International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Mathematical Modelling and Implementation of NLP for Prediction of Election Results based on social media Twitter Engagement and Polls

Tenneti Ram prasad[1], Mukesh Kumar Tripathi[2], CH V K N S N Moorthy*[3], Vidya A. Nemade[4], Jyotsna Vilas Barpute[5], Sanjeev Kumar Angadi[6]

Submitted: 08/12/2023 **Revised:** 19/01/2024 **Accepted:** 29/01/2024

Abstract: Leveraging Twitter data for predicting election outcomes has become a notable trend in political research. Researchers use sentiment analysis tools to categorize tweets into positive, negative, or neutral sentiments. This helps gauge the public's overall mood regarding political candidates or issues. Analyzing sentiment patterns over time provides insights into the evolving dynamics of public opinion during the election cycle. Network analysis identifies influential users, key themes, and the structure of political discourse on Twitter. Integrating diverse data sources provides a more comprehensive and accurate picture of public Twitter data analysis can offer valuable insights for election campaigns, allowing them to tailor their strategies based on real-time public sentiment. Twitter users may only be partially representative of the general population. Identifying genuine sentiments from tweets can be challenging due to sarcasm, irony, and other nuances of language. Advancements in machine learning algorithms and natural language processing contribute to developing more accurate models.

Keywords: *Twitter, election result prediction, recursive neural tensor network, Natural Language Processing*

1. Introduction

Leveraging machine learning for election prediction aligns with the current data science and predictive analytics trend. It allows for developing models to learn patterns from large datasets and make informed predictions. Addressing the collection and pre-processing of a large volume of tweets is crucial. Proper data cleaning and organization are essential to ensure the accuracy and reliability of the machine learning model. Recognizing and mitigating the issue of plagiarism in Twitter data is a valuable contribution. Providing that the tweets used in the analysis are original enhances the integrity of the study and the reliability of the findings.

A key aspect is highlighting the potential of using social media data for election prediction. Demonstrating how insights extracted from Twitter can contribute to accurate forecasts provides a compelling narrative for the study's significance. Emphasizing the importance of incorporating relevant features and selecting appropriate machine learning algorithms reflects understanding the nuances involved in model development. Feature selection is crucial for identifying the most influential variables, and choosing the correct algorithm enhances prediction accuracy. The ultimate goal of election prediction studies is accurate forecasting. By focusing on feature selection and algorithm choice and demonstrating the potential of social media data, your study aligns with the broader aim of providing reliable election forecasts.

Government-sponsored schemes and campaigns play a crucial role in fostering a country's citizens' social, economic, and personal development. Government-sponsored schemes can cover many areas, such as employment generation, poverty alleviation, education, healthcare, housing, and more. The diversity of these programs reflects the government's commitment to addressing various aspects of citizens' well-being. Social media platforms are effective communication channels for government entities to disseminate information about them schemes and policies. Platforms like Twitter, Facebook, and Instagram allow real-time updates and engagement. Government entities use social media to engage directly

[1]*Department of Mathematics,Vasavi College of Engineering, Hyderabad, India,*
Email: ramprasad@staff.vce.ac.in
[2]*Department of Computer Science & Engineering, Vardhaman College of Engineering, Hyderabad, India,*
Email: mukeshtripathi016@gmail.com
orcid.org/0000-0001-5031-8947
[3]*Department of Mechanical Engineering, Vasavi College of Engineering, Hyderabad, India,*
Email: krishna.turbo@gmail.com,
orcid.org/0000-0001-7209-8017
[4]*Department of Computer Engineering, PES's Modern college of Engineering, Pune*
Email: vidya.nemade@gmail.com
[5]*Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology Pimpri Pune,*
Email: barputejyotsnav@gmail.com
orcid.org/0009-0008-5655-7242
[6]*Department of Computer Science and Engineering, Nutan College of Engineering and Research, Pune,*
Email: angadi.sanjeevkumar@gmail.com
Correspondence author: CH V K N S N Moorthy
krishna.turbo@gmail.com

with citizens. They share updates, explain the details of schemes, and provide relevant information. This engagement helps in building awareness and understanding among the public. Social media accounts owned by government entities act as information hubs, sharing details about eligibility criteria, application processes, and deadlines related to various schemes. This ensures that citizens are well-informed about the available opportunities. Social media platforms enable citizens to share their views, perceptions, feedback, and experiences regarding government schemes. This two-way communication allows for a more dynamic interaction between the government and the public. Social media provides a platform for citizens to come together as a community. They can discuss the impact of government schemes, share success stories, and support each other. This community building fosters a sense of shared purpose and collaboration. Social media allows for the rapid dissemination of information.

