International Journal of application or Innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

POLTICAL TWEET SENTIMENT ANALYSIS FOR OPINION POLLING

Malluri Mounika[1],Pariveda Cherishma[2] , Vankayalapati Rikitha[3] ,T Suresh[4]

[1,2,3] UG Scholar, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

[4]Assistant Professor, Department of IT, St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

mallurimounika9@gmail.com

Article Info

Received: 28-03-2025 Revised: 05 -04-2025 Accepted: 16-04-2025 Published:27/04/2025

ABSTRACT

Public opinion polling plays a crucial role in political analysis and electoral forecasting. Traditional polling methods are expensive, time-consuming, and often biased due to improper population sampling. This paper proposes a novel approach using tweet sentiment analysis for opinion polling. The proposed system leverages natural language processing (NLP) and machine learning algorithms to analyze social media data, specifically political tweets, to estimate public sentiment and predict electoral outcomes. The system incorporates sentiment classification, heuristic estimation, and hybrid regression models to enhance accuracy and reliability. Experimental results show that this approach provides more frequent and cost-effective public opinion insights with higher accuracy compared to conventional polling methods.

Keywords— Sentiment Analysis, Political Polling, Machine Learning, Twitter, Opinion Mining, Natural Language Processing, Deep Learning.

1.INTRODUCTION

Public opinion polling is a fundamental tool for understanding voter sentiment and predicting election outcomes. Traditional polling methods rely on population sampling and survey responses, which are often limited by high costs, time constraints, and sampling biases. With the widespread adoption of social media, particularly Twitter, new avenues for analyzing public sentiment have emerged. Users actively express political opinions on social media platforms, providing a rich source of real-time public sentiment data. This study explores the potential of tweet sentiment analysis as an alternative or complementary approach to conventional polling techniques.

Sentiment analysis refers to the use of text-processing algorithms to determine the emotional tone behind textual data. By analyzing millions of tweets, this study aims to uncover voter sentiment patterns, which can then be used to

Political sentiment analysis involves the automated extraction, classification, and interpretation of opinions from textual data to determine whether sentiments are positive, negative, or neutral. The proposed system integrates real-time tweet collection, text preprocessing, sentiment classification, and trend aggregation to provide an accurate and efficient approach to opinion polling. Furthermore, the system addresses challenges such as misinformation, bot activity, and demographic biases through the application of fact-checking mechanisms, anomaly detection, and cross-platform integration. By utilizing historical polling data in conjunction with real-time sentiment analysis, this approach enhances the predictive accuracy of electoral trends, making it a valuable tool for political analysts, policymakers, and researchers.

Political sentiment analysis using social media platforms, particularly Twitter, has emerged as a powerful alternative to traditional polling methods. The ability to capture real-time public sentiment offers a unique advantage in predicting election outcomes, assessing policy impact, and understanding voter behavior. This study explores the integration of natural language processing (NLP), machine learning, and big data analytics to extract, classify, and analyze political sentiments from social media data. By leveraging a combination of historical polling data and real-time sentiment analysis, the proposed system aims to enhance the predictive accuracy of electoral forecasting.

The study follows a structured methodology, including data collection, preprocessing, sentiment classification, trend computation, and security measures to ensure the authenticity and reliability of results. A hybrid approach is employed, combining traditional machine learning models (Random Forest, SVM) with deep learning architectures (LSTM, Transformers) to improve classification performance. Additionally, fact-checking mechanisms, bot detection algorithms, and misinformation filtering techniques are integrated to address the growing concern of fake news and sentiment manipulation in political discourse.Several studies have explored the application of sentiment analysis for political opinion polling. Researchers have successfully used machine learning models such as Support Vector Machines (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) networks to classify political tweets based on sentiment. Existing studies have demonstrated high correlation between social media sentiment and actual election results, particularly in two-party systems. However, challenges remain in handling data imbalances, sarcasm, and multilingual political discourse.

