Election Result Prediction using Opinion Mining – A Review

[1]Ashweeni A. Kulkarni, [2]Mukta Dhopeshwarkar
[1]PG Student, [2]Assistant Professor
[1]Department of Computer Science & Information Technology,
[1]Dr. Babasaheb Ambedkar Marathwada University, Aurangabad(MS), India

Abstract : Rapid growth in web-based Applications like Blogs, News-Portal, Social Media, E-commerce Sites etc. It has created countless opportunities for the public to vocalize their opinions. Now a day's one of the web application like Social Media has become the mirror of the society. People conferring opinion about the movie or discussion about political parties during election, predicting box-office result or checking fame of any celebrity these all things become easy now because of Social Media. This review paper focuses on the prediction of election result using opinion mining, Presently these day's many researchers approach to utilizing Twitter to monitor people opinion around political activities like campaigns or debates. Analysis of the prediction of elections result using tweets related to political parties or politician. Opinion Mining is used as text classification in which snippets (like sentence, paragraphs, or phrases) are categorized into being objective or subjective and in the letter case positive, negative or neutral.

IndexTerms **- Social Media, Opinion Mining, Twitter, Election campaign, Machine learning.**

I. INTRODUCTION

In today's world an expeditious development in Social Networking websites and micro blogging site . It becomes the biggest web destination for people to communicate with each other and to express their opinion or thoughts about product or movie, and also share their daily experiences and putting their own opinion about the upcoming events like sports or political issues or the upcoming elections etc. Twitter is a micro blogging website that allows user to post brief short 140 character messages called "tweets" and also quick real time updates regarding different activities. Twitter allows the quick spread of news or information as they facilitates sharing, forwarding and replying messages quickly [1]. Political party use this platform for campaigns. Huge investments by political party in social media for campaigns before elections along with arguments or debates between their competitor and supporter to enhance the claim that views and opinion posted by users[5]. There are various sentiment analysis algorithms can be used to identify the attitude of the user regarding to election candidate or political party.

Using opinion mining, in natural language process there are some techniques which used to find opinions. Opinions are expressed in large text through text-processing. In current scenario to categorize opinions as positive or negative there are methods to find the opinion polls.

II. OPINION MINING

Opinion mining consist of Natural Language Processing and Text mining, is used to tracking the mood of public about any particular product or upcoming events or to analyze text and classify them into positive or negative. [3] Opinions of others have highly impact on the human behavior and decision making activities. Opinion mining is also named as sentiment analysis.

Opinion mining is a type of natural language processing, it is also the part of a text mining and web mining for tracking the mood of public. The goal is to make computer able to recognize and express emotions [2]. Opinions are of different type, it can be direct, indirect and comparative [3].

- Direct Opinion

Direct Opinion indicates the view about an particular object like products, events, topics, persons. Example " the picture quality of the camera is great ".

- Comparative Opinion

Comparative Opinion describe the like relations expressing similarities or diffrences of more than one object. Example " car x is cheaper than car y ".

2.1 Working of opinion mining

As shown in following Fig 1 Opinion mining is start by dragging user data generated in the form of blogs or comments or tweets etc. Next step is preprocessing, to perform the task input data is taken from dataset as csv file. In this task tokenization of the text and filtering of data takes place. Then sentiment task is done by using some machine learning methods or algorithms to classify the data into positive, negative opinions [2].

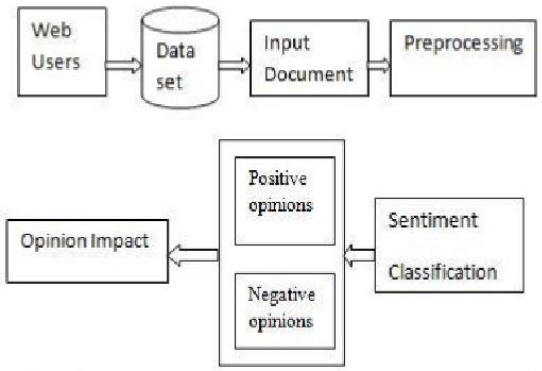

Fig 1: Roadmap of opinion mining

III. RELATED WORK

In this section, we explain the related study of work about predicting the result of an election using sentiment analysis or opinion mining. Many researchers provided research work in sentiment analysis they mainly focused on generating statistical inference from social media.

