Sentiment Analysis for Political Campaigns

Devisetti Guru Sai,
Apex Institute Of Technology,
Chandigarh University,
Punjab, India
gurusaigurusai143@gmail.com

Madishetty Soumith,
Apex Institute Of Technology,
Chandigarh University,
Punjab, India
madishettysoumith21@gmail.com

Deshaveni Akshar Rao,
Apex Institute Of Technology,
Chandigarh University,
Punjab, India
aksharraodeshaveni2004@gmail.com

Kothalanka Siva Ganesh,
Apex Institute Of Technology,
Chandigarh University,
Punjab, India
kothalankasivaganesh@gmail.com

DR. Navjeet Kaur,
Apex Institute Of Technology
Chandigarh University,
Punjab, India
navjeet.e16069@cumail.in

Abstract—The rise of digital communication technology motivates political campaigns to use social media as a channel for public engagement and emotional evaluation of voters. Sentiment analysis functions as a necessary NLP component for efficiently sorting textual content into positive-negative-neutral opinion categories through its public opinion evaluation process. This research focuses on sentiment analysis of social media tweets which pertain to Indian political parties during election campaigns. The task requires extensive data processing alongside TF-IDF and BOW feature extraction, followed by sentiment classification through Linear SVM and Logistic Regression and Decision Trees machine learning models. Research data provides essential patterns of public opinion that serve as a tool to forecast political strategy modifications and electoral results. The research results prove that sentiment analysis proves its worth in tracking public sentiment throughout political campaigns. The combination of the Decision Tree classifier with TF-IDF features produced 93.40% accuracy, which served as the most precise predictive model. The models Logistic Regression and Linear SVC obtained similar results as they returned 82.75% and 82.42% accuracy results respectively.

***Keywords**: Sentiment analysis, political campaigns, social media, Twitter, Facebook, computational political analysis.*

INTRODUCTION

In the digital age social media functions as a crucial space for political conversation which provides politicians along with political parties and voters an opportunity to distribute information and debate and make their viewpoints known. Facebook together with Instagram and Twitter play a substantial role in transforming voter conduct and shaping public perception during political election periods. Political strategists rely on social media discussion analysis to build their decisions and acquire public sentiment information.

Textual sentiment analysis for positive and negative or neutral opinions now trends as a popular analytical approach. Based on systematic analysis of vast social media data sentiment analysis generates vital knowledge about voter opinions to enhance campaign communication and direct addressing of public matters as well as election result prediction.

Research in political studies uses sentiment analysis more frequently yet implementation barriers consistently affect its implementation. The identification of emotions becomes complex with political tweets because these posts contain sarcastic language as well as abbreviations and distinctive local linguistic elements. Social media trends need immediate waveform analysis as well as flexible model structures because these formats have a dynamic structure format. This research tackles the existing problems through analysis of Indian election-related tweets guided by machine learning mathematical patterns. The research utilizes data preparation tasks that remove stop words, tokenize words and stem words together with the feature extraction techniques which combine Term Frequency-Inverse Document Frequency (TF-IDF) and BOW. Evaluation of Naive Bayes, Logistic Regression and Linear SVM classification models identifies which one shows the most potential as a sentiment prediction model. This article explains how sentiment analysis helps number political election outcomes alongside aiding election plan development and advancing democratic choices through better information.

This paper discusses how sentiment analysis enables election result predictions and defines election strategies and facilitates more educated democratic choices.

The sequence of sections in this research features these elements: Section II delivers a review of current studies which showcase the body of understanding regarding sentiment analysis for political texts. This study reveals the research methodology in Section III which details its data acquisition process and model development alongside data preprocessing techniques. Section IV displays the analytical and experimental results. The last section discusses crucial discoveries alongside research boundaries and proposed future investigation.

II. LITERATURE REVIEW

During political campaigns, sentiment analysis is becoming an essential tool for determining public opinion, especially since social media platforms have become more and more popular as venues for political discourse. The findings, methods, and gaps in the literature are examined in this review of the literature, which looks at significant studies in this area.

Key Findings

1. P. Sharma [2] analysed public opinion and forecasting for the 2019 Indian general elections and predicted election results using data from Twitter. High prediction accuracy was attained by using Vader for sentiment analysis. Showed a high degree of agreement between the sentiment expressed on Twitter and the actual election outcomes.
2. Using Twitter data for opinion mining during the Indian elections, Bharat R. Naiknaware [15] highlighted the difficulties presented by India's linguistic diversity and lexicon-based sentiment analysis. Emphasized how crucial language-specific preprocessing is to accuracy.
3. Arpit Khare and Amisha Gangwar [10] conducted an analysis of political sentiment in Indian tweets that concentrated on sentiment polarity in regional languages with LSTM models. Discovered that tweets in regional languages offer a more in-depth analysis of the local political environment.
4. Ankita Sharma [16], in Examining Twitter Sentiments to Forecast Politics, examined machine learning techniques including Naïve Bayes and SVM for political sentiment analysis. Decided that voter preferences could be indicated by sentiment patterns.
5. Sentiment Analysis in Many Languages for Indian Politics, conducted by Payal Khurana Batra [11], used translation tools and cross-lingual models to handle the complexity of multilingual tweets. Tokenization specific to a language increased sentiment accuracy.
6. Mohd Zeeshan Ansari [12], in Election-related Sentiment Trends in Tweets, observed changes around significant election events, like rallies and manifesto releases, using time-based sentiment analysis.
7. Twitter's Influence on Political Conversation, investigated by Jyoti Ramteke and Samarth Shah [15], explored the ways in which public narratives and political discourse are influenced by Twitter hashtags and content. It showed how Twitter can be used to set agendas in real time.
8. Applying Machine Learning to Political Twitter Sentiment, conducted by Prerna Mishra [6], used ensemble methods (SVM + Random Forest) to improve sentiment classification. She found user accounts and trending hashtags that are influencing political sentiment.
9. Sentiment Analysis of Big Data for Indian Elections, conducted by Deepak Kumar Gupta [10], launched Hadoop and Spark, two big data platforms, to handle enormous Twitter datasets. This made it possible for political sentiment analysis to be scalable and done almost instantly.
10. Classifying Election Sentiment Using Deep Learning, conducted by Ghanshyam Parmar [8], used LSTM and CNN models to extract tweet features. These models surpassed conventional ML models and brought attention to issues like unbalanced datasets

Gaps Identified

1. Absence of Datasets on Indian Political Sentiment: Indian political parties like the BJP and Congress have few resources accessible, and sentiment research concentrates on global databases. Using a labelled, curated collection of tweets regarding the Congress and BJP, India-specific political sentiment analysis is made possible.
2. Lack of Party-Wise Sentiment Comparison: Most studies examine sentiment generically without discriminating between political entities. To better comprehend popular opinion towards the BJP and Congress, a party-level sentiment comparison is presented.
3. Undiscovered usage of TF-IDF for Political Content: TF-IDF is commonly utilised in NLP tasks, but its use in categorising political tweets has gotten less attention. The approach uses TF-IDF to extract relevant traits from political tweets, demonstrating its application in this sector.
4. There is a lack of focus on text normalisation for political slang. Political tweets typically contain slang, hashtags, or sarcasm, all of which require proper normalisation. Text cleaning is done with regular expressions and tokenisation to decrease noise and increase analytical accuracy.

III. METHODOLOGY

This research adopts a strategy which converts the unsupervised nature of raw social media content into a supervised learning task for analysis. The analysis of tweets uses sentiment classifications including positive and negative and neutral categories enabling machine learning models to achieve proper training results. The transformed data enables the model to both track the political emotion spread and identify recurring sentiment forms and generate predictions about citizen response patterns.

A. Dataset Collection

All users of the Kaggle website can access the Indian Election Tweets (2019) dataset that contains a handpicked selection of tweets related to the 2019 Indian General Elections. The tweets mainly stem from Indian voters who discuss their support for the two principal national parties: the Indian National Congress (INC) and the Bhartiya Janata Party (BJP). The data collection involved relevant political hashtags and Twitter accounts #BJP, #Congress, #Modi, and #RahulGandhi until the 79,730 tweets were obtained for the dataset [1].

A separation exists between two CSV files for the collection which features separate tweets about two major Indian political parties. One file presents 49,478 tweets about the Bhartiya Janata Party (BJP) whereas the second file carries 30,253 tweets focused on the Indian National Congress (INC). The partition helps researchers carry out sentiment analysis for individual parties which supports analysis comparisons to better understand public opinion during the 2019 Indian election year. Preprocessing applications and analysis tasks become simpler because of the CSV format which allows data exchange with diverse processing systems.

The distressing sentiment of Twitter users regarding political parties BJP or Congress is noted as positive or negative through this dataset. Each tweet received its score through Vader sentiment analysis tool by counting positive and negative phrases it contained. The platform classified scores that exceeded the threshold as Positive tweets whereas those below the threshold were classified as negative.

B. Data Pre-Processing

The dataset needed complete cleaning at the same time as preprocessing for sentiment analysis purposes. The text definition together with noise removal constitutes a vital step that affectively determines model classification performance. The content of all tweets underwent conversion to lowercase before processing to prevent tokenization errors between "India" and "india.". Text simplification occurred through regular expression removal of URLS and user tags (@PMOIndia) and hashtags alongside emojis and digits and punctuation marks and unique characters.

The data refinement process removed standard terms which provide no emotional value through predetermined lists provided by the Natural Language Toolkit (NLTK). Through the use of NLTK or spaCy tokenization techniques the cleaning process split the text into individual word units. A normalized text base with wider applicability could be obtained through lemmatization that transforms words into dictionary bases as well as through stemming that removes word suffixes. The base word "run" collects all its forms, which include running, runs and ran under a single grouping.

The cleaned text data underwent further processing to get saved in the processed text column for visible traceability purposes throughout routine procedures like feature extraction and model building. Higher-quality input depended on this structured sanitized data so the machine learning models could identify political conversation dynamics better.

C. Sentiment Annotation (Labelling)

A sentiment categorisation system needs labelled data for both its training and evaluation processes. The researchers conducted sentiment annotation on the Indian Election Tweets (2019) dataset because the data lacked initial sentiment labels. This research implemented an automated sentiment analysis technique known as VADER (Valence Aware Dictionary and Sentiment Reasoner) for the evaluation. VADER functions excellently on Twitter content and other similar social media platforms due to its rules-based sentiment analysis structure. Each tweet receives a compound score from the system which represents overall sentiment through evaluation of positive and negative and neutral language elements in the text. The developed compound score serves as an indicator to classify tweets into specific sentiment categories.

The selection of VADER-based technique for the Indian Election Tweets dataset occurred because it handled large data quantities efficiently and at scale. The dataset received additional sentiment labels which became the target variable to evaluate models after training.

D. Exploratory Data Analysis (EDA)

The initial phase of data study depended on Exploratory Data Analysis (EDA) to learn about statistical composition within twitter data records. The examination helps detect hidden behavioural patterns and discover inconsistent patterns together with critical textual aspects before constructing the model. The first visualization showed the distribution of political party tweets in order to detect biases and evaluate the tweet balance between BJP and Congress. A boxplot and histogram analysis of tweet lengths was performed in order to understand the range of tweet verbosity because longer texts potentially affect sentiment expression.

A separate set of word clouds gave visual representation of predominant phrases to help readers understand key campaign topics and focus areas within Congress and BJP tweets. Frequency plots indicated the most popular phrases and hashtags throughout competitive activities between each party and their supporters. The distribution of positive and negative and neutral sentiment was plotted as a first step either through automated VADER tool annotation or manual pre-annotation methods were used for sentiment labeling. Selecting appropriate features and creating powerful sentiment classification methods started from the EDA phase.

E. Feature Extraction

Raw textual input needs feature extraction for converting it into numerical data which machine learning algorithms understand. The analysis employed Word Embeddings alongside TF-IDF together with Bag of Words (Bow) as its main representations.

1. Bag of Words (Bow):

A sentiment classification data preparation stage utilized Bag of Words (BoW) methodology which operated through the Count Vectorizer function at max_features 5000. The BoW technique converts content into a fixed-length vector through word frequency counts while disregarding both grammatical structure and word placement. BoW technique represents each tweet as a vector that shows presence or absence of rules about frequently occurring phrases from all tweets within the lexicon. BoW provides suitable support to numerous NLP tasks particularly sentiment categorization since it successfully reads word frequency patterns which can relate to sentiment direction despite its straightforward nature. The reduction of the feature space to include 5000 words enables better preservation of important textual information at lower computational cost.

2. **TF-IDF (Term Frequency–Inverse Document Frequency):**

Vectorisation of previous text data occurred through TF-IDF (Term Frequency-Inverse Document Frequency) which processed textual patterns for machine learning model usage. The TfidfVectorizer maintained 5000 most informative words after its initialization with the max_features parameter. Using this processing method reduced the features space dimensions to maintain essential information. The vectorizer was then fitted and applied to the cleaned_text column of the dataset using the fit_transform( ) method. In this step, a sparse matrix was produced, where each column represented a term and each row represented a tweet. The sparse matrix's values show how important the term is in that particular tweet in comparison to the entire corpus. The resulting TF-IDF feature matrix was utilised as input by many machine learning classifiers, which improved the models' ability to learn sentiment label patterns by down-weighting frequently used words and highlighting contextually meaningful ones.

