International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

SEMANTIC-SBERT: LEVERAGING SENTENCE TRANSFORMERS FOR GRANULAR PRODUCT REVIEW SENTIMENT ANALYSIS

K. Balakrishna[1], B. Poojitha[1], K. Chiranjeevi[1], N. Siva Nagamani[2]
[1]Assistant Professor, [2]Associate Professor, [1,2]Department of Computer Science and Engineering (AI & ML)
[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

ABSTRACT
The rapid expansion of the global e-commerce ecosystem, projected to exceed USD 7 trillion by 2030, has significantly increased the influence of customer reviews on consumer decision-making, with a majority of users relying on online feedback before making purchases. However, extracting meaningful insights from large-scale review data remains challenging due to its volume, variability, and time-sensitive nature. Manual evaluation is inefficient and inconsistent, while many existing approaches struggle to capture contextual semantics and often perform poorly on imbalanced datasets. To address these challenges, this study proposes an advanced framework based on Natural Language Processing (NLP) for automated sentiment classification of product reviews using annotated datasets. The workflow begins with structured preprocessing and Exploratory Data Analysis (EDA) to clean and analyze the data for meaningful patterns. Sentence Bidirectional Encoder Representations from Transformers (SBERT) is used to generate context-aware embeddings that effectively capture semantic relationships. To handle class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to create a balanced training distribution, improving model robustness. In contrast to conventional models such as Random Forest Classifier (RFC), Light Gradient Boosting Machine (LGBM), and Extreme Gradient Boosting (XGB), the proposed framework integrates Deep Neural Network (DNN)-based feature extraction with a Boosted Rules Classifier (BRC) to enhance both prediction accuracy and interpretability. The system classifies reviews into Negative, Neutral, and Positive sentiments, enabling a deeper understanding of customer opinions. Additionally, the trained model is deployed using Django, a Python-based web framework, allowing users to upload data, generate predictions in real time, and manage interactions through an intuitive interface. Experimental results demonstrate that the proposed system achieves higher accuracy, improved scalability, and reduced bias, making it a reliable solution for extracting actionable insights and supporting data-driven decision-making.

Keywords: E-commerce, Text Classification, Natural Language Processing, BERT architecture, Boosted rules classifier, Synthetic Minority Over-sampling Technique.

1. INTRODUCTION
Over the past decade, the volume of product reviews on online retail platforms has grown rapidly, contributed by both individual users and professional reviewers. These reviews have become a key source of information that helps reduce uncertainty during purchasing decisions. A significant proportion of consumers rely on customer feedback before buying a product, with studies reporting that close to 90% of users consider reviews an essential part of their decision-making process [1]. As a result, reviews now hold substantial importance for both consumers and businesses. For consumers, they provide insights into product quality and usability, while for businesses, they influence brand perception, customer trust, and overall sales performance [2]. On large e-commerce platforms such as Amazon, product reviews play a crucial role in guiding purchasing behavior by offering authentic user experiences and detailed evaluations. However, the sheer volume of reviews makes it difficult for users to manually analyze them efficiently. Reviews are equally valuable for sellers, as they help in

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

understanding customer preferences, improving product quality, and enhancing marketing strategies. Additionally, product ratings associated with reviews simplify decision-making by summarizing overall customer satisfaction and shaping the perception of a product [3]. The significance of online reviews increased notably during the COVID-19 pandemic, when digital transactions surged due to restrictions on physical shopping. This transition led to a considerable rise in the number of online reviews, reinforcing their role in e-commerce ecosystems. Reviews and ratings provide a convenient way for customers to evaluate products quickly, while businesses leverage this data to gain insights into customer expectations and improve their offerings [4].

With the expansion of online platforms and social media, the volume of user-generated content has grown exponentially, creating challenges such as information overload and the presence of misleading or biased reviews. This makes it increasingly difficult for consumers to differentiate between genuine feedback and promotional content. To address these challenges, researchers have applied Machine Learning (ML) techniques to analyze large-scale review data and predict review usefulness. These approaches focus on identifying meaningful patterns and extracting actionable insights from complex datasets [5].

Given the continuous generation of massive amounts of review data, manual sentiment interpretation is no longer practical. Studies indicate that up to 95% of consumers read online reviews before making purchases, and businesses with higher ratings often experience noticeable improvements in sales performance [6]. Therefore, there is a growing demand for automated and scalable sentiment analysis solutions. Such systems enable organizations to efficiently interpret customer opinions, evaluate product performance, and identify areas for improvement. Consequently, sentiment analysis has become an essential component in modern e-commerce, supporting informed decision-making and enhancing customer trust.

2. LITERATURE SURVEY

Recent advancements in sentiment analysis (SA) have evolved from traditional machine learning approaches to hybrid deep learning architectures, transformer-based models, and context-aware embeddings. Researchers are increasingly focusing on improving contextual understanding, handling complex textual data, and developing domain-specific solutions.

