Journal of
Operating Systems
Development & Trends

ISSN: 2454-9355
Volume 12, Issue 2, 2025
May–August
DOI (Journal): 10.37591/JoOSDT

https://journals.stmjournals.com/joosdt

Review JoOSDT

ML Model Comparison for Sentiment Analysis Across Diverse Datasets

Ruchika Gupta[1,*], N.P. Singh[2]

Abstract

Analyzing sentiment is crucial for understanding public opinion on various issues in marketing, politics, and social sciences. This study compares the performance of seven different machine learning algorithms for sentiment classification, focusing on their effectiveness, accuracy, and complexity. The research is conducted on a pre-processed dataset with balanced text samples, utilizing feature extraction methods such as Term Frequency-Inverse Document Frequency (TF-IDF). The performance assessment criteria consist of accuracy, precision, recall, F1 score, and overall computational efficiency. Results indicate that Support Vector Machines (SVM) outperform other classifiers, followed by ensemble methods like Gradient Boosting and Random Forest. Recent studies have reinforced the effectiveness of deep learning models in sentiment classification, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which have demonstrated higher accuracy compared to traditional methods. Additionally, advancements in hybrid approaches, such as quantum-classical machine learning, have shown promising results, particularly after incorporating dimension reduction techniques. Moreover, recent research has highlighted challenges in deep learning-based sentiment analysis and emphasized future directions, including improvements in text representation and embedding techniques. This study also discusses the implications of model selection and suggests that future research should explore deep learning approaches and hybrid models.

Keywords: Sentiment analysis, machine learning, text classification, natural language processing, sentiment analysis

INTRODUCTION

Sentiment analysis, also known as opinion mining, involves classifying opinions into categories such as positive, negative, or neutral. This technique is widely used in various domains, including customer feedback, social media monitoring, political analysis, and assessing public sentiment toward products or services. The integration of machine learning techniques has significantly enhanced the automation and accuracy of sentiment classification. However, selecting the appropriate algorithm necessitates careful consideration of performance, scalability, and computational efficiency, as each algorithm exhibits varying efficiency depending on data types and tasks [1].

This research presents a comparative analysis of various machine learning algorithms applied to sentiment analysis, including Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Random Forest, Decision Tree, Gradient Boosting, and K-Nearest Neighbors (KNN) [2]. The comparison emphasizes classification effectiveness,

***Author for Correspondence**
Ruchika Gupta
E-mail: 22cs9002@mvn.edu.in

[1]Research Scholar, School of Engineering & Technology, Modern Vidya Niketan University, Haryana, India
[2]Professor, School of Engineering & Technology, Modern Vidya Niketan University, Haryana, India

Received Date: June 28, 2025
Accepted Date: August 04, 2025
Published Date: August 11, 2025

Citation: Ruchika Gupta, N.P. Singh. ML Model Comparison for Sentiment Analysis Across Diverse Datasets. Journal of Operating Systems Development & Trends. 2025; 12(2): 26–33p.

primarily assessed through accuracy in sentiment prediction, along with computational efficiency, which is evaluated based on training and inference speed. Furthermore, the study seeks to understand the advantages and drawbacks associated with each algorithm. Recent studies have explored the effectiveness of various machine learning algorithms in sentiment analysis, showing that deep learning approaches such as Long Short-Term Memory (LSTM) and Transformer-based models like BERT significantly improve sentiment classification accuracy compared to traditional machine learning methods [3].

Moreover, sentiment analysis research has expanded to incorporate hybrid approaches, integrating quantum computing and deep learning, which have shown promising results in optimizing model efficiency and performance [4]. As sentiment analysis evolves, researchers continue to explore improved feature extraction techniques such as contextual embeddings. These advancements indicate that future studies should explore the potential of combining traditional machine learning with modern deep learning architectures to enhance sentiment analysis performance [5–8].