3. Word Embeddings:

In order to capture the semantic links between words, Word Embeddings, such Word2Vec, were optionally examined. Similar words are clustered together in a high-dimensional space in the dense vector representations that these models produce. Because embeddings maintain the context and meaning of words, unlike Bow or TF-IDF, they are especially useful for understanding sentiment and opinion-based things in political tweets.

F. Model Building

To classify political tweets based on their emotional content, a number of traditional machine learning approaches were applied. Each model adds unique characteristics and benefits to the text classification process. Among the models that have been applied are XG Boost, Naive Bayes, Support Vector Machine (SVM), and Logistic Regression.

1. Logistic Regression

Logistic regression was one of the primary machine learning methods utilised in this study for sentiment classification. This supervised learning method is used for binary and multiclass classification tasks.The sigmoid function in logistic regression enables probability score prediction to transform continuous values into discrete categories while linear regression produces continuous outputs. The experimental data taken from Twitter went through pre-processing and then got converted to numerical features through TF-IDF and Bag of Words algorithm applications. The logistic regression model received feature vectors for training the sentiment association between word frequency and importance ratings with their corresponding classification types (negative, positive or neutral). The model computes sentiment class probabilities by first using the sigmoid function to perform weight-based mathematical operations on input characteristics. The prediction model selects the class which has the greatest probability value. Maximum likelihood estimation serves the model to adjust its weights during training to minimize prediction labels mismatches with the actual sentiment values.

2.Decision Tree Classifier

The Decision Tree classifier implements a decision tree structure which defines feature conditions regarding each internal node and connects branches to sentiment classes through leaf nodes. This research made use of the TF-IDF together with the Bag of Words (BoW) methods for converting tweets into structured vector inputs. The Decision Tree algorithm processed this input to develop sequence of if-else rules that progressively created homogenous sentiment-based subgroups. The model needed to achieve better sentiment group separation by reducing split-level impurities through Gini index or entropy measurements. Her ability to handle intricate decision boundaries and identify non-linear data relationships makes decision trees an excellent tool for situations where text sentiment arises from term groupings rather than single words. These models provide benefits through their visual interpretability because users need to understand forecasting mechanisms. The model functions optimally for political sentiment analysis because it tracks meaningful word associations and the decision-making paths which serve as indicators of prevailing public opinion. The model structure enables users to find out which content elements specifically contribute most to sentiment detection results.

3.Naive Bayes Classifier

The research utilized Naive Bayes classifier as its fundamental method for sentiment classification. The probabilistic analysis based on Bayes' Theorem determines the likelihood of any class when operating with specific attributes. Highlighting independence between features produces superior text classification outcomes even though such relationships cannot exist in language data. A transformation of cleaned Twitter text into numerical attributes occurred through the combination of TF-IDF and Bag of Words vectorisation approaches. With the obtained characteristics the Naive Bayes model calculated how likely each tweet was to belong to a particular attitude class based on phrase distribution patterns and frequency. The Multinomial Naive Bayes variant was chosen because it shows outstanding performance for modelling attributes that consist of word counts or TF-IDF scores. Training enables the classifier to understand the base probability of sentiment classes besides discovering word probabilities linked to each sentiment class. A prediction involves calculating the posterior probability of every class then selecting the label showing the highest value. The authors implemented Naive Bayes at the beginning of their study since it provides a basic yet efficient model due to its simplicity and low execution costs. Naive Bayes functions optimally under two conditions that frequently occur in tweet sentiment analysis which are large and sparse text data as well as linearly separable data.

4. XG Boost (Extreme Gradient Boosting)

XGBoost (Extreme Gradient Boosting) functioned as the most advanced classification model for sentiment analysis within this research. The gradient boosting ensemble approach creates a sequence of decision trees through its framework to form a large final model. The model accuracy becomes stronger after each sequenced tree training session since they learn to fix errors from preceding models. The cleaned Twitter data underwent vectorization processing through the combination of TF-IDF and Bag of Words methods to generate input characteristics for this project. The XGBoost learning method optimizes its prediction tasks by maximizing a function which imposes model complexity limitations alongside accuracy metrics on false predictions. XGBoost achieves excellent results when analyzing data sets that are extensive and contain mainly empty values like those derived from text documents. Tree pruning together with L1 and L2 regularisation and parallel processing make it a powerful and efficient method. The various characteristics of XGBoost position it well to predict the complex lexical patterns discovered in political tweets.

