International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Natural Language Processing for Prediction of Election Results on Twitter Engagement and Polls

Tilottama Goswami [1], Mukesh Kumar Tripathi*[2], Erappa G[3], Shivakumar Swamy N[4], Manohar koli[5], Niranjan R Chougala[6]
[1]Department of Information Technology,
Vasavi College of Engineering, Hyderabad, India
Email: tgoswami@staff.vce.ac.in
[2]Department of Computer Science & Engineering,
Vardhaman College of Engineering, Hyderabad, India
Email: mukeshtripathi016@gmail.com
[3]Department of Information Science and Engineering,
R.R Institute of Technology, Bengaluru, India
Email: erappa.uppar@gmail.com
[4]Department of Computer Science and Engineering,
R.R Institute of Technology, Bengaluru, India
Email: nsshivakumar519@gmail.com
[5]Department of Computer Science,
Karnataka University Dharwad, India
Email: koli.manohar@gmail.com
[6]Department of Computer Science and Engineering,
R.R Institute of Technology, Bengaluru, India
Email: drnrchougala@gmail.com

Abstract—With the ability to predict political outcomes and provide insights into public opinion, using 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. Researchers use sentiment analysis tools to categorize tweets as good, harmful, or neutral and follow sentiment patterns over time. 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. The paper shows the results of election prediction for Indian political parties based on Twitter data

Keywords- Twitter, election result prediction, Recursive Neural Tensor Network, Natural Language Processing

I. INTRODUCTION

Election forecasting is challenging, and reliable predictions are essential for political campaigns and election analysts. Massive volumes of user-generated data about political candidates and movements have been produced due to the rise of social media platforms like Twitter. As such, social media data provides an unprecedented opportunity to extract insights and accurately predict 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 forecasts.

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 forecasts. The rise of social media sites like Twitter has produced a wealth of user-generated data that offers insightful information on various 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 election outcomes accurately [1-3]. 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 make accurate election predictions [4].

The use of Twitter data for predicting political outcomes and understanding public opinion has become a notable trend in recent years. The massive user base and real-time nature of Twitter make it a valuable source for researchers and analysts aiming to gain insights into political dynamics. Here are some key points related to the use of Twitter data for predicting election results:

- ***Sentiment Analysis:***
  - Researchers employ sentiment analysis tools to categorize tweets based on their sentiment, such as positive, negative, or neutral.
  - By tracking sentiment patterns over time, analysts can gauge the evolving public opinion regarding political candidates, parties, or issues.
  - By tracking sentiment patterns over time, analysts can gauge the evolving public opinion regarding political candidates, parties, or issues.
- ***Network Analysis:***
  - Network analysis helps identify influential users or key nodes in the Twitter ecosystem.
  - Understanding the dynamics of interactions and connections within the political discourse on Twitter provides valuable insights into the spread of information and the formation of opinions.
- ***Combining Traditional Polling Data with Machine Learning:***
  - Integrating traditional polling data with machine learning techniques enhances the accuracy of predictions.
  - Machine learning models can learn from historical data and Twitter patterns to make more informed predictions about election outcomes.
- ***Improving Predictions:***
  - Ongoing research focuses on refining methodologies to enhance the accuracy of predictions.
  - Researchers explore ways to mitigate issues related to representativeness, bias, and the challenge of identifying genuine sentiment in tweets.
- ***Campaign Insights and Political Strategies:***
  - The insights derived from Twitter data analysis can offer valuable information for election campaigns.
  - Political strategists can tailor their messaging and outreach efforts based on the real-time feedback and sentiment expressed by Twitter users.
- ***Challenges and Limitations:***
  - Despite its potential, using Twitter data for election prediction has its challenges. Representativeness is a significant concern, as Twitter users may not fully reflect the diversity of the broader population.
  - Distinguishing between genuine sentiment and noise, such as sarcastic or ironic expressions, poses a challenge.
- ***Advancements and Future Directions:***
  - Ongoing research aims to address these limitations and advance the reliability of election predictions using Twitter data.
  - The field continues to evolve, with researchers exploring new techniques and approaches to extract meaningful insights from the vast amount of Twitter data.

