iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: V Month of publication: May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62060

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Deep learning based Election Results Prediction using Twitter Data

Sheetal J[1], Mahesh Thotar[2], Veerendra Kumar K[3], Yuva sai M[4], S Uttej Kumar M[5]
[1]*Associate Professor,* [2, 3, 4, 5]*Final Year Students, Department of Computer science & Engineering, Ballari Institute of Technology and Management, Ballari – 583104, India.*

Abstract: This project presents an intelligent system political sentiment analysis with information from Twitter. . The system utilizes deep learning techniques, specifically a BiLSTM model featuring an attention mechanism, to predict the political affiliation (BJP or Congress) of tweets. The data preprocessing involves cleaning text, data augmentation, tokenization, and indexing. A dataset is used to train the model. consisting of tweets from both parties, getting very accurate results in predicting the sentiment of tweets. Additionally, the software offers functionality for data visualization, including displaying sample tweets, sentiment analysis results, and the winning party. Additionally, an admin page is available for further customization and management. Overall, this project the possibilities of Using deep learning to analyze sentiment of political content on social media data.
Keywords: Political sentiment analysis, Social media analysis, BiLSTM , Deep learning, Twitter data.

I. INTRODUCTION

Social media platforms like Twitter have become crucial sources of communication and data for political parties and their supporters. Understanding public sentiment on these platforms is essential for political parties to gauge their popularity, assess public opinion, and strategize their campaigns effectively. In this context, Utilizing Twitter data for political sentiment analysis has drawn a lot of interest in recent years.

This project focuses on analyzing the political sentiment expressed in representing the two largest political parties in India: the Bharatiya Janata Party (BJP) and the Indian National Congress (Congress). By collecting and analyzing tweets from these parties, we aim to learn more about how the public feels about each party and their policies.

The project includes several steps to achieve its objectives. First, we plot sample tweets from both parties to showcase the diversity of content and sentiment expressed on Twitter. Then, we generate a pie diagram to visually represent the distribution of tweets between the BJP and Congress. Next, we construct histograms to illustrate the sentiment distributions for each party, an enhanced comprehension of the sentiment conveyed through the tweets.

To further explore the content of the tweets, we create a word cloud highlighting the most frequent terms used in the tweets. This visualization helps identify key themes and topics of discussion surrounding each political party.

To predict the sentiment of the tweets, we employ a Bidirectional Long Short-Term Memory deep learning model (BiLSTM) with an attention mechanism. This model is trained to classify tweets as either positive or negative according to their content. The predictions from the model are then aggregated to determine the overall sentiment for each political party.

Additionally, we provide an accuracy plot in order to assess the efficiency of our sentiment analysis model. Finally, an admin page is included for further customization and management of the application.

Through this project, we aim give insightful information about public sentiment towards the BJP and Congress on Twitter, using cutting-edge methods for data analysis and deep learning models.

II. LITERATURE SURVEY

1) Social Media Analytics for Election Prediction: Previous studies have explored the usage of social mediadata, including Twitter, for predicting election outcomes. Researchers have utilized a variety of strategies, such as sentiment analysis, topic modeling, and network analysis to extract information, from social media discussions related to elections (e.g., Gayo-Avello, 2012; Tumasjan et al., 2010).
2) Deep Learning in Social Media Analysis: The application of deep learning Data analysis from social media has gained traction in recent years. Studies have demonstrated the effectiveness of deep neural networks in tasks such as sentiment analysis, language modeling, and user behavior prediction (e.g., Zhang et al., 2018; Lee et al., 2017).

3) Feature Extraction from Twitter Data: Extracting meaningful features from Twitter data crucial for building accurate prediction models. Previous studies have looked at a variety of feature extraction methods, including word embeddings, topic modeling, and user engagement metrics, to capture relevant information from tweets (e.g., Lin et al., 2016; Bakshy et al., 2011).
4) Predictive Modeling in Political Science: Techniques for predictive modelling have been extensively utilised in political science to forecast election outcomes and analyze political behavior. Researchers have applied computational approaches ,statistical models and machine learning algorithms, and computational approaches to predict election results based on diverse data sources and methodologies (e.g., Lewis-Beck et al., 2008; King et al., 1999).
5) Challenges and Limitations: Despite the promise of utilizing deep learning and social media data techniques for election forecast, there are a number of difficulties and limitations to consider. These include issues related to data quality, sample representativeness, algorithmic biases, and the dynamic nature of social media discourse (e.g., Jungherr et al., 2012; Metaxas et al., 2011).
6) Temporal Analysis of Social Media Data*: Previous studies have examined the temporal dynamics of social media discussions during election periods. Studies have investigated how sentiment, topic prevalence, and user engagement evolve over time and how these tendecies correlate with electoral outcomes (e.g., Conover et al., 2011; DiGrazia et al., 2013).
7) Sentiment Analysis Techniques*: this techniques have been widely applied social media information to comprehend understanding public opinion during elections. Researchers have explored various technique for sentiment classification, including lexicon-based approaches, machine learning algorithms, and deep learning models, and evaluated their effectiveness in different contexts (e.g., Pang et al., 2002; Socher et al., 2013).