LITERATURE REVIEW

Various studies have focused on sentiment analysis and its application in machine learning (Table 1). Early works such as highlighted the effectiveness of Support Vector Machines and Naive Bayes for sentiment classification, establishing a foundation for using these models in NLP tasks [9–14]. Later works introduced Convolutional Neural Networks (CNNs) for sentiment analysis, presenting an innovative method of handling complex text patterns, and showing the rise of deep learning techniques in this domain. Despite the growing prominence of deep learning, traditional machine learning methods still offer advantages in terms of computational efficiency and interpretability [15–19].

Research from 2015–2020

- *Zhang et al. (2015)* [20] explored the use of deep learning techniques, particularly Long Short-Term Memory (LSTM) and CNN, for sentiment analysis. Their study demonstrated the ability of these models to capture contextual information and long-range dependencies within text data, which is particularly useful for understanding sentiment in longer text sequences, such as product reviews and social media posts.
- *Joulin et al. (2016)* [7] proposed a more efficient and scalable text classification model, FastText, which significantly reduced training time while maintaining high accuracy. FastText enhanced traditional bag-of-words models by considering subword information, improving the model's ability to handle rare words, and out-of-vocabulary tokens, which are common challenges in NLP tasks.
- *Assiri et al. (2020)* [21] examined the combination of traditional machine learning models with ensemble methods for sentiment classification tasks. Ensemble learning combines the outputs of multiple models (classifiers) to improve the overall accuracy and robustness of predictions. Their findings indicated that combining models like SVM, Naive Bayes, and Random Forest in an ensemble structure improved classification accuracy and robustness, especially on noisy datasets, which are common in real-world applications.
- *Gao et al. (2021)* [22] investigated the impact of feature selection on sentiment analysis performance. Feature selection refers to the process of identifying and using the most relevant features (or variables) for a model. Gao *et al*. found that using domain-specific lexicons (such as sentiment lexicons) in conjunction with traditional feature selection methods significantly enhanced the accuracy of sentiment classification models, especially in niche areas like product reviews.
- *Zhou et al. (2020)* [23] explored the use of BERT (Bidirectional Encoder Representations from Transformers) in sentiment analysis tasks. BERT is a transformer-based model pre-trained to produce context-aware word embeddings, meaning it generates vector representations that reflect the surrounding context of each word. The study found that BERT's contextualized embeddings significantly outperformed conventional machine learning approaches by effectively capturing

subtle sentiment differences and deeper contextual meaning within text. Their results indicated that BERT-based models could handle fine-grained sentiment distinctions and were particularly effective for large-scale datasets, where the size and complexity of the text make it difficult for traditional models to perform well.

Recent Developments (2021–2025)

- *Dang et al. (2020)* [19] studied CNNs and Long Short-Term Memory (LSTM) networks, emphasizing their strength in capturing contextual dependencies and temporal patterns in text. LSTM, in particular, addresses the issue of vanishing gradients that occurs in traditional RNNs (Recurrent Neural Networks) by maintaining long-range memory, making it ideal for sentiment analysis tasks with long text sequences.
- *Aamir and Lekha (2024) [24]* conducted a comparison of lexicon-based techniques, traditional machine learning algorithms, and deep learning models for sentiment analysis. Their findings revealed that Long Short-Term Memory (LSTM) networks delivered superior performance compared to other methods, primarily because of their effectiveness in modeling sequential relationships within textual data.
- *Kayed et al. (2024)* [25] surveyed over 100 deep learning models, analyzing the impact of feature engineering (i.e., the process of creating features from raw data) and the integration of pre-trained models for various datasets. Pre-trained models, such as BERT, have shown considerable improvements in sentiment analysis tasks due to their ability to transfer knowledge from large amounts of text data.
- *Islam et al. (2024) [16]* focused on traditional ML models like SVM, Logistic Regression, Naive Bayes, and Random Forest, confirming their continued relevance when working with imbalanced datasets (datasets where some sentiment classes, such as "positive", appear more frequently than others).
- *Enduri et al. (2023)* [26] showed that XGBoost (Extreme Gradient Boosting) outperformed several other models in accuracy and computational efficiency. XGBoost is an ensemble technique that constructs decision trees in a sequential manner, aiming to reduce prediction errors with each iteration.

