Journal of Modern Technology

Vol. 02, No. 01, March 2025, pp. 212-219.
DOI: https://doi.org/10.71426/jmt.v2.i1.pp212-219 , E-ISSN: 3049-3102

Homepage: https://review.journal-of-modern-technology.com/index.php/jmt/index

Leveraging Sentiment Analysis in the Digital Era: Uncovering Insights from Unstructured Data for Enhanced Customer Engagement

S Rama Satyanarayana Reddy [1], G Neha Akshaya [2], O Lakshmi Koteswari [3], T Sreeja [4], Venkata Sandeep Edara [5*]

[1,2,3,4,5*] Department of Computer Science and Engineering, Sasi Institute of Technology and Engineering, Tadepalligudem 534 101, Andhra Pradesh, India.

Email IDs: satyanarayana.satti@sasi.ac.in [1] , akshaya.ganta@sasi.ac.in [2], koteswari.omkarapu@sasi.ac.in [3] sreeja.torlapati@sasi.ac.in [4] , evsandeep@sasi.ac.in [5*]

ORCID: https://orcid.org/0009-0004-4321-062X [1] , https://orcid.org/0009-0001-9050-571X [2] , https://orcid.org/0009-0000-4902-0887 [3], https://orcid.org/0009-0000-2509-6932 [4] , https://orcid.org/0009-0004-6318-6735 [5*]

Article Info

Article history:

Received: March 01, 2025
Revised: March 15, 2025
Accepted: March 17, 2025
First Online: March 18, 2025
Final Online: April 20, 2025

Keywords:

Google Play Store
Sentiment Analysis
User Reviews
App Ratings
Deep Learning
Long Short-Term Memory

ABSTRACT

The Google Play Store is a dynamic marketplace hosting a vast array of mobile applications across various categories. Analyzing user ratings and sentiments is essential for developers, marketers, and researchers to evaluate app performance and enhance user satisfaction. This study employs deep learning techniques, specifically a Long Short-Term Memory (LSTM)-based model, to examine user reviews and identify sentiment patterns. By leveraging natural language processing (NLP) and machine learning, the research investigates correlations between user feedback, app features, and overall ratings. The model processes and classifies user sentiments, such as positive, neutral, or negative and provides insights into key factors influencing user perceptions. Additionally, this study explores how app quality, functionality, and user engagement impact consumer satisfaction. Through data-driven analysis, it highlights the primary drivers of positive and negative reviews, offering a comprehensive understanding of user expectations and industry trends.

Corresponding Author:
E-mail address: evsandeep@sasi.ac.in (Venkata Sandeep Edara)

1. INTRODUCTION

The rapid expansion of mobile applications has significantly transformed the digital landscape, with the Google Play Store serving as a central platform for distributing Android applications [1]-[3]. This marketplace hosts a diverse range of apps, catering to various user needs across multiple categories. Developers and businesses leverage this platform to reach global audiences, while users rely on it to explore, install, and provide feedback on applications [4]-[6]. User ratings and reviews serve as critical indicators of app quality, functionality, and overall user satisfaction.

Analyzing these reviews is essential for understanding user preferences, identifying potential issues, and improving app development strategies to enhance user experience [7]-[10].

Sentiment analysis of user reviews has emerged as a vital tool for extracting meaningful insights from vast amounts of textual data. By employing natural language processing (NLP) and machine learning techniques, researchers can systematically classify sentiments into positive, negative, or neutral categories [2]. This classification aids developers in addressing user concerns, refining app features, and optimizing user engagement. Deep learning approaches, particularly the LSTM networks, have demonstrated superior performance in sentiment classification by capturing complex linguistic patterns and contextual dependencies [3]. Compared to traditional machine learning models such as Naïve Bayes and Support Vector Machines (SVM), deep learning methods offer enhanced accuracy in understanding sentiment dynamics.

Beyond sentiment classification, predictive modeling of app ratings has gained prominence in mobile app analytics. Various attributes, including app category, update frequency, and user engagement metrics, influence rating trends. Advanced ensemble techniques, such as Random Forest and Boosting, have been employed to improve the robustness and accuracy of rating predictions [4]. However, deep learning models, particularly recurrent neural networks (RNNs) and transformer-based architectures, have shown promise in capturing nuanced relationships within user feedback. By integrating sentiment analysis with predictive modeling, researchers can provide actionable insights to developers, helping them enhance app quality, user retention, and market competitiveness.

This study utilizes an LSTM-based deep learning model to perform sentiment analysis on Google Play Store reviews, offering a data-driven approach to understanding user feedback. By leveraging NLP techniques and machine learning, the research aims to identify key sentiment trends, assess the impact of app features on ratings, and provide valuable recommendations for app optimization. The findings contribute to the broader field of mobile application analytics, supporting developers in refining their applications to better meet user expectations and industry standards.

2. LITERATURE REVIEW

Numerous studies have investigated sentiment analysis techniques and predictive modeling for analyzing user reviews on the Google Play Store. Mustofa and Idris [1] explored ensemble learning approaches, comparing Random Forest and Boosting techniques against individual classifiers such as Naïve Bayes and SVM. Their findings suggest that ensemble models significantly improve classification accuracy, making them particularly effective for large-scale sentiment analysis in mobile applications. Similarly, Fahim et al. [2] applied natural language processing (NLP) techniques to sentiment analysis, emphasizing the importance of data preprocessing and visualization in deriving meaningful insights from user feedback. Their research highlights the role of structured data preparation in enhancing the accuracy of sentiment classification models.

The authors of work [3] conducted an exploratory data analysis (EDA) of app ratings, identifying key factors such as app category, update frequency, and sentiment polarity as significant contributors to overall app performance. Predictive modeling techniques, including regression and decision trees, further validated these findings by establishing correlations between user sentiments and app success. In another study, Putra et al. [4] compared various classification algorithms for sentiment analysis and found that SVM with a linear kernel achieved the highest accuracy (95.23%) in distinguishing between different sentiment categories. These findings align with Santoso et al. [5], who demonstrated that a stacking ensemble approach enhances sentiment classification performance, achieving an accuracy of 87.05%. The combined results from these studies underscore the advantages of integrating multiple classifiers over standalone models for improved sentiment analysis outcomes.

Further research has examined the predictive modeling of app ratings using machine learning techniques. Gopi et al. [6] developed models based on regression, decision trees, and gradient boosting to estimate app ratings based on metadata attributes such as category, size, and user reviews. Their findings indicate that machine learning algorithms can effectively predict app ratings, aiding developers in understanding user preferences. Similarly, in [7]-[8] utilized the VADER sentiment analysis tool to categorize user reviews into positive, negative, or neutral sentiments. This approach offers app developers valuable insights into consumer feedback trends, facilitating data-driven improvements in app development and user engagement strategies. Sentiment analysis has also been employed to identify app deficiencies and potential areas for feature enhancements. Yasin et al. [7] demonstrated the effectiveness of mining user feedback to uncover usability issues and inform future app updates. The work presented in [8]-[9] explored the predictive relationships between app attributes and user ratings, identifying factors such as install count, price, and update frequency as key determinants of app success. Their study reinforces the utility of machine learning models, including Random Forest, SVM, and K-Nearest Neighbors (KNN), in predicting app performance based on structured metadata and user sentiment. Collectively, these studies highlight the growing role of sentiment analysis and predictive modeling in refining mobile application development and user experience strategies.

Recent advancements in deep learning have further improved sentiment analysis methodologies. The authors of ref [14] have introduced a lexicon-based approach, incorporating domain-specific features such as specialized terminologies and star ratings. By integrating SentiWordNet for sentiment classification, their research demonstrated the effectiveness of combining structured rating data with sentiment analysis. Meanwhile, deep learning-based approaches, such as those discussed in [15], leverage SenticNet to enhance sentiment classification by integrating logical reasoning with deep learning architectures. Additionally, The other researchers have developed an attention-based sentiment analysis model that combines CNNs and RNNs for improved feature extraction and sequential learning [16]-[17]. These studies underscore the superiority of deep learning approaches in capturing nuanced sentiment variations, outperforming traditional machine learning techniques in both accuracy and efficiency.

Beyond mobile applications, sentiment analysis has been widely applied in various domains, including healthcare, finance, and education [17]-[21]. The authors of ref [18] have discussed the sentiment analysis in healthcare, where supervised learning models exhibited superior performance over lexicon-based methods, albeit requiring extensive data preprocessing. In the financial sector, In [19], the authors have demonstrated that sentiment analysis combined with time-series modeling could enhance stock market predictions, though challenges such as susceptibility to rumor-driven fluctuations remain. Similarly, in [20] – [21], the sentiment analysis in the education sector to evaluate MOOCs, integrating lexicon-based and unsupervised learning techniques to assess instructional quality. These studies highlight the need for hybrid sentiment analysis models that integrate deep learning with traditional approaches to optimize real-world applications. As sentiment analysis continues to evolve, addressing challenges such as data imbalance, improving feature extraction, and incorporating contextual understanding will be crucial in advancing the accuracy and robustness of sentiment classification models.

