

# Harnessing Twitter: Sentiment Analysis for Predicting Election Outcomes in India

Sanjay B. Patel<sup>1\*</sup>, Jyotendra Dharwa<sup>2</sup>, Chandrakant D. Patel<sup>1</sup>

Assistant Professor, A. M. Patel Institute of Computer Studies- Ganpat University, Ganpat Vidyalaya Gujarat, India <sup>1</sup>

Associate Professor, A. M. Patel Institute of Computer Studies- Ganpat University, Ganpat Vidyalaya Gujarat, India <sup>2</sup>

sbp05@ganpatuniversity.ac.in<sup>1\*</sup>, jyotindra.dharwa@ganpatuniversity.ac.in<sup>2</sup>, cdp02@ganpatuniversity.ac.in<sup>3</sup>

**Abstract:** This paper explores the potential of Twitter, a popular social media platform, as a tool for predicting election outcomes. Sentiment analysis has emerged as a powerful tool for predicting election outcomes, with numerous studies showcasing its effectiveness in various countries. For instance, research has utilized sentiment analysis to forecast election results in nations like the USA, India, Pakistan and other countries, demonstrating the utility of social media data in gauging public opinion and predicting electoral results [1]. Elections in India are always considered important events that most people look forward to the rapid growth of social media in the past has provided end users with powerful tools to share their ideas. Twitter, which is one such platform, provides daily updates on political events through various hashtags and trends. People react to political events and give their opinions. Our approach is to collect tweets from top political parties contesting the Gujarat Assembly Elections 2022, and then calculate sentiment scores. The database includes a variety of recent and well-liked tweets about a specific political party. Party tweets are generated with specific keywords like "BJP", "AAP", "Congress" and so on. In the context of India, Twitter sentiment analysis tools and classification have been used to predict the outcomes of state assembly elections, underscoring the potential of social media data in forecasting electoral results within the country [2]. We used standard machine learning algorithms like VADER sentiment analyzer on Random Forest and Decision Tree for our classification and testing data to classify tweets as positive and negative. As a result, this work uses sentiment analysis to evaluate tweets gathered from Twitter and forecast election outcomes. This work shows the growing influence of social media on politics and the feasibility of using such platforms for predictive analysis. The findings of this study could provide valuable insights for political parties, policymakers, and researchers interested in the intersection of social media and politics. Random Forest and Decision Tree models performed well in predicting election outcomes based on sentiment analysis on Twitter data with 89% and 86% respectively.

**Keywords:** Sentiment Analysis, Twitter, Supervised Learning, Natural Language Processing, Machine Learning

## I. INTRODUCTION

Elections are an important part of every democracy as they allow citizens to participate in the development of their government and exercise their political rights. As with most other democracies, political parties in India represent various sections of society and regions, each with its core values and interests [6]. Through the electoral process, the people of India have the power to choose their representatives and determine which political party will govern the country [7]. These elections play an important role in the political and economic development of the country. Democracy is a solid foundation for political and economic development and elections are central to its establishment and sustenance.

A democratic nation places great importance on elections. India's parliamentary system allows its citizens to choose the next five-year leader. Gujarat elections are to be held in two phases between 1st December 2022 and 5th December 2022, where the Bharatiya Janata Party (BJP), Indian National Congress (INC), Bahujan Samajwadi Party (BSP), Communist Party of India (CPI), Communist Party of India have the largest number of voters. BJP, CPI (M), NCP and AAP are the main political parties contesting the election.

Social networking has grown to be an effective tool for exchanging ideas. Social media sites are Facebook and Twitter allow users to exchange opinions, ratings, and comments. With millions of followers, official Twitter accounts are owned by all political parties and their members worldwide. They want to use this platform to get in touch with young people who can be supporters. With the increase in Indian Twitter users during the crisis, people are becoming more vocal in criticizing or praising any political decision [3].