The paper is organized as follows – Section II focuses on literature survey. Section III describes the proposed methodology. Section IV discusses all the functional components and techniques mentioned in the methodology in details. Section V mentions the result and analysis done on the Indian political parties for election using Twitter data. Section VI briefly discusses the future work and the paper finally concludes with Section VII.

II. LITERATURE SURVEY

The study focuses on state-of-art research work done on tweet data mining, developing a credibility ranking algorithm for tweets during major events, specifically elections. The algorithms take into account several factors to assess the credibility of tweets and aims to enhance the accuracy of election result predictions. They consider the reputation of the user who posted a particular tweet. User reputation may be determined by factors such as the user's past behavior, engagement history, or credibility based on previous posts [5]. The number of retweets a tweet receives is often considered as a measure of its impact or popularity. In this context, it could be used as an indicator of credibility, with more retweets potentially suggesting higher reliability or importance. The sentiment analysis plays an important role to understand the emotional tone or attitude expressed in tweets. Sentiment analysis helps in gauging whether a tweet is positive, negative, or neutral, and this information could contribute to assessing the overall credibility of the content. By combining these factors, the algorithm aims to rank tweets based on their credibility, providing a tool for better distinguishing reliable information from less reliable or potentially misleading content, especially during critical events like elections.

The authors in [5] investigate the utility of Twitter data for predicting election results by employing sentiment analysis

to analyze political tweets, aiming to understand the sentiment expressed on Twitter regarding political candidates or issues. The primary goal of the research is to explore whether Twitter data, specifically the sentiments expressed in political tweets, can be used as a predictive tool for election outcomes. Sentiment analysis is employed as the analytical tool. This involves using natural language processing and machine learning techniques to determine the emotional tone or sentiment expressed in tweets. The author provides a demonstration of the relationship between Twitter sentiment and election outcomes. This likely involves analyzing the sentiment of tweets leading up to the election and comparing it to the actual election results. The study aims to provide insights into the potential of Twitter data as a predictive tool. This could include understanding whether Twitter sentiment aligns with or predicts election results, and if so, to what extent. Overall, this type of research is valuable for exploring the role of social media, particularly Twitter, as a potential source of information for predicting public opinion and election outcomes. It reflects the growing interest in leveraging online platforms for understanding and forecasting political dynamics.

The authors in [6] predict elections with Twitter sentiment analysis. This study likely involves predicting elections through the analysis of Twitter sentiment. It discusses limitations which include methodological constraints, challenges in the data used, or potential shortcomings in the application of sentiment analysis to Twitter data for predicting election outcomes. The research emphasizes the importance of considering contextual factors when using Twitter data for election predictions. This suggests the broader political and social context can significantly impact the accuracy and relevance of sentiment analysis on social media platforms. The examination likely discusses potential biases in analyzing political sentiment on social media platforms like Twitter. Biases can arise due to factors such as the demographics of Twitter users, the nature of political discourse on the platform, and the algorithmic parameters in sentiment analysis. In essence, the critical examination article appears to contribute to the scholarly discourse by evaluating and shedding light on the challenges and limitations associated with using Twitter data for predicting election results through sentiment analysis. This type of critical analysis is essential for advancing the field, ensuring a thorough understanding of the strengths and weaknesses of such predictive methods.

Another study [7] focuses on the utilization of aggregated Twitter sentiment to characterize the performance of political candidates during debates. The primary goal of the study is to assess the performance of political candidates during debates. Instead of relying solely on traditional methods, the study uses aggregated Twitter sentiment as a novel approach to understanding public perception. The research involves the analysis of tweet sentiment and topic distributions. Sentiment analysis is likely employed to determine the overall emotional tone of tweets related to political candidates, while topic distributions provide insights into the issues or subjects that are most discussed during debates. By analyzing aggregated Twitter sentiment and topic distributions, the study aims to provide insights into public perception of political candidates. This can include understanding how positive or negative sentiment correlates with candidate performance and how different topics are associated with public discussions during debates. The use of Twitter sentiment as a tool for characterizing candidate performance suggests that the study may explore whether there is a relationship between public sentiment expressed on Twitter and the perceived success or effectiveness of political candidates in debates. The mention of "potentially predicting election outcomes" suggests that the study may investigate whether the insights gained from Twitter sentiment during debates can be indicative of broader public opinion, with implications for election results. In summary, this research contributes to the exploration of social media data, specifically Twitter sentiment, as a valuable source of information for assessing political events such as debates. It highlights the potential of novel approaches in gauging public perception and understanding how online discussions may be linked to broader political dynamics, including potential implications for election outcomes.