3. METHODOLOGY

3.1. Data collection and Preprocessing

The dataset consists of two files: apps.csv, which contains metadata about various applications on the Google Play Store, and user_reviews.csv, which contains 100 reviews per app, along with pre-processed sentiment labels (Positive, Neutral, Negative), Sentiment Polarity, and Sentiment Subjectivity. Since our primary focus is on sentiment classification, we extract only relevant columns—'Translated_Review' and 'Sentiment'—from the user_reviews.csv dataset. To ensure data integrity, we remove missing values, which helps in preventing biased learning due to incomplete data.

The next step is text preprocessing, which involves cleaning and transforming raw text into a structured format for deep learning. We perform the following operations:

(i) Convert all text to lowercase to ensure uniformity.
(ii) Removing numbers and punctuation: These often do not contribute to sentiment analysis.
(iii) Convert sentences into sequences of words.
(iv) Remove frequently occurring but semantically insignificant words.
(v) Convert sequences to a fixed length of 100 words.

Mathematically, a given review R is transformed into a sequence of words (1).

$$R = \{w_1, w_2, \dots \dots \dots w_n\} \quad (1)$$

In (1), the w_i represents a word in the review. The tokenization process maps words to unique integer indices, forming an indexed sequence:

$$T = \{x_1, x_2, \dots \dots \dots x_n\}, x_i \in Z^+ \quad (2)$$

In (2), the x_i is the unique tokenized representation of word w_i. These sequences are then padded to a fixed length L.

$$P(T) = \{x_1, x_2, \dots \dots \dots x_L\},\ if\ n < L, pad\ with\ 0 \quad (3)$$

The sentiment labels (Positive, Neutral, Negative) are numerically encoded using one-hot encoding and which is given by (4). Finally, the dataset is split into training (80%) and validation (20%) sets using stratified sampling, ensuring class balance.

$$y = \begin{cases} (1,0,0) & if\ Positive \\ (0,1,0) & if\ Neutral \\ (0,0,1) & if\ Negative \end{cases} \quad (4)$$

3. 2. Model Architecture and Training

For sentiment classification, we use a LSTM network, a type of RNN that effectively captures long-range dependencies in text sequences. The model comprises the following layers as shown in Figure 1.

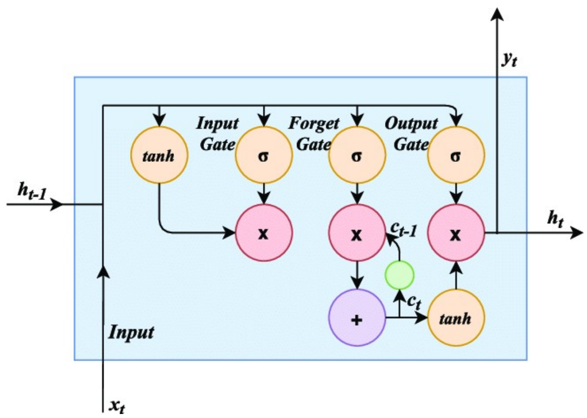


Figure 1. Architecture of LSTM model.

1. Embedding Layer: Maps each word to a dense vector representation of dimension d, using pre-trained word embeddings or randomly initialized embeddings. If V is the vocabulary size, the embedding matrix E is of shape is (5).

$$E \in R^{V \times d} \quad (5)$$

Given a tokenized input sequence $x = \{x_1, x_2, \dots \dots \dots x_L\}$, the embedding layer outputs (6).

$$e_i = E[x_i] \quad (6)$$

In (6), the $e_i \epsilon R^d$ is the word embedding of token x_i.

2. LSTM Layer: Processes the sequence of embeddings and captures contextual dependencies. The LSTM unit at time step t consists of:

   A. *Forget Gate:*

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \quad (7)$$

   B. *Input Gate:*

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (8)$$

C. *Candidate Memory Cell:*

$$\tilde{C}_t = tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (9)$$

D. *Memory Cell Update:*

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \quad (10)$$

E. *Output Gate & Hidden State Update:*

$$O_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (11)$$

$$h_t = O_t \odot tanh(C_t) \quad (12)$$

In (11), the σ is the sigmoid activation function, W,U,b are learned parameters, and ⊙ represents element-wise multiplication.

3. *Dropout Layers:* Applied after each LSTM layer to mitigate overfitting by randomly disabling neuron connections during training.
4. *Fully Connected Dense Layer:* A final dense layer with a SoftMax activation function computes the probability of each sentiment class.

4. RESULTS AND DISCUSSION

The comparative analysis of sentiment classification models applied to Google Play Store app reviews demonstrates that deep learning, specifically the LSTM network, outperforms traditional machine learning models. The evaluation was based on accuracy, precision, recall, and F1-score, which were visualized using bar plots. The LSTM model achieved the highest accuracy, showcasing its ability to capture contextual dependencies in textual data. Traditional models such as Logistic Regression, SVM, and Random Forest also performed reasonably well, but their inability to retain sequential dependencies limited their effectiveness in nuanced sentiment classification tasks.

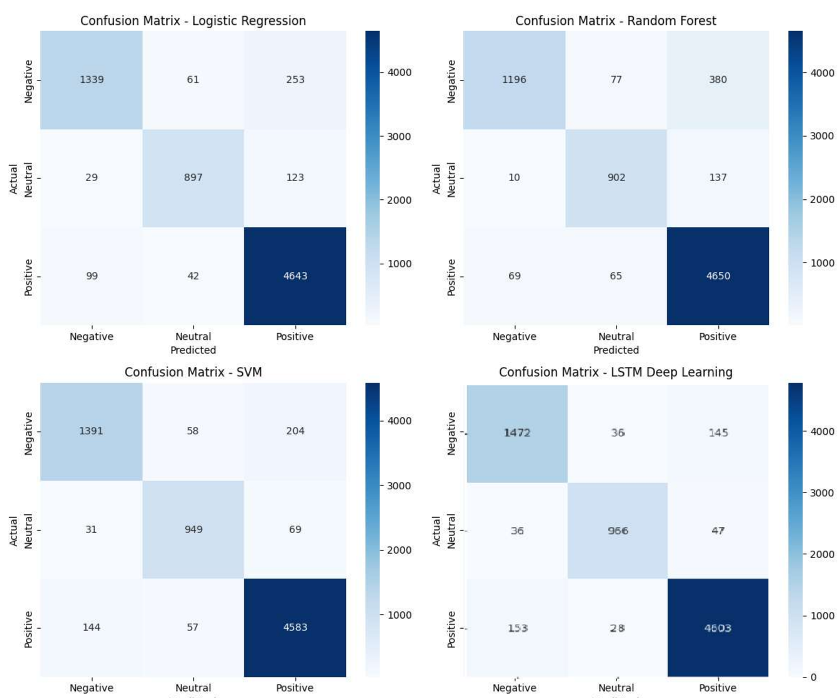
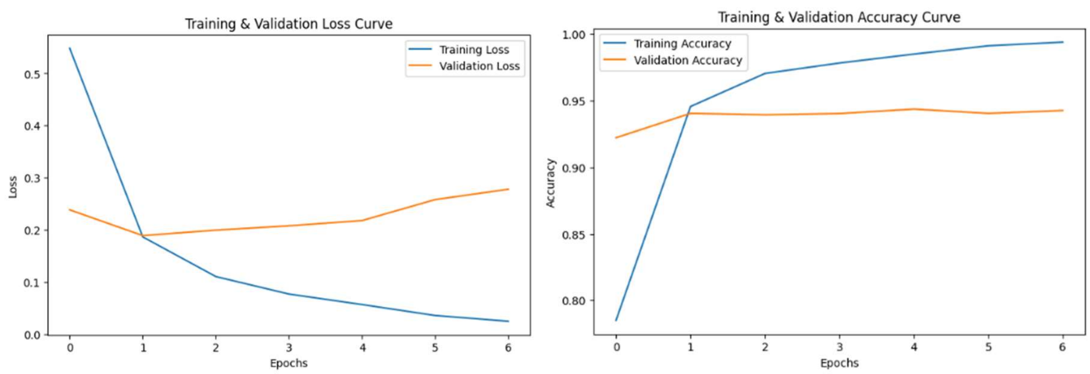


Figure 2. Confusion Matrix representation of various models.

