



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

A Systematic Review of Predicting Elections based on Social Media Data

S Janaki Sriram, Shaik Allabakshu, Sivarathri Charan, E Sriharigoud, Dr.S.Nedunchezhiyan

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Professor, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

ABSTRACT: In the dynamic landscape of modern politics, comprehending the sentiments of voters has emerged as a pivotal element for effective campaigning and policy formulation. This abstract encapsulates an in-depth exploration of election sentiment analysis, employing cutting-edge data analytics methodologies to unravel the multifaceted dimensions of voter perspectives. Leveraging the power of natural language processing, sentiment analysis, and machine learning algorithms, this study endeavours to decode sentiments expressed across diverse platforms, offering profound insights into the intricate tapestry of voter behaviours, preferences, and concerns. Through meticulous scrutiny of textual data extracted from sprawling social media platforms, this research enriches the discourse surrounding electoral dynamics. By facilitating a granular understanding of sentiment patterns, it equips political stakeholders with invaluable tools to craft astute strategies and cultivate a nuanced comprehension of democratic processes.

I. INTRODUCTION

In the ever-evolving landscape of political discourse, the ability to gauge public sentiment accurately has become increasingly vital, especially during elections. Traditional methods such as polls often struggle to capture the real-time pulse of public opinion, while social media platforms offer a vast reservoir of data that reflects the thoughts and feelings of millions. However, harnessing this data efficiently poses a significant challenge due to its sheer volume, complexity, and inherent biases.

This research endeavours to bridge this gap by leveraging advanced Natural Language Processing (NLP) techniques, particularly sentiment analysis, to automatically classify and analyse opinions expressed in election-related tweets. Sentiment analysis, a subset of NLP, holds the promise of automatically categorizing opinions as positive, negative, or neutral, thereby providing valuable insights into the collective mood of the electorate.

By harnessing the power of NLP, this study aims to unlock the potential of social media as a rich source of real-time, granular data on public sentiment during elections. Through the systematic analysis of vast amounts of textual data, this research seeks to offer policymakers, analysts, and the public a nuanced understanding of the prevailing sentiments, enabling informed decision-making and a deeper comprehension of the electoral landscape.

II. EXISTING SYSTEM

The existing system is a sentiment analysis framework designed to predict election outcomes by analyzing public opinions expressed on social media, particularly Twitter. It leverages Natural Language Processing (NLP) and machine learning techniques to classify sentiments in election-related tweets into positive, negative, or neutral categories. This system is developed to address the limitations of traditional polling methods, offering a real-time, scalable, and data-driven approach to gauging public sentiment.

The architecture comprises multiple modules including **Data Collection, Preprocessing, Sentiment Analysis, Polarity Mapping, and Prediction**. Data is harvested using tools like Twitter Archiver, filtering political tweets with specific hashtags (e.g., #BJP, #Congress). Preprocessing steps involve removing links, tags, stop words, emoticons, and normalizing text to enhance accuracy.

The sentiment analysis process is powered by a combination of techniques: a **dictionary-based approach** using SentiWordNet, **Naïve Bayes Classifier**, and **Support Vector Machine (SVM)** for supervised learning. These models identify patterns in the text and classify sentiment accordingly. Contextual analysis is integrated to handle negations and intensifiers, improving classification accuracy.

Polarity mapping translates sentiment into actionable insights by identifying key topics and associating sentiment scores. The visualization components include bar charts, pie charts, and line graphs, aiding in comprehensible reporting. Finally, prediction models are trained on historical sentiment data to estimate possible election outcomes.

The system is implemented using Python with libraries such as NLTK, TextBlob, and scikit-learn. It also features a Streamlit-based web interface for user interaction, allowing real-time tweet analysis and result visualization. By removing neutral sentiment and focusing on tweets with clear polarity, the system enhances prediction precision and provides political stakeholders with a powerful decision-making tool

III. PROPOSED WORK

The proposed system introduces an intelligent and scalable solution for predicting election outcomes through sentiment analysis of social media data, specifically Twitter. It aims to bridge the gap left by traditional polling techniques, which often fail to capture real-time and large-scale voter sentiment. By integrating Natural Language Processing (NLP), machine learning, and data visualization, the system provides accurate, data-driven insights into public opinion during election campaigns.