The paper [8] puts up questions on reliability to predict election results using Twitter platform and address several limitations associated with it. The article challenges the reliability of Twitter as a predictive tool for election outcomes. These skepticisms arise from the fact that the dynamics on social media platforms may not fully represent the complexity of factors influencing election results. The study discusses various limitations associated with using Twitter data. These limitations include biases inherent in the user demographics of Twitter, the potential lack of representativeness of Twitter users compared to the broader population, and challenges in capturing a comprehensive view of public sentiment. The study cautions against over-reliance on Twitter predictions for election outcomes. This cautionary approach suggests that while Twitter data may provide valuable insights, it should not be the sole or definitive source for making predictions about electoral results. In summary, this paper contributes to the ongoing conversation about the utility of social media data, particularly Twitter, in predicting election outcomes. By highlighting the limitations and urging caution in interpretation, it provides a balanced perspective on the use of Twitter as a predictive tool.

The research work in [9] investigates political polarization on Twitter, focusing on the ideological positions of users and their interactions. The primary objective of the

research is to examine political polarization on Twitter. Political polarization refers to the extent to which individuals or groups diverge in their political beliefs and preferences. The study involves analyzing the ideological positions of Twitter users. This analysis includes determining the political leanings or affiliations of users based on their tweets, interactions, or other relevant data. The research explores how users with different ideological positions interact on Twitter. This interaction analysis could shed light on the nature of political discourse, the prevalence of echo chambers, and the potential for cross-ideological engagement. The study discusses the impact of political polarization on the reliability of using Twitter data for election predictions. Political polarization can influence the dynamics of information dissemination, sentiment analysis, and the overall landscape of political discussions on the platform. The research likely discusses the implications of political polarization for predicting election outcomes based on Twitter data. Polarization may introduce challenges or biases that need to be considered when using social media as a source for understanding public opinion.

The study in [10] explores the role of influential users in online protest diffusion, with potential relevance to election-related discussions on Twitter. The primary objective of the study is to investigate the role of influential users in the diffusion of online protests. Influential users likely refer to individuals or accounts with a high degree of reach, impact, or centrality within the online network being studied. The research examines how information about protests spreads online. This could involve understanding the patterns of information flow, the dynamics of discussions, and the factors influencing the diffusion of protest-related content. The study likely explores the network dynamics of online platforms, such as Twitter. Network dynamics could encompass the relationships and interactions between users, the formation of communities or clusters, and the overall structure of the online social network. The research specifically focuses on users with high centrality within the network. Users with high centrality often play a crucial role in information diffusion, as they have the potential to reach a large audience and influence the spread of content. In summary, this research contributes to the understanding of how influential users shape the diffusion of online protests and may have implications for the study of election-related discussions on platforms like Twitter. By examining network dynamics and the role of high centrality users, the study aims to provide insights into the mechanisms that drive information spread during significant political events.

The primary goal of the research in [11] is to examine the role of Twitter as a medium for political communication during the campaign for the German federal election. This involves understanding how Twitter is used by various stakeholders, including political actors, voters, and other participants in the political process. The study involves the analysis of the content and sentiment of tweets. Content analysis encompass examining the topics, themes, and messaging present in political tweets, while sentiment analysis aims to understand the emotional tone or attitude expressed in the tweets, such as positivity, negativity, or neutrality. The research aims to highlight the ways in which Twitter functions as a platform for political discourse. This includes exploring how political actors utilize Twitter to communicate their messages, engage with the public, and participate in broader political conversations. The study investigates how Twitter contributes to the dissemination of political information. This could involve understanding the reach of political messages, the patterns of information flow, and the role of retweets and mentions in amplifying content. By analyzing the content and sentiment of tweets, the research seeks to provide insights into the broader political dynamics surrounding the German federal election. This includes understanding public sentiment, identifying key issues, and assessing the impact of Twitter on shaping political narratives. In summary, this research contributes to our understanding of the role that Twitter plays in the political communication landscape during the German federal election campaign. By analyzing content and sentiment, the study aims to uncover patterns and dynamics that highlight the platform's significance in facilitating political discourse and information sharing.