The introduction effectively sets the stage by emphasizing the wealth of user-generated data on social media platforms like Twitter and its potential for providing insights into politics, including election outcomes. Highlighting the unique opportunity presented by the vast amount of data generated on Twitter during an election season is crucial. This underscores the significance of social media as a rich source for understanding public sentiment and political discourse. Addressing concerns about the authenticity of Twitter data due to issues like fake news and plagiarism is a crucial aspect of your study. Acknowledging and mitigating these concerns is essential for establishing the reliability of your findings. Proper data cleaning is fundamental to the accuracy of subsequent analyses [1-3]. These techniques provide valuable insights into the Twitter data's emotional tone, influential figures, and critical themes. The explicit mention of addressing the problem of plagiarism in Twitter data adds an essential layer of credibility to your methodology. Employing measures to ensure the originality of tweets contributes to the robustness of your analysis. A key point emphasizes the study's role in demonstrating the possibilities of using social media data for election prediction. This provides a clear objective and potential contributions to political analysis. The study recognizes the importance of incorporating relevant features and selecting appropriate machine learning algorithms for accurate predictions. This understanding is crucial for any machine learning-based predictive model [4].

Indeed, using Twitter data to predict political outcomes and understand public opinion has become a notable trend in recent years. Including the statistic about 79% of India's population using social media in 2019, adds a quantitative dimension to emphasize the ubiquity of social media in daily life [5]. The mention of various forms of interaction on social networks, highlights the diverse ways people engage with each other and content on these platforms. Providing statistics about Twitter's user base, such as over 48 million monthly active users in India [6], underscores Twitter's global reach and influence as a significant social networking platform. Describing Twitter's functionalities, including the ability to share news, events, and information through brief messages, photos, videos, and links, helps set the stage for understanding how information is disseminated on the platform. Highlighting the retweeting functionality as a mechanism for exponential growth in the popularity of messages emphasizes the viral nature of content on Twitter. This is particularly relevant for political campaigns seeking widespread visibility. The platform's real-time nature and broad user base make it an assertive political communication, engagement, and mobilization tool. Since the statistics provided are from 2019, it is beneficial to provide more recent data or acknowledge any potential changes or trends in social media usage that have occurred since then.

2. Literature Survey

The study on developing a credibility ranking algorithm for tweets during significant events, particularly elections, addresses an important and timely issue. The emphasis on creating a credibility ranking algorithm aligns with the critical need for discerning trustworthy information, especially during major events like elections where misinformation can be prevalent. Considering the user's reputation as a factor in the credibility algorithm is a thoughtful approach. User behaviour, engagement history, and credibility based on previous posts can provide valuable insights into the reliability of the information shared. You are acknowledging the number of retweets as a potential indicator of credibility is reasonable. High retweet counts often suggest that a tweet has resonated with a larger audience, which can be correlated with credibility, reliability, or importance. Integrating sentiment analysis is a valuable addition. Understanding the emotional tone expressed in tweets can provide context. Positive sentiment might indicate support, while negative sentiment could suggest scepticism or criticism. It's crucial to validate and test the algorithm rigorously. Consider using historical election data or simulated scenarios to assess the accuracy and effectiveness of your credibility ranking algorithm. The sentiment analysis enhances the comprehensiveness of your credibility ranking algorithm. Ensure thorough testing, validation, and ethical considerations to contribute robust insights to the field.

The research proposal is well-structured and addresses a relevant and intriguing question about the utility of Twitter data. This clarity sets the foundation for the study[7-8]. The use of sentiment analysis, along with natural language processing and machine learning techniques, is appropriate for extracting emotional tones from tweets. Consider elaborating on the specific methods and tools you plan to use for sentiment analysis. It is crucial to mention the likely temporal aspect of the research . This temporal dimension adds depth to the study, capturing the evolving sentiment dynamics. Highlighting the comparative analysis between Twitter sentiment and election outcomes strengthens the study. This comparison will provide valuable insights into the predictive potential of Twitter data.