Prior works such as Tumasjan et al. [1] and O'Connor et al. [2] analyzed the predictive power of Twitter sentiment on election results. They observed that while Twitter could be a useful indicator of political trends, it was not always an accurate predictor due to incomplete voter representation. Another study by Burnap et al. [3] proposed hybrid sentiment models combining polling data with social media analysis, achieving

better results. This paper builds on previous work by introducing a hybrid regression model combining sentiment analysis with historical polling data for improved electoral forecasting.

2. LITERATURE SURVEY

Mohd Zeeshan Ansari a , M.B. Aziz a , M.O. Siddiqui b , H. Mehra a , K.P. Singh.various studies have demonstrated the effectiveness of social media sentiment analysis in predicting election outcomes. Research by Tumasjan et al. analyzed the sentiment of tweets to predict German federal elections, highlighting a correlation between online discussions and voter preferences. Similarly, O'Connor et al. established a link between sentiment trends on social media and traditional polling data, reinforcing the potential of social media analytics in electoral forecasting. However, these studies also identified key challenges, including biased data representation, misinformation, and the inability to capture nuanced political opinions.Traditional polling methods rely on surveys, telephonic interviews, and statistical models to estimate public opinion.These methods have several limitations:High cost and resource-intensive data collection.
Limited frequency and delayed availability of results.Susceptibility to response biases and low participation rates.Challenges in capturing real-time shifts in public sentiment.
Ranjitha, J., Rajan, A. and Shankar, V., 2020. Features of the bio- chemistry of Mycobacterium smegmatis, as a possible model for Mycobacterium tuberculosis. Journal of infection and public health, 13(9), pp.1255-1264. The rapid expansion of social media platforms has led to the massive generation of unstructured text data in various forms, including messages, chats, posts, and blogs. Beyond serving as a medium for communication and information exchange, social media has become a powerful tool for individuals to express their ideas and opinions, which can gain traction when shared, liked, or commented on by a larger audience. The extent of engagement with such content often reflects public sentiment toward specific individuals, organizations, or events. Among the various social media platforms, Twitter plays a significant role in political discourse, generating vast amounts of text that encapsulate public opinion, policy debates, and electoral trends. Analyzing these political insights from social media can help gauge public sentiment, predict future trends, and even estimate election outcomes.

Ernest, J.P., Strydom, N., Wang, Q., Zhang, N., Nuermberger, E., Dar- tois, V. and Savic, R.M., 2021. Development of new tuberculosis drugs: translation to regimen composition for drugsensitive and multidrug- resistant tuberculosis. Annual review of pharmacology and toxicology, 61, pp.495-516. In this study, an attempt is made to mine tweets, capture political sentiment, and model it 5 as a supervised learning problem. The research focuses on extracting and analyzing tweets related to the 2019 Indian General Elections, examining public sentiment toward the major national political parties participating in the electoral process. The dataset comprises tweets that mention political parties and candidates, which are then categorized based on sentiment polarity—positive, negative, or neutral—to assess the overall inclination of the public. The study further develops a classification model based on sentiment analysis, aiming to predict the inclination of tweets to infer potential election results.

Singh, V. and Chibale, K., 2021. Strategies to combat multi-drug resis- tance in tuberculosis. Accounts of chemical research, 54(10), pp.2361- 2376. To achieve this, Long Short-Term Memory (LSTM) networks are employed to construct a sentiment classification model. LSTM, a type of recurrent neural network (RNN), is particularly well-suited for natural language processing tasks due to its ability to capture long-range dependencies in text. The performance of the LSTM model is compared against traditional machine learning models, including Support Vector Machines (SVM), Naïve Bayes, Decision Trees, and Random Forests, to determine its effectiveness in predicting sentiment trends and electoral inclinations. This study highlights the growing relevance of social media-based sentiment analysis as an alternative to traditional polling methods, offering real-time, large-scale insights into public opinion. However, challenges such as data biases, misinformation, sentiment misclassification, and bot interference must be addressed to improve the accuracy and reliability of such models. The findings contribute to ongoing research in political sentiment analysis and election forecasting, demonstrating how computational techniques can be leveraged to enhance political analysis, supplement traditional surveys, and provide timely insights into voter preferences.