III. PROPOSED MODEL

This paper aims to develop framework for an election results prediction algorithm using Twitter data. The proposed methodology combines elements of sentiment analysis, network analysis, and machine learning. The implementation results vary based on the specifics of the election, the available data, and the context. Additionally, ethical considerations and data privacy is taken into account throughout the process. 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 dynamic nature of social media platforms. It provides insights into the complexities of election prediction using Twitter data. Flow chart for Election Results Prediction Algorithm by Twitter Data is depicted in Figure 1.

- **Data Collection:**
  - ***Define Query Parameters:*** Use the Twitter API to gather tweets related to the election. Define search queries based on relevant hashtags, candidate names, political parties, and election-specific keywords.
  - ***Temporal Scope:*** Collect data over a defined period, including the campaign period and election day.
  - ***Geographical Scope:*** If applicable, filter tweets based on the geographical region of the election.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

➢ ***Data Preprocessing:***

- ***Text Cleaning*:** Remove noise, such as retweets, duplicates, and irrelevant content. Clean text data by handling special characters, emojis, and URLs.
- ***Sentiment Analysis:*** Apply sentiment analysis to categorize tweets as positive, negative, or neutral. This analysis helps gauge the overall sentiment of Twitter users regarding candidates and election-related topics.
- ***Tokenization and Lemmatization:*** Tokenize the text into words and apply lemmatization to reduce words to their base or root form.

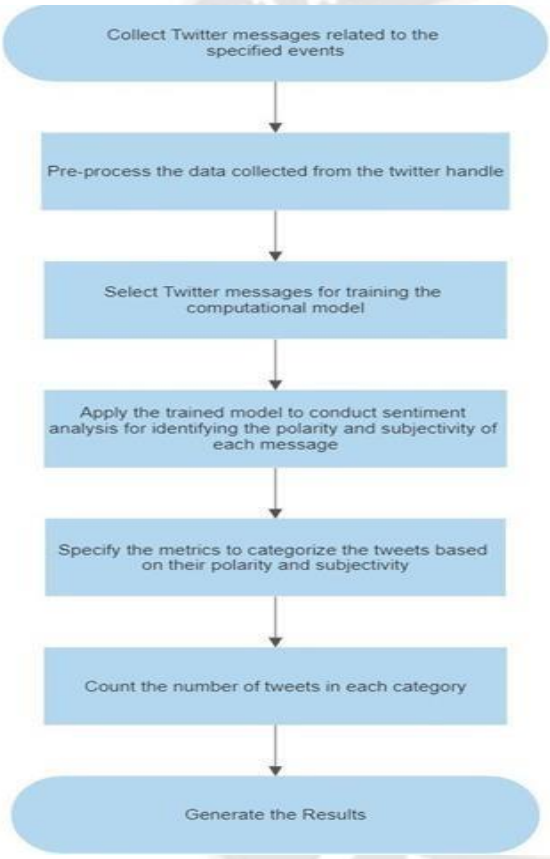

Figure 1: Flow chart of Election Results Prediction Algorithm Using Twitter Data

➢ ***Feature Engineering:***

- ***User Features:*** Extract features such as user engagement history, follower count, and account verification status. Identify influential users.
- ***Temporal Features:*** Consider temporal features like tweet frequency, hourly patterns, and sentiment changes over time.

➢ ***Network Analysis:***

- ***User Interaction Graph:*** Build a graph representing interactions between users (mentions, retweets). Identify clusters and influential users within the network.

➢ ***Machine Learning Model:***

- ***Training Data:*** Prepare labeled training data with election outcomes (win/loss) and corresponding features extracted from Twitter data.
- ***Model Selection:*** Choose a suitable machine learning model (e.g., logistic regression, random forest, or neural networks) for binary election outcome prediction.
- ***Training and Validation:*** Train the model on historical data, validate its performance, and fine-tune hyperparameters.

➢ ***Real-time Prediction:***

- ***Scraping Real-time Data:*** Continuously collect real-time tweets during the election day.
- ***Feature Extraction:*** Extract features from real-time data, including sentiment, user features, and temporal features.
- ***Prediction:*** Apply the trained machine learning model on the real time test data to predict election outcomes in real-time.