The study aims to investigate the alignment and predictive potential of Twitter sentiment regarding election results, and you acknowledge the critical importance of defining the extent of alignment and addressing reliability and validation aspects. Additionally, the emphasis on contextual factors, recognition of potential biases, and a critical examination of the challenges associated with using Twitter data for election predictions contribute to the scholarly discourse. Clearly define how success or alignment will be measured. Whether it's accuracy, precision, recall, or other metrics, having a well-defined set of success criteria will enhance the rigor of your analysis. Elaborate on the validation process for sentiment analysis. Describe how you plan to validate results, the requirements for accuracy, and the potential role of human annotators in comparison. This transparency will reinforce the reliability of your findings. Since the author [9] mentioned predicting elections using Twitter sentiment analysis, consider incorporating insights or conclusions from that study into your critical analysis. This comparative perspective could enrich your examination. Expand on the discussion of biases in sentiment analysis. Consider how demographic factors of Twitter users, the nature of political discourse on the platform, and challenges in sentiment analysis algorithms may introduce biases and impact the reliability of predictions. Explore the algorithmic challenges specific to sentiment analysis on social media platforms. Discuss issues like sarcasm, irony, or the brevity of tweets and how these challenges might affect sentiment classification accuracy.

The summary provides a comprehensive study overview, highlighting its primary goal, methodology, and potential implications. The description of aggregated Twitter sentiment and topic distributions as a novel approach is straightforward. Consider providing additional details on the specific methods used for sentiment analysis and how topic distributions are extracted from tweets. Elaborate on the methods employed for sentiment analysis. Emphasize the study's contribution to providing insights into public perception of political candidates. Explain how the combination of sentiment analysis and topic distributions enhances our understanding of how the public engages with and perceives candidates during debates[10-12]. The author Discusses in more detail the potential for predicting election outcomes. If the study indicates a relationship between Twitter sentiment during debates and broader public opinion, delve into the implications for election predictions. Discuss the practical implications if the study suggests a link between Twitter sentiment during debates and election outcomes. This could include considerations for political campaigns, media coverage, or public engagement strategies. The summary effectively captures the essence of the research. Expanding on certain aspects will provide a more detailed and nuanced understanding of the study's contributions and implications.

The main objectives is to provide a clear understanding of its purpose. Emphasize the importance of understanding the patterns of information flow in the context of online protests. Discuss whether the study investigates how information spreads, who the key disseminators are, and any notable characteristics of the diffusion process. Provide more insights into the factors the study explores regarding the diffusion of protest-related content. This could include user characteristics, the nature of the content, external events, or other variables that influence the spread of information. Also, Explore in more detail how the findings of the study might be relevant to election-related discussions on platforms like Twitter[13]. The study explores community formation within the online network, discusses how communities are identified, what defines them, and whether there are patterns of interaction or shared characteristics among members. The summary provides a strong foundation, and incorporating these additional points can enhance the understanding of the study's objectives and potential contributions.

This paper provides a comprehensive overview of the research focusing on the role of Twitter . Consider elaborating on how various stakeholders engage with Twitter, including political actors, voters, and other participants. Discuss these groups' specific roles and behaviours within the platform's political communication ecosystem. Provide insights into the methodological approach used for content and sentiment analysis [14-15]. This will add depth to the understanding of the research process. Emphasize the potential insights the study might provide into public sentiment. How does Twitter reflect the feelings of the electorate? Explore whether sentiment analysis can uncover trends or patterns in public attitudes toward political events, candidates, or issues. Explore the study's findings regarding how Twitter contributes to shaping political narratives. Discuss whether specific themes or topics gain prominence and how the platform influences public discourse on critical political issues. In summary, your description effectively outlines the research goals and methods. Expanding on these additional points will provide a more detailed and nuanced understanding of the study's contributions to our knowledge of political communication on Twitter during election campaigns.