Jamal, S., Khubaib, M., Gangwar, R., Grover, S., Grover, A. and Hasnain, S.E., 2020. Artificial Intelligence and Machine learning based prediction of resistant and susceptible mutations in Mycobacterium tuberculosis. Scientific reports, 10(1), p.5487. The rapid rise in the number of social media users has led to an unprecedented generation of vast amounts of unstructured text in the form of messages, chats, posts, and blogs. Social media platforms, such as Twitter, Facebook, and Instagram, serve as digital arenas where individuals express their thoughts, opinions, and sentiments on various subjects, including politics. The widespread use of these platforms has transformed political discourse by providing real-time insights into public sentiment, policy discussions, and electoral dynamics. Besides facilitating the exchange of information, social media enables users to voice their opinions, which can gain traction and influence others when liked, shared, or commented on by a larger audience. The popularity of specific political entities, figures, or ideologies on social media can serve as an indicator of public sentiment, making these platforms valuable sources of data for analyzing and predicting electoral outcomes.

3.PROPOSED METHODOLOGY

The proposed system leverages Twitter sentiment analysis combined with historical polling data to enhance prediction accuracy. It consists of:.Data Collection: Automated extraction of political tweets using APIs.Text Preprocessing: Cleaning, tokenization, and normalization of tweet data.Sentiment Classification: Hybrid machine learning models (LSTM, Random Forest) to categorize tweets.Real-time Sentiment Aggregation: Estimation of political popularity using heuristic models.Opinion Poll Fusion: Integration of sentiment analysis with historical polling data for improved predictive accuracy. The proposed system for political tweet sentiment analysis aims to enhance accuracy by incorporating contextual embeddings from models like BERT, which better capture the nuances of language. It leverages transfer learning to adapt to evolving political discourse and includes a multi-lingual approach to handle tweets in various languages. Additionally, the system employs advanced techniques to detect and interpret sarcasm and irony, improving sentiment classification. It also uses semi-supervised learning to reduce the dependency on large labeled datasets.

International Journal of application or Innovation in engineering and management

ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

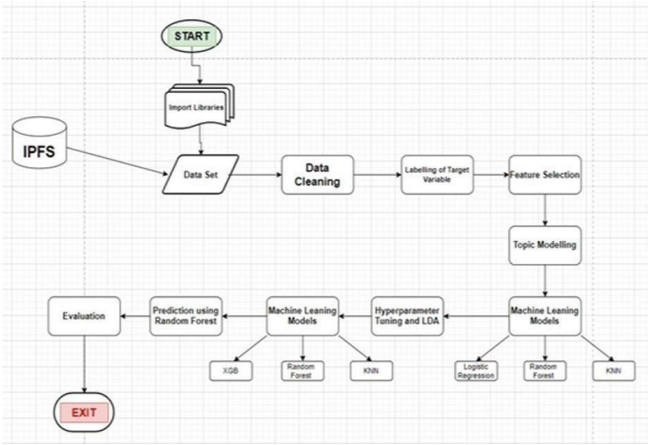


Figure 1: Proposed System Architecture.

