Analysing Political Party Influence Using Data Mining

M.P. Vani
Department of Smart Computing
School of Information Technology and Engineering,
Vellore Institute of Technology University, Vellore, India
Email: mpvani@vit.ac.in

***Abstract*—** The ongoing progression in the usage of computers and smartphone along with cheaper rates for wireless internet has led India to have a firm foot in the modern technology era. As the number of people using digital platforms is increasing, especially the younger generation, they share their feeling, opinion, perspective, problems that are almost all their life's story in the social networking platforms in the form of tweets on Twitter, post on facebook etc. This all information can be used to know the opinion of a person or a community on a certain topic. Along with the information found on the internet, newspaper, websites can also be used as the analysis for the issues faced by certain community.

***Keywords*—** Progression, smartphone, twitter, facebook, wireless, internet.

I. INTRODUCTION

The aim of this paper is as the above mentioned data from social networking websites, and information from the online newspaper to be considered as big data and to perform a subsequent analysis on the given data to know which political party has stronghold in which region, and if a party has negative or weaker impact on the citizen of a region, provide a way to improve their position in the community. The public opinion of a political party can be improved, if the party is seen to be socially aware of the problems and dilemmas of that region, this information can be found by the current news articles which are provided in the online news websites. All these information can help a political party to have a good lead from their opposition during the election and also they can maintain a good public opinion and support after the election period is over. Even when the party doesn't win the given information can help them to improve party's current position in the region, so they gain party support in a community and can have a better chance of winning in the next time round. This method is also a good way to attract the younger generation to vote, mostly for those who are first-time voters or hesitant to vote, by launching a television and online advertising campaigns for the party and party policies which will target the current topics which are trending in the social platforms.

The data can be mined using classification and clustering the data according to location, language and month. As the data are majourly text which is subjective in nature the main mining method is to classify, it can be text categorization with the discriminative classifier and opinion mining with sentiment analysis with sentiment classification which is known as ordinal logistic regression.

The fundamental objective and test of the framework are the investigating twitter information and news information for Indian population to see the effect of social on Indian or on the specific state's sentiment on a political gathering. The proposed calculation is comprised of four primary advances :

II. METHODLOGIES

Algorithm 1: K-Means Clustering Algorithm

At first, subject savvy tweets are set as a jog of clusters. For each emphasis, distance amongst focus and test is checked and the test is added to separate cluster. Separation amongst focus and test is estimated utilizing TF-IDF.Clusters are refreshed at each cycle. Based on TF-IDF weight.

Steps:

Input: let N be set of data points where $N=n_1,n_2,n_3,\ldots\ldots n_n$ and let X be set of centers where $X=x_1,x_2,x_3,\ldots\ldots X_n$

Step 1: Pick number of clusters to be settled

Step 2: Set initial centers for the clusters, (centroid is chosen randomly)

Step 3: Repeat the following until there is no change in cluster centers or clusters do not change:

- Designate each of the objects to the closest cluster center using Euclidian distance.
- Calculate new cluster centers by calculating mean

Algorithm 2: Naïve Bays Classifier[4]

Political introduction of clients towards the party, points can be dissected from tweets. Map Reduce form of naïve Bayes calculation will be executed to group tweets into classes of positive, negative and neutral.

Step 1: a list of positive and negative tweets are enlisted separately from each other.

Step 2: Convert this two list into single list having two parts one is a comment and other is the type.

Step 3 : Design Classifier :

- Take the word that is featured on the list with its frequency count
- Using this list of words a feature extractor is created (contains the words which matched with words given as input)

Step 4: Training dataset is used for training the classifier

- Create a label_list which contains positive as well as negative labels.
- Create feature_list containing feature words.

Step 5: Calculate the probability of the positive and negative label.

Step 6: compare above probability to identify tweets to categorize as positive or negative or neutral.

Algorithm 3: Sentiment Analysis[5]

Sentiment analysis could be utilized to examine or anticipate tweets for a political party and the sentiment of the user for that party. It has seen countless in late time as 80% of information accessible is unstructured as content, sounds, recordings and so on.

Step 1: Gather examples of the sentiment

With respect to each managed learning issue, the calculation should be prepared from marked cases keeping in mind the end goal, to sum up to new information.

