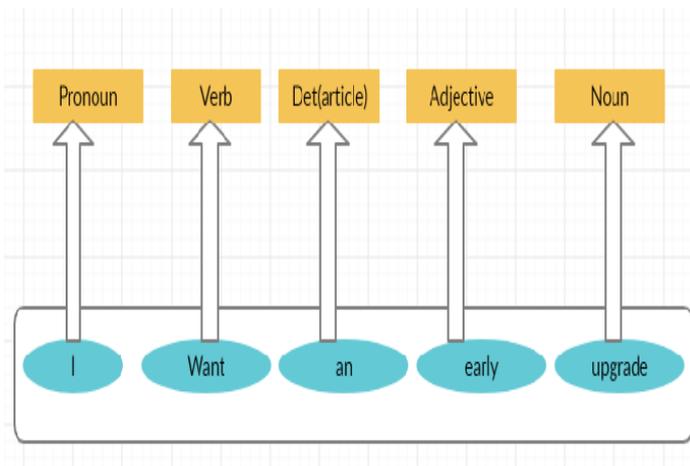

Fig. 3 Example of Parts of speech tagging

Collapsed sampling using Latent Dirichlet Allocation (LDA).The below diagram is the o diagram of the LDA system. This diagram gives detailed information about the proposed system. This diagram clarifies the actual working of the system [15]. Here is the actual working of the system is that firstly how the topics and aspects are obtain from the twitter data.Natural language processing helps to extract the main topic from the twitter data set with the help of the Latent dirichlet allocation (LDA)
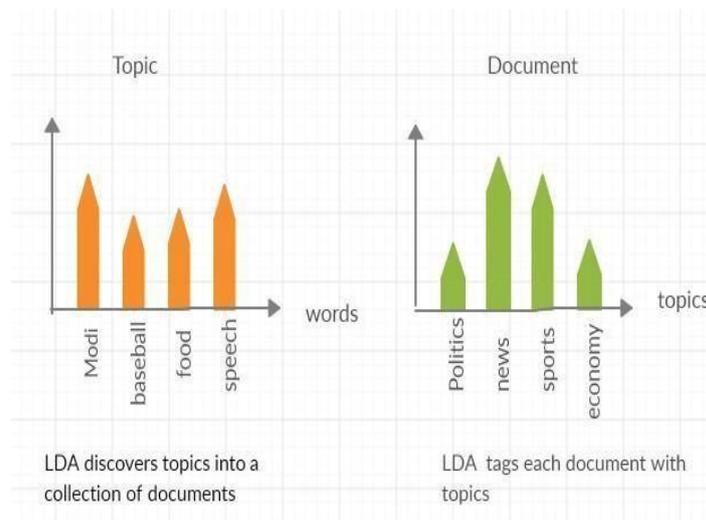

Fig. 4 latent Dirchlet Allocation (LDA)

 Then polarity is obtain from the topics and aspects .from the obtained polarity it is clear that the result of the elections prediction

IV.COMPARISON OF METHODOLOGIES

Sr.No	Parameter	Natural language Processing(NLP)	Machine Learning Technique(MLT)	Decision Tree Learning(DTL)
I	Parent	1-Computer science 2-Artificial Intelligence 3-Information Engineering 4-Linguistics	1-Artificial Intelligence	1- DataMining 2- MachineLearning
II	Purpose	Learn Specific pattern from huge data	Perform Intelligent Things/automatically learn without any help of programmer	Based on different inputs and obtained the desired output
III	Technical Algorithm	1- Hidden MarkorvModel 2- Continuous bag ofwords 3-N gramAlgorithm	1-Supervised i)Support Vector Machine ii) Naïve Bayes 2-Unsupervised i)Neural network ii)Hidden Markov iii)ICA,PCA	1-C 4.5 Algorithm 2-ID3 Algorithm 3-CART Algorithm
IV	Used For	NLP basically used to understanding the language	Machine learning working on Learning things from Experience	Decision tree learning basically working on classification
V	Processing Time	Based on the algorithm data and structure of the program .It takes less time to perform operation	Based on the supervised and Unsupervised .On On an average it takes less time in unsupervised learning but in supervised it consume more time to perform operation	Based on the algorithm of size of data, it takes more time compared to others
VI	Learning Frequency	We have to feed first and solution may or may not be useful to other problem	Self-learning and learn each and every cycle of the solution which is further useful	We have to feed first and part of solution is useful
VII	Working On	To understand the structure and meaning of the complex problem	To learn the data complex problem by observation or by experience	Observed the data or the specific data of complex problem
VIII	Problem Solving Capacity	Solve complex problem based on the data size	Solve complex problem based on technique. Once setup then large dataset also process in less time	Solve problem based on data representation in the form of tree
IX	Application	1- Informationretrieval 2- Sentiment analysis 3-Information	1-sentiment analysis 2-fraud prediction 3-election prediction	1-Rapid miner 2-matlab

Table 1: Comparative study

Time required to process data? What is their learning frequency and problem solving capacity? This comparison shows why we combine two Methodologies Natural language processing and Machine Learning Technique to get better result.[8],[10],[11],[12],[13],[14],[15],[16],[17].

If we see the comparisons we found that the natural language processing is better to understand the language but every time we not feed data to the NLP. We required the machine that learn or understand the data by itself and those parts comes under unsupervised machine learning. It gives better output, better execution time and accurate results rather than single methodology. This combination of methodologies handles the complex problem very easily and takes less time to perform operation. Decision tree learning is not enough to fulfill requirements of this project.