2.1 Hybrid and Deep Learning-Based Models

Several studies have proposed hybrid architectures that combine multiple deep learning techniques to enhance sentiment classification performance. Ahanin et al. [8] introduced a hybrid feature extraction model that integrates human-engineered features with deep learning models such as Bi-LSTM and BERT. Their approach improves multi-label emotion classification by leveraging both contextual embeddings and lexical knowledge. Tan et al. [10] proposed a hybrid RoBERTa-GRU model, where RoBERTa captures contextual semantic representations and GRU models sequential dependencies. Their model achieved high accuracy across multiple benchmark datasets and demonstrated robustness in handling imbalanced data through augmentation. Trisna et al. [12] developed a fusion-based text representation model combining GloVe and Word2Vec embeddings with a biGRU classifier. This approach enhances contextual understanding while maintaining computational efficiency, achieving strong F1-scores in sentiment classification tasks.

2.2 Transformer Models and Large Language Models

Transformer-based models and large language models (LLMs) have significantly improved sentiment analysis capabilities. Serna et al. [6] applied deep learning techniques by fine-tuning a pretrained language model on transport-related user-generated content. Their approach demonstrated robustness in processing large-scale opinion data and highlighted its usefulness in sustainability analysis. Ghatora et al. [7] conducted a comparative study between traditional machine learning models and GPT-based

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

large language models for sentiment analysis. Their findings showed that while traditional models perform well on short texts, LLMs outperform them in analyzing complex and context-rich data, providing higher precision and better detection of mixed sentiments. Oprea et al. [13] focused on fine-grained emotion classification using DistilBERT. Their two-phase approach—feature extraction followed by fine-tuning—resulted in improved performance, particularly for informal and noisy customer review data.

2.3 Advanced Embedding and Context-Aware Approaches

Improving word representations to capture both semantic and emotional context has been a key research focus. Mao et al. [11] proposed sentiment-aware word embeddings that integrate emotional lexicons with traditional embeddings, enhancing both interpretability and classification accuracy. Chu et al. [9] introduced a Complex-valued Quantum-enhanced LSTM (CQLSTM), which incorporates concepts from quantum theory to model interactions between words and sentences. This method improves the model's ability to capture deeper contextual relationships that are often missed by conventional approaches.

2.4 Domain-Specific Applications and Data-Centric Approaches

Domain-specific sentiment analysis plays a crucial role in practical applications. Serna et al. [6] focused on the transport domain by creating an annotated dataset for sustainability-related sentiment analysis, demonstrating the importance of domain-specific corpora in improving classification performance. Additionally, recent studies emphasize the importance of data representation and model adaptation techniques such as embedding fusion, lexicon integration, and fine-tuning strategies, which significantly enhance the robustness and generalization capability of sentiment analysis models across different domains.

3. PROPOSED METHODOLOGY

The proposed system architecture integrates a complete NLP-driven sentiment analysis pipeline with a web-based deployment layer for real-time predictions and management. It starts with structured data ingestion and preprocessing, followed by feature extraction using SBERT to capture contextual semantics. The architecture incorporates class balancing using SMOTE to handle data imbalance effectively. Multiple models including RFC, LGBM, XGB, and a DNN-based classifier are trained and compared to identify optimal performance. The final prediction layer combines deep feature extraction with BRC to ensure both high accuracy and interpretability. As depicted in Fig. 1, the system provides an end-to-end flow from raw data processing to deployment through a Django-based interface. The inclusion of evaluation metrics and visualization modules further strengthens performance monitoring and model validation.

Step 1: Data Input and Preprocessing: The system begins by accepting CSV input data, which undergoes preprocessing steps such as tokenization, stopword removal, lemmatization, and POS tagging. Text normalization ensures consistency and removes irrelevant noise from the dataset. Label encoding is applied to transform categorical sentiment labels into numerical format. This stage prepares structured and clean data for further processing.

Step 2: EDA and Visualization: EDA is performed to analyze the dataset and identify important patterns and distributions. Visualization techniques such as word frequency plots, POS tag distributions, and document-level statistics are generated. These insights help in understanding feature importance and class imbalance. It also guides decisions for feature engineering and model selection.

Step 3: Feature Extraction and Class Balancing: SBERT is used to generate high-quality contextual embeddings that capture semantic meaning effectively. These embeddings are further processed and cached for efficient reuse. SMOTE is applied to balance the dataset by generating synthetic samples for minority classes. This step improves model robustness and reduces bias toward dominant classes.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