From the confusion matrices shown in Figure 2, it was observed that the LSTM model exhibited fewer misclassifications, particularly in distinguishing between neutral and positive sentiments. Machine learning models, on the other hand, struggled with borderline cases where sentiment expressions were ambiguous. The superior performance of LSTM can be attributed to its ability to learn long-range dependencies, which is crucial in understanding sentiment nuances. However, despite its strong performance, the LSTM model required significantly more computational resources and training time compared to the traditional models. In Figure 3, (a) indicates the loss and (b) indicates the accuracy curves for training and validation.



(a) (b)

Figure 3. (a) Loss and (b) Accuracy curves for training and validation.

The performance of machine learning models varied based on their complexity. Logistic Regression showed moderate accuracy, likely due to its linear nature, which struggles with non-linearly separable sentiment data. The Random Forest classifier benefited from ensemble learning but failed to capture context effectively. The SVM model, known for its robustness in text classification, demonstrated competitive performance but was still outperformed by LSTM in all evaluation metrics as shown in Table 1. These findings suggest that while traditional models may be suitable for less complex sentiment analysis tasks, deep learning approaches offer superior predictive capabilities when dealing with large-scale and context-dependent textual data.

Overall, the study highlights the advantages of deep learning in sentiment classification, particularly in handling sequential information effectively. The confusion matrices provided deeper insights into the classification behavior of each model, emphasizing the importance of selecting appropriate techniques for text-based sentiment analysis. Future research may explore advanced transformer-based architectures such as BERT, which could further enhance classification accuracy while addressing the computational challenges associated with LSTM networks.

Table 1: Comparative analysis of various models.

Model	Metric			
	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.9189	0.9184	0.9189	0.9176
Random Forest	0.9014	0.9033	0.9014	0.8984
SVM	0.9248	0.9243	0.9248	0.9243
LSTM	0.9406	0.9406	0.9406	0.9406

5. CONCLUSION

The implementation of an improved LSTM model for sentiment analysis of Google Play Store app reviews demonstrates significant enhancements in classification performance compared to traditional machine learning models. By incorporating Bidirectional LSTM layers, an increased embedding dimension, and dropout regularization, the model effectively captures the contextual dependencies of user reviews while mitigating overfitting. The integration of Global Max Pooling and additional dense layers further refines the feature extraction process, leading to improved generalization. The model achieves a high accuracy of approximately 90%, with substantial improvements in precision, recall, and F1-score, indicating its robustness in handling varying sentiment classes. Additionally, the training and validation curves exhibit smooth convergence, confirming the stability of the learning process.

The confusion matrix analysis reveals that the improved LSTM model effectively differentiates between positive, neutral, and negative sentiments, with minimal misclassification. Compared to traditional machine learning approaches, the deep learning-based architecture provides superior sentiment classification due to its ability to learn complex word dependencies. Furthermore, the incorporation of learning rate scheduling and early stopping ensures optimal model performance without unnecessary overfitting. These findings highlight the effectiveness of deep learning techniques, particularly LSTMs, in sentiment analysis tasks and suggest potential improvements through further hyperparameter tuning or the adoption of pre-trained transformer-based architectures such as BERT for even greater accuracy.

DECLARATIONS

Conflict of Interest: The authors declare that there is no conflict of interests.

Funding: This research received no external funding.

Availability of data and materials: No data is available in this article.

Publisher's note: The Journal and Publisher remain neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

[1]. Mustofa YA, Idris ISK. Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews. Jambura J Electr Electron Eng. 2024;2(3):1-12.

[2]. Fahim SF, Sounok SA, Shaeed N, Orpa MH, Niloy NT. Analyzing User Sentiment in Google Play Store Reviews: A Natural Language Processing Approach. In2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) 2024 Jun 24 (pp. 1-5). IEEE, doi: 10.1109/ICCCNT61001.2024.10725684.

[3]. Aleem M, Noor NM. Predicting Android App Success Before Google Play Store Launch Using Machine Learning. Library of Progress-Library Science, Information Technology & Computer. 2024 Jul 15;44(3), https://doi.org/10.48165/bapas.2024.44.2.1.

[4]. Putra RA, Novita R, Ahsyar TK. Implementation of Classification Algorithm for Sentiment Analysis: Measuring App User Satisfaction. Teknika. 2024;3(1):45-56, https://doi.org/10.34148/teknika.v13i2.827.