Mathematical models in Natural Language Processing (NLP):

The significance of mathematical models always inevitable in Natural Language Processing (NLP). These mathematical ideas are used in numerous NLP tasks and models. These mathematical foundations are frequently used by probabilistic models such as Hidden Markov Models (HMMs), statistical language models, and machine learning techniques such as neural networks. Symbolic techniques, rule-based systems, and formal language theory also contribute to some elements of NLP, such as syntax and grammar analysis. NLP's interdisciplinary nature entails utilizing mathematical principles from computer science, linguistics, and statistics to construct successful language processing systems. The mathematics used in NLP can be broadly categorized into several areas like Probability and Statistics, Linear Algebra, Set Theory, Information Theory, Optimization, Bayesian Inferences…etc. For example, modeling language can be thought of as computing the probability of sequences of words. Consider a sequence of k words $(\mathrm{x}_1, \mathrm{x}_2, \dots \mathrm{x}_\mathrm{k})$where x is a word from finiteset of vocabulary words through Statistical Language Modeling we use conditional probability here. The probability of this random sequence of words to appear in a text is given by

$$P(X_1, X_2 \dots X_k) = P(X_1). P(X_2 \,/\, X_1). P(X_k/X_1, X_2 \dots X_{k-1})$$

The premise is that the current word in the sequence is determined by its entire history. Estimating conditional probabilities gets difficult when you have very long sequences of data, particularly from genuine languages with a huge but finite vocabulary set. To address this issue, the standard approach to modeling word sequences frequently assumes that the current word is simply dependent on the prior word. The Markov chain model is a traditional method for modelling and predicting event sequences. This fundamental model is most famous for weather prediction and biological sequence analysis [21].

Support Vector Machines (SVMs) involve a large number of mathematical concepts. The essential mathematical components of SVMs, notably in the context of NLP, are Decision Function, Hyper plane and Margins, Optimization Objective, Kernel Trick, Soft Margin SVM. A soft-margin SVM is used to solve the dual problem, particularly in the case of non-linearly separable data or to deal with outliers. The optimization objective is adjusted to add a penalty term for misclassifications is given by

$$\text{Minimize } \frac{1}{2}\|v\|^2 + S\sum_{i=1}^{m} \max\,(0{,}1 - y_i\,.\,(w.t_i + b)).$$

'S' is the regularization parameter that governs the trade-off between maximizing the margin and minimizing the classification error when given a set of input features and a weight vector 'v'. Also, another important mathematical tool in NLP is logistic regression, which is used for data processing, feature extraction, model training, model evaluation, feature importance, predictions, and hyper parameter adjustment based on the NLP task. In scenario there are few authors [22-24] who worked extensively on NLP using Mathematical models.

3. Proposed System:

Given a list of parties Pi$\in P$, P0 stands for the Bhartiya Janta Party (BJP), P1 stands for Indian National Congress (INC), and P2 stands for other parties. Also, given a user u and a tweet t, say you prefer pp if the user has at least one tweet that contains a term from the preference profile of pp. Given a set P of parties, associate with each party Pi$\in$ P, a document dp that is modelled as a bag of words. Let D = $\{$dp | Pi$\in PP\}$. Term probabilities in the corpus are calculated using tf-idf scores. The goal is to capture party-specific topics like unigrams, bigrams, or trigrams. The approach involves associating each party with a document modelled as a bag of words and calculating term probabilities in the corpus using tf-idf scores.

Data Collection:

The data collection process and the sentiment analysis focus on two main national parties (P0 and P1), with the rest of the parties grouped as Others (P2). Collected a dataset of 3896 tweets prominently significant to the two main national parties (P0 and P1) and others (P2). Identified relevant tweets through queries for party names, abbreviations, famous leaders' names, and election-related hashtags. Categorized parties into three classes: P0, P1, and Others (P2). This grouping is done collectively to facilitate sentiment analysis that considers the combined effect of the parties. Chose to measure sentiment collectively for P0 and P1, P0 and P2, P1 and P2, none, and unrelated. This approach aims to capture the inclination of tweets towards or against political parties collectively rather than in isolation.

Preprocessing:

The pre-processing steps have outlined in each step:

Filtering Non-Standard Lexical Tokens:

After tokenization, non-standard lexical tokens such as mentions, hashtags, emoticons, and unconventional punctuation are filtered out. This step helps focus on the main content of the tweets by removing elements specific to Twitter conventions.

Eliminating Duplicate Tweets and Retweets:

Duplicate tweets and retweets are eliminated to ensure the uniqueness of each tweet in the dataset. This step helps in maintaining a clean and representative set of tweets without redundancy.

Removing Standard Stop words:

Standard stop words, common wosrds that typically do not carry significant meaning, are removed from the tweets. This step helps reduce noise in the data and focuses the analysis on more informative content.

Case Folding:

Case folding is performed to convert all tokens into lowercase letters. This normalization step ensures consistency in the representation of words, as variations in letter case are ignored. This is particularly important for text analysis, as it treats "word" and "Word" as the same.

Handling Out of Vocabulary (OOV) Words:

A fair number of out-of-vocabulary words are observed in the data, and these words are kept intact. Preserving OOV words may be crucial for maintaining the richness and authenticity of the language used in the tweets. By incorporating these pre-processing steps, you aim to create a cleaned and standardized dataset that is better suited for subsequent analysis or machine learning tasks. The steps reflect an awareness of the specific characteristics of Twitter data and the need to tailor the pre-processing approach accordingly. This attention to detail in pre-processing contributes to the overall quality of the data used for analysis or modelling.

Annotation:

Annotators were instructed to identify the opinion or speculation present in each tweet rather than assigning an absolute positive or negative sentiment. This approach reflects a focus on capturing nuanced opinions rather than polarized sentiments. The annotation categories consisted of eight classes: P0, P1, P2, (P0 or P1), (P0 or P2), (P1 or P2), none, and non-relevant. These classes allow for a fine-grained tweet categorization based on content and political relevance. This decision enables the analysis of both political and non-political content within the dataset. A specific definition of sentiment is proposed, focusing on identifying opinions or speculations within the tweets rather than a binary positive or negative sentiment. This nuanced approach aligns with the complexity of political discourse.

Feature extraction:

In these four different types of feature vectors based on Term Frequency (tf) and Term Frequency-Inverse Document Frequency (tf-idf) for unigrams, bigrams, and trigrams.

Feature Vectors:

Four types of feature vectors are considered:

- Term-Frequency (tf)
- Term Frequency-Inverse Document Frequency for unigrams (tf-idf1)
- Term Frequency-Inverse Document Frequency for bigrams (tf-idf2)
- Term Frequency-Inverse Document Frequency for trigrams (tf-idf3)

Matrix Representation:

- TF and TF-IDF are used to convert a collection of raw documents into a matrix format, where rows represent documents, columns represent terms, and the matrix values represent the corresponding term frequencies or TF-IDF scores.

Weightage Based on Rarity:

- TF-IDF assigns higher weightage to words that are rare in a document but common across the entire document collection. This helps in emphasizing terms that are distinctive to a particular document.

Statistical Significance:

- TF-IDF is a significant statistical measure because it not only captures the local importance of a term in a document but also considers its global rarity across the entire corpus.

4. Functional Component:

➢ ***Natural Language Processing***

In NLP algorithms in predicting election results using Twitter data are pretty accurate. Indeed, Natural Language Processing (NLP) techniques play a crucial role in extracting valuable insights from the vast amount of textual data generated on social media platforms like Twitter during election campaigns. Analysing the sentiments in tweets related to political candidates and issues, one can gauge public opinion and identify trends in voter sentiment over time. Historical election data combined with Twitter data can be used to train machine learning models. Supervised learning algorithms can learn patterns from past election results and associated Twitter sentiments to predict future outcomes' models can quickly adapt to changes in sentiment by processing and analysing large volumes of data in real time. This adaptability is crucial during an election campaign when public sentiment can shift rapidly in response to debates, scandals, or other current events. The insights derived from NLP analysis can guide political analysts and campaign strategists in crafting messages that resonate with the concerns and priorities of the voters. NLP algorithms are well-suited for handling the massive amounts of textual data generated on social media platforms, allowing for efficient and accurate analysis. It's important to note that while NLP can provide valuable insights, predicting election outcomes is inherently complex due to the multitude of factors influencing voters. Social media sentiment may not always accurately reflect the overall sentiment of the population, and careful consideration of biases and limitations is necessary[16-17].

