iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** I **Month of publication:** January 2025

DOI: https://doi.org/10.22214/ijraset.2025.66229

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

AI Driven Sentiment News Curation

Devaraj F. V.[1], Nagadeepa S. M.[2], Neha B. Nanjegowda[3], Rakshitha D. H.[4]
"Department of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering"

Abstract- This paper provides an in-depth review of sentiment analysis techniques applied in various domains, focusing on methodologies such as the VADER sentiment analysis model and Long Short-Term Memory (LSTM) networks. The survey discusses their respective advantages, including VADER's efficiency in handling real-time and news articles text and LSTM's ability to capture long-term dependencies in sequential data. Additionally, the paper explores the use of Bidirectional LSTM (Bi-LSTM) for improving sentiment classification accuracy and the Natural Language Toolkit (NLTK) for enabling diverse natural language processing tasks.

I. INTRODUCTION

Sentiment analysis plays a key role in understanding public opinion, especially in digital spaces like social media. It categorizes sentiments into positive, negative, or neutral, helping businesses and organizations gauge customer feedback and brand perception. Methods like VADER (Valence Aware Dictionary and Sentiment Reasoner) and LSTM (Long Short-Term Memory) networks have been widely used for sentiment classification. VADER is particularly effective in analyzing social media text, capturing both sentiment polarity and intensity, while LSTM networks address long-term dependencies in text. These models, tested on datasets like fetched news articles, demonstrate high accuracy and are critical for real-time sentiment analysis applications in industries like customer feedback monitoring and brand management. Sentiment analysis has also proven useful in analyzing political opinions, market trends, and societal issues through the lens of digital conversations. Further advancements in hybrid models and integration with deep learning techniques promise improved accuracy, enabling better insights into dynamic online conversations.

II. LITERATURE SURVEY

In this section, various authors have presented diverse methods and techniques for sentiment analysis across different domains.

In [1] Antony Samuels and John McGonical proposed a lexicon-based approach to analyze sentiments in news articles, classifying them as positive, negative, or neutral. Using a BBC News dataset of 2,225 articles, they used RapidMiner for preprocessing and WordNet for sentiment scoring. This method is efficient for large datasets and useful for public sentiment analysis and news categorization.

In [2] Dr. G. S. N. Murthy et al. used LSTM networks for sentiment analysis to handle long-term dependencies in text. They tested their model on the IMDB dataset and achieved 85% accuracy, proving its effectiveness in sentiment classification. They also suggested using advanced embeddings and larger datasets to further improve performance.

In [3] Dr. Manjula Bairam et al. discussed the importance of sentiment analysis in processing large-scale data. Their study reviewed key approaches, including machine learning, hybrid, and lexicon-based methods, and addressed challenges like corpus-based and dictionary- based text analysis. This research gives a comprehensive overview of the strengths and limitations of current sentiment analysis methods. The study also highlighted the growing need for multilingual sentiment analysis to cater to diverse global audiences. Furthermore, it emphasized the potential of combining advanced techniques to improve the accuracy and scalability of sentiment classification systems.

In [4] Greg Van Houdt et al. explored LSTM models, focusing on their ability to overcome issues like vanishing gradients and handle long-term dependencies. The study discussed improvements like Extended LSTM and LSNN, as well as hybrid models like CNN-LSTM for better performance in tasks like sentiment analysis and time-series prediction. The research also emphasized the importance of optimizing hyperparameters to improve the efficiency and accuracy of LSTM-based models. Additionally, it highlighted the potential of hybrid approaches to adapt to varying data complexities in real-world applications.

In [5] Mërgim H. Hoti and Jaumin Ajdari used the VADER model with a custom Albanian lexicon to analyze social media posts. Their study evaluated companies like Vala Telecommunication and Art Motion, achieving 89%-95% accuracy in sentiment classification. This work demonstrates how VADER can be adapted for localized and multilingual sentiment analysis.

In [6] Douglas C. Youva explored the VADER model, designed for analyzing social media text. The study explained how VADER uses lexicons, sentiment scoring, and rules to handle informal content like slang and emoticons.

©IJRASET: All Rights are Reserved

It highlighted the model's use in marketing and social media monitoring, showing its ability to work well with unstructured data. The research emphasized VADER's simplicity and ease of implementation compared to more complex deep learning models. It also demonstrated the model's scalability for analyzing large volumes of social media posts in real-time. Additionally, the study noted that VADER's rule-based approach performs effectively even in the absence of extensive training data.

In [7] J. Reshma et al. applied the VADER model to analyze news articles from inshorts.com, classifying them as positive, negative, or neutral. The study showcased VADER's speed and efficiency, proving it useful for real- time sentiment analysis across different industries. It also highlighted the model's ability to handle diverse vocabulary and short-form content commonly found in news snippets.

In [8] U. B. Mahadevaswamy and Swathi P. used a Bi-LSTM network for sentiment analysis on the Amazon Product Review dataset. Their model achieved 91.4% accuracy by considering both past and future contexts, making it effective for classifying reviews. They also pointed out the potential for further improvements, such as finer sentiment classification. The study emphasized the advantages of bidirectional architectures in capturing nuanced sentiment patterns in text.

In [9] Steven Bird and colleagues introduced the Natural Language Toolkit (NLTK), an open-source tool for learning computational linguistics. It provides modules and tutorials for tasks like parsing and morphological analysis, making NLP concepts easier to understand. This tool supports hands-on learning and continues to evolve with community contributions, encouraging community contributions. The study highlighted NLTK's role in bridging the gap between theoretical concepts and practical applications in NLP.

In [10] C.J. Hutto and Eric Gilbert introduced the VADER model for sentiment analysis of social media text. VADER uses a lexicon with valence scores to determine sentiment polarity and intensity. It performs better than similar models like LIWC in real-time scenarios and does not require labeled data, making it a fast and reliable option for sentiment analysis in various fields. The study also highlighted VADER's ability to handle informal language, including slang, emoticons, and punctuation, commonly found in social media posts.

Table 1: Summarization of various Authors

Authors	Title	Research Focus	Remarks
Antony Samuels and John McGonical[1], 2020	Sentiment Analysis of News Articles Using a Lexicon-Based Approach	Sentiment analysis of news articles by categorizing sentiments into positive, negative, or neutral using a lexicon-based approach.	This study utilized the BBC News dataset and highlighted the method's efficiency for large datasets. It provided insights into public sentiment and news categorization using a lexicon-based approach.
Dr. G. S. N. Murthy et al.[2], 2020	Sentiment Analysis Using LSTM Networks	The application of Long Short-Term Memory (LSTM) networks for sentiment analysis, specifically targeting long-term dependencies in text data.	The model achieved 85% accuracy on the IMDB dataset, demonstrating LSTM's effectiveness in sentiment classification. Future improvements aim to enhance the model's performance using advanced embeddings and larger datasets.
Dr. Manjula Bairam et al.[3], 2019	A Study of Sentiment Analysis: Concepts, Techniques, and Challenges	An exploration of sentiment analysis concepts, techniques, and challenges, with emphasis on machine learning, hybrid, and lexicon-based approaches.	The paper offers valuable insights into different sentiment classification methods and the challenges of sentiment analysis, highlighting the importance of corpus-based and dictionary-based lexicon approaches.

