International Journal for Modern Trends in Science and Technology
Volume 12, Special Issue 01, pages 815-820.
ISSN: 2455-3778 online
Available online at: http://ijmtst.com/ICRCSET2026.html
DOI: https://doi.org/10.5281/zenodo.19613135

Smart Review Analyser: A Multiplatform AI System for Aspect-Based Sentiment Analysis on E-Commerce Products

Challa Vasishta, Gochipathala Yona, Danaboyina Triveni, Desireddy Vinodreddy, J. Sudhir Kumar

Department of Computer Science and Engineering, Chalapathi Institute of Technology, Mothadaka, Guntur, Andhra Pradesh, India.

To Cite this Article
Challa Vasishta, Gochipathala Yona, Danaboyina Triveni, Desireddy Vinodreddy & J. Sudhir Kumar (2026). Smart Review Analyser: A Multiplatform AI System for Aspect-Based Sentiment Analysis on E-Commerce Products. International Journal for Modern Trends in Science and Technology, 12(SI01), 815-820. https://doi.org/10.5281/zenodo.19613135

Article Info
Received: 12 March 2026; Revised: 07 April 2026; Accepted: 10 April 2026.

KEYWORDS

Sentiment Analysis, Amazon Product Reviews, Machine Learning Models, TF-IDF Vectorizer, Aspect-Based Sentiment Classification

ABSTRACT

The rapid expansion of user-generated content on e-commerce platforms like Amazon has heightened the importance of sentiment analysis for understanding consumer behavior and guiding business strategies. Traditional methods of textual analysis have progressed from basic keyword matching to advanced machine learning models, yet existing systems often emphasize fine-tuning specific architectures rather than comparing diverse approaches for overall sentiment classification. This project presents the Systematic Algorithmic and Feature Representation Comparison (SAFReC) Framework, designed to evaluate multiple machine learning models for sentiment classification of Amazon product reviews. The framework integrates models such as Bi-LSTM, Cat Boost, XG Boost, and Light GBM with two feature representation techniques Count Vectorizer and TF-IDF Vectorizer to optimize text preprocessing and vectorization. Using the Amazon Product Reviews dataset, the framework systematically compares the models' performance to identify the most effective approach for aspect-based sentiment analysis. The results provide actionable insights for businesses to better assess customer opinions and enhance product offerings.

I. INTRODUCTION

In today's rapidly evolving digital marketplace, e-commerce platforms have transformed the way consumers purchase products and share their experiences. Customer reviews have become a powerful influence on purchasing decisions, brand perception, and competitive positioning. Platforms such as Amazon and Flipkart host millions of user-generated reviews that reflect customer satisfaction, complaints, suggestions, and overall product experiences. These reviews provide valuable insights for businesses seeking to improve product quality, enhance customer service, and strengthen brand loyalty. However, the massive volume of reviews generated daily makes manual analysis

inefficient, time-consuming, and impractical. As a result, automated sentiment analysis systems have become essential for extracting meaningful information from unstructured textual data.

Sentiment analysis, a subfield of Artificial Intelligence (AI) and Natural Language Processing (NLP), focuses on identifying emotions, opinions, and attitudes expressed in textual content. Over the years, sentiment analysis techniques have evolved from simple keyword-based approaches to advanced machine learning and deep learning models capable of understanding contextual and nuanced expressions. Algorithms such as Bi-LSTM, Cat Boost, XG Boost, and Light GBM have demonstrated strong performance in text classification tasks. Additionally, feature representation techniques like Count Vectorizer and TF-IDF Vectorizer play a crucial role in converting textual reviews into numerical formats suitable for machine learning models.

Despite significant advancements, many existing studies focus on optimizing or fine-tuning individual models rather than systematically comparing multiple machine learning paradigms and feature representation methods. There is a need for a comprehensive framework that evaluates diverse algorithms and identifies the most effective approach for sentiment classification in e-commerce reviews. To address this gap, the proposed project introduces the Systematic Algorithmic and Feature Representation Comparison (SAFReC) Framework. This framework integrates multiple machine learning models with different text vectorization techniques to assess their comparative performance on Amazon product reviews.

