Submitted By:
1. Md. Zubair Rahman - 0242220005101325
2. Mahadi Hasan Rakib– 0242220005101321
3. Ashraful Alam Chowdhury – 0242220005101317

Submitted To:
Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Date: 24-12-2025

A Sentiment Analysis Framework for Customer Product Reviews in the Bengali Language

Md. Zubair Rahman [1], Mahadi Hasan Rakib[2], Ashraful Alam Chowdhury [3]

Citation: Md. Zubair Rahman, Mahadi Hasan Rakib, Ashraful Alam Chowdhury
Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh;
rahman22205101325@diu.edu.bd
rakib22205101321@diu.edu.bd

Abstract

Sentiment analysis plays a critical role in understanding customer opinions, yet research on Bengali language remains limited due to scarce annotated datasets, preprocessing challenges, and the frequent mixing of Bengali and English text. These limitations hinder the ability of businesses and researchers in Bangladesh to extract actionable insights from large volumes of user generated content. To address this gap, this study presents an integrated sentiment analysis framework specifically designed for Bengali product reviews collected from the Daraz e-commerce platform. A dataset of 3,072 raw reviews was constructed, cleaned, and manually annotated into three sentiment categories: positive, negative, and neutral. The preprocessing pipeline included numeral standardization, removal of non-Bengali text, normalization of review structure, and categorical and encoded labeling. Two modeling strategies were investigated: traditional machine learning using TF-IDF features (K-Nearest Neighbors, Logistic Regression, Random Forest, and Naïve Bayes) and a deep learning approach based on tokenized sequences with a Convolutional Neural Network (CNN). Experimental evaluation revealed that the CNN achieved the highest accuracy (79%), outperforming classical models while maintaining balanced classification across minority classes. In addition, a web-based system was developed that enables users to upload datasets, automatically process reviews, and visualize results through performance metrics and feature-importance analyses. Overall, the proposed framework demonstrates the feasibility and practical value of automated Bengali sentiment analysis, offering a scalable solution for e-commerce analytics and future NLP research.

Keywords: Bengali Sentiment Analysis, E-commerce Reviews, Daraz, Natural Language Processing, Text Preprocessing, TF-IDF, Convolutional Neural Network (CNN), Machine Learning, Deep Learning, Bangla NLP, Dataset Annotation, Web-based Sentiment System.

1. Introduction

Customer reviews play a critical role in shaping purchasing decisions and business strategies in modern e-commerce ecosystems. With the rapid expansion of online marketplaces in Bangladesh, platforms such as Daraz now generate thousands of user written reviews daily. However, extracting meaningful insights from this growing volume of text remains challenging, particularly when reviews are predominantly written in Bengali. While sentiment analysis has been widely studied for high-resource languages such as English, Bengali remains comparatively underexplored, primarily due to the lack of standardized datasets, preprocessing tools, and language specific models [1], [4]. As a result, organizations frequently rely on manual review inspection, which is time consuming, subjective, and impractical at scale. These limitations restrict data driven decision making and reduce the ability of businesses to understand customer satisfaction trends and product performance.

Unlike English, Bengali exhibits rich morphology, free-word order, and frequent incorporation of transliterated Banglish text, making direct application of English-trained models ineffective [2], [5]. Previous studies have explored Bengali sentiment analysis using machine learning, lexicon-based methods, and transfer learning approaches such as multilingual BERT and Bangla-BERT, showing promising results but also highlighting persistent challenges including domain dependence, lack of balanced datasets, and difficulty handling neutral sentiment [3], [5], [6], [9]. Other works have focused on specific domains such as political discourse or social media comments [4], [7], yet these datasets are often noisy and not directly applicable to structured e-commerce environments. Consequently, there remains a significant gap in scalable, accurate, and interpretable sentiment analysis solutions specifically tailored to Bengali product reviews.

This research aims to address these challenges by developing a comprehensive sentiment analysis framework designed for Bengali e-commerce data. Building on curated datasets and modern natural language processing techniques, the proposed system integrates preprocessing, feature engineering, and both traditional and deep learning models to achieve reliable sentiment classification. Reviews collected from Daraz were cleaned, standardized, and manually annotated into positive, negative, and neutral categories, ensuring linguistic consistency and high-quality supervision. The framework further incorporates a web-based interface that enables users to upload datasets, automatically process reviews, visualize outputs, and download labeled results — making the system practical for real-world deployment.

The major contributions of this study are summarized as follows:

• Develops a curated, manually annotated Bengali sentiment dataset derived from real e-commerce product reviews, addressing the scarcity of domain-specific labeled resources for Bengali NLP [1], [4].

• Proposes a unified sentiment analysis pipeline combining traditional TF-IDF–based machine learning models with a deep learning Convolutional Neural Network (CNN), enabling comparison between approaches and improving robustness across sentiment classes [2], [3], [6].

• Introduces preprocessing strategies tailored to Bengali text, including numeral normalization, removal of Banglish content, and structured encoding of labels, enhancing model generalization and consistency [4], [5].

• Incorporates performance visualization and feature-importance analysis to improve interpretability and decision support an aspect often overlooked in prior Bengali sentiment research [3], [7].

• Implements a scalable, web-based platform that allows automated dataset processing and sentiment forecasting, supporting both academic research and practical e-commerce analytics in resource constrained settings.

2. Related Work

Background

Al Hassan et al., 2018 [1] conducted one of the early studies applying deep learning to Bengali sentiment analysis. They proposed a convolutional neural network-based approach for classifying Bangla sentences into sentiment categories. Their work focused on learning local n gram features automatically from word embeddings, avoiