SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V12I17P101
Volume 12 Issue 7, 1-5, July 2025

Original Article

Assessing the Role of Positive and Negative Tweets in Predicting Electoral Outcomes: A Study of the 2019 Lok Sabha and 2020 U.S. Elections

Nirvaan Malhotra

Genesis Global School, Delhi, India.

Corresponding Author : nirvaan2607@gmail.com

Received: 04 May 2025 Revised: 13 June 2025 Accepted: 01 July 2025 Published: 16 July 2025

Abstract - *This research paper aims to ascertain whether positive or negative tweets are a more effective indicator for predicting electoral outcomes. The investigation primarily focuses on two major elections: the 2019 Indian Lok Sabha election and the 2020 U.S. presidential election. The study compares sentiment trends of Twitter users against real-world election results. Tweets about the main political parties in each election were collected from Kaggle, a public data source. Each tweet was analyzed using the RoBERTa sentiment analysis model, which assigned positive, negative, and neutral sentiment scores. The strongest positive or negative score was recorded as the final sentiment for each tweet. After classifying the tweets, the ratio of positive tweets between the two parties was calculated and compared to the ratio of votes won in the actual election. Similarly, the ratio of negative tweets between the two parties was calculated and compared to the ratio of votes lost. The findings show that while both forms of sentiment are relevant, positive tweets align more closely with electoral success. In contrast, negative tweets do not reliably predict losses. This suggests that supportive sentiment expressed online may be a stronger indicator of real-world outcomes than negative sentiment alone.*

Keywords - *Electoral prediction, Machine learning, RoBERTa, Sentiment analysis, Social media analytics.*

1. Introduction

Since the widespread inclusion of the internet into our communities and the rise of online platforms, the sphere of political discourse in modern society has transformed dynamically. Social media applications, especially Twitter, have become leading mediums of communication, wherein millions of individuals express their political opinions through brief yet impactful tweets. The interconnected and fast-paced nature of these services has allowed users to instantly react to geopolitical events, comment on political agendas, and engage in debates over laws and approaches to governance. Compared to traditional media such as newspapers, television, and radio, Twitter has revitalized interest and attraction towards political affairs by inviting both leaders and citizens to engage in public discussions inside a quick, direct and highly interactive space.

During major elections and significant political events, Twitter often sees a substantial surge in user engagement, as the public readily expresses its sentiments and perspectives. While politicians use the platform to announce policies, promote campaigns and counter opponents, ordinary citizens use it to express support, raise key concerns or criticize parties and candidates. This endless dialogue, which brims with unique thoughts, ideas, and opinions, helps create a vast collection of digital opinions that can be analyzed to understand how people feel in the lead-up to an election. While some studies have attempted to predict election outcomes using tweet volumes or engagement rates, less attention has been paid to comparing positive versus negative sentiment and how each relates to real-world electoral results.

Numerous studies have explored the use of Twitter sentiment analysis in the context of election predictions. Many of these studies, however, have focused primarily on positive sentiment and overall tweet volume as potential indicators of electoral outcomes. For example, one of the author study by Soni and Mali used sentiment analysis to predict the Karnataka state elections, focusing on positive, negative, and neutral sentiment scores sourced from tweets.

While their research successfully demonstrated that Twitter has the ability to provide crucial insights into understanding public mood, their primary emphasis was on predicting outcomes based on tweet sentiment distribution, relying on the number of positive versus negative tweets to forecast a winner. This approach does not cover investigating whether positive or negative sentiment aligned more strongly with actual electoral success or failure.

Likewise, previous works such as Bermingham and Smeaton [1] and Choy et al. [2] focused on using tweet frequency and engagement to predict election trends. While they used metrics such as number of positive and negative tweets and factored in metrics such as likes, retweets and comments to mirror public alignment better, they did not compare positive versus negative sentiment ratios against real-world election results like vote shares or seats won. Additionally, many of these studies primarily analyzed tweets from official accounts or political figures, rather than capturing the general public's voice. This limits our understanding of how ordinary users express support or dissent, especially during politically sensitive periods like elections.

Moreover, there is a scarcity of research that directly compares the ratios of positive to negative sentiments and correlates these with actual election outcomes. Most existing studies do not employ a structured approach to quantify how each sentiment (positive and negative) aligns with electoral results; they focus on directly comparing the number of positive and negative tweets.

This study aims to fill these gaps by focusing on tweets from regular users, analyzing both positive and negative sentiments using the RoBERTa language model. By comparing the sentiment ratios to actual election outcomes in a structured manner, this research seeks to determine whether positive or negative sentiments on Twitter serve as better indicators of electoral success. This approach introduces a novel perspective on digital political discourse, shedding necessary light on the significance that support (positive tweets) and dissent (negative tweets) have in understanding public opinion during elections.

This research matters because it explores the connection between digital political expression and real-world democratic outcomes. In a world that is ever so complicated, dominated by contrasting opinions and countless pieces of political data flooding the internet, it is essential to understand which of the two, positive or negative online sentiment, better reflects real election results. This key analysis will help political analysts, campaigners, and researchers interpret social media more accurately and determine what some of the content or perspectives shared on these platforms mean for political parties.

The primary research question addressed in this study is: Which type of Twitter sentiment, positive or negative, is a better indicator of electoral outcomes?

2. Methodology

2.1. Research Design

This research paper aims to explore the relationship between public user sentiment on Twitter and real-world electoral outcomes. As social media becomes an increasingly integral aspect of modern political discourse, the need to analyze and dissect this environment and draw accurate conclusions becomes more pronounced. It becomes crucial to understand whether expressions of support or dissent are more closely aligned with how people vote. By focusing on tweets from regular, ordinary users in the period leading up to an election, the study uses sentiment analysis to better understand how digital opinions map onto electoral behaviour, and the distinction that positive and negative tweets signify for elections.

2.2. Sample Selection and Inclusion Criteria

This study focuses on analyzing tweets of major political parties during two significant political elections: the 2019 Indian Lok Sabha election and the 2020 U.S. presidential election. The incorporated sample of tweets deliberately included comments solely from ordinary Twitter users, excluding discourse from any verified political accounts and politicians, to better reflect public opinion through an unfiltered lens.

Tweets were included based on the following criteria:

The tweets mentioned or were tagged with keywords or hashtags related to major political parties (e.g., BJP and Congress for India; Democrat and Republican for the U.S.)

They were posted within a specific time window leading up to the election (2 weeks prior to the election)

They contain actual user-generated original content and are not comments or retweets.

The tweets are in English or translated into English.

2.3. Data Collection

The tweet datasets were taken from Kaggle, an open-source data platform. For each election, two separate datasets were used, one for each major political party:

- Lok Sabha 2019: Tweets related to BJP and Congress
- U.S. Presidential 2020: Tweets related to the Democratic and Republican parties

After downloading the datasets, the following data cleaning steps were applied:

- Removal of blank, duplicate, or non-text tweets
- Filtering of non-English tweets
- Equalization of tweet counts per party within each election to avoid class imbalance.

This cleaning process ensured the dataset was balanced and suitable for comparative sentiment analysis across elections and parties.

2.4. Data Analysis: Training/Classification Techniques

To analyze sentiment, the study used the RoBERTa (Robustly Optimized BERT Pre-training Approach) model through the Hugging Face Transformers library. RoBERTa is a transformer-based language model known for its strong performance in natural language understanding tasks, especially for short, informal texts like tweets.

Each tweet was passed through the model using a pre-trained sentiment classification pipeline, which outputs three probabilities: positive, negative, and neutral. The dominant sentiment (positive or negative) was assigned as the tweet's final classification. Neutral tweets were excluded from the final analysis to maintain focus on clear expressions of support or dissent.

The rationale for selecting RoBERTa includes:

- It outperforms older models (like Naive Bayes or SVM) in understanding context and sarcasm, which is common in political tweets
- It requires no additional training or fine-tuning for general sentiment classification
- It is widely validated in sentiment analysis research, making it a reliable, interpretable tool for this study

To assess the predictive value of sentiment ratios, the study calculated:

- Positive tweet ratios between competing parties
- Negative tweet ratios between competing parties
- Deviations between sentiment ratios and actual electoral outcomes (votes and seats)
- Mean absolute deviations to quantify prediction accuracy

3. Results

3.1. 2019 Indian Lok Sabha Election Analysis

Table 1 presents the sentiment analysis results for the 2019

Lok Sabha election, comparing the Bharatiya Janata Party (BJP) and the Indian National Congress.

Table 1. Sentiment analysis of 2019 Lok Sabha election tweets

Party	Positive Tweets	Negative Tweets	Total Analysed
BJP	7,029	22,970	30,000
Congress	2,392	27,607	30,000

The positive tweets ratio between the BJP and Congress was 2.94:1, indicating that the BJP received nearly three times more positive tweets than Congress. Conversely, the negative tweets ratio was 0.83:1, showing that the BJP received fewer negative tweets compared to Congress. When compared to actual election results:

- Seats Won Ratio: 5.8:1 (BJP: 303 seats, Congress: 52 seats)
- Votes Won Ratio: 1.9:1 (BJP: 229,076,879 votes, Congress: 119,495,214 votes)

The positive tweet ratio (2.94:1) more closely aligned with the votes won ratio (1.9:1) than the negative tweet ratio (0.83:1) did with any electoral outcome metric.

3.2. 2020 U.S. Presidential Election Analysis

Table 2 shows the sentiment analysis results for the 2020 U.S. Presidential Election between the Democratic and Republican parties.

Table 2. Sentiment analysis of 2020 U.S. Presidential election tweets

Party	Positive Tweets	Negative Tweets	Total Analysed
Republican	64,788	285,172	350,000
Democrat	100,193	249,782	350,000

The positive tweets ratio was 1.55:1 in favor of Democrats, while the negative tweets ratio was 0.88:1.

Actual election results showed:

- Electoral Votes Won Ratio: 1.32:1 (Democrat: 306, Republican: 232)
- Popular Votes Won Ratio: 1.1:1 (Democrat: 81,283,501, Republican: 74,223,975)

Again, the positive tweet ratio (1.55:1) demonstrated stronger alignment with electoral outcomes than the negative tweet ratio.

3.3. Correlation Analysis

Table 3 presents a comprehensive comparison of tweet sentiment ratios and electoral outcomes across both elections.

Table 3. Correlation between tweet sentiment ratios and electoral outcomes

Election	Metric	Winning Party Ratio	Positive Tweet Ratio	Negative Tweet Ratio	Deviation from Positive	Deviation from Negative
2019 Indian Lok Sabha	Votes Won	1.9:1 (BJP: Congress)	2.94:1 (BJP: Congress)	0.83:1 (BJP: Congress)	+0.54	-2.07
	Seats Won	5.8:1 (BJP: Congress)	2.94:1 (BJP: Congress)	0.83:1 (BJP: Congress)	-2.86	-4.97
2020 U.S. Presidential	Votes Won	1.1:1 (Dem: Rep)	1.55:1 (Dem: Rep)	0.88:1 (Rep: Dem)	+0.45	-1.98
	Electoral	1.32:1 (Dem: Rep)	1.55:1 (Dem: Rep)	0.88:1 (Rep: Dem)	+0.23	-2.20

The correlation analysis reveals several key patterns:

- Positive Sentiment Alignment: Across both elections, positive tweet ratios consistently aligned with electoral success. The average deviation between positive tweet ratios and actual vote ratios was only 0.50, indicating strong predictive value. In both cases, the party with the higher positive tweet ratios won the election.
- Negative Sentiment Divergence: Negative tweet ratios showed substantial deviation from electoral outcomes, with an average deviation of 2.81 from actual results. The inverse relationship expected between negative tweets and electoral success was not observed consistently.
- Cross-Cultural Validity: The pattern held true across both Indian and American electoral contexts. In the Indian election, the positive tweet ratio (2.94:1) was closer to the actual votes ratio (1.9:1) than the negative tweet ratio (0.83:1). Similarly, in the U.S. election, the positive tweet ratio (1.55:1) closely matched both the popular vote ratio (1.1:1) and electoral vote ratio (1.32:1).
- Statistical Significance: The mean absolute deviation for positive sentiment predictions was 0.77 compared to 3.05 for negative sentiment predictions, demonstrating that positive tweets are approximately 4 times more accurate in predicting electoral outcomes.

4. Discussion

4.1. Interpretation of Results

The findings demonstrate that positive tweets serve as a more reliable indicator of electoral success than negative tweets. This pattern was consistent across both the 2019 Indian Lok Sabha election and the 2020 U.S. Presidential election, despite significant cultural and political system differences. As shown in Table 3, positive tweet ratios had an average deviation of only 0.77 from actual electoral outcomes, while negative tweet ratios showed a much larger average deviation of 3.05. This four-fold difference in predictive accuracy strongly suggests that positive sentiment on Twitter is a more reliable electoral indicator.