By providing a structured evaluation of various algorithms and feature engineering methods, the SAFReC framework aims to deliver accurate sentiment classification and actionable business insights. The system not only detects positive and negative sentiments but also supports aspect-based analysis, enabling companies to understand specific product attributes influencing customer opinions. Ultimately, this research contributes to improved decision-making processes in e-commerce environments while advancing methodological understanding in sentiment analysis.

2. LITERATURE SURVEY

Sentiment analysis, particularly on e-commerce platforms like Amazon, has become a critical area of research due to the rapid growth of user-generated content in the form of product reviews. Over the years, sentiment analysis techniques have evolved from simple rule-based methods to more sophisticated machine learning models. The aim of this literature review is to explore key research papers that have contributed to the field of sentiment analysis, particularly in the context of Amazon product reviews, and to highlight the different approaches and challenges associated with them.

Early studies in sentiment analysis were predominantly based on lexicon-based or rule-based approaches. For example, Turney (2002) introduced the use of semantic orientation to classify sentiment in text. Turney's method used a set of predefined lists of positive and negative words and their respective frequencies to determine the sentiment of a sentence. While this approach was effective for simple sentiment classification, it struggled with handling complex language constructs like sarcasm, irony, and negations. Moreover, it required extensive manual effort to create and maintain word lists, limiting its scalability for large datasets like Amazon reviews.

With the rise of machine learning, a shift toward supervised learning models for sentiment classification emerged. Pang et al. (2002) conducted seminal work on sentiment analysis by applying Naive Bayes, Support Vector Machines (SVM), and Maximum Entropy to classify movie reviews as positive or negative. This approach marked a significant improvement over lexicon-based methods, as it allowed models to learn from labeled data and capture more complex relationships between words and sentiment. However, these models had limitations in handling contextual information, and they struggled with ambiguities, such as distinguishing between positive sentiment in one context and negative sentiment in another. This challenge became more pronounced with product reviews, where the context and phrasing could vary significantly.

3 SYSTEM ARCHITECTURE

The system architecture of the SMART REVIEW ANALYSER is designed as a streamlined, end-to-end pipeline for sentiment analysis of Amazon product reviews. It begins with data collection, where reviews are gathered from datasets or APIs, followed by data preprocessing, which cleans and prepares the text through tokenization, stop word removal, stemming or lemmatization, and balancing class. The processed text is

then transformed into numerical representations using feature extraction techniques like Count Vectorizer and TF-IDF. These features are fed into multiple machine learning models—BiLSTM, CatBoost, XGBoost, and LightGBM—for training. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1 score, and the best-performing model is selected for deployment. In the real-time sentiment analysis phase, new reviews are classified using this model. To ensure the system remains accurate and up-to-date, a retraining mechanism periodically incorporates new data into the model, forming a continuous cycle of learning and improvement.

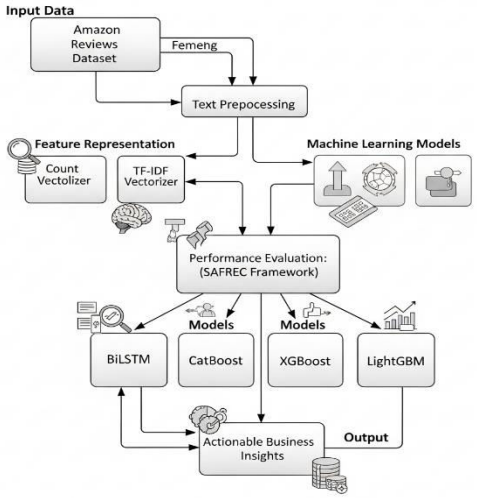

Fig1: System Architecture

4.METHODOLOGY

The proposed Smart Review Analyzer is developed using the Systematic Algorithmic and Feature Representation Comparison (SAFReC) Framework to evaluate multiple machine learning models and feature extraction techniques for sentiment classification of Amazon product reviews. The methodology consists of structured stages including data acquisition, preprocessing, feature engineering, model development, evaluation, and comparative analysis.

