Sentiment Analysis of Social Media For Indian Politics Using PNR Indicator

Daxita Rajput[1], Prof Trupesh Patel[2]
[1, 2] Guajrat Technological University, Ahmedabad, Guarat, India

Abstract- *Social media platforms have become a major influence on elections: they are increasingly being used to shape political opinion and beliefs generally, and in electoral periods they influence voter choices. Reports from many countries have shown that disinformation attempts to manipulate elections, be it via discrediting campaigns, external influence or trying to suppress voter turnout. Some may represent violations of electoral or other rules, while others may not be illegal in national legislation, but are nevertheless inconsistent with the idea of fair campaigning as outlined in international law. Besides disinformation, social media also facilitates the placement of paid political and issue-based ads, targeted to the preferences of different groups of voters. By weighting the tendency of tweets, we were able to obtain a proposed indicator of support: the positiveness ratio (PR). The results suggest that PR is a feasible barometer to demonstrate the measurable patterns of support tendency regarding political parties and users' behavioral activity to track their affinity on Twitter, Facebook and other social media. The findings indicate consistent support behavior by users toward traditional parties and optimistic users' behavior regarding emerging political parties.*

Keywords- Sentimental Analysis, Positiveness Ratio, Social Media Sentiments, Indian Election. Natural Language Processing

I. INTRODUCTION

Social media data has been important in every segment in current scenario. Data collected from social media represents feeling and review of the crowd. Views of people from particular area for any particular event or particular topic can be easily extracted using Machine Learning. As election is main event for any democratic country and views of people are very important at the time of election. In our research our main goal is to extract the positive and negative feeling and views of people in India at the time of election. We extended "Indicator Proposal for Measuring Regional Political Support for the Electoral Process on Twitter: The Case of Spain's 2015 and 2016 General Elections" [1] for our research purpose. As existing work has been proven better for Spain Election result analysis, we decided to extend the work for Indian Election. As twitter is an important source of social media worldwide. But as we are going to work for Indian election, twitter cannot consider as enough data. We have also used Facebook data for the same purpose also. Election like event generate vast amount of discussion and comment data in both platform. We have considered text data and ignored video and image contents. Through social media, researchers can analyze trends, evaluate public opinion, gauge reactions and appraise voters. [3][4].

Based on the social media content analysis, online discussion forums offer great source for political debate to occur due to easy access to low-cost discussion platforms. Consequently, new media has become increasingly important during election campaigns [5], [6]. Generally, users shares shorter text to demonstrate or relate their opinion and views. Mostly user ignores writing longer excerpts or documents [7], and this is one of the main reasons to choose Twitter and Facebook as a research platform

Up to now so much work related to Social media impact and analysis or Social media data has been carried out during election time by various researchers. We have studied some of them. As "Indicator Proposal for Measuring Regional Political Support for the Electoral Process on Twitter: The Case of Spain's 2015 and 2016 General Elections" [1] has done work for Spain Election. They have introduced an indicator named PR (Positiveness Ratio) Indicator. In their work they have collected data from twitter. They have collected 2, 50,000 tweets related to Spanish Election during 2015 and 2016. They have divided data based on geolocation.

Their result suggests that PR is feasible barometer to demonstrate the measurable patterns of support tendency regarding political parties and users' behavioral activity to track their affinity on Twitter. The model applied for their system has been shown in figure 1.

As users of Social Media has been increased in last decade in Indian and all over the world; users share their views toward various political points and parties. Analysis of their views are important for any political party. Positive and Negative sentiments detection from various posts can lead a proper decision for an political party, so that we can say it is

very important to analyze sentimental from Social Media in politics [11-13].

