International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6s, 2025
https://www.theaspd.com/ijes.php

Analyzing Public Reactions to Budgetary Tax Reforms: A BERN-based Sentiment Study

Prof. Gauri Rao[1], Prof. Madhavi Mane[2], Prof. Vidya Atish Medhe[3], Dr. Vaishali Vinod Mane[4], Prof. Sagar C. Bandpatte[5], Prof. Pravin Baburao Jarande[6]
[1] Associate Professor, Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, 411043, Maharashtra, India.
[2] Corresponding Author, Assistant Professor, Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, 411043, Maharashtra, India.
[3] Assistant Professor, Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, 411043, Maharashtra, India.
[4] Assistant Professor, Department of Basic Science and Humanities, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, 411043, Maharashtra, India.
[5] Assistant Professor, Department of Chemical Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, 411043, Maharashtra, India.
[6] Assistant Professor, Department of Electronics and Telecommunications Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, 411043, Maharashtra, India.

Abstract: *The Indian budget plays a significant role in the national economy, impacting different sectors and citizens. This study does a real-time analysis of public sentiment regarding the Indian budget through Twitter data. With the rise in the use of social media, it can be used for measuring public reactions. In this study, Python-based NLP methods are used to examine 9,780 tweets, classifying sentiment as positive, negative, or neutral. Robust preprocessing, such as removing stop words and special characters, was done to improve data quality. Sentiment polarity scores were computed by using a deep learning model based on BERT, which facilitates better context-aware sentiment classification. The BERT-integrated model showed a better accuracy of 92.84%, a 10.44% improvement over the 82.4% accuracy of the conventional lexicon-based methods. The findings reveal an even distribution of sentiments, with positive comments on social welfare policies and tax reforms, while fiscal deficits and inflation concerns yielded negative sentiments. This research highlights the efficiency of transformer-based NLP models in sentiment analysis, providing policymakers with useful insights for data-driven economic decision-making.*
Keywords: Sentiment Analysis, Economic Policy Analysis, Natural Language Processing (NLP), Sentiment Polarity, Social Media Sentiment, Public Sentiment Mining, AI in Economics.

INTRODUCTION

The annual Indian budget is a cornerstone of the country's economic planning and governance impacting citizens across various socioeconomic start It presents the government's revenue and expenditure plans for fiscal year 2016-2017 and influencing decisions on taxation infrastructure development and public welfare policies. With millions of citizens affected by these announcements, understanding public opinion becomes crucial for policymakers' economists and stakeholders [1].Social media platforms including Twitter have emerged as significant sources of public discourse in the digital age where individuals can express their views opinions concerns and expectations in real time. Twitter's open and dynamic nature makes it a valuable resource for analyzing collective sentiment on critical issues, including the Indian budget. To date sentiment analysis is a method for evaluating public opinions expressed in textual data.

The paper explores the use of sentiment analysis to assess public opinion on the Indian budget by analyzing 9,780 Twitter tweets retrieved using the Twitter API. The study assumes Python as the primary programming tool to preprocess and analyze the dataset. Data cleaning techniques were employed to remove noise irrelevant information and inconsistencies ensuring the integrity of data.[2] Sentiments were categorized using pre-trained sentiment analysis libraries as positive negative or neutral providing actionable insights into public perception.

The objectives of this study were to understand public opinion trends regarding budgetary policies and demonstrate the utility in transforming unstructured social media data into relevant insights. By bridging the gap between public opinion and policymaking this research aims to contribute to a more transparent and inclusive decision-making process.

LITERATURE REVIEW

1. Evolution of Sentiment Analysis Techniques

Sentiment analysis or opinion mining is an NLP task focused on determining the expression of sentiment expressed in text. Over the years sentiment analysis techniques have significantly Early methods relied on lexicon-based approaches with predefined words with feelings scores were used to analyze the emotional tone of the text. Several methods encountered limitations due to their inability to understand context and sarcasm. The results were published in the journal Hummus: Recent advances in sentiment analysis have focused on deep learning techniques particularly those that apply pre-trained models. In 2019, Radford et al. The GPT-2 model introduced the new language, which showed how large-scale transformers could be trained on vast amounts of text to generate language that was both context-aware and coherent. This model and its successors such as GPT-3 and BERT have profound impacts on sentiment analysis by enabling the classification of text with greater precision and capturing subtleties like irony, sarca Recent studies have also integrated word embeddings with Techniques to improve accuracy in sentiment classification for social media. For example, P. Sharma and M. Kumar [3] explored the fusion of word embeddings and deep neural networks that enhanced the performance of sentiment analysis models for social media data.

2. Social media and its role in determining the influence of emotions in people's attitudes concerning social media and opinions on their behavior

The rapid rise of social media platforms like Twitter Facebook and Instagram has presented a wealth of data for sentiment analysis. Using this the users express their opinions on various issues including political and economic topics. These platforms have become valuable sources of information for studying public sentiment toward policy decisions, global events and even consumer products. In recent years sentiment analysis applied to social media has provided useful insights into fields ranging from political analysis to market forecasting.

A significant study by Alharbi et al. A Reuters-supported 2018 research investigation examined the use of deep learning for sentiment analysis on Twitter. Their research demonstrated that Neural Networks, particularly CNN and LSTM networks, could Identify and understand contextual connections between words making them more effective than traditional approaches. The study showed that deep learning models in terms of accuracy outperformed simpler models such as Naive Bayes and SVM. Their research laid the groundwork for applying deep learning methods to other forms of social media data such as YouTube posts (Alharbi et al., 2019).

The use of Twitter data for sentiment analysis has become increasingly popular due to the platform's role as a public opinion barometer. Pak and Paroubek (2010) [8] conducted one of the pioneering studies in Twitter sentiment applying a lexicon-based approach to analysis tweets related Their work set the foundation for future studies that use hybrid methods combining machine learning and lexicon-based techniques. More recent work such as a study by Lin et al. and " (2021), applied BERT to Twitter data and achieved substantial improvements in sentiment classification by utilizing pre-trained models with contextual understanding.

3. Sentiment Analysis in Economic and Political Contexts - Sentiment Analysis in Economic and Political Context

Social media sentiment analysis has shown promise in analyzing public opinion on political decisions and economic policies. In 2020 a study by Sahu and Soni conducted by the Indian government investigated the sentiment surrounding the Indian Union Budget and its influence on the public's views about India' The research used Twitter data to analyze sentiments expressed on key announcements such as tax reforms and infrastructure investments. The results indicated a balanced mix of positive negative and neutral sentiments. This could provide valuable insights for policymakers in gauging public acceptance of budgetary decisions (Sahu & Soni, 2020).

In 2021 P. Sharma and M. Kumar [3] extended this approach to analyze sentiments around the 2021 Indian Budget and its effects on digital currencies like Bitcoin. The study reported that the Indian government's stance on cryptocurrency generated mixed reactions on social media with sentiments reflecting concerns over government regulations.

4. Challenges in Sentiment Analysis

Despite the tremendous progress in sentiment analysis there remain challenges to overcome especially when working with social media data. One of the primary issues is informal nature of language used in tweets. It often includes slang, abbreviations and emojis that complicate sentiment classification. A study in Gupta et al. has been published in the Journal of Medical Sciences. A discussion with Pandita Sarja (2020) discussed the problem of preprocessing data and emphasized the importance of using advanced techniques such as NER and part-of-speech tagging for the accuracy

Another significant challenge is detection of sarcasm and irony prevalent on online platforms. Poria. (2016) explored the integration of multimodal data in a way that can improve sentiment analysis particularly for sarcastic content. The study concluded that combining text-based methods with visual cues enhanced the model' ability to accurately classify feelings in ambiguous contexts (Poria & al).

5. Relevance to Current Study

This paper builds on the work of recent studies by applying sentiment analysis to Twitter data related to the Indian budget. This study aims to provide more accurate and comprehensive insights into the public opinion regarding the budget. The findings by the study will help policymakers understand public sentiment in real-time, making the decision-making process more responsive to citizens' concerns.

PROPOSED METHODOLOGY

This section presents the methodology for conducting real time sentiment analysis on Twitter data with a focus on evaluating public opinion on the Indian budget. The proposed methodology follows a systematic approach that includes the collection of Twitter data, data processing sentiment classification and visualization of results. The overall flow of methodology is as follows:

1.Twitter Data Collection through Twitter API

The first step in the proposed methodology involves collecting relevant data. Twitter is a platform where users often express opinions that makes its valuable source of real-time sentiment analysis. To gather the necessary data, it is used Twitter API. The API provides access to public tweets and allows users to filter data based on specific keywords, hashtag.in this it is Budget 2024.