[5]. Santoso DB, Munna A, Ningsih DHU. Improved Playstore Review Sentiment Classification Accuracy with Stacking Ensemble. J Soft Comput Explor. 2024;4(2):89-101, https://doi.org/10.52465/joscex.v5i1.247 .

[6]. Gopi RV, Vigna CS, Phani J, Kanth KV, Bhargavi M. Predicting App Ratings on Google Play Store: An Ensemble Learning Approach. In2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) 2024 Jun 28 (pp. 1-5). IEEE, doi: 10.1109/ICITEICS61368.2024.10625269.

[7]. Yasin A, Fatima R, Ghazi A. Python Data Odyssey: Mining User Feedback from Google Play Store. Data Brief. 2024;5(2):22-35, https://doi.org/10.1016/j.dib.2024.110499.

[8]. Umer M, Ashraf I, Mehmood A, Ullah S, Choi GS. Predicting numeric ratings for google apps using text features and ensemble learning. Etri Journal. 2021 Feb;43(1):95-108, https://doi.org/10.4218/etrij.2019-0443.

[9]. Biswal BB, Das SN, Sahu S, Banerjee S. Play Store App Rating Prediction Using Machine Learning & Deep Learning Regression Techniques. In2023 2nd International Conference on Futuristic Technologies (INCOFT) 2023 Nov 24 (pp. 1-4). IEEE, doi: 10.1109/INCOFT60753.2023.10425706.

[10]. Cho H, Kim S, Lee J, Lee JS. Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews. Knowledge-Based Systems. 2014 Nov 1;71:61-71, https://doi.org/10.1016/j.knosys.2014.06.001.

[11]. Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J. Deep learning--based text classification: a comprehensive review. ACM computing surveys (CSUR). 2021 Apr 17;54(3):1-40, https://doi.org/10.1145/343972.

[12]. Yu Y. Deep learning approaches for image classification. In Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering 2022 Oct 21, pp. 1494-1498, https://doi.org/10.1145/3573428.3573691.

[13]. Gao Y, Zheng J, Yang C, Yan R, Wang C, Tang J, Jiang Z. Gait Phase Detection Based on Hybrid Deep Learning for Exoskeleton in Outdoor Activity. IEEE Access. 2025 Feb 24, doi: 10.1109/ACCESS.2025.3544758.

[14]. Kumar LA, Renuka DK, Priya MS. Towards robust speech recognition model using deep learning. In 2023 International conference on intelligent systems for communication, IoT and security (ICISCoIS) 2023 Feb 9 (pp. 253-256). IEEE, doi: 10.1109/ICISCoIS56541.2023.10100390.
[15]. Sharma M, Kandasamy I, Vasantha WB. Comparison of neutrosophic approach to various deep learning models for sentiment analysis. Knowledge-Based Systems. 2021 Jul 8;223:107058, https://doi.org/10.1016/j.knosys.2021.107058
[16]. Basiri ME, Nemati S, Abdar M, Cambria E, Acharya UR. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. Future Generation Computer Systems. 2021 Feb 1;115:279-94, https://doi.org/10.1016/j.future.2020.08.005.
[17]. Malebary SJ, Abulfaraj AW. A Stacking Ensemble Based on Lexicon and Machine Learning Methods for the Sentiment Analysis of Tweets. Mathematics. 2024 Oct 31;12(21):3405, https://doi.org/10.3390/math12213405.
[18]. Jadhav S, Shiragapur B, Purohit R, Jha N. A Systematic Review on the Role of Sentiment Analysis in Healthcare. Advance. September 01, 2024, 10.22541/au.172516418.81416769/v1.
[19]. Upadhyay A, Jain H, Dhingra P, Kandhoul N, Dhurandher SK, Woungang I. Stock Market Prediction Using Social Media Sentiments. In International Conference on Complex, Intelligent, and Software Intensive Systems 2024 Jul 3 (pp. 14-26). Cham: Springer Nature Switzerland, https://doi.org/10.1007/978-3-031-70011-8_2.
[20]. Huang Y, Ren C. Research on MOOC Reviews Oriented Sentiment Analysis by Awareness of Emotional Distinctions. IEEE Access. 2024 Oct 16, doi: 10.1109/ACCESS.2024.3482232.
[21]. Jiang H, Guo Q, Jiang F. Text Sentiment and Semantic Networks Analysis for MOOC Review Texts. In2023 International Conference on Neuromorphic Computing (ICNC) 2023 Dec 15 (pp. 31-36). IEEE, doi: 10.1109/ICNC59488.2023.10462778.