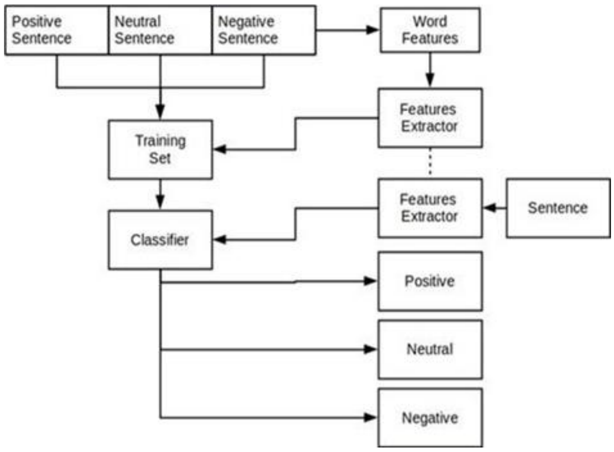

Fig 1: Flow chart for proposed model

Support Vector Machines (SVM):

Support Vector Classifier (SVC) is a discriminative classifier called Support Vector Machine (SVM). A discriminative classifier, specifically the Support Vector Classifier (SVC), is a machine learning algorithm [18-19]. Support Vector Machines, including the Support Vector Classifier, are widely used in machine learning for their effectiveness in handling various tasks. The focus on finding an optimal hyper-plane for classification is a critical feature that allows SVMs to generalize well to different data types. The optimum separation is obtained by finding the hyper-plane with the most significant distance from the nearest training data points. A discriminative classifier, specifically the Support Vector Classifier (SVC), is a machine learning algorithm.

Decision Tree:

A decision tree is a decision support tool represented as a tree-like graph. It models decisions and their consequences, considering chance event outcomes, resource costs, and utility [20]. Utilizes a tree-like structure to describe an algorithm. Performs recursive binary partitioning of the feature space. It breaks a dataset into smaller subsets while simultaneously building an incremental decision tree. Decision trees include decision nodes and leaf nodes. The decision node has two or more branches. The leaf node represents decisions or classifications. The topmost node in the tree is called the root node. The tree is incrementally developed through recursive binary partitioning. The goal is to find the most miniature tree that fits the data, often yielding the lowest cross-validation error. The decision tree classifier (DTC) was chosen for its ability to empower predictive models with high stability. Decision trees are famous for their simplicity, interpretability, and effectiveness in various applications. The emphasis on strength, interpretability, and accuracy makes DTC suitable for predictive modelling tasks.

5. Results and Discussion:

LSTM (Long Short-Term Memory) is used for evaluation. Other machine-learning algorithms include Support Vector Machines, Decision Trees, Logistic Regression, and Random Forests. Stratified 10-fold cross-validation is employed over the entire dataset. Precision is measured, with Random Forest (tf-idf-unigram) and LSTM (tf-idf-trigram) exhibiting the highest precision (0.77 and 0.76, respectively). The recall is considered, with Random Forest (tf-idf-trigram) achieving the highest recall of 0.77, comparable with its term-frequency and tf-idf-unigram models. F1-Score is calculated, and both Random Forest and LSTM show the highest F1-Score of 0.74 for tf-idf-unigram and tf-idf-trigram models. Support Vector Machine is noted to have the poorest performance among all classifiers. A comparative accuracy analysis is presented in Table 1, where LSTM with tf-idf-trigram achieves the highest accuracy. This summary provides a comprehensive understanding of the relative performance of different classifiers, highlighting the strengths and weaknesses of each algorithm in the context of the given corpus and evaluation metrics. The information is valuable for choosing an appropriate algorithm based on specific performance requirements.

