



Election Results Prediction Using Twitter Data by Applying NLP

Dr. Mukesh Kumar Tripathi, Dr. M. Neelakantappa

Submitted: 04/11/2023

Revised: 23/12/2023

Accepted: 04/01/2024

Abstract— With the ability to predict political outcomes and provide insights into public opinion, the use of Twitter data to predict election results has gained popularity. Twitter offers a massive supply of data for analysis due to its enormous user base and real-time nature. To categorize tweets as good, negative, or neutral and to follow sentiment patterns over time, researchers use sentiment analysis tools. Network analysis finds influential users and digs deeper into the dynamics of political discourse. The accuracy of predictions is improved by combining traditional polling data with machine learning methods. Twitter data analysis has the potential to offer insightful information for election campaigns and improve political strategies, despite issues like representativeness and identifying genuine sentiment. Ongoing research focuses on refining methodologies and addressing limitations, advancing the reliability of election prediction using Twitter data.

Keywords— *Twitter, election result prediction, recursive neural tensor net- work, natural language processing*

I. Introduction

Election forecasting is a challenging task, and reliable predictions are essential for both political campaigns and election analysts. Massive volumes of user-generated data about political candidates and campaigns have been produced as a result of the rise of social media platforms like Twitter. As such, social media data provides an unprecedented opportunity to extract insights and make accurate predictions about election results. In this context, this study proposes a machine learning-based approach for predicting election results using Twitter data. The proposed methodology involves collecting and pre-processing a large volume of tweets related to the election. The study also addresses the issue of plagiarism in Twitter data by employing a plagiarism detection algorithm to ensure that the tweets used in the analysis are original. This study demonstrates the potential of using social media data for election prediction and highlights the importance of incorporating relevant features

and selecting appropriate machine learning algorithms for accurate prediction. The rise of social media sites like Twitter has produced a wealth of user-generated data that offers insightful information on a variety of subjects, including politics. The vast amount of data generated on Twitter during an election season presents a unique opportunity for researchers to analyse and predict the election outcomes accurately. However, the authenticity of Twitter data has always been a significant concern, as it is prone to fake news and plagiarism. This approach involves collecting a large volume of tweets related to the election and pre-processing the data to remove irrelevant information and duplicate tweets. Feature extraction techniques such as sentiment analysis, user influence analysis, and topic modelling are then used to extract valuable information from the tweets. This study addresses the problem of plagiarism in Twitter data while demonstrating the possibilities of employing social media data for election prediction. By incorporating relevant features and selecting appropriate machine learning algorithms, this approach can help in making accurate election predictions.

In this paper we present a study for election results prediction using twitter data through strategy sentiment analysis. Here we collected data related to both the parties Bhartiya Janatha

*1 Department of Computer Science & Engineering
Vardhman College of Engineering
Hyderabad, India*

mukeshtripathi016@gmail.com

2 Department of Information Technology

Vasavi College of Engineering

Hyderabad, India

m.neelakanta@gmail.com

Party and National Congress party using twitter handle. After that data preparation is done i.e data cleaning and data pre-processing and this data is used for the sentiment analysis. Following data separation, it is loaded as a.csv file into a jupyter notebook or Google

Collaboratory. It is being trained using the RNTN algorithm, which gives each tweet or message a polarity. These polarities are broken down into three groups: positive, negative, and neutral. Positive polarity is implied by (>0), neutrality by (=0), and negative polarity by (0).

