FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 9, May, 2026, pp 95 – 102
DOI: https://doi.org/10.33003/fjs-2026-1009-4892

Performance Evaluation of Some Selected Ensemble Classifiers For Users' Online Book Reviews Based On Sentiment Analysis

***Fatimah Binta Abdullahi and Olufunwa Great**

Department of Computer Science, Faculty of Physical Sciences, University of Abuja, FCT Abuja, Nigeria.

*Corresponding authors' email: fatimah.abdullahi@uniabuja.edu.ng

ABSTRACT

User reviews on platforms such as Amazon have become a focal point due to their extensive use in sentiment analysis, which provides valuable feedback to the public, private companies, and governments. Analysing these reviews not only contributes to enhancing the quality of products and services but also supports the development of marketing and financial strategies aimed at boosting profitability and customer satisfaction. Even with several models developed for this task, there is room to enhance the processing, classification, and interpretation of user feedback, thereby assisting product managers in refining product quality. This paper introduces ensemble classifiers designed to categorize reviews as positive, negative, or neutral. The study first assesses the performance of widely used ensemble methods, including Adaptive Boosting (AdaBoost), Categorical Boosting (CatBoost), Gradient Boosting (GBM), and Extreme Gradient Boosting (XGBoost). It then evaluates stacking ensemble classifiers, where outputs from Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM) are combined with CatBoost, XGBoost, AdaBoost, and GBM. Using dataset sourced from Amazon, experimental results demonstrate that stacked ensemble classifiers achieve superior accuracy, 88.23%, 86.01%, 84.56%, and 85.12%, compared to traditional ensemble classifiers.

Received: 16 May 2026

Accepted: 01 June 2026

Published: 16 June 2026

Keywords: Ensemble Classifier, Sentiment Analysis, Stacking Ensemble Classifiers, User Reviews

INTRODUCTION

The rapid emergence and expansion of online platforms such as Twitter, Facebook, Google, and Amazon have transformed them into tools that enable users to brainstorm, exchange information, and articulate opinions through reviews. These reviews constitute a valuable source of user feedback, which can be leveraged to revise existing policies and devise new strategies aimed at improving the overall product quality and services. Nevertheless, extracting meaningful insights from user-generated opinions remains a complex and non-trivial task.

A specialized domain known as *sentiment analysis* (Khalid et al., 2020) provides a range of techniques and tools designed to extract vital information that reflects users' perspectives. These approaches are usually situated within the broader domain of natural language processing (Liu, 2015; Lui, 2010). Within this context, the steps of extracting opinions from user reviews are referred to as sentiment analysis. Opinion mining seeks to develop systems capable of extracting, classifying, and interpreting reviews, thereby facilitating the identification of sentiments embedded within textual data.

Conventionally, the term sentiment analysis focuses on classifying opinion polarity into either positive, negative, or neutral categories (Cambria et al., 2017). Polarity determination typically refers to the presence of emotions or specialized lexical cues within user reviews. These reviews are of considerable importance to both the public and private sectors, as they encapsulate user preferences, including likes and dislikes usually provide vital information for improving product and service quality. Furthermore, such feedback can inform the formulation or revision of organizational policies.

Textual comments are as important as numerical ratings when compared to them, as they capture the nuanced perspectives of users regarding specific products (Zhang et al., 2018). Despite their potential value to governments, businesses, and individuals, the vast volume of user-generated content necessitates the application of text mining and sentiment analysis techniques. Nevertheless, sentiment analysis is an inherently complex task, facing numerous obstacles that impede the precise determination of sentiment polarity (Yadav & Singh, 2020).

One major challenge lies in reviewing texts, which are often semi-structured or unstructured. Since most reviewers are non-expert or non-professional writers, their expressions typically lack standardized linguistic or structural conventions, resulting in irregular data formats (Hang et al., 2018). Additional difficulties arise from domain dependence, variability in review structures, and the intricacies of language semantics that complicate sentiment interpretation. For thorough reviews of these challenges, readers are referred to Yadav & Singh (2020).

In general, sentiment analysis seeks to identify positive or negative expressions of opinion concerning individuals, events, or digital content. Accurate interpretation of user sentiment enables businesses, service providers, and policymakers to reassess strategies, enhance offerings, and obtain critical feedback regarding product performance and areas requiring improvement. Against this backdrop, the present study makes the following contributions:

i. To investigate how ensemble classifiers perform sentiment analysis on user online reviews.
ii. To investigate stacking ensemble using Machine Learning classifiers (RF, NB, SVM) on Adaboost, CatBoost, GBM,

and XGBoost to determine their classification performance.

The subsequent section of this paper is structured as follows: Section 2 provides a review of related/recent literature. Section 3 outlines the methodology underlying the proposed approach. Section 4 presents the experimental results and their analysis, while Section 5 offers a detailed discussion of these findings. Finally, Section 6 concludes the findings of the paper and suggests directions for future research.

Related Work

Vidyashree et al. (2024) proposed a tweet sentiment classification framework employing an ensemble classifier, emphasizing sentiment analysis as a critical tool for capturing public opinion in applications that utilize ensemble learning methods. However, the research lacks interpretability as it was difficult to understand. Ramdani (2025) introduced natural language processing and sentiment analysis by defining their key concepts, exploring their historical evolution, and how they relate to each other. In addition, various types, architectures, and tools are used. Zhengbing et al. (2024) proposed the integration of advanced Natural Language Processing (NLP) models and classifiers in sentiment analysis for binary text classification, which demonstrated significant improvement in prediction accuracy. The research uses ensemble learning data augmentation that is limited to handling dynamic and evolving language patterns such as emojis, slang, and stickers. Tiwari et al in 2023 provided a comprehensive overview of ensemble classifiers using various ensemble methods. Bhowmic et al (2024) leverage ensemble learning for enhanced sentiment classification, which achieved higher accuracy; however, they lack exploration of a varied e-commerce domain. Guru et al. (2024) propose the use of sentiment analysis on transfer learning for e-commerce that provides insights into effective classification techniques using SVM and Random Forest. Basha and Rajput (2017) investigated the role of sentiment analysis feature selection process, and their analysis demonstrated that incorporating feature selection techniques leads to a significant improvement in classification accuracy. In a comparative study, Abheek and Preeti (2021) highlighted the respective advantages of lexicon driven strategies and machine learning techniques. Srivastava and Bhatia (2016) provided an overview of the benefits of using ensemble methods for sentiment analysis of online micro-texts across diverse domains. Joshi & Kadu in 2024 designed an ensemble sentiment analysis model feedback evaluation via a multimodal feature selection process that achieved high accuracy. Tan et al. (2022) introduced a sentiment analysis framework based on a hybrid ensemble deep learning model that demonstrated improved accuracy in sentiment classification. Similarly, Thein and Phan (2020) explored the combination of deep learning techniques within ensemble learning approaches, reporting high classification accuracy and underscoring the effectiveness of leveraging deep learning and ensemble methodologies. Kelvin et al. (2026), proposed ensemble learning integrated with advanced feature selection techniques for lightweight, efficient, and adaptive machine learning framework that enhances ransomware classification. Dahiru et al (2026), examines ensemble explainability frameworks for multimodal Chest X-ray (CXR) classification using SHAP, Grad-CAM, and LIME.

MATERIALS AND METHODS

Sentiment analysis serves as an important tool for capturing public opinion in domains such as online book reviews. The present study focuses on extracting sentiment from reviews posted on the Amazon platform. To accomplish this, an ensemble-based classifier is employed to categorize sentiments expressed in textual data. The overall sentiment classification framework using the ensemble approach is illustrated diagrammatically in Figure 1

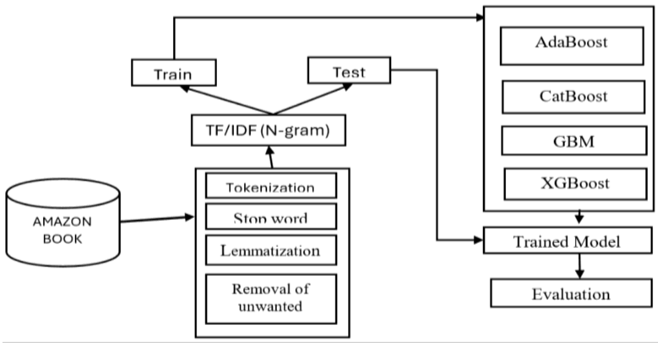

Figure 1: Workflow for Sentiment Classification of Online Reviews (Olufunwa et al, 2025)

Data Pre-processing

The raw data obtained from the Amazon dataset (Zhao et al., 2020) undergoes pre-processing to eliminate irrelevant or noisy information that complicates the classification process. Amazon's online book review data is often unsuitable for direct analysis; therefore, normalization is required to transform it into a more structured format prior to the application of analytical methodologies. In the proposed framework, pre-processing techniques are employed to remove extraneous content, thereby ensuring that the data can be effectively utilized by various learning algorithms.

A critical aspect of data preparation involves converting raw, unstructured input into a comprehensible and standardized format. Real-world data frequently contains inconsistencies, fragmentation, and irregular patterns, which introduce errors and hinder accurate analysis. Pre-processing thus serves as a fundamental strategy to address these challenges in sentiment analysis, ultimately enhancing the accuracy of classification models. The specific steps involved in the pre-processing stage are outlined as follows:

Removal of Stop Words

Frequently occurring but semantically redundant words, such as adverbs, prepositions, and articles, are excluded from the dataset. Eliminating these stop words reduces the dimensionality of the data and enhances the efficiency of the classification process.

Elimination of Blank Spaces

Superfluous blank spaces within the text are removed, as their presence increases computational overhead and slows down the classification process.

Feature Extraction

Following pre-processing, the data undergoes feature extraction, wherein relevant attributes are derived using established techniques such as the Bag-of-n-grams model and Term Frequency–Inverse Document Frequency (TF-IDF). The Bag-of-n-grams approach is employed to capture the continuity of words within textual data, which is essential for sentiment analysis. This model is typically represented in three forms: unigrams, bigrams, and trigrams. A unigram corresponds to a single word, a bigram represents a sequence of two consecutive words, and a trigram denotes a sequence of three or more words.

Classification using Ensemble Methods

The features selected through the wrapper-based technique are subsequently processed for sentiment classification using an ensemble approach. Ensemble techniques that integrate diverse algorithms are widely employed to improve classification performance. In this work, notable machine learning classifiers are integrated, and classification is performed using the Adaptive Boosting (AdaBoost) method. Specifically, sentiment classification of tweets is carried out through a stacked ensemble model that integrates the Random Forest (RF), Naive Bayes (NB), and the Support Vector Machine (SVM) classifiers. Stacked ensemble models are known to achieve higher accuracy as compared to the conventional ensemble techniques. The workflow of the stacked ensemble model is illustrated in Figure 2.

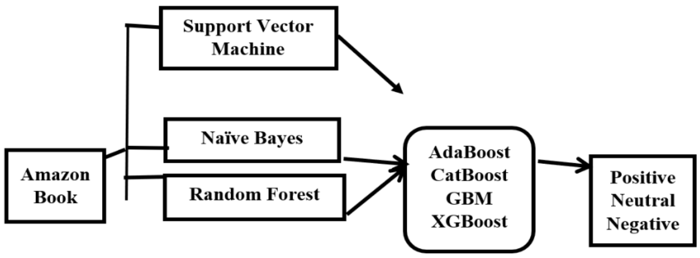

Figure 2: Workflow of the Ensemble Classification Model

Random Forest

is a widely recognized model used for classification and regression tasks. It has evolved into an ensemble learning approach that combines different decision trees during the training phase. Each tree contributes to the overall classification, with the model aggregating their outputs to produce a final prediction. By introducing randomness in tree selection and reducing correlations among individual trees, the Random Forest effectively mitigates overfitting and enhances predictive accuracy.

Naive Bayes

is a classification technique that is based on the use of Bayes' theorem, assuming all estimators are independent of each other. In Bayesian classification, the basic idea is to find the posterior probability, that is, the probability that the label will achieve some agreement.

Support Vector Machine

is a robust algorithm designed to perform binary classification in both linear and non-linear contexts. Since many datasets exhibit near-linear separability, the main goal of SVM is to identify optimal hyperplanes that distinguish data points into different categories. In sentiment analysis, this enables the classification of samples into positive, neutral, or negative classes, guided by the principle of margin maximization and error minimization.

Ensemble Models

Ensemble models are widely used in machine learning and pattern recognition tasks, where they combine the outputs of multiple weaker algorithms to enhance both accuracy and classification efficiency. By integrating diverse classifiers, ensemble approaches generally achieve superior performance compared to individual models. In this paper, the ensemble framework evaluates the predictive capabilities of individual classifiers and aggregates them using techniques such as AdaBoost, CatBoost, Gradient Boosting Machine (GBM), and XGBoost, thereby improving overall sentiment classification accuracy.

i. AdaBoost is a boosting ensemble learning technique that develops a strong predictive model by sequentially integrating different weak learners, most often simple decision trees. On each iteration, greater weights are assigned to misclassified instances, enabling subsequent learners to concentrate on correcting the errors of their predecessors.
ii. GBM extends the idea of boosting by constructing decision trees sequentially, with each tree designed to correct the errors of its predecessors through the minimization of a specified loss function, such as log-loss in binary classification tasks. Unlike AdaBoost, which adjusts instance weights, GBM employs gradient descent optimization to iteratively reduce residual errors, thereby refining the model's predictive performance at each stage.
iii. XGBoost is an extension of gradient boosting that emphasizes efficiency and high performance. Building upon GBM's sequential tree-construction framework, it incorporates enhancements such as L1 and L2 regularization to reduce overfitting, parallelized processing to accelerate training, and native mechanisms for handling missing values.
iv. CatBoost is a gradient boosting framework specifically designed to efficiently handle datasets with numerous categorical features. Compared to other models that require extensive preprocessing steps, such as one-hot encoding, CatBoost natively processes categorical variables with reduced complexity. It employs an ordered boosting strategy, which minimizes overfitting by processing data in a defined order, thereby ensuring more stable and reliable predictions.

Three base models were trained, and their classifications served as input to the ensemble methods, which learn to combine these classifications to improve overall performance. For example, the SVM, NB, and RF were the base models trained. After training, the outputs from the base classifiers were utilized as input features for subsequent models, including AdaBoost, CatBoost, GBM, and XGBoost. These ensemble models combined the predictions of the base classifiers to derive the final sentiment classification, which could be positive, negative, or neutral, as presented in the figure above.

Dataset

The Amazon online book dataset consists of reviews regarding users' reviews, book titles, and user ratings. The Amazon online book dataset is a publicly available resource that provides valuable insights for real-world applications. This dataset comprises approximately 500 reviews, many of which contain lengthy sentences, stop words, misspellings, and other inconsistencies. To address these issues and facilitate effective classification, a pre-processing stage was applied to refine the data and ensure it was suitable for subsequent analysis.

Performance Metrics

To assess sentiment classification, the ensemble classifiers' performance is evaluated using standard metrics, namely the accuracy, precision, recall, and F1-score, as defined in Equations (1) - (4).

Accuracy

Accuracy is defined as the percentage of correctly classified instances out of all instances in the dataset.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \qquad (1)$$

Where TN represents *true negatives*, TP represents *true positives*, FP represents *false positives*, and FN represents *false negatives*.

Precision

Precision is defined as the fraction of correctly identified positive instances (true positives) to all instances classified as positive (true positives plus false positives).