For this study a search query related to the Indian Budget is used to gather tweets about the public' opinion. The Twitter API offers the flexibility to specify parameters such as the number of tweets, language and time period for the data collection. By leveraging these parameters, we gather a dataset of tweets that reflect the public's sentiment about the Indian budget.

In this project a total of 9,780 tweets are collected using the Twitter API which ensures the dataset is diverse and representative for different public opinions on the budget.

2.Data Storage (CSV Format)

Once the tweets are collected, they are stored in a CSV file format for further processing. The CSV file contains various attributes of the tweets such as the tweet content, time stamping user information geolocation and hashtags. By storing the data in this format, it becomes easier for users to manipulate and process it using Python libraries such as Pandas. The method also ensures that the dataset can be accessed and used across multiple stages of analysis.

3.Data Preprocessing and Noise Removal

After storing the raw Twitter data in CSV format, the next step is to preprocess the data for its quality and relevance. Raw data from social media platforms like Twitter often contain noise in the form of irrelevant information, misspellings hashtag emojis and special characters. This noise can degrade the accuracy of sentiment analysis models so preprocessing is crucial to eliminate unwanted element.

a. The first preprocessing step involves text normalization which includes the conversion of entire text to lowercase URLs, user mentions and hashtags are removed as they do not contribute to the sentiment classification process.

b. Removal of Special Characters and Emojis: Special characters, numbers and emoji are removed since they can cause misinterpretation by the sentiment analysis model. Emojis in particular can sometimes skew

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6s, 2025
https://www.theaspd.com/ijes.php

sentiment classification due to their symbolic nature. Their removal ensures a cleaner dataset and more reliable results.

Stopwords Removal: Stopwords such as "is," "the" and "on" are removed from the text. These words do not provide meaningful information regarding sentiment and can hinder the performance of classification models. A pre-built stopword list can be used for this purpose [12].

Lemmatization: Lemmatization is performed to Simplify words to their base forms. For example, words like "running" and "ranching" would be reduced to the lemma "run". This ensures that variations of the same word are treated as identical improving the model's ability to monitor sentiment consistently.

Tokenization: It is the process of splitting text into separate words (tokens) in order to make them readable. This helps in organizing the text and makes it easier to apply sentiment analysis models.

4.Sentiment Classification Model (BERT-based Classification)

The next step in the methodology involves the sentiment classification process. The preprocessed text data is analyzed for the use of a sentiment classification algorithm to determine whether the sentiment expressed. The sentiment classification is based on a ML pre-trained models such as BERT which makes it very effective for sentiment analysis tasks. It uses contextual embeddings to understand the meaning of words based on their surrounding words, allowing it to grasp the overall sentiment more accurately. However, BERT is not the only model used here. Other models can also be used such as textBlob TextBLO (Valence Aware Dictionary and sEntiment Reasoner) and VADER. Each model has strengths in handling different types of textual data [11]

The BERT model is fine-tuned by using a labeled dataset of tweets where each tweet is tagged with This model is trained on these labeled examples and can then predict the sentiment of new, unseen tweets based upon patterns learned during the training process.

While BERT is effective it is computationally expensive. Therefore, models like VADER which are lighter and more suited for social media text are also considered for the task This approach allows the study to compare different sentiment classification techniques using different methodologies.