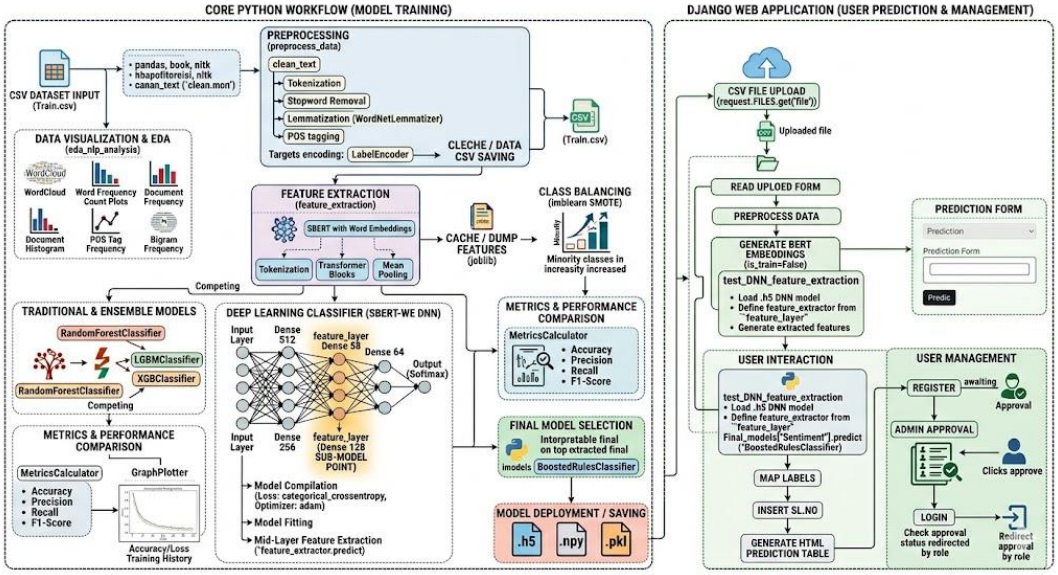


Fig. 1: System architecture of product sentiment analysis with SBERT word embeddings.

Step 4: Model Training and Comparison: Multiple models including RFC, LGBM, XGB, and a DNN classifier are trained using the extracted features. The DNN performs deep feature learning through multiple dense layers and dropout mechanisms. Each model is evaluated using metrics such as accuracy, precision, recall, and F1-score. This comparison helps in selecting the most effective approach.

Step 5: Final Model Selection and Deployment: The best-performing model is integrated with BRC to enhance interpretability and decision-making. The finalized model is saved in formats such as .h5, .npy, and .pkl for deployment. This ensures flexibility and efficient reuse in different environments. The deployment-ready model supports scalable and reliable predictions.

Step 6: Web Application and User Interaction: A Django-based web application enables users to upload input data and obtain sentiment predictions in real time. The system processes the input, generates embeddings, and applies the trained model pipeline for classification. It also includes user management features such as registration, admin approval, and login handling. This ensures a user-friendly interface and seamless system interaction.

3.1 Boosted Rules Classification

The BRC is a hybrid, rule-based ensemble method designed to enhance classification performance, particularly for imbalanced datasets. In the proposed methodology, it is applied to the DNN-extracted features to improve sentiment prediction by combining interpretable rules with boosting techniques.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

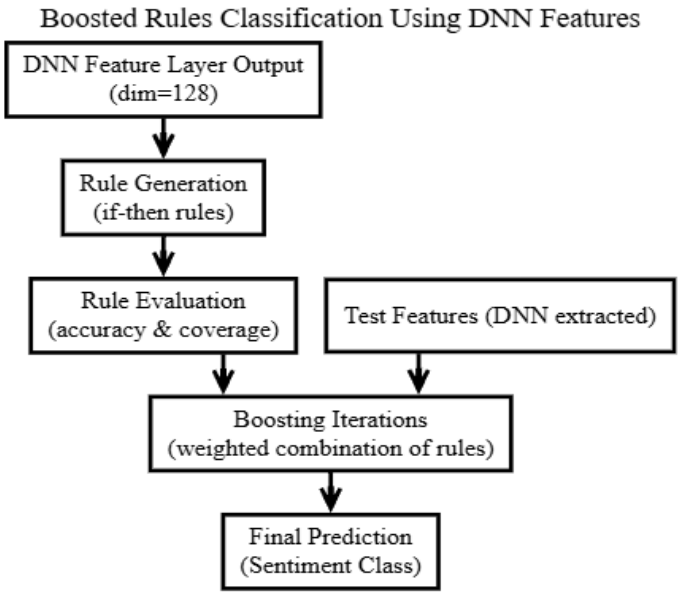


Fig. 2: Internal operations of BRC on DNN-extracted features.

Internal Operation

1. **Input Features:** Receives 128-dimensional features extracted from the DNN feature_layer. These features are compact, discriminative, and optimized for sentiment classification.
2. **Rule Generation:** The algorithm automatically generates a set of if-then rules from the input features. Each rule identifies patterns in the feature space that correspond to specific sentiment classes.

   Example: "If feature_12 > 0.5 and feature_45 < 0.3 → class = Positive."
3. **Rule Evaluation:** Each generated rule is evaluated based on its classification accuracy and coverage over the training data. Only rules meeting minimum thresholds for support and confidence are retained.
4. **Boosting Iterations:** Weak rules are combined into a strong classifier using boosting. Misclassified samples in each iteration are given higher weight, guiding the generation of new rules in subsequent iterations. This iterative process continues until the ensemble achieves optimal performance or reaches a predefined number of boosting rounds.
5. **Prediction:** For each test sample, all rules in the ensemble are evaluated. Each rule casts a weighted vote for a class based on its boosting score. The final class label is assigned based on the aggregated weighted votes.

