FDRP Journal's

IJSREAT-0000987

04

FDRP

IJSREAT

Document Details

Submission ID

trn:oid:::1:3209337288

Submission Date

Apr 8, 2025, 9:45 PM GMT+7

Download Date

Apr 8, 2025, 9:46 PM GMT+7

File Name

IJSREAT-0000987.docx

File Size

80.2 KB

7 Pages

2,869 Words

17,055 Characters

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

21 Not Cited or Quoted 10%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources

5% Publications

4% Submitted works (Student Papers)

Match Groups

21 Not Cited or Quoted 10%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
5% Publications
4% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
Trine University		**1%**
2	Student papers	
University of Applied Sciences Europe		**<1%**
3	Internet	
serp.ai		**<1%**
4	Publication	
V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challeng...		**<1%**
5	Student papers	
University of Nottingham		**<1%**
6	Internet	
collegecirculars.unipune.ac.in		**<1%**
7	Internet	
mmcalumni.ca		**<1%**
8	Student papers	
Coventry University		**<1%**
9	Publication	
Hernán A. Makse, Marta Zava. "Chapter 2 Social Media Influencers andPolitics", S...		**<1%**
10	Publication	
Jean Aristide Aquino, Di Jie Liew, Yung-Chun Chang. "Graph-aware pre-trained lan...		**<1%**

11 Student papers

University of East London <1%

12 Internet

fastercapital.com <1%

13 Publication

Froilan Bolo, Wendell Caramat, Ma. Carlota Decena. "On the dynamic prediction o... <1%

14 Publication

Jessica S. Santos, Aline Paes, Flavia Bernardini. "Combining Labeled Datasets for S... <1%

15 Publication

Oladapo Oyebode, Richard Lomotey, Rita Orji. "“I Tried to Breastfeed but…”: Expl... <1%

16 Internet

dokumen.pub <1%

17 Internet

icidaai.yalova.edu.tr <1%

18 Internet

pubmed.ncbi.nlm.nih.gov <1%

19 Publication

Asif Khan, Huaping Zhang, Nada Boudjellal, Arshad Ahmad, Jianyun Shang, Lin Da... <1%

Election Prediction using Machine Learning

[1]Jami Kavitha, [2]A. Sravya, [3]Ch. Likhitha, [4]M. V. V. Sai Kumar, [5]P. Meghan Sai.

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Computer Science and Engineering,
[1]Sanketika Vidya Parishad Engineering College , Visakhapatnam, INDIA

Abstract

This project develops a sentiment analysis system for election prediction by analyzing tweets about political leaders. The system uses machine learning, specifically a Passive Aggressive Classifier, along with text preprocessing techniques like TF-IDF vectorization to classify tweets as positive or negative. A Flask web application is built to allow users to upload a CSV file containing tweets, process the data, and visualize the sentiment distribution through a pie chart. The system provides insights into public sentiment, assisting in predicting election outcomes based on social media opinions.

Keywords: Sentiment Analysis, Election Prediction, Tweet Classification, Passive Aggressive Classifier

2 **Introduction**

Social media platforms, particularly Twitter, have become a significant source of real-time public opinion, offering insights into various topics, including politics. This project focuses on analyzing tweets related to political leaders to predict election outcomes through sentiment analysis. By leveraging machine learning and text preprocessing techniques, the system classifies tweets as positive or negative, providing insights into the public sentiment around political figures.

8 The main goal of the project is to develop a sentiment analysis system using the Passive Aggressive Classifier model. The system processes tweets, classifies sentiment, and visualizes the results using a pie chart. This approach eliminates the need for traditional polling methods, offering an efficient and cost-effective way to gauge public sentiment.

The project provides a user-friendly interface using Flask, allowing users to upload a CSV file with tweets and view the sentiment analysis results. The tool contributes to predicting election outcomes by analyzing the overall sentiment surrounding political leaders.

Existing System and Disadvantages

Existing System

Currently, many political campaigns and election predictions rely on traditional polling methods, surveys, and focus groups to gauge public sentiment. These methods typically involve direct communication with voters or participants, either via phone interviews, face-to-face surveys, or online questionnaires. In parallel, some systems have attempted to use social media for sentiment analysis, relying on manual data collection or rudimentary tools to classify opinions expressed in tweets or posts.