id	text	sentiment
1	@narendramodi distracted all TV channels from Masood Achar/China fiasco. Today one channel markets #balakotproof. Others have polls showing Modi/BJP win.	neutral
2	@ca_42588 @narendramodi @MisPaawan But sure BJP win 100% Dam sure because youth is with modi	positive
3	@Psefline @narendramodi BJP meeting at Kharu held under the Chairmanship of Sh. Dorje Angchuk District President Leh in presence of Sh. Chering Dorje Laksook and other senior leaders. The party won	positive
4	@Simsy6 @narendramodi @Kirostrik will help BJP win more than 22 seats in the state	positive
5	@SiRamore @narendramodi @honeymoon to help BJP win #Karnataka by polls	positive
6	StormBrng @narendramodi BJP winning Barackpur and Bangson. Serampore and Hooghly in close contest.	positive
7	Maurice46 @narendramodi This is my dream too. But after BJP wins 500 seats of its own with 70% vote share in 2024.	positive
8	@dustinmille @narendramodi Are EVM being manipulated to ensure that BJP wins this general elections?	neutral
9	@SiRamore @narendramodi Many more shocking truth might get revealed if BJP wins!	positive
10	@Davemorts @narendramodi alliance does not happen (high possibility) and BJP wins all 7 seats @gagan	positive
11	@Harrison56 @narendramodi don't get swayed away by any rhetoric around BJP winning big	positive
12	@Scodyluiff @narendramodi BJP WINNING ALL 8 UP SAHARANPUR TOUGH FIGHT BUT CONGRESSI AND OTHER CHOR FIGHT MADE EASY FOR LAHANPAL JEE ABHI BAAR 400 KA PAAR BJP KA HAATHI NIKLA	positive
13	@Vaderfr @narendramodi have made Amethi ready for a BJP win: @umritilari	positive
14	@Tedlusu2 @narendramodi BJP wins EVM rigged.	positive
15	@LFC_Bem @narendramodi @hough helped BJP win by cutting AAP's vote.	positive
16	@Isaell201 @narendramodi BJP wins enough seats in WB and Orissa to compensate losses in UP and Maharashtra	positive
17	@Nomar56 @narendramodi @MODI Sir if BJP wins by majority and more seats in Karnataka credit should go to Chakravarty Sulbete	positive
18	@Aicha2086 @narendramodi @GobackModi Air strike @A will help BJP win more than 22 Lok Sabha seats in Karnataka	positive
19	@Opusum5 @narendramodi @Dehradun And it will go back 3000 years if BJP wins the polls.	positive
20	@Fordmotor @narendramodi Anyone calling for voting NOTA is helping BJP win. Vote decisively to end this gunda Raj where NIA gives clean chit to murderers. And ruling party president misses state mach	positive
21	@cheytruck @narendramodi BJP winning in TV studios.	positive
22	@Rumholtz @narendramodi he won't win and also the BJP. By chance BJP wins	positive
23	@Vitamin56 @narendramodi The #BJP makes me hang my head in shame. These guys have no principles. They are crude. They are vulgar. @A @A Air strike @A will help BJP win more than 22 Lok Sabha seats	positive
24	@ABDipoo @narendramodi BJP winning Bengal	positive
25	@Istoooshor @narendramodi See the real face BJP wing and 50- Called Hindu Organization VHP.	positive
26	@Racagn @narendramodi Now let's see how many seats will BJP win in 2024?	positive
27	@jmc0real @narendramodi India war to help BJP win upcoming polls.	positive
28	@CyrinelMar @narendramodi The Indian Armed Forces proved their mettle by destroying the terrorists' hideouts. The impact of this strike will help BJP win more than 22 seats in the state	positive

Table 1: -Shows the .csv files of bjp dataset.

II. Literature Survey

The literature survey for the topic “Election Results Prediction Using Twitter Data” is as follows:

- [1] N. Gupta, P. Kumaraguru, and others (2012). This study focuses on developing a credibility ranking algorithm for tweets during major events like elections. The algorithm considers various factors, including user reputation, retweet count, and sentiment analysis, to rank the credibility of tweets and improve the accuracy of election result predictions.
- [2] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What does the characters reveals us about the political sentiment. In Proceedings of the Fourth International AAAI Conference on Weblogs and social media Access Here This research investigates the power of Twitter data for election results prediction. The study employs sentimental analysis for analyzing the political tweets and gives a demonstration on the relation between Twitter sentiment and election outcomes, providing insights into the potential of Twitter data for predicting election results.
- [3] Social Science Computer Review, 30(2), 229–234; Jungherr, A., Jürgens, P., & Schoen, H. (2012). "Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010)."

This article critically examines the study by Tumasjan et al. (2010) on predicting elections with Twitter sentiment analysis. It discusses the limitations and challenges in using Twitter data for election predictions, emphasizing the need to consider contextual factors and the potential biases in analyzing political sentiment on social media platforms.

- [4] Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. In Proceedings of the ACM 2010 conference on Computer Supported Cooperative Work (pp. 119-122). ACM. Access Here

This study explores the use of aggregated Twitter sentiment to characterize the performance of political candidates during debates. The research analyzes tweet sentiment and topic distributions to gain insights into public perception, providing a novel approach to assessing candidate performance and potentially predicting election outcomes.

