Grenze International Journal of Engineering and Technology, June Issue

Deep Learning-based Sentiment Analysis on Social Media

D.V.L.N. Sastry[1], Swathi Jallu[2], B. Sashank[3], H. Bindu Madhavi[4], M. Praveen[5] and P. Baby Prathyusha[6]
[1-2]Asst.Professor, ECE Dept., Aditya Institute of Technology and Management, Tekkali, AP, India-532201
Email: dvlnsastri@gmail.com, jalluswathi@gmail.com
[3-6]Students, ECE Dept., Aditya Institute of Technology and Management, Tekkali, AP, India-532201
Email: borasashank03@gmail.com, hbindu1509@gmail.com, mattapraveen110@gmail.com, panukubabyprathyusha@gmail.com

***Abstract*— Understanding public sentiment on social media is essential for businesses, researchers, and organizations to track trends, monitor brand reputation, and make data-driven decisions. This study explores sentiment analysis using a combination of machine learning and deep learning techniques, focusing on NLP classifiers, Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) models. By analyzing tweets, posts, and comments, we classify sentiments as positive, negative, or neutral. NLP techniques like tokenization, stopword removal, and vectorization help preprocess the text, making it suitable for model training. Our results show that while traditional models provide decent accuracy, deep learning models, especially LSTMs, excel at capturing context and long-term dependencies in text. This research highlights the power of combining NLP with deep learning to improve sentiment analysis accuracy, providing valuable insights for businesses, social research, and customer engagement strategies.**

***Index Terms*— Sentiment Analysis, Social Media, Machine Learning, Deep Learning, Natural Language Processing (NLP) Classifiers, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM).**

I. INTRODUCTION

In the digital age, sentiment analysis on social media has become a crucial tool for tracking trends, managing brand reputation, and understanding public opinion. With millions of users expressing their thoughts, opinions, and feedback on platforms such as Facebook, Instagram, and Twitter, businesses, researchers, and organizations can extract valuable insights from this vast pool of user-generated content. Sentiment analysis helps classify text as positive, negative, or neutral by identifying its emotional tone. To prepare text data for effective model training, machine learning and natural language processing (NLP) techniques, such as tokenization and stopword removal, are applied. By comparing traditional models with advanced deep learning techniques like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, we demonstrate that while traditional models perform well, deep learning models are superior in capturing context and long-term dependencies in text. This study highlights how the integration of NLP and deep learning methods significantly enhances sentiment analysis accuracy. By automating the processing and analysis of large volumes of social media data, sentiment analysis can uncover trends, gauge public opinion, and support strategic decision-making.

Grenze ID: 01.GIJET.11.2.807_22

As social media content continues to expand, deep learning algorithms are increasingly being employed to improve sentiment analysis efficiency, making it a vital tool for market research, customer engagement, and brand management.

II. NLP AND DEEP LEARNING CLASSIFIER'S

NLP techniques—like Bag of Words (BoW) and Word Embeddings—are crucial methods for turning text into numerical data. BoW is a straightforward technique that counts how often each word appears in a document, ignoring grammar and word order. Word Embeddings, like Word2Vec or GloVe, represent words as dense vectors, understanding relationships and meanings by learning from the context. This makes embeddings much more effective than BoW for capturing deeper language features, especially in deep learning tasks.

A. Bag of Words

A straightforward but powerful NLP method for text classification is called Bag of Words (BoW). It portrays text as a collection of words, concentrating only on word frequency and disregarding grammar and word order. Using this method, a text corpus is transformed into a matrix, with each row denoting a document and each column a distinct word from the whole dataset. The values in the matrix indicate the occurrence of words in each document. BoW lacks context awareness, which makes it less useful for comprehending complex language structures than more sophisticated methods like word embeddings, even though it is simple to use and performs well for simple text analysis.

