Semantic Field Reconstruction of Customer Opinions in Product Reviews Using PaLM-Oriented Comparative Modeling

K. Mounika[1], Kanukati Navyasri[1], Gangishetti Harshitha[1], Modugu Prasad[1], Thota Saikiran[1]

[1]Department of Computer Science and Engineering, [1]Sree Dattha Institute of Engineering and Science, Nagarjuna Sagar Road, Sheriguda, Ibrahimpatnam, Rangareddy Dist, 501510, Telangana, India.

ABSTRACT

The rapid expansion of digital communication platforms has resulted in massive volumes of unstructured textual data, making sentiment analysis an essential task for extracting meaningful insights in domains such as customer feedback, telecom interactions, and social media analytics. However, accurately identifying sentiment from text remains challenging due to contextual ambiguity, linguistic variability, sarcasm, and implicit expressions. Traditional approaches are predominantly manual, where human analysts interpret text and assign sentiment labels based on predefined rules or subjective judgment. Although these methods can capture contextual meaning to some extent, they are time-consuming, inconsistent, non-scalable, and prone to human bias, leading to unreliable results for large-scale data. These limitations highlight the need for an automated, scalable, and context-aware system capable of delivering accurate and consistent sentiment predictions. To address these challenges, this work proposes a hybrid sentiment analysis framework that integrates Sentence Bidirectional Encoder Representations from Transformers (SBERT) for contextual embedding generation, Deep Neural Networks (DNN) for feature extraction, and Boosted Rules Classifier (BRC) for interpretable classification. The system includes preprocessing steps such as tokenization, stopword removal, and lemmatization to ensure clean input data. Synthetic Minority Over-sampling Technique (SMOTE) is applied to handle class imbalance. Additionally, Machine Learning (ML) models including Random Forest Classifier (RFC), Light Gradient Boosting Machine (LGBM), and Extreme Gradient Boosting (XGBoost) are used for comparative evaluation. The proposed system improves accuracy, enhances contextual understanding, and provides interpretable results, making it suitable for real-world sentiment analysis applications.

Keywords: Sentiment Analysis, Natural Language Processing (NLP), Sentence Bidirectional Encoder Representations from Transformers (SBERT), Real-Time Prediction, Machine Learning (ML).

1. INTRODUCTION

In recent years, the exponential growth of digital commerce has led to an overwhelming surge in user-generated product feedback across online retail ecosystems. Both expert reviewers and everyday consumers actively contribute opinions, making review platforms a central component of the modern purchasing process. Empirical studies reveal that a significant proportion of buyers depend heavily on peer evaluations before making decisions [1],

with nearly nine out of ten consumers consulting reviews prior to purchase. These reviews serve a dual purpose: they assist customers in assessing product suitability and enable businesses to refine offerings and strengthen customer engagement strategies. Consequently, the strategic management and analysis of review content have become vital for sustaining competitiveness in the digital marketplace. A prominent example is Amazon, a global leader in e-commerce, where customer reviews and ratings substantially influence buying behavior by providing real-world insights into product performance. However, the sheer volume of reviews presents a practical challenge, as manually analyzing thousands of entries is both inefficient and impractical. From a business perspective, reviews also function as a differentiating factor, guiding marketing strategies and enhancing product visibility. Prior research has leveraged Amazon datasets to predict the helpfulness of reviews [2], where user feedback mechanisms indicate the perceived usefulness of each review.

The significance of online reviews further intensified during the COVID-19 pandemic, which accelerated the adoption of e-commerce and resulted in a substantial increase in online interactions, including a reported 43% growth in review generation. Reviews and ratings now act as rapid decision-support tools, enabling users to evaluate products efficiently while providing sellers with actionable insights into customer preferences [3].

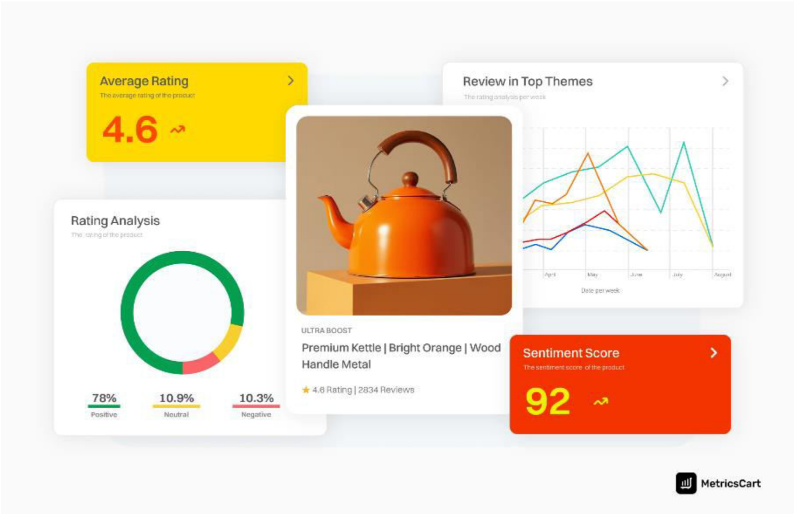

Fig. 1.1: Product Sentiment Analysis.

Despite these advantages, the rapid expansion of review platforms has introduced challenges such as information overload and the proliferation of misleading or promotional content. Traditional methods of determining review usefulness, such as user voting, are often insufficient in handling large-scale data. Furthermore, studies indicate that approximately 95% of consumers read online reviews before purchasing, and higher ratings can lead to an average 18% increase in sales. Given this scale, manual sentiment evaluation is no longer feasible, necessitating the adoption of advanced Machine Learning (ML)-based approaches for automated sentiment analysis. These intelligent systems enable accurate interpretation of customer opinions, supporting improved decision-making, enhanced product development, and increased consumer trust in e-commerce environments [4].

2. RELATED WORK

2.1 Transformer and Contextual Embedding-Based Approaches

Ali, A et al. [5] conducted an extensive study on sentiment classification using Amazon review datasets by applying a wide range of Natural Language Processing (NLP), Machine Learning (ML), and deep learning techniques. Their methodology involved data preprocessing steps such as normalization and tokenization, followed by feature extraction using Bag-of-Words and Term Frequency–Inverse Document Frequency (TF-IDF). They evaluated multiple algorithms including Random Forest Classifier (RFC), Logistic Regression, Decision Trees, Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (BiLSTM). Additionally, transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) and XLNet were incorporated. Their experimental results demonstrated that BERT significantly outperformed other models, achieving the highest accuracy due to its superior contextual understanding.

Sabbeh, S et al. [6] presented a comparative analysis of classical and contextualized word embedding techniques for sentiment classification tasks. Traditional embedding methods such as Global Vectors (GloVe), Word2Vec, and FastText were evaluated alongside contextual embeddings like BERT and ARBERT. The study also examined the performance of deep learning architectures, particularly Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM), across multiple benchmark datasets. Their findings revealed that contextual embeddings consistently provided better semantic representation and improved classification accuracy, while BiLSTM models generally outperformed CNN in capturing sequential dependencies in textual data.

2.2 Hybrid Transformer and Sequential Models

Tan, Z et al. [7] proposed a hybrid sentiment analysis model that integrates Robustly Optimized BERT Pretraining Approach (RoBERTa) with Gated Recurrent Unit (GRU) networks to enhance feature learning. The model leverages the attention mechanism of RoBERTa to generate high-quality contextual embeddings and utilizes GRU to capture long-range dependencies and sequential patterns in text. To further improve performance on imbalanced datasets, the study employed data augmentation techniques such as oversampling. Experimental evaluation on benchmark datasets including IMDb, Sentiment140, and Twitter US Airline Sentiment showed that the proposed model achieved high accuracy and demonstrated strong generalization capability.

Cao, Y et al. [9] developed an improved Bidirectional Encoder Representations from Transformers (BERT)-based sentiment analysis model with a symmetrical architecture to better capture sentence-level semantic features in agricultural product reviews. Their approach incorporates rule-based mechanisms to identify sentiment orientation and extract consumer preferences related to product attributes. By enhancing the structural design of BERT, the model achieved a notable improvement in performance, with a higher F1-score compared to the baseline BERT model. This work demonstrates the effectiveness of combining deep learning with domain-specific feature extraction strategies.