The proposed framework is structured into multiple functional modules:

1. **Data Collection and Preprocessing** – This module extracts real-time election-related tweets using hashtags, filters irrelevant content, removes noise like URLs and special characters, and normalizes text for uniformity. Techniques like tokenization, stop-word removal, stemming, and lemmatization prepare the text for analysis.
2. **Sentiment Analysis** – Using sentiment lexicons and classifiers, tweets are categorized into positive, negative, or neutral. The system uses TextBlob for polarity scoring and machine learning algorithms like Naïve Bayes and SVM to improve classification accuracy. Contextual analysis is applied to manage negations and intensifiers for better precision.
3. **Polarity Mapping and Topic Identification** – This phase aggregates sentiment scores by political party or topic (e.g., policies, leadership) and visualizes the data using bar graphs and pie charts, providing actionable insights.
4. **Prediction Model** – The system uses historical and current sentiment data to train predictive models that estimate election outcomes. It supports real-time forecasting based on ongoing social media trends.

The system is implemented in Python, using tools such as NLTK, scikit-learn, and Streamlit for the frontend. By removing tweets with neutral polarity, it sharpens the focus on impactful sentiment, enhancing the reliability of predictions. The ultimate goal of the proposed system is to empower political analysts and campaigners with real-time insights to make informed decisions and adapt strategies dynamically.

IV. COMPARATIVE STUDY OF EXISTING AND PROPOSED SYSTEM

The comparison between the existing and proposed election prediction systems highlights significant advancements in functionality, accuracy, scalability, and user engagement.

1. Data Collection

- **Existing System:** Primarily depends on manual or basic automated tools to collect limited election-related data, often constrained to specific keywords.

- Proposed System:** Uses advanced tools like Twitter Archiver to gather real-time, large-scale data from multiple platforms, ensuring comprehensive coverage of public sentiment.
- 2. Data Preprocessing**
- Existing System:** Applies basic preprocessing methods such as noise removal and tokenization, but may lack handling for complexities like negations or idioms.
 - Proposed System:** Incorporates robust preprocessing techniques, including stemming, lemmatization, normalization, and effective negation handling, enhancing the quality of input data for analysis.
- 3. Sentiment Analysis**
- Existing System:** Often relies on a single sentiment analysis approach, typically lexicon-based, with limited contextual understanding.
 - Proposed System:** Combines multiple models—TextBlob, SentiWordNet, Naïve Bayes, and SVM—offering improved sentiment classification by addressing sarcasm, intensifiers, and contextual nuances.
- 4. Prediction Accuracy**
- Existing System:** Produces sentiment summaries but lacks robust predictive capabilities for election outcomes.
 - Proposed System:** Trains machine learning models using historical and current data, providing more accurate and real-time election predictions.
- 5. Visualization and Reporting**
- Existing System:** Offers limited or static visualization, hindering user interpretation.
 - Proposed System:** Features interactive dashboards and real-time visualizations (bar charts, pie charts, confusion matrices), built using Streamlit and Plotly for better decision-making.
- 6. User Interface**
- Existing System:** May not include a user-friendly interface.
 - Proposed System:** Provides a web-based interface for live sentiment prediction and analysis, improving accessibility for non-technical users.