The work adopted XGBoost to extract intricate patterns between sentiment-carrying words as well as their interconnections. The increased training duration compared to basic models does not reduce the value of this model for sentiment analysis because it competently handles complex non-linear relationships and improves performance.

5.Linear Support Vector Classifier (Linear SVC)

Linear SVC operates as an SVM variant to analyze high-dimensional sparse datasets and applies to text data that uses TF-IDF or Bag of Words feature extraction approaches. The Linear SVC model processed vectorised twitter data as it was trained to split attitudes into three groups labeled positive, negative and neutral. The model determines the best hyperplane among all possibilities in order to separate sentiment classes. The decision boundary of the hyperplane enables exact classification of new data based on metrical relations between tweets and their positioning along the optimal boundary. The linear SVC algorithm excels at processing wide data sets because it specializes in classification duties while other forecasting models provide probabilities. Hinge loss training makes the model excel at linearly separable data while decreasing misclassification mistakes.

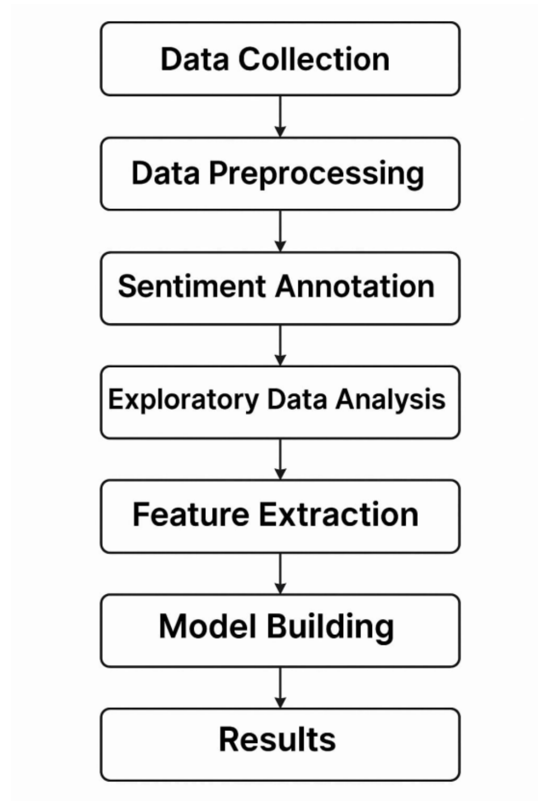

Fig: Flow chart

IV. RESULTS AND DISCUSSIONS

To identify the best method for classifying the sentiment of political tweets, the labelled dataset was used to assess the performance of several ml models. Among the models were XG Boost, Decision Tree, Linear SVC, Naive Bayes, and Logistic Regression.

Model Evaluation Metrics:

Each model was assessed using significant categorisation metrics such as these.

1. **Accuracy:** Shows the proportion of emotions that were correctly predicted.
   Accuracy = (TP + TN) / (TP + TN + FP + FN)

TP (True Positive): The correct identification of positive cases between predicted and actual instances. The model projection of "positive sentiment" matched with the actual analysis result of positive sentiment.

TN (True Negatives): A correct prediction of negative instances counts among true negative cases. The model correctly identified "negative sentiment" while the true sentiment existed at this level.

FP (False Positives): Positive instances which were mis predicted as contrary to actual truth lead to Type I errors. The model indicated "positive sentiment" despite the situation having a negative actual sentiment.

FN (False Negatives): A Type II error reflects the number of mis predicted negative instances. The model identified "negative sentiment" while the true sentiment held a positive tone.

Performance Overview:

The models with the greatest prediction performance on the political twitter dataset were Decision Tree and Logistic Regression, and Linear SVC. These models were trained using TF-IDF vectorisation, which provided more meaningful representation by considering the importance of words throughout the corpus.

Figure 1 and Table 1 illustrate the accuracy of TF-IDF features with five different classification models. The **Decision Tree** model achieved the highest accuracy of **93.40%**, followed by **Logistic Regression** at **82.74%**, and **Linear SVC** at **82.42%**. In contrast, **XG Boost** and **Naive Bayes** yielded lower accuracies of **79.85%** and **74.10%**, respectively. These results indicate that tree-based and linear models are particularly effective in capturing sentiment cues from political text data when combined with TF-IDF features.