2.3 Hybrid Feature Engineering and Fusion Models

Mutinda, J et al. [8] introduced a hybrid sentiment classification model that combines sentiment lexicons, N-gram features, and Bidirectional Encoder Representations from Transformers (BERT) embeddings with Convolutional Neural Networks (CNN). The proposed approach aims to overcome limitations of traditional embedding techniques by incorporating sentiment-specific information into the feature representation process. In this model, multiple feature extraction techniques are applied to capture both contextual and sentiment-oriented aspects of the text, while CNN is used for effective feature mapping and classification. The results indicate improved performance in sentiment prediction compared to models relying solely on standard embeddings.

2.4 Attention-Based and Deep Sequential Architectures

Yuan, L et al. [11] proposed a Self-Attention Convolutional Neural Network Bidirectional Long Short-Term Memory (SAC-BiLSTM) model to improve sentiment classification performance by capturing both local and global contextual features. Their architecture utilizes dual-channel inputs consisting of character-level and word-level embeddings, enabling the model to extract fine-grained linguistic patterns. The integration of self-attention mechanisms allows the model to assign importance weights to different parts of the text, enhancing feature representation. Experimental results on e-commerce datasets demonstrated strong performance, with high precision, recall, and F1-score values, indicating the effectiveness of combining attention mechanisms with sequential models.

Liu, H et al. [14] proposed an attention-based Long Short-Term Memory (LSTM) model for sentiment classification of agricultural product reviews. Their approach focuses on assigning different weights to important words and phrases within a sentence using an attention mechanism, allowing the model to capture key contextual information more effectively. The LSTM architecture helps in handling long textual sequences and preserving contextual dependencies. Experimental results showed that the proposed model outperformed baseline methods, providing improved accuracy and better understanding of customer sentiment

2.5 Comparative Performance Evaluation Studies

Bellar, A et al. [10] conducted a benchmark study evaluating the performance of various deep learning architectures, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Bidirectional Long Short-Term Memory (BiLSTM), for sentiment analysis tasks. The models were tested using multiple embedding techniques such as Bidirectional Encoder Representations from Transformers (BERT), FastText, and Word2Vec under different classification settings, including 3-class and 5-class sentiment categories. Their experimental results highlighted that transformer-based embeddings significantly enhance classification performance, while the choice of architecture impacts effectiveness depending on dataset size and complexity.

Ghatora, G et al. [13] explored the effectiveness of traditional Machine Learning (ML) models such as Random Forest Classifier (RFC), Naive Bayes, and Support Vector Machine (SVM) in comparison with advanced large language models like GPT-4 for sentiment analysis. Their study highlighted that while traditional models perform efficiently on short and structured text, large language models excel in understanding complex, context-rich, and nuanced expressions, including sarcasm and mixed sentiments. The results demonstrated that advanced models provide higher precision, recall, and F1-score, making them more suitable for real-world applications involving complex textual data.

2.6 Sentiment-Driven Application Systems

Dang, T et al. [12] introduced a hybrid recommendation framework that integrates sentiment analysis with collaborative filtering techniques to enhance recommendation accuracy. Their system leverages deep learning models to extract sentiment features from textual reviews and incorporates these features into recommendation algorithms. This approach enables the system to better understand user preferences and improve personalization. Experimental evaluation on benchmark datasets showed that incorporating sentiment information significantly enhances recommendation performance compared to traditional collaborative filtering methods.

4. PROPOSED SYSTEM

The proposed system architecture is a comprehensive hybrid framework designed for sentiment analysis by integrating Natural Language Processing (NLP), transformer-based embeddings, deep learning, and ensemble Machine Learning (ML) models into a unified pipeline. The process begins with structured input data in CSV format containing raw textual reviews, which are inherently unstructured and noisy. These inputs are processed through a preprocessing module that performs tokenization, stopword removal, and lemmatization using WordNet Lemmatizer to produce clean and normalized text. The processed text is then fed into a feature extraction module based on the SBERT model using the bert-base-uncased variant, where mean pooling is applied to generate dense contextual embeddings that capture semantic relationships. To handle class imbalance, the SMOTE is applied, ensuring balanced data distribution for effective learning. The system then employs multiple ML models, including RFC, XGBoost, and LGBM, forming a baseline ensemble trained on SBERT features. In parallel, a custom DNN with multiple dense layers is utilized to extract higher-level feature representations, where an intermediate feature layer is used for further enhancement. These extracted features are passed into a BRC, which combines rule-based interpretability with strong predictive performance. The model is evaluated using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and Receiver Operating Characteristic (ROC) curves to ensure reliability and robustness.