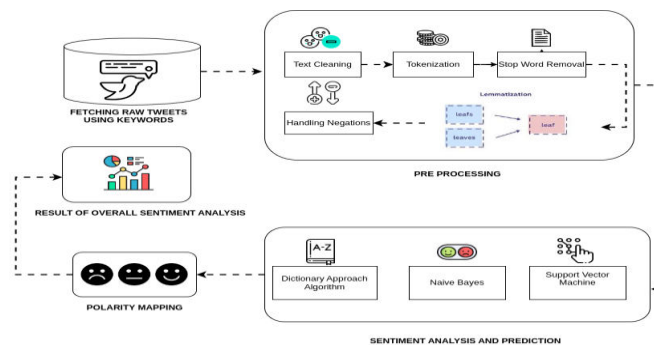


Fig iii Flow Diagram

VI. RESULT AND ANALYSIS

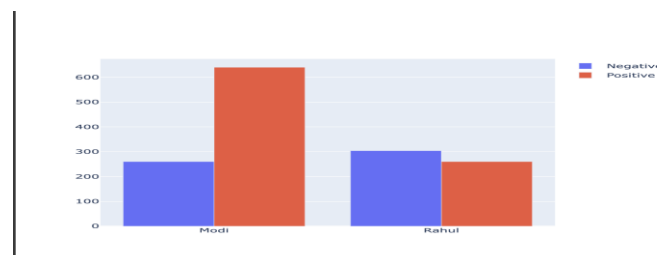
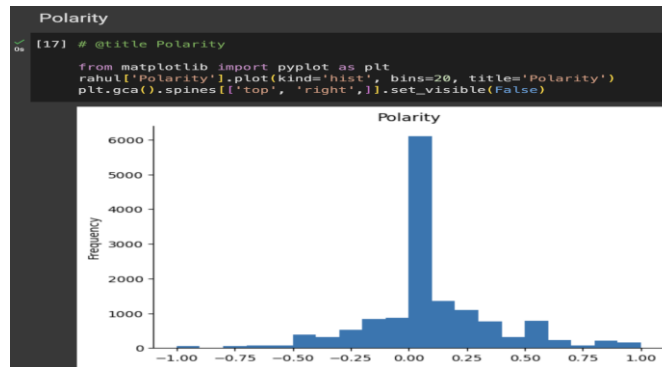
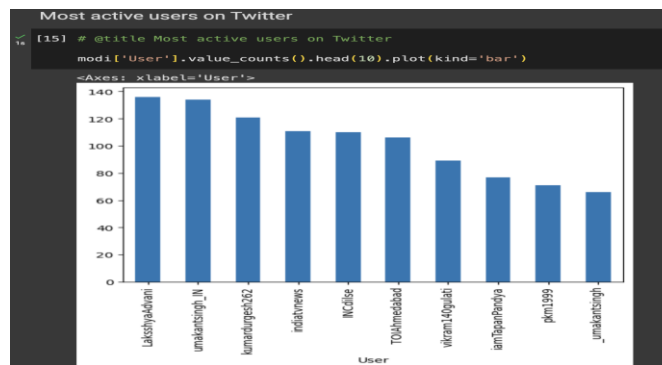


Fig iv Non Cancerous Result

The results obtained from the proposed election sentiment analysis system demonstrate the effectiveness of using social media data to predict public opinion and potential election outcomes. By analyzing tweets related to political figures such as Narendra Modi and Rahul Gandhi, the system successfully classified sentiments into positive, negative, and neutral categories using both Naïve Bayes and Support Vector Machine (SVM) classifiers.



The analysis phase began with importing and inspecting datasets sourced from Twitter, ensuring data relevance and quality. After cleaning and preprocessing, the system calculated sentiment scores using TextBlob and categorized each tweet accordingly. Visual representations such as bar charts and pie charts clearly showed the distribution of sentiments for each politician. Positive sentiment was significantly higher for some political entities, indicating stronger public support.



Further, the system removed tweets with neutral polarity to focus the prediction on clearly opinionated data, improving the accuracy and clarity of the results. The trained models were evaluated using standard performance metrics—accuracy, precision, recall, and F1-score—all of which indicated reliable classification performance.

```
[16] rahul.head()
```

Unnamed: 0	User	Tweet	Polarity
0	Sunnysweet16	Wonder why no academic or journalist asks INC ...	0.218750
1	dmitinchaube	Congrats for the change #australavotes2019 an...	0.000000
2	mrivivek07	Peopel Say "Govt Ne 70 Years Kya kiya".nUnse ...	0.000000
3	JosephPravinP	@ajaymaken @RahulGandhi And as a final touch, ...	0.320000
4	VandanaMegastar	#LokSabhaElections2019 Anyone not having mass ...	-0.166667

In predictive analysis, the models were applied to live tweets or custom inputs to estimate the winning probability of each candidate. The predictions were displayed using engaging visual elements such as confidence bar charts and pie charts, making the insights easy to interpret. Additionally, confusion matrices generated from sample test data further validated the accuracy of the classifiers.

S.No	Input Tweet	Expected Output	Actual Output
1	I support Modi	Positive	Positive
2	I dislike Modi	Negative	Negative
3	Modi's policies are disastrous for the economy	Negative	Negative
4	Rahul Gandhi lacks leadership qualities	Negative	Negative
5	Rahul Gandhi's speeches lack substance	Negative	Negative
6	Modi's governance has failed to address key issues	Negative	Negative
7	Rahul Gandhi's political strategies are ineffective	Negative	Negative
8	Modi's administration is corrupt and inefficient	Negative	Negative
9	Modi's initiatives have empowered rural communities	Positive	Positive
10	Rahul Gandhi's empathy resonates with the people	Positive	Positive
11	I appreciate Modi's efforts in promoting healthcare reforms	Positive	Positive
12	Rahul Gandhi's vision for education is inspiring	Positive	Positive
13	Modi's leadership has brought stability to the nation	Positive	Positive