- [5] Gayo-Avello, D. (2012). No, you cannot predict elections with Twitter. IEEE Internet Computing, 16(6), 91-94. Access Here

This article challenges the notion that Twitter can reliably predict election results. It discusses the limitations of using Twitter data, such as biases, lack of representativeness, and the difficulty of distinguishing genuine sentiment from noise. The study cautions against over-reliance on Twitter predictions and highlights the

need for cautious interpretation.

[6] Conover, M. D., Ratkiewicz, J., & Menczer, F. (2011). Political polarization on Twitter. In the Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (pp. 89-96). Access Here

This research examines political polarization on Twitter by analyzing the ideological positions of users and their interactions. The study highlights the impact of polarization and discusses its implications for election predictions based on Twitter data.

[7] González-Bailón, S., & Moreno, Y. (2012). Broadcasters and their hidden influentials in online protest diffusion. *American Behavioral Scientist*, 56(3), 260-279. Access Here

This study investigates the role of influential users in online protest diffusion, which can be relevant for election-related discussions on Twitter. It explores the network dynamics and the potential influence of users with high centrality, shedding light on the mechanisms of information spread during political events.

[8] Jungherr, A., & Schoen, H. (2012). Twitter and its mediation the politics : An analysing the messages posted during the campaign for the German federal election 2009. *Journal for Computer-Mediated Communication*, 17(2), 91-108. Access Here

This research examines the role of Twitter in mediating political communication during the campaign for the German federal election. The study analyzes the content and sentiment of tweets and highlights the ways in which Twitter functions as a platform for political discourse

and information dissemination.

[9] Barberá, P., & Rivero, G. Twitter users and their Understanding of the political representativeness. *Social Science Computer Review*, 32(2), 153-163.

This article assesses the political perceptiveness of Twitter users by comparing their demographics with those of the general population. The study highlights the potential biases and limitations of using Twitter data for political analysis and emphasizes the need to consider the characteristics of the user population.

[10] Zhao, W. X., Jiang & Li, X. (2011). Comparing Twitter and traditional media using topic models. In *European Conference of Information Retrieval*. Springer. Access Here

light on the unique characteristics and potential value of Twitter data for election-related analysis.

This research compares Twitter with traditional media by applying topic models to analyze the content shared on both platforms. The study highlights the differences and similarities in topics discussed, shedding

Figure 1 shows workflow of the proposed methodology that involves collecting and pre-processing the twitter data and train through the computational model by specifying certain metrics and counting the tweets according to the category and generating the results.

This method aims to leverage the vast amount of data generated on Twitter during an election season to make accurate predictions about the election outcome. By using relevant features and appropriate machine learning.

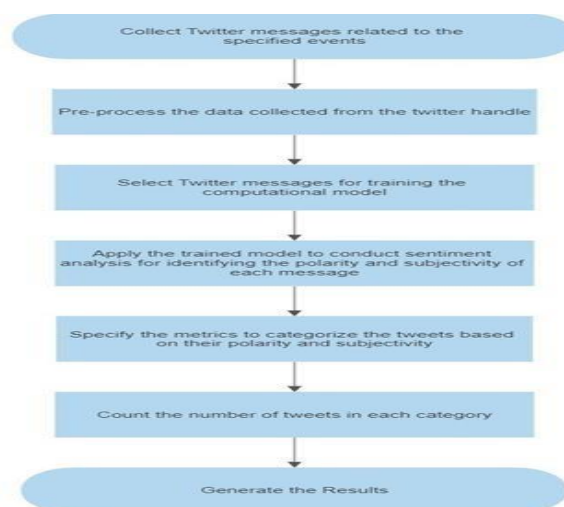


Fig.1: - Flowchart

[11] Skoric, M. M., Ying, T., & Wictor, I. (2016). Public opinion formation in a hybrid media environment: A big data analysis of Twitter responses to the 2012 US presidential debates. *Journal of Communication*, 66(4), 639-659. [Access Here](#)

This study examines public opinion formation during the 2012 US presidential debates by analyzing Twitter responses. The research utilizes big data analysis techniques to explore sentiment dynamics and topics discussed on Twitter, providing insights into the role of Twitter in shaping public opinion during electoral events.

