

Survey Paper on Sentiment Analysis and Detection of Indian Election 2019 on Tweets

Ruby Jaiswal*,

***¹(Department of CSE, Global Nature Care Sangathan's Group of Institutions, Jabalpur)**

Abstract

Social media usage is at an all-time high today the digital world. The vast majority of the population uses social media tools (Twitter, Facebook, YouTube, etc.) to share their thoughts and experiences with Society. It is important to analyze the feelings and opinions of ordinary people for government and business. This is the reason behind the fact that many media agencies have been active during the election period to conduct various types of opinion polls. We have worked on this paper analyzes the sentiments of the people of India during the Lok Sabha elections 2019 uses Twitter data for that period. We made an automatic tweet Analyzer that uses the transfer learning technique to handle unauthorized behaviour of this issue. We used the linear support vector classifier method in Machine Learning Model and Term Frequency Inverse Document Frequency (TF-IDF) method for handling text data of tweets. More, we increased the model's ability to address ridiculous tweets Posted by some users.

Keywords: Sentiment Analysis, Sarcasm Detection, Linear SVC, TF-IDF, Political Tweets.

Introduction

In today's world, to share our opinion with other people, we use online social networking platforms like Twitter, Facebook, YouTube, etc. Very often, people use these platforms to share their experiences about some incidents or some issues they are facing. The usage of these platforms allows the user to increase the range of audience to which they can share their thoughts, as compared to the offline method for opinion sharing. The analysis of the opinion of the common public is very important for a wide range of organisations, let it be a political party, businessman, investors, and working professionals, everyone tries to understand the opinion and trends getting followed by the public.

When it comes to politics and elections, many parties try to recognise the trend of opinions getting followed among the citizens. Even news channels and different media houses, perform various kinds of opinion polls and survey polls to understand the opinions and sentiments of the common person.

elections are a periodic activity that will occur regularly, even during the pandemic time, so to reduce the risk of the field workers of the polling agency for performing survey polls, sentiment analysis of political tweets is very significant. In this paper, we have used the political tweets of the Lok Sabha Elections – 2019 for getting the political sentiments of the people by performing the sentiment analysis techniques along with handling the sarcastic tweets which is not yet considered in the state-of-the-art works in this domain. Training and testing are done on Election – 2019 data where tweets are taken from the data science platform 'Kaggle' [2]. The paper presents the analysis of the tweets of the major political parties participating in the elections.

Literature Review

Pang and Lee [3] in 2008, developed an algorithm to mine the opinions of different people and analyse types of words.

Jhanwar and Das [4] proposed a technique for studying the sentiments of the people of India by analysing the texts written in mixed languages of Hindi-English.

Go et al. [5] tried to train a sentiment analysis algorithm to detect whether a tweet is a positive tweet about any specific subject by using emoticons. In the same domain

Pak and Paroubek [6] tried to expand the functionality by means of subjectively and objectively of different words which can create classifiers. These classifiers can be used to collect data to regulate a tweet's sentiment.

In [7], an open-source application called VADER [7] was introduced. This is a rule-based application for finding the sentiment scores for textual data.

Zhang et al. [8] examined the COVID-19 tweets in four cities of Canada and four cities of the USA. Sentiment intensity scores are examined through VADER [7] and NRC [9] methods and visualized the data during the pandemic.

Das and Bandyopadhyay [10] tried many approaches for concluding the sentiments behind various tweets. They used the approaches of interactive game, bilingual dictionary, and WordNet library.

In the literature [11] and [12], the authors have used a dictionary-based approach in which they created a small set of opinion words and then searched in a large collection of texts for growing the dictionary, in the same way as is done in WordNet [13].

Taboada et al. [14] also introduced a sentiment analysis method using the lexicon-based approach. The authors used a dictionary of positive or negative visualized words to analyse the tweets.

Bhadane et al. [15] reviewed various methods used to perform analysis on some natural language textual data, as per the sentiments expressed in that data, i.e. whether the text is of positive emotion or negative emotion.

Hamling and Agrawal [16] have analysed the 2016 US election related tweets. These tweets were used to find a correlation between the tweet sentiments and the election results. They used the Sentiwordnet library [17] for finding out the sentiment scores. Values for positive words were ranged from +0.0625 to +1.0; and -0.0625 to -1.0 for negative values. For instance, positive keyword like "helpful" would receive a value of +0.125 from the sentiment algorithm, or a strongly negative word like "unhappy" would receive a value of -0.25. In their algorithm, they were not able to handle the sarcastic tweets done by some of the users, since sarcasm is also a general form of expression of opinion, so the importance of sarcasm detection can't be overlooked in this domain.

In [18], the authors tried to extract the positive and negative sentiments and emotions of the people for the common political parties by computing a distance measure. It denotes the closeness of various tweets. More closely the tweets are for different political parties, it indicates the chances of a tough and close fight among the different parties. What people think is directly related to what they post over social media, many

investigations have proved to predict election results by performing sentiment analysis of Twitter data, such as, using the lexicon method for the Swedish elections [19].

Jose and Chooralil [20] have implemented a new method by using Twitter's streaming API for the data collection process. They tried to extract the sentiments and information from the tweets by using lexical tools like WordNet [13] and SentiWordNet [10]. Also, to increase their efficiency, they used a method for handling negation in the data pre-processing stage.

The usage of Twitter by politicians and their political campaigns has been a subject of interest for researchers. In the 2008 USA elections, the campaign routines of Barack Obama have increased the interest in the role of Twitter in political battles [21][22]. In [23] and [24], it is reported that many US Congress members in their Twitter venture posted their opinions about political issues on Twitter and the issues related to their election area.

Sharmistha Chatterjee [25] used crawling twitter data through API for performing the sentiment analysis of the two major parties BJP and INC. She used standard ML and Deep learning algorithms for mood classification of the two major parties. She crawled and merged the tweets every week for some months.

The WordCloud and N-gram Model [26] was used for Sentiment Representation. She also added an additional location mapping feature for the tweets and used the retweet frequency distribution.

Sharma and Moh [27] attempted to predict Indian election results using sentiment analysis on Hindi Twitter data. They fetched a total of 42,345 tweets in Hindi language. Then they performed data cleaning to remove irrelevant tweets and left with total 36,465 tweets, and after that labeled the data manually on these 36,465 tweets, making the unsupervised problem a supervised one. Then they used I Bayes, support vector machine on the Twitter data.

Gaikar and Sapare [28] also predicted the results of General Elections-2019 in India using the LSTM Neural Network approach. They used over 1500 labeled tweets, which were labeled with labels like positive, negative, and neutral for the training of their model. In real-time, they used the Twitter API to extract a total of 40,000 tweets from Jan 2019 to Mar 2019 related to elections to test their model. They also visualized their results using word clouds and compared them with ABP-C and India Today Survey results for the elections.

Ansari et al. [29] also performed a classification of the tweets related to the General Elections-2019 in India, they employed the LSTM model to perform the classification process of the Twitter text. They used the classification model to predict the inclination of tweets to infer the results of elections.

With respect to the Indian election, Sharma and Ghose [30] performed text mining using the named entity recognition to filter out the unrelated tweets. For performing sentiment analysis on the related tweets they used the model Rapid-Miner AYLIEN [31].

Naiknaware and Kawathekarm [32] used the Sentiment analysis score method of R programming language to execute the sentiment analysis on the election related tweets for General Election 2019 of India.

Bose et al. [33] in the context of political tweets, used the NRC emotion lexicon approach for finding the overall tone of the event. Then they used the deep learning tool ParallelDots that can categorize the tweets into positive, negative, and neutral categories.

A strategy known as Adaptive Neuro-Fuzzy Inference System (ANFIS) is proposed by Katta and Hegde [34]. The Fuzzy based ontology is made by implementing Non-Linear SVM classifier analysis to improve the fuzzy principles. They concluded that an ANFIS Non-linear SVM-based model for sentiment analysis of social media text is less complex and provides high accuracy.