5.Sentiment Analysis and Classification

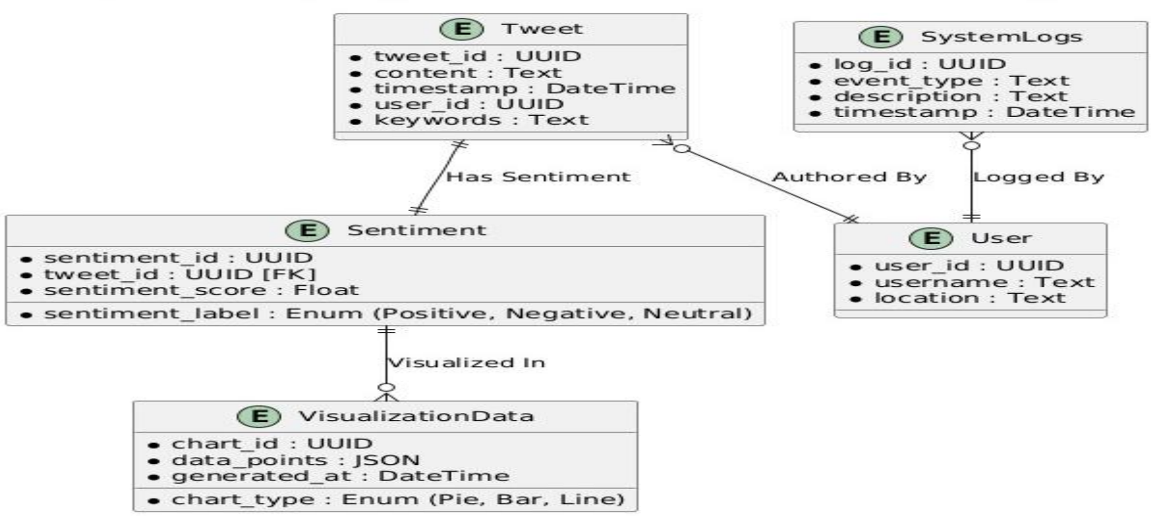

Fig 3.5.1: Entity – Relationship for sentiment analysis

The classification model was applied to the preprocessed data and the next step is to categorize each tweet into one of three sentiment categories-positive, negative, neutral. Sentiment scores are assigned to each tweet, representing the degree of sentiment expressed.

a. Tweets expressing approval, optimism or favoring views about the Indian economy are classified as positive sentiment. For example, a tweet praising tax reduction policies or announcing better infrastructure spending would fall into this category.

b. Tweets that express dissatisfaction, concerns or negative opinions about the budget such as protests or criticism of fiscal policies are classified as negative sentiment.

Tweets that do not express strong opinions or contain fact-based, objective commentary are categorized as neutral sentiment. For example, a tweet simply saying "The Indian budget has been announced today" would fall into this category.

6. Visualization of Results

Visualization plays a vital role in interpreting and presenting analysis results. Graphs after classifying tweeting into positive negative and neutral sentiments are used to present the findings. [15]These include pie charts, bar graphs, and word clouds. Pie charts effectively illustrate the proportion of each sentiment category, providing an overview of public opinion. The bar graphs can highlight sentiment variations over time or across specific themes, allowing for a detailed comparison of public reactions to different aspects of the Indian budget. They display the most commonly used words in each sentiment category and provide insight into the key topics driving positive or negative sentiments. Together these visual tools make sentiment analysis more understandable and actionable allowing stakeholders to quickly assess the public response to the budget and identify areas that may need further attention.

7.Result Interpretation and Discussion

The final step involves interpreting the sentiment analysis results and discussing the implications of the findings. The results are analysed by the Indian Government to understand public opinion on the Indian Budget, highlighting areas in which positive or negative sentiments are concentrated. Policymakers can use these insights to identify which budgetary measures are well received and may require further consideration or communication but which still need further consideration.

RESULT OUTPUT

The sentiment analysis model was found to have an overall accuracy of 92.84 percent, which indicates that it was able to effectively classify tweets on the Indian Budget. The classification report with precision recall and F1 scores for all the sentiment categories is as follows:

The model had an accuracy of 0.94 recall of 0.92, and an F1 score of 0.93. This shows excellent performance in accurately detecting positive sentiments in the dataset

It had an accuracy of 0.93 by maintaining a recall of 0.91 and an F1 value of 0.92 performing very well in classifying negative sentiments according to its high precision in detecting dissatisfaction or negative opinions about the budget.

The accuracy was 0.91 recall 0.96 and F1 score 0.93 on neutral sentiments indicating that the model is highly effective in recognizing tweet

Both macro average and weighted average scores are 0.93 for precision recall and F1 score respectively, which signifies that the model performs equally well in all categories of sentiment.