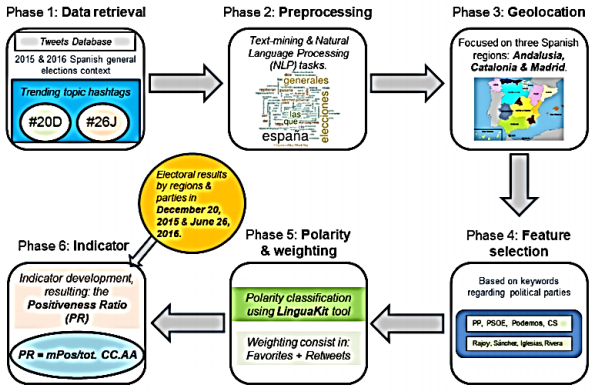

Figure 1 Model of Existing System [1]

As in Phase1 they are collecting the data from social media platform twitter. In second phase they have preprocessed the data. In phase 3 the divided data in 3 parts based on Geolocation. Then in phase 4 they have applied Feature selection based on keywords. In phase 5 they have applied polarity classification using Lingua Kit. In last phase 6 they have developed PR Indicator to extract the ratio of Positiveness. As a result they have concluded their method to visualize PR matches the result of election. In another research "Sentiment Analysis of Tweets for the 2016 US Presidential Election" [2] was carried out for US Election 2016. In their research work they collected tweets related to election in US and labeled manually based on Hash Tags. They have applied lexicon and Naive Bayes Machine Learning Algorithm to calculate the sentiment. As US has presidential election system they have classified the tweets as negative or positive for both candidates. In research "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis" [8] they have also used micro blogging website twitter for data collection. They have considered Tweets, Retweets, Word count etc. features for their Sentimental analysis. We have also reviewed the research work "Casting Online Votes: To Predict Offline Results"[9], as this work has been carried out for New Delhi General Election, they have collected the data from Delhi location. They have used NLTK for their research implementation purpose. They have compared various classifiers. The comparison result for their system is shown in Table 1.

Table 1 Comparision Result of Research [8]

Sr No	Classifier	Accuracy	Classification Error
1	Original Naive Bayes	72.00	0.28
2	MultinomialNB	80.00	0.20
3	BernoulliNB	75.00	0.25
4	Linear Regression	76.00	0.24
5	SGDC	69.00	0.31
6	Linear SVC	71.00	0.29
7	NuSVC	78.00	0.22

II. BACKGROUND

2.1 Election System In India

As India is the largest democracy in the world. Election in India in not a single day program. In India there are 2599 political parties registered in 2019. Among them 8 national parties are there. In India important of rural parties cannot be ignored when we are working for political research. There are also alliance system exists in India. In last election there were three major alliances were existing. Among them UPA and NDA are main alliance those are ruling in India since more than 20 years. For analysis purpose we have to also consider this alliance system.

2.2 Social Media in India

As India is a large country and so many languages are there. There are 22 official languages in India, so social media data collection method should be modified to collect the best data. Another point to be note is nearly 68% population of India belongs to villages and Twitter is not a platform that is used by most of part of India, so that another social media platform Facebook is also included in our research work.

2.3 Recent Work Toward Sentimental Analysis in Politics

In recent years; after introduction of Social Media in technical world; there is constant work being done to mine the related information. Some of the recent work has been reviewed in this paper earlier. Most of work are depending on Text Classification, Text Clustering and other machine learning techniques.

III. PROPOSED METHOD

As our proposed method is extension and modification to research work [1] with aspect of Indian Election System. We have divided our work in 6 Phases. First of all we collect the data from social media platform Twitter and Facebook. As collected data may be in region languages so that we have converted the data into English using Google Translate. Then we cleaned the dataset. In this step we have

removed Spam Tweets and other Preprocessing are done. In next phase we have generated the dataset based on clean data. We applied Feature selection in next phase. After feature selection polarity classification has been performed for Alliance wise and Party wise. In last phase Positive and Negative Ratio (PNR) is calculated. The detailed model is shown in figure 2.

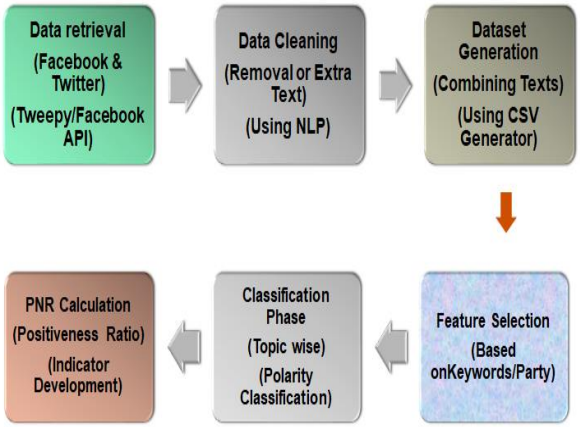

Figure 2 Proposed Model

2.4 Data Collection

As stated earlier we have used Twitter and Facebook data for our research purpose. For data extraction from twitter we have used Twitter API. We have also used indian-political-tweets-2019-feb-to-may from kaggle. This dataset contains 1,17,576 tweets related to Indian Election collected between February 2019 to May 2019. To collected historical data from Facebook we have used GraphAPI.