Step 2: Extracting features from the given example

Change every case into a component vector. The most straightforward approach to do it is to have a vector where each measurement speaks to the recurrence of a given word in the record.

Step 3: Parameters are trained accordingly

This is the place the model will gain from the information. There are various methods for utilizing highlights to produce a yield, however, one of the least complex calculations is a strategic relapse. In the least complex shape, each element will be related to a weight. On the off chance that you have in excess of 2 yield classes, for instance in the event that you need to order between "positive", "impartial" and "negative", each element will have the same number of weights as there are classes, and the class with the most noteworthy weighted component total wins.

Step 4: Testing model

After preparing the parameters to fit the preparation information, ensure the model sums up to new information since it's extremely simple to overfit. The general method for regularizing the model is to keep parameters from having outrageous esteems.

Algorithm 4: Trend Analysis[3]

A trend analysis is a part of the specialized analysis that tries to foresee the future development of a gathering in view of past information. Trend analysis depends on the possibility that what has occurred in the past gives dataset a thought of what will occur later on. There is three fundamental sorts of trends: short, intermediate and long haul.This analysis can be used know what is the current issues trending among the population of certain place through a newspaper article.

Step 1: collect a list of news headlines.

Step 2: give a set of pre-determined words for analysis

Step 3: separate the list according to the word features

Step 4: count the total number of headlines in each word list.

Step 5: The word with a maximum number of headline is the more pressing issue concerning the citizens.

Results

Sentiment Analysis with Naïve Bayes:

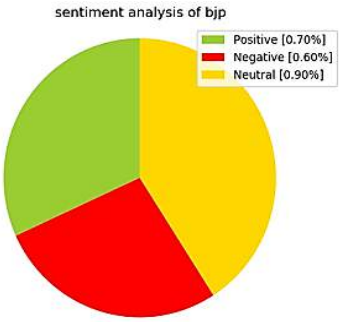


Fig 1. Sentiment Analysis for BJP

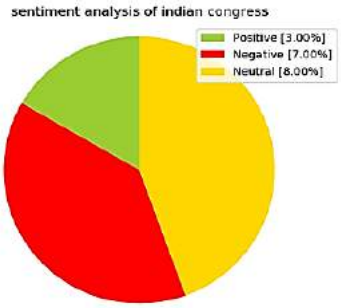


Fig 2. Sentiment Analysis for Indian National Congress

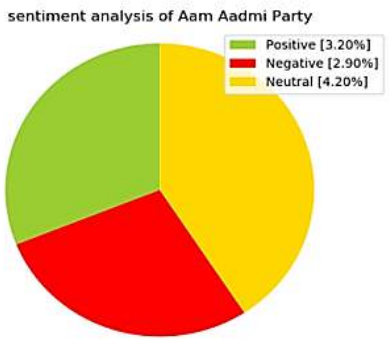


Fig 3. Sentiment Analysis for Aam Aadmi Party

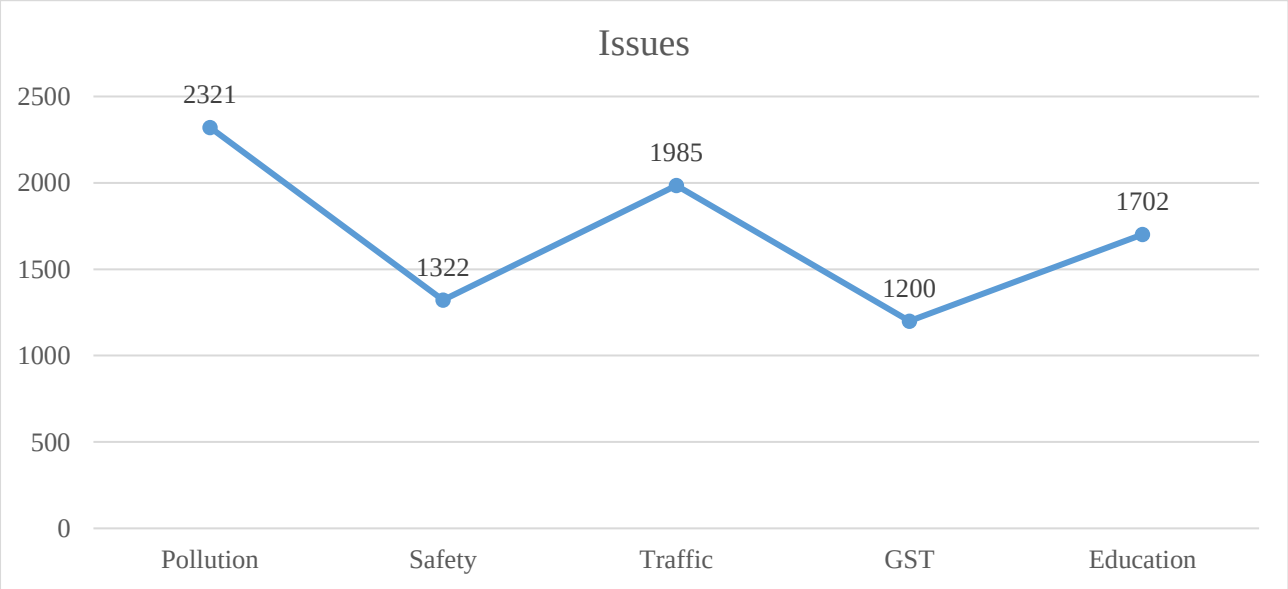
In the above three figures, (Fig. 1,2,3) a sentiment analysis using naïve Bayes theorem was used to know the sentiments and thoughts of the users in twitter about BJP (Fig 1), Indian National Congress (Fig 2) and Aam Aadmi Party (Fig 3). The pie chart in the figures is divided into three-part positive, negative and neutral. The result is drawn from the data of 1000 unique tweets for each party. The neutral tweets in the data can be considered as imperial tweets or tweets that can't be differentiated by the API as either positive or negative as some tweets can be Hindi or any other native language. In Fig 1, BJP has 10 % more positive tweets than negative tweets (60%), but for Indian National Congress the tweets are more towards the negative side with negative tweets 40% more than the positive tweets. For Aam Aadmi Party in Fig 3, the tweets have a similar trend as BJP's. Here it can be said that if we ignore neutral tweets, it can be said that BJP has more support from the users than Indian Congress while the Aam Aadmi Party is in the middle.

Table 1. Clustering tweets to understand users' orientation

Cluster 0	**Value : 1000**	**Len : 9651**
32390397609791488, 32390397618196483, 32390397735641088, 32390397853073410, 32390397962121216, 32390398012461057, 32390398230544385, 32390398314438656, 32390398352195585, 32390398826164225, 32390398993932289, 32390399149109248, 32390699295926273, 32390399300100096, 32390656318676993, 32390787551836160, 32390771256938496, 32390845545049088		
Cluster 1	**Value : 940**	**Len : 8773**
32390483584655360, 32390648529789952, 32911610236293120, 32391000567697409, 32391605161438368, 32920146454425600, 32392151028702848, 32392355979996417, 32925264352546816		

In the above table, we have taken two clusters each for one party. Here we have only chosen the BJP and the Indian Congress party as these two parties were popular in among the users. In the above cluster 0 belongs to the BJP party as it can be seen any numbers of user belong to the BJP party while the second cluster cluster1 belongs to the Indian National Congress Party it can be seen the number is significantly lower than the BJP party. It can be said that the BJP party among the three parties are more popular than the three given popular parties (BJP, INC, AAP).