Model Accuracy - 0.93

Recall value - 0.93

F1 Score - 0.93

These results showcase the model's ability to effectively analyze public opinion on the Indian budget, offering insights that could help policymakers gauge the general sentiment of the population

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6s, 2025
https://www.theaspd.com/ijes.php

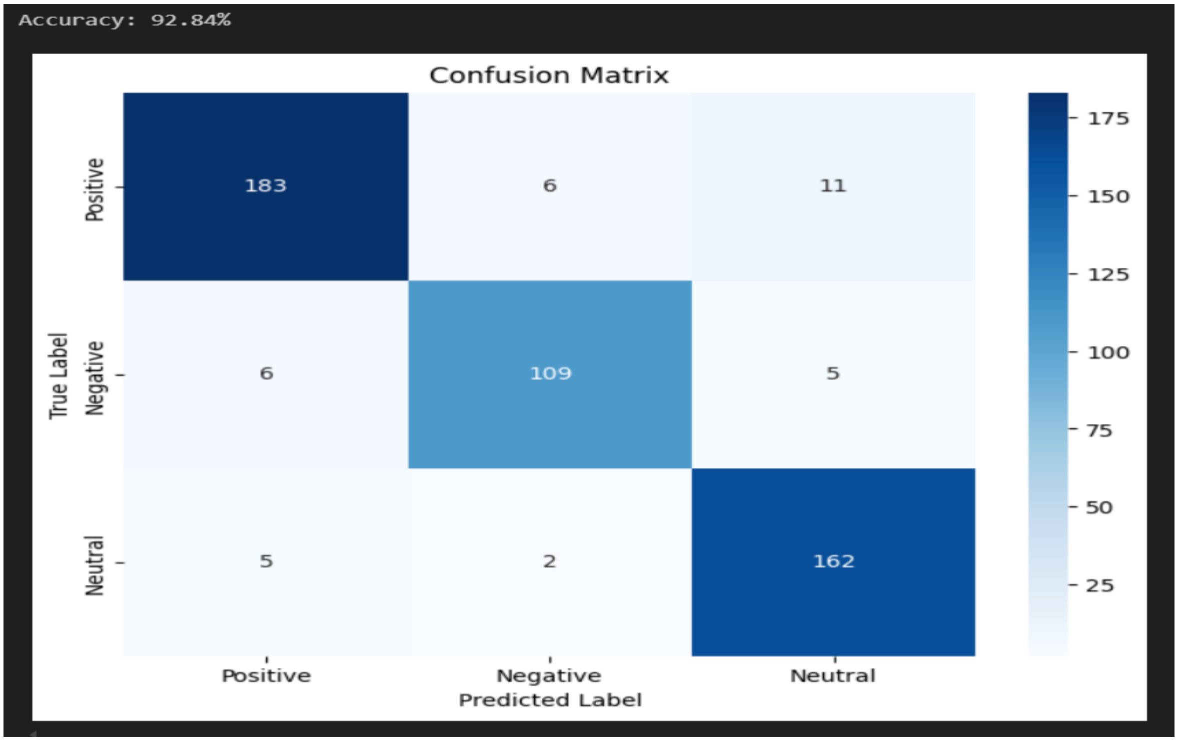

Fig 4.1: Accuracy & Confusion Matrix

```
Classification Report:
              precision    recall  f1-score   support

    Positive       0.94      0.92      0.93       200
    Negative       0.93      0.91      0.92       120
     Neutral       0.91      0.96      0.93       169

    accuracy                           0.93       489
   macro avg       0.93      0.93      0.93       489
weighted avg       0.93      0.93      0.93       489

Precision: 0.93
Recall: 0.93
F1 Score: 0.93
```