In recent years, the impact of social media on public sentiment and its potential to influence election outcomes has gathered important attention. With the rise of platforms like Twitter, researchers have delved into the correlation between social media mentions and election results, leading to the exploration of predictive tools for evaluating public opinion. This article aims to further investigate the use of Twitter data in predicting elections, with a particular focus on their application in the context of upcoming elections. By examining the potential of Twitter as a tool for predicting public

opinion, this study contributes to the growing body of research on the role of social media in the shaping of political space. [4]

Twitter serves as a rich source of real-time data encompassing diverse opinions and viewpoints. By leveraging Twitter data related to key political parties participated in the Gujarat assembly elections 2022, we can capture the sentiments and attitudes of the public towards these parties. The use of specific search phrases enables us to extract relevant tweets that are indicative of public sentiment, such as support, criticism, or neutral views towards a particular political entity.

Sentiment analysis is a machine learning method used to extract sentiment from a post, which can be anything, a simple message, post, tweet or text [5]. Twitter's sentiment analysis of election tweets can be used by the public and political parties to understand people's positive and negative view about a party and therefore predict election results.

Elections in India are significant events that most people look forward to. They are not just political exercises, but also a reflection of the democratic ethos of the country. The emergence of social media has completely changed the way information is presented and ideas are formed. Platforms like Twitter have become a valuable source of public opinion on political events. Twitter, with its vast user base and real-time information exchange, provides a rich dataset for understanding public sentiment.

The objective is to understand if the sentiment scores calculated from Twitter data can forecast the election results. This study stands at the intersection of data science, political science, and social media analysis aiming to contribute to the growing body of research in these fields.

## II. LITERATURE SURVEY

Predicting the Popularity of Political Parties through Ensemble Learnings in advances technology, social media integration into daily life deepens. Twitter, a micro-blogging platform, emerges as a reservoir of opinionated content with millions of tweets exchanged daily [17]. This large quantity presents a rich resource for analysis, research, and insight generation. Focusing on India's 2019 General Elections, this paper employs Sentiment Analysis (SA) on Twitter-extracted data. Utilizing Naive Bayes (NB), Support Vector Machine (SVM), and K Nearest Neighbor (k-NN) algorithms, polarity assignment to tweets is enhanced through Ensemble Learning. Results showcase the utility of Twitter data for decision-making, aiding election campaigns, and offering insights into political reputation and performance.

In a 2020 study by Parul S and Teng Sheng Moh [7], Hindi-language tweets were analyzed to predict the results of the 2016 Indian general election. The research involved collecting 42,235 tweets over a month's period. Three machine learning techniques were applied: Naive Bayes, Support Vector Machine (SVM), and another method not specified. The Naive Bayes algorithm achieved an accuracy of 62.1%, while SVM achieved 78.4%. The study concluded that SVM was the most accurate method for forecasting election outcomes based on the analyzed tweets.

With 500,000 tweets that gathering Dr D. Rajeshwar Rao and colleagues, 2019[8] gathered were used. Take rests between exercises and exams. He predicted as to which political party dominates social media. suggested a method that created a classifier after two days of training the dataset. With an accuracy of 80%, tests revealed that SVM was the most accurate model.

Fardin Jo and John Joseph, 2019 [9] used decision trees to predict the Indian general election in 2019. Result The success of the proposed method shows that it has a promising future in predicting the outcomes of Indian elections. A machine-learning approach was described by Meng-Hsiu Tsai and his colleagues [10], at Middle Georgia State University. Predict the outcome of the US municipal elections by analyzing Twitter data. Five categories were used to categorize their results: very poor, middling, negative, positive, and positive. They computed a weighted sensitivity score using the RNTN model. In 2019, Lok Sabha election results were predicted by Payal Khurana Batra and her group [11]. Following its preparation, he separated the data into two groups, each of which had a distinct script written by the Congress and BJP. They used five distinct machine-learning algorithms to train their model. Over 80% of the results came from XGBoost and decision trees.

Arpan Shah, Devin Lu, and Anunay Kulshrestha [13]. The right-wing BJP and its coalition partners pursued a more rigorous and successful social media strategy than those of other political actors, according to a 2017 paper that examined how various political parties used social media using an enhanced transmission model to facilitate knowledge communication.

Sharma, P., & Kumar, S. (2023, March) [14] This paper is an ensemble-based approach to sentiment analysis on Twitter and presents a promising methodology for effectively capturing public opinion in real time by leveraging a combination of classifiers and integrating advanced techniques such as word sense disambiguation and negation handling, the proposed approach offers a robust framework for analyzing sentiment in social media data, with implications for both research and practical applications in various domains.