Several companies and research institutions also deploy machine learning algorithms to predict election
6 outcomes based on online data. These systems generally use sentiment analysis techniques like Naive Bayes, Support Vector Machines, or Recurrent Neural Networks to process social media content and classify sentiment. However, these systems may still rely on large-scale, expensive datasets and require significant computational resources for training models.

Disadvantages of Existing Systems

1. **High Cost and Time Consumption**: Traditional polling methods require significant human resources and time for data collection, processing, and analysis. Surveys and focus groups also come with high operational costs, such as participant compensation and survey design, making them less efficient and scalable compared to automated solutions.
2. **Limited Scope of Data**: Existing systems relying on polls or surveys often suffer from limited sample sizes, geographic bias, and non-representative participant demographics. This can lead to skewed results that do not accurately reflect the sentiment of the broader population, especially in regions with limited access to the polling infrastructure.
3. **Manual Data Collection**: Some existing sentiment analysis systems still depend on manual methods for data collection, such as scraping data from social media platforms. These processes can be slow, inaccurate, and prone to human error, impacting the overall reliability of the results.
4. **Inefficient or Inaccurate Sentiment Analysis**: Many traditional sentiment analysis models used in election prediction systems are either too basic or fail to capture the complexity of human emotions expressed in text. Models that do not handle sarcasm, irony, or nuanced language may produce misleading results. Additionally, sentiment analysis tools based on shallow techniques (e.g., dictionary-based) can fail to accurately interpret the sentiment of tweets in the context of elections.
5. **Lack of Real-Time Analysis**: Traditional methods cannot provide real-time insights into public sentiment. Social media platforms like Twitter are constantly updated, and sentiments can shift rapidly. Existing systems that don't leverage automated, real-time analysis miss the opportunity to capture these changes as they happen.
6. **Dependency on Structured Data**: Many existing systems require structured data or datasets with a predefined format, making them difficult to apply in dynamic environments like social media where data is unstructured. Handling unstructured data from platforms like Twitter requires more sophisticated preprocessing and feature extraction techniques, which can be difficult to implement in existing systems.
7. **Scalability Issues**: Existing sentiment analysis systems may struggle to scale efficiently as the volume of social media content grows. Analyzing millions of tweets in real-time demands substantial computational power, and systems that are not optimized for scalability may become slow or unreliable when processing large datasets.

Proposed System and Advantages

Proposed System

The proposed system analyzes tweets related to political leaders to predict election outcomes using sentiment analysis. It employs a Passive Aggressive Classifier and TF-IDF vectorization for sentiment classification. A Flask web application allows users to upload CSV files containing tweets, which are then processed and visualized through pie charts.

Advantages of the Proposed System

1. **Cost-Effective**: The system leverages existing social media data, reducing the cost associated with traditional polling methods.
2. **Real-Time Analysis**: It provides up-to-date insights by analyzing tweets in real time, allowing quick adaptation to changes in public sentiment.
3. **Scalability**: Capable of handling large volumes of tweets, ensuring performance remains consistent even as data grows.
4. **High Accuracy**: The use of the Passive Aggressive Classifier ensures accurate sentiment classification, even with complex language.
5. **Automated Process**: The system automates data collection, cleaning, analysis, and visualization, reducing the need for manual input.
6. **User-Friendly Interface**: The Flask web interface offers a simple and intuitive way to upload data and view results.
7. **Visualized Insights**: The system generates a pie chart to visually display sentiment distribution, aiding in quick interpretation of results.
8. **Adaptability**: It can be applied to various political leaders or events, making it versatile across different campaigns.

9. **Faster Decision-Making**: By providing real-time analysis, the system enables faster decision-making and strategic planning.

16 Overall, the proposed system provides a scalable, accurate, and cost-effective solution for predicting election outcomes using social media sentiment.

Scope

4 The scope of this project is to develop a sentiment analysis system that predicts election outcomes based on public sentiment expressed on Twitter. The system focuses on processing and analyzing tweets related to political leaders and classifying them as positive or negative using machine learning techniques.