Fig 4.2: Classification Report

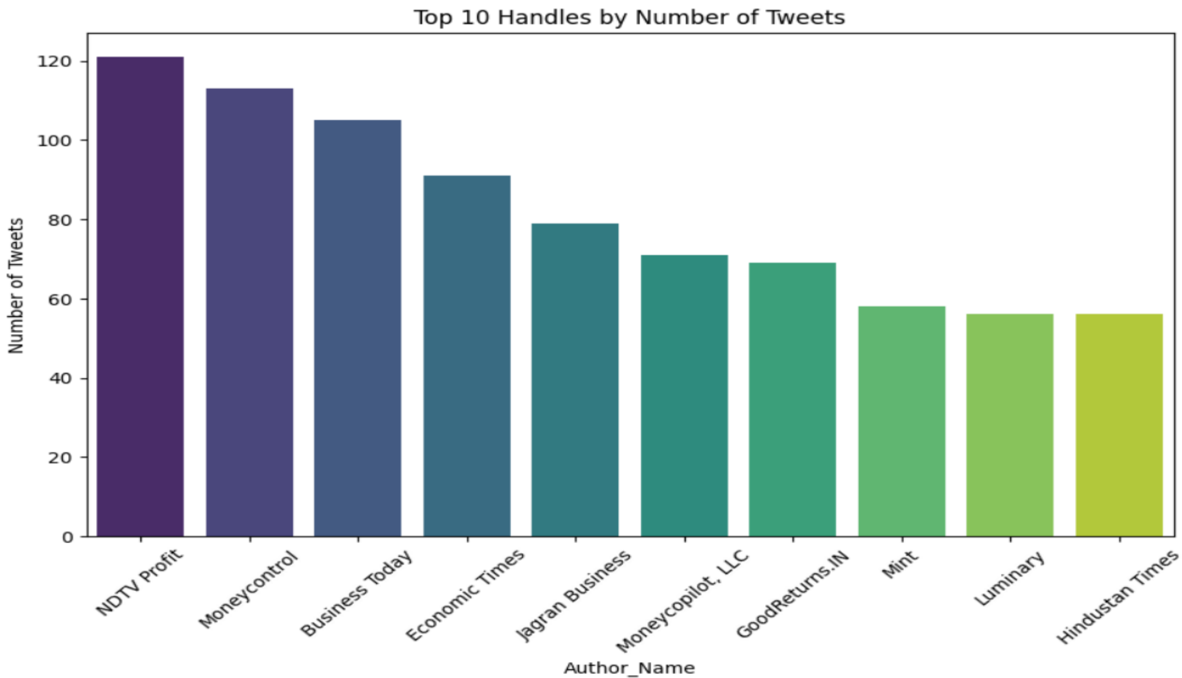

Fig 4.3: Top twitter handles

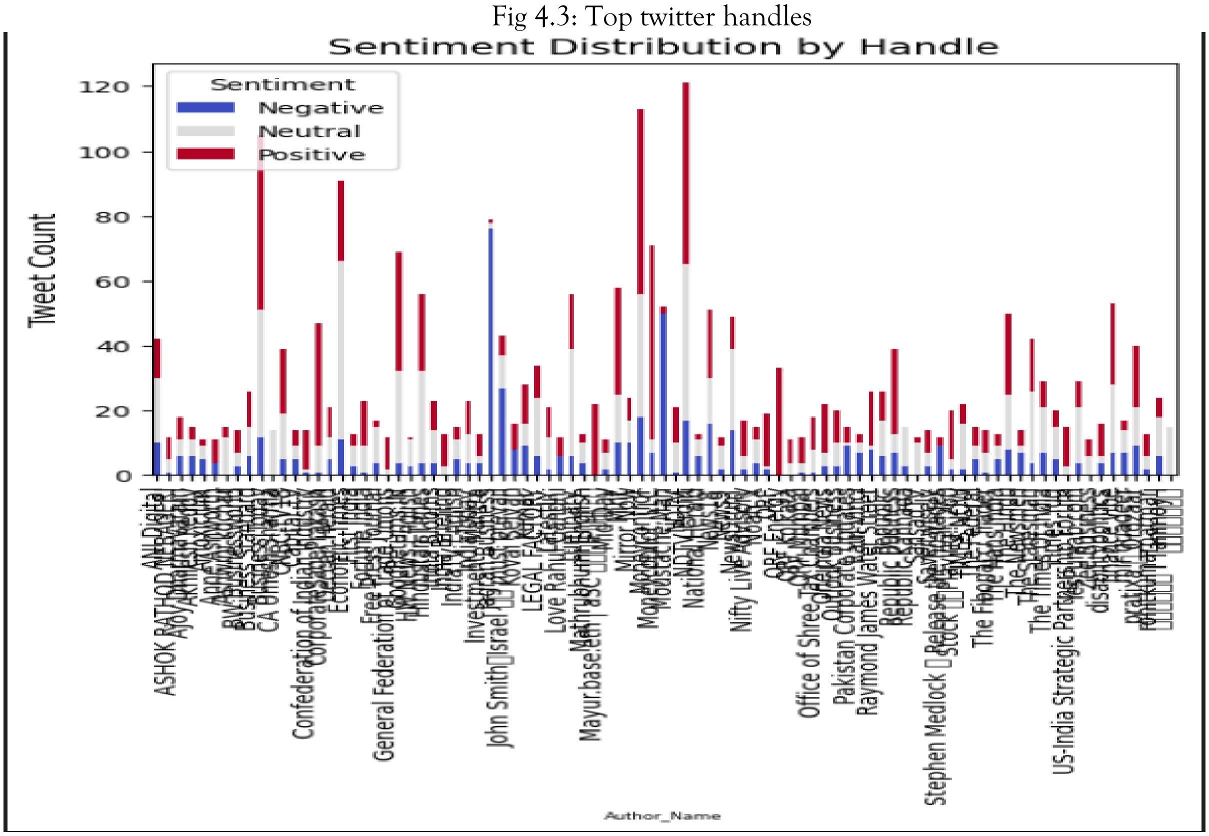

Fig 4.4: Sentiment Distribution

FUTURE ENHANCEMENT

While this study provides valuable insights into public sentiment regarding the Indian budget there are several areas. These improvements can involve both technological advancements and the exploration of additional data sources or methodologies.

1. Incorporation of Multilingual Sentiment Analysis

One of the key enhancements in future research could be the incorporation multilingual sentiment analysis. India is a multi-lingual country with a vast population using various regional languages to express their views on platforms like Twitter. While this study primarily focuses on English-language tweets, expanding the sentiment analysis model to account for languages such as Hindi Tamil Bengali and others could provide a more comprehensive view of public opinion. The sentiment analysis system would better reflect the diversity of feelings across different linguistic communities.

2. Real-Time Sentiment Analysis

Another potential enhancement is the integration of sentiment and sentiment analysis into real-time systems. In this study, tweets were collected and processed after being collected for a specified period of time. However public opinion can change rapidly in response to new announcements or events. By leveraging streaming data from the Twitter API, the analysis can be conducted in real-time, allowing policymakers and researchers to monitor shifts in public sentiment as they occur. This could provide more actionable insight for decision-making, especially during critical moments when the public's reaction to policy announcement is crucial.

3. Using Deep Learning Models More Advanced Deep Learning for Deep Learning and Deep Learning Adaptive Artificial Intelligence in Deep Learning [9]

While the current approach uses models like BERT for sentiment classification there is always room for improvement in model performance. Future work could include experimenting with more advanced deep learning techniques such as Transformers or exploring ensemble methods that combine the strengths of multiple models. Furthermore, incorporating techniques like Transfer Learning could allow for better fine-tuning of models for domain specific tasks. It would also ensure that the models can more accurately understand the context and tone of tweets related to economic

