User-Centric Sentiment Analysis of DeepSeek R1: Leveraging Google Play Store Reviews through Feature-Rich Machine Learning Models

Prof. Simran Thakur
Department of CSE
Chandigarh University
Mohali,India
thakursimi392@gmail.com.

Prof. Anuradha Devi
Department of CSE
Chandigarh University
Mohali,India
anuradha.e17284@cumail.in

*Prof. Priyanka Kumari
Department of CSE
Chandigarh University
Mohali,India
priyankapatiyal578@gmail.com

* Corresponding author

doi: https://doi.org/10.21467/proceedings.7.6.28

Abstract

In the modern era, the rapid development of AI-based models has made sentiment analysis an essential tool for understanding user perspectives. One such model, Deep Seek (RI version), has gained significant importance in natural language processing tasks. This study conducts a comprehensive sentiment analysis on Deep Seek R1 using four machine learning models—Logistic Regression, Random Forest, Support Vector Machine (SVM), and Naïve Bayes—along with two lexicon-based tools, VADER and Text Blob. Two feature extraction methods, Count Vectorizer and Term Frequency-Inverse Document Frequency (TF-IDF), were employed to evaluate the performance of the machine learning models. The results indicate that Logistic Regression achieves the highest accuracy when using Count Vectorizer, while SVM outperforms other models when using TF-IDF. However, among all approaches, VADER sentiment analysis achieves the highest accuracy at 99%, whereas Text Blob records the lowest accuracy. These findings highlight the effectiveness of lexicon-based approaches in sentiment classification while also demonstrating the impact of feature extraction techniques on machine learning model performance.

Keywords: Deep seek R1, ChatGPT, sentiment analysis.

1. Introduction

1.1 Artificial Intelligence

The modern era leads to different technical developments in Artificial Intelligence which has spawned multiple generative AI models including ChatGPT alongside Deep Seek and Gemini while these programs employ expansive language models for decoding and creating human language to accomplish various text-related operations together with interpretation and Question ,Answer systems[1]. This platform produces textual Contents including articles and blog posts for educational purposes as well as research activities while creating codes through spoken instructions. The software solution provides these comprehensive features to boost human creativity and enhance decision-making abilities. Companies across the market launched considerable generative ai model developments during the past years which led to intensive competition in this field. Also a distinguishing characteristic exists in every developed model which make all model unique from one another.

1.2 Deep Seek

Deep seek stands as one of the deep models that drew significant attention because it offers improved reasoning features and an open-source design framework. By using this model users can obtain instant relevant information in every response which integrates web search analysis beyond basic chatbot functions. The principal value of this model arises from its capability to retrieve updated internet data which becomes integrated into its response system. The system draws from real-time data so it can provide detailed explanations with appropriate up-to-date information to strengthen its accurate and relevant answers. One more standout functionality of this tool includes its ability to perform multiple tasks including the file upload system for text extraction which extends its practical value. The system lets people request document assessment after uploading files which enables Deep Seek to extract and analyse content and compare it with fresh information sources for deeper results. All of its functionalities make this tool more powerful and unique from other models. The Chinese startup developed this advanced ai platform in Dec 2023 yet scholars have paid less attention to it compared to its commercial competition. Academic research about ai usability development concentrates on the subscription and general-purpose systems like ChatGPT but ignores accessible domain-specific open-source models despite their importance in user analysis perspectives. Our

Proceedings DOI: 10.21467/proceedings.7.6; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-989164-9-5

study performed sentiment analysis on deep seek R1 version which is one of the famous ai generative model. We analyse it's reviews to found the user perception about this model are they liking this or not within the google play store platform. As this platform empowers users to exchange their application experience and rate them also. Alongside offering essential feedback to developers who aim to enhance their products[2]. Understanding user sentiment about specific apps represents an essential requirement because any new application serves only users so without positive feedback there exists no app advantage therefore, we perform sentiment analysis within in our study. Sentiment analysis is a subtle sentiment category labelling task [3].It centers on identifying the polarity of sentiment of aspect terms within a sentence. Itis a significant task within the area of Natural Language Processing (NLP). SA ordinarily takes on a general structure which encompasses gathering raw data, pre-processing the gathered data to eliminate noise, converting the preprocessed data to computational appropriate form, referring to training data as "labeling."[4]. Also, as in general with the emergence of multimedia technology, artificial intelligence is slowly moving into the multimodality era [5]. The main approach of conventional sentiment models limits its analysis to assign texts into three distinct categories such as positive, neutral and negative. This study conducts a review analysis of Deep Seek R1 equipment and delivers multiple important findings:

- The initial analysis was run through five emotional sentiment groups including Strongly Negative and Slightly Negative as well as Neutral and Slightly Positive and Strongly Positive.
- The project examines Google Play Store review classification through Deep Seek with different feature extraction approaches before performing a comparison analysis.
- Our team assessed numerous machines learning model through a process of training and testing for classification evaluation.
- The investigated results will enhance understanding of user feelings about this generative AI model.

2. Literature review

Author Name	Year	Models and methods Used	Inference
Abdur et al.[6]	2025	Deep Seek and Chatbots.	The paper proposes an Emotion-Aware Embedding Fusion framework to enhance LLMs' empathy and contextual comprehension in psychotherapy chatbots. By integrating emotion lexicons, attention mechanisms, and hierarchical fusion, the system outperformed baseline models in generating emotionally intelligent responses with enhanced empathy, coherence, fluency, and informativeness.
Souza et al.[7]	2025	Grok, Gemini, ChatGPT, Deep Seek	The article provides a comparative overview of four large conversational AI models—Grok, Gemini, ChatGPT, and Deep Seek—and their architectures, strengths, and areas of use. It draws attention to the special capabilities of each model (e.g., multimodality of Gemini, versatility of ChatGPT, research orientation of Deep Seek) and addresses problems such as bias and hallucination.
Krishnaveni et al.[8]	2025	Chat GPT and Deep Seek.	The authors evaluate reddit user 0emotions regarding LLM models ChatGPT and Deep Seek through negative, neutral and positive categorizations of user responses. The study explores issues about trust and AI biases together with misuse potential and societal implications.
Rahul et al.[9]	2025	Deep Seek vs Chat GPT	A study conducted here compares the two models regarding their structural design platform alongside execution abilities and real-world examples of project developments shows superior performance of Chat GPT for interactive duties. The deep seeker system delivers better efficiency during real-time queries and fact verification and financial datasets which make it suitable for academic fields and financial operations. and other data-driven applications.

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

Anichur et al.[10]	2025	Deep Seek, ChatGPT, Gemini	Three AI models including DeepSeek and ChatGPT and Google Gemini underwent evaluation through research papers for their specific strengths and potential applications. Through these technologies the authors foretold substantial AI advancements in various business fields.
Zonghao et al.[11]	2025	Deep Seek LLMs Deep seek MLLM Deep seek T21	Deep seek models require safety measures according to this paper since they demonstrate substantial security weaknesses which include jailbreaking susceptibility and unsafe content production and chain of Thought safety problems.
Sarah et al.[12]	2025	Deep Seek V3, Deep Seek R1,	The research focuses on Deep seeks R1 model as an important developmental achievement within generative AI by using deep seek v3 as a base model.

3. METHODOLOGY

3.1. Data Collection - The research data for user reviews taken from 15435 feedback entries found on the Kaggle platform under the Google play store. We gathered data from Google play store review platform because it serves as a top distribution system that permits users to look for and obtain apps and games and provide feedback about their experiences.