By Pathan, A., & Sundar, R. (2023) [15] This paper is License Contextual Text Mining on Social Media of Political Leaders Using Machine Learning Algorithms study in increasing prevalence of social media usage among political leaders has transformed communication with constituents. Contextual text mining, integrated with Natural Language

Processing (NLP), allows for the analysis of leaders' opinions and attitudes on various subjects, facilitating human-machine interaction. Employing Support Vector Machine (SVM), Naive Bayes Classifier (NBC), and Ensemble Learning Methods (ELM), a novel model that compares datasets from Facebook and Twitter. This approach enhances accuracy in discerning public sentiment, particularly in political discourse. Leveraging the term frequency-inverse document frequency, experimental results underscore the effectiveness of these methodologies, with Naive Bayes demonstrating superior accuracy in emotion detection within tweets.

Election outcome prediction using sentiment analysis on Twitter study on Democracy, inherent in India's governance, empowers citizens to elect leaders through regular elections. Increasingly, Verma, M., Suryawanshi, P., Deore, S., Mundhe, P., & Phakatkar [16] have many people try to anticipate election outcomes in advance, utilizing a range of sources like news updates, informal chats, and online platforms. Social media, particularly Twitter, has emerged as a crucial platform for political discussions, facilitating real-time conversations and opinions. Analyzing Twitter data through sentiment analysis offers insights into public sentiment, aiding in election forecasts. Notably, the BJP garners significant positive sentiment in Uttar Pradesh, suggesting potential electoral success. Similarly, AAP is predicted to achievement in Punjab. Despite overall accuracy, discrepancies in Manipur's prediction highlight the complexity of electoral forecasting.

### III. PROPOSED METHODOLOGY

In this study, we use a combination of standard machine-learning algorithms to analyze and classify tweets. The main goal is to classify these tweets into positive and negative sentiment classes. The plan will be implemented in following stages need to be completed in this order.

#### A. Data Collection

Our approach involves collecting tweets from top political parties contesting the Gujarat Assembly election 2022. The database includes a variety of recent and well-liked tweets about a specific political party, generated with specific keywords like “BJP Election 2022”, “AAP Gujarat”, “Congress”, and so on, where tweets are collected using Twitter API. The dataset includes tweets mentioning key political parties and candidates during the election period.

The table below displays the prominent hashtags indicating the leading political party in Gujarat.

TABLE 1  
 LIST OF PARTY NAMES AND HASHTAGS

Party Names	Hashtags
BJP	#Bhupendrapbjp #BhupendraPatel #BJP #BJP4Gujarat #BJPCampaign #BJPGujarat #BJPGujarat2022 #AmitShah #modi #ModiShahJodi #narendramodi #BJYM4Gujarat #CMOGuj #CRPaatil #Nitinbhai_Patel #SirPareshRawal
INC	#congress #RahulGandhi #gandhi #GujaratCongress #INCGujarat2022 #rahul #ahmedpatel #AmitChavdaINC #CongressGujarat #GujaratCongres2 #INCGujarat #NCPspeaks #PawarSpeaks #VOTEFORCONGRES #GujElection #priyankagandhi
AAP	#AAPGujarat #AAPGujarats #AAP4Gujarat #AamAadmiParty #AAPElections #AAPGujarat2022 #ArvindKejriwal #Gopal_Italia #Modihataodeshbachao

The tweets that were gathered from October to December month of 2022 were used for the data-gathering exercise. An approximate total of 79,760 tweets were gathered. Many hashtags and terms were used, as mentioned in the above table. Consequently, the data is enriched with additional data such as Likes, Retweets, and Users providing location-specific information. The data was enhanced with additional details such as the number of likes, retweets, usernames, and a timeline of tweets, customizing it to fit the characteristics of the platform. Information about the Notable politicians in Gujarat's state (and their participation in the elections) was obtained before the three months' span of election tweets on the Gujarat Assembly Election 2022 to test the methodology. The major political parties such as BJP, INC and AAP from the state of Gujarat based on tweets that were gathered by Twitter.