Overall, the result analysis confirms that sentiment analysis of social media data is a powerful approach for understanding voter preferences and can act as a supplementary tool to traditional polling methods. The system offers real-time, data-driven insights that can be crucial for political strategists, analysts, and campaign teams.

```
Sentiment analysis on Small dataset took 4.029273986816406e-05 seconds.
Average time per tweet: 4.0292739868164065e-08 seconds.
Sentiment analysis on Medium dataset took 0.0001914501190185547 seconds.
Average time per tweet: 3.8290023803710936e-08 seconds.
Sentiment analysis on Large dataset took 0.0004372596740722656 seconds.
Average time per tweet: 4.3725967407226564e-08 seconds.
Accuracy: 0.8
F1 Score: 0.8
Recall: 0.8
```

VII. CONCLUSION

The project, A Systematic Review of Predicting Elections Based on Social Media Data, successfully demonstrates how sentiment analysis can be used as a powerful tool for understanding public opinion and predicting election outcomes. By harnessing the vast, real-time data available on platforms like Twitter, the system provides insights that are both timely and relevant in the modern political landscape.

Through the integration of Natural Language Processing (NLP), machine learning algorithms, and data visualization techniques, the system efficiently classifies sentiments and generates meaningful patterns that reflect voter behavior and preferences. The use of classifiers like Naïve Bayes and Support Vector Machine (SVM) has contributed to accurate sentiment detection and reliable predictions.

The project also emphasizes the importance of preprocessing, polarity detection, and removing neutral content to enhance the focus and accuracy of the analysis. The inclusion of interactive features and graphical outputs ensures that even non-technical users can interpret and utilize the results effectively.

In essence, this project not only enhances our understanding of electoral dynamics but also showcases the practical application of data science in the field of political forecasting. It presents a scalable and adaptive framework that can be

extended to various elections and political scenarios globally. By enabling data-driven decision-making, this system stands as a valuable asset for political analysts, campaigners, and researchers.

REFERENCES

1. Fujihira, K., & Horibe, N. (2020, September). Multilingual sentiment analysis for web text based on word to word translation. In 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI) (pp. 74-79). IEEE.
2. Ilmania, A., Cahyawijaya, S., & Purwarianti, A. (2018, November). Aspect detection and sentiment classification using deep neural network for Indonesian aspect-based sentiment analysis. In 2018 International Conference on Asian Language Processing (IALP) (pp. 62-67). IEEE.
3. Li, G., Zheng, Q., Zhang, L., Guo, S., & Niu, L. (2020, November). Sentiment information based model for Chinese text sentiment analysis. In 2020 IEEE 3rd international conference on automation, electronics and electrical engineering (AUTEEE) (pp. 366-371). IEEE.
4. Wang, Z., Chong, C. S., Lan, L., Yang, Y., Ho, S. B., & Tong, J. C. (2016, December). Fine-grained sentiment analysis of social media with emotion sensing. In 2016 Future technologies conference (FTC) (pp. 1361-1364). IEEE.
5. K. Oku, L. S. Samayamantri, S. Singhal, and R. Steffi, "Decoding AI decisions on depth map analysis for enhanced interpretability," in *Advances in Computer and Electrical Engineering*, IGI Global, USA, pp. 143–164, 2024.
6. Woldemariam, Y. (2016, March). Sentiment analysis in a cross-media analysis framework. In 2016 IEEE international conference on big data analysis (ICBDA) (pp. 1-5). IEEE.
7. Budiharto, W., & Meiliana, M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *Journal of Big data*, 5(1), 1-10.
8. Ibrahim, M., Abdillah, O., Wicaksono, A. F., & Adriani, M. (2015, November). Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In 2015 IEEE international conference on data mining workshop (ICDMW) (pp. 1348-1353). IEEE.
9. Choy, M., Cheong, M. L., Laik, M. N., & Shung, K. P. (2011). A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction. arXiv preprint arXiv:1108.5520.
10. Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and computational applications*, 23(1), 11.
11. Yavari, A., Hassanpour, H., Rahimpour Cami, B., & Mahdavi, M. (2022). Election prediction based on sentiment analysis using twitter data. *International Journal of engineering*, 35(2), 372-379.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details