➢ ***Post-Processing and Evaluation:***

- ***Post-Processing:*** Implement any necessary post-processing steps, such as threshold tuning or ensemble methods, to enhance prediction accuracy.
- ***Evaluation Metrics:*** Assess the algorithm's performance using metrics like accuracy, precision, recall, and F1-score.

➢ ***Ethical Considerations and Validation:***

- ***Ethical Guidelines:*** Ensure compliance with ethical standards, including user privacy and consent.
- ***Validation:*** Validate the model's predictions against official election results and analyze any discrepancies.

This proposed methodology integrates data preprocessing, feature engineering, network analysis, and machine learning to develop an algorithm capable of predicting election outcomes using Twitter data. Adaptations may be necessary based on specific election characteristics and data availability. Additionally, ethical considerations and transparency in reporting are crucial throughout the process.

IV. FUNCTIONAL COMPONENTS

➢ ***Natural Language Processing:***

In the context of election results prediction using Twitter data, NLP algorithms can identify tweet patterns and sentiments related to political candidates and issues. By analyzing tweets about political events, candidates, and issues, NLP techniques

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

such as sentiment analysis and topic modeling can 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 valuable feedback on the most critical issues 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 [11-12].

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 represent the population as a whole. Therefore, it is necessary to validate the predictions made by NLP models with other data sources to ensure their accuracy. Overall, NLP is a powerful tool for analyzing and interpreting natural language data and has excellent potential in predicting election results using Twitter data. As the technology improves 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.

- ***The Recursive Neural Tensor Network***

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 analyse the sentiment expressed in tweets and understand the overall feeling towards political candidates or parties. The RNTN model comprises three main components: word embeddings, compositional functions, and a tensor layer [13-14]. Word embeddings represent words as continuous vectors that capture their semantic meaning. Compositional procedures 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 sketches 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 introduced, 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 feeling 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.

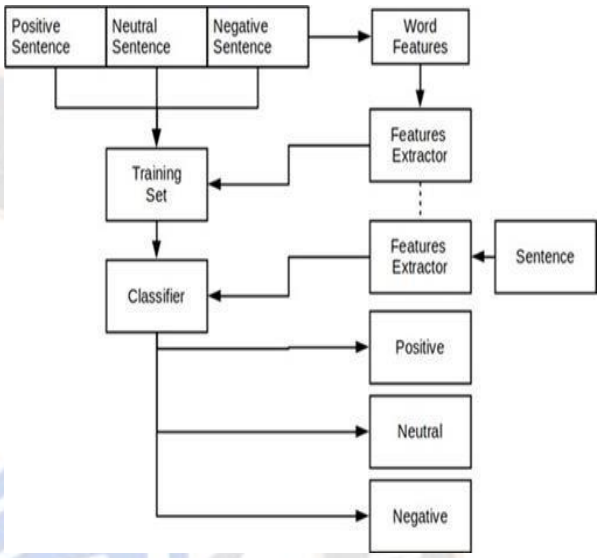

Figure 2: RNTN Architecture

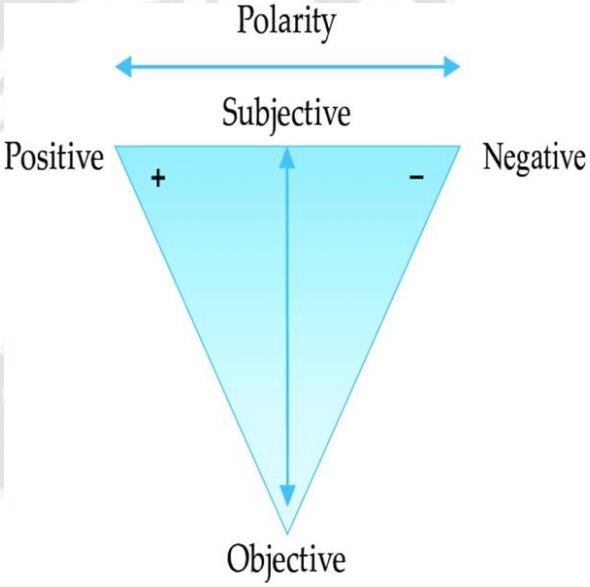

Figure 3: Polarity and subjectivity

Figure 2 shows the RNTN architecture, which 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