4. RESULTS AND DISCUSSION

The dataset used for this research consists of product-related textual data collected from online platforms, encompassing a total of 6,364 records**.** Each record corresponds to a single product review or post and contains multiple attributes relevant for sentiment analysis. The primary columns in the dataset are Text_ID, Product Description, Product Type, and Sentiment**.** The Text_ID serves as a unique identifier for each record, ensuring traceability and proper indexing of the dataset. The Product Description column contains the actual review text or description of the product, which can range from a few words to multiple sentences. This textual data forms the core input for natural language processing and is used to generate embeddings for model training. The Product Type column categorizes products into different classes, which can be leveraged for downstream analysis such as product-specific sentiment trends. Finally, the Sentiment column encodes the sentiment associated with each review, with numeric labels representing categories such as Positive, Negative, Neutral, or Cannot Say. This

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

column serves as the target variable for supervised model training. The dataset is rich in real-world textual variations, including abbreviations, mentions and special characters, which require careful preprocessing. It is structured to support both exploratory data analysis and feature extraction workflows, making it ideal for testing both existing machine learning models and the proposed BRC driven by SBERT embeddings.

Fig. 3 bar chart illustrates the distribution of sentiment classes, with three categories labeled 0, 1, and 2. Class 2 has the highest count, exceeding 3500, indicating a significant prevalence. Class 1 follows with a count around 2000, while Class 0 has the lowest count, just above 0. This suggests an imbalanced dataset with a dominant presence of Class 2.

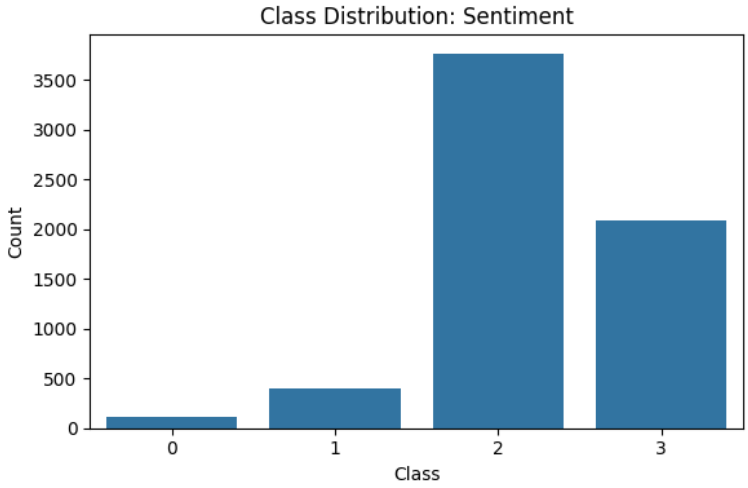
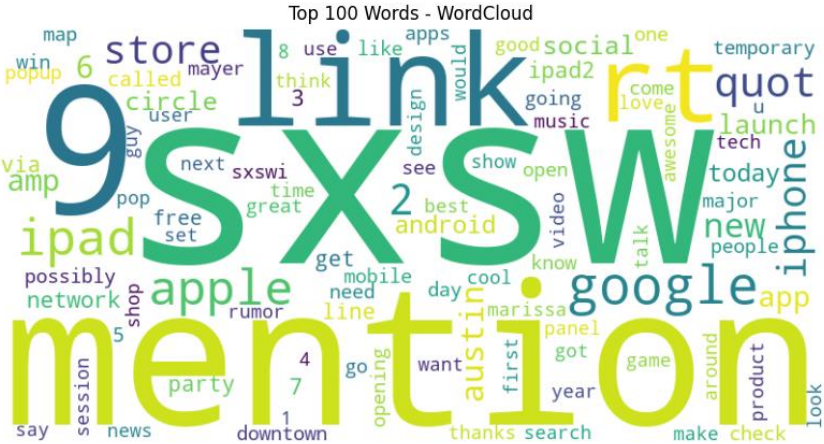


Fig. 3: Class distribution for sentiment attribute.

Fig. 4 showcases word cloud highlights the most frequently occurring words, with size and prominence reflecting their frequency. Words like "sxsw," "link," "mention," "google," "ipad," and "apple" stand out, indicating key topics or entities. The use of various colors helps distinguish the words, with "sxsw" and "link" appearing most dominant. Fig. 5 represents bar chart ranks the top 20 most frequent words, with "sxsw" leading at over 6000 counts, followed by "mention" and "link" with counts around 5000 and 4000, respectively. Other words like "rt," "google," "ipad," and "apple" also show notable frequencies, decreasing gradually, reflecting their relative importance in the dataset. Fig. 6 illustrates histogram, overlaid with a curve, shows the frequency of document lengths in words. The distribution peaks at around 15-20 words with a frequency of over 1200, indicating the most common document length. The curve suggests a roughly normal distribution, with frequencies dropping off on either side, especially beyond 25 words.