B. Word Embeddings

In Natural Language Processing (NLP), word embedding refers to converting words into numerical representations that computers can process and understand. Instead of treating words as isolated entities, word embedding places them in a continuous vector space, where similar words are positioned closer together. Popular techniques like Word2Vec, GloVe, and FastText utilize contextual information from large text datasets to learn word meanings. This improves various NLP tasks, including sentiment analysis, chatbots, and text classification, by enabling models to capture relationships between words, such as synonyms and related concepts. As a result, word embeddings enhance a model's ability to understand language and make more accurate predictions. Deep learning, a branch of machine learning, employs multi-layered neural networks—hence the term "deep"—to automatically recognize and extract patterns from large datasets. Unlike traditional machine learning models that rely on manual feature extraction, deep learning models learn directly from raw data, making them highly effective in areas such as image recognition, speech processing, and natural language understanding. These models consist of multiple interconnected layers of nodes, where each layer progressively learns more abstract features of the input data. For sequential data like text, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are commonly used, while Convolutional Neural Networks (CNNs) are well-suited for image-related tasks.

C. Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are commonly used in sentiment analysis on social media due to their ability to process text sequentially while maintaining context from previous words. Unlike traditional models, RNNs utilize a hidden state to capture the flow of meaning in tweets, comments, and posts, making them particularly effective in analyzing informal language, slang, and emerging trends.

However, to address the vanishing gradient problem, which limits their ability to retain long-range dependencies, more advanced models like Long Short-Term Memory (LSTM) networks are often employed to improve sentiment classification accuracy..

D. Long Short-Term Memory Networks (LSTM)

Long Short-Term Memory (LSTM) networks are an advanced form of Recurrent Neural Networks (RNNs) designed to manage long-range dependencies in text. Traditional RNNs often struggle to retain past information due to the vanishing gradient problem, but LSTMs overcome this by using memory cells that preserve essential context while filtering out irrelevant details. This makes them particularly effective for sentiment analysis on social media, where sarcasm, slang, and subtle emotions are prevalent. By processing text sequentially and utilizing techniques like tokenization and word embeddings (such as Word2Vec and GloVe), LSTMs can accurately classify sentiments as positive, negative, or neutral. Although they require significant computational power, their ability to analyze emotions in real-time makes them a valuable tool for businesses and researchers studying consumer behavior and public opinion.

III. ABOUT SENTIMENT ANALYSIS ON SOCIAL MEDIA

Sentiment analysis on social media is the process of analyzing posts, comments, and tweets to determine public opinion and emotions. It helps businesses track brand reputation, understand customer feedback, and monitor trends in real time. Natural Language Processing (NLP) techniques like tokenization, stopword removal, and word embeddings prepare text for classification into positive, negative, or neutral sentiments. Machine learning and deep learning models, including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, improve accuracy by capturing context and sentiment nuances. With the massive amount of social media data available, sentiment analysis is crucial for market research, customer engagement, and decision-making.

IV. ABOUT DATASET

The Twitter US Airline Sentiment dataset, which categorizes tweets about major U.S. airlines as positive, neutral, or negative, and Sentiment140, which comprises 1.6 million tweets labeled as positive or negative using emoticons, are popular datasets for sentiment analysis on social media using deep learning techniques like LSTM, RNN, and NLP. Some noteworthy datasets are SemEval (Semantic Evaluation) datasets, which offer benchmark datasets for sentiment tasks on Twitter; IMDb Reviews, which classifies movie reviews according to sentiment; and SST (Stanford Sentiment Treebank), which provides fine-grained sentiment analysis. Deep learning models are frequently trained using these datasets to identify sentiment patterns in social media discussions.