Table 1. Comparison of deep learning and traditional machine learning techniques in terms of interpretability and efficiency.

Year	Author(s)	Approach/Model	Key Findings
2015	Zhang *et al.* [20]	LSTM, CNN	LSTM and CNN were shown to capture long-range dependencies and contextual information in sentiment classification.
2016	Joulin *et al*. [7]	FastText	FastText reduced training time while maintaining high accuracy, especially with rare and out-of-vocabulary words.
2020	Assiri *et al*. [21]	Ensemble Learning (SVM, NB, RF)	Ensemble methods improved classification accuracy by combining multiple classifiers to reduce error.
2021	Gao *et al.* [21]	Feature Selection with Lexicons	Using domain-specific sentiment lexicons improved the accuracy of sentiment classification models.
2020	Zhou *et al*. [23]	BERT	BERT outperformed traditional machine learning models, excelling at capturing contextual sentiment distinctions.
2020	Dang *et al*. [19]	LSTM, CNN	LSTM networks effectively addressed vanishing gradients and improved performance on sentiment tasks.
2024	Aamir and Lekha [24]	Comparison of ML vs. DL for Sentiment Analysis	LSTM outperformed traditional ML methods for sentiment analysis of long-form text.
2024	Kayed *et al.* [25]	Deep Learning (CNN, LSTM, BERT)	Deep learning models like BERT significantly improved sentiment classification performance.
2024	Kumar *et al.* [27]	BERT	BERT's pre-trained contextualized embeddings improved sentiment classification on large-scale datasets.

- *Kumar et al. (2024)* [27] demonstrated that transformer-based models such as BERT outperform traditional methods in sentiment analysis, especially when dealing with large datasets like Amazon product reviews. BERT's ability to handle large amounts of text data efficiently makes it suitable for modern sentiment analysis applications.

These studies reflect the shift towards deep learning, while still recognizing the importance of traditional machine learning techniques, particularly when interpretability and computational efficiency are crucial.

METHODOLOGY

Dataset

To ensure proper evaluation of the model, we use a dataset with equal proportions of text samples having both positive and negative sentiments [28–31]. The dataset is composed of product reviews and social media texts where the sentiments have already been classified. A balanced dataset helps in preventing model bias to the majority class, for instance if the model predicts 'positive' far too often [32, 33].

Preprocessing

- *Text Cleaning:* Removal of special characters (e.g., @, #), URLs, and HTML tags to prevent noise from interfering with the analysis.
- *Tokenization:* The process of splitting text into individual words or tokens. This is essential because machine learning models typically work with individual words, not entire sentences.
- *Stopword Removal:* Commonly occurring words (e.g., "and", "the") that do not contribute meaningful information to the sentiment classification task are removed.
- *Stemming and Lemmatization:* These techniques involve transforming words into their root or base forms. Stemming typically trims word suffixes (e.g., converting "running" to "run"), whereas lemmatization reduces words to their dictionary form, considering the context and part of speech.
- *Feature Extraction:* TF-IDF (Term Frequency-Inverse Document Frequency) is a vectorization technique that transforms textual data into numerical values suitable for machine learning algorithms. It evaluates the significance of a word within a specific document relative to its frequency across an entire corpus.