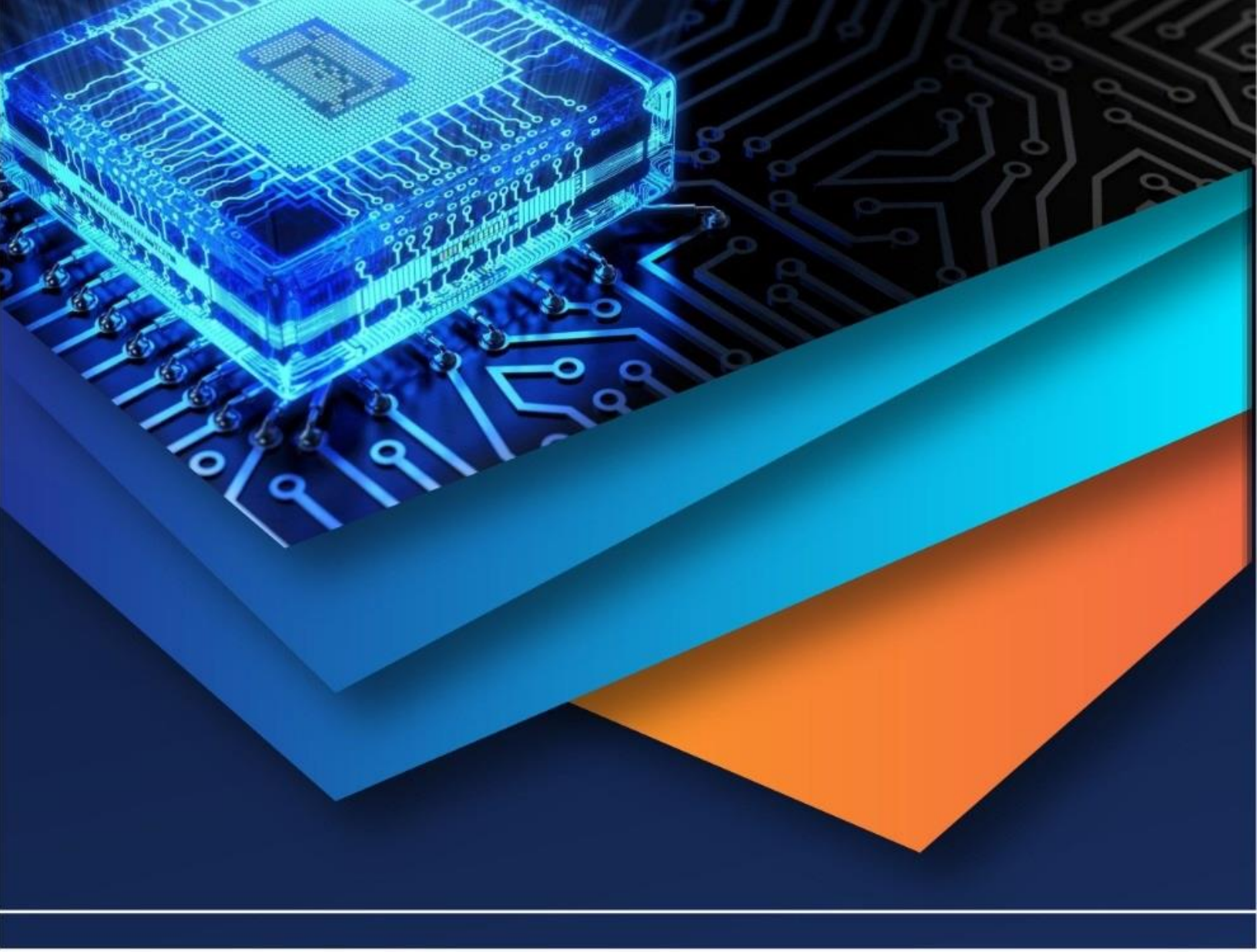
Greg Van Houdt et al.[4], 2019	A Review on the Long Short-Term Memory Model	A review of LSTM models and their application in sentiment analysis, emotion recognition, and time-series prediction.	The paper discusses advancements like Extended LSTM and LSNN and compares various architectures, emphasizing LSTM's relevance in solving temporal sequence problems.
Mërgim H. Hoti and Jaumin Ajdari[5], 2023	Sentiment Analysis Using the VADER Model for Assessing Company Services Based on Posts on Social Media	Evaluating customer satisfaction by analyzing social media comments using the VADER sentiment analysis model.	The study achieved high accuracy (89%-95%) in analyzing posts related to companies in Kosovo, emphasizing the model's effectiveness with comprehensive preprocessing steps.
Douglas C. Youva[6], 2024	Understanding Sentiment Analysis with VADER:A Comprehensive Overview and Application	A detailed exploration of the VADER sentiment analysis model, including its lexicon creation and application to social media text.	The paper underscores VADER's strengths in handling social media text, highlighting its simplicity, interpretability, and performance, especially in real-time applications.
J. Reshma et al.[7], 2022	Sentiment Analysis on News Articles	Sentiment analysis of news articles across various domains (sports, politics, etc.) using the VADER model.	The study demonstrated VADER's ability to classify sentiments accurately in large datasets of online news articles, emphasizing its suitability for real-time sentiment analysis in multiple industries.
U. B. Mahadevaswamy and Swathi P.[8], 2023	Sentiment Analysis Using Bidirectional LSTM Network	The use of Bi-LSTM networks for sentiment analysis on product reviews, specifically on the Amazon Product Review dataset.	The study demonstrated the effectiveness of Bi-LSTM in improving accuracy by capturing both past and future context in text, achieving an accuracy of 91.4%.
Steven Bird et al.[9], 2022	Natural Language Toolkit (NLTK)	A comprehensive open-source toolkit for teaching natural language processing (NLP) using both symbolic and statistical methods.	NLTK helps students develop programming skills and offers tools for various NLP tasks. It is continuously evolving, supporting complex NLP tasks and projects.
C.J. Hutto and Eric Gilbert[10], 2023	VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of social media Text	The VADER sentiment analysis model, designed for social media text, emphasizing sentiment polarity and intensity.	VADER outperforms other sentiment lexicons, especially in social media, and is suitable for real- time applications. Its simplicity and speed make it ideal for large-scale, dynamic environments.