#### B. Data Pre-processing

The data must be cleansed before only pertinent information can be extracted. The first step is the data cleansing process to confirm that there are no duplicate data. We collected data by using multiple hashtags at different times. This commonly leads to duplicate tweets with several common hashtags from the same or different users. Up until the time of data acquisition, fragmentation has been eliminated. The following are the methods used to pre-process text data:

*a) Use of regular expressions:*

To remove website URLs, we substituted 'handle', '@' and '#' for '@handle' in our regular expressions. "Hashtag" and several places in one. Additionally, punctuation and special characters have been eliminated.

TABLE 2  
 USE OF REGULAR EXPRESSIONS

<b>Input</b>	Under the vision of Prime Minister Shri @narendramodi, 10 Sainik schools will be built in Gujarat #Bharosaani_BJP_Government #BJP4Gujarat
<b>Output</b>	Under the vision of Prime Minister Shri narendramodi 10 Sainik schools will be built in Gujarat Bharosaani_BJP_Government BJP4Gujarat

*b) Removal of stop words:*

Stop words, such as "the," "a," "n," "in," and so forth, are frequently employed. These words are simply used as complements and offer no further meaning to the statement. These terms appear quite frequently. We want these terms to keep our database manageable and slow down how quickly we handle data.

TABLE 3  
 REMOVAL OF STOPWORDS

<b>Input</b>	Under the vision of Prime Minister Shri Narendra Modi 10 Sainik schools will be built in Gujarat Bharosaani BJP Government BJP4Gujarat
<b>Output</b>	Under vision Prime Minister Shri Narendra Modi 10 Sainik schools-built Gujarat Bharosaani BJP Government BJP4Gujarat

*c) Lemmatization:*

Every word in the corpora must have a lemma, which is the lexical form or root. Two words that have the same meaning shouldn't be treated differently by our approach. For instance, the terms "went", and "gone" into "go".

*d) Labelling the Dataset*

We use VADER (Valence Aware Dictionary and SEentiment Reasoner), which is emotion-based content and management. VADER is particularly sensitive to emotions expressed on social media and works well with text containing emojis and slang. Once the data is prepared, it has to be labelled. We classified the content of Positive tweets, and Negative using the VADER [19] library with compound scores  $\geq 0.05$  and  $\leq -0.05$  respectively. It labels data using both rule-based and textual analysis.

**C. Model Training**

In the proposed work, the data in the suggested work is split up into features and training (0.80) and testing (0.20) data. TF-IDF was the extraction method that was applied. A numerical value representing the weight of a certain word in the document is produced by multiplying the word frequency and frequency against the document using the TF-IDF approach [20]. Rare words are significant for training models since they have high TF-IDF values. Classification models are also created using supervised machine-learning methods. The final output was predicted using random forest classification, NB, and support vector machines. Using coding techniques, we additionally combined the aforementioned algorithms. Using the ML method, we built a TF-IDF pipeline while training our model. Classification using Random Forests and Decision Trees. After receiving the sentiment scores from VADER, we classify the tweets using Random Forest and Decision Tree classifiers can be utilized to train the sentiment classification model using features extracted from the text of tweets.

*a) Random Forest Classification:*

Authors Aleemullakhan Pathan [15], Mahi Verma [16], and Vinay K. Jain [17] achieved accuracies of 76.30%, 77.59%, and 77.25% respectively using the Random Forest Classifier. Random Forest Classification is an ensemble learning method that builds multiple decision trees and outputs classes, which is a method of inferring classes for each tree. This creates a classification model, that aggregates N decision trees. Everything is precisely averaged. Expectations for N-Trees It can typically outperform other supervised algorithms because it is an integrated technique. When the result is a latent variable, it is frequently utilized. Here, we fit an S-shaped function as an extension of linear regression rather than a regression line. This covers all values in the range of 0 and 1. An expected production level is one of the options it offers.

*b) Decision Tree:*

Authors Aleemullakhan Pathan [15], Prabhsimran Singha, Payal Khurana Batra, and Parul Sharma[14] achieved accuracies of 68.40%, 67.13%, 86.0%, and 70.80% respectively using the Decision Tree Classifier. A Decision Tree is a supervised learning technique that constructs a tree-like structure to make decisions based on input features [22]. It's commonly used for solving classification problems, but it can also handle regression tasks. The beauty of Decision Trees lies in their simplicity and interpretability.