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

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

- ***Text Blob***

Text Blob 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 Text Blob 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 [15]. Text Blob provides convenient functions for these pre-processing tasks. Once the data is prepared, sentiment analysis is performed using Text Blob's pre-trained sentiment classifier. The sentiment polarity score ranges between [-1, +1], where values that are close to -1 represent (-ve) sentiment, values that are near to 1 indicate (+ve) sentiment, and values showing 0 signify neutral feeling. By aggregating these sentiment scores across tweets, an overall opinion towards candidates or parties can be estimated, providing insights for election results prediction. Text Blob also offers additional features like speech tagging, phrase extraction, and language translation, which can further enhance Twitter data analysis for election prediction. These features enable a more comprehensive understanding of the text and facilitate the identification of critical topics and sentiments Twitter users express. It's important to consider that Text Blob'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 carefully considering any biases or limitations associated with the sentiment classifier. Text Blob 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 analyzing sentiment, estimating overall feelings towards candidates or parties, and gaining insights into public opinion during elections.

- ***Polarity and subjectivity***

Polarity and Subjectivity are essential for sentiment analysis in predicting election results using Twitter data. Polarity refers to the sentiment expressed in a text, whether positively biased, negatively biased, or unbiased. It allows the quantification of feelings towards political candidates or parties based on tweets collected from Twitter. It is represented as a score ranging from 0 to 1, with 0 indicating objectivity and 1 representing high Subjectivity. Subjectivity analysis helps understand the level of personal judgment or bias in tweets related to elections. Using polarity as a metric in sentiment analysis can help identify patterns. As shown in Figure 3, Subjectivity is a metric used in sentiment analysis to determine the degree to which a text expresses an opinion rather than a fact. In 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 many tweets about a specific candidate or issue, political analysts and campaign strategists can gain insights into the public's opinions and attitudes toward that candidate or issue. This can help them predict an election's outcome 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 election outcome. 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.

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, as seen in Figure 4. The dataset used for the study in this paper is [16.], representing the tweets of two very big political parties, namely BJP and Congress. The cloud offers a quick and intuitive overview of the prominent themes and keywords emerging from the Twitter conversations about elections.

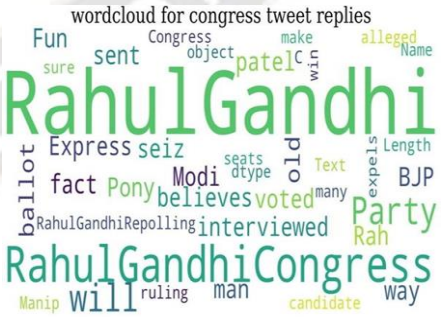

Figure 4: Word Cloud for Congress party tweets

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

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

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 election-related tweets. The word cloud can be generated for a specific period leading up to the election or during crucial events such as debates or rallies. Analysing multiple word clouds across different timeframes can reveal shifts in public discourse and sentiment, aiding in tracking the dynamics of electoral campaigns.

Word clouds can be a starting point for more in-depth analyses, such as sentiment analysis, topic modelling, 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 to make 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.

V. Results and Analysis

The prediction uses sentiment analysis tools like the RNTN Algorithm and Text Blob 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 predict an election's outcome. Another important metric used in analyzing 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 sentiments and opinions of the public and make more informed decisions about how to respond to public opinion. The implementation is done in Python using NLTK and scikit- learn libraries.

In the real time 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. Figure 5 shows the polarity classification of tweets to positive, negative, and neutral for the Congress party. 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 views 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 essential 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, the 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 view. 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.