i. Data Collection and Preparation

The Amazon Product Reviews dataset is used as the primary data source. It contains review text, ratings, and sentiment labels. The dataset is first cleaned to remove duplicate entries and incomplete records. The data is then divided into training and testing sets using an 80:20 split to ensure unbiased performance evaluation.

ii. Text Preprocessing

Preprocessing improves data quality and reduces noise. The steps include converting text to lowercase, removing punctuation, stop words, and special characters, followed by tokenization and lemmatization. This ensures consistent textual representation before vectorization.

iii. Feature Representation

To transform textual data into numerical form, two techniques are applied:

(a) Term Frequency (TF):

$$TF(t,d) = \frac{f(t,d)}{\sum_k f(k,d)}$$

Where $f(t,d)$is the frequency of term t in document d.

(b) TF-IDF Weighting:

$$TFIDF(t,d) = TF(t,d) \times \log\left(\frac{N}{df(t)}\right)$$

Where N is the total number of documents and $df(t)$is the document frequency of term t.

Count Vectorizer and TF-IDF Vectorizer are implemented to compare representation effectiveness.

iv. Model Development

Multiple machine learning models are trained using both feature sets, including BiLSTM (deep learning), CatBoost, XGBoost, and LightGBM. For classification, the general prediction function is expressed as:

(c)Classification Function:

$$\hat{y} = f(X;\theta)$$

Where Xrepresents feature vectors and θdenotes learned model parameters.

v. Model Evaluation and Comparison

Models are evaluated using Accuracy, Precision, Recall, and F1-score. Cross-validation ensures robustness and prevents over fitting. Finally, the SAFReC framework systematically compares all model–feature combinations to identify the optimal configuration for sentiment classification. The best-performing model is selected for deployment, enabling businesses to extract actionable insights and enhance decision-making processes in e-commerce environments.

5. DESIGN AND CONSTRUCTION

The design of the Smart Review Analyzer is structured around a modular and scalable architecture that integrates data processing, machine learning, and visualization components. The system is designed to automatically collect, preprocess, analyze, and classify customer reviews from e-commerce platforms. It follows

a layered architecture consisting of a data input layer, preprocessing layer, feature extraction layer, model training and prediction layer, and a results visualization layer. This structured design ensures flexibility, easy integration with other platforms, and efficient handling of large volumes of review data.

The construction phase begins with implementing the data preprocessing module, where raw textual reviews are cleaned and standardized. Next, feature representation techniques such as Count Vectorizer and TF-IDF Vectorizer are constructed to convert text into numerical vectors. These vectors are then fed into multiple machine learning models including BiLSTM, CatBoost, XGBoost, and LightGBM. The SAFReC framework is integrated to systematically compare algorithmic performance. Finally, a user-friendly dashboard is developed to display sentiment classifications, graphical summaries, and key insights. The system is built to support scalability, real-time analysis, and future expansion with advanced deep learning or transformer-based models for enhanced sentiment prediction accuracy.

6.RESULTS AND DISCUSSION

The SMART REVIEW ANALYSER demonstrates effective performance in classifying Amazon product reviews into positive, negative, and neutral sentiments using multiple machine learning and deep learning models. Experimental results show that all evaluated models Bi-LSTM, Cat Boost, XG Boost, and Light GBM successfully perform sentiment classification, with varying levels of accuracy and efficiency. Among them, Bi-LSTM achieves the best performance due to its ability to capture contextual relationships and long-range dependencies in textual data.