Campbell and Lewis-Beck investigates the previous research work of Lee sigleman on prediction and conclude that, for data collection we need lot of refinements. They used some methods for Us Presidential Election but due to lack of data it shows improper result [5].

Yulan He, Hassa Saif, Zhongyu Wer, Kam-Fai Wong [6] developed a statistical model for sentiment detection, they also proposed the offline sentiment in real world for public opinion in twitter.

In Akcora, Cuneyt Gurcan, et al [7] they planned a method to determine the changes in public opinion over the time, and also identify the news that led in public opinion for identifying break point.

Brendan O'Connor, et al [8], presents sentiments of public opinion on sentence level classification, they derived from polls from micro blogging site like Twitter. They find both subjective and objective sentence. They also used to demonstrate the consumer confidence on political opinion for linking text sentiment.

Nishantha Medagoda, Subana Shanmuganathan, Jacqeline Whalley [9] they evaluates opinion mining and sentiment classification in non-english language to find the efficiency of each algorithm.

P Bhoir et al in [10] this they used lexicon approach for movie review to implemented method to find the subjectivity of sentence and used rule based system to determine feature-opinion pair and also used another technique orientation of extract opinion is revealed.

The second approach by using selected data from days or weeks before election. The earliest research is done on prediction of result derived by comparison between positive sentiments towards each candidate or the number of tweets mentioning each candidate[11].

Ms. Farha Nausheen and Ms. Sayyada Hajera Begam [12] proposed lexicon based sentiment analyzer which classifies the tweets based on its sentiment value. Polarity and subjectivity measures such as positive, negative or neutral attitude of users towards a particular candidate is done by sentiment analysis.

Jyoti Ramteke et al., in [13] performed data set creation by first collecting data using twitter streaming API . VADAR tool is used for sentiment analysis which is lexicon and rule based analysis tool. VADAR tool is useful and show more accurate result for the prediction of result using sentiment analysis.

Bouazizi et al., proposed pattern based approach for sentiment quantification in twitter. They defined two metrics to measure the correctness of sentiment detection and provide multi-class classification[14].

IV. METHODS

In this section we discuss methods to perform opinion mining to predict election results.

Data Collection

The data collection step is the initial phase in the research, where data is collected from Twitter. Twitter is provided various API to access data. By using the API we can fetch tweets from twitter.

Data Cleaning

To filter out spam tweets, non-English tweets, Abbreviations, Twitter bots to ensure reliability of the database. Preprocessing is used to remove URLs, @tags, hashtags, emoticons, trailing whitespace, non-English words, abbreviations to filter dataset. This method is used for retrieving the meaningful part of the tweet by removing the unnecessary content.

Machine Learning Supervised Classification

i. **Naïve Bayes Classifier**

Naïve bayes classification algorithm is a simple but effective machine learning classifier. This classifier is used for sentiment analysis to find opinion polls sentiments such as positive or negative for the given opinion[15].

$$C^* = \text{ARG}\text{MAX}_C P_{NB}(C/D)$$

$$P_{NB}(C/D) = \frac{(P(c)\sum_{i=1}^{n} P(f/c)^{,(d)})}{P(d)}$$

In this above mathematical expression,

Where, n is the no. of features and f is the feature vector.

Suppose consider a training method consisting of relative frequency estimation P(c) and P(f | c) [16]. Specifically, Naïve bayes classifier classifies a document based on the count or relative frequency with which word appears in a document.

ii. Support Vector Machine

The support vector machine is a learning machine classification [17]. It is used to classify the text as positive or negative and also for regression and classification challenges. SVM works for text classification. Svm has best advantage such as its potential to handle large feature. Another advantage of svm is robust when there is sparse set of example. Svm gives More accuracy then it gives in Naïve Bayes and maximum entropy classification [18].

IV. RESULTS AND DISCUSSION

With the immense use of social media, this paper is focused mainly on use of social media as a tool for election campaign and the techniques used for predicting result of the election. Here we discussed about the political inclination like the user behavior about election prediction. Here we provided a view regarding to the user behavior about any product or entity on the basis or their reply or comments. The different approach are discussed for the analysis of the data and also discuss some machine learning classification algorithms like Naïve bayes classification and SVM for classification of data. We discuss and conclude that It gives more accuracy of data as comparing to the classification techniques.