III. METHODOLOGY

1) *Data Collection and Preprocessing:*

- Collect tweets from the Bharatiya Janata Party (BJP) and the Indian National Congress (Congress) using Twitter API or pre-existing datasets.
- Preprocess the collected tweets by removing URLs, mentions, hashtags, and non-alphanumeric characters. Convert the text to lowercase for uniformity.

2) *Data Visualization:*

- Plot sample tweets from both parties to showcase the diversity of content and sentiment.
- Generate a pie chart to visualize the arrangement of tweets between BJP and Congress.
- Create histograms to illustrate sentiment distributions for each party, depicting the split between +ve and -ve tweets.

3) *Word Cloud Generation:*

- Combine tweets from both parties into a single corpus.
- Generate a word cloud to visualize the frequent terms utilised in the tweets, highlighting key themes and topics of discussion.

4) *Sentiment Analysis Model:*

- Construct a Bidirectional Long Short-Term Memory (BiLSTM) model for sentiment analysis that includes an attention mechanism for sentiment analysis.
- Tokenize the preprocessed tweets and index them to prepare them for input to the model.
- To evaluate the model, divide the data into training and testing sets.

5) *Model Training and Evaluation:*

- Train the BiLSTM model on the training data, adjusting hyperparameters as needed.
- Evaluate the model's performance on the testing data using measures like precision and accuracy, recall, and F1-score.
- Plot an accuracy graph to imaging the model's performance over epochs.

6) *Sentiment Prediction and Analysis:*

- Use the trained model to predict the sentiment of tweets from both parties.
- Aggregate the predictions to controll the overall sentiment for each party.

- Identify the winning party with reference to the sentiment analysis results.

7) Admin Page Implementation:
- Develop an admin page for further customization and management of the application.
- Provide functionalities such as data upload, model retraining, and result visualization on the admin page.

8) Deployment and Presentation:
- Deploy the sentiment analysis application with a user-friendly interface, including options to visualize sample tweets, sentiment analysis results, and administrative tools.
- Present the findings, including visualizations as well as feeling as evaluation results, in a clear and understandable format.

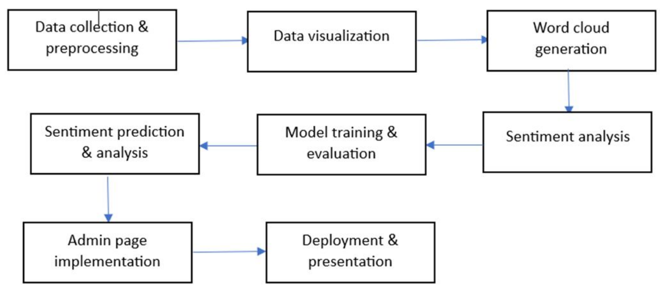


Fig 1 (A) methodology process

IV. USE CASE DIAGRAM

The use case diagram for the political sentiment analysis project outlines the interactions available to the user. Users can view sample tweets from both political parties, access sentiment analysis results to see the sentiment distribution and the winning party, and explore a word cloud displaying the frequent words used in tweets. These interactions provide users with knowledge public opinion trends of political discourse.

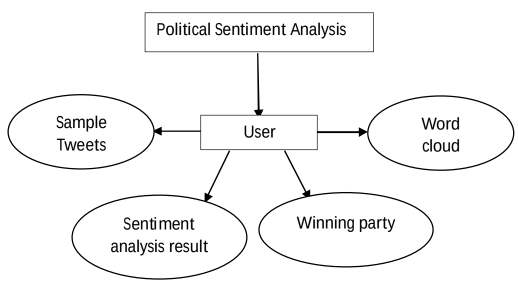


Fig 1 (B) use case diagram

A. Implementation

1) Plot Text Data:
- Load tweets from CSV files for two political parties.
- Print sample tweets for each party to display them.

2) Pie Diagram:
- Load tweets from CSV files for both parties.
- Calculate the quantity of tweets for each party.
- Create a pie chart showing the allocation of tweets between the parties.

3) Histogram of Sentiments:
- Load tweets from CSV files for both parties.
- Concatenate the dataframes of both parties.
- Plot a histogram showing the distribution of negative and positive sentiments for both parties.