International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Fig. 4: Top 100 Words – WordCloud

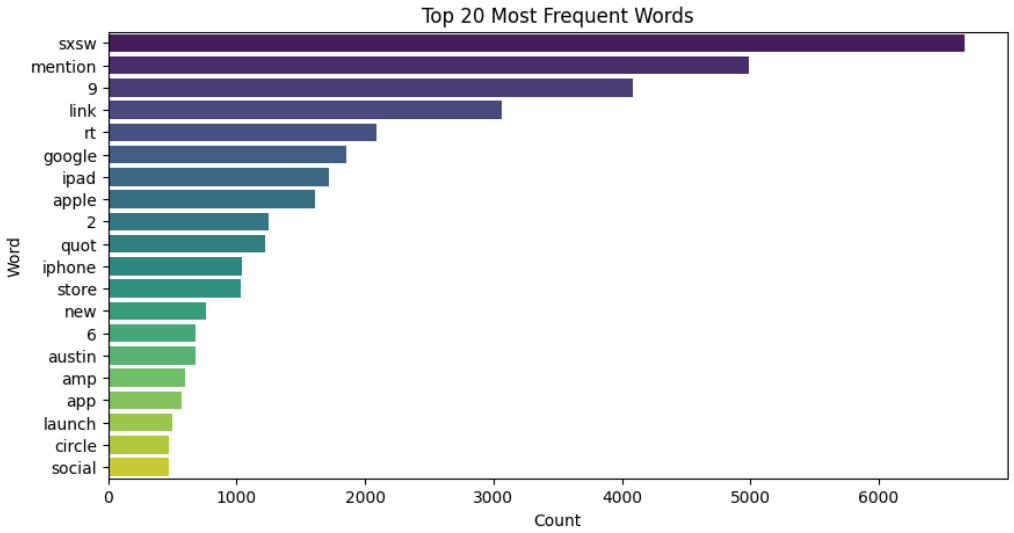


Fig. 5: Top 20 most frequent words.

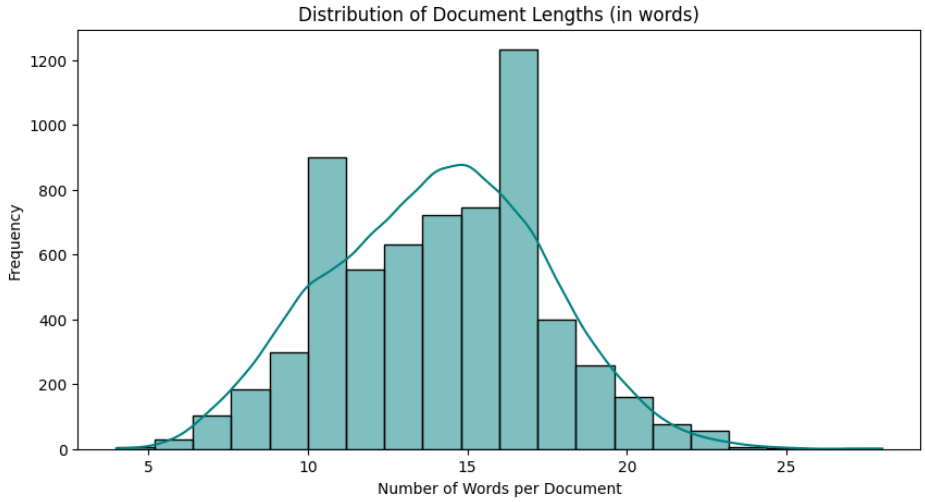


Fig. 6: Distribution of document lengths (in words).

Fig. 7 displays the frequency of different parts of speech tags. The tag "NN" (noun) dominates with a frequency above 4000, followed by "CD" (cardinal number) and "VB" (verb) with frequencies around 2000 and 1000, respectively. Other tags have significantly lower frequencies, showing a skewed distribution favoring nouns. Fig. 8 displays the bar chart lists the top 20 most frequent bigrams, with "rt mention" leading at over 2000 counts, followed by "9 rt" and "sxsw link" with counts around 1500 and 1200, respectively. Bigrams like "ipad 2," "apple store," and "mention mention" also appear frequently, indicating common word pairs in the dataset. Fig. 9 represents dual-plot chart tracks training and validation accuracy (left) and loss (right) over 50 epochs. Accuracy rises sharply to around 0.95 and stabilizes, while loss drops from 0.8 to below 0.2, with both training and validation curves converging, indicating a well-fitted model with high performance.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