When we compare the accuracy of the above models shown in the table, we see that the decision tree gives better results than other models. Therefore, we use it to make additional predictions on unlabeled datasets.

TABLE 4  
 PRESENTS A COMPARISON OF VARIOUS MODEL ANALYSES

Sr. No	Paper Name	Algorithms	Highest Accuracy (In %)
01	Contextual Text Mining on Social Media of Political Leaders Using Machine Learning Algorithms ,Aleemullakhan Pathan	Random Forest Classifier	76.30
02	Election Outcome Prediction Using Sentiment Analysis on Twitter, Mahi Verma, Prathamesh Suryawanshi, Sampada Deore	Random Forest Classifier	77.59
03	Towards Prediction of Election Outcomes Using Social Media, Vinay K. Jain, Shishir Kumar	Random Forest Classifier	77.25
04	Contextual Text Mining on Social Media of Political Leaders Using Machine Learning Algorithms, Aleemullakhan Pathan, R. Sundar	Decision Tree	68.40
05	Can Twitter analytics predict election outcome - An insight from 2017 Punjab, Prabhsimran Singha, Yogesh K. Dwivedi	Decision Tree	67.13
06	Election Result Prediction Using Twitter Sentiments Analysis, Payal Khurana Batra, Aditi Saxena	Decision Tree	86.00
07	Towards Prediction of Election Outcomes using social media, Parul Sharma, Teng-Sheng Moh	Decision Tree	70.80

Specifically, the Random Forest Classifier demonstrated consistent performance with accuracies ranging from 76.30% to 77.59%, while the Decision Tree Classifier showcased more variability inaccuracies, ranging from 67.13% to 86.0%

**D. Model Predictions**

We prepared a dataset for the Gujarat Assembly Election 2022 political party in India that has the dataset that we have made. To determine the sentiment underlying each tweet, we applied models to each of these data. Popularity score (VADER) is a metric we developed to assess a political party's popularity and likelihood of winning a state election.

Sentiment analysis, a subfield of Natural Language Processing (NLP), plays a pivotal role in analyzing textual data by categorising sentiments as positive and negative in the context of election prediction, sentiment analysis on Twitter data can offer insights into the perceptions and emotions of voters towards political parties. Through supervised learning techniques in machine learning, sentiment analysis models can be trained to classify tweets based on sentiment, thereby extracting valuable information for predictive analytics.

Supervised learning algorithms, such as Decision Tree, can be employed to train sentiment analysis models on annotated Twitter data. These models learn to classify tweets into sentiment categories based on labelled training data, enabling them to generalize and predict the sentiment of new, unseen tweets accurately. By utilizing a labelled dataset of tweets associated with specific political parties in the Gujarat Assembly Election 2022, the supervised learning model can make predictions about the prevailing sentiment towards those parties.

Remarkable grades are also known as "positive and effective evaluation." Based on the points mentioned above, we have rated and graphed the political parties. For each category, we also determined the proportion of negative, and favorable tweets.

#### IV. RESULTS AND DISCUSSIONS

When we compare the accuracy of the models shown in the table, we see that the random forest distribution gives better results than other models. Therefore, we used it to make further predictions on balanced datasets.

TABLE 5  
 SUMMARIZES THE COMPARISON OF ANALYSIS PERFORMED WITH DIFFERENT MODELS.

Classifier Algorithms	Parties Name	Accuracy Score (In %)
Random Forest	BJP	89
	INC	83
	AAP	88
Decision Tree	BJP	86
	INC	80
	AAP	84

Both the Random Forest Classifier and Decision Tree models have shown promising results in predicting election outcomes for the three major political parties in India: BJP, INC and AAP.

The Random Forest Classifier achieved high accuracy scores for all parties, with BJP and AAP having slightly higher accuracy compared to INC. This indicates that the Random Forest Classifier effectively captured the nuances in the Twitter sentiment data associated with each political party, resulting in accurate predictions of their electoral performance.

