Election result prediction using Sentiment Analysis

Madhura Shirodkar[1], Sayali Nimbalkar[2], Ashlesha Ingole[3], Pragati Mishra[4], Sandeep Sahu[5]

[1]*Professor, Dept. of Electronics and Telecommunication, Xavier Institute of Engineering, Mumbai, Maharashtra, India*

[2,3,4,5]*Student, Dept. of Electronics and Telecommunication, Xavier Institute of Engineering, Mumbai, Maharashtra, India*

---***---

Abstract - *In the recent year, social media has provided end users a powerful platform to voice their opinions. Opinion of people matters a lot to analyze how the propagation of information impacts the lives in a large-scale network like Twitter. Businesses need to identify the polarity of these opinions in order to understand user orientation and thereby make smarter decisions. One such application is in the field of politics, where political entities need to understand public opinion to determine their campaigning strategy. Twitter is indeed used extensively for political deliberation. We find that the mere number of messages mentioning a party reflects the election result. This data is used to predict outcome of election by using sentiment analysis. Sentiment analysis of the tweets determine the polarity and inclination of vast population towards specific topic, item or entity. Popular text classification algorithms like Naive Bayes and SVM are Supervised Learning Algorithms which require a training data set to perform Sentiment analysis. These algorithms are utilized to build classifier and classified the test data as positive, negative and neutral. A two stage framework can be formed to create a training data from the mined Twitter data and to propose a scalable machine learning model to predict the election result.*

Key Words: Social Media, Twitter, Politics, Sentiment Analysis, Naive Bayes, Support Vector Machine

1. INTRODUCTION

An election is a most important part in the democracy. It is the instrument of democracy where the voters communicate with the representatives. Due to their important role in politics, there has been a big interest in predicting an election outcome. Traditional polls are too costly and still accuracy won't be achieved. To overcome this problem social media has been used as it is easy and freely available. Social media sites have become valuable sources for opinion mining because people post everything right from the details of their life to the products and services they use, to give opinions about the current issues such as political and religious views. Millions of messages are being posted every day on popular social media sites such as Twitter, Instagram and Facebook. Twitter is an online social networking service that enables users to send and read short 240-character messages called "tweets". Along with the short messages, users can use the hashtags before a relevant keyword or phrase in their Tweet to categorize those Tweets and help them show more easily in Twitter Search. The use of hash tags makes the problem of text classification relatively easier since the hash tag itself can convey an emotion or opinion. Currently around 6500 tweets are published per second, which results in approximately 561.6 million tweets per day. Facebook is the most popular and well known Social Networking Service (SNS) all over the world. According to reports 86.3 percent people reported using the Internet in the previous six months. So Facebook is also one of the good sources among various social media's available to be considered as a data source. In June 2016, the number of active Instagram users, those who use the application on a weekly basis, surpassed the number of active users on Twitter reaching a total of more than 500 million, of these some 300 million use their accounts at least once every day. This total has reached even greater heights in 2017 with the social media company boasting a total of 700 million active users, making Instagram the second most widely used social media platform after Facebook. Since Instagram attracts in surplus of 700 million users worldwide, allowing them to share and promote material at little cost, it would seem rational for political parties to want to take advantage of this communication tool. This is an interesting research area that combines politics and social media which both concern today's society.

2. LITERATURE REVIEW

This section summarizes some of the scholarly articles and research works in the field of Machine Learning and data mining to analyze sentiments on Twitter and preparing prediction model for various applications. The use of social media in last few decades have been helpful in determining people's attitude with respect to specific topics or events, a wide research interest in natural language processing and determining the sentiments based on it. In [1], the collected tweets are analyzed using lexicon based approach to determine the sentiments of public and a comparison is made among the candidates over the type of sentiment. Also, a word cloud is plotted representing most frequently appearing words in the tweets. Sentiment analysis on social media data has been done by authors of [2] as it is an effective tool to monitor user preferences and inclination. Popular text classification algorithms like Naive Bayes and SVM are Supervised Learning Algorithms which require a training data set to perform Sentiment analysis. Corpus collection, linguistic analysis and Training a classifier was performed step by step in [3]. Corpus is a collection of written texts. They collected a corpus of 300,000 text posts

from Twitter then evenly split into three sets of texts: Positive, Negative and Neutral. It is based on the Naive Bayes classifier that uses N-gram and POS-tags as features. With respect to [4], the mere number of messages reflects the election result and even comes close to traditional election polls. Using LIWC text analysis software, they conducted analysis of over 100,000 messages containing a reference to either a political party or a politician. Their result shows that Twitter is indeed used extensively for political deliberation and mere number of messages mentioning a party reflects the election result. Utilization of Dictionary Based, Naive Bayes and SVM algorithm to build the classifier was done by [5] and classified the test data as positive, negative and neutral. They identified the sentiment of Twitter users towards each of the considered Indian political parties. The results of the analysis for Naive Bayes was the BJP (Bhartiya Janta Party), for SVM it was the BJP (Bhartiya Janta Party) and for the Dictionary Approach it was the Indian National Congress. In [6], authors have used Lexicon based approach with machine learning to find emotions in tweets and predict sentiment score. The research also showed that lexicon based sentiment analysis improves the prediction result, but the improvements also vary in different states. The authors of [7], used fast and in memory computation framework 'Apache Spark' to extract live tweets and perform sentiment analysis. This paper provides a method for analyzing sentiment score in noisy twitter streams and reports on the design of a sentiment analysis by classifying user's perception via tweets into positive and negative. The result of Indonesian Election was predicted by authors of paper [8]. Indonesian people had not elected a president until 2004, so Presidential candidate in Indonesia become a hot and interesting conversation among Indonesian citizen, and many of them expressed it through social media. In this research, the authors focused on tweets related to 2019 Presidential election with top keywords. Among Jokowi and Probowo result is produced by using R language showed that Jokowi leads the election.