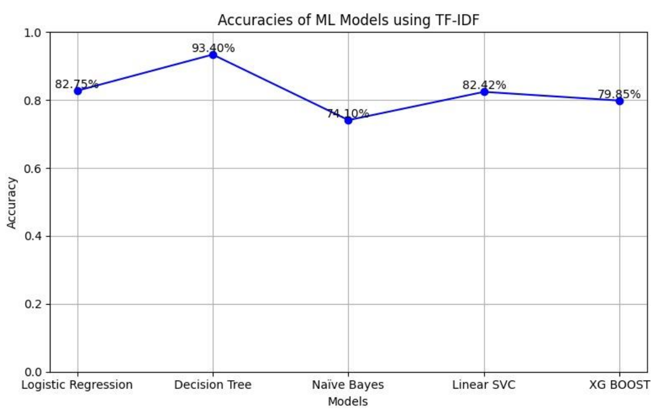

Fig1 Comparison of ML Models using TF-IDF

TABLE-I

ML Models(combined with TF-IDF)	**Accuracy**
Logistic Regression	82.75%
Decision Tree	93.40%
Naïve-Bayes	74.10%
Linear SVC	82.42%
XG BOOST	79.85%

The accuracy of five different ml models using the Bow feature extraction technique is displayed in Figure 2 and Table 2. The accuracy of the Bow model was 82.74% with Logistic Regression, 74.10% with Naive Bayes, 82.42% with Linear SVC, 79.85% with XG Boost, and an incredible 93.40% with the Decision Tree model. In conjunction with this traditional feature extraction strategy, the Decision Tree classifier demonstrated remarkable performance in political tweet classification, yielding the highest accuracy across all model combinations utilising the Bow methodology.

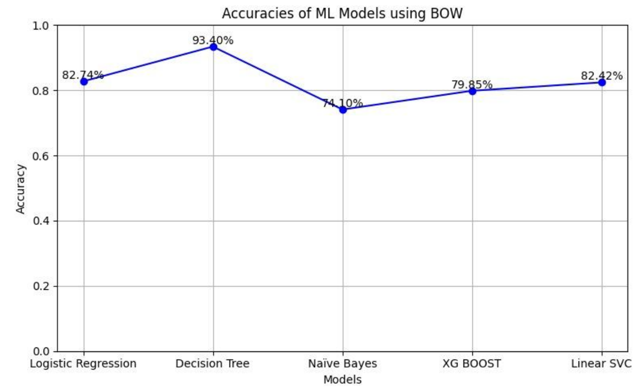

Fig2: Comparison of ML Models using BOW

TABLE-II

ML Models(combined with BOW)	**Accuracy**
Logistic Regression	82.74%
Decision Tree	93.40%
Naïve-Bayes	74.10%
XG BOOST	79.85%
Linear SVC	82.42%

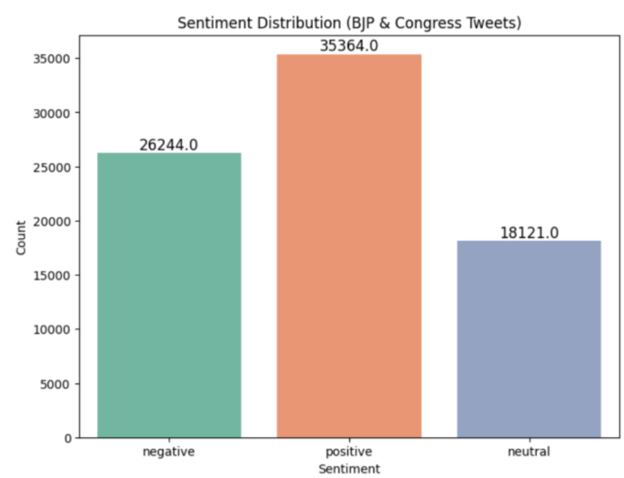

Figure 3 shows the sentiment distribution of tweets related to BJP and Congress. Out of the total tweets, 35,364 are positive, 26,244 are negative, and 18,121 are neutral. This indicates that most political tweets during the 2019 Indian Elections conveyed a positive sentiment, followed by a significant portion of negative and fewer neutral tweets.

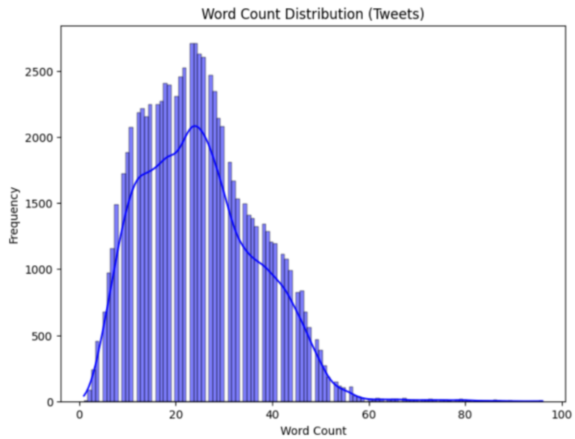

Figure 4 displays the distribution of word counts across all tweets. Most tweets contain between 10 to 30 words, with the peak frequency around 20 words. The distribution is right-skewed, indicating that while most tweets are short, a few contain a significantly higher number of words.