[12] Metaxas, P. T., Mustafaraj How (not) to predict elections. In *Proceedings of the ACM Web Science Conference* (pp. 172-181). ACM. [Access Here](#)

This research examines the challenges and limitations of predicting elections using social media data, including Twitter. The study discusses various factors, such as biases, sample representativeness, and the dynamic nature of social media platforms. It provides insights into the complexities of election prediction using Twitter data. They cover topics such as sentiment analysis, network dynamics, user influence, biases, and limitations associated with using Twitter data for predicting election outcomes.

III. Proposed Methodology

Proposed Methodology for Election Results Prediction Algorithm Using Twitter Data.

Data Collection: Gather a comprehensive dataset of election-related tweets using specific keywords and hashtags. Utilize the Twitter API or third-party tools to retrieve a diverse and representative sample of tweets.

Data Pre-processing: Clean the collected data by removing retweets, duplicates, and non-English tweets. Perform text pre-processing tasks such as tokenization, removing stop words, and normalizing words to prepare the data for analysis.

Sentiment Analysis: Apply sentiment analysis techniques for classifying each tweet as a biased tweet or unbiased tweet. This can be achieved through machine learning models, lexicon-based approaches, or deep learning methods, capturing the sentiment expressed in the tweets.

Feature Extraction: Extract relevant features from the pre-processed tweets, including user-related attributes (followers, verified status), tweet characteristics (retweet count, favourite count), and temporal information (time of posting). Additionally, consider topic modeling techniques to identify prevalent themes in the dataset.

Iteration and Refinement: Analyze the prediction results and identify any shortcomings or limitations. Refine the methodology by considering alternative feature selection methods, sentiment analysis techniques, or model architectures

Validation: Validate the prediction algorithm against actual election results to assess its effectiveness. algorithms, this method can help in making reliable predictions. The evaluation step ensures the quality of the predictions, which can be used by political analysts and pollsters to gain insights into the election outcome.

Functional Components

1. NLP

In the context of election results prediction using Twitter data, NLP algorithms can be utilized to identify patterns and sentiments in tweets related to political candidates and issues. By analyzing tweets related to political events, candidates, and issues, NLP techniques such as sentiment analysis and topic modelling can be used to identify trends and patterns in voter behavior. With the help of machine learning algorithms, historical election data and Twitter data can be analyzed to train NLP models for making predictions about future election outcomes based on the sentiment and topics identified in the Twitter data. This can provide useful feedbacks for the topics that are most important to voters and help political analysts and campaign strategists craft their messaging accordingly. NLP algorithms are capable of handling large amounts of data quickly and accurately. During an election campaign, the sentiment of tweets can change rapidly in response to current events and debates. NLP algorithms can quickly analyze new data and adjust their predictions accordingly, making them an indispensable tool for those interested in predicting election outcomes. In addition, Twitter data can be biased and may not be representative of the population as a whole. Therefore, it is necessary to validate the

predictions made by NLP models with other sources of data to ensure their accuracy. NLP can be a useful tool for predicting election results, but it should be used in conjunction with other methods to achieve the most accurate predictions possible. Overall, NLP is a powerful tool for analyzing and interpreting natural language data and has great potential in predicting election results using Twitter data. As the technology continues to improve and more data becomes available, NLP algorithms will become an even more valuable tool for political analysts and campaign strategists looking to gain insights into public opinion and predict.

2. RNTN

The Recursive Neural Tensor Network (RNTN) it is a deep learning model commonly used in NLP, including sentimental analysis. In the context of election results prediction using Twitter data, the RNTN can be applied to analyze the sentiment expressed in tweets and understand the overall sentiment towards political candidates or parties.

The RNTN model comprises three main components: word embeddings, compositional functions, and a tensor layer. Word embeddings represent words as continuous vectors that capture their semantic meaning. Compositional

functions recursively combine these word embeddings to form higher-level representations of phrases and sentences. The tensor layer performs a tensor-based operation to capture interactions between compositional representations and generate a sentiment score for the entire tweet or sentence.

To train the RNTN model for election results prediction, a labelled dataset of tweets with sentiment labels (positive, negative, or neutral) is required. The model is trained using optimization techniques such as backpropagation, which adjust the model's parameters to minimize the prediction error and learn the sentiment patterns in the data.