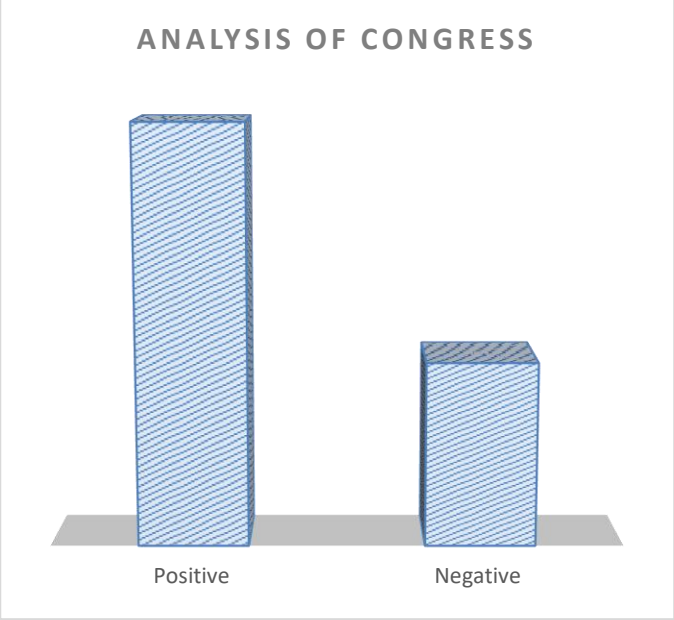

Figure 5: -Analysis for Congress

Figure 5 shows the positive and negative classification of tweets for both the parties, Congress and Bhartiya Janta Party, in percentages. The positivity for Congress is around 31 percent, and negativity is about 13 percentage, whereas for

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

the Bhartiya Janta Party, the positivity is about 34 percentage, and negativity is about 12 percentage. Another factor influencing 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 they are appropriate for the task. Figure 6 also shows the overall final result of the project by adding the negativity of the Bhartiya Janta Party to the positivity of Congress and the negativity of Congress to the positivity of the Bharthi Janta Party. The final count depicts that the Bharthi Janta Party holds the majority, with 35 percent of the tweets in favor, and Congress has 32 percent. According to the Twitter data analysis, the Bharti Janta Party can form the government again.

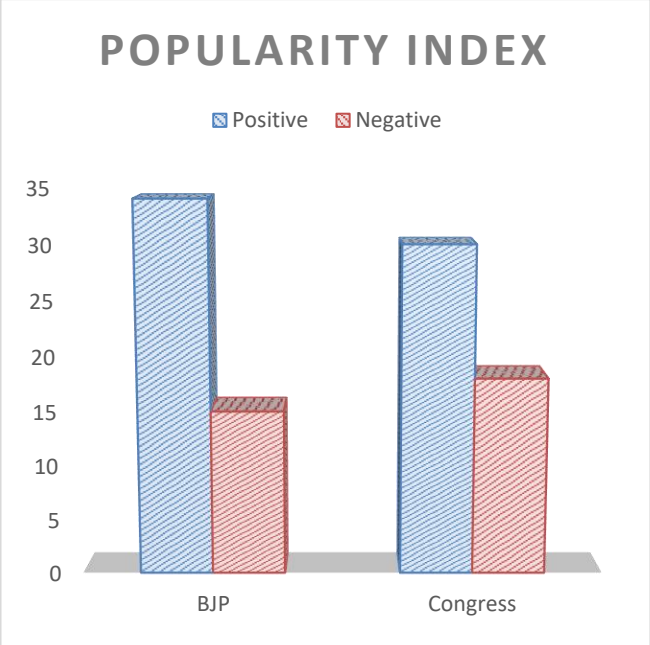

Figure 6: Polarity classification among both parties

Figure 6 shows the positive and negative classification of tweets for both the parties, Congress and Bhartiya Janta Party, in percentages. The positivity for Congress is around 31 percent, and negativity is about 13 percentage, whereas for the Bhartiya Janta Party, the positivity is about 34 percentage, and negativity is about 12 percentage. Another factor influencing 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 they are appropriate for the task. Figure 6 also shows the overall final result of the project by adding the negativity of the Bhartiya Janta Party to the positivity of Congress and the negativity of Congress to the positivity of the Bharthi Janta Party. The final count depicts that the Bharthi Janta Party holds the majority, with 35 percent of the tweets in favor, and Congress has 32 percent. According to the Twitter data analysis, the Bharti Janta Party can form the government again.

VI. 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:

- **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 platforms, with Twitter data analysis.
- **Multilingual analysis:** As Twitter continues to grow globally, the ability to analyses sentiment in multiple languages will become increasingly important for accurate election predictions.
- **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.