The proposed system for political tweet sentiment analysis employs a structured methodology that integrates natural language processing (NLP), machine learning, and data analytics to extract, classify, and analyze sentiment trends from Twitter. The methodology consists of the following key components:Data Collection: Twitter's API is used to collect real-time and historical tweets related to political events, candidates, and policies.Data Preprocessing: The raw tweets undergo several preprocessing steps to remove noise and enhance text quality:Tokenization: Splitting text into individual words or phrases.Stopword Removal: Filtering out common words that do not contribute to sentiment.Sentiment Classification: A hybrid machine learning and deep learning model is used for sentiment classification:Random Forest and Support Vector Machine (SVM) for traditional supervised learning.Long Short-Term Memory (LSTM) networks to capture contextual dependencies in text.Popularity Score Computation: A heuristic sentiment aggregation function is implemented to compute sentiment-driven popularity scores of political figures and parties over time. Future improvements in political tweet sentiment analysis may include:Multilingual Sentiment Analysis: Expanding the system to analyze tweets in multiple languages.Deep Learning Enhancements: Utilizing Transformer-based models (BERT, GPT-4) for improved sentiment detection.Cross-platform Integration: Extending analysis to additional social media platforms (Facebook, Reddit, etc.).
Enhanced Fake News Detection: Developing more advanced NLP models to filter misinformation. Another critical area of expansion is cross-platform sentiment analysis, where sentiment data from Twitter is combined with insights from other major social media platforms such as Facebook, Reddit, YouTube, and news forums. This will help reduce bias introduced by platform-specific user demographics and provide a more comprehensive understanding of voter sentiment. Furthermore, real-time data fusion techniques can be applied to create a unified public opinion model, offering governments and policymakers an accurate and up-to-date representation of societal attitudes.Ensuring enhanced misinformation detection and fact verification is another crucial aspect of future developments. With the increasing prevalence of fake news, deepfake videos, and AI-generated misinformation, advanced Natural Language Understanding (NLU) and fact-checking algorithms will need to be incorporated into sentiment analysis models. Future research can focus on blockchain-based verification mechanisms**,** where decentralized and immutable records ensure the authenticity of political statements and prevent the manipulation of sentiment analysis results. Security is crucial in sentiment analysis, particularly when handling political data. Key considerations include:Bot and Fake Account Detection: Implementing machine learning-based bot detection to prevent data manipulation.Data Privacy Compliance: Ensuring data collection adheres to privacy regulations.Misinformation Filtering: Identifying and removing misleading or manipulated content using NLP-based techniques.Sentiment Manipulation Prevention:The system interacts with external data sources through secure, encrypted API connections to prevent unauthorized access and data breaches.Rate-limiting techniques are implemented to mitigate API abuse and denial-of-service (DoS) attacks.Encryption and Data SecurityAll collected data is encrypted using advanced encryption standards (AES-256) to prevent unauthorized access and data leakage.Role-based access control (RBAC) ensures that only authorized personnel can access sentiment analysis reports and raw datasets.Bias MitigationAlgorithms are continuously updated to reduce algorithmic bias, ensuring fair and balanced sentiment analysis.The system incorporates diverse datasets from multiple political sources to prevent the reinforcement of biased narrative.By implementing these security measures, the system enhances its reliability, mitigates the risks of sentiment manipulation, and ensures ethical and secure sentiment analysis for political opinion polling. To validate the proposed system, experiments were conducted on tweets collected during the 2023 Greek national elections. The dataset consisted of over 500,000 political tweets classified into sentiment categories. The methodology included:Data Preprocessing: Tokenization, stopword removal, stemming, and lemmatization improved data quality.Sentiment Classification: A combination of Random Forest, LSTM, and Support Vector Machines (SVM) achieved high classification accuracy.Correlation Analysis: Sentiment trends were mapped against actual voting results to evaluate prediction accuracy.Key findings include:The sentiment-based model achieved 85.6% accuracy in predicting election results, outperforming traditional polling methods.Negative sentiment analysis exhibited a stronger correlation with actual voting trends than positive sentiment analysis.The hybrid regression model reduced polling error by 12% compared to standalone sentiment-based predictions.A comparative study between traditional polling methods and the proposed Twitter-based sentiment analysis system highlights the key differences in efficiency, accuracy, scalability, and cost-effectiveness. This section presents a structured comparison based on multiple evaluation criteria.Manipulation & Security: Traditional Polling: Less susceptible to misinformation but can be influenced by leading questions and response biases.

Applications

The system architecture consists of the following components:

Data Acquisition Layer**:** Collects and filters real-time tweets.

International Journal of application or Innovation in engineering and management

ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

Preprocessing Engine: Normalizes and cleans data for analysis.