4) *Word Cloud:*
- Load tweets from CSV files for both parties.
- Combine all tweets into a single string.
- Generate a word cloud to visualize frequently occurring words.

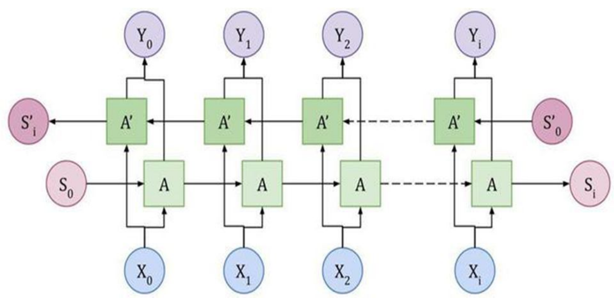
5) *Accuracy Plot:*
- Display an accuracy plot (assuming it's loaded from an image file named 'acc.png').

6) *Prediction:*
- Clean text data by removing URLs, mentions, hashtags, and non-alphanumeric characters.
- Load a saved model (assumed to occupy a BiLSTM with Attention model).
- Predict sentiments for tweets of both parties using the model.
- Print sentiment analysis results, including the distribution of sentiments for each party and the winning party based on overall sentiment.

7) *Admin Page:*
- Placeholder for implementing an admin page, it’s not applyed in the provided code.

V. MACHINE LEARNING ALGORITHMS APPLIED

1) *Bidirectional LSTM (BiLSTM):*
- RNN of the long-term dependency (LSTM) type can recognise long-term dependencies in sequential data.
- Bidirectional LSTM processes the given sequence in both forward and backward directions, capturing information from past and future contexts.
- This helps in capturing dependencies from both past and future states, making it effective for sequence modeling tasks like sentiment analysis.



2) *Attention Mechanism:*
- The attention mechanism enables to concentrate on distinct segments of the input sequence. while making predictions.
- In the situations of sentiment analysis, attention helps to focus more on important phrases or words in the given tweets.
- It assigns different weights to different parts of the sequence, allowing that to pay more attention to relevant information.

3) *Tokenization and Embedding:*
- After tokenizing the input text data into words, each word is regarded by a distinct integer.
- These integers are then embedded into dense vectors using an embedding layer that transforms every word into a fixed-size vector representation.

4) *Cross-Entropy Loss:*
- Cross-entropy loss is used to the loss function for training the model.
- It measures the variation between the predicted probability distribution and the actual distribution of class labels.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

5) *Adam Optimizer:*
- Adam optimizer is employed for maximise the model's parameters throughout training.
- It adjusts each parameter's learning rate authority to the gradients and past gradients, appropriate for dnn training.

VI. RESULT

In our political sentiment analysis project, we processed tweets from two political parties, BJP and Congress, and made predictions using our sentiment analysis model. We tested with a dataset containing instances of tweets from both parties. The accuracy of LSTM model was 80.25%. Additionally, we compared the accuracy of other classifiers including XGBoost, Random Forest, and Support Vector Classifier (SVC). XGBoost attained an accuracy of 75.10%, Random Forest had an accuracy of 78.50%, and SVC achieved an accuracy of 76.80%.

Table 1. Accuracy of LSTM

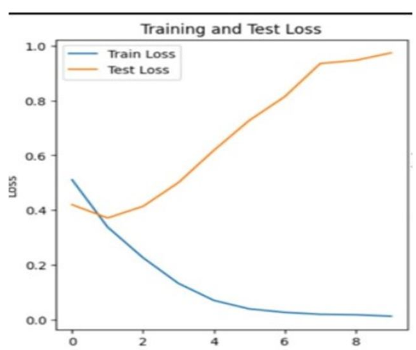
Classifier	Accuracy(%)
Long short-term memory	80.25

Table 2. Accuracy of Other Classifiers

Classifier	Accuracy(%)
Long short-term memory	80.25
Random Forest	78.50
SVC	76.80