VII. Conclusion

The paper aimed to develop an algorithm that could effectively make the Prediction of Election Results on Twitter Engagement and Polls based on Natural language Processing techniques and RNTN model. 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 emotion 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. The paper shows the result analysis of election prediction for Indian political parties based on Twitter data. 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.

Acknowledgments:
The authors would like to acknowledge Vasavi College of Engineering, Hyderabad for supporting the work.

Reference:

[1] Olabanjo, Olusola, Ashiribo Wusu, Oseni Afisi, Mauton Asokere, Rebecca Padonu, Olufemi Olabanjo, Oluwafolake Ojo, Olusegun Folorunso, Benjamin Aribisala, and Manuel Mazzara. "From Twitter to Aso-Rock: A sentiment analysis framework for understanding Nigeria 2023 presidential election." Heliyon 9, no. 5 (2023).

[2] Singh, Prabhsimran, Yogesh K. Dwivedi, Karanjeet Singh Kahlon, Annie Pathania, and Ravinder Singh Sawhney. "Can twitter analytics predict election outcome? An insight from 2017 Punjab assembly elections." Government Information Quarterly 37, no. 2 (2020): 101444.

[3] Dhiman, Aarzoo, and Durga Toshniwal. "AI-based Twitter framework for assessing the involvement of government schemes in electoral campaigns." Expert Systems with Applications 203 (2022): 117338.

[4] Antonakaki, Despoina, Paraskevi Fragopoulou, and Sotiris Ioannidis. "A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks." Expert Systems with Applications 164 (2021): 114006.

[5] Orellana-Rodriguez, Claudia, and Mark T. Keane. "Attention to news and its dissemination on Twitter: A survey." Computer Science Review 29 (2018): 74-94.

[6] Georgiadou, Elena, Spyros Angelopoulos, and Helen Drake. "Big data analytics and international negotiations: Sentiment analysis of Brexit negotiating outcomes." International Journal of Information Management 51 (2020): 102048.

[7] Jain, Vinay Kumar, Shishir Kumar, and Steven Lawrence Fernandes. "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics." Journal of computational science 21 (2017): 316-326..

[8] Kondamudi, Medeswara Rao, Somya Ranjan Sahoo, Lokesh Chouhan, and Nandakishor Yadav. "A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches." Journal of King Saud University-Computer and Information Sciences 35, no. 6 (2023): 101571.

[9] Parackal, Mathew, Damien Mather, and David Holdsworth. "Value-based prediction of election results using natural language processing: A case of the New Zealand General Election." International Journal of Market Research 60, no. 2 (2018): 156-168.

[10] Mohd, Mudasir, Saheeba Javeed, Mohsin Altaf Wani, Hilal Ahmad Khanday, Abid Hussain Wani, Umar Bashir Mir, and Sheikh Nasrullah. "poliWeet—Election prediction tool using tweets." Software Impacts 17 (2023): 100542.

[11] Skoric, Marko M., Jing Liu, and Kokil Jaidka. "Electoral and public opinion forecasts with social media data: A meta-analysis." Information 11, no. 4 (2020): 187.

[12] Parackal, Mathew, Damien Mather, and David Holdsworth. "Value-based prediction of election results using natural language processing: A case of the New Zealand General Election." International Journal of Market Research 60, no. 2 (2018): 156-168.

[13] Tsai, Meng-Hsiu, Yingfeng Wang, Myungjae Kwak, and Neil Rigole. "A machine learning based strategy for election result prediction." In 2019 international conference on computational science and computational intelligence (CSCI), pp. 1408-1410. IEEE, 2019.

[14] Alvi, Quratulain, Syed Farooq Ali, Sheikh Bilal Ahmed, Nadeem Ahmad Khan, Mazhar Javed, and Haitham Nobanee. "On the frontiers of Twitter data and sentiment analysis in election prediction: a review." PeerJ Computer Science 9 (2023): e1517.

[15] Gayo-Avello, Daniel. "A meta-analysis of state-of-the-art electoral prediction from Twitter data." Social Science Computer Review 31, no. 6 (2013): 649-679.

[16] Khare, Arpit, Amisha Gangwar, Sudhakar Singh, and Shiv Prakash. "Sentiment Analysis and Sarcasm Detection in Indian General Election Tweets." In Research Advances in Intelligent Computing, pp. 253-268. CRC Press, 2023.