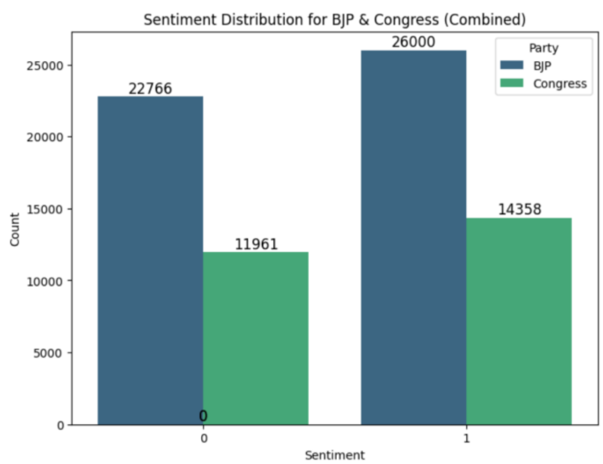

Figure 5: Illustrates the distribution of sentiments (0 = Negative, 1 = Positive) for BJP and Congress tweets. BJP tweets show a higher volume of both positive (26,000) and negative (22,766) sentiments compared to Congress, which has 14,358 positive and 11,961 negative tweets. This indicates a higher level of engagement and sentiment polarity in BJP-related content.

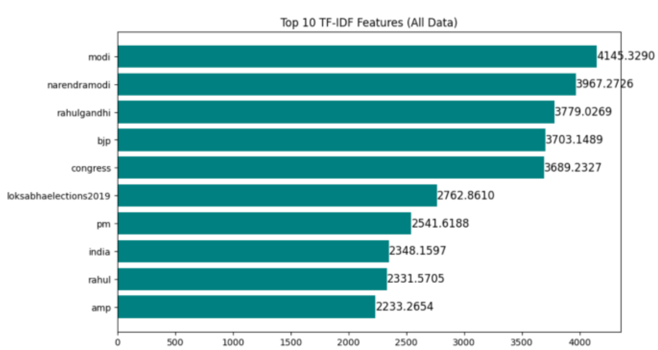

Figure 6: Top 10 TF-IDF Features

Presents the top 10 TF-IDF features extracted from all tweets. The most significant term is "Modi", followed by "Narendra Modi", "Rahul Gandhi", "bjp", and "congress", indicating that the discussion heavily revolves around political leaders and parties. Terms like "loksabhaelections2019", "pm", and "India" also rank high, reflecting the relevance of election-related topics. The high TF-IDF scores suggest that these terms are both frequent and distinctive across the dataset.

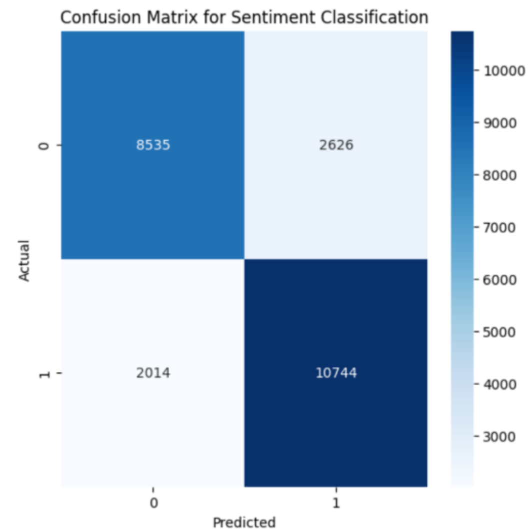

Fig 7: Confusion Matrix for sentiment classification

The confusion matrix in Figure 6 illustrates the performance of the sentiment classification model. While 8,535 tweets with a negative attitude (label 0) were correctly recognised, 2,626 tweets were mistakenly labelled as positive. 10,744 tweets with positive emotion (label 1) were properly identified, whereas 2,014 tweets were mistakenly classified as unfavourable. This demonstrates an excellent overall performance, especially in recognising positive attitudes, despite some misclassification in negative predictions.