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (2)$$

Recall

is defined as the ratio of correctly detected positive instances that are correctly identified as positive, to the sum of true positive and false negative instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (3)$$

F1-Score

The F1-score is a metric that integrates both precision and recall, offering a balanced measure of classification performance. It is computed as the harmonic mean of precision and recall. When precision and recall both equal 1, the F1-score also equals 1, signifying perfect classification.

$$\text{F1 score} = 2 \times \frac{Precision * Recall}{Precision + Recall} \qquad (4)$$

RESULTS AND DISCUSSION

The effectiveness of the ensemble method was assessed and compared against existing approaches. The performance of four ensemble classifiers: AdaBoost, GBM, CatBoost, and XGBoost, was systematically evaluated. Table 1 presents the results of these classifiers in sentiment classification of online book reviews. As illustrated in Table 1 and Figure 3, the AdaBoost ensemble classifier obtains the highest accuracy of 86.68%, outperforming the other three ensemble methods as well as the individual base classifiers.

Table 1: Performance of different Ensemble Classifiers with Existing Individual Classifiers on Amazon Datasets

Classifiers	Accuracy	Precision	Recall	F1-Score
AdaBoost	86.98	88	86	87
CatBoost	84.12	86	83	84
GBM	81.34	83	80	81
XGBoost	82.56	84	81	82
Existing models				
RF	78.12	80	77	78
SVM	80.23	82	79	80
NB	85.34	87	84	85

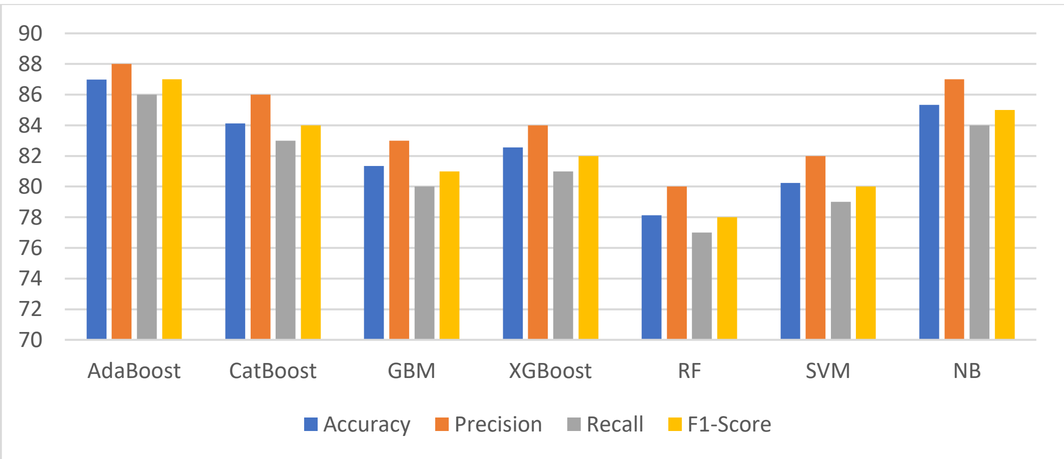

Figure 3: Graphical Representation of Sentiment Classification Performance for Online Book Reviews

Table 2: Performance of Stacking Ensemble Classifiers on Amazon Datasets

Classifiers	Accuracy	Precision	Recall	F1 Score
Stacking AdaBoost	88.23	90	87	88
Stacking CatBoost	86.01	88	85	86
Stacking GBM	84.56	86	83	84
Stacking XGBoost	85.12	87	84	85

The stacked ensemble classifier integrates SVM, NB, and RF as base learners. The classifiers performance was systematically compared, with results presented in Table 2 and Figure 4. As shown, the AdaBoost ensemble classifier performs better than the other classifiers with an accuracy of 88.23%, followed by CatBoost with 86.01%. This improvement is based on the ensemble's ability to combine the individual predictions of RF, SVM, and NB. By aggregating these outputs, the ensemble framework provides input to adaptive, categorical, gradient, and extreme gradient boosting techniques, which leverage the principle of boosting to construct a robust classifier from weaker learners.