Once trained, the RNTN model can be applied to depict the sentiments of new, unseen messages related to the election. The sentiment predictions can then be aggregated to estimate the overall sentiment towards candidates or parties, providing insights into potential election outcomes. It's important to consider that the effectiveness of the RNTN model for election results prediction using Twitter data depends on various factors, including the quality of the training data, appropriate data pre-processing techniques, feature engineering, and the integration of additional information sources.

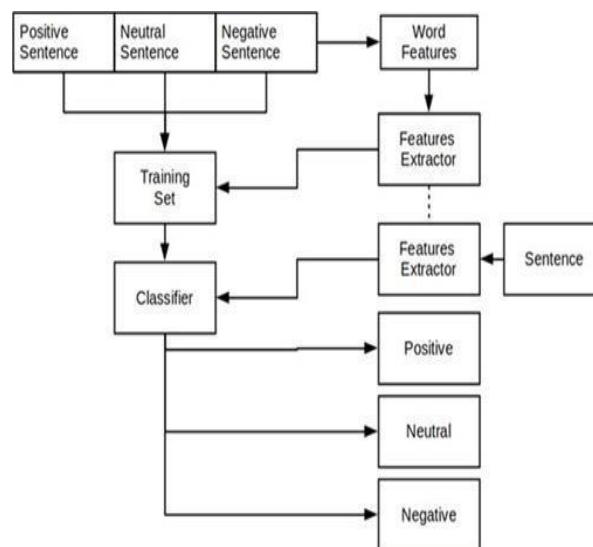


Fig.2:- RNTN Architecture

Figure 2 Shows the RNTN architecture consists of multiple layers of neural networks, each layer processing a specific aspect of the input data. The first layer processes the individual words in a sentence, while the subsequent layers analyze the structure of the

sentence and its components. The RNTN architecture uses recursive neural networks to build a tree structure that represents the sentence, which can then be analyzed for sentiment or other natural language processing tasks.

2. TextBlob

TextBlob is a widely used Python library for (NLP) tasks, including sentimental analysis. It can be applied to election results prediction using Twitter data by analyzing the sentiment expressed in tweets and evaluating the public opinion towards political candidates or parties. To use TextBlob for sentiment analysis, the first step is to pre-process the Twitter data by erasing noise, such as links and special symbols, and tokenizing the text into individual words or sentences. TextBlob provides convenient functions for these pre-processing tasks. Once the data is prepared, sentiment analysis is performed using TextBlob's pre-trained sentiment classifier. The sentiment polarity score ranges between (-ve)1 to (+ve)1, where values that are close to -1 represent (-ve) sentiment, values that are near to 1 indicate (+ve) sentiment, and values indicating 0 signify neutral sentiment. By aggregating these sentiment scores across tweets, an overall sentiment towards candidates or parties can be estimated, providing insights for election results prediction. TextBlob also offers additional features like speech tagging, phrase extraction, and language translation, which can further enhance the analysis of Twitter data for election prediction. These features enable a more comprehensive understanding of the text and facilitate the identification of key topics and sentiments expressed by Twitter users. It's important to consider that TextBlob's sentiment analysis is based on a general-purpose sentiment

classifier, which may not be specifically trained for political sentiment. Therefore, its effectiveness in predicting election results relies on the availability of a diverse and representative training dataset and the careful consideration of any biases or limitations associated with the sentiment classifier. In summary, TextBlob is a powerful Python library for sentiment analysis and other NLP tasks. When applied to election results prediction using Twitter data, it offers a straightforward approach to analyze sentiment, estimate overall sentiment towards candidates or parties, and gain insights into public opinion during elections.