III. CONCLUSIONS

The This paper presented a survey of sentiment analysis techniques, focusing on the use of VADER and LSTM models. These models have shown significant promise in classifying sentiments from various sources, such as social media, reviews, and news articles, into positive, negative, or neutral categories. The VADER model is particularly efficient for real-time social media analysis, capturing sentiment intensity, while LSTM networks excel at handling long-term dependencies in textual data, offering high accuracy in sentiment prediction. By reviewing the strengths and limitations of these models, this paper provides a comprehensive understanding of their applications in sentiment analysis. Future research could focus on refining these models, exploring hybrid approaches, and expanding their capabilities to handle.

REFERENCES

[1] Antony Samuels and John McGonical, "Sentiment Analysis of News Articles Using a Lexicon-Based Approach," BBC News Dataset Analysis, 2020. Available On: https://www.researchgate.net/search
[2] Dr. G. S. N. Murthy, K. Shalini, and S. Raghavendra, "Sentiment Analysis Using LSTM Networks," Proceedings of the 2020 International Conference on Machine Learning, 2020. Available On: https://www.researchgate.net/search
[3] Dr. Manjula Bairam, S. V. Dinesh, and S. Sharma, "A Study of Sentiment Analysis: Concepts, Techniques, and Challenges," Journal of Machine Learning and Applications, Vol. 12, Issue 4, 2019. Available On: https://www.researchgate.net/search
[4] Greg Van Houdt, M. C. Stan, and F. Q. Wang, "A Review on the Long Short-Term Memory Model," Neural Networks and Applications, Elsevier, 2019. Available On: https://www.researchgate.net/search
[5] Mërgim H. Hoti and Jaumin Ajdari, "Sentiment Analysis Using the VADER Model for Assessing Company Services Based on Posts on Social Media," Proceedings of the Social Media and AI Applications Symposium, 2023. Available On: https://www.researchgate.net/search
[6] Douglas C. Youva, "Understanding Sentiment Analysis with VADER: A Comprehensive Overview and Application," AI and Data Science Journal, 2024. Available On: https://www.researchgate.net/search
[7] J. Reshma, K. Meghana, and S. Ram, "Sentiment Analysis on News Articles," International Conference on Text Analytics and Natural Language Processing, 2022. Available On: https://www.researchgate.net/search
[8] U. B. Mahadevaswamy and Swathi P., "Sentiment Analysis Using Bidirectional LSTM Network," Proceedings of the Amazon Dataset Sentiment Analysis Challenge, 2023. Available On: https://www.researchgate.net/search
[9] Steven Bird, Ewan Klein, and Edward Loper, "Natural Language Toolkit (NLTK)," NLP Education and Tools, Elsevier, 2022. Available On: https://www.researchgate.net/search
[10] C. J. Hutto and Eric Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Proceedings of the 2023 Sentiment Analysis Symposium, 2023. Available On: https://www.researchgate.net/search



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)