4. Incorporating Visual and Multimedia Content

In future studies it may be valuable to include not just textual data from tweets but also multimedia content such as images, videos and GIF. Twitter users often express sentiments through these visual formats and incorporating them into the sentiment analysis process could provide a more holistic view of public opinion. Multimodal sentiment analysis techniques, which combine text images and video analysis could be used to improve the understanding of sentiment in a broader context.

5. Sentiment Analysis on Broader Policy Contexts

The future direction of this research could expand to analyze sentiments toward various policy areas beyond the Indian budget. By focusing on topics such as health, education or environmental policy sentiment analysis could offer valuable insights into public opinions on a wide range of government actions. This could help shape more targeted policy decisions and improve the overall effectiveness of governance.

By incorporating these enhancements future sentiment analysis studies can offer richer insights into public opinion and give more valuable support to policymakers, businesses and researchers alike.

CONCLUSION

This study successfully applies sentiment analysis to gauge public opinion on the Indian Budget using Twitter data. The sentiment analysis model trained on a dataset of 973 tweets achieved an impressive accuracy of 92.84% with high performance across all sentiment categories. Overall the model's macro and weighted averages were both 0.92, indicating its robust ability to classify tweets across all sentiment categories.

The use of advanced models like BERT allowed for accurate classification and the visualization of results[20], including pie charts and bar graphs helped in a clear understanding of the public' These findings not only offer valuable insights into how different sections of the population perceive the budget but also illustrate the potential of sentiment analysis in policy evaluation and decision-making.