In [35], Bansal and Srivastava used a Lexicon-based approach for Twitter sentiment analysis for the vote share prediction using the emojis and n-gram features.

Hitesh et al. [36] performed real-time sentiment analysis of the 2019 General elections in India using Word2vec and Random Forest Model

Joseph [37] used the Decision Tree classifier approach for predicting the outcomes for Indian general elections-2019. He considered the tweets of the English language. His approach was to perform the mood mapping of people over a timely basis during different phases of elections

Prediction of Indonesia's election results was performed by Kristiyanti et al. [38] using the Support Vector Machine (SVM) with selection features of Particle Swarm Optimisation (PSO) and Genetic Algorithm (GA). They tried to perform predictions for the post of President and Vice president of Indonesia.

Hidayatullah et al. [39] also performed the prediction of Indonesia's election results using the deep learning approach using various algorithms like Convolutional Neural Network (CNN), Long short-term memory (LSTM), CNN-LSTM, Gated Recurrent Unit (GRU)-LSTM, and Bidirectional LSTM. They compared the results with various traditional machine learning algorithms and concluded that the Bidirectional LSTM achieved the best accuracy.

The various methodologies have been introduced for opinion and sentiment mining, but all these are broadly classified into 2 major groups. One is a Machine learning approach and the second one is a Lexicon-based method, a linguistically-inclined method [40]. In this paper, we have used the Machine learning approach for opinion mining. There are two significant types of machine learning algorithms, supervised learning and unsupervised learning. The supervised learning approach requires labeled data for the particular domain targeted to build a machine learning model. In contrast, the unsupervised approach is used when there is no labeled data for training the model. Sentiment analysis of tweets lies in the category of unsupervised learning since we do not have any assigned labels to the 2019 Loksabha election-related tweets dataset. To solve this lack of having the desired labeled dataset, we have used the transfer learning approach. Transfer learning is a type of approach for solving unsupervised problems. In this approach, the model is trained using some other problems' dataset (labeled dataset), and then the trained model is used for predictions on some new dataset which is not having any class labels attached to it. Although several works have been done on sarcasm detection [41], none of them have been performed in the domain of political tweets, so, this paper contributes in that direction. It applies sarcasm detection in the domain of election related tweets.

Proposed Methodology

We have tried to counter the insecure nature of analyzing the sentiments of election-2019 tweets using transfer learning. We used the standard Twitter review dataset available on Kaggle [2] to train our model using Linear SVC and then analyze sentiment on the actual data. Since data is a textual format, we have used the TFIDF method to create the term frequency inverse document frequency matrix to deal with textual data.

Conclusion

In this paper proposed an idea of transfer learning for handling sarcastic tweets and analysing the tweets for positive and negative polarities. We also observed that our trained models are work properly or not. we compared our model's results with the actual election results. The preparation of the annotated corpus from Twitter automatically for political sentiment analysis of national political parties of India with respect to the forthcoming general elections in India is carried. Inspection of the sentiment keywords pertaining to political preferences in the tweets is considered during annotation in order to mark class labels. Several significant features based on term frequency inverse document frequency are extracted from the corpus to examine the various learning models. The data analysis of the acquired corpus reflects the presence of a single party on Twitter which is more than fifty percent among all the extracted politically motivated tweets. The evaluation of the annotated corpus over itself using LSTM and several machine learning models are obtained and promising results are achieved with LSTM and Random Forests.