3.2. Data Cleaning and Preprocessing -

3.3. Sentiment analysis Classification: After pre-processing the data, we carried out sentiment analysis to identify the polarity of each feedback. We used the Sentiment Intensity Analyzer from the NLTK package to do this. The sentiment polarity scores for each feedback, including compound polarity, neutral, slightly negative, strongly negative Strongly positive and slightly positive, were generated by using this analyzer. Based on these polarity ratings, we applied sentiment labels to establish sentiment kinds. In order to understand the public's perception about Deep Seek, emphasize the importance of this categorization. The appeal of sentiment analysis classification is that it can convert the complex web of human expression found in reviews into actionable knowledge. It enables us to give emotional context to specific statements and viewpoints, turning abstract text into quantifiable sentiment scores. Additionally, if a company or organization wants to respond to consumer feedback and public sentiment, sentiment analysis classification is a crucial tool. It can influence strategic decisions and provides insightful information about user experiences. For example, reviews with a sentiment _compound _polarity value is less than or equal to -0.5 were labelled as " STRONGLY NEGATIVE," score between 0 and -0.5 are considered "SLIGHTLY NEGATIVE," score of exactly 0 signifies a "NEUTRAL". When the compound score reaches a value above 0.5 the system indicates "STRONGLY POSITIVE" to represent maximal positive sentiment. And if scores lie between 0 and 0.5 are classified as "SLIGHTLY POSITIVE" to indicate their mild positive. After conducting sentiment analysis, we gather the number of neutral, slightly negative, strongly negative, strongly positive and slightly positive, review.

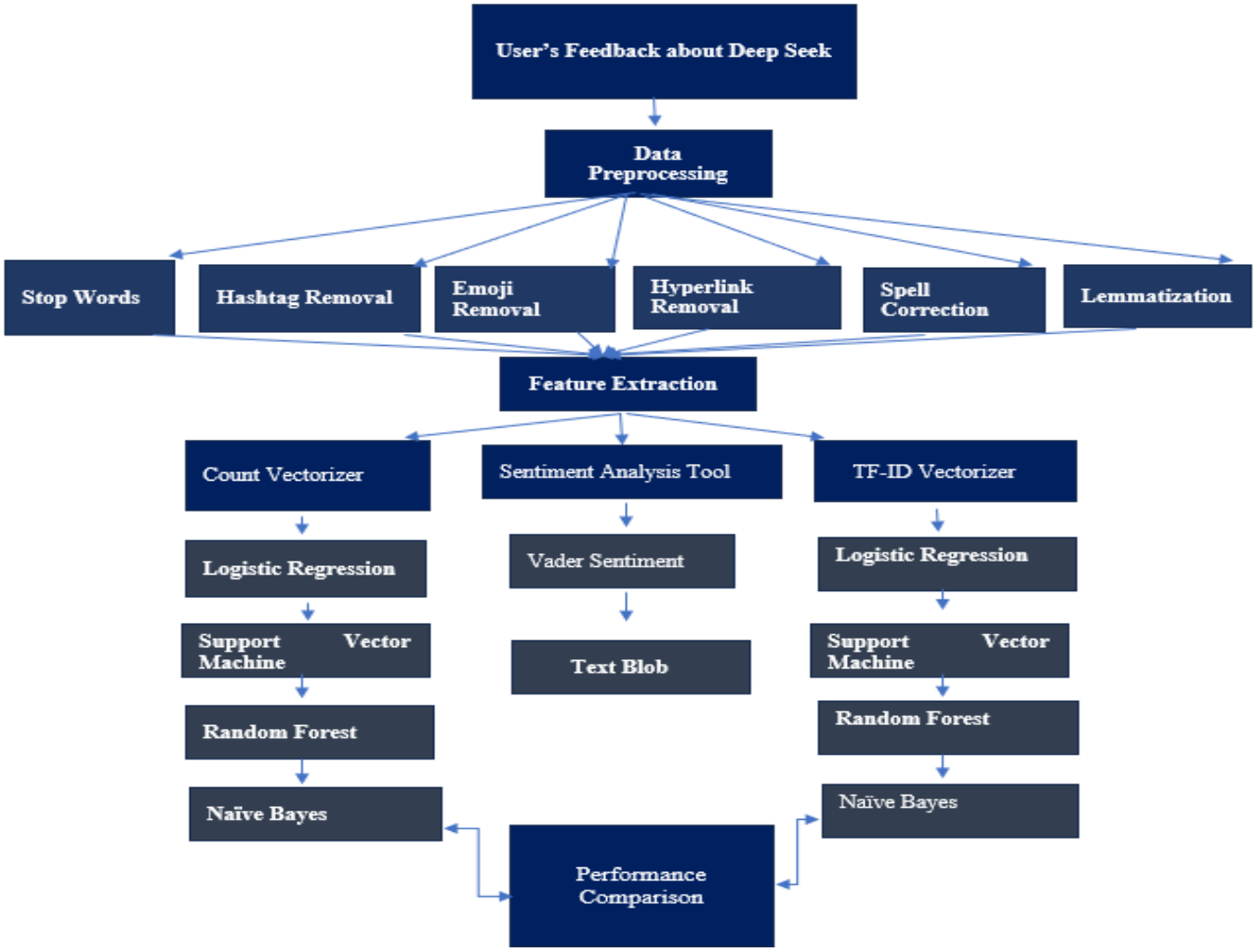
3.4. Feature Extraction**:** The machine learning model applied for analyzing Google Play Store reviews needs numerical data to classify instead of textual information. Therefore, we selected various features for text representation during our analysis.