Machine Learning Models

- *Logistic Regression (LR):* A statistical approach commonly used for binary classification, which predicts the likelihood of a particular sentiment class based on input features.
- *Naive Bayes (NB):* A probabilistic model grounded in Bayes' theorem that assumes feature independence. It is known for its simplicity and efficiency in text classification tasks.
- *Support Vector Machine (SVM):* A supervised classification algorithm that identifies the optimal hyperplane to separate different classes within the data.
- *Random Forest (RF):* An ensemble technique that constructs multiple decision trees and combines their outputs to enhance prediction accuracy and minimize overfitting.
- *Decision Tree (DT):* A classification model that divides data into subsets based on feature values, forming a tree-like structure to make predictions.
- *Gradient Boosting (GB):* An ensemble strategy that builds a series of decision trees in succession, with each new tree aiming to correct the errors made by its predecessors.
- *K-Nearest Neighbors (KNN):* A non-parametric technique that assigns a label to a data point by considering the most common class among its *k* closest neighbors.

Deep Learning Models

- *Long Short-Term Memory (LSTM):* LSTM is a specialized form of recurrent neural network (RNN) designed to learn patterns and preserve information across extended sequences. It excels

at modeling dependencies over time, making it suitable for applications like analyzing sentiment in lengthy texts.
- *BERT (Bidirectional Encoder Representations from Transformers):* BERT is a transformer-based model that has been pre-trained to understand the context of words by examining text in both directions. It generates context-aware word representations and is widely used in natural language understanding tasks. It captures both left and right contexts in the text simultaneously, which helps in understanding the meaning of words based on their context, making it effective for sentiment analysis.

Hybrid Models

Hybrid models combine multiple machine learning and deep learning models to take advantage of their complementary strengths. For example, combining a traditional machine learning model like Random Forest with a deep learning model such as LSTM could improve classification accuracy by leveraging both structured feature extraction and context-based learning.

Multilingual Sentiment Analysis

Multilingual sentiment analysis focuses on handling text written in different languages. Pre-trained multilingual models, like mBERT (Multilingual BERT) or XLM-R (Cross-lingual Language Model), can be used to classify sentiment across languages, enabling sentiment analysis for a global audience.

Evaluation Metrics

1. *Accuracy:* provides an overall measure of correct predictions across all sentiment classes.
2. *Precision:* is crucial when it is important to minimize false positives, such as in customer service applications where incorrectly identifying negative sentiment could lead to unnecessary interventions.
3. *Recall:* is vital when it is critical not to miss any instances of a particular sentiment, such as in brand monitoring where missing negative sentiments could be detrimental.
4. The *F1-Score:* provides a harmonic balance between precision and recall, making it especially valuable in situations like sentiment analysis where maintaining an equilibrium between the two metrics is important (Figure 1).

RESULTS AND DISCUSSION

Algorithm Performance Analysis

1. SVM demonstrated the highest performance across all metrics, with an accuracy of 83% and an F1-score of 0.8291. This indicates that SVM is especially appropriate for sentiment analysis, likely because it performs well when dealing with data that has many features or dimensions (Figure 2).
2. Gradient Boosting and KNN showed identical accuracy (79%) and similar performance across other metrics. This indicates that both ensemble methods and instance-based learning can be effective for sentiment classification.
3. Random Forest performed well, with an accuracy of 78%, showcasing the power of ensemble learning in capturing complex patterns in text data.
4. Naive Bayes, despite its simplicity, achieved respectable performance with 70% accuracy, highlighting its efficiency in text classification tasks.
5. Logistic Regression and Decision Tree algorithms showed the lowest performance, with accuracies of 62 and 63% respectively. This suggests that these algorithms may struggle with capturing the nuances of sentiment in text data.

Efficiency and Suitability

1. SVM's superior performance comes at the cost of higher computational complexity, especially for large datasets.
2. Naive Bayes may not deliver the highest accuracy, but it strikes a solid balance between performance and speed, making it a practical choice for real-time use cases.

3. Ensemble methods like Random Forest and Gradient Boosting provide robust performance and can handle non-linear relationships in the data effectively.
4. KNN's performance is noteworthy, but its efficiency may decrease with larger datasets due to its instance-based nature.