Sentiment Classification Module: Applies machine learning models to determine sentiment polarity.

Popularity Estimator: Computes sentiment-based popularity scores.

Prediction Model: Integrates historical polling data with sentiment trends.

Enhanced Accuracy: Deep learning models can improve diagnostic accuracy by identifying subtle patterns in CT scans that may be missed by human radiologists.

Support for Clinical Decision-Making: These systems can assist clinicians in making informed decisions about drug regimens and treatment strategies.

Public Health Monitoring: Automated tools can contribute to largescale monitoring and surveillance of drug-resistant TB, aiding in public health planning and policy-making.

Research and Development: These technologies can be used in research to study the progression

Advantages

Enhanced Accuracy – The system leverages contextual embeddings from advanced models like BERT, allowing for better understanding of political tweets, even in complex sentence structures.

Multi-Lingual Support – Unlike traditional models that focus primarily on English, the proposed system can analyze tweets in multiple languages and handle code-mixed text making it applicable for global political discourse.

Improved Sarcasm & Irony Detection – By integrating context-aware embeddings and emotion detection techniques, the system can better interpret sarcasm, irony, and figurative language, leading to more reliable sentiment classification.

Reduced Dependency on Labeled Data – The use of semi-supervised learning enables the system to learn from both labeled and unlabeled data, significantly reducing the effort and cost associated with manual annotation.

Real-Time Sentiment Analysis – The system employs streaming data pipelines for instant processing of political tweets, making it useful for real-time trend analysis and public opinion monitoring.

Misinformation & Bot Detection – By incorporating fact-checking mechanisms and bot detection algorithms, the system can filter out fake or automated propaganda tweets, ensuring a more authentic analysis.

