IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, November 2024

Sentiment Analysis on X (Formerly Twitter)

Rajput Sarang Sangram[1], Sonawane Yash Prashant[2], Bhosale Dipak Balasaheb[3], Kadam Gaurav Bhausaheb[4], Prof. V. S. Walke[5]
Students, Diploma of Computer Engineering[1,2,3,4]
Guide, Diploma of Computer Engineering[5]
Guru Gobind Singh Polytechnic, Nashik, India

Abstract: *The current research extends beyond traditional methods by integrating transformer models and exploring domain-specific applications to enhance practical usage. Further challenges in sarcasm detection and context sensitivity are discussed. Sentiment analysis, also known as opinion mining, is a process that aims to determine the sentiment behind a piece of text. With the rapid rise of social media platforms like X (formerly known as Twitter), vast amounts of user-generated content are created daily. This paper explores the application of sentiment analysis on X to extract opinions, emotions, and trends from tweets. We implemented various machine learning (ML) and natural language processing (NLP) techniques to classify tweets as positive, n....*

Keywords: integrating transformer models

I. INTRODUCTION

Moreover, this paper underscores the importance of real-time data in business intelligence and public sentiment assessment, showcasing how applications can be developed for monitoring trends effectively.

X is a microblogging platform where users express their opinions on various topics, ranging from daily activities to global events. The short, informal nature of tweets makes them an excellent source of real-time data for sentiment analysis. Businesses, governments, and researchers can utilize sentiment analysis on X to gauge public opinion, monitor brand reputation, or track the emotional response to events.

II. LITERATURE REVIEW

Recent advancements include the use of GPT models, which offer enhanced comprehension by incorporating larger training datasets and improved architectural designs. Studies have also shown success in domain adaptation techniques to fine-tune models for specific sectors, further improving sentiment accuracy.

Sentiment analysis has been a well-researched topic in the domain of natural language processing. Early approaches used simple lexicon-based methods, where the sentiment was determined by the occurrence of positive or negative words. However, these methods struggled with context, sarcasm, and negations. Modern approaches leverage machine learning, particularly deep learning techniques like LSTMs, CNNs, and Transformer-based models (e.g., BERT), which can understand the nuances of language better.

III. METHODOLOGY

Data Collection

The dataset for this project was obtained using the X API. A collection of tweets from various topics (such as technology, politics, sports, and entertainment) was extracted. The data was labeled as positive, negative, or neutral based on the sentiment expressed in the tweet.

Data Preprocessing

Since tweets are often informal, noisy, and contain abbreviations, the following preprocessing steps were applied: - Tokenization, Stopword Removal, Lemmatization, Handling Emoticons and Special Characters, URL and Username Removal.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, November 2024

Feature Extraction
Techniques included TF-IDF and Word Embeddings (Word2Vec, GloVe).

Machine Learning Models
We explored Naive Bayes, SVM, LSTM, and BERT for classification.

Model Training and Evaluation
The dataset was split into training, validation, and test sets. Metrics used include Accuracy, Precision, Recall, and F1-Score.

IV. RESULTS AND DISCUSSION

The success of BERT in this study aligns with findings from previous research where deep bidirectional representations excel in context understanding. Analyzing misclassifications showed that sarcasm and mixed sentiments were leading causes of errors, suggesting a need for models designed to handle such complexities.
Experiments showed BERT achieved the highest accuracy of 85%, demonstrating superior context understanding.
Simpler models like Naive Bayes and SVM provided decent performance, but deep learning models significantly enhanced results.

V. CONCLUSION

The study also highlights potential for expansion using ensemble methods that combine model strengths and exploring fine-tuning strategies for real-time deployment.
This study successfully implemented a sentiment analysis system for X. The results suggest that transformer-based models are effective in understanding complex language patterns. The tool can be refined for better context-awareness and domain-specific vocabulary.

VI. FUTURE WORK

- Handling Sarcasm: Improved models for sarcastic comments.
- Multilingual Support: Extend to multiple languages.
- Real-time Analysis: Develop a real-time monitoring system.
- Improving Accuracy: Use more advanced models like GPT-4.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers forLanguage Understanding," arXiv preprint arXiv:1810.04805, 2018.
[2] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in Advances in Neural Information Processing Systems, 2013.
[3] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval,2008.