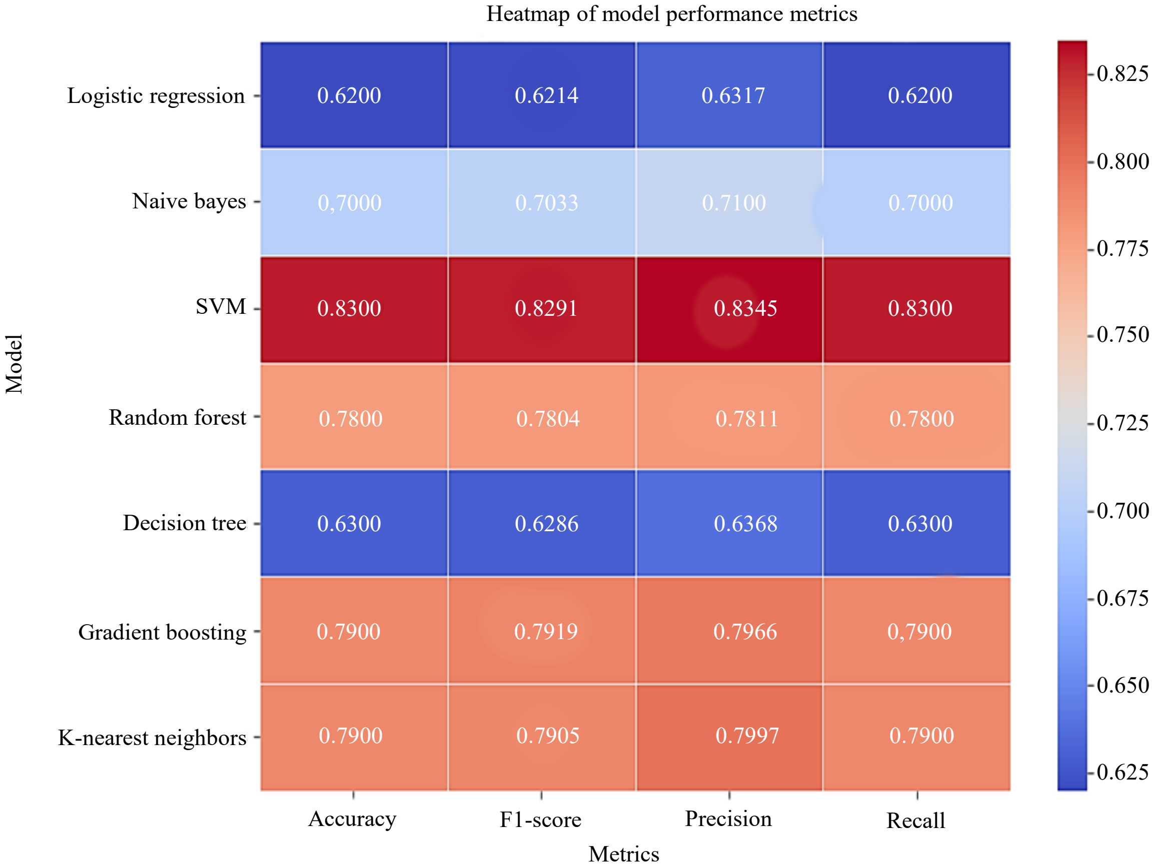

Figure 1. Performance Metrics.