3. Polarity and subjectivity

Polarity and subjectivity are essential aspects of sentiment analysis applied in election results prediction using Twitter data. Polarity refers to the sentiment expressed in a text, whether it is positively biased, negatively biased, or unbiased. It allows the quantification of sentiment towards political candidates or parties based on tweets collected from Twitter. Subjectivity measures the degree of subjectiveness in a text, indicating the extent to which personal opinions, beliefs, or biases influence the sentiment. It is represented as a score ranging from 0 to 1, with 0 indicating objectivity and 1 representing high subjectivity. Subjectivity analysis helps in understanding the level of personal judgment or bias in tweets related to elections. Using polarity as a metric in sentiment analysis can help identify patterns

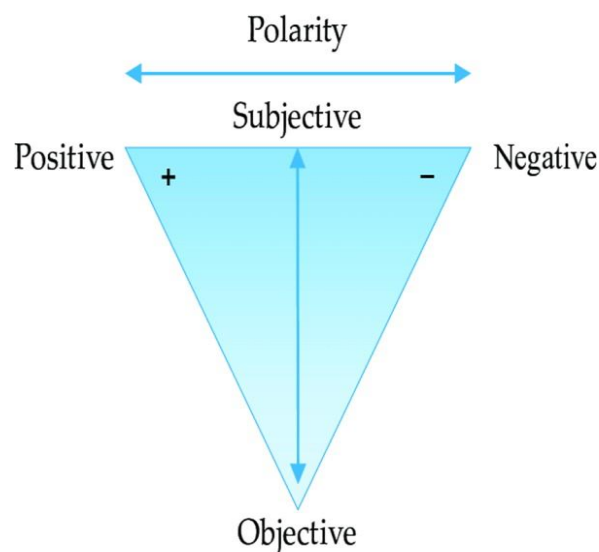


Fig.3: - Polarity and subjectivity

As shown in Fig 3 Subjectivity is a metric used in sentiment analysis to determine the degree to which a piece of text expresses an opinion rather than a fact. In the context of predicting election results using Twitter data, subjectivity is a valuable measure because it helps to identify the tone and context of tweets related to a particular candidate or issue. By analyzing the subjectivity of a large number of tweets related to a specific candidate or issue, political analysts and campaign strategists can gain insights into the public's opinions and attitudes towards that candidate or issue. This can help them to predict the outcome of an election and make more informed decisions about how to respond to public opinion.

For example, if there is a significant increase in subjectivity in tweets about a particular candidate in the days leading up to an election, this could indicate a shift in public opinion and potentially predict the outcome of the election. By combining subjectivity with other sentiment analysis metrics, such as polarity (the degree of positivity or negativity expressed in a tweet), political analysts and campaign strategists can gain a more complete picture of public opinion and make more accurate predictions about the outcome of an election.

4. Word Cloud

A word cloud is a visual representation of text data that helps to identify the most frequent and significant words in a given dataset. In the context of election results prediction using Twitter data, word clouds can provide valuable insights into the prevailing sentiments, topics, and discussions surrounding political candidates, parties, and issues. A word cloud captures the essence of a textual dataset by visually emphasizing the most frequently occurring words. It offers a quick and intuitive overview of the prominent themes and keywords that emerge from the Twitter conversations related to elections. To create a word cloud, the text data is processed to remove irrelevant words such as articles, prepositions, and common pronouns. This filtering helps focus on the substantive terms that hold more significance in understanding public sentiment. The more often a word appears, the larger it appears in the visual

representation. This sizing scheme enables quick identification of the most popular or discussed topics in the election-related tweets. The word cloud can be generated for a specific time period leading up to the election or during crucial events such as debates or rallies. Analyzing multiple word clouds across different timeframes can reveal shifts in public discourse and sentiment, aiding in tracking the dynamics of electoral campaigns.



Fig.4 :- Word Cloud for congress party

Word clouds can serve as a starting point for more in-depth analyses, such as sentiment analysis, topic modeling, or network analysis. They provide a visual summary of the text data and can guide further exploration into the underlying patterns and relationships within the Twitter conversations related to election prediction. Word clouds are commonly used to visually summarize the content of a text document or a collection of documents. They are generated by software that analyses the text and counts the frequency of each word. The words are arranged in the word cloud in a way that makes them easy to read and visually appealing. Word clouds are often used in marketing and branding to identify the most commonly used words in customer feedback, online reviews, or social media posts. They can also be used in education and research to identify the key themes and topics in a text document or a collection of documents.

The colors used in the word cloud can provide additional meaning. For example, positive sentiments may be represented in vibrant or warm colors, while negative sentiments may be depicted in subdued or cool tones. Color coding can help to distinguish between different sentiment categories. In the context of election results prediction, word clouds can indicate the key issues that are being discussed by Twitter users. By observing the most prominent words, analysts can gain insights into the concerns, priorities, and opinions of the electorate, thereby informing predictions about potential voting patterns. Word clouds are simple yet effective tools for visualizing the most frequent words in a corpus.