3.4.1 Count Vectorizer: The Count Vectorizer class transforms textual documents into word frequency matrices which display the word occurrences in original text.

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

3.4.2 **Term Frequency**: The method uses TF-IDF scoring for corpus words to assign higher values to words that provide unique illustration for specific documents



.

Fig.1. Proposed Methodology

4. Sentiment Analysis Model*:* The reviewed sentiments evaluated by machine learning models working together with natural language processing to analyze sentiment. The choice of modeling method was Logistic Regression: An algorithm which performs binary classification tasks is the most frequently applied algorithm. Naïve Bayes: The probabilistic text classification tool uses a Baye's theorem model to operate under an assumption of feature independence. Support Vector machine: Achieves outstanding results in classifying multiple text groups while also being valuable for sentiment analysis and topic tagging and spam filtering through its ability to locate the optimal hyperplane between different text classes Random Forest: It is aa statistical model used for both classification and regression purposes together with sentiment analysis applications. During training Random Forest constructs numerous "forest" trees that generate predictions which become combined into final outputs by averaging in regression tasks and through collective voting in classification tasks. Vader Sentiment: Stands as a rule-based sentiment analysis solution. It applies a sentiment score database containing predetermined words and operates with specific rules to interpret strength modifiers that include "very," "not," and "extremely." It shows excellent ability to analyze emotions in brief informal texts including text with sarcasm and slang together with emoticons which makes it ideal for real-time user-generated content assessment. Text Blob. It functions as a text processing library which constructs its structure upon NLTK and Pattern. Through its easy-to-use API Text Blob enables users to complete basic NLP operations such as speech tagging and identification of noun phrases and sentiment assessment. Text Blob employs a multifunctional lexicon that functions effectively with formal documents including news reports, academic papers and scholarly publications and essays.

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

4. RESULT

The experiment code was run on Google Collab because its cloud environment optimizes model training and evaluation processes. Machine learning models built through Kera's within TensorFlow received their data manipulation through Pandas. Our research examined a sentiment classification task that comprised five distinct polarity categories between strongly positive to strongly negative sentiment along with slightly positive and neutral sentiments. The model required numerical sentiment values for effective training and evaluation processes. The assessment of model performance occurred through evaluation of F1 Score along with Precision , Recall and Accuracy scores.