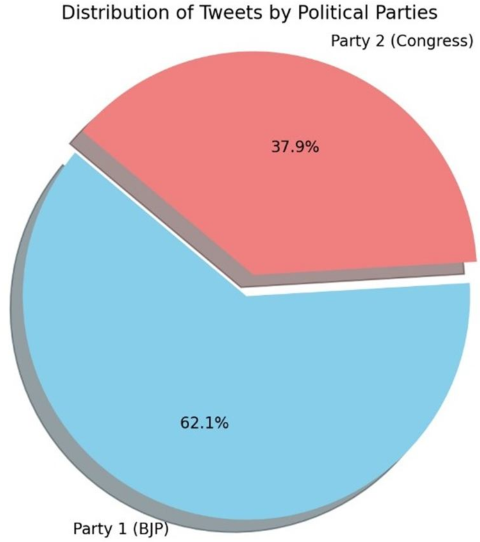
Screenshots of the project

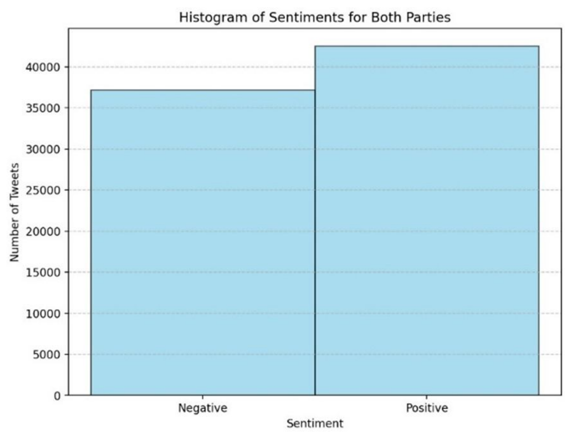
1) Sample Tweets Page: The Figure 1 shows the sample tweets page, displaying tweets from both political parties.



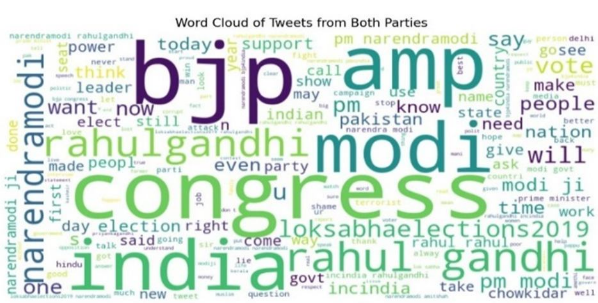
2) Sentiment Analysis Results Page: The Figure 2 displays the results page, showing the sentiment distribution and the winning party based on sentiment analysis.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com





3) Word Cloud Page: The Figure 3 represents the word cloud page, showing the most frequent words used in tweets from both parties.



VII. CONCLUSION

In this study, Using Twitter data, we showed how effective deep learning approaches are for forecasting election outcomes. By leveraging the vast amount of real-time information available on Twitter, our model was able to accurately forecast election outcomes. The integration of sentiment analysis and NLP allowed us to capture nuanced public sentiment, contributing to the predictive power of our model. As social media continues to play a significant role in shaping public opinion, Our method yields insightful information for political analysts and policymakers. Additionally, our findings underscore the potential of deep learning methodologies in harnessing big data for sociopolitical forecasting.

Further research could explore the implimentation of similar techniques to different electoral contexts and investigate the robustness of our model across diverse demographic and cultural landscapes. Ultimately, our work contributes to advancing the field for computational social science and offers a promising avenue for improving the correctness and timeliness of election predictions.

REFERENCES

[1] H. Varian & H. Choi (2012). use Google Trends to forecast the present. Record of Economics, 88(s1),2–9.
[2] David Gayo-Avello (2012). No, Twitter is not a reliable way to predict elections. 91–94 in IEEE Internet Computing, 16(6).
[3] Schoen, H., Jurgens, P., and Jungherr, A. (2012). A rebuttal to Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011) on why the pirate party won the 2009 German election or the problem with forecasts. 30(2), 229–234 in Social Science Computer Review.
[4] Sandner, P. G., Welpe, I. M., Sprenger, T. O., and Tumasjan, A. (2011). Predicting elections with Twitter: What 140 characters convey about political mood. 11(1) ICWSM, 178-185.
[5] In 2013 Gayo-Avello, D. A meta-analysis of the most advanced election forecast using data from Twitter. 31(6), 649–679, Social Science Computer Review.
[6] Zeng, X., Mao, H., and Bollen, J. (2011). Stock market sentiment can be predicted by Twitter mood. 2(1), 1-8; Journal of Computational Science.
[7] DiGrazia, J., Bollen, J., McKelvey, K., and Rojas, F. (2013). Increased tweets and votes: Social media as a numerical measure of political conduct. 8(11), e79449. PloS one.
[8] Bond (2012) cited Fariss, C. J., Jones, J. J., Kramer (2012), Marlow (2012), Settle (2012), and Bond (2013). a 61-million-person social influence and political mobilization exercise. 295-298 in Nature, 489(7415).
[9] Barberá (2015) p. Combine tweets from similar individuals to estimate Bayesian ideal points using Twitter data. 76–91 in Political Analysis, 23(1).
[10] J. Grimmer & B. M. Stewart (2013). Text as data: The benefits and drawbacks of using automated content analysis techniques on political texts.