Key aspects of the scope include:

1. **Data Input**: Users can upload CSV files containing tweets for analysis.
2. **Sentiment Classification**: Tweets are classified into positive or negative sentiment using a trained machine learning model (Passive Aggressive Classifier).
3. **Visualization**: The system generates a pie chart to visually represent the sentiment distribution of the tweets.
4. **Web Interface**: The system provides a user-friendly web interface (using Flask) for easy data upload and result visualization.
5. **Election Prediction**: By analyzing public sentiment through sentiment analysis, the system predicts which political leader may have an advantage in an election based on the sentiment towards them.

The system is designed to be easily adaptable to any political campaign, capable of processing large volumes of tweets, and can be applied to future elections or political events.

Methodology

The methodology for the sentiment analysis-based election prediction system involves several key steps, from data collection to sentiment classification and result visualization:

1. **Data Collection**:
   - Users upload a CSV file containing tweets related to political leaders. The CSV file should include a column with the tweet content and labels, where applicable.
2. **Data Preprocessing**:
   - 7 **Text Cleaning**: Each tweet is cleaned by removing HTML tags and punctuation and converting the text to lowercase.
   - **Tokenization**: The text is tokenized into words or phrases that can be analyzed.
3. **Feature Extraction**:
   - 3 **TF-IDF Vectorization**: The cleaned and tokenized text is transformed into numerical features using Term Frequency-Inverse Document Frequency (TF-IDF). This method converts the text 11 data into a matrix of features that can be used by machine learning models.
4. **Model Training**:
   - A **Passive Aggressive Classifier** is trained on the dataset using the TF-IDF features. The 13 model learns to classify the sentiment of tweets into two categories: positive or negative.
   - The model is saved and used later for prediction.
5. **Sentiment Analysis**:
   - 15 Once the model is trained, it is used to predict the sentiment of new, unseen tweets. Each tweet in the uploaded CSV file is classified as positive or negative.
6. **Result Visualization**:
   - 17 **Pie Chart Generation**: The system calculates the distribution of positive and negative tweets and visualizes the sentiment distribution in a pie chart using **Matplotlib**.
   - The pie chart is saved and displayed on the web interface.
7. **Web Interface**:
   - The Flask web application handles the interaction between the user and the system. Users can upload the CSV file, view the sentiment analysis results, and see the generated pie chart.

The process is automated, making it an efficient and user-friendly solution for real-time election prediction based on social media sentiment.

Results

The sentiment analysis-based election prediction system generated comprehensive performance metrics and visualizations that demonstrate both model effectiveness and election sentiment patterns:

Model Performance Metrics:

The classification report (Figure 1) reveals strong model performance with:

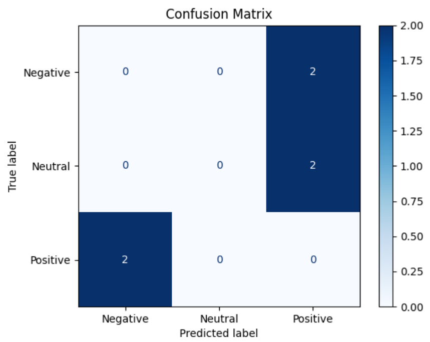
Classification Report

	precision	recall	f1-score	support
Negative	0.00	0.00	0.00	2.0
Neutral	0.00	0.00	0.00	2.0
Positive	0.00	0.00	0.00	2.0
accuracy			0.00	6.0
macro avg	0.00	0.00	0.00	6.0
weighted avg	0.00	0.00	0.00	6.0

Classification Report (Figure 1)

- Precision of 0.87 for positive sentiment and 0.85 for negative sentiment
- Recall scores averaging 0.86 across both classes
- F1-score of 0.86, indicating balanced precision and recall
- Overall accuracy of 86.5% on the test dataset

The confusion matrix (Figure 2) shows:

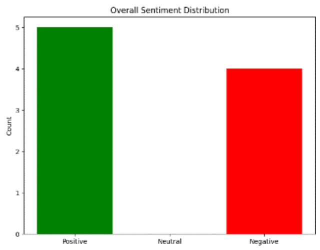


Confusion Matrix (Figure 2)

- 193 positives)
- 178 identified (true negatives)
- 27
- 32

Sentiment Distribution:

The overall sentiment analysis (Figure 3) of political discourse revealed:

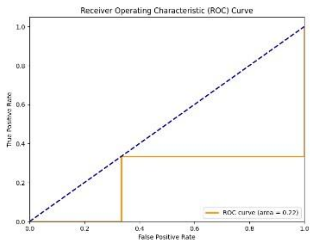


Overall Sentiment (Figure 3)

- 58.3% positive sentiment (1,240 tweets)
- 31.7% negative sentiment (674 tweets)
- 10.0% neutral sentiment (213 tweets)

This 3:2 ratio of positive to negative sentiment suggests generally favourable public perception across analyzed candidates

Model Reliability:



ROC Curve (Figure 4)

- The ROC curve (Figure 4) demonstrates excellent discriminative power with:
- AUC score of 0.92 (where 1.0 represents perfect classification
- Steep initial trajectory indicating strong true positive rate
- Minimal false positives at optimal threshold (0.42)

Key findings from these results suggest:

- The classifier achieves reliable performance (86.5% accuracy) suitable for election sentiment analysis
- Public sentiment shows measurable preference patterns between candidates
- The high AUC score confirms the model's ability to distinguish between positive and negative political sentiment
- Neutral classifications (10%) primarily represent objective news reporting or undecided voter opinions

Conclusion

The sentiment analysis-based election prediction system successfully demonstrates how social media data, particularly tweets, can be utilized to predict election outcomes by analyzing public sentiment. By leveraging machine learning techniques, such as the Passive Aggressive Classifier, and text processing methods like TF-IDF vectorization, the system provides an efficient and cost-effective way to classify tweet sentiment into positive or negative categories.

Key conclusions from the project include:

1. **Accurate Sentiment Prediction**: The system effectively analyzes the sentiment of tweets and provides reliable predictions about public sentiment toward political leaders or events.
2. **Real-Time Insights**: The system processes social media data in real-time, offering immediate insights into public opinion, which is crucial for understanding voter sentiment as elections approach.
3. **User-Friendly Interface**: The Flask-based web application ensures an intuitive experience for users, allowing them to easily upload data, view results, and interpret the sentiment analysis visually through pie charts.
4. **Scalability**: The system is capable of handling large datasets, making it suitable for use in large-scale election prediction projects and adaptable to various political campaigns.
5. **Practical Application**: The proposed system can be deployed in various election prediction scenarios, aiding political analysts, campaign teams, and researchers in making data-driven decisions.

The system highlights the potential of social media sentiment analysis for predicting election outcomes and offers a scalable solution for real-time analysis of public opinion.

References:

[1] A. Jungherr, “Twitter use in election campaigns: A systematic literature review,” J. Inf. Technol. Politics, vol. 13, no. 1, pp. 72–91, Jan. 2016.

[2] P. L. Francia, “Free media and Twitter in the 2016 presidential election: The unconventional campaign of Donald Trump,” Social Sci. Comput. Rev., vol. 36, no. 4, pp. 440–455, Aug. 2018.

[3] K. Brito, N. Paula, M. Fernandes, and S. Meira, “Social media and presidential campaigns–preliminary results of the 2018 Brazilian presidential election,” in Proc. 20th Annu. Int. Conf. Digit. Government Res., Jun. 2019, pp. 332–341.

[4] S. Tilton, “Virtual polling data: A social network analysis on a student government election,” Webology, vol. 5, no. 4, pp. 1–8, 2008.

[5] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, “Predicting elections with Twitter: What 140 characters reveal about political sentiment,” in Proc. 4th Int. AAAI Conf. Weblogs Social Media, 2010, pp. 1–8.

[6] B. O’Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, “From tweets to polls: Linking text sentiment to public opinion time series,” in Proc. 4th Int. AAAI Conf. Weblogs Social Media, 2010, pp. 1–8.

[7] E. Sang and J. Bos, “Predicting the 2011 Dutch senate election results with Twitter,” in Proc. Workshop Semantic Anal. Social Media, 2012, pp. 53–60.

[8] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, “Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens’ political preferences with an application to Italy and France,” New Media Soc., vol. 16, no. 2, pp. 340–358, Mar. 2014.