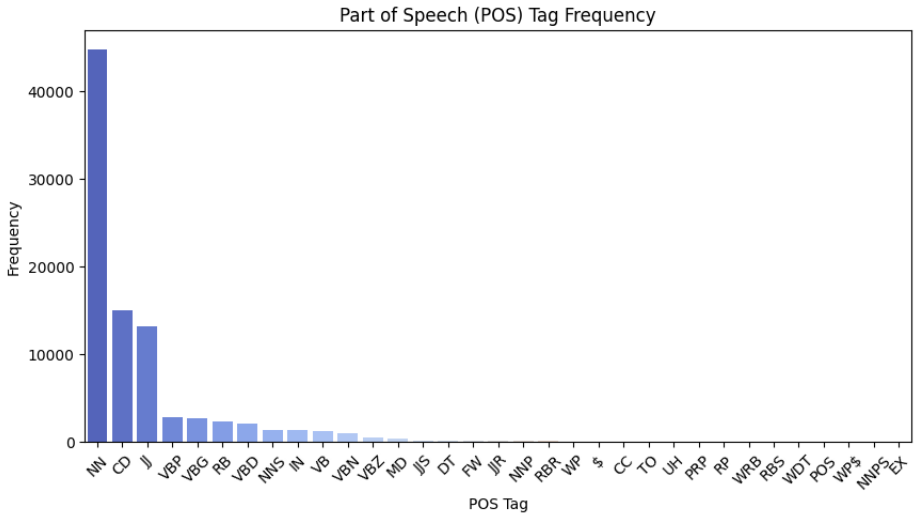


Fig. 7: Part of Speech (POS) tag frequency

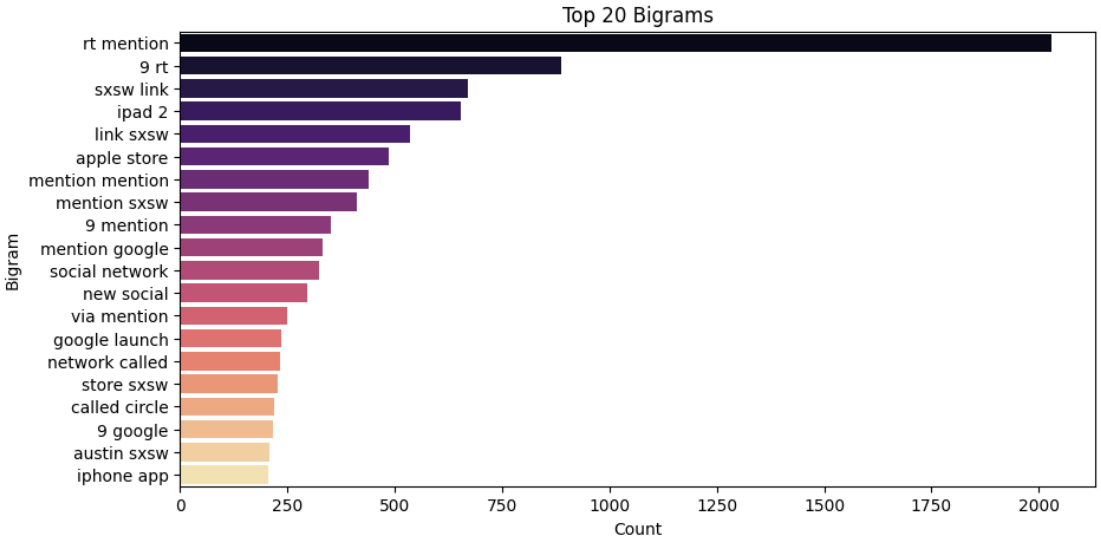


Fig. 8: Top 20 bigrams.

Fig. 10 (left) compares confusion matrix true and predicted sentiment classes. The "Cannot Say" class has 751 correct predictions, "Negative" has 751, and "Neutral" has 740, with "Positive" showing 739 correct predictions. Off-diagonal values (e.g., 14 misclassifications from Neutral to Positive) are minimal, suggesting good classification performance. In Fig. 10 (right), ROC curve plot shows true positive rates against false positive rates for each class. All classes ("Cannot Say," "Negative," "Positive," "Neutral") achieve an AUC of 1.0, indicating perfect classification, while the random guess line (AUC = 0.99) serves as a baseline.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

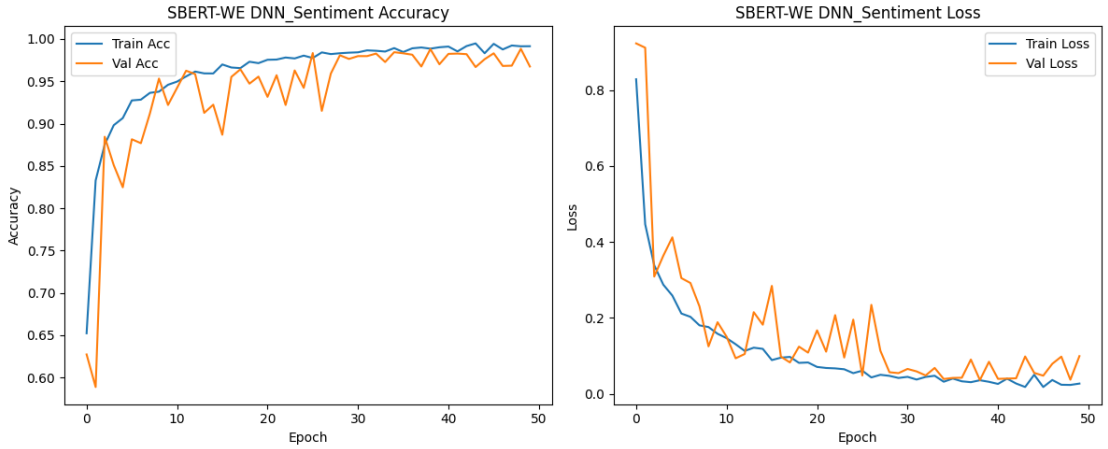


Fig. 9: SBERT-WE DNN sentiment accuracy and loss.

