International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 2, February-2025, http://ijmec.com/, ISSN: 2456-4265

# Predicting Elections With Social Media And ML

**P.Aditya**
*Department of Electronics and Communication, Kalasalingam Academy of Research and Education, Madurai,India*
*Pamuaditya83@gamil.com*

**M.Pranav**
*Department of Electronics and Communication, (Kalasalingam Academy of Research and Education,Madurai,INDIA)*
*Pranavmavilla24@gmail.com*

**P.Vatsav Naidu**
*Department of Electronics and Communication, (Kalasalingam Academy of Research and Education,Madurai,INDIA)*
*Vatsavnaidu12@gamil.com*

**L.Sunil Gopi**
*Department of Electronics and Communication, (Kalasalingam Academy of Research and Education,Madurai,INDIA)*
*Sunilgopi773@gmail.com*

**Dr. Abha Dargar**
*Department of Electronics and Communication, (Kalasalingam Academy of Research and Education,Madurai,INDIA)*
*Abha.nyati@gmail.com*

***ABSTRACT***

***Millions of people are expected to participate in the greatest democratic exercise in human history—the general election in India. One of the most widely utilized social media sites for sharing opinions on subjects like politics, sports, and entertainment is Twitter. One of the most important channels for politicians and political parties to advertise themselves in the run-up to elections is social media. The aim of this paper is to do sentiment analysis of the Social Media network in the months leading up to the election and build a model that will be able to interpret the political subject matter. The analysis will be based on the politicians and the parties involved in the elections. This analysis can be used to acquire knowledge of the sentiments involved in the elections and predict the outcome of the upcoming elections. The models built are a Bidirectional Recurrent Neural Network model. The model is optimized using varying neural architectures and hyper parameter tuning. Kernel Regularization methods are used to address over fitting to get the best performance.***

***Keywords: Sentiment Analysis, Bidirectional Recurrent Neural Network Model, Neural Architectures, Social media (SM)***

## I. INTRODUCTION

The use of machine learning (ML) to predict election outcomes based on social media data has gained significant interest in recent years. As social media platforms like twitter, face book, and instagram have become central to public discourse, they offer valuable data on voter opinions, political sentiment, and engagement trends, making them an appealing resource for predictive analysis.

Modelling a complicated event like an election is neither straightforward nor easy but extracting public opinion from social media writing, offers a rich and demanding setting for investigating computer models of natural language processing. Politicians and political parties frequently utilize twitter, one of the most prominent social networks, to advertise themselves. On various social media platforms, including the well-known microblogging website Twitter, a large number of individuals shared their opinions about each candidate. The term “Tweet” refers to each microblog, which is limited to 140 characters. Many tweets include a label for other Twitter users in addition to a hashtag, which frequently indicates the topic of the message.

Social media has played a central role in politics and elections throughout this decade. We have entered a

new era mediated by SM in which politicians conduct permanent campaigns without geographic or time constraints, and additional information about them can be obtained not only by the press but also directly from their profiles on social networks (SNs) and through other people sharing an amplifying their voices on SM. In practice, recent examples of SM engagement and electoral success include the 2016 U.S. presidential election, when Donald Trump focused his campaign on free-media marketing, and the 2018 Brazilian presidential election, when the candidate with more SM engagement but little exposition on traditional media was elected. It is possible to measure how a politician's message is spreading over SM and try to estimate how much attention a candidate is receiving or how many people are talking about a candidate. Thus, considering a large amount of data available in real time and the low cost of their acquisition, combined with the advances of techniques for processing them, a new research subject has emerged, focusing on using the SM data to predict election outcomes.

## II. LITERATURE SURVEY

Sumukh Chourasia et.al completed the election analysis by using NLP and ML's to predict the election outcome using Twitter data [1]. NLP is used to clean the data and eliminate punctuation and the Valence-Aware Dictionary for Analysis of Sentiment Reasoning is done on the processed tweets.

Mirna Adriani et.al proposed the method for forecasting the outcome of the Indonesian presidential election using twitter and buzzer detection model sentiment analysis as our main source to predict presidential election results in a twitter nation [2]. We look into the available twitter data as a survey support tool to gauge public sentiment.

John Q Gan et.al proposed a method which is used to forecast the results of the 2017 French presidential election, the suggested method performed better than both Tumasjan's method and another method in which the authors put forth in 2017 [3].Rincy Jose et.al completed the election prediction outcome using a classifier ensemble method and improved sentiment analysis on twitter data. It leads to a classification problem where the main objective is to identify and then categorize them as a positive or negative sentiment [4].

Ravi Kumar et. al proposed the election prediction using GIS and twitter and this twitter is a well-liked social networking site for sharing thoughts on a range of subjects, such as politics, sports, and entertainment. In case of elections, social media has become a crucial tool for politicians and political parties to advertise themselves [5].

Shimei Pan et. al proposed examining and avoiding prejudice in text-based personal characteristic prediction systems because in recent years have seen a surge in interest in personality prediction based on textual data because of the possibility for extensive personalized applications like political campaigns and social media marketing [6]. Peter Meissner et. al proposed the 2014 academy awards prediction using twitter, millions of individuals regularly tweet, offering a wealth of information on people's thoughts, emotions, and actions. Elections, influenza epidemics, movie sales, and the stock market have all been predicted using data from Twitter [7].

Mucahid Kutlu et.al proposed the social media prediction of election outcomes an analysis of the 2018 Turkish Presidential Election started by examining social media account's retweets to determine their position for prediction challenges [8].

Dimitris Nikoglou et.al proposed a hybrid method to

sentiment analysis of tweets associated with elections. So, recently in this the use of social media content for political sentiment analysis has grown in popularity just like the analysis of Greek tweets regarding the most recent European elections is the main topic of this article [9].

Chin-Laung Lei et.al proposed the Boosting election prediction accuracy by crowd wisdom on social forums Researchers investigate approaches to understand how humans demonstrate their political tendencies based on the growing amount of data available through online social services [10].

Neerav Nishant et.al proposed a novel approach to election forecasting using hybrid A-BiCNN-RNN approach where the capacity of various techniques, including sentiment, volumetric, and social network analytics are used to forecast important outcomes with social media data is evaluated [11].

Adam Tsakalidis et. Al completed using polls and twitter to predict elections in several countries prior research in this area they just used data from twitter has not worked. Others have questioned the use of twitter data for this task, citing polls as their source of information [12]. Joseph D. Prusa et. al proposed the social media for polling and forecasting the results of the US election, In this the goal is to find out how well social media works as a polling and election outcome prediction tool and they compile a dataset of roughly 3 million tweets in order to achieve this [13].

Dimitrios Rousidis et. al discovered a literature review and social media prediction is a novel and effective tool that is becoming more and more well-liked by scholars and professionals. With this many benefits, it has certain disadvantages, such as bias and noise in the data, a lack of reliable predictions that are not generalizable [14].

**PROPOSED METHOD**

The process begins with the upload of a dataset, which is then preprocessed to clean and prepare the data for analysis. Subsequently, the preprocessed data is used to train various machine learning algorithms. Once the models are trained, they are employed to make predictions on new, unseen data. Finally, the generated results, which may include classifications, predictions, or insights, are produced.

The proposed system employs several machine learning models, including Decision Trees DT, Random Forests RF, Artificial Neural Networks ANN, and Multilayer Perceptron's MLP, to classify statements as positive or negative and identify instances of winning or losing. Previous research has not adequately addressed the issue of misdiagnosis, particularly when dealing with heterogeneous and large datasets. The proposed models aim to overcome this limitation by considering these factors.

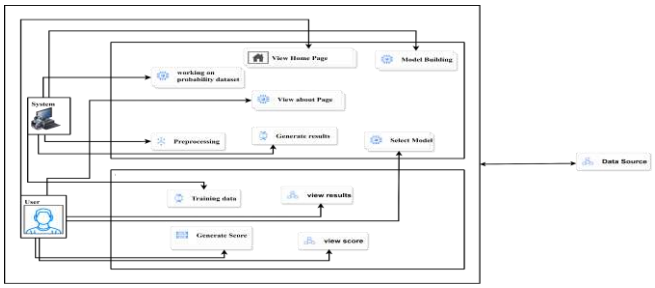



Fig. 1 Data flow diagram for election analysis using Machine Learning

## III.METHODOLOGY

A systematic review on using machine learning to predict elections through social media data should start with clear research objectives and guiding questions. These questions could address which machine learning models have been effective in election prediction, how their accuracy compares to traditional forecasting methods, and which social media platforms and data features are most commonly analyzed. The next step involves a comprehensive search of academic databases such as IEEE Xplore and Google Scholar, using keywords like “machine learning,” “election prediction,” “social media,” and “sentiment analysis.” To provide a detailed explanation, the following steps can be expanded upon:

Load data:

Loading data is the process of bringing data from a source into a system, tool, or environment so that it can be analyzed, visualized, or modeled further. Loading is usually a phase in the larger ETL (Extract, Transform, Load) process.

Input Model:

In election prediction models, the input model refers to the data and elements utilized to forecast election results. Election prediction models are statistical or machine learning models that use a number of inputs to estimate a candidate's chances of success. These inputs are carefully selected to reflect elements that have historically correlated with election results.

**View Results:**

In election prediction models, viewing outcomes entails analyzing and displaying the model's output, which is intended to communicate the expected outcome in a comprehensible and practical manner. The output of these models might range from simple percentages to elaborate representations. And it types are: categories of results, visualization of results, detailed metrics in results, sources of uncertainty, comparison to historical data, paths to victory, presentation in different contexts.

**Working on dataset:**

When working with a dataset, you must first load and examine the data, then clean it up to deal with missing values, fix data types, and get rid of duplicates. To comprehend distributions, correlations, and patterns, do exploratory data analysis. Feature engineering prepares data for analysis or modeling and generates new, relevant variables. If predictive modeling is being done, divide the dataset into training, validation, and test sets, then assess the outcomes. Lastly, record findings, analyze outcomes, and illustrate important takeaways for reporting.

**Pre-processing:**

Pre-processing is the process of cleaning and altering raw data in order to get it ready for analysis. This covers fixing data types, eliminating duplication, and dealing with missing values (such as filling or dumping). While encoding transforms categorical data into numerical form (e.g., one-hot encoding), scaling or normalizing data brings numerical features into a consistent range. To avoid skewed findings, outliers are dealt with, and skewed distributions are managed by applying transformations like log or square root. After completing these procedures, the dataset is consistent and prepared for efficient modeling or analysis.

Training the data:

Fitting a model that recognizes patterns in the data requires using a subset of the dataset, or training set. In order to reduce error, the selected model is exposed to the input features and the associated target labels while modifying internal parameters. Hyperparameters, which regulate the complexity and performance of the model, may need to be adjusted during this process. Another option is cross-validation, which divides the data several times to guarantee reliable learning and avoid overfitting. To determine whether the model can generalize to new data, it is tested on a different validation or test set once it has been trained.

Model Building:

The first step in developing a model is selecting the appropriate algorithm, such as decision trees for classification or linear regression for continuous data, depending on the task and type of data. By modifying internal parameters to reduce prediction error, the model is trained to identify patterns in the data. To increase accuracy and avoid overfitting, the model is refined through hyperparameter adjustment. Lastly, test data is used to evaluate the model to make sure it works well with unknown data and is prepared for practical application.

Generated Score:

The performance indicator used to assess a model's predictions is called a "generated score." Depending on the job (classification or regression), this score is usually calculated using metrics like accuracy, precision, recall, F1-score, or mean squared error. This score is determined by contrasting the actual results in the test or validation set with the predictions made by the model. Better model performance and generalization are indicated by a higher score. It directs future developments and aids in evaluating the model's performance in practical situations.

Output and Interpretation:

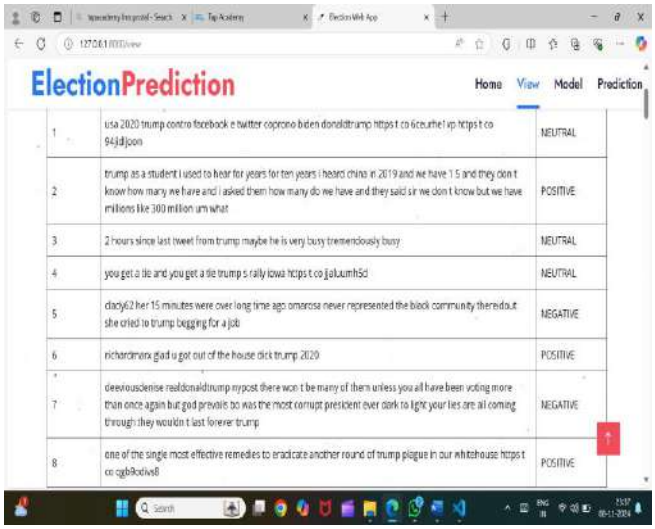
After processing the input data—which may contain values such as class labels, probabilities, or numerical forecasts—a model's output shows its predictions or outcomes. Interpretation is the process of examining these results to determine their meaning in relation to the issue, such as seeing patterns, coming to conclusions, or comprehending important influencing elements. It assists in converting unprocessed model output into useful insights or suggestions. Verifying the model's applicability and precision in addressing the intended issue may also be part of the interpretation process.

## IV. EXPERIMENTAL SETUP AND IMPLEMENTATION

The image appears to be the homepage of a website called "Election Prediction" and the title of the page is "A Systematic Review of Predicting Elections Based on Social Media Data". The image shows a silhouette of a crowd of people, suggesting the topic of the website is related to elections and predicting election outcomes based on social media data.


**Fig.5.1** Home page of developed Election Prediction website

The table analyzes election-related statements and comments on social media, categorizing them as Neutral, Positive, or Negative, highlighting Trump's interactions, China, and US population size.




| | | |
|---|---|---|
| 1 | usa 2020 trump contro facebook e twitter coprono biden donaldtrump https t co 6ceurthe l vp https t co 94jdjjoon | NEUTRAL |
| 2 | trump as a student i used to hear for years for ten years i heard china in 2019 and we have 1 5 and they don t know how many we have and i asked them how many do we have and they said sir we don t know but we have millions like 300 million um what | POSITIVE |
| 3 | 2 hours since last tweet from trump maybe he is very busy tremendously busy | NEUTRAL |
| 4 | you get a die and you get a die trump s rally iowa https t co jjaluumh5d | NEUTRAL |
| 5 | clady62 her 15 minutes were over long time ago omarosa never represented the black community thereidout she cried to trump begging for a job | NEGATIVE |
| 6 | richardmerx glad u got out of the house dick trump 2020 | POSITIVE |
| 7 | deevrousdenise realdonaldtrump nypost there won t be many of them unless you all have been voting more than once again but god prevails bo was the most corrupt president ever dark to light your lies are all coming through they wouldn t last forever trump | NEGATIVE |
| 8 | one of the single most effective remedies to eradicate another round of trump plague in our whitehouse https t co cgb9odivs8 | POSITIVE |

**Fig.5.2** Sentiment analysis of collected election data

The website uses machine learning algorithms to analyze user input, such as tweets, to predict elections based on negative sentiment.

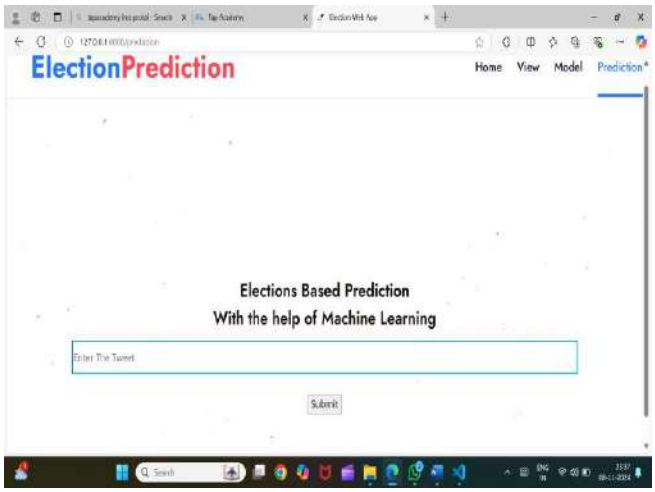



**Fig.5.3** Responses of election polling

The website uses a Random Forest Classifier machine learning model for election predictions, achieving an accuracy of 69.87%, with a dropdown menu for user selection.

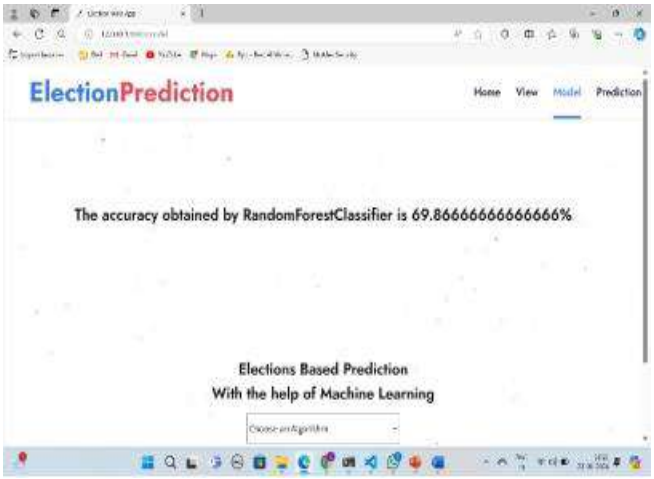



**Fig.5.4** Election Based prediction with the help of ML

The website uses machine learning algorithms to predict elections based on input data like constituency, party, and votes, utilizing various election-related data points.

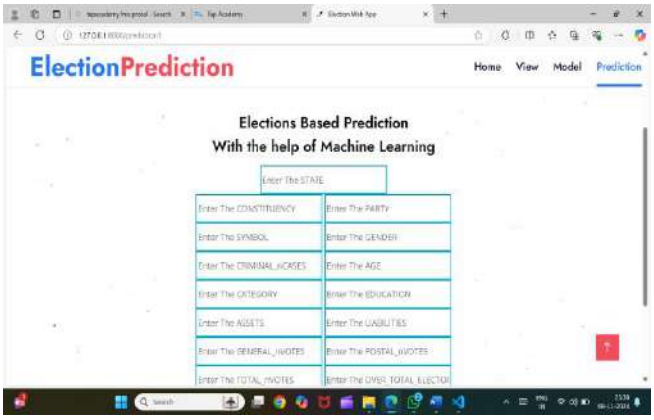



**Fig.5.5** Prediction of election result on basis collected comments

## V. CONCLUSION

The model that we selected is the Bidirectional RNN. The test data consists of two files. The first file contains tweets related to the party BJP and the second file contains tweets related to the Indian National Congress. The number of tweets in both files is the same. The tweets are preprocessed in the same way the tweets for the training data are processed. The model is used to make predictions on the tweets contained in both the files. The positive tweets for each file are calculated. The party with the greatest number of positive tweets can be considered the most likely to win the elections. According to the model prediction, the party BJP had more positive tweets. The model predicted that BJP has more positive tweets out of 2000 tweets comparison to INC has fewer positive tweets out of 2000.For the analysis, it can be inferred that the BJP would not have a clear majority like they had in the 2014 elections. However, they are still the most likely to win the 2019 elections but with a lesser margin.

## VII. REFERENCES

[1] Ramteke J, Shah S, Godhia D, Shaikh A. "*Election result prediction using Twitter sentiment analysis. In international conference on inventive computation technologies*" (Vol. 1, pp. 1-5) IEEE Conference Coimbatore, India 26 Aug 2016

[2] Wang L, Gan JQ. "*Prediction of the 2017 French election based on Twitter data analysis. In 9th Computer Science and Electronic Engineering*" (pp. 89-93) IEEE, Conference Location: Colchester, UK 27 Sep 2017

[3] Sharma P, Moh TS. "*Prediction of Indian election using sentiment analysis on Hindi Twitter. In IEEE international conference on big data"* (pp. 1966-1971). Washington, DC, USA. 5 Dec 2016

[4] Tumasjan A, Sprenger T, Sandner P, Welpe I. "*Predicting elections with twitter: What 140 characters reveal about political sentiment. In proceedings of the international AAAI conference on web and social media"* (Vol. 4, No. 1, pp. 178-185) 16 May 1010

[5] Kassraie P, Modirshanechi A, Aghajan HK. *"Election Vote Share Prediction using a Sentiment-based Fusion of Twitter Data with Google Trends and Online Polls*". (pp. 363-370), 24 Jul 2017

[6] Heredia B, Prusa JD, Khoshgoftaar TM. "*Location-based twitter sentiment analysis for predicting the US 2016 presidential election*. *In The Thirty-First International Flairs Conference".* 2018 May 10.

[7] G.P. Wani, N.V. Alone, *"Analysis of Indian election using twitter, International Journal of Computer Applications*," vol. 121, no. 22, pp. 120-127, Jan 2015.

[8] Bhola A. Twitter and Polls: *"Analyzing and estimating political orientation of Twitter users in Indian General Elections 2014"* 19 Jun 2024

[9] Kumar R, Kumar S, Soni A. *"Election prediction using twitter and GIS. In International Conference on Advance Computing and Innovative Technologies in Engineering "* (pp. 306-311) 4 Mar 2021

[10] M. Ibrahim, O. Abdillah, A. F. Wicaksono and M. Adriani, *"Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation, IEEE International Conference on Data Mining Workshop"*, Atlantic City, NJ, USA, 14 Nov, 2015.

[11] L. Wang and J. Q. Gan, *"Prediction of the 2017 French Election Based on Twitter Data Analysis Using Term Weighting, 2018 10th Computer Science and Electronic Engineering"*, Colchester, UK, 19 Sep 2018

[12] R. Jose and V. S. Chooralil, *"Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach, In International Conference on Data Mining and Advanced Computing*", Ernakulam, India. 16 Mar 2016

[13] C. Bayrak and M. Kutlu, *"Predicting Election Results Via Social Media: A Case Study for 2018 Turkish Presidential Election, in IEEE Transactions on Computational Social Systems"*, vol. 10, no. 5, pp. 2362-2373, Oct. 2023

[14] I. D. Y. Kiliç and S. Pan, *"Analyzing and Preventing Bias in Text-Based Personal Trait Prediction Algorithms, IEEE 28th International Conference on Tools with Artificial Intelligence*", San Jose, CA, USA, 6 Nov,2016.

[15] Simon Munzert; Christian Rubba; Peter Meissner; Dominic Nyhuis, *"Predicting the 2014 Academy Awards Using Twitter, in Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining"*, 18 Dec ,2014.

[16] D. Beleveslis, C. Tjortjis, D. Psaradelis and D. Nikoglou, *"A Hybrid Method for Sentiment Analysis of Election Related Tweets,4th South-East Europe Design Automation, Computer*

*Engineering, Computer Networks and Social Media Conference*, Piraeus, Greece, 21 Nov, 2019.

[17] *M. -H. Wang and C. -L. Lei, "Boosting election prediction accuracy by crowd wisdom on social forums, 13th IEEE Annual Consumer Communications & Networking Conference"*, Las Vegas, NV, USA, 9 Jan, 2016.

[18] P. B. Achar Jee, A. A. Megadump, M. Thejovathi, R. Jain, K. Umarani and N. Nishant, *"An Innovative Method for Election Prediction using Hybrid A-BiCNN-RNN Approach, 2nd International Conference on Automation, Computing and Renewable Systems"*, Pudukkottai, India,11 Dec,2023.

[19] A. Tsakalidis, S. Papadopoulos, A. I. Cristea and Y. Kompatsiaris, *"Predicting Elections for Multiple Countries Using Twitter and Polls in IEEE Intelligent Systems"*, 26 Jan, 2015.

[20] R. Vijayan and G. Mohler, *"Forecasting Retweet Count during Elections Using Graph Convolution Neural Networks, IEEE 5th International Conference on Data Science and Advanced Analytics"*, Turin, Italy, 1 Oct, 2018.

[21] Y. Gupta and P. Kumar, "*Real-Time Sentiment Analysis of Tweets: A Case Study of Punjab Elections, IEEE International Conference on Electrical, Computer and Communication Technologies"*, Coimbatore, India, 20 Feb, 2019.

[22] A. S. Jaiswal, D. V. Bhavsagar, K, S. Chaurasia, M. Daph and S. Chourasia, *"Sentiment Analysis of Election Result Prediction using Twitter Data by NLP and ML, International Conference on Innovations and Challenges in Emerging Technologies"*, Nagpur, India, 7 Jun, 2024.

[23] A. Surolia and S. Mehta, *"Optimal Feature Selection for Retweet Prediction in Indian Election, 2nd International Conference on Informatics"*, Noida, India, 23 Nov, 2023.