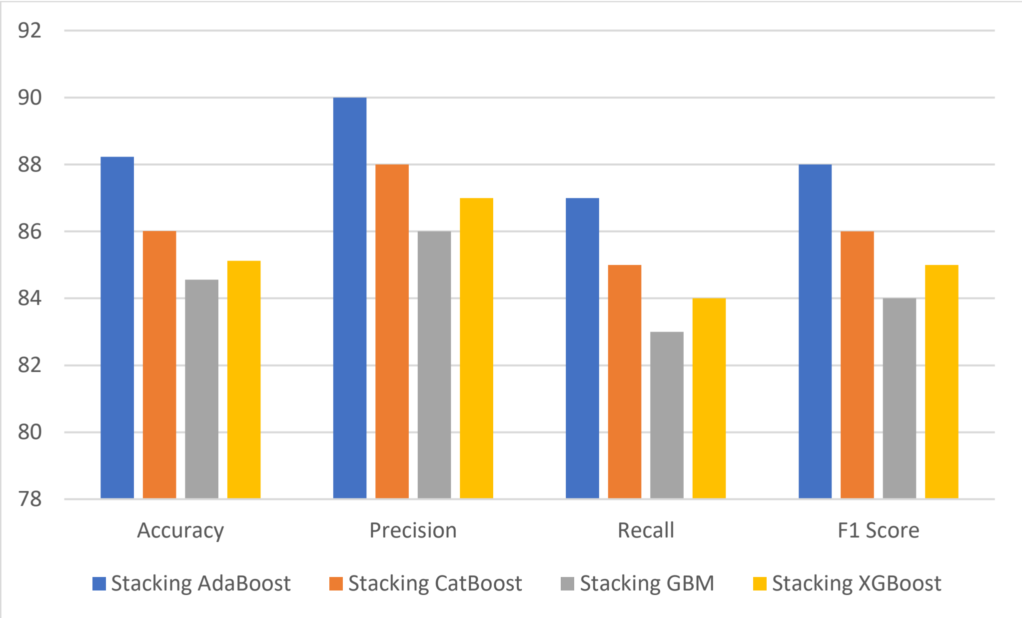

Figure 4: Graphical Representation for Comparison of Ensemble Classifiers for Amazon Online Book Review

The results from the performance of the ensemble classifiers presented in Table 1 and Figure 2 indicate that the Stacked AdaBoost classifier demonstrates the highest overall accuracy at 88.23%, paired with commendable precision (90%), recall (87%), and an F1 score of (88%). These metrics demonstrate that the Stacked AdaBoost classifier not only obtains a high rate of correctly identified true positive instances but also effectively reduces the occurrence of false positives. This makes it particularly well-suited for applications where accurate detection of the positive class is of critical importance. In contrast, the Stacked CatBoost model, while also performing well with a better accuracy of 86.01%, shows a notable drop in recall (85%), indicating a higher proportion of actual positive cases were missed as compared to the CatBoost. Meanwhile, both Stacked GBM and Stacked XGBoost classifiers exhibit lower accuracies at 84.56% and 85.12%, respectively, and their recall values (83% and 84%) indicate challenges in detecting the positive class. This variation in performance highlights the necessity of selecting an appropriate classifier for specific use cases, particularly in contexts where achieving an optimal balance between precision and recall is essential.

The high F1 score achieved by Stacked AdaBoost reflects its effectiveness in maintaining such a balance, essential for applications such as medical diagnostics or fraud detection, where false negatives can have serious consequences. Overall, Stacked AdaBoost stands out as the most reliable choice among the tested classifiers for effectively handling the positive class, thereby highlighting the advantages of utilizing advanced ensemble methods in classification tasks.