Several factors may explain why positive sentiment better predicts electoral outcomes:

- Engagement Asymmetry: Supporters may be more likely to express positive sentiment on social media, while critics might choose other forms of political expression or remain silent online.
- Mobilization Effect: Positive tweets may reflect and reinforce voter enthusiasm, translating more directly into actual voting behavior.
- Echo Chamber Dynamics: Positive sentiment may create self-reinforcing networks that strengthen supporter commitment and attract undecided voters.

4.2. Implications for Political Analysis

These results suggest that political analysts and campaign strategists should prioritize monitoring positive sentiment metrics over negative ones when gauging electoral momentum on social media. However, this should not lead to dismissing negative sentiment entirely, as it may still provide valuable insights into specific policy concerns or campaign vulnerabilities.

4.3. Limitations

Several limitations should be acknowledged:

- Platform Bias: Twitter users may not represent the general voting population, potentially skewing sentiment analysis results.
- Temporal Constraints: The study analyzed tweets from only two weeks before elections, potentially missing longer-term sentiment trends.
- Language Limitations: Analysis was restricted to English-language tweets, excluding significant portions of the Indian electorate who communicate in regional languages.
- Bot Activity: The study did not account for potential bot activity that could artificially inflate sentiment metrics.

4.4. Future Research Directions

Future studies should:

- Extend the analysis to multiple social media platforms
- Incorporate multilingual sentiment analysis capabilities
- Develop methods to detect and filter bot-generated content
- Examine sentiment dynamics over longer time periods
- Investigate the causal mechanisms linking online sentiment to voting behaviour

5. Conclusion

This research provides empirical evidence that positive Twitter sentiment serves as a stronger predictor of electoral outcomes compared to negative sentiment. Across two major democratic elections in different cultural contexts, parties that received higher ratios of positive tweets consistently achieved better electoral results. The analysis revealed that positive tweet ratios deviated by an average of only 0.77 from actual electoral outcomes, while negative tweet ratios showed a mean deviation of 3.05, making positive sentiment approximately four times more accurate as a predictive indicator. These findings contribute to the growing body of literature on social media's role in political discourse and electoral prediction. While Twitter sentiment analysis cannot replace traditional polling methods, it offers a complementary tool for understanding public opinion dynamics in real-time. The consistency of results across Indian and American elections suggests that the relationship between positive online sentiment and electoral success may reflect fundamental aspects of how political support manifests in digital spaces.

As social media continues to shape political communication, understanding these patterns becomes increasingly crucial for researchers, analysts, and democratic participants. This study demonstrates that in the digital age, expressions of support may speak louder than expressions of opposition when it comes to predicting electoral outcomes.

Acknowledgments

The author would like to thank Kaggle for making the tweet datasets publicly available.

References

[1] Adam Bermingham, and Alan F. Smeaton, "On Using Twitter to Monitor Political Sentiment and Predict Election Results," *Workshop at the International Joint Conference for Natural Language Processing*, Chiang Mai, Thailand, pp. 1-9, 2011. [Google Scholar] [Publisher Link]

[2] Murphy Choy et al., "A Sentiment Analysis of Singapore Presidential Election 2011 using Twitter Data with Census Correction," *arXiv preprint arXiv:1108.5520*, pp. 1-11, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[3] Suryansh Bhatnagar, Effectiveness of Sentiment Analysis for Predicting Election Outcomes, ResearchGate, pp. 1-8, 2023. [Online]. Available: https://www.researchgate.net/publication/376230037_Effectiveness_of_Sentiment_Analysis_for_predicting_Election_Outcomes

[4] Manuel Garcia-Herranz et al., "Using Friends as Sensors to Detect Global-Scale Contagious Outbreaks," *PLoSONE*, vol. 9, no. 4, pp. 1-7, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kokil Jaidka et al., "Predicting Elections from Social Media: A Three-Country, Three-Method Comparative Study," *Asian Journal of Communication*, vol. 29, no. 3, pp. 252-273, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yu Wang, Yuncheng Li, and Jiebo Luo, "Deciphering the 2016 U.S. Presidential Campaign in the Twitter Sphere: A Comparison of the Trumpists and Clintonists," *Journal of Big Data*, vol. 5, no. 34, pp. 723-726, 2018. [CrossRef] [Google Scholar] [Publisher Link]