2.5 Data Cleaning

In this module collected data has been cleaned using Text Libraries of Python. We have removed non-informational contents. We have applied Stop Word Removal techniques. After data cleaning length of the tweet or post is included as one of the features for further processing.

2.6 PNR Calculation

As we stated earlier we are using PNR indicator. As research[1] they have used only positiveness ratio (PR) we have also negativeness also. PNR is a modified indicator, which was originally developed by researchers[1] that undertook an in-depth examination of Twitter users' support by weighting the tendency of tweets.

PR=mPos/tot.

For each political party/Alliance based on the weighting of the positive amounts divided by the sum of the positive accumulation of tweets.

NR=mNeg/tot.

For each political party/Alliance based on the weighting of the negative amounts divided by the sum of the negative accumulation of tweets.

As we are using negative and positive contents for our research work we have extracted Positive Text and Negative text with party and alliance wise. We have ignored the social media content that is marked as Neutral using our sentimental analysis. We have also included the leaders of various parties like 'Modi', 'Rahul' etc. for better result.

III. RESULTS

3.1 Experimental Setup

We have implemented our research work using Python 3.0 on Windows Machine having 16GB of Internal memory. Various libraries like Tweely, NLP, Pandas are used for our implementation purpose.

3.2 Result Statistics

We have collected Twitter and Facebook data and extracted some numerical features from the dataset. Below tables displays some of the data extracted from our collected social media content.

Table 2 Data Collection

Type of Documents	Frequency
Total Documents (Facebook + Twitter)	205832
Total Positive	65332
Total Neg	67240
Neutral	73260

Below Table 3 shows party wise occurrence of keywords in our collected data.

Table 3 Frequency of Important Terms

Keyword	Frequency
BJP	78642
Congress	68306
NDA	10906
UPA	3520
AAP	4321
TDP	452
TMC	563
BSP	431
SP	321
DMK	389

As parties rather than Congress and BJP do not share much amount of data, we have ignored the same for PNR Calculation.

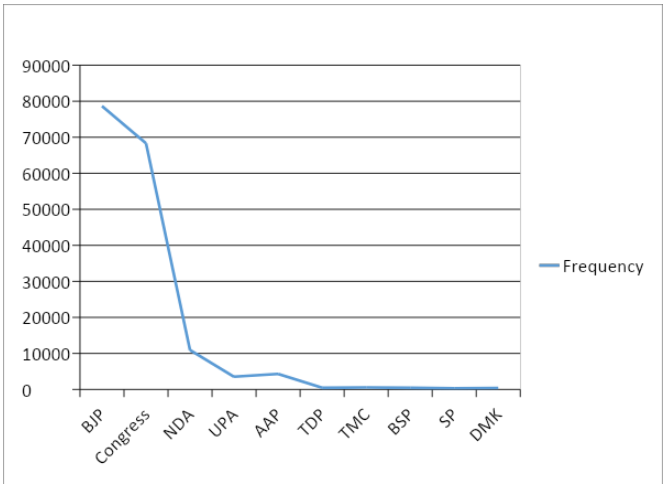

Figure 3 Frequency Graph for Important Terms

As stated above we have ignored other political parties and neutral contents, we have calculated the PR and NR ratio and founded the following result in table 4

Table 4 PR and NR Result of Proposed System

Party/Alliance	PR	NR
BJP + NDA	0.400388618	0.311173896
Congress + UPA	0.330715897	0.305738869