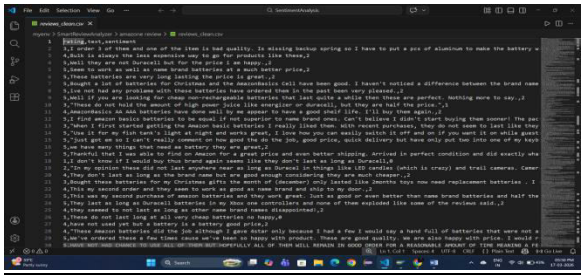

Fig 2: Amazon Reviews Dataset

The dataset used for training is presented in Figure 2, which contains a large collection of customer reviews. Proper preprocessing and feature extraction techniques such as TF-IDF and Count Vectorizer are applied to convert text into meaningful numerical representations. The results indicate that TF-IDF slightly outperforms Count Vectorizer in capturing word importance and improving classification accuracy.

The model development and hybrid approach where a combination of ALBERT and Bi-LSTM is used to enhance contextual understanding. This hybrid model improves sentiment detection by leveraging both transformer-based embedding's and sequential learning capabilities.

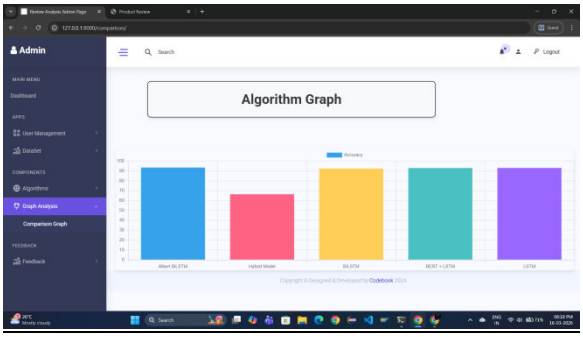

Fig 3: Model Comparison Graph

The performance comparison is illustrated in Figure 3, where Bi-LSTM achieves the highest accuracy, while Light GBM and XG Boost demonstrate faster execution and efficiency on large datasets. Cat Boost also provides competitive performance with balanced results.

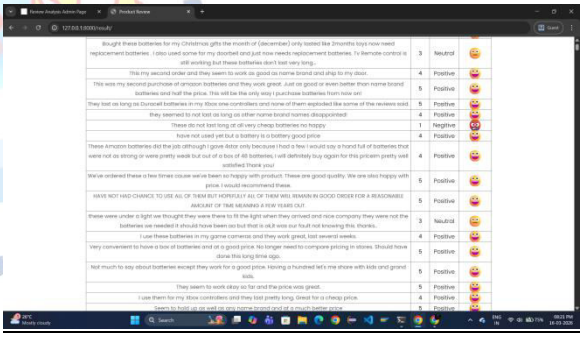

Fig 4: Sentiment Analysis Output

The final prediction output is depicted in Figure 4 where customer reviews are classified into positive, negative, or neutral categories. This enables businesses to quickly analyze customer feedback and gain insights into product performance.

Overall, the system achieves high accuracy and efficiency in sentiment classification. The combination of advanced preprocessing, TF-IDF feature extraction, and hybrid deep learning models significantly enhances performance. These results confirm that the Smart Review Analyser is a reliable and scalable solution for real-time sentiment analysis, helping e-commerce platforms improves decision-making and customer satisfaction.

7. CONCLUSION

The sentiment analysis system developed for Amazon product reviews leverages multiple machine learning models such as Bi-LSTM, Cat Boost, XG Boost, and Light GBM to efficiently classify reviews into positive, negative, or neutral categories. The system employs various preprocessing techniques to clean and transform the raw data, followed by feature extraction using Count Vectorizer and TF-IDF. After training and evaluating these models, the best-performing one is selected for real-time sentiment analysis. This allows businesses to gain actionable insights into customer feedback, helping them improve their products and services. The system's accuracy, scalability, and real-time capabilities make it a powerful tool for sentiment analysis of large-scale, unstructured text data.

FUTURE SCOPE

- **Multilingual Support**: Extend the system to analyze reviews in multiple languages, making it more adaptable to global e-commerce platforms.