4.1 .Results Using Counter Vectorizer :

Among the models evaluated, Logistic Regression demonstrated the highest performance, achieving an accuracy of approximately 86%, with precision, recall, and F1-score all consistently around 85-86%. This indicates that Logistic Regression maintains a strong balance between correctly classified positive instances (precision) and its ability to identify all relevant instances (recall), making it the most reliable and well-rounded model. The Support Vector Machine (SVM) performed slightly below Logistic Regression, with accuracy, precision, and recall values ranging between 81-82%. While still a strong performer, it falls short of Logistic Regression in maintaining an optimal balance between precision and recall, as evidenced by its marginally lower F1-score. Similarly, Random Forest exhibited performance close to SVM, with scores between 80-81% across all four metrics. The slight reduction in F1-score suggests potential inconsistencies in handling imbalanced data or minor variations in classification performance. In contrast, Naïve Bayes emerged as the weakest model, with accuracy and precision around 68-69%, recall close to 70%, and F1-score at 66%. These lower scores indicate that Naïve Bayes struggles in balancing precision and recall, which is a common limitation due to its strong assumption of feature independence. The reduced F1-score highlights its inefficiency in scenarios where both precision and recall need to be optimized. Based on the comparative analysis, Logistic Regression is the most effective model, offering the highest accuracy and balanced performance across all metrics. Naïve Bayes, on the other hand, is the least effective model, as it fails to maintain an optimal trade-off between precision and recall. These results suggest that for datasets with similar characteristics, Logistic Regression would be the most suitable choice for achieving high classification performance, while Naïve Bayes should be used with caution, especially in cases where feature independence assumptions do not hold.

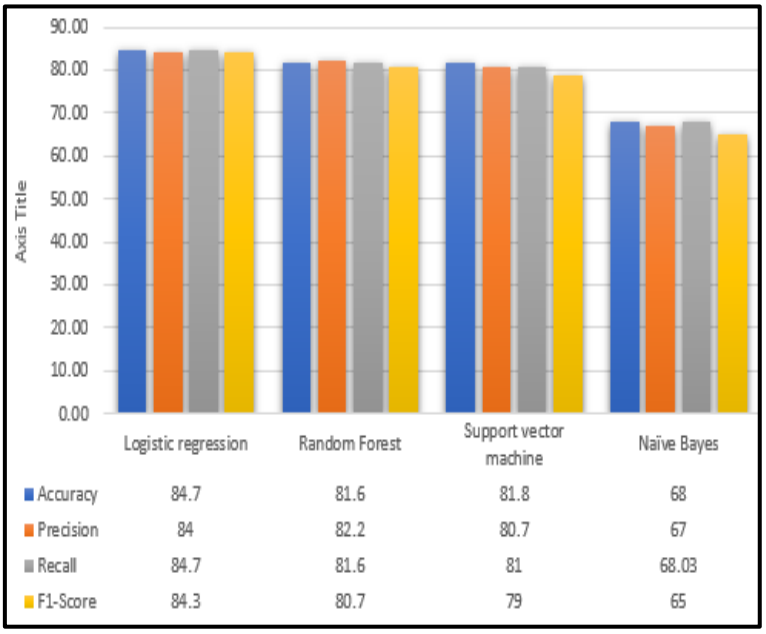


Fig. 2. Bar plot using the count-vectorizer feature extraction

4.2. Results Using TF-IDF Vectorizer:

In the comparison of machine learning models based on Accuracy, Precision, Recall, and F1-Score, the Support Vector Machine (SVM) demonstrated the highest performance, achieving an accuracy of approximately 83%, along with consistently high values across all other metrics. Logistic Regression and Random Forest also performed well, both achieving an accuracy of around 81%, with minor variations in their Precision, Recall, and F1-Score. In contrast, the Naïve Bayes model exhibited the lowest performance, with an accuracy of approximately 64% and a

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

noticeably lower F1-Score, indicating weaker predictive capabilities. This suggests that while SVM is the most reliable model for this dataset, Logistic Regression and Random Forest remain competitive options, whereas Naïve Bayes may not be suitable for this specific classification task due to its lower overall performance.