References

1. Mohamed Ridhwan, K., Hargreaves, C.A.: Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *Int. J. Inf. Manag. Data Insights.* 1, 100021 (2021). <https://doi.org/10.1016/J.JJIMEI.2021.100021>.
2. Indian Political Tweets 2019 (Feb to May) | Kaggle.
3. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. *Found. Trends® Inf. Retr.* 2, 1–135 (2008). <https://doi.org/10.1561/15000000011>.
4. Jhanwar, M.G., Das, A.: An Ensemble Model for Sentiment Analysis of Hindi-English Code-Mixed Data. (2018).
5. Go, A., Bhayani, R., Huang, L.: Twitter Sentiment Classification using Distant Supervision.
6. Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. undefined. (2010).
7. Hutto, C., Gilbert, E.: VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proc. Int. AAAI Conf. Web Soc. Media.* 8, 216–225 (2014).
8. Zhang, Q., Yi, G.Y., Chen, L.-P., He, W.: Text mining and sentiment analysis of COVID19 tweets. (2021).
9. Mohammad, S.M., Kiritchenko, S., Zhu, X.: NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. **SEM 2013 - 2nd Jt. Conf. Lex. Comput. Semant.* 2, 321–327 (2013).
10. Das, A., Bandyopadhyay, S.: SentiWordNet for Indian Languages. 21–22 (2010).
11. Kim, S.-M., Hovy, E.: Determining the sentiment of opinions. 1367-es (2004). <https://doi.org/10.3115/1220355.1220555>.
12. Emotions in social psychology: Essential readings. - PsycNET.
13. WordNet | A Lexical Database for English.

14. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-Based Methods for Sentiment Analysis. *Comput. Linguist.* 37, 267–307 (2011). https://doi.org/10.1162/COLI_A_00049.
15. Bhadane, C., Dalal, H., Doshi, H.: Sentiment Analysis: Measuring Opinions. *Procedia Comput. Sci.* 45, 808–814 (2015). <https://doi.org/10.1016/J.PROCS.2015.03.159>.
16. Hamling, T., Agrawal, A.: Sentiment Analysis of Tweets to Gain Insights into the 2016 US Election.
17. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, (2010).
18. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting Elections with 16 Twitter: What 140 Characters Reveal about Political Sentiment. (2010).
19. Liu, B.: Sentiment Analysis and Opinion Mining. <http://dx.doi.org/10.2200/S00416ED1V01Y201204HLT016>. 5, 1–184 (2012). <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>.
20. Jose, R., Chooralil, V.S.: Prediction of election result by enhanced sentiment analysis on Twitter data using Word Sense Disambiguation. 2015 Int. Conf. Control. Commun. Comput. India, ICCCI 2015. 638–641 (2016). <https://doi.org/10.1109/ICCCI.2015.7432974>.
21. Abrams, L.C., Lefebvre, R.C.: Obama’s Wired Campaign: Lessons for Public Health Communication. <http://dx.doi.org/10.1080/10810730903033000>. 14, 415–423 (2009). <https://doi.org/10.1080/10810730903033000>.
22. Jarvis, S.E.: Communicator-in-Chief: How Barack Obama Used New Media Technology to Win the White House – Edited by John Allen Hendricks and Robert Denton, Jr. *Pres. Stud. Q.* 40, 800–802 (2010). <https://doi.org/10.1111/J.1741-5705.2010.03815.X>.
23. Glassman, M.E., Straus, J.R.: CRS Report for Congress Social Networking and Constituent Communication: Member Use of Twitter During a Two-Week Period in the 111 th Congress. (2009).
24. Golbeck, J., Grimes, J.M., Rogers, A.: Twitter use by the U.S. Congress. *J. Am. Soc. Inf. Sci. Technol.* 61, 1612–1621 (2010). <https://doi.org/10.1002/ASI.21344>.
25. Twitter Sentiment Analysis for the 2019 Lok Sabha Elections | Hacker Noon.
26. Cavnar, W.B., Trenkle, J.M.: N-Gram-Based Text Categorization.
27. Sharma, P., Moh, T.S.: Prediction of Indian election using sentiment analysis on Hindi Twitter. *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016.* 1966–1971 (2016). <https://doi.org/10.1109/BIGDATA.2016.7840818>.
28. Gaikar, D., Sapare, G., Vishwakarma, A., Parkar, A., Professor, A.: Twitter Sentimental Analysis for Predicting Election Result using LSTM Neural Network. *Int. Res. J. Eng. Technol.* (2019).
29. Ansari, M.Z., Aziz, M.B., Siddiqui, M.O., Mehra, H., Singh, K.P.: Analysis of Political Sentiment Orientations on Twitter. *Procedia Comput. Sci.* 167, 1821–1828 (2020). <https://doi.org/10.1016/J.PROCS.2020.03.201>.
30. Sharma, A., Ghose, U.: Sentimental Analysis of Twitter Data with respect to General Elections in India. *Procedia Comput. Sci.* 173, 325–334 (2020). <https://doi.org/10.1016/J.PROCS.2020.06.038>.

