Available online at: http://www.ijmtst.com/vol6issue02.html

International Journal for Modern Trends in Science and Technology

ISSN: 2455-3778 :: Volume: 06, Issue No: 02, February 2020

A Survey Paper on Selection of Candidates Using Machine Learning

Prof. Mohan Nikam[1] | MuskaanShaikh[1] | Rutuja Bhusal[1] | Rajkumar Singh[1]

[1]Department of Information Technology, Sandip Institute of technology and research centre, Nashik, Maharashtra,India

To Cite this Article

Prof. Mohan Nikam, MuskaanShaikh, Rutuja Bhusal and Rajkumar Singh, "A Survey Paper on Selection of Candidates Using Machine Learning", *International Journal for Modern Trends in Science and Technology*, Vol. 06, Issue 02, February 2020, pp.-14-16.

Article Info

Received on 15-January-2020, Revised on 28-January-2020, Accepted on 06-February-2020, Published on 09-February-2020.

ABSTRACT

In present era, the use of social media is on boom. Due to its real-time and interactive nature, social media has received much attention. The main challenges with social media are security, privacy, accuracy. Still social media has given us new opportunities for sharing, collaboration, and engagement of users. It provides the platform for politics as well. Social media platforms, like Twitter, Facebook and YouTube motivate people to get involved in all the political activities by sharing their view about party and candidates.

Through Twitter, Facebook and other social media, analyzing the candidates based on their past experience, finding out how compatible is the candidate for winning the elections in a particular region is the motive of our project. Through this people will be able to understand which person is worthier to be their representative. It is beneficial for the public as well as for the election members.

Sentiment information mined using machine learning from twitter data models was the most accurate predictor of election outcomes and hence can be useful for selecting a candidate as well based on the negative or positive image he has maintained in the eyes of the people.

KEYWORDS: *Sentiment analysis, Candidate selection, political candidate, Social media, Tweets*

I. INTRODUCTION

As compared to traditional methods of public opinion measurement, social media allows time- and cost-effective data collection and analysis with less human effort. Social media can be used as a real-time complement to traditional surveys, to monitor the day-by-day sentiments of voters towards electoral candidates and to identify trends in user's political preferences, which take time and effort to collect and collate from survey responses.

This project is made for helping the public to know the candidate well for elections and to know who more deserves to win the election. The constituency will be able to know the probability that their candidate can win the election. The aim is to build a system that will help the public to nominate a deserving candidate for elections and find the chances of winning of that candidate.

The proposed system collects data from Twitter social network site and does NLP techniques to extract features from the tweets. Then various Ensemble methods of classification are applied to classify the data as Positive, Negative and Neutral. It has been observed that the ensemble method outperforms the traditional classification methods.

The ensemble methods Extremely Randomized Trees classification performs better than others.

We find that predictive power of social media performs well for India. Overall, we find that it is useful to consider the Twitter posts while using it to predict a real outcome, such as selection of candidate of election. Sentiment information mined using machine learning models was the most accurate predictor of election outcomes. Social network information is stable despite sudden surges in political discussions.

Methods combining sentiment and volume information, or sentiment and social network information, are effective at predicting smaller vote shares, for e.g. vote shares in the case of independent candidates and regional parties. We conclude with a detailed discussion on the users of social media analysis for predicting real-world outcomes and recommendations for future work.

Extraction of Tweets using Tweepy:

Twitter is a popular social network where users share messages called tweets. Twitter allows us to mine the data of any user using Twitter API or Tweepy. The data can be tweets extracted from the user. The first thing to do is get the consumer key, consumer secret, access key and access secret from twitter developer available easily for each user. These keys will help the API for authentication.

II. LITERATURE SURVEY

Most of the authors discuss the opinions and attitudes expressed in Twiter are growing in abundance nowadays and people decide everything based on these tweets what users make. People express their views via tweets and because of this, it leads to the creation of an enormous amount of data. Huge storage of data becomes a major problem for companies nowadays. Here comes the area for analyzing data mostly on popular platforms such as twitter. There are several methods are available for analyzing the data. In paper [1] author discusses the Twitter API.

This paper evaluates the people's sentiment about a person, trend, product or brand. Using the Twitter API, it is possible to collect twitter data. RAuth is used in performing authentication by giving in the keys. Consumer Secret, Consumer Key, Access Token and Access Token Secret for twitter application and perform Handshake protocol. After completion of this, the certificate is downloaded and PIN is generated for the application to access tweets. After this, Twitter data is classified based on sentiments. This paper [2] shows sentiment analysis types and techniques used to perform the extraction of sentiment from tweets. Sentiment Analysis, Opinion Mining, Social Media, Twitter Data these techniques are explained.