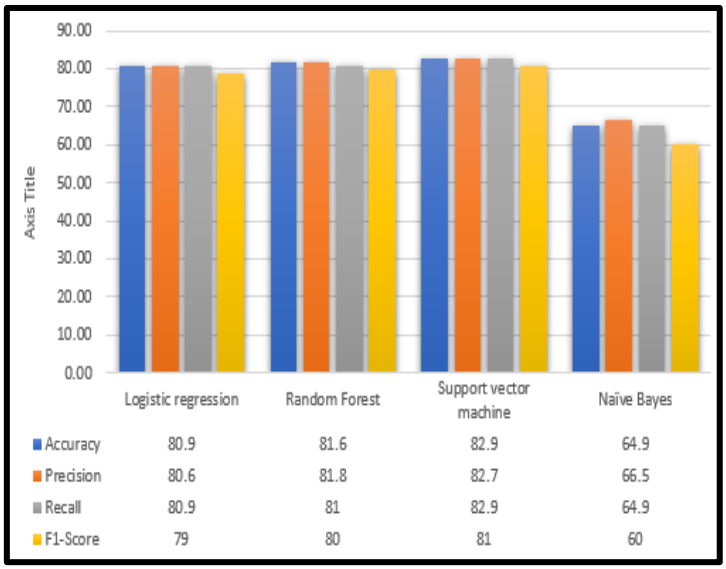


Fig. 3. Bar plot using the TF-IDF-vectorizer feature extraction

4.3. Using Vendor sentiment analyses and text blob:

Based on the evaluation metrics, the model that performs better would be determined by comparing the accuracy and classification reports of both VADER and Text Blob. Typically, VADER tends to outperform Text Blob in sentiment analysis, especially for social media text or short content, due to its ability to handle nuances like emoticons, slang, and context-specific sentiment expressions. VADER's compound score calculation is more refined for detecting sentiment intensity, which often results in higher accuracy and better F1-scores across different sentiment classes. VADER's attains an accuracy level of 99% despite surpassing the results obtained in several previous research projects this is because we have applying preprocessing techniques to our dataset serve as the main reason for achieving high accuracy. Our methodology included stop word removal followed by lemmatization then lowercasing and removal of special characters and symbols that had no relevance. The processing steps effectively minimized text noise which enabled VADER to function with normalized standardized input dataset. On the other hand, Text Blob, while effective for general-purpose text, may struggle with the subtleties of complex or informal language, leading to slightly lower accuracy and less precision in certain classes. However, the final verdict depends on the specific dataset; if the dataset contains highly contextual or domain-specific language, the performance difference may vary.

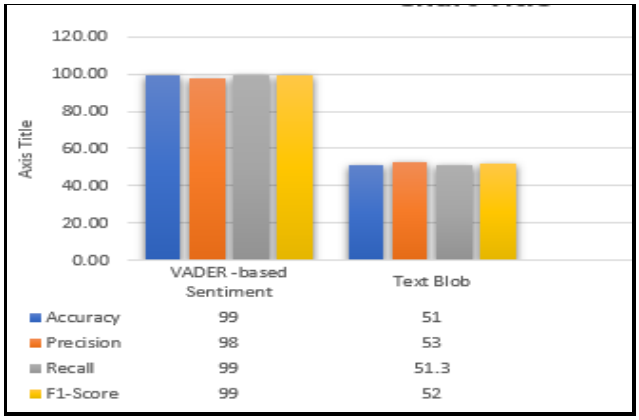


Fig. 4. Bar plot using the Sentiments tool

5. CONCLUSION AND FUTURE SCOPE

In our research we evaluated user feedback from Google Play Store about the Deep Seek application formed the focus of this examination to assess overall user attitudes which might be either positive or negative or neutral. with Logistic Regression and Random Forest algorithms and SVM classifiers and Naive Bayes systems through VADER