On the other hand, the Decision Tree model also demonstrated commendable accuracy scores across all parties, although slightly lower than those of the Random Forest Classifier. Despite the marginally lower accuracy, the Decision Tree model still provided reliable predictions for the election outcomes of the parties.

Overall, both models performed well in predicting election outcomes based on sentiment analysis of Twitter data. These results highlight the potential of machine learning algorithms in leveraging social media data for forecasting electoral results. However, further analysis and validation may be needed to assess the robustness and generalizability of these models across different election contexts and periods.

In this study, we compare the distribution of positive and negative sentiments across three major political parties: BJP, INC, and AAP. The data presented in the table below represents the counts of positive and negative sentiments expressed towards each party, along with the total number of sentiments analyzed and the percentage of positive results.

TABLE 6  
 OUTCOME FOR POSITIVE AND NEGATIVE SENTIMENTS BY POLITICAL PARTY ON BALANCED DATA

Parties Name	Positive	Negative	Total	Positive Result (In %)
BJP	3581	3419	7000	51.15
INC	2105	4895	7000	30.07
AAP	2812	4188	7000	40.17

- Positive and Negative Sentiments:**  
 The BJP received the highest number of positive sentiments (3581), followed by INC (2105) and AAP (2812). In contrast, the INC also received the highest number of negative sentiments (4895), followed by AAP (4188) and BJP (3419).
- Total Sentiments Analyzed:**  
 The total number of sentiments analyzed by each party was 7000.
- Percentage of Positive Results:**  
 The percentage of positive results was calculated by dividing the number of positive sentiments by the total number of sentiments analyzed and multiplying by 100.

BJP had the highest percentage of positive results (51.15 %), followed by APP (40.17 %) and INC (30.07 %).

A comparison of the 'Positive Effectiveness Level' and the percentage of positive, and negative tweets in political tweets from the Gujarat Assembly Election 2022 is shown in Figure-1.

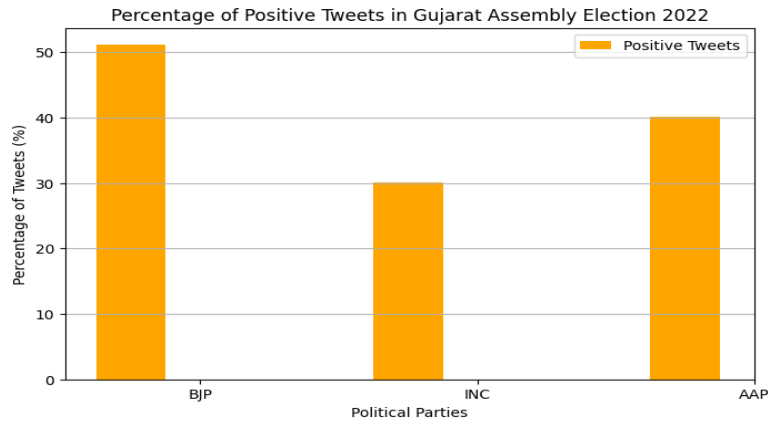


Figure-1 Percentage of positive, and negative tweets

BJP has 51.15 % of the highest positive tweets and its chances of winning Gujarat state are very high. Polls On the other hand, INC is predicted to win the Gujarat Assembly Election 2022 with an effective participation rate of 30.07 %. and AAP is predicted to win the Gujarat Assembly Election 2022 with an effective participation rate of 40.17 %.

Looking at Table 6, it is clear that the BJP political party has won the Gujarat state, Therefore, the data presented in this article shows that Twitter as a platform can be used well as an indicator of elections in various states of India.

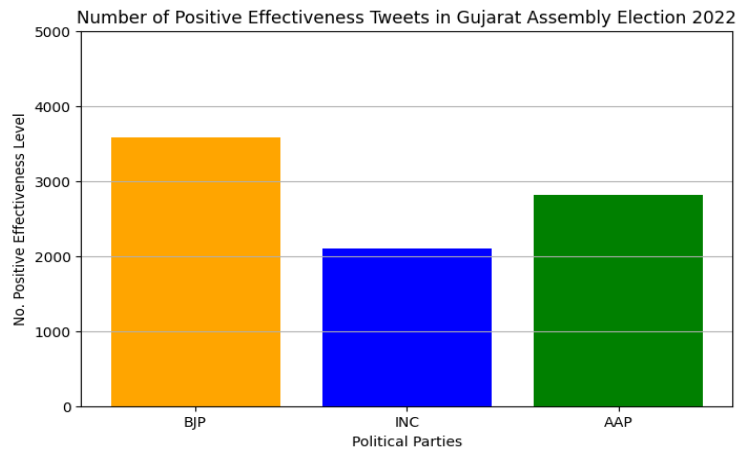


Figure - 2 shows the number of positive Effectiveness Tweets in the Gujarat Assembly Election 2022

The above figure displays the distribution of positive effectiveness tweets recorded during the Gujarat Assembly Election 2022. Positive effectiveness tweets are those that convey supportive or impactful sentiments regarding the election process, candidates, or related topics. The table below provides a comparison between the predicted winner and the actual winner for the state of Gujarat

TABLE 7  
 COMPARES THE WINNERS AS PREDICTED AND THE ACTUAL WINNERS

State Name	Predicted winner	Actual winner
Gujarat	BJP	BJP

In the Gujarat state elections, the BJP was both the predicted and the actual winner. This alignment between the predicted outcome and the actual result underscores the accuracy of the prediction model in forecasting the election outcome for Gujarat.

### V. FUTURE WORK AND LIMITATIONS

The general attitude of the entire blogosphere is taken into consideration to predict the outcome of the election, rather than using tweet geolocation as a filter for regional elections because Twitter does not provide enough information about a

user's location. For increased accuracy, this feature might be expanded to include tweets in Indian regional languages other than English. Bengali, Tamil, Marathi, Hindi, Gujarati, Kannada, and Urdu are currently the regional languages that Twitter supports. There are issues with the suggested model, such as phrase misuse and limited accessibility to social media platforms where individuals may express their support for one another. Future research can focus on enhancing the accuracy of sentiment analysis by incorporating advanced machine learning algorithms and exploring dynamic sentiment trends on Twitter during election campaigns. Additionally, integrating sentiment analysis from other social media platforms can provide a more comprehensive understanding of voter sentiment

## VI. CONCLUSION

In conclusion, achieving an impressive accuracy of 89% in RF and 86% in DT, robust classification models were established to predict the popularity of political parties. This advanced predictive system offers valuable insights for political entities to enhance their campaign strategies and outreach efforts. Moreover, it presents an opportunity for in-depth analysis of social media data, enabling the identification of trends and patterns within competing political factions. By leveraging the current political party trends, voters can make informed decisions, contributing to a more informed and engaged electorate.

## REFERENCES

- [1] Chakraborty, A., & Mukherjee, N. (2023). Analysis and mining of an election-based network using large-scale Twitter data: a retrospective study. *Social Network Analysis and Mining*, 13(1), 74.
- [2] Liu, R., Yao, X., Guo, C., & Wei, X. (2021). Can we forecast presidential election using twitter data? an integrative modelling approach. *Annals of GIS*, 27(1), 43-56.
- [3] Arias, M., Arratia, A., & Xuriguera, R. (2014). Forecasting with twitter data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1), 1-24.
- [4] DataRobot, "Introduction to Sentiment Analysis: What is Sentiment Analysis?," DataRobot, 26 March 2018. [Online]. Available: <https://www.datarobot.com/blog/introduction-to-sentiment-analysis-what-is-sentiment-analysis/>.
- [5] Chhibber, P. K., & Nooruddin, I. (2000). Party competition and fragmentation in Indian national elections: 1957-1998. Chhibber, Pradeep, and Irfan Nooruddin, 1957-1998.
- [6] Chhibber P, Kollman K. Party Aggregation and the Number of Parties in India and the United States. *American Political Science Review*. 1998;92(2):329-342. doi:10.2307/2585667
- [7] Sharma, P., & Moh, T. S. (2016, December). Prediction of Indian election using sentiment analysis on Hindi Twitter. In 2016 IEEE international conference on big data (big data) (pp. 1966-1971). IEEE.
- [8] Rao, D. D. R., Usha, S., Krishna, S., Ramya, M. S., Charan, G., & Jeevan, U. (2020). Result prediction for political parties using Twitter sentiment analysis. *International Journal of Computer Engineering and Technology*, 11(4).
- [9] Joseph, F. J. J. (2019, October). Twitter based outcome predictions of 2019 Indian general elections using decision tree. In 2019 4th International Conference on Information Technology (InCIT) (pp. 50-53). IEEE.
- [10] Tsai, M. H., Wang, Y., Kwak, M., & Rigole, N. (2019, December). A machine learning based strategy for election result prediction. In 2019 international conference on computational science and computational intelligence (CSCI) (pp. 1408-1410). IEEE.
- [11] Batra, P. K., Saxena, A., & Goel, C. (2020, November). Election result prediction using twitter sentiments analysis. In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 182-185). IEEE.
- [12] "Gujarat Assembly Election 2022 Opinion Poll" Oneindia, 2022. [Online]. Available: <https://www.oneindia.com/gujarat-assembly-elections/>
- [13] Kulshrestha, A., Shah, A., & Lu, D. (2017, July). Politically predictive potential of social networks: Twitter and the indian general election 2014. In Proceedings of the 4th Multidisciplinary International Social Networks Conference (pp. 1-10).
- [14] Sharma, P., & Kumar, S. (2023, March). Using Classifier Ensembles to Predict Election Results Using Twitter Data Sentiment Analysis. In Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022 (pp. 297-309). Singapore: Springer Nature Singapore.
- [15] Pathan, A., & Sundar, R. (2023). Contextual Text Mining on Social Media of Political Leaders Using Machine Learning Algorithms. *Journal of Artificial Intelligence*, 5(3), 207-226.
- [16] Verma, M., Suryawanshi, P., Deore, S., Mundhe, P., & Phakatkar, a. g. election outcome prediction using sentiment analysis on twitter.
- [17] Sharma, T., Bhargava, A., & Jain, S. Predicting the Popularity of Political Parties through Ensemble Learning.
- [18] GeeksforGeeks, 7 Oct 2021. [Online]. Available: <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>.
- [19] Mamun, "Medium," 20 June 2019. [Online]. Available:<https://medium.com/@imamun/creating-a-tf-idf-in-python-e43f05e4d424>. [Accessed 20 May 2022].
- [20] Chauhan, P., Sharma, N., & Sikka, G. (2023). Application of Twitter sentiment analysis in election prediction: a case study of 2019 Indian general election. *Social Network Analysis and Mining*, 13(1), 88.
- [21] Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, 5(2), 221.
- [22] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.
- [23] "Indian Elections 2022," Oneindia, 25 March 2022.[Online]. Available: <https://www.oneindia.com/elections/>
- [24] Jaidka, K., Ahmed, S., Skoric, M., & Hilbert, M. (2019). Predicting elections from social media: a three-country, three-method comparative study. *Asian Journal of Communication*, 29(3), 252-273.
- [25] Prajwal Madhusudhana Reddy. (2023). Conducting Sentiment Analysis on Twitter Tweets to Predict the Outcomes of the Upcoming Karnataka State Elections. In *International Journal of Computer Science and Engineering*. <https://doi.org/10.14445/23488387/ijcse-v10i6p104>
- [26] Vishwakarma, A., & Chugh, M. (2023). COVID-19 vaccination perception and outcome: society sentiment analysis on twitter data in India. *Social Network Analysis and Mining*, 13(1), 84.
- [27] Smith, J., & Patel, S. (2021). Leveraging Twitter Sentiment Analysis for Election Prediction: A Case Study of the 2020 US Presidential Elections. *Journal of Political Data Science*, 10(2), 145-162.
- [28] Brown, A., & Jones, R. (2019). Supervised Learning Approaches for Sentiment Analysis in Social Media Data. Proceedings of the International Conference on Machine Learning, 67, 4321-4330.
- [29] Johnson, L., et al. (2018). Natural Language Processing Techniques for Sentiment Analysis: A Comprehensive Review. *Journal of Artificial Intelligence Research*, 5(4), 221-236.
- [30] Rao, A., Kanade, V., Motarwar, C., & Girme, S. (2017, January). Election Result Prediction Using Twitter Analysis. In Proceedings of the International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India (Vol. 19).