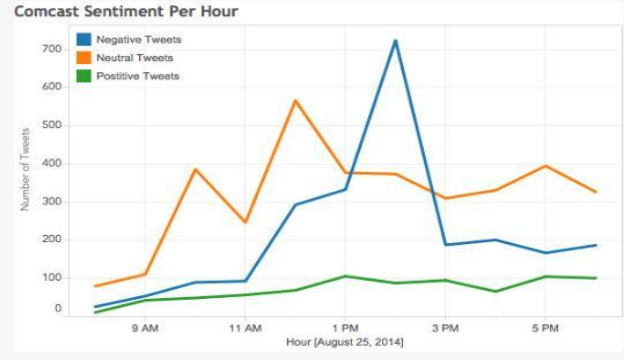
They have taken the comparative study of different techniques and approaches of sentiment analysis having twitter as data.The user's tweets from twitter are collected based on input specified by user in the form of hashtags. The process of classifying tweet is initiated by first collecting the tweets.Using twitter API, it is achievable to collect the twitter data. Library named RAuth is used in performing authentication by giving in the keys. Consumer Key, Consumer Secret, Access Token and Access Token Secret for twitter application and perform Handshake protocol. After which, certificate is downloaded and PIN is generated for the application to access tweets. In paper [3], various techniques used for Sentiment Analysis i.e. Natural Language Processing, Case Bases Reasoning (CBR), Artificial Neural Network (ANN), Application Programming Interface, Python, Sentiment Analysis. This paper reports on the design of sentiment analysis, extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers' perspectivevia tweets into a positive or negative, which is represented in a pie chart. Initial steps are the same for all papers, In paper [4] after the pre-processing of the tweets featured vector is created. In this paper, Sentiment classification is done by three classification Technique i.e. Naive Base Classifier, SVM Classifier, Maximum Entropy Classifier. Paper [4] and Paper [5] are similar. Both use the same methods but Paper [5] used a supervised machine learning algorithm for sentiment analysis. The outcome of the analysis is depicted for positive, negative and neutral remarks about their opinions using visualization techniques such as histogram and Pie chart.

III. PROPOSED SYSTEM

A trend analysis and a popular or trending candidate and a sentiment analysis to actually bifurcate the positive and negative tweets for the candidate so that making trend analysis on this tweets can help this party or the candidate to act accordingly to improve their reputation at the same time and it might help user to actually make a clear opinion about any party/candidate. This will be conducted in three phases. In detail about it the phase one is connecting with tweeter and downloading the tweets. The second phase deals

with loading these tweets on HDFS for further process of analysis and the third phase is the actual analysis of tweets and they are the Volume analysis, Trend Analysis and Sentiment analysis.

The today's comparative study compares the performance of three major approaches for predicting the vote share in the general elections of three Asian countries. It also assesses the utility of social media models for such predictions, by benchmarking them against previous election results and traditional opinion polls. It implements all three commonly used approaches found in previous work volume-based models, sentiment analysis based on the lexicon and probabilistic models, and social network analysis.



In characteristic manner, the sociological approach is represented by the well-known writings of Seymour M. Lipset, for whom 'in every modern democracy conflict among different groups is expressed through political parties which basically represent a democratic translation of class struggle'. This viewpoint, can be contrasted through the experience raised by Spanish democratic elections and, surprisingly enough, can be thought as still valid always keeping in mind the changes in social classes configuration that arise in advanced societies of information.

Social states are not defined now as in Lipset's writings years, but a common rationale can be found in relation to the influence of social differences.Cleaning of tweets is done by removing additions of text in the tweets like URL's, numbers and special characters which shortens the size of tweets for comparison. In phase III, the pre-processed tweets are compared with the available bag of words (BoW) and are classified as positive, negative and neutral. Two files under bag of words are used, one for positive and one for negative.

The important goal and the challenge of the system is analysing twitter data for Indian election to see the actual impact of tweeter on Indian or particular state of election. The proposed system is consists of three main steps. The public orientation towards these parties can be studied using the tweets the people have posted about the parties or candidates on the Tweeter.

Tweeter is generally consumed by academicians, journalists and Politicians for its potential political value. Most of the politicians make use of this micro blogging site to express themselves in the limit of 140 characters generally. All these tweets can be differentiated and categorized on various policies such as relocation analysis to analyses the peoples view for that particular area which can help parties to design their winning strategy. The dataset comprises 3.4 million tweets related to the general elections in Malaysia, India, and Pakistan. It also evaluates the impact of different pre-processing and temporal weighting steps on prediction accuracy.

IV. CONCLUSION

With the increased use of social media in today's world, the current paper focused election campaign. India is well known to be one of the wired countries in the world ith having more than 65 % of its youth below age-group of 35.

Social media plays very important role in the life of this young youth. The proposed systems will try to analyze the Maharashtra State Assembly Election. To study the impact of social media on Maharashtra Politics System found people can express their views in 140 characters most efficiently and openly.

REFERENCES

[1] Prakhurti V, Sindhu D, and Dr S Anupam Kumar, " Real Time Sentiment Analysis Of Twitter Posts" in: Department of MCA, RashtreeyaVidhyalaya College of Engineering, Bengaluru. ISBN: 978-1-5386-6078-2 © 2018 IEEE

[2] Rasikawagh, and PayalPunde, "Survey on Sentiment Analysis using Twitter Dataset " in: (Mtech): Department of Computer Science and IT, Dr. BAMU Aurangabad, India. 978-1-5386-0965-1/18/$31.00 ©2018 IEEE.

[3] AlizaSarlan, ChayanitNadam, and ShuibBasri, "Twitter Sentiment Analysis" in: Computer Information Science UniversitiTeknologi PETRONAS Perak, Malaysia. 978-1-4799-5423-0/14/$31.00 ©2014 IEEE.

[4] Neethu M S, and Rajshree R, "Sentiment Analysis in Twitter using Machine Learning Technique" in: Department of Computer Science and Engineering College of Engineering Trivandrum, 695016, India. 4th ICCCNT 2013 July 4 - 6, 2013, Tiruchengode, India IEEE – 31661.

[5] Mitali Desai, and Mayur A. Mehta, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey" in: Computer Engineering Department Sarvajanik College of Engineering and Technology Surat, India. ISBN: 978-1-5090-1666-2/16/$31.00 ©2016 IEEE.