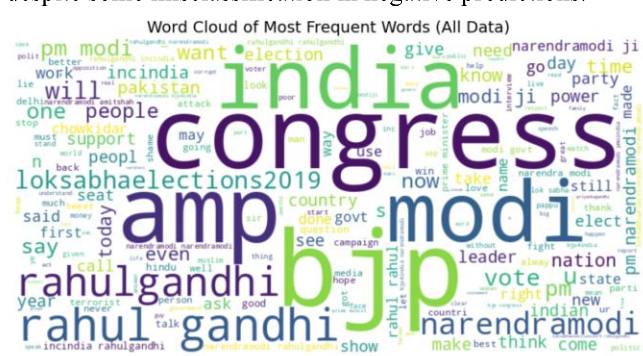

Fig 8: Word Cloud of Most Frequent Words

Figure 8 depicts the main words which appear throughout the entire dataset through a word cloud presentation. The words "congress," "india," "modi," "bjp," and "rahulgandhi" represent the most recurring terms which indicate the significance of political figures and organizations. The election theme of tweets becomes more apparent through usage of words including "loksabhaelections2019" and "amp" and "vote." The widespread use of these particular phrases unveils what major concepts and disagreements shaped public opinion in the 2019 Lok Sabha elections.

V. CONCLUSIONS

The research evaluated the 2019 Lok Sabha elections public sentiment through an NLP examination of tweets from Congress and BJP political parties. The research implementation included the combination of Bow and TF-IDF features with several machine learning models that included Logistic Regression and Naive Bayes alongside XG Boost and Decision Tree and Linear SVC.

The implementation of Decision Tree classifier with TF-IDF produced the highest levels of sentiment classification accuracy at 93.40%. The proportion of positive tweets surpassed neutral ones in the mood distribution indicating overall positive sentiments in political discussions. Senders' public attention toward the election season became evident through word frequency analysis combined with TF-IDF scores which demonstrated crucial political figures and parties were prominent. Public opinion monitoring on social media becomes feasible through combining machine learning systems with appropriate feature extraction approaches according to the findings. The gathered data enables voter mood evaluation improvements for data scientists and political strategists who use media analysts to create enhanced campaign strategies.

VI. FUTURE WORK

The present study creates a strong foundation for Twitter data-based political sentiment analysis even though additional advancements are achievable. Advanced transformer-based models like BERT or deep learning models like LSTM and Bi-LSTM can significantly improve the accuracy of sentiment classification and contextual understanding.

Furthermore, incorporating tweets in regional Indian languages into the study will provide a more comprehensive and inclusive view of public sentiment across a variety of demographic groups. The development of real-time sentiment monitoring systems might provide campaign managers and political analysts access to information instantly during election seasons.

Additionally, aspect-based sentiment research can help determine how the public views specific political individuals, policies, or events. Finally, the sentiment analysis would be guaranteed to reflect real public opinion by eliminating artificial or spam content with the use of bot recognition technologies. The scope, precision, and use of sentiment analysis in political science should all be enhanced by these possible avenues.

REFERENCES

[1] R.Vyas, “kaggle,” 2023. Available:https://www.kaggle.com/datasets/rishabhvyas/indian-election-tweets.

[2] P.sharma, Prediction of Indian election using sentiment analysis on Hindi Twitter, Washington, 2016.

[3] Elghazaly, “Political sentiment analysis using twitter data.,” 2016.

[4] A.Bhowmik, “Opinion Mining and Sentimental Analysis of Twitter Data During India’s General Elections,,” 2024.

[5] Ahmed, “The 2014 Indian elections on Twitter : A comparison of campaign strategies of political parties.,” 2016.

[6] Tarek, “Political sentiment analysis using twitter data.,” 2016.

[7] Bhola, “Twitter and Polls : Analysing and estimating political orientation of Twitter users in India General #Elections 2014.,” 2014.

[8] Gupta, Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly, 2017.

[9] Kawathekar, Prediction of 2019 Indian Election Using Sentiment Analysis, 2018.

[10] Khare, Sentiment Analysis and Sarcasm Detection of Indian General Election Tweets.

[11] Khurana, “Election Result Prediction Using Twitter Sentiments Analysis,,” 2020.

[12] Z. Ansari, Analysis of Political Sentiment Orientations on Twitter, 2020.

[13] Batra, Election Result Prediction Using Twitter Sentiments Analysis, 2020.

[14] Passi, Twitter Sentiment Analysis of the 2019 Indian Election.

[15] Ramteke, Election result prediction using Twitter sentiment analysis, 2016.

[16] Sharma, Sentimental Analysis of Twitter Data with respect to General Elections in India, 2020.

[17] Mishra, Sentiment analysis of Twitter data: Case study on digital India, 2016.