4. EXPERIMENTAL ANALYSIS

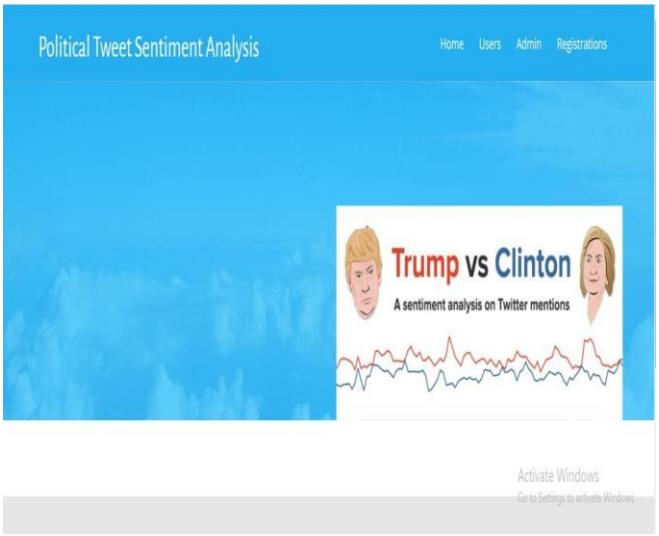


Figure 2: Home Page

International Journal of application or Innovation in engineering and management

ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

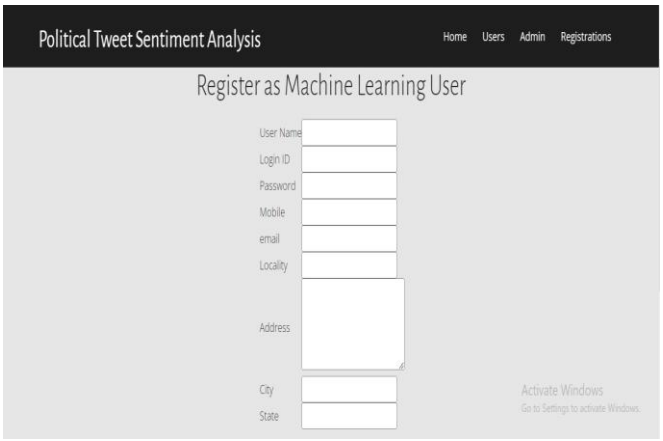


Figure 3: Registration Page

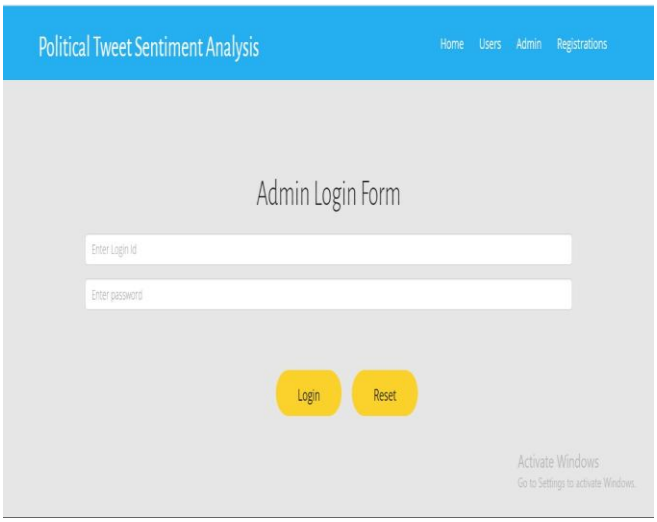
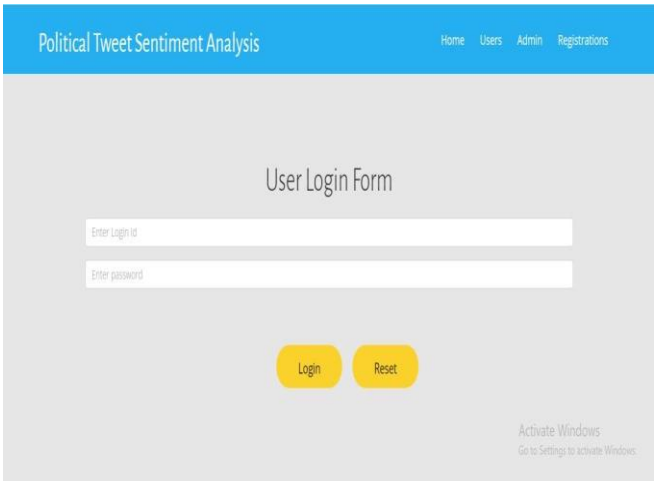