CONCLUSION

The sentiment analysis of online book reviews using ensemble learning models produces low scores. So, this paper introduced a stacked ensemble classifier that classifies the sentiment of the book reviews based on their polarities. This research used a wrapper-based technique for selecting the relevant features and the Wrapper technique evaluates a score for each feature. The feature with a low score is ignored and a high score is processed for classification using a stacked ensemble classifier. The proposed stacked ensemble classifier integrates multiple machine learning algorithms, specifically Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB). The individual prediction values gathered from these ML classifiers are fed to the AdaBoost, CatBoost, GBM, and XGBoost techniques that combine the poor classifiers and extract the better prediction value to make a better classifier. The stacked ensemble classifier effectively classifies the sentiment reviews. The developed stacked ensemble model achieves a better accuracy of 88.23 %, 86.01%, 84.56%, and 85.12%, which is comparatively better than the individual ensemble models without stacking. For future direction, the authors will be integrating deep learning architectures, such as LSTMs, CNNs, and transformer-based models within ensemble frameworks to capture both statistical and contextual representations of text. For future research, the authors plan to incorporate deep learning architectures such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and transformer-based models within ensemble frameworks. This integration aims to capture both statistical patterns and contextual representations of text, thereby enhancing the robustness and accuracy of sentiment classification.

REFERENCES

Abheek, N & Preeti, S (2021). Interdisciplinary Research in Technology and Management Proceedings of the International Conference on Interdisciplinary Research in Technology and Management (IRTM, 2021), 26-28 February,2021, Kolkata, India. Doi: https://doi.org/10.1201/9781003430469

Basha, S. M., & Rajput, D, S. (2017). Evaluating the Impact of Feature Selection on Overall Performance of Sentiment Analysis, Proceedings of the 2017 International Conference on Information Technology, pp. 96-102. https://doi.org/10.1145/3176653.3176665

Bhowmic, A., Parua, S., Mandal, D., Ghosh, B. K., and Karmakar, S. (2024) "Leveraging Ensemble Learning for Enhanced Sentiment Analysis," International Conference on Distributed Computing and Optimization Techniques (ICDCOT), Bengaluru, India, 2024, pp. 1-6, Doi: https://doi.org/10.1109/ICDCOT61034.2024.10515983.

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2017). Sentiment analysis and effective Computing: A review. Emotion, Sentiment, and Affect in Human-Computer Interaction, 1–18.

Dave, K.; Lawrence, S.; Pennock, D.M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, 20–24 May 2003; pp. 519–528. DOI: https://doi.org/10.1145/775152.775226

Guru. P. M. S., Praveen. G. J, Rishi, K, Prasanna. K. H. R., Vinay. S. K and Aryan, D (2024), Sentiment Analysis Using Transfer Learning for E-Commerce Websites, International Conference on Advances in Modern Age Technologies for Health and
Engineering Science (AMATHE), Shivamogga, India, 2024, pp. 1-5, doi: https://doi.org/10.1109/AMATHE61652.2024.10582143.

Hussein, D.M.E.D.M. A survey on sentiment analysis challenges. J. King Saud Univ.-Eng. Sci. 2018, 30, 330–338. http://dx.doi.org/10.1016/j.jksues.2016.04.002

Kasper, W., Vela, M. (2011) Sentiment analysis for hotel reviews. In Proceedings of the Computational Linguistics-Applications Conference, Jachranka, Poland, 17–19 October 2011; Volume 231527, pp. 45–52.

Kadu, S., Joshi, B. (2024). ESFMS: Design of an Ensemble Sentiment Analysis Model for Feedback Evaluation via Multimodal Feature Selection Process. In: Nanda, S.J.,

Yadav, R.P., Gandomi, A.H., Saraswat, M. (eds) Data Science and Applications. ICDSA 2023. Lecture Notes in Networks and Systems, vol 818. Springer, Singapore. https://doi.org/10.1007/978-981-99-7862-5_30

Khalid, M., Ashraf, I., Mehmood, A., Ullah, S., Ahmad, M., & Choi, G. S. (2020). GBSVM: Sentiment Classification from Unstructured Reviews Using an Ensemble Classifier. Applied Sciences, 10(8), 2788. Doi: 10.3390/app10082788

Liu, B. (2010). Handbook Chapter: Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing. In Handbook of Natural Language Processing, Marcel Dekker, Inc.: New York, NY, USA, 2009.