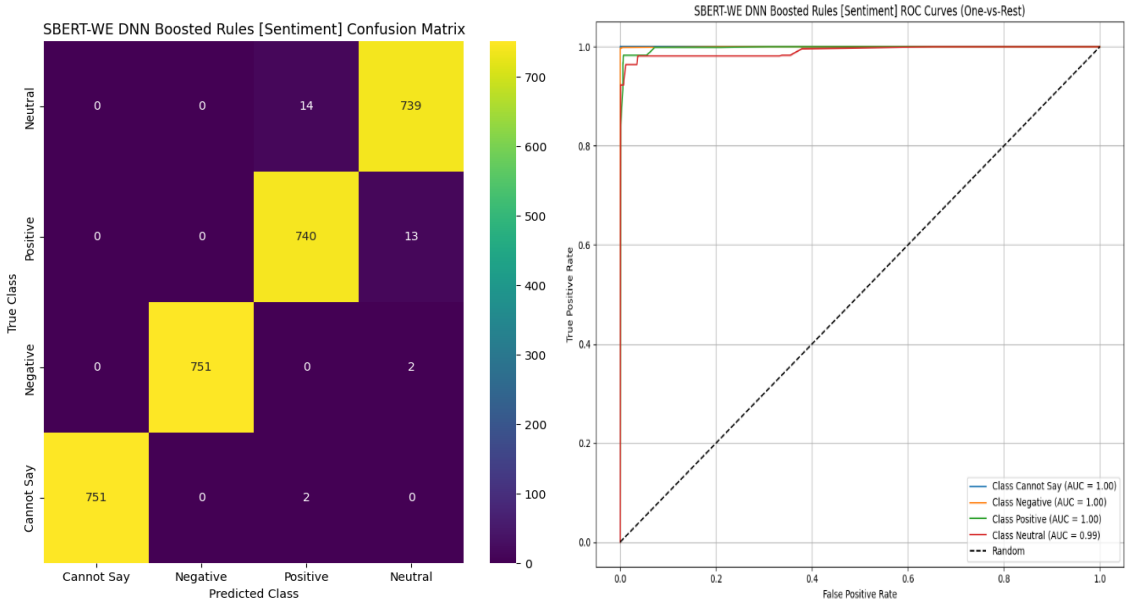
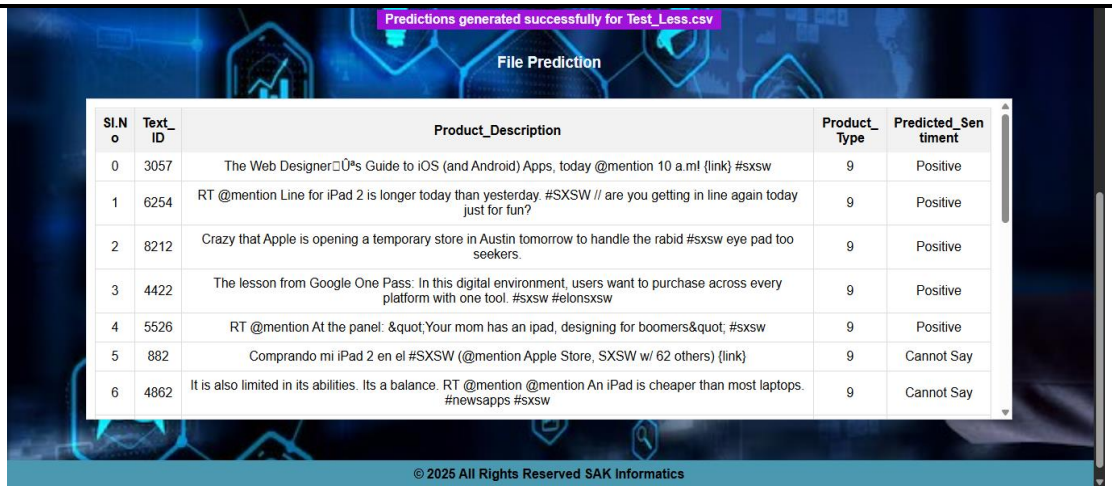


Fig. 10: SBERT-WE DNN boosted rules (Sentiment) confusion matrix (left). ROC curves (One-vs-Rest) (right).

Table 1: Model performance comparison.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SBERT-WE RFC [Sentiment]	92.065	92.055	92.065	92.055
SBERT-WE LGBM [Sentiment]	93.260	93.248	93.260	93.252
SBERT-WE XGB [Sentiment]	93.526	93.481	93.526	93.482
SBERT-WE DNN boosted rules [Sentiment]	98.971	98.974	98.971	98.972

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



Sl.No	Text_ID	Product_Description	Product_Type	Predicted_Sentiment
0	3057	The Web Designer□Ûªs Guide to iOS (and Android) Apps, today @mention 10 a.m! {link} #sxsw	9	Positive
1	6254	RT @mention Line for iPad 2 is longer today than yesterday. #SXSW // are you getting in line again today just for fun?	9	Positive
2	8212	Crazy that Apple is opening a temporary store in Austin tomorrow to handle the rabid #sxsw eye pad too seekers.	9	Positive
3	4422	The lesson from Google One Pass: In this digital environment, users want to purchase across every platform with one tool. #sxsw #elonsxsw	9	Positive
4	5526	RT @mention At the panel: "Your mom has an ipad, designing for boomers" #sxsw	9	Positive
5	882	Comprando mi iPad 2 en el #SXSW (@mention Apple Store, SXSW w/ 62 others) {link}	9	Cannot Say
6	4862	It is also limited in its abilities. Its a balance. RT @mention @mention An iPad is cheaper than most laptops. #newsapps #sxsw	9	Cannot Say

Fig. 11: Real time predictions of Product sentiment analysis.