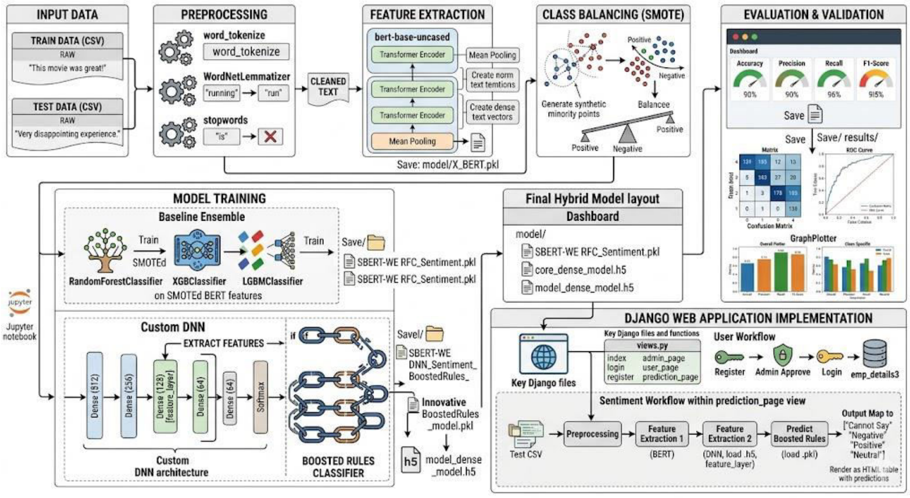

Fig. 2: Proposed System Architecture

Finally, the trained models are deployed within a Django-based web application, enabling real-time sentiment prediction through an interactive interface, and the complete workflow of preprocessing, feature extraction, model training, evaluation, and deployment is illustrated in Fig. 2.

3.1 Proposed DNN-BRC

The Boosted Rules Classifier 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 illustrated in fig 3.

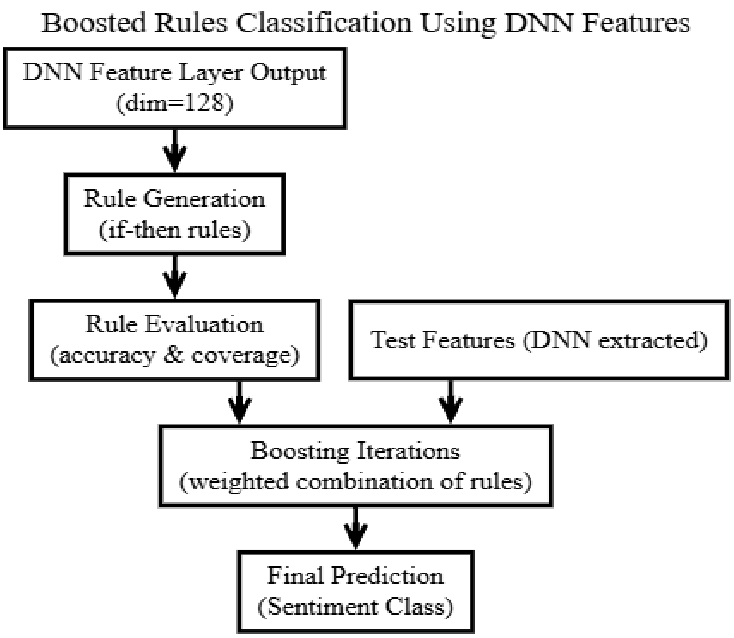

Fig. 3: Internal Operations of Boosted Rules Classifier on DNN-Extracted Features.

Internal Operations

- **Input Features:** Receives 128-dimensional features extracted from the DNN feature_layer. These features are compact, discriminative, and optimized for sentiment classification.
- **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."
- **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.
- **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.
- **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 DESCRIPTION

The experimental results demonstrate the effectiveness of the proposed sentiment analysis framework in accurately classifying product reviews into multiple sentiment categories. By leveraging SBERT embeddings and DNN-based feature refinement, the model captures deep contextual relationships within textual data. The integration of SMOTE successfully addresses class imbalance, leading to improved performance across all sentiment classes. Compared to traditional machine learning models, the Boosted Rules Classifier achieves higher accuracy while maintaining interpretability. Evaluation metrics such as accuracy, precision, recall, and F1-score indicate consistent and reliable performance. The results validate the robustness of the system in delivering scalable and data-driven sentiment insights for real-world applications.