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

analysis, Text Blob tools and Count Vectorizer and TF-IDF Vectorizer methods. The research demonstrated that Count Vectorizer worked best with Logistic Regression although Naive Bayes generated substandard results using both extraction techniques. In TF-IDF context SVM managed better performance than its competitors yet Naive Bayes demonstrated no effective results. The highest accuracy rate was achieved by VADER which proved its effectiveness in sentiment analysis over Text Blob that showed inferior results. The research demonstrates that VADER proves to be an effective algorithm for sentiment analysis projects particularly when working with user-made content. In Future researchers can expand this work by applying deep learning methods which include Recurrent Neural Networks (RNNs) or Transformers to extract advanced patterns from user feedback. Real-time sentiment analysis should be included because it generates immediate perspectives from user opinions. Additional reviews extracted from different platforms together with domain-specific model training will boost the system's accuracy levels. Business strategies gain value from sentiment analysis outcomes which enables the development of enhanced app features that produce better user satisfaction.

REFERENCES

[1] W. Feng, Y. Li, C. Ma, and L. Yu, "From ChatGPT to Sora: Analyzing public opinions and attitudes on generative artificial intelligence in social media," *IEEE Access*, Jan. 2025.

[2] J. Yu and C. Qi, "Machine learning-based sentiment analysis in English literature: Using deep learning models to analyze emotional and thematic content in texts," *IEEE Access*, Mar. 2025.

[3] Y. Yang, X. Dong, Y. Qiang, and W. Si, "SKE-MSA: Enhancing representation learning with VAD lexicon for multimodal sentiment analysis," in *ICASSP 2025 – 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, 2025,

[4] S. Velampalli, C. Muniyappa, and A. Saxena, "Performance evaluation of sentiment analysis on text and emoji data using end-to-end, transfer learning, distributed and explainable AI models," *arXiv preprint arXiv:2502.13278*, Feb. 2025.

[5] P. Yu, J. Gu, D. Pi, Q. Zhou, and Q. Wang, "Aspect-aware graph interaction attention network for aspect category sentiment analysis," *IEEE Transactions on Emerging Topics in Computational Intelligence*, Jan. 2025.

[6] A. Rasool, M. I. Shahzad, H. Aslam, V. Chan, and M. A. Arshad, "Emotion-aware embedding fusion in large language models (Flan-T5, Llama 2, DeepSeek-R1, and ChatGPT 4) for intelligent response generation," *AI*, vol. 6, no. 3, pp. 56, Mar. 2025.

[7] M. E. de Carvalho Souza and L. Weigang, "Grok, Gemini, ChatGPT and DeepSeek: Comparison and applications in conversational artificial intelligence," *Inteligencia Artificial*, vol. 2, pp. 1, 2025.

[8] K. Katta, "Analyzing user perceptions of large language models (LLMs) on Reddit: Sentiment and topic modeling of ChatGPT and DeepSeek discussions," *arXiv preprint arXiv:2502.18513*, Feb. 2025.

[9] R. V. Dandage, "A comparative analysis of ChatGPT and DeepSeek: Capabilities, applications, and future directions," 2025.

[10] A. Rahman, S. H. Mahir, M. T. Tashrif, A. A. Aishi, M. A. Karim, D. Kundu, T. Debnath, M. A. Moududi, and M. D. Eidmum, "Comparative analysis based on DeepSeek, ChatGPT, and Google Gemini: Features, techniques, performance, future prospects," *arXiv preprint arXiv:2503.04783*, Feb. 2025.

[11] Z. Ying, G. Zheng, Y. Huang, D. Zhang, W. Zhang, Q. Zou, A. Liu, X. Liu, and D. Tao, "Towards understanding the safety boundaries of DeepSeek models: Evaluation and findings," *arXiv preprint arXiv:2503.15092*, Mar. 2025.

[12] S. Mercer, S. Spillard, and D. P. Martin, "Brief analysis of DeepSeek R1 and its implications for generative AI," *arXiv preprint arXiv:2502.02523*, Feb. 2025.

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922