Prediction of elections result could be made by number of positive sentiments towards political party.

V. ACKNOWLEDGMENT

I would like to acknowledge Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad for providing me all facilities related to my research work; I would also like to acknowledge my research guide Dr. Mukta Dhopeshwarkar for her valuable guidance.

REFERENCES

[1] Godbole, Namrata, Manja Srinivasaiah, and Steven Skiena. "Large-Scale Sentiment Analysis for News and Blogs." ICWSM 7.21 (2007): 219-222.

[2] G.Angulakshmi1, Dr.R.Manicka Chezin," An Analysis on Opinion Mining: Techniques and Tools" ,Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, India, International Journal of Advanced Research in Computer and Communication Engineering Vol.3, Issue 7,July 2014.

[3] Gautami Tripathi and Naganna S," Opinion Mining: A Review", International Journal of Information & Computation Technology.ISSN 0974-2239 Volume 4, Number 16 (2014), pp. 1625-1635 © International Research Publications House.

[4] Mondher Bouazizi, Tomoaki Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", Access IEEE, vol. 5, pp. 20617- 20639, 2017, ISSN 2169-3536.

[5] James E. Campbell and Michael S. Lewis-Beck, " US presidential election forecasting: An introduction", Elsevier, 2008

[6] Yulan He,Hassan Saif,Zhongyu Wei,Kam-Fai Wong "Quantising Opinions for Political Tweets Analysis"www.researchgate.net/publication/233934221,July-2015.

[7] Akcora, Cuneyt Gurcan, et al. "Identifying breakpoints in public opinion." Proceedings of the first workshop on social media analytics. ACM, 2010.

[8] Brendan O"Connor, et al. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series" (2010). Association for the Advancement of Artificial Intelligence 2010.

[9] Nishantha Medagoda,Subana Shanmuganathan, Jacqueline Whalley,"A Comparative Analysis of Opinion Mining and Sentiment Classification in non- English Languages ",Auckland University of Technology-2013.

[10] Bhoir, Purtata, and Shilpa Kolte. "Sentiment analysis of movie reviews using lexicon approach." Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on. IEEE.

[11] Tumasjan, A. S. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. ICWSM, 10, 178-185.

[12] Ms. Farha Nausheen and Ms. Sayyada Hajera Begum presented "Sentiment Analysis to Predict Election Results Using Python" International Conference on Inventive Systems and Control (ICISC 2018) IEEE International Conference on. IEEE, 2018.

[13] Ramteke, Jyoti, et al. "Election result prediction using Twitter sentiment analysis." Inventive Computation Technologies (ICICT), International Conference on. Vol. 1. IEEE, 2016.

[14] Bouazizi, Mondher, and Tomoaki Ohtsuki. "Sentiment analysis in twitter: From classification to quantification of sentiments within tweets." Global Communications Conference (GLOBECOM), 2016 IEEE. IEEE, 2016.

[15] Nishantha Medagoda,Subana Shanmuganathan, Jacqueline Whalley,"A Comparative Analysis of Opinion Mining and Sentiment Classification in non- English Languages ",Auckland University of Technology-2013.

[16] Jayashri Khairnar,Mayura kinikar,Departmentof Computer Engineering, Pune University, NIT Academy of Engineering,Pune,"Machine Learning Algorithms for Opinion Mining and Sentiment

[17] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning,vol. 20, no. 3, pp. 273–297, 1995 T. Joachims, "Text categorization with support vector machines: LearNing with many relevant features," in Machine Learning: ECML-98, ser. Lecture Notes in Computer Science, C. N´ edellec and C. Rouveirol, Eds. Springer Berlin Heidelberg, 1998, vol. 1398, pp. 137–142.

[18] T. Joachims, "Text categorization with support vector machines: Learn-ing with many relevant features," in Machine Learning: ECML-98, ser.Lecture Notes in Computer Science, C. N´edellec and C. Rouveirol, Eds.Springer Berlin Heidelberg, 1998, vol. 1398, pp. 137–142.