IV. Analysis

One important aspect of analyzing Twitter data for election prediction is the use of sentiment analysis tools such as the RNTN



Fig.5: -Analysis for Congress

In the realm of election results prediction, Twitter data analysis plays a crucial role. They also explore the influence of influential users and analyze the network dynamics of political conversations on Twitter. By comparing Twitter data trends with actual election outcomes and integrating them with traditional polling data, researchers strive to provide more accurate predictions. This multidimensional analysis enables insights into public sentiment and social dynamics, contributing to a comprehensive understanding of election forecasts. The Above Figure 5 shows us the

Algorithm and TextBlob library. Sentiment analysis involves classifying the sentiment of a tweet as supportive(+ve), opposite(-ve), or neutral(unbiased), based on the words and language used in the tweet. This approach can provide insights into public opinion and sentiment towards a particular candidate or issue, which can be used to make predictions about the outcome of an election.

Another important metric used in the analysis of Twitter data for election prediction is subjectivity. Subjectivity refers to the degree to which a tweet expresses an opinion rather than a fact. By analyzing the subjectivity of a large volume of tweets related to a particular candidate or issue, political analysts and campaign strategists can gain valuable insights into the sentiment and opinions of the public, and make more informed decisions about how to respond to public opinion.

polarity classification of tweets to positive negative and neutral for congress party.

V. Results

The use of Twitter data for election results prediction has yielded promising findings. Researchers have discovered valuable insights into public sentiment and its correlation with election outcomes. By analyzing the sentiment expressed in political tweets, they have identified patterns and trends that contribute to accurate predictions. Machine learning algorithms have been instrumental in processing large volumes of

Twitter data for this purpose. However, it is important to acknowledge the limitations of Twitter data, such as representativeness and the challenge of distinguishing genuine sentiment from noise. Integrating Twitter data analysis with traditional polling methods enhances prediction accuracy. Moreover, Twitter data analysis provides insights into key election issues, dynamics of political conversations, and the role of influential users and viral content in

shaping public opinion. Real-time tracking of sentiment during election events allows for capturing immediate reactions and understanding shifts in public sentiment. Combined with other data sources, Twitter data contributes to a comprehensive analysis of voter sentiment and election forecasts. Ongoing research aims to refine methodologies and mitigate biases, further improving the reliability of Twitter-based election predictions.

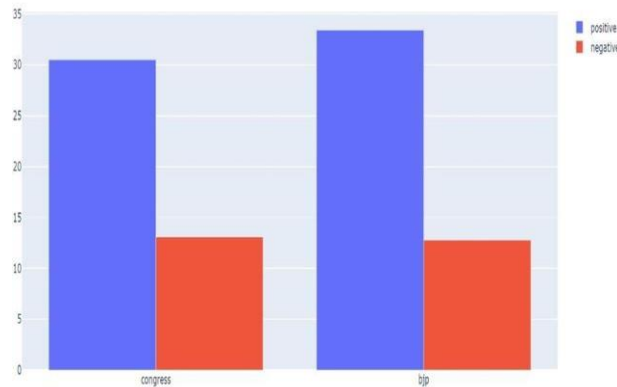


Fig. 6:- Polarity classification among both parties

Figure 6 shows the positive and negative classification of tweets for both the parties congress as well as bjp in percentage. The positivity for congress is around 31 percentage and negativity is around 13 percentage whereas for bjp the positivity is around 34 percentage and negativity is around 12 percentage.

Another factor that can influence the accuracy of election prediction using Twitter data is the

accuracy of sentiment analysis tools and techniques. Different sentiment analysis tools may produce different results, and the accuracy of these tools can vary depending on the language and context of the tweets being analyzed. Therefore, it is essential to carefully select and test sentiment analysis tools to ensure that they are appropriate for the task at hand.