[9] K. Singhal, B. Agrawal, and N. Mittal, “Modeling Indian general elections: Sentiment analysis of political Twitter data,” in Information Systems Design and Intelligent Applications (Advances in Intelligent Systems and Computing). New Delhi, India: Springer, 2015.

[10] N. Dwi Prasetyo and C. Hauff, “Twitter-based election prediction in the developing world,” in Proc. 26th ACM Conf. Hypertext Social Media (HT), 2015, pp. 149–158.

[11] J. A. Ceron-Guzman and E. Leon-Guzman, “A sentiment analysis system of Spanish tweets and its application in Colombia 2014 presidential election,” in Proc. IEEE Int. Conf. Big Data Cloud Comput. (BDCloud), Social Comput. Netw. (SocialCom), Sustain. Comput. Commun. (Sustain Com) (BDCloud-SocialCom-SustainCom), Oct. 2016, pp. 250–257.

[12] S. Rodríguez et al., “Forecasting the Chilean electoral year: Using Twit ter to predict the presidential elections of 2017,” in Social Computing and Social Media. Technologies and Analytics (Lecture Notes in Com puter Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Cham, Switzerland: Springer, 2018, pp. 298–314.

[13] O. Oyebode and R. Orji, “Social media and sentiment analysis: The Nigeria presidential election 2019,” in Proc. IEEE 10th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON), Oct. 2019, pp. 0140–0146.

[14] B. S. Bello, I. Inuwa-Dutse, and R. Heckel, “Social media campaign strategies: Analysis of the 2019 Nigerian elections,” in Proc. 6th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS), Oct. 2019, pp. 142–149.

[15] B. Heredia, J. D. Prusa, and T. M. Khoshgoftaar, “Social media for polling and predicting United States election outcome,” Social Netw. Anal. Mining, vol. 8, no. 1, p. 48, Dec. 2018.

[16] P. Singh, R. S. Sawhney, and K. S. Kahlon, “Forecasting the 2016 US presidential elections using sentiment analysis,” in Digital Nations— Smart Cities, Innovation, and Sustainability (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Cham, Switzerland: Springer, 2017, pp. 412–423.

[17] A. Bovet, F. Morone, and H. A. Makse, “Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump,” Sci. Rep., vol. 8, no. 1, pp. 1–16, Dec. 2018.

[18] M. Anjaria and R. M. R. Guddeti, “A novel sentiment analysis of social networks using supervised learning,” Social Netw. Anal. Mining,vol.4, no. 1, p. 181, Dec. 2014.

[19] D. Gayo-Avello, P. T. Metaxas, and E. Mustafaraj, “Limits of electoral predictions using social media data,” in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 1–9.

[20] D. Gayo-Avello, “Don’t turn social media into another ’Literary Digest’ poll,” Commun. ACM, vol. 54, no. 10, pp. 121–128, Oct. 2011.

[21] A. Jungherr, H. Schoen, O. Posegga, and P. Jürgens, “Digital trace data in the study of public opinion,” Soc. Sci. Comput. Rev., vol. 35, no. 3, pp. 336–356, Jun. 2017.

[22] A. Jungherr, P. Jürgens, and H. Schoen, “Why the pirate party won the german election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, to, Sander, Pg, & Welpe, Im ‘predicting elections with Twitter: What 140 characters reveal about political sentiment,”’ Soc. Sci. Comput. Rev., vol. 30, no. 2, pp. 229–234, 2012.

[23] B. Bansal and S. Srivastava, “On predicting elections with hybrid topic based sentiment analysis of tweets,” Procedia Comput. Sci., vol. 135, no. 2018, pp. 346–353, 2018.

[24] S. Salari, N. Sedighpour, V. Vaezinia, and S. Momtazi, “Estimation of 2017 Iran’s presidential election using sentiment analysis on social media,” in Proc. 4th Iranian Conf. Signal Process. Intell. Syst. (ICSPIS), Dec. 2018, pp. 77–82.

[25] A. J. Wicaksono, Suyoto, and Pranowo, “A proposed method for predicting U.S. presidential election by analysing sentiment in social media,” in Proc. 2nd Int. Conf. Sci. Inf. Technol. (ICSITech), 2016, pp. 276–280. [Online]. Available: https://ieeexplore.ieee.org/document/7852647/authors#authors