V. PROPOSED METHOD

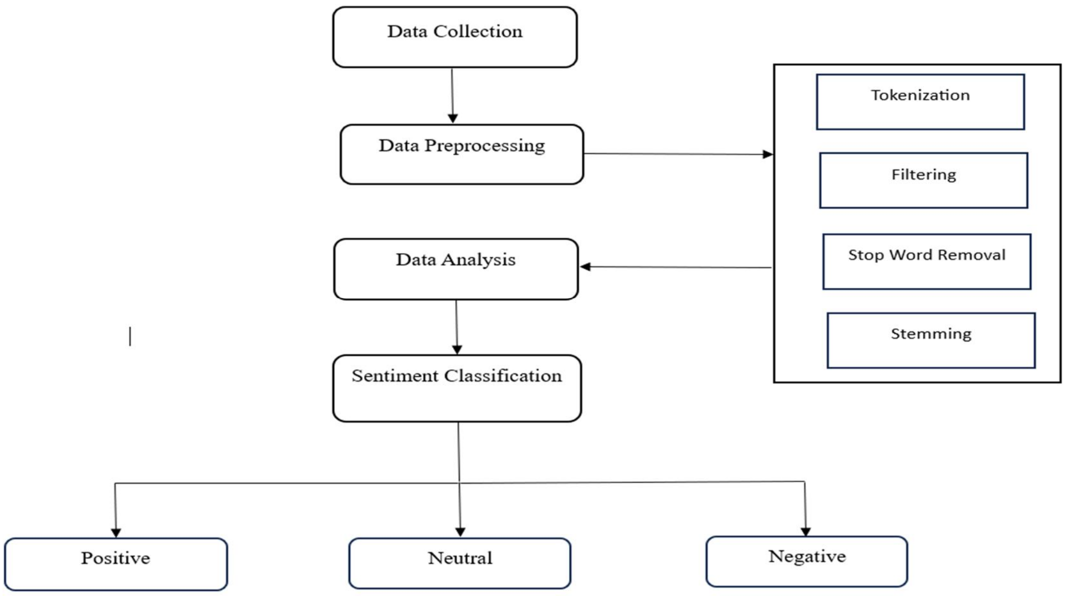


Fig 1

VI. IMPLEMENTATION & RESULTS

The dataset consists of 35,910 records and 59 columns in total. Out of these, 58 columns serve as independent features, while one column is designated as the target variable, which represents different emotions. Each record corresponds to a specific emotion, labeled with a value from 0 to 7. The emotion categories include: 0 for happy, 1 for fear, 2 for disgust, 3 for anger, 4 for sadness, 5 for neutrality, 6 for surprise, and 7 for calm. This structure provides a detailed representation of emotional states across the dataset. In this study, the model was trained using 75% of the dataset, and its performance was tested using the remaining 25%.

A. Bag of Words

Accuracy of 66.2%, with lower precision and recall, especially for neutral and positive sentiments.
Accuracy: 66.2 %

Classification Report:

TABLE I: CLASSIFICATION REPORT FOR BAG OF WORDS

Sentiment Analysis	Precision	Recall	f1-Score
Negative	0.72	0.85	0.78
Neutral	0.48	0.34	0.40
Positive	0.57	0.41	0.47

B. Word Embeddings

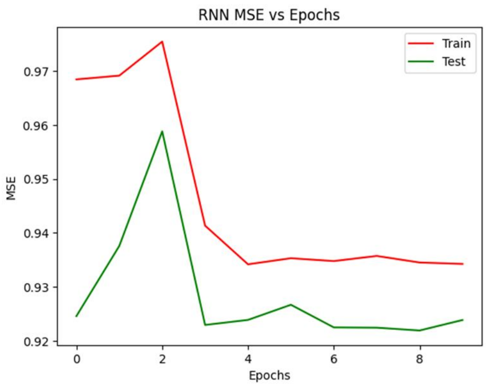
Accuracy improved to 68.3%, showing better context understanding than BoW but still struggling with neutral sentiment classification.
Accuracy: 68.3 %
Classification Report:

TABLE II: CLASSIFICATION REPORT FOR WORD EMBEDDING

Sentiment Analysis	Precision	Recall	f1-Score
Negative	0.68	0.93	0.78
Neutral	0.50	0.22	0.31
Positive	0.63	0.23	0.34

C. Recurrent Neural Networks (RNN)

Demonstrating a significant improvement over traditional models by capturing sequential text dependencies.
Accuracy: 76.6%