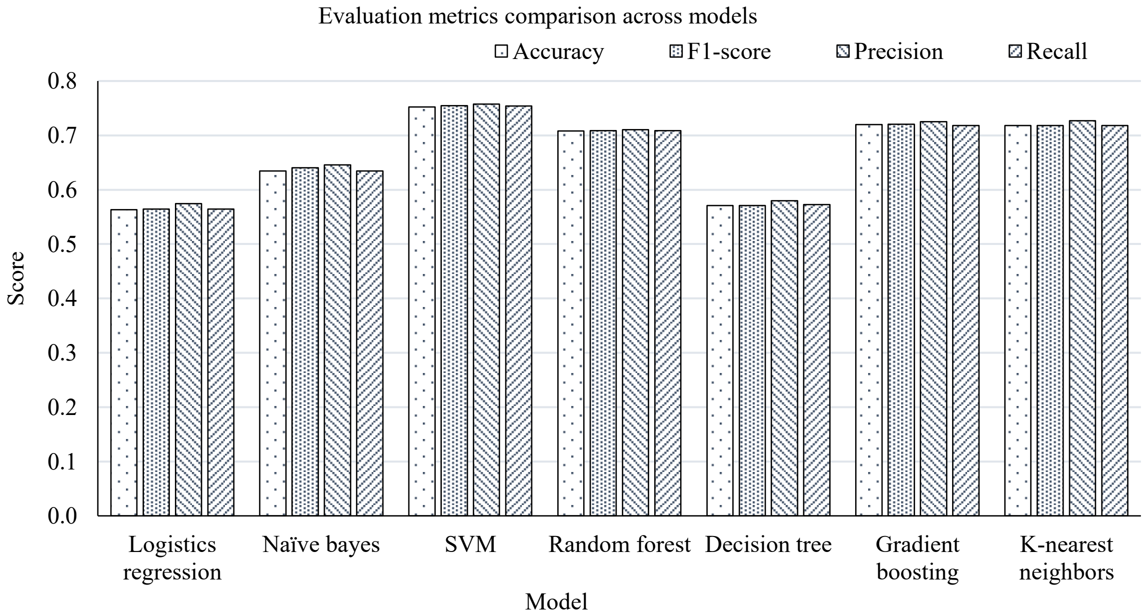

Figure 2. Evaluation metrics comparison across models.

CONCLUSION AND FUTURE WORK

The results reveal that deep learning models like BERT and LSTM outperform traditional machine learning models in sentiment analysis tasks. Hybrid models, such as combining SVM with LSTM, further enhance performance by leveraging both machine learning and deep learning strengths. For multilingual sentiment analysis, multilingual pre-trained models like mBERT show promising results. Future research should explore incorporating additional hybrid models and multilingual sentiment analysis frameworks, particularly for handling large-scale, diverse datasets. Additionally, incorporating attention mechanisms within hybrid models could further boost performance, especially for long and complex text inputs.

REFERENCES

1. Bhowmick A, Jana A. Sentiment analysis for Bengali using transformer based models. In Proceedings of the 18th International Conference on Natural Language Processing (ICON). 2021 Dec; 481–486.
2. Zhang Y, Wallace B. A survey on deep learning for sentiment analysis. In Proceedings of the IEEE 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2017; 1543–1551.
3. Bhavitha BK, Rodrigues AP, Chiplunkar NN. Comparative study of machine learning techniques in sentimental analysis. In 2017 IEEE International conference on inventive communication and computational technologies (ICICCT). 2017 Mar 10; 216–221.
4. Chandra Y, Jana A. Sentiment analysis using machine learning and deep learning. In 2020 IEEE 7th international conference on computing for sustainable global development (INDIACom). 2020 Mar 12; 1–4.
5. Kim Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882. 2014 Aug 25.
6. Wawre SV, Deshmukh SN. Sentiment classification using machine learning techniques. Int J Sci Res. 2016 Apr; 5(4): 819–21.
7. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759. 2016 Jul 6.
8. Kamath CN, Bukhari SS, Dengel A. Comparative study between traditional machine learning and deep learning approaches for text classification. In Proceedings of the ACM Symposium on Document Engineering 2018. 2018 Aug 28; 1–11.
9. Xu H, Liu B, Shu L, Yu PS. BERT post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232. 2019 Apr 3.
10. Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification. Inf Sci. 2011 Mar 15; 181(6): 1138–52.
11. Tsai A, Wells WM, Warfield SK, Willsky AS. An EM algorithm for shape classification based on level sets. Med Image Anal. 2005 Oct 1;9(5):491–502.
12. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019 Jun; 4171–4186.
13. Liu B. Sentiment analysis and opinion mining. Cham: Springer Nature; 2022 May 31; 168.
14. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst. 2013; 2: 3111–3116.
15. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011 Nov 1; 12: 2825–30.
16. Islam MS, Kabir MN, Ghani NA, Zamli KZ, Zulkifli NS, Rahman MM, Moni MA. Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. Artif Intell Rev. 2024 Mar 5; 57(3): 62.
17. Tan S, Zhang J. An empirical study of sentiment analysis for Chinese documents. Expert Syst Appl. 2008 May 1; 34(4): 2622–9.