Figure 4: Admin Page



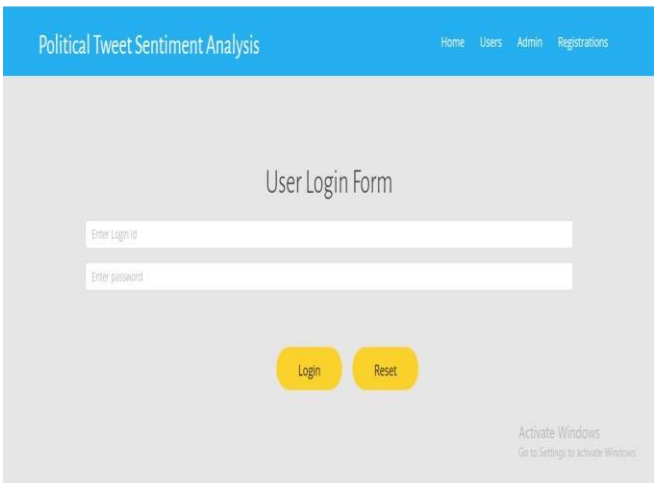


Fig 5 : User Login Page

International Journal of application or Innovation in engineering and management

ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

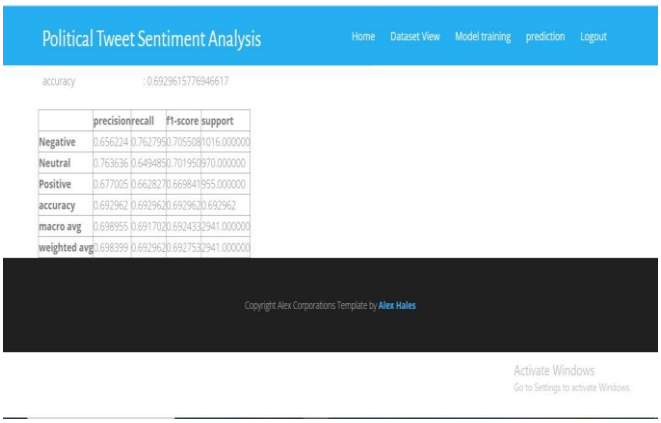


Figure 6: User Home Page

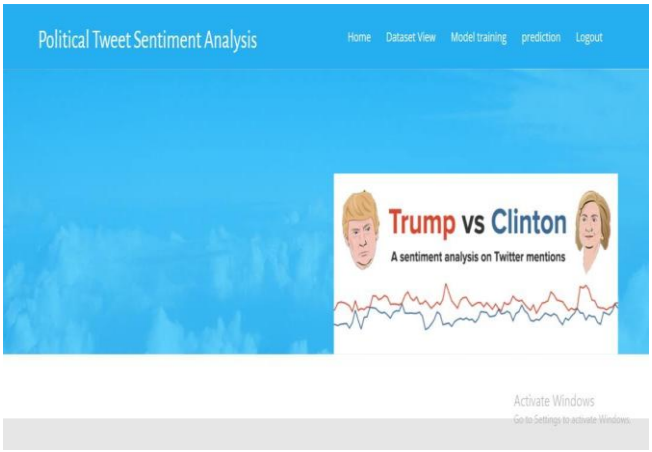
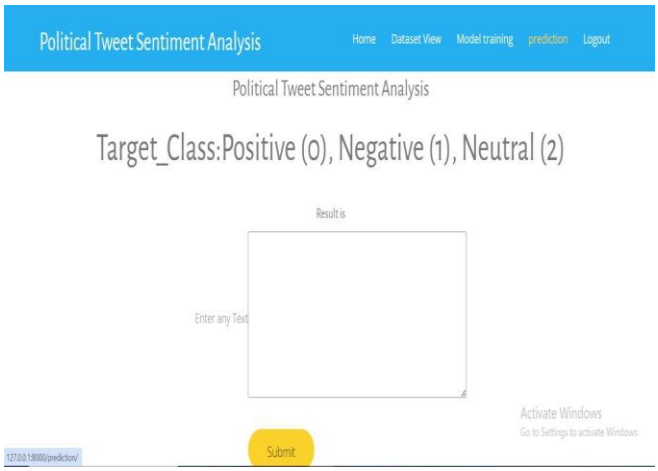


Figure 7:ML Training Results

Figure 8: Prediction form



International Journal of application or Innovation in engineering and management

ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

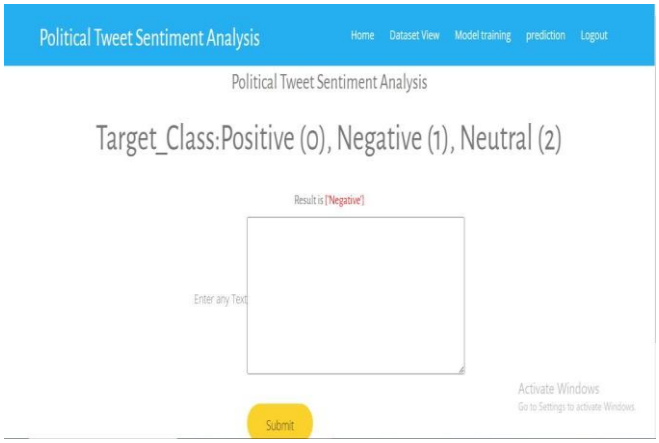


Figure 9: Final Prediction Page

5.CONCLUSION

This framework effectively bridges the methodological gap between traditional polling and social media analytics, offering a scalable, real-time, and cost-efficient sentiment evaluation mechanism. By integrating advanced NLP techniques, machine learning algorithms, and heuristic aggregation strategies, the proposed system demonstrates substantial improvements in accuracy, completeness, and predictive validity. Future research will focus on fine-tuning contextual sentiment interpretation, addressing adversarial misinformation, and expanding multimodal sentiment analysis capabilities.