Liu, B. (2015). Sentiment analysis: A literature review. Computer Science Review, 5(1), 1- 10.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of Word representations in Vector Space, International Conference on Learning Representations. https://api.semanticscholar.org/CorpusID:5959482

Mukherjee, A., Venkataraman, V.; Liu, B., Glance, N. What Yelp fake review filter might be doing? In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, USA, 8–11 July 2013.

Olufunwa, G, Abdullah**i**, F.B, Belgore, A (2025), Sentiment Classification for Online Book Reviews Using Ensemble Classifiers, TAA International Journal of Science Research And Technology (TIJSRAT), Vol. 8(9), pages 239-259 **DOI:** https://doi.org/10.70382/tijsrat.v08i9.056

Ramdani, F. (2025). Natural Language Processing and Sentiment Analysis. In: Data Science: Foundations and Hands-on Experience. Springer, Singapore. https://doi.org/10.1007/978-981-96-4683-8_9

Srivastava, R., and Bhatia, M. P. S. (2016), "Ensemble methods for sentiment analysis of online micro-texts," *2016* International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2016, pp. 1-6, Doi:
https://doi.org/10.1109/ICRAIE.2016.7939525.

Tan, K. L., Lee, C. P., Lim, K. M., and Anbananthen, K. S. M. (2022) "Sentiment Analysis With Ensemble Hybrid Deep Learning Model," in IEEE Access, vol. 10, pp. 103694-103704, 2022, Doi: https://doi.org/10.1109/ACCESS.2022.3210182.

Thien, K. T. & Phan, T. T, (2019). Deep Learning Application to Ensemble Learning: The Simple. but Effective, Approach to Sentiment Classifying, Applied Science, 9(13), 2760 https://doi.org/10.3390/app9132760

Tiwari, D., Nagpal, B., Bhati, B.S. (2023). A Systematic Review of Social Network Sentiment Analysis with Comparative Study of Ensemble-based Techniques. Artificial Intelligence Review, 56, 13407–13461, https://doi.org/10.1007/s10462-023-10472-w

Usman, H, Abubakar, S. H., and Idi, M**.** (2026), An Ensemble Explainability Framework for Multimodal Chest X-ray Disease, UDMA Journal of Science (FJS), Vol. 10(3), pages 389-396, DOI: https://doi.org/10.33003/fjs-2026-1003-4826

Victor U. A, Olalekan J. A, Lasisi A. M, Saifulahi S. S, Amina, A, Catherine A. A, Badamasi A. M and Ejima, U I. (2026), Enhancing Ransomware Classification Using

Lightweight Machine Learning Algorithm and Ensemble Methods, FUDMA Journal of Science (FJS), Vol. 10(5), pages 2645-2914, DOI: https://doi.org/10.33003/fjs-2026-1005- 5000

Vidyashree, K P., Rajendra, A B., Gururaj, H. L., Vinayakumar, R, Moez, K (2024). A tweet Sentiment classification approach using an ensemble classifier, International Journal of Cognitive Computing in Engineering, Vol. 5, pp 170-177, ISSN 2666-3074, https://doi.org/10.1016/j.ijcce.2024.04.001.

Yadav, V., & Singh, M. (2020). Sentiment analysis of social media data using word embeddings and classifier algorithms. International Journal of Computer Applications, 975, 8887

Zhao, Y., Qin, B., & Liu, T. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery DOI: https://doi.org/10.1002/widm. 1253

Zhang, L., Wang, S., and Lui, B (2018). Deep Learning for Sentiment Analysis: A Survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol 8, pp. 1-34 https://api.semanticscholar.org/CorpusID:10694510

Zhenbing, H., Ivan, D., Kateryna, P., Vasyi, M. (2024), Augmenting Sentiment Analysis Prediction in Binary Text Classification through Advanced Natural Language Processing Models and Classifiers, International Journal of Information and Computer Science (IJITCS) Vol. 16(2), pp. 16-31, DOI: https://doi.org/10.5815/ijitcs.2024.02.02