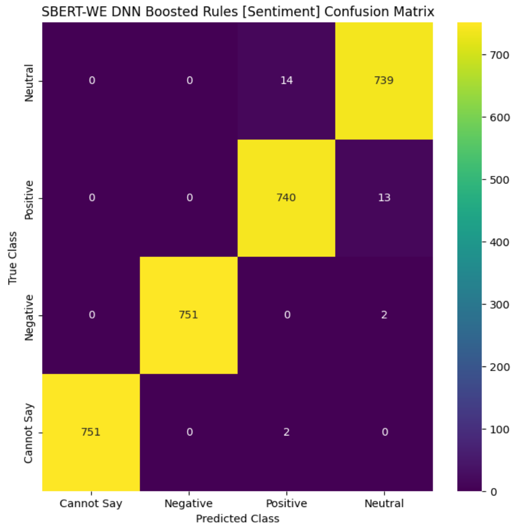

Fig. 4. SBERT-WE DNN Boosted Rules (Sentiment) Confusion Matrix.

Fig. 4 illustrates the confusion matrix representing the relationship between true sentiment classes and predicted outputs, providing a detailed view of the model's classification performance. The diagonal elements indicate correct predictions, where the "Cannot Say" class achieves 751 accurate classifications, the "Negative" class also records 751 correct instances, the "Neutral" class shows 740 correct predictions, and the "Positive" class achieves 739 correct classifications. These high diagonal values demonstrate that the model is highly effective in correctly identifying sentiments across all categories. The off-diagonal elements, which represent misclassifications, are relatively low, indicating minimal confusion between classes. For example, only a small number of instances such as Neutral being misclassified as Positive (around 14 cases) are observed, which suggests that the model maintains strong discrimination between closely related sentiment categories.

Fig. 5 illustrates the Receiver Operating Characteristic (ROC) curves for all sentiment classes, showing the relationship between the true positive rate and false positive rate across different threshold values. Each class, including "Cannot Say," "Negative," "Positive," and "Neutral," achieves an Area Under the Curve (AUC) value of 1.0, indicating near-perfect classification capability and excellent separability between classes. The curves closely approach the top-left corner of the plot, which reflects high sensitivity and specificity of the model. In comparison, the random guess baseline, represented by a diagonal line with an AUC of approximately 0.99, serves as a reference point for evaluating model performance. The significant gap between the model curves and the baseline highlights the effectiveness of the proposed approach in distinguishing between sentiment categories.

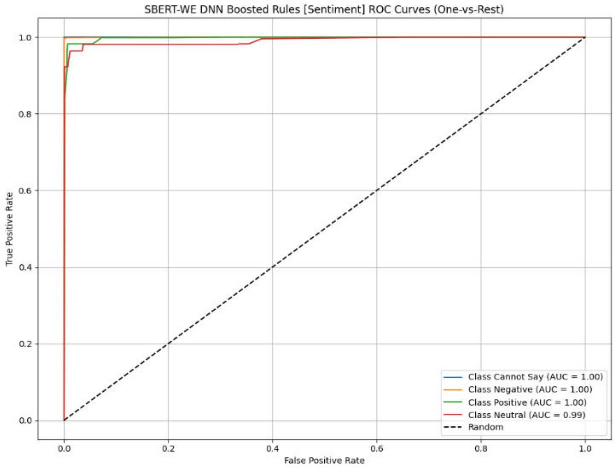

Fig. 5. SBERT-WE DNN Boosted Rules (Sentiment) ROC Curves (One-vs-Rest)