Trend Analysis



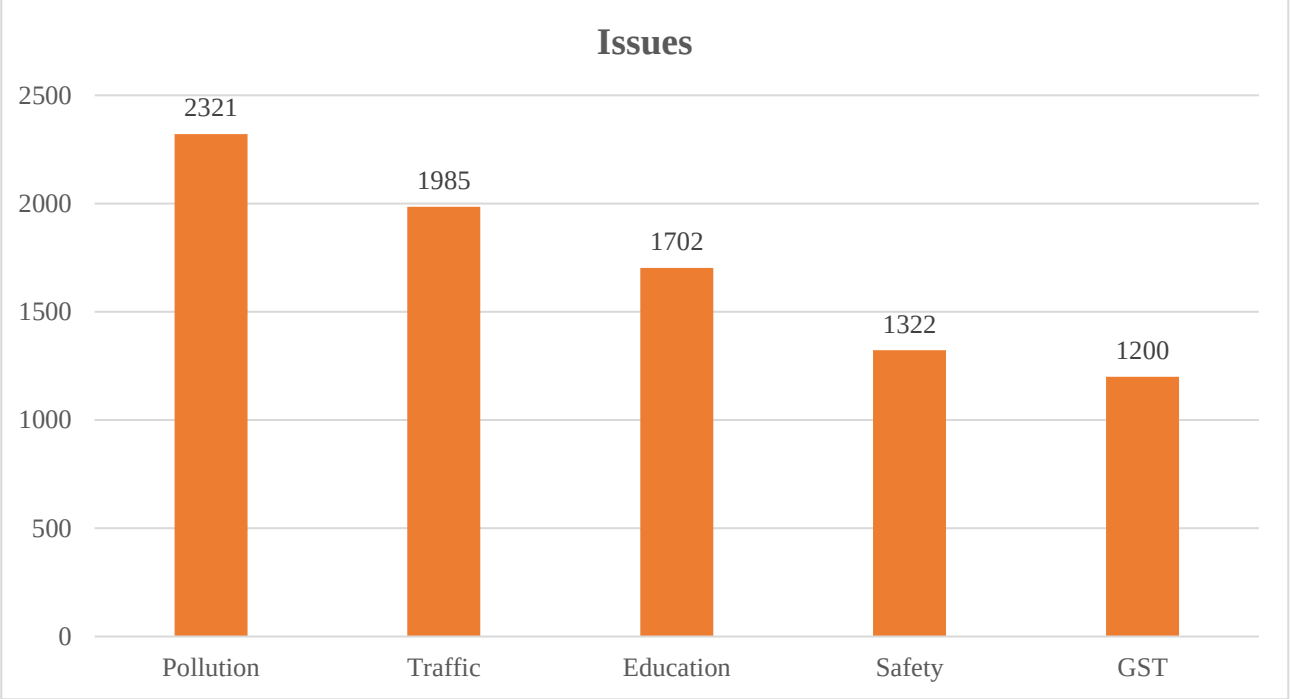


Fig 5. Bar Chart Showing Top 5 issues (in decreasing order) concerning the population

In the above chart, it is seen the top five concerns of the citizens are about pollution, traffic, education, safety, and GST. While Pollution, Traffic, and Education are the top three main concerns of the citizens. It is seen the pollution ranks much higher than any other concerns. It can be said that the party who is trying to gain the favor of the citizen can take the issue of pollution along with other four issues given and make their policies around it according to the needs of the citizens and propagate these agendas around for more favorable popularity among the population.

III CONCLUSION

In the present paper the expanded utilization of online networking is concentrated on utilization of web based social networking as an apparatus for decision making. India which is known to be one of the wired nation on the planet consisting of 65% of its youth beneath age group of 35, social media assumes as an indispensible part in life is youthful youth.

The proposed frameworks will attempt to investigate the political party impact and if the gathering's impact is less good then they can know about the issues that leads to anxiety and pick up the notoriety of the populace by influencing the strategies around these issues. In order to concentrate the effect of web based social networking on political perspective in individual networking on political perspective individuals can express their perspective on 140 characters more proficiently and straight forwardly.

IV FUTURE WORK

The examination can be improved to really group the gender orientation of the client and to co-ordinate it with genuine socio economic, it can be incorporated with degree for checking whether the cliemt is human or robot. The examination can be incorporated in a machine learning method of preparing framework to consequently tweet and complete a sentiment analysis of the tweets.

REFERENCES

[1] Sharma P., Moh T. S., Prediction of Indian election using sentiment analysis on Hindi Twitter, Big Data (Big Data), 2016 IEEE International Conference, pg. 1966-1971, IEEE. (2016 December).

[2] Wani G. P., Alone, N. V., Analysis of Indian Election using Twitter, International Journal of Computer Applications, 121[22], (2015).

[3] Bhola, A., Twitter, and Polls: Analyzing and estimating the political orientation of Twitter users in India General # Elections 2014, arXiv:1406.5059, (2014).

[4] Song M., Kim M. C., Jeong Y. K., Analyzing the political landscape of 2012 Korean presidential election on Twitter, IEEE Intelligent Systems, 29[2],(2014).

[5] Almatrafi O., Parack S., Chavan B., Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014, 9th International Conference on Ubiquitous Information Management and Communication, pg. 41, (2015 January).