Aspect-Based Sentiment Analysis: Implement aspect-based sentiment analysis to identify and analyze sentiments regarding specific product features, providing more detailed insights.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

1. K. Mokgwatjane, T. Paepae, "An explainable ensemble machine learning approach for multi-domain, multiclass sentiment analysis in Amazon product reviews," Machine Learning with Applications, 2025.

2. I.M. Abbas, İ. Atacak, S. Toklu, N. Barışçı, İ.A. Doğru, "A Hybrid Deep Learning Model Based on Local and Global Features for Amazon Product Reviews: An Optimal ALBERT-Cascade CNN Approach," Applied Sciences, 2025.

3. Z. Jiang, C. Xu, "A Multi-Scale Feature Fusion Linear Attention Model for Movie Review Sentiment Analysis," Big Data and Cognitive Computing, 2025.

4. K. Balasaranya, P. Ezhumalai, "Sentiment analysis of Amazon product reviews using an inception-based recurrent residual CNN approach," International Journal of Data Science and Engineering, 2025.

5. M.H. Hosen, Y. Haque, D.C. Nijhum, "XLM-RoBERTa-LIME: A Novel Explainable Framework for Sentiment Analysis on Bangla–English–Hindi Code-Mixed Text," Proceedings of the 12th International Conference, 2025.

6. M. Kumar, L. Khan, M.Z. Khan, A.A. Alhussan, "AGF-HAM: Adaptive Gated Fusion Hierarchical Attention Model for Explainable Sentiment Analysis," Mathematics, 2025.

7. K. Natarajan, V. Nirmalrani, S. Gowri, "Leveraging Word2Vec-Enhanced CNN-LSTM Hybrid Architecture for Sentiment Analysis in E-Commerce Product Reviews," International Journal of Electronics and Computer Science Engineering, 2026.

8. I. Duru, A.S. Sunar, "Transformer and Pre-Transformer Model-Based Sentiment Prediction with Various Embeddings:

9. B. Borhan, Y. Ortakci, A. Lathigara, "Comparative Performance of Machine Learning Algorithms for Sentiment Analysis:

10. J.J.H. Hu, F. Ahmad, M. Bader-El-Den, "Decoding brand sentiments: Leveraging customer reviews for insightful brand perception analysis using natural language processing and Tableau," PLoS One, 2025.

11. G. Liste, M. Plumed Lasarte, "Milk labelling and consumer opinion: a data-driven analysis from e-commerce reviews," British Food Journal, 2025.

12. M.S. Rahdi, F.Z. Boroujeni, "Enhancing Recommendation Systems with Autoencoder-SVD and Transformer-Based Summarization: A Sentiment-Aware Approach Using GPT-2 and VADER," Qubahan Academic Journal, 2025.

13. J. Mondal, "Practical application of Sentiment Analysis," OER Commons, 2025.

14. P. Sharma, S. Susan, "Holistic Evaluation of 7B-Scale LLMs for Sentiment Analysis: Accuracy Meets Explainability," ResearchGate, 2025.

15. N.N.I. Prova, V. Ravi, M.P. Singh, V.K. Srivastava, "Multilingual sentiment analysis in e-commerce customer reviews using GPT and deep learning-based weighted-ensemble model," International Journal of Applied Engineering Research, 2025.

16. M. Reddy, C.H. Srilakshmi Kaza, A. Sri, "Cuda-Enhanced Yolo Feature Extraction with Rf–Catboost Ensembles For High-Performance Product Review Sentiment,"

17.K. M S, R. R. G and S. Karthik, "Streamlining Load Scheduling in Cloud Computing: A Thorough Performance Assessment and Development of Effective Methods for Design," 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), Shivamogga, India, 2024, pp. 1-7, doi: 10.1109/AMATHE61652.2024.10582239.

18. Sai Srinivas Vellela, Roja D, NagaMalleswara Rao Purimetla, SyamsundaraRao Thalakola, Lakshma Reddy Vuyyuru, Ramesh Vatambeti, Cyber threat detection in industry 4.0: Leveraging GloVe and self-attention mechanisms in BiLSTM for enhanced intrusion detection, Computers and Electrical Engineering,Volume 124, Part A,2025,110368,ISSN 00457906,https://doi.org/10.1016/j.compeleceng.2025.110368.