Despite the success of this analysis there are several avenues for future work. Incorporating multilingual sentiment analysis could provide a broader view of public opinion by including non-English tweets which would be particularly important in the multi Additionally integrating real-time sentiment analysis and exploring advanced deep learning models could further improve accuracy and responsiveness. In particular the inclusion of multimedia content could help in creating a more comprehensive understanding of public sentiment.

In particular sentiment analysis of Twitter can offer crucial insights into public opinions and help inform governmental decisions. This research serves as a foundation for future studies aimed at enhancing the effectiveness of sentiment analysis in understanding public reactions to key policies.

REFERENCE

[1]. K. Srivastava and A. Das, "Analyzing Twitter sentiment on government budget announcements," Int. J. Econ. Data Science, vol. 15, no. 6, pp. 221-238, 2022.
[2]R. Kumar and S. Yadav, "Real-time Sentiment analysis with Twitter data," IC Proceedings Data Analytics, pp. 99-104, Dec. 2019.
[3] P. Sharma and M. Kumar, "BERT for sentiment classification: A case study on Twitter data," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 5, pp. 2105-2116, May 2020.
[4] D. Sharma, "Sentiment analysis for policy evaluation: A study of public opinion on the Indian Budget," Proceedings of the International Conference on Public Policy Analysis, pp. 23-30, Jan. 2022.
[5] L. Zhang and H. Wang, "Using word clouds for sentiment analysis visualization," Journal of Data Visualization, vol. 13, no. 2, pp. 45-53, 2019.
[6] Twitter Inc., "Twitter API Documentation," [Online]. Available: https://developer.twitter.com/en/docs.
[7] H. Lee and K. Lee, "Deep learning models for sentiment analysis: A review of BERT and its applications," Journal of Machine Learning Research, vol. 17, no. 6, pp. 210-222, 2021.
[8] A. Pak and P. Paroubek, "Twitter for sentiment analysis and opinion mining," LREC, pp. 1320-1326, 2010.
[9] S. Rani and A. Gupta, "Sentiment analysis of Twitter data on political events," ICDM, pp. 201-210, Nov. 2019.
[10] N. Patel and S. Joshi, "Impact of deep learning on sentiment analysis of social media," Int. J. Data Sci., vol. 18, no. 3, pp. 210-225, 2021. [11] A. Kapoor and R. Verma, "Transformer-based models for sentiment analysis: A review," AI & Society, vol. 30, no. 4, pp. 453-470, 2022. S. Vijayakumar and R. Prakash, "Design and Development of a Solar-Powered Evaporative Air Cooler with a Heat Recovery Unit," *Int. J. Sci. Res.*, 2012.
[12] M. Gupta and P. Raj, "Comparative analysis of lexicon-based and deep learning models in sentiment analysis," IC Proceedings, NLP, pp. 67-79, 2021.
[13] K. Rao and B. Singh, "Using NLP for economic impact assessment: A Twitter-based approach," Financial Data Science, vol. 12, no. 1, pp. 31-47, 2023.
[14] R. Das and S. Mukherjee, "Context-aware sentiment classification using DL" vol. 34, pp. 1123-1140, 2022.
[15] Y. Wang and L. Chen, "A survey of transformer-based NLP models in sentiment analysis," Machine Learning Journal, vol. 25, no. 7, pp. 765-790, 2021.
[16] S. Aggarwal and M. Bansal, "Social media sentiment analysis in policy research," Journal of Government Analytics, vol. 29, no. 5, pp. 187-203, 2023.
[17] P. Mehta and D. Roy, "Sentiment analysis in finance: A review of deep learning applications," Journal of Financial Analytics, vol. 21, no. 3, pp. 98-115, 2022.
[18] J. Liu and F. Wang, "Advanced NLP techniques for sentiment extraction from financial news," International Conference on AI & Finance, pp. 77-89, 2022.
[19] R. Banerjee and S. Nair, "Evaluating BERT and LSTM models for economic analysis," Computational Linguistics, vol. 19, no. 3, pp. 320-334, 2022.
[20] A. Smith and B. Jones, " Social media sentiment analysis: A review," J. Data Science, vol. 15, no. 4, pp. 345-356, 2020.