31. Das, S., Kolya, A.K.: Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm. Proc. - 2017 3rd IEEE Int. Conf. Res. Comput. Intell. Commun. Networks, ICRCICN 2017. 2017-Decem, 239–244 (2017). <https://doi.org/10.1109/ICRCICN.2017.8234513>.
32. Naiknaware, B.R., Kawathekar, S.S.: Prediction of 2019 Indian Election using sentiment analysis. Proc. Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2018. 660– 665 (2019). <https://doi.org/10.1109/I-SMAC.2018.8653602>.
33. Bose, R., Dey, R.K., Roy, S., Sarddar, D.: Analyzing Political Sentiment Using Twitter Data. Smart Innov. Syst. Technol. 107, 427–436 (2019). https://doi.org/10.1007/978-981-13-1747-7_41.
34. Katta, P., Hegde, N.P.: A Hybrid Adaptive Neuro-Fuzzy Interface and Support Vector Machine Based Sentiment Analysis on Political Twitter Data. Int. J. Intell. Eng. Syst. 12, (2019). <https://doi.org/10.22266/ijies2019.0228.17>.
35. Bansal, B., Srivastava, S.: Lexicon-based Twitter sentiment analysis for vote share prediction using emoji and N-gram features. Int. J. Web Based Communities. 15, 85–99 (2019). <https://doi.org/10.1504/IJWBC.2019.098693>.
36. Hitesh, M.S.R., Vaibhav, V., Kalki, Y.J.A., Kamtam, S.H., Kumari, S.: Real-time sentiment analysis of 2019 election tweets using word2vec and random forest model. 2019 2nd Int. Conf. Intell. Commun. Comput. Tech. ICCT 2019. 146–151 (2019). <https://doi.org/10.1109/ICCT46177.2019.8969049>.
37. Joseph, F.J.J.: Twitter Based Outcome Predictions of 2019 Indian General Elections Using Decision Tree. Proc. 2019 4th Int. Conf. Inf. Technol. Encompassing Intell. Technol. Innov. Towar. New Era Hum. Life, InCIT 2019. 50–53 (2019). <https://doi.org/10.1109/INCIT.2019.8911975>.
38. Kristiyanti, D.A., Normah, Umam, A.H.: Prediction of Indonesia presidential election results for the 2019-2024 period using twitter sentiment analysis. Proc. 2019 5th Int. Conf. New Media Stud. CONMEDIA 2019. 36–42 (2019). <https://doi.org/10.1109/CONMEDIA46929.2019.8981823>.
39. Hidayatullah, A.F., Cahyaningtyas, S., Hakim, A.M.: Sentiment Analysis on Twitter using Neural Network: Indonesian Presidential Election 2019 Dataset. IOP Conf. Ser. Mater. Sci. Eng. 1077, 012001 (2021). <https://doi.org/10.1088/1757-899X/1077/1/012001>.
40. Mehta, P., Pandya, S.: A Review On Sentiment Analysis Methodologies, Practices And Applications. Int. J. Sci. Technol. Res. 9, 2 (2020).
41. JoshiAditya, BhattacharyyaPushpak, J., C.: Automatic Sarcasm Detection. ACM Comput. Surv. 50, (2017). <https://doi.org/10.1145/3124420>.