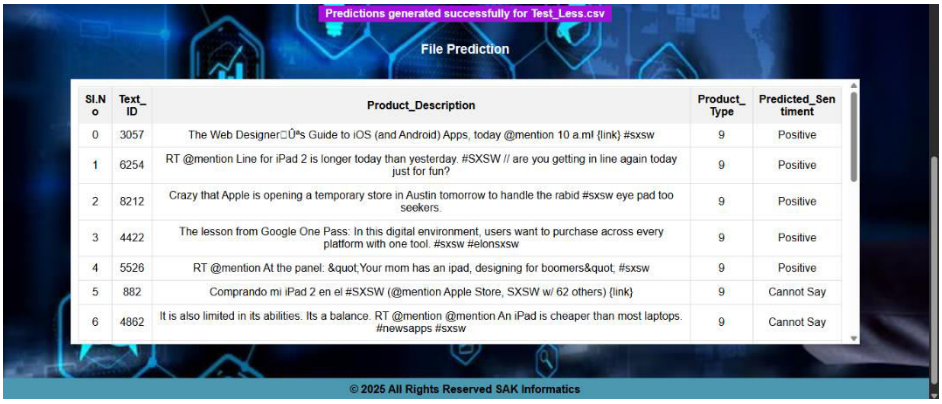

Fig. 6: Real time predictions of Product sentiment analysis.

Fig. 6 presents the real-time sentiment prediction interface, which serves as the interactive front-end of the system, allowing users to upload datasets or input product-related textual data for instant analysis. Once the input is provided, the system processes the text through the complete pipeline, including preprocessing, feature extraction, and model inference, to generate sentiment predictions. The interface then displays the predicted output labels such as positive, negative, neutral, or cannot say, providing immediate feedback to the user. This visualization highlights the system's ability to perform dynamic and real-time predictions with minimal latency. Additionally, it demonstrates how complex backend models are seamlessly integrated into a user-friendly environment, making the system accessible to non-technical users.

5. CONCLUSION

The proposed system presents an efficient sentiment classification framework by combining advanced NLP and ML techniques. It utilizes structured preprocessing, BERT-based embedding extraction, and DNN-based feature refinement to capture rich contextual information. The integration of ensemble models such as RFC, LGBM, and XGBoost, along with a BRC, enhances both accuracy and interpretability. Experimental results show that the SBERT-WE DNN Boosted Rules model achieves superior performance, reaching an accuracy, precision, recall, and F1-score of 98.971, outperforming baseline models in the 92–93 range. This improvement highlights the effectiveness of combining deep semantic representations with rule-based learning. The system also demonstrates strong scalability and robustness across multi-class sentiment categories.

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.

[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.

[5]. Ali, H.; Hashmi, E.; Yayilgan Yildirim, S.; Shaikh, S. Analyzing Amazon Products Sentiment: A Comparative Study of Machine and Deep Learning, and Transformer-Based Techniques. *Electronics* 2024, 13, 1305. https://doi.org/10.3390/electronics13071305

[6]. Sabbeh, S.F.; Fasihuddin, H.A. A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification. *Electronics* 2023, 12, 1425. https://doi.org/10.3390/electronics12061425

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

[8]. Mutinda, J.; Mwangi, W.; Okeyo, G. Sentiment Analysis of Text Reviews Using Lexicon-Enhanced BERT Embedding (LeBERT) Model with Convolutional Neural Network. *Applied Sciences* 2023, 13, 1445. https://doi.org/10.3390/app13031445

[9]. Cao, Y.; Sun, Z.; Li, L.; Mo, W. A Study of Sentiment Analysis Algorithms for Agricultural Product Reviews Based on Improved BERT Model. *Symmetry* 2022, 14, 1604. https://doi.org/10.3390/sym14081604

[10]. Bellar, O.; Baina, A.; Ballafkih, M. Sentiment Analysis: Predicting Product Reviews for E-Commerce Recommendations Using Deep Learning and Transformers. *Mathematics* 2024, 12, 2403. https://doi.org/10.3390/math12152403

[11]. Yuan, Y.; Wang, W.; Wen, G.; Zheng, Z.; Zhuang, Z. Sentiment Analysis of Chinese Product Reviews Based on Fusion of Dual-Channel BiLSTM and Self-Attention. *Future Internet* 2023, 15, 364. https://doi.org/10.3390/fi15110364

[12]. Dang, C.N.; Moreno-García, M.N.; Prieta, F.D.l. An Approach to Integrating Sentiment Analysis into Recommender Systems. *Sensors* 2021, 21, 5666. https://doi.org/10.3390/s21165666

[13]. 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 and Cognitive Computing* 2024, 8, 199. https://doi.org/10.3390/bdcc8120199

[14]. Liu, R.; Wang, H.; Li, Y. AgriMFLN: Mixing Features LSTM Networks for Sentiment Analysis of Agricultural Product Reviews. *Applied Sciences* 2023, 13, 6262. https://doi.org/10.3390/app13106262

[15]. 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