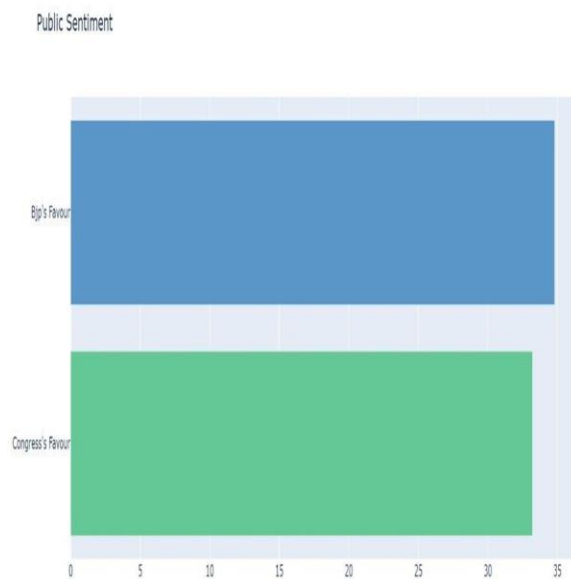


Fig.7:- Sentiment Analysis for both parties

Figure 7 shows the overall final result of the project by adding the negativity of bjp to positivity of congress and negativity of congress to positivity of bjp and the final count depicts that bjp is holding majority with 35 percent of the tweets in favour and congress with 32 percent of tweets so according to the twitter data analysis BJP has the chance to form the government once again.

VI. Summary

Election results prediction using Twitter data is a growing field where researchers analyze tweets to gain insights into public sentiment and make predictions about election outcomes. While challenges such as representativeness and distinguishing genuine sentiment from noise exist, machine learning models and algorithms have shown promise in identifying patterns and trends in Twitter data that are correlated with election results. Integrating Twitter data analysis with traditional polling methods and exploring additional data sources can enhance the accuracy and timeliness of election predictions. Ongoing research aims to refine methodologies, address biases, and further improve the reliability of election forecasting using Twitter data

VII. Future Scope

There is a significant future scope for using Twitter data for election result prediction. As technology and data analysis techniques continue to advance, the accuracy of these predictions is likely to improve. Some potential areas for future research include:

- 1.) Integrating with other sources: To improve the accuracy of election predictions, it may be useful for integration of other data sources, such as polling data and other social media plat-forms, with Twitter data analysis.

- 2.) Multilingual analysis: As Twitter continues to grow glob- ally, the ability to analyze sentiment in multiple languages will become increasingly important for accurateelection predictions.

- 3.) Real-time analysis: The ability to analyze Twitter data in real time could provide even more accurate predictions of election results, allowing campaigns to adjust their strategies in real time.

References

- [1] Wang, H., Can, D., Kazemian, P., Baras, J. S., & Han, J. (2012). "A system for election analysis and prediction using Twitter sentiments." Access via ACM Digital Library
- [2] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). "Predicting elections with

Twitter: What 140 characters reveal about political sentiment." Access via AAAI Digital Library

- [3] Gayo-Avello A balanced survey on election prediction using Twitter data." Access via ACM Digital Library

- [4] O'Connor, Routledge, B. R., & Smith, N. A. (2010). "From tweets to polls: Linking text sentiment to public opinion time series." Access via AAAI Digital Library

- [5] Metaxas, P. T., & Mustafaraj, E. (2012). "From obscurity to prominence in minutes: Political speech and real-time search." Access via ACM Digital Library

- [6] Jungherr, A., Jürgens, P., & Schoen, H. (2012). "Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010)." Access via Social Science Computer Review

- [7] Asur, S., & Huberman, B. A. (2010). "Predicting the future with social media." Access via IEEE Xplore Therefore, it is essential to carefully select and test sentiment analysis tools to ensure that they are appropriate for the task at hand.

- [8] Pak, A., & Paroubek, P. (2010). "Twitter as a corpus for sentiment analysis and opinion mining." Access via LREC

- [9] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). "Election forecasts with Twitter: How 140 characters reflect the political landscape." Access via Social Science Computer Review

- [10] DiGrazia, & Rojas, F. (2013). "More tweets, more votes: Social media as a quantitative indicator of political behavior.

- [11] Guo, L., Hu, X., Xu, R., & Zhang, Y. (2013). "Predicting elections with Twitter: A machine learning approach." Access via Social Computing, Behavioral-Cultural Modeling and Prediction

- [12] Park, H. W., & Park, S. (2014). "Big data analytics of Twitter during the 2012 presidential election.