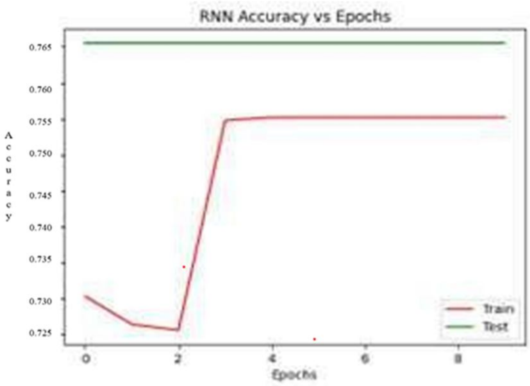
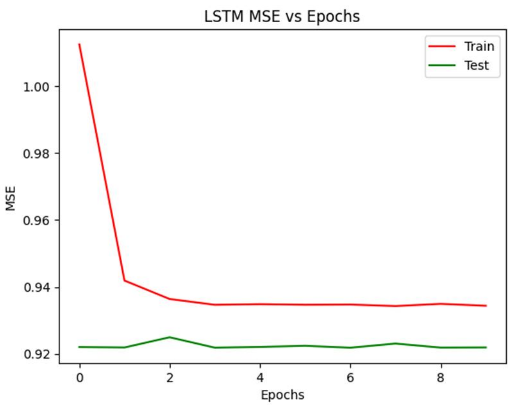


Fig 2: RNN MSE vs Epochs Graph

Fig 3: RNN Accuracy vs Epochs Graph

D. Long Short-Term Memory Networks (LSTM)

Achieved the highest accuracy of 88.2%, effectively handling long-range dependencies and sentiment nuances.
Accuracy: 88.2%



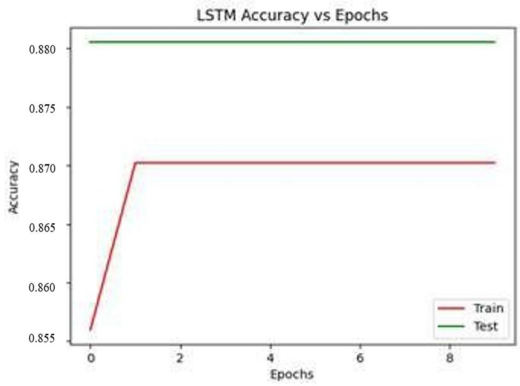


Fig 3: LSTM MSE vs Epochs Graph   Fig 4: LSTM Accuracy vs Epochs Graph

VII. Conclusion

The results reaffirm the significance of combining NLP preprocessing techniques with deep learning to achieve higher accuracy in sentiment classification. The study shows that deep learning models outperform traditional NLP approaches in sentiment analysis on social media. While BoW and Word Embeddings offer a foundation, they lack contextual understanding, RNNs perform better but have vanishing gradient issues, and LSTMs are the most effective way to capture sentiment in social media text.

References

[1] Malliga Subramanian, Veerappampalayam EaswaramoorthySathiskumar, G. Deepalakshmi, Jaehyuk Cho, G. Manikandan, “A survey on hate speech detection and sentiment analysis using machine learning and deep learning models”,2023, DOI: https://doi.org/10.1016/j.aej.2023.08.038

[2] Shini Renjith , Annie Abraham, Surya B. Jyothi, Lekshmi Chandran, Jincy Thomson, “An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms”, 2021, DOI: https://doi.org/10.1016/j.jksuci.2021.11.010

[3] Kakuthota Rakshitha, Ramalingam H M, M Pavithra, Advi H D, Maithri Hegde, “Sentimental analysis of Indian regional languages on social media”, 2021 DOI: https://doi.org/10.1016/j.gltp.2021.08.039

[4] Sayyida Tabinda Kokab , Sohail Asghar, Shehneela Naz, “Transformer-based deep learning models for the sentiment analysis of social media data, 2022, DOI: https://doi.org/10.1016/j.array.2022.100157

[5] Yichiet Aun1, Ming-Lee Gan1 , Nur Haliza Binti Abdul Wahab and Goh Hock Guan, “Social Engineering Attack Classifications on Social Media Using Deep Learning”, 2022, DOI: https://doi.org/10.32604/cmc.2023.032373

[6] Amit Kumar Sharmaa,∗ , Sandeep Chaurasiaa , Devesh Kumar Srivastava, “Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec”, 2020, DOI: http://creativecommons.org/licenses/by-nc-nd/4.0/

[7] Hanan M. Alghamdi, Saadia H.A. Hamza, Aisha M. Mashraqi and Sayed Abdel-Khalek, “Seeker Optimization with Deep Learning Enabled Sentiment Analysis on Social Media”, 2022, DOI: https://doi.org/10.32604/cmc.2022.031732