3. PROPOSED METHOD

Researchers use a different approach for election result prediction. There are researchers who tried to discover the political or ideology preference of a user, then relate it to the election and there are others who used tweets and polling system related to the upcoming election and figured out vote preference of the user. Some of them have used polling system which has less number of opinions and could not be used for prediction of accurate result. Most of the researchers have used Twitter as their source of data but not everyone is active on Twitter. To overcome these drawbacks considering social media sites such as Facebook, Instagram and Twitter simultaneously can be an advantage. Twitter allows its users to retrieve tweets by using twitter API. Twitter API is simply a set of URLs that take parameters. The URLs let you access many features of Twitter, such as posting a tweet or finding tweets that contain related information on a specific topic. Facebook has complicated, highly granular,

poorly-documented privacy settings. Working with it requires trial and error and a ton of error handling code. It has received less attention because it is complicated and different from other social media APIs. Data extraction from Instagram is also done with the help of various API. Panoply has recently integrated their data warehouse with Instagram API to collect data. After extraction most important process is data preprocessing. The basic flow of this project is as shown below,
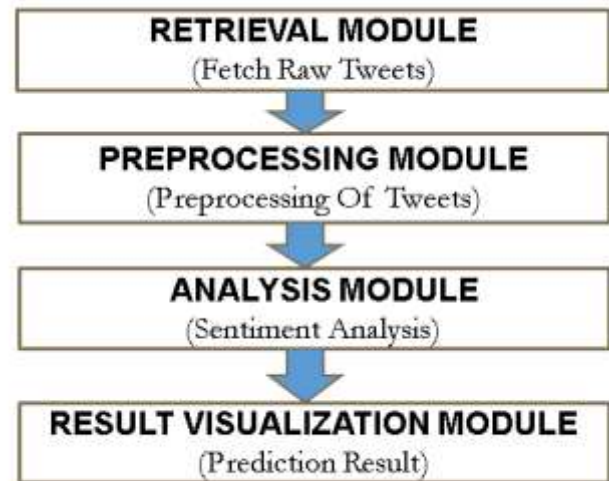

Fig -1: Proposed Model

As explained earlier, Data is retrieved by using different API. Next stage is preprocessing. In this stage, special characters like '@' and URLs can be stripped off to overcome noise. One of the most important goals of preprocessing is to enhance the quality of the data by removing noise. It is a technique which is used to transform the raw data in a useful and efficient format. After Preprocessing, applying SVM and Naive Bayes Algorithm to classify them. A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes. Naive bayes is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. These algorithms will classify the data according to related party. Sentiment analysis can be done on classified data to determine polarity of the word. Researches have done sentiment analysis on words but by applying sentiment analysis on whole sentence may provide better result. The classified data is then given polarity depending upon the sentiment whether it is positive, negative or neutral for each existing party. The result can be represented in the form of pie chart, graph or tables. The comparison between various parties can be done by plotting graphical figures.

4. CONCLUSION

Naive Bayes and SVM are supervised learning algorithms used to classify data according to parties. Most of the researches have extracted only twitter data but we can also use other social media sites like Facebook and Instagram to fetch data. The data used so far is in the form of words and sentiment analysis is applied on it to determine the polarity of the word. But applying sentiment analysis on sentences it may provide better results than on words. Final result can be obtained by comparing Sentiment Percent of various political parties obtained by using above algorithms.

REFERENCES

[1] Ms. Farha Nausheen et al., "Sentiment Analysis to Predict Election Results Using Python", Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018).

[2] Jyoti Ramteke et al., "Election Result Prediction Using Twitter Sentiment Analysis", 2016 International Conference on Inventive Computation Technologies(ICICT).

[3] Alexander Pak, et al., "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", Proceedings of the Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS.K. Elissa, "Title of paper if known," unpublished.

[4] Andranik Tumasjan et al., "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

[5] Parul Sharma et al., "Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter", 2016 IEEE International Conference on Big Data (Big Data).

[6] Pritee Salunkhe et al., "Twitter Based Election Prediction and Analysis", International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 10 | Oct -2017.

[7] Amandeep Kaur et al., "Sentiment Analysis On Twitter Using Apache Spark", Carleton University Project Report.

[8] Widodo Budiharto et al., "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis", J Big Data (2018) 5:51.