Fig. 11 illustrates the real-time sentiment prediction interface where users upload or input product-related data to obtain instant sentiment results. The figure presents how the system processes text through the model and displays outputs such as positive, negative, or neutral sentiment. It represents the core analytical feature of the system, showcasing immediate and dynamic prediction capability.

5. CONCLUSION

The proposed approach presents an effective and scalable framework for sentiment classification by combining advanced NLP techniques with ML strategies. The methodology incorporates structured preprocessing, context-aware embedding generation using SBERT, feature optimization through DNN, and final prediction using ensemble methods along with BRC. This integrated pipeline significantly enhances model performance compared to conventional techniques. Experimental comparisons show that the SBERT-WE DNN BRC model achieves superior results over baseline models such as RFC, LGBM, and XGB. The proposed model attains an accuracy, precision, recall, and F1-score of 98.971, notably exceeding the 92–93 performance range observed in standard ensemble approaches. These results highlight the strength of combining semantically enriched embeddings with deep feature learning and rule-based boosting to effectively capture both contextual meaning and classification logic. Furthermore, the integration of multiple modeling techniques not only improves predictive performance but also preserves interpretability, which is essential for practical deployment. The system efficiently handles multi-class sentiment classification and maintains consistent performance across evaluation metrics, demonstrating its reliability, scalability, and suitability for real-world applications.

REFERENCES

[1]. Y. Jiang, A. Huang, S. Gao, and S. Yu, “Relationship between the terminal-built environment and airport retail revenue,” J. Air Transp. Manage., vol. 116, Apr. 2024, Art. no. 102568.

A. Alabaidi, “The impact of work life balance on employee attitudes and behaviour in health care sector,” Tech. Rep., 2024.

[2]. X. Zhao and Y. Sun, “Amazon fine food reviews with BERT model,” Proc. Comput. Sci., vol. 208, pp. 401–406, Jan. 2022.

[3]. J. Ballerini, A. Ključnikov, D. Juárez-Varón, and S. Bresciani, “The e-commerce platform conundrum: How manufacturers’ leanings affect their internationalization,” Technol. Forecasting Social Change, vol. 202, May 2024, Art. no. 123199.

[4]. M. S. Akin, “Enhancing e-commerce competitiveness: A comprehensive analysis of customer experiences and strategies in the Turkish market,” J. Open Innov., Technol., Market, Complex., vol. 10, no. 1, Mar. 2024, Art. no. 100222.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[5]. Khan, L.; Amjad, A.; Afaq, K.M.; Chang, H.-T. Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media. Appl. Sci. 2022, 12, 2694. https://doi.org/10.3390/app12052694

[6]. Serna, A.; Soroa, A.; Agerri, R. Applying Deep Learning Techniques for Sentiment Analysis to Assess Sustainable Transport. Sustainability 2021, 13, 2397. https://doi.org/10.3390/su13042397

[7]. Ghatora, P.S.; Hosseini, S.E.; Pervez, S.; Iqbal, M.J.; Shaukat, N. Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM. Big Data Cogn. Comput. 2024, 8, 199. https://doi.org/10.3390/bdcc8120199

[8]. Ahanin, Z.; Ismail, M.A.; Singh, N.S.S.; AL-Ashmori, A. Hybrid Feature Extraction for Multi-Label Emotion Classification in English Text Messages. Sustainability 2023, 15, 12539. https://doi.org/10.3390/su151612539

[9]. Chu, Z.; Wang, X.; Jin, M.; Zhang, N.; Gao, Q.; Shao, L. An Effective Strategy for Sentiment Analysis Based on Complex-Valued Embedding and Quantum Long Short-Term Memory Neural Network. Axioms 2024, 13, 207. https://doi.org/10.3390/axioms13030207

[10]. Tan, K.L.; Lee, C.P.; Lim, K.M. RoBERTa-GRU: A Hybrid Deep Learning Model for Enhanced Sentiment Analysis. Appl. Sci. 2023, 13, 3915. https://doi.org/10.3390/app13063915

[11]. Mao, X.; Chang, S.; Shi, J.; Li, F.; Shi, R. Sentiment-Aware Word Embedding for Emotion Classification. Appl. Sci. 2019, 9, 1334. https://doi.org/10.3390/app9071334

[12]. Trisna, K.W.; Huang, J.; Liang, H.; Dharma, E.M. Fusion Text Representations to Enhance Contextual Meaning in Sentiment Classification. Appl. Sci. 2024, 14, 10420. https://doi.org/10.3390/app142210420

[13]. Oprea, S.-V.; Bâra, A. Extracting Emotions from Customer Reviews Using Text Mining, Large Language Models and Fine-Tuning Strategies. J. Theor. Appl. Electron. Commer. Res. 2025, 20, 221. https://doi.org/10.3390/jtaer20030221