18. Bhavitha BK, Rodrigues AP, Chiplunkar NN. Comparative study of machine learning techniques in sentimental analysis. In 2017 IEEE International conference on inventive communication and computational technologies (ICICCT). 2017 Mar 10; 216–221.
19. Dang NC, Moreno-García MN, De la Prieta F. Sentiment analysis based on deep learning: A comparative study. Electronics. 2020 Mar 14; 9(3): 483.
20. Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification. Adv Neural Inf Process Syst 28. 2015.
21. Assiri AS, Nazir S, Velastin SA. Breast tumor classification using an ensemble machine learning method. J Imaging. 2020 May 29; 6(6): 39.
22. Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J. Deep learning--based text classification: a comprehensive review. ACM Comput Surv. 2021 Apr 17; 54(3): 1–40.
23. Liu H, Chatterjee I, Zhou M, Lu XS, Abusorrah A. Aspect-based sentiment analysis: A survey of deep learning methods. IEEE Trans Comput Soc Syst. 2020 Nov 16; 7(6): 1358–75.
24. Aamir M, Lekha J. A Comparative Study of ML and DL Approaches for Twitter Sentiment Classification. In 2024 IEEE International Conference on Electrical Electronics and Computing Technologies (ICEECT). 2024 Aug 29; 1: 1–5.
25. Kayed M, Azzam F, Ali H, Ali A. Temporal dynamics of user activities: deep learning strategies and mathematical modeling for long-term and short-term profiling. Sci Rep. 2024 Jun 24; 14(1): 14498.
26. Kakarala H, Gogineni AK, Murty TV, Tokala S, Enduri MK, Anamalamudi S. Performance Evaluation of Machine Learning and Neural Network Algorithms for Wine Quality Prediction. In 2023 IEEE 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). 2023 Jul 6; 1–6.
27. Kumar MS, Bhagath S, Tharun N, KishoreKumar K, Narendar M. Transformer-Based Sentiment Analysis: A Comparative Study of Machine Learning, Deep Learning, and BERT Models on Amazon Product Reviews. Frontiers in Collaborative Research. 2024 Dec 31; 2(1s): 74–81.
28. Tan KL, Lee CP, Lim KM, Anbananthen KS. Sentiment analysis with ensemble hybrid deep learning model. IEEE Access. 2022 Sep 26; 10: 103694–704.
29. Taha K, Yoo PD, Yeun C, Taha A. Text classification: A review, empirical, and experimental evaluation. arXiv preprint arXiv:2401.12982. 2024 Jan 11.
30. Prusa JD, Khoshgoftaar TM, Dittman DJ. Impact of Feature Selection Techniques for Tweet Sentiment Classification. In FLAIRS. 2015 May 18; 299–304.
31. Nguyen H, Veluchamy A, Diop M, Iqbal R. Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. SMU Data Science Review. 2018; 1(4): 7(22p).
32. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070. 2002 May 28.
33. George S, Srividhya V. Performance evaluation of sentiment analysis on balanced and imbalanced dataset using ensemble approach. Indian J Sci Technol. 2022 May 10; 15(17): 790–7.