Table 1: Classifier training time using tf-idf3

Classifier	Training time (s)
Support Vector Machines	4.4699
Decision Tree Classifier	0.1215
Logistic Regression	**0.0519**
Random Forest Classifier	0.8587
Long Short Term Memory	82.7816

Fig 2: Comparison of all machine Learning Techniques

6. Conclusion

The paper outlines the process and findings related to preparing an annotated corpus from Twitter for political sentiment analysis, mainly focusing on national political parties in India in the context of forthcoming general elections. The annotated corpus is prepared from Twitter data, specifically for political Sentiment keywords related to political preferences, which are considered during annotation to assign class labels. Significant features are extracted based on the term frequency-inverse document frequency (TF-IDF) from the corpus. The analysis of the acquired corpus indicates the dominance of a single political party on Twitter, constituting more than fifty percent of all extracted politically motivated tweets. Promising results are achieved with LSTM and Random Forests. To enhance the analysis and inference, it is recommended to increase the number of extracted tweets to a reasonable size. Consideration of semi-supervised techniques for corpus construction is suggested to build a robust corpus. The findings highlight the potential of LSTM and Random Forests for political sentiment analysis on Twitter. Suggestions for improvement focus on increasing the size of the dataset and employing balanced sampling techniques. This summary provides an overview of the methodology, results, and recommendations, offering insights into the process and potential enhancements for future analyses.

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] Tripathi, Mukesh Kumar, M. Neelakantapp, Anant Nagesh Kaulage, Khan Vajid Nabilal, Sahebrao N. Patil, and Kalyan Devappa Bamane. "Breast Cancer Image Analysis and Classification Framework by Applying Machine Learning Techniques." International Journal of Intelligent Systems and Applications in Engineering 11, no. 3 (2023): 930-941.

[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] Tripathi, Mukesh Kumar, and Dhananjay D. Maktedar. "Internal quality assessment of mango fruit: an automated grading system with ensemble classifier." The Imaging Science Journal 70, no. 4 (2022): 253-272.

[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] Shivendra, Kasa Chiranjeevi, and Mukesh Kumar Tripathi. "Detection of Fruits Image Applying Decision Tree Classifier Techniques." In Computational Intelligence and Data Analytics: Proceedings of ICCIDA 2022, pp. 127-139. Singapore: Springer Nature Singapore, 2022.

[7] 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.

[8] 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.

[9] Tripathi, Mukesh Kumar, Dhananjay Maktedar, D. N. Vasundhara, CH VKNSN Moorthy, and Preeti Patil. "Residual Life Assessment (RLA) Analysis of Apple Disease Based on Multimodal Deep Learning Model." International Journal of Intelligent Systems and Applications in Engineering 11, no. 3 (2023): 1042-1050.

[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] Tripathi, Mukesh Kumar. "Neutrosophic approach based intelligent system for automatic mango detection." Multimedia Tools and Applications (2023): 1-23.

[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] Tripathi, Mukesh Kumar, and Dhananjay D. Maktedar. "Optimized deep learning model for mango grading: Hybridizing lion plus firefly algorithm." IET Image Processing 15, no. 9 (2021): 1940-1956.

[14] Moorthy, CH VKNSN, Mukesh kumar Tripathi, Nilesh Pandurang Bhosle, Vishnu Annarao Suryawanshi, Priyanka Amol Kadam, and Suvarna Abhimunyu Bahir. "Investigating Advanced and Innovative Non-Destructive Techniques to Grade the Quality of Mangifera Indica." International Journal of Intelligent Systems and Applications in Engineering 12, no. 1 (2024): 299-309.

[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] 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.

[17] Tripathi, Mukesh Kumar, Praveen Kumar Reddy, and Madugundu Neelakantappa. "Identification of mango variety using near infrared spectroscopy." Indonesian Journal of Electrical Engineering and Computer Science 31, no. 3 (2023): 1776-1783.

[18] 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.

[19] 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.

[20] 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.

[21] Tripathi, Mukesh Kumar, and Dhananjay D. Maktedar. "Optimized deep learning model for mango grading: Hybridizing lion plus firefly algorithm." IET Image Processing 15, no. 9 (2021): 1940-1956.

[22] P. A. Gagniuc. Markov chains: from theory to implementation and experimentation. John Wiley & Sons, 2017.

[23] D. I. Martin and M. W. Berry. "Mathematical foundations behind latent semantic analysis". In. 2007.

[24] M. Gerlach and F. Font-Clos. "A standardized Project Gutenberg corpus for statistical analysis of natural language and quantitative linguistics". Entropy 22.1 (2020), p. 126.

[25] T. Jolliffe and J. Cadima. "Principal component analysis: a review and recentdevelopments". Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374.2065 (2016), p.20150202.