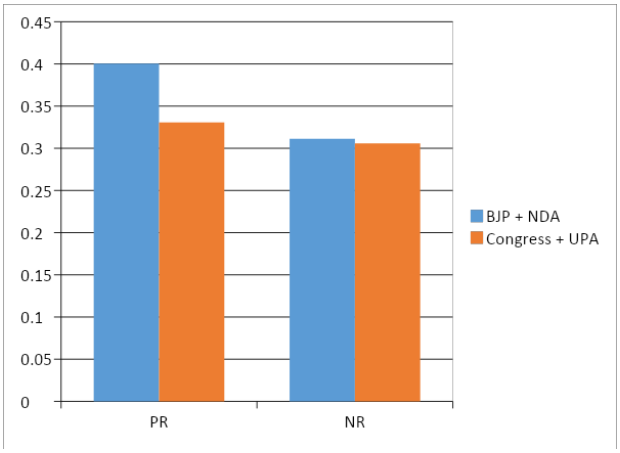

Figure 4 Comparative graph for PR and NR

Earlier system was not including negative feelings of users toward any political parties. In politics negative feelings toward any point or any political party plays an important role for opposite party. In Indian election system it is hard to decide to whom negative sentiments toward one party will benefit. So that Positive Ratio is also important to calculate.

IV. CONCLUSION AND FUTURE WORK

After analyzing the data we have founded that peoples sentiment were positive toward BJP and NDA (BJP Lead Alliance) .As positive ratio toward NDA and BJP Led alliance is 0.1 higher than opposition. But Negative ratio is also higher toward the same alliance compare to other alliance. We have also come to know that negative content for the same party and alliance is also higher compare to other parties. And election result is also in favor of same party so that we can summarize that generally people's sentiment on social media indicates similar direction to the actual result. In future we intend to detect paid or promoted contents in social media by various parties and find out real user data.

REFERENCES

[1] J. N. Franco-Riquelme, A. Bello-Garcia and J. Ordieres-Meré, "Indicator Proposal for Measuring Regional Political Support for the Electoral Process on Twitter: The Case of Spain's 2015 and 2016 General Elections," in IEEE Access, vol. 7, pp. 62545-62560, 2019, doi: 10.1109/ACCESS.2019.2917398.

[2] B. Joyce and J. Deng, "Sentiment analysis of tweets for the 2016 US presidential election," 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, 2017, pp. 1-4, doi: 10.1109/URTC.2017.8284176.

[3] B. Pang, L. Lee, and S. Vaithyanathan, ‘‘Thumbs up?: Sentiment classification using machine learning techniques,’’ in Proc. ACL Conf. Empirical Methods Natural Lang. Process. (EMNLP), vol. 10, 2002, pp. 79–86.

[4] B. Pang and L. Lee, Opinion Mining and Sentiment Analysis, vol. 2, nos. 1–2. Boston, MA, USA: Now, 2008.

[5] S. Kruikemeier, ‘‘How political candidates use Twitter and the impact on votes,’’ Comput. Human Behav., vol. 34, pp. 131–139, May 2014.

[6] S. Quinlan, M. Shephard, and L. Paterson, ‘‘Online discussion and the 2014 Scottish independence referendum: Flaming keyboards or forums for deliberation?’’ Elect. Stud., vol. 38, pp. 192–205, Jun. 2015.

[7] D. Jiang, X. Luo, J. Xuan, and Z. Xu, ‘‘Sentiment computing for the news event based on the social media big data,’’ IEEE Access, vol. 5, pp. 2373–2382, 2017

[8] Budiharto, W., Meiliana, M. Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *J Big Data* **5,** 51 (2018). https://doi.org/10.1186/s40537-018-0164-1

[9] P. Juneja and U. Ojha, "Casting online votes: To predict offline results using sentiment analysis by machine learning classifiers," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8203996.

[10] Ramteke, Jyoti & Shah, Samarth & Godhia, Darshan & Shaikh, Aadil. (2016). Election result prediction using Twitter sentiment analysis. 1-5. 10.1109/INVENTIVE.2016.7823280.

[11] K. Chauhan et al., "Automated Machine Learning: The New Wave of Machine Learning," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 205-212, doi: 10.1109/ICIMIA48430.2020.9074859.

[12] Vachhani H. et al. (2020) Machine Learning Based Stock Market Analysis: A Short Survey. In: Raj J., Bashar A., Ramson S. (eds) Innovative Data Communication Technologies and Application. ICIDCA 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 46. Springer, Cham

[13] Karimkhan, Mo, and Jitendra B. Bhatia. "Sentiment Analysis and Big Data Processing." *Computer Science and Electronic Journal* (2014): 136-142.