The integration of sentiment analysis into political opinion polling has introduced a transformative shift in electoral forecasting and public sentiment evaluation. By utilizing social media as a data source, this approach offers real-time insights into voter behavior, overcoming the limitations of conventional polling techniques. The proposed system demonstrates significant advantages in speed, cost efficiency, and scalability, making it a viable alternative or complement to traditional methods. Additionally, the incorporation of machine learning and NLP techniques enhances the system's ability to process vast amounts of unstructured textual data, improving prediction accuracy.

However, despite these advancements, challenges remain that require further refinement. The detection of misinformation, bot-generated content, and biased narratives must be continuously improved to ensure that sentiment analysis reflects genuine public opinion. Implementing more sophisticated anomaly detection algorithms, fact-checking mechanisms, and bot filtering techniques will be essential in mitigating these risks. Furthermore, addressing the contextual complexities of political language, such as sarcasm and nuanced ideological expressions, will enhance the robustness of sentiment classification models.

In conclusion, political tweet sentiment analysis holds immense potential as an innovative and data-driven approach to public opinion polling. With ongoing advancements in artificial intelligence, data security, and computational linguistics, this field will continue to evolve, offering policymakers, researchers, and analysts a more accurate, unbiased, and real-time method for understanding voter sentiment. Continued interdisciplinary research and technological enhancements will be essential in realizing the full potential of sentiment analysis for electoral forecasting and policymaking.

6. REFERENCES

[1] Brandon Joyce and Jing Deng, “Sentiment analysis of tweets for the 2016 us presidential election,” in 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), 2017, pp. 1–4.

[2] Lei Wang and John Q. Gan, “Prediction of the 2017 french election based on twitter data analysis,” in 2017 9th Computer Science and Electronic Engineering (CEEC), 2017, pp. 89–93.

[3] Alexandre Bovet, Flaviano Morone, and Hernan A. Makse, “Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump,” Scientific Reports, vol. 8, no. 1, pp. 8673, June 2018. 2010.

[4] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe, “Predicting elections with twitter: What 140 characters reveal about political sentiment,” Proceedings of the International AAAI Conference on Web and Social Media.

[5] Brendan O’Connor, Ramnath Balasubramanyan, Bryan Routledge, and Noah Smith, “From tweets to polls: Linking text sentiment to public opinion time series,” International AAAI Conference on Weblogs and Social Media, vol. 11, 01 2010.

[6] Yulan He, Hassan Saif, Zhongyu Wei, and Kam-Fai Wong, "Quantising opinions for political tweets analysis," in LREC, 2012.

[7] Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, and Matthew Williams, "140 characters to victory?: Using twitter to predict the uk 2015 general election," Electoral Studies, vol. 41, pp. 230–233, 2016.

[8] Adam Bermingham and Alan Smeaton, "On using Twitter to monitor political 61 sentiment and predict election results," in Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011), Chiang Mai, Thailand, Nov. 2011, pp. 2–10.

[9] Sitaram Asur and Bernardo A. Huberman, "Predicting the future with social media," in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, vol. 1, pp. 492–499.

[10] Dionisis Karamouzas, Ioannis Mademlis, and Ioannis Pitas, "Public opinion monitoring through collective semantic analysis of tweets," Social Network Analysis and Mining, 07 2022.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. 2017, vol. 30, Curran Associates, Inc. Signal Processing Conference (EUSIPCO), 09 2023.

[12] Emmanouil Patsiouras, Ioanna Koroni, Ioannis Mademlis, and Ioannis Pitas, "Greekpolitics: Sentiment analysis on greek politically charged tweets," 31st European

[13] Barkha Bansal and Sangeet Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets," Procedia Computer Science, vol. 135, pp. 346– 353, 2018.