This part of the paper is focused on the different methodologies used for Sentiment Analysis. In this section we check the methodologies capabilities in different parameters like what is the purpose of the particular Methodology? What algorithm they used? How much time they required?

V.CONCLUSION

In this paper, we saw that how social media influence the people to take their decision and also saw the local example of social influence.After seen some related work we decided our proposed system which can be modify as per the requirement of this project. The important part of this project is which methodology we are using for election prediction. There are many methodologies present in the market for prediction or for sentiment analysis but only few of them work efficiently and gives the accurate results. Such methodologies are NLP, MLT and DTL. NLP and MLT are the mostly used methodologies in modern era.

But with good methodologies of combination there are some limitations like NLP and MLT not perform well when data is wrong or incomplete also if data is not properly arrange.

VI.FUTURE SCOPE

In Future work, we try to solve this problem by using better algorithms or techniques and implement stable and better sentimental analysis model which is useful in real time environment for prediction purpose. If this model works fine then it is helpful for small scale industries and private and government organizations too.

REFERENCES

[1]JiezhongQiu,JianTang,HaoMa,YuxiaoDong,Kuansan Wang and Jie Tang,"Deeplnf:Social Influence Prediction with Deep Learning",The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data mining,August19-23,2018,London,United Kingdom.ACM,NewYork,NY,USA

[2]Mark Carmen,markKoerber,JiuyongLi,Kim-Kwang Raymond Choo and Helen Ashman,"Manupulating Visibility Of Political and Apolitical Threads On Reddit Via Score Boosting", 2018 17th IEEE International Conference On Trust,Security and Privacy in Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering.

[3]Steven T. Smith,Edward K. Kao,Danelle C. Shah,OlgaSimek and Donald B. Rubin,"Influence Estimation On Social Media Networks Using Inference",2018 IEEE Workshop on Statistical Signal Processing(SSP)

[4]ZhenhuaTan,(Member,IEEE),DankeWu,GuangmingYang,And Zhang Bin,"APMSID:Activated Probability for Multi-Score Information Diffusion in Online Social Networks,Volume6,2018"

[5]QuanzenYou,LiangliangCao,YangCong,Member,IEEE,XianchaoZhang,andJieboLuo,Fellow,IEEE,"A Multifaceted Approach to Social Multimedia-based Prediction of Elections ",IEEE Transactions on Multimedia (Volume: 17 Issue:12, Dec. 2015)

[6]Parul Sharma and Teng-Sheng Moh,"Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter",2016 IEEE International Conference On Big Data (BigData)

[7]RohiniAshokPandit,Dr.GulveAvinashKushal,Dr.PraveenShetiye,"Implementing Feature Selection Technique to estimate influence from voting records",International Journal of Research and Analytical Reviews(IJRAR) 2019 IJRAR June 2019,Volume 6,Issue 2

[8]InternetLink:https://en.m.wikipedia.org/wiki/Natural_language_processing

[9]InternetLink:https://en.m.wikipedia.org/wiki/Machine_learning

[10]InternetLink:https://www.quora.com/Whats-the-difference-between-machine-learning-AI-and-NLP

[11]Rekha S. Thorat,Dr. Praveen C. Shetiye,Dr. Avinash K. Gulve,"Prediction Using Satellite Imagery And Machine Learning ",International Journal of Scientific And Technology Research Volume 9,Issue2,FEB2020

[12]InternetLink:https://rutumulkar.com/blog/2016/NLP-ML/

[13]InternetLink:https://www.google.com/search?q=latent+dirichlet+allocation+explained&sxsrf=ALeKk00eOH0epoqKh37vw XnWdyXwNs0VQQ:1583255113047&tbm=isch&source=iu&ictx=1&fir=G0tsIyD6o71fM%253A%252CvNBqmfQ2m8O4vM %252C%252Fm%252F0cc726&vet=1&usg=AI4_kTeCAul3I7dOcz0STMpU6tdbajiLQ&sa=X&ved=2ahUKEwicOf05P7nAh Vf7HMBHa8bCEwQ_B0wHHoECAsQAw#imgrc=G0tsIyD6o-71fM

[14]InternetLink:https://www.google.com/search?q=part+of+speech+tagging&sxsrf=ALeKk01uMmVBoIvn0hn6b4s2tymdbc1 kAA:1583255227189&tbm=isch&source=iu&ictx=1&fir=Km4FsjGqh56Y6M%253A%252C1EX4mmNBzdWDSM%252C%2 52Fm%252F037xvs&vet=1&usg=AI4_kSCIrquHcKPJYcwQ6mpsRP_DOrB8w&sa=X&ved=2ahUKEwjf5J6r5f7nAhUU6nM BHTDcBZEQ_B0wGXoECAsQAw#imgrc=Km4FsjGqh56Y6M:

[15]SupriyaB.Moralwar ,Sachin N.Deshmukh,"Different Approches of Sentiment Analysis",JCSE International Journal of Computer Science AndEngineering,Volume-3,Issue-3

[16]NikhilWagh,VikulPawarandKailashKharat,"Implementation of stable Private Cloud using Openstack with Virtual Machine Results",International Journal of Computer Science And Engineering & Technology(IJCET-19) Volume 10,Issue 2,and March-April2019