19.S. S. Vellela, L. R. Vuyyuru, K. B. S. K, N. MalleswaraRaoPurimetla, L. Dalavai and M. V. Rao, "A Novel Approach to Optimize Prediction Method for Chronic Kidney Disease with the Help of Machine Learning Algorithm," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1677-1681, doi: 10.1109/IC3I59117.2023.10397974.

20.Kavitha Mettupalayam Subramaniam, Ramachandra Rao Goli, Karthik Subburathinam, Srihari Kannan,Optimization of pyrolysis parameters for enhanced biochar production from agricultural biomass: A study on energy efficiency and carbon sequestration

potential,Science of The Total Environment, Volume 1015,2026,181362,ISSN 00489697,https://doi.org/10.1016/j.scitotenv.2026.181362.

21.K. K. Kumar, S. G. B. Kumar, S. G. R. Rao and S. S. J. Sydulu, "Safe and high secured ranked keyword searchover an outsourced cloud data," *2017 International Conference on Inventive Computing and Informatics (ICICI)*, Coimbatore, India, 2017, pp. 20-25, doi: 10.1109/ICICI.2017.8365348.

22.R. K. Yarava, G. R. C. Rao, Y. Garapati, G. C. Babu and S. D. V. Prasad, "Analysis on the Development of Cloud Security using Privacy Attribute Data Sharing," *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichy, India, 2022, pp. 1-5, doi: 10.1109/ICEEICT53079.2022.9768608.

23. K. K. . Kommineni and A. . Prasad, "A Review on Privacy and Security Improvement Mechanisms in MANETs", Int J Intell Syst Appl Eng, vol. 12, no. 2, pp. 90–99, Dec. 2023.

24.Kommineni, K.K., Prasad, A. Enhancing Data Security and Privacy in SDN-Enabled MANETs Through Improved Data Aggregation Protection and Secrecy. Wireless Pers Commun 139, 855–882 (2024). https://doi.org/10.1007/s11277-024-11635-w

25."Blockchain-Enabled Secure Data Aggregation for SDN-Enabled Ad-Hoc Networks," International Journal of Intelligent Engineering and Systems, vol. 18, no. 5, pp. 704–717, Jun. 2025, doi: https://doi.org/10.22266/ijies2025.0630.49.

26.K. K. Kommineni, P. Ande, "Blockchain-driven key management and privacy-preserving data Aggregation Scheme for SDN-enabled MANETs," International Journal of Intelligent Engineering and Systems, vol. 18–18, no. 9, pp. 601–615, 2025, doi: 10.22266/ijies2025.1031.39.

27.P. Anusha and J. R. Babu, "Enhancing Radiographic Diagnosis: A Novel AI-based Bone Fracture Detection System," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2025, pp. 1262-1266, doi: 10.1109/ICSCDS65426.2025.11167456.

28. K. N. Rao, B. R. Gandhi, M. V. Rao, S. Javvadi, S. S. Vellela and S. Khader Basha, "Prediction and Classification of Alzheimer's Disease using Machine Learning Techniques in 3D MR Images," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 85-90, doi: 10.1109/ICSCSS57650.2023.10169550.

29. S. S. Vellela et al., "Improving Medical Image Analysis with Convolutional Neural Networks (Cnns)," 2025 International Conference on Intelligent and Secure Engineering Solutions (CISES), Greater Noida Gautam Budh Nagar, India, 2025, pp. 579-584, doi: 10.1109/CISES66934.2025.11265231.

31. V. Khedkar, N. Vullam, J. R. Babu, U. Bhagyalatha, S. Babu Vadde and A. Lakshmanarao, "Hybrid Classification Approach for Heart Disease using Few Shot Inspired Machine Learning Models," 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2025, pp. 01-05, doi: 10.1109/ICICACS65178.2025.10968965