

Twitter Sentiment Analysis Based on Political Tweets

Jyoti Nagnath Gavhane

CSE Department MIT Art, Design and Technology Pune, Maharashtra, India
jyoti.gavhane@mituniversity.edu.in

Ranjana Gore

CSE Department MIT Art, Design and Technology Pune, Maharashtra, India
goreranjana123@gmail.com

Shubhra Mukherjee Mathur

CSE Department MIT Art, Design and Technology Pune, Maharashtra, India
disha.gabhane@mituniversity.edu.in

Abstract

We are playing a major role in data generation. There is a tremendous amount of data available on the web. Social media is evolving as a platform for online education, entertainment, the exchange of ideas, and the sharing of opinions. Social networking services such as WhatsApp, Twitter, LinkedIn, Facebook, Instagram, and Google+ are experiencing rapid growth due to their capacity for users to express and share their viewpoints. These services facilitate many subjects to converse with diverse cultures and disseminate information internationally. Twitter is one of the popular platforms among all age groups to put their opinions spontaneously. This study principally focuses on sentiment analysis of twitter data, which is useful for analyzing information in tweets when opinions are very unstructured, varied, and occasionally neutral. There are numerous algorithms existing to analyze the data. The study is based on XGBoost algorithm which has been proven to be the most accurate and fast algorithm until now.

Keywords—Sentiment Analysis, XG-Boost, Politics

Introduction

Sentiment analysis is a natural language processing (NLP) technique that aims to automatically extract attitudes, opinions, perspectives, and emotions from various textual sources, including text, speech, social media posts such as tweets, and data stored in databases. This process involves categorizing textual opinions into labels such as "positive," "negative," and "neutral," and it goes by various names, including subjectivity analysis, opinion mining, and sentiment assessment extraction. While the terms "opinion,"

"sentiment," "view," and "belief" are sometimes used interchangeably, it's important to note that each of these terms carries a distinct meaning. To learn opinions about the political leaders' it is needed to engage the public through TV programs, YouTube, etc. Many reporters have interviewed them. Regarding the political discussions that were included on the TV shows, people converse with each other about their viewpoints. The effort to use surveys and polls to research people's opinions is highly time-consuming and expensive. One of the important platforms where individuals may engage with various communities and clearly express their thoughts is Twitter. The primary objective of the study is to foresee the outcome of the election using public opinion.

India's political climate is most uncertain since it is the most populous nation and the largest populated democracy in the world. The opposition would have different perspectives on every move made by the ruling party. However, today's general public share their thoughts about every government decision on social media (such as the demonetization of all Indian 500 and 1000 rupee notes from the Mahatma Gandhi Series). Therefore, our goal is to analyze how online users feel about each political party, their leaders, and their actions based on social media tweets.

II. Literature Review

A. Analyzing Political Sentiment using Twitter Data In India, just 12.18% of the population communicates in English, while most of the ordinary people express their ideas in their local tongue. Therefore, it is not always worthwhile to analyze sentiment using solely English. The implementation of the vocabulary of many native Indian languages (such as Hindi, Gujarati, etc.) is the main goal covered in this paper.

B. Analyzing sentiment and examining the extent during U.S.

presidential election of year 2016. In this research, the study focused on investigating political homophily on Twitter during the 2016 U.S. presidential election by employing sentiment analysis techniques. Over 122 days, they gathered user profiles, contact networks, and tweets. They classified Twitter users into six categories using sentiment analysis: It categorized the Twitter users into six groups: "Neutral," "Negative," "Positive," "Trump Supporter," "Hillary Supporter," and "Whatever." Afterward, conducted an analysis to explore the extent of political like-mindedness or homophily within these defined classes. To accomplish this task, two types of networks were defined: uniplex and multiplex. A multiplex network is one having more types of connections between nodes whereas an uniplex network has one form of connection between nodes. They listed the following actions as examples of links between users on Twitter: following, mentioning, and retweeting. Both unidirectional and reciprocal links are possible. To determine the homophily in the identified groupings, they employed the measure IHi. Finally, they examined homophily in three contexts: (i) friendship with a similar speech, (ii) multiplex connections, and (iii) uniplex relationships.

C. A real-time investigation of political sentiment polarity

Two analytic models based on unique events and temporal periods are provided in this research, which focuses on dynamic political sentiment polarity analysis. BERT pre-training model used as a fine-tune a classifier for analyzing tweets related to American politicians. The tweets were fed into the classifier to see how well it could understand and classify them. The results turned out to be quite promising and effective. After evaluating it has been observed how well the traditional Convolutional Neural Network (CNN) model and the Recurrent Neural Network (RNN) model performed on the dataset in terms of classification. The dynamic political sentiment polarity of individual tweets was analyzed and exhibited using multi-classifiers in conjunction with analytical models, offering a fresh approach to the gathering, acquisition, and analysis of open source information.

D. Analyzing the Impact of Sentiment on Political Communication

This paper examined the relationship between emotion and virality regarding political tweets. They succeeded in finding an effective sentiment classifier after a thorough search and now have a reliable multilingual model that can reliably detect sentiment in

tweets from politicians. This was accomplished by using cutting-edge transformer-based language models that were also tailored to the job at hand. According to their data, there is a significant correlation between the popularity of a politician's tweet and its emotion, with negative tweets having a higher network penetration than good tweets. All three of the studied sovereign nations—Greece, Spain, and the United Kingdom—seem to exhibit the same tendency. The control trials carried out are used to confirm their conclusions even further. A historical examination of these control trials indicates that the tendency of negative tweets having greater influence is ongoing and intensifying with time.

E. Sentiment Analysis of Political Tweets From the 2019 Spanish Elections

In our research, our objective was to examine how sentiment measurements change over time, particularly in the context of highly charged political elections. While many recent studies have emphasized the application of machine learning techniques, our approach was to extract knowledge using specific indices rooted in information theory, as well as intrinsic features derived from nonlinear transformations to create lower-dimensional representations. This strategy seeks to provide benefits for interpretability and the development of temporal dynamics.

F. Research Questions

Twitter sentiment analysis has made significant advancements in recent years, there are still some research gaps that exist. Here are a few areas where further research could be conducted:

1. Contextual Understanding: Twitter messages often contain sarcasm, irony, and other forms of contextual nuances that can affect sentiment analysis. Improving algorithms to better understand the context of tweets would enhance the accuracy of sentiment analysis.
2. Handling Linguistic Variations: Twitter users employ various forms of linguistic variations such as abbreviations, slang, and misspellings. These variations can pose challenges for sentiment analysis models. Exploring techniques to handle these linguistic variations and developing robust models that can handle diverse language usage on Twitter is an area that requires further investigation.
3. Domain Adaptation: Sentiment analysis models trained on general datasets might not perform as well when applied to specific domains. Adapting sentiment analysis models to different domains, such as politics, finance, or healthcare, is crucial to improve accuracy and applicability in specialized contexts.

4. Handling Multilingual Data: Twitter is a global platform with users from diverse linguistic backgrounds. Expanding sentiment analysis research to encompass multilingual data and developing models capable of analyzing sentiments in multiple languages would be valuable.

5. Bias and Ethics: Sentiment analysis models can be susceptible to biases embedded in their training data, potentially resulting in unfair outcomes or the reinforcement of existing biases. Further research is needed to develop techniques that alleviate bias and confirm ethical use of sentiment analysis in sensitive contexts.

6. Long-Term Sentiment Analysis: Twitter sentiment analysis has primarily focused on short-term analysis of individual tweets. However, understanding sentiment trends over longer periods, such as weeks, months, or years, can provide valuable insights. Investigating methods to capture and analyze sentiment patterns over extended periods would contribute to a more comprehensive understanding of user sentiment on Twitter.

7. User-Level Sentiment Analysis: Most sentiment analysis models analyze sentiment at the tweet level, but Twitter users often have varying sentiments expressed in different tweets. Developing techniques to aggregate sentiments at the user level and understanding sentiment changes over time for individual users could offer deeper insights into user behavior and preferences.

8. Real-Time Analysis: Twitter generates an enormous volume of data in real-time. Conducting sentiment analysis on such data in real-time poses challenges in terms of computational efficiency, scalability, and maintaining high accuracy. Exploring techniques for efficient real-time sentiment analysis on Twitter streams is an important research direction.

By addressing these research gaps, the field of Twitter sentiment analysis can advance and provide more accurate, nuanced, and context-aware insights from the vast amounts of user-generated content on the platform.

We have decided to focus on user-level sentiment analysis and real-time analysis in terms of computational efficiency, scalability, and maintaining high accuracy.

G. Objectives of Sentiment Analysis:

The application should analyze Indian political tweets to determine sentiment.

Sentiment analysis should cover various aspects such as positive, negative, and neutral sentiments.

The sentiment analysis algorithm should be accurate and reliable.

H. Technologies Used

Python is a high-level, interpreted, and versatile programming language that prioritizes code manageability and readability, making it a prime choice for Machine Learning applications.

Django, one of the most popular Python web frameworks, empowers the rapid development of secure and maintainable websites. It stands out for its ease of maintenance, debugging capabilities, and its comprehensive documentation.

Tweepy, an open-source Python package, provides a convenient means to access the Twitter API in Python. Tweepy's classes and methods represent Twitter's models and API endpoints, simplifying tasks like data encoding and decoding.

Apache SQL runs on XAMPP server for Windows and employs phpMyAdmin within the localhost. This database serves as a repository for storing, verifying, and retrieving user login credentials.

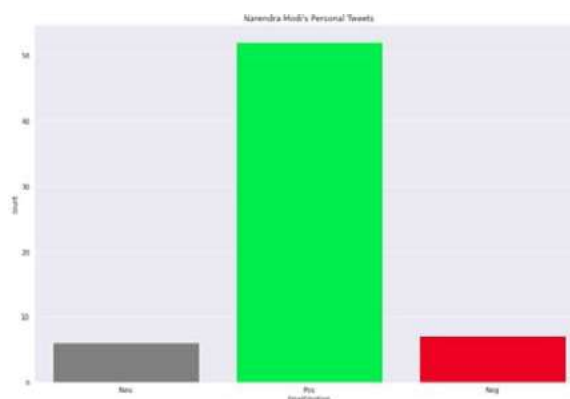
Bootstrap enhances the user interface, enabling the creation of modern, intuitive, and responsive web pages.

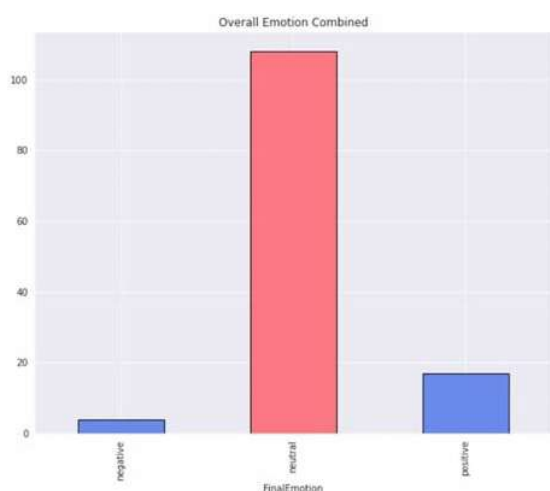
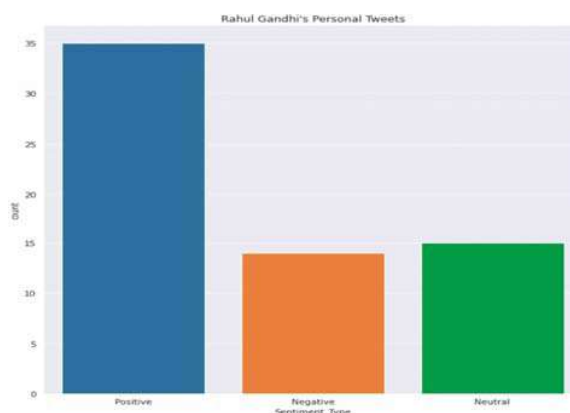
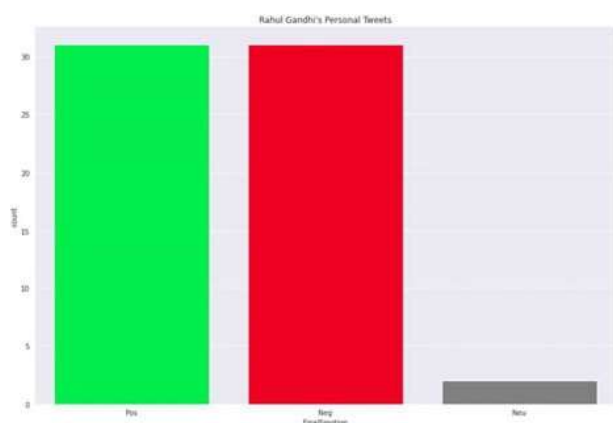
jQuery plays a pivotal role in sending and retrieving data to and from the Tweepy API and displaying the results on the webpage.

HTML and CSS form the foundational elements for designing the website's front-end.

III. Results And Analysis

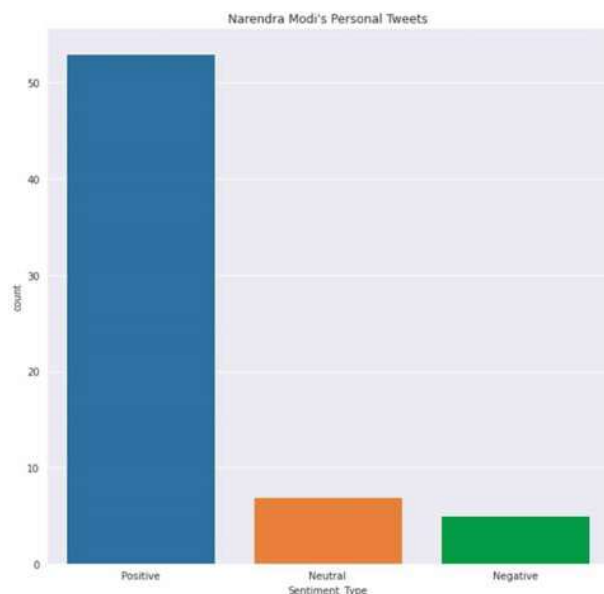
Using XGBoost, we analyzed the following sentiments. As the dataset did not have already assigned sentiments, we cannot exactly compare the accuracy. Efforts have centered on the following task: To evaluate a given message and categorize it as either having a positive, negative, or neutral sentiment. In instances where a message conveys both positive and negative sentiments, the classifier should determine the predominant sentiment.





We obtained the following results from TextBlob:

Comparative Study of Tweets sentiments of tweets from Narendra Modi & Rahul Gandhi:



Future Scope

By harnessing the power of sentiment analysis algorithms, the application provides valuable insights into the positive, negative, and neutral sentiments expressed in Indian political tweets. The integration of a Map API allows users to visualize sentiment analysis results on a geographical map, providing a unique perspective on sentiment variations across different regions of India. This visual representation enhances the understanding of political sentiment and enables users to identify trends and patterns in public opinion.

The inclusion of a Twitter chat box fosters real-time discussions among users, creating a dynamic environment for political discourse and engagement. Users can actively participate, share their views, and engage with other like-minded individuals, contributing to a vibrant and interactive community focused on Indian politics.

Data privacy and security are prioritized in the application, ensuring that user information is protected and adhering to regulatory standards. User authentication mechanisms and personalized user profiles enhance the security and tailored experience for each user.

IV. Conclusion

Both candidates have relatively the same no. of tweets so we can assume somewhat of the same engagement from both sides. Though if we see the negative polarity score is more for Rahul Gandhi's tweets than Narendra Modi's tweets, it may be beneficial for both the sides as Rahul Gandhi is the leader of opposition and Modi was present PM at that time, but positive approach of Modi's tweets may be slightly more than Gandhi's negative tweets.

Overall, the Twitter Sentiment Analysis on Indian Political Tweets web application, with its combination of sentiment analysis, map visualization, and real-time discussions, empowers users to gain a deeper understanding of public sentiment and facilitates informed discussions on Indian politics. It serves as a valuable tool for individuals, researchers, and policymakers to stay informed about the evolving political landscape and make data-driven decisions.

Overall, the Twitter Sentiment Analysis on Indian Political Tweets web application, with its combination of sentiment analysis, map visualization, and real-time discussions, empowers users to gain a deeper understanding of public sentiment and facilitates informed discussions on Indian politics. It serves as a valuable tool for individuals, researchers, and policymakers to stay informed about the evolving political landscape and make data-driven decisions.

References

- [1] Sharma, N., Pabreja, R., Yaqub, U., Atluri, V., Chun, S. A., & Vaidya, J. (2018, May). Web-based application for sentiment analysis of live tweets. In Proceedings of the 19th Annual International Conference on Digital government research: Governance in the data Age (pp. 1-2).
- [2] Nahar, L., Sultana, Z., Iqbal, N., & Chowdhury, A. (2019, May). Sentiment analysis and emotion extraction: A review of research paradigm. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (pp. 1-8). IEEE.
- [3] Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017, April). Deep learning for hate speech detection in tweets. In Proceedings of the 26th international conference on World Wide Web companion (pp. 759-760).
- [4] Nithin P, Owais Ahmed, Aditya Jason Hans, Amaan Faraaz, Prof. Madhusudhan M.V. (2021). Depression Detection Model Based on Sentiment Analysis on Twitter API, International Journal of New Technology and Research (IJNTR), Volume-7, Issue-5, pp. 16-20.
- [5] Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M., & Lee, S. (2017, October). Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. In 2017 international conference on information and communication technology convergence (ICTC) (pp. 138-140). IEEE.
- [6] Roy, S. (2018). Analyzing Political Sentiment Using Twitter Data.
- [7] Caetano, J. A., Lima, H. S., Santos, M. F., & Marques-Neto, H. T. (2018). Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election. *Journal of internet services and applications*, 9(1), 1-15.
- [8] Chang, C., & Wang, X. (2020, November). Research on Dynamic Political Sentiment Polarity Analysis of Specific Group Twitter Based on Deep Learning Method. In *Journal of Physics: Conference Series* (Vol. 1651, No. 1, p. 012108). IOP Publishing.
- [9] Antypas, D., Preece, A., & Camacho-Collados, J. (2023). Negativity spreads faster: A large-scale multilingual twitter analysis on the role of sentiment in political communication. *Online Social Networks and Media*, 33, 100242.
- [10] Rodríguez-Ibáñez, M., Gimeno-Blanes, F. J., Cuenca-Jiménez, P. M., Soguero-Ruiz, C., & Rojo-Álvarez, J. L. (2021). Sentiment analysis of political tweets from the 2019 Spanish elections. *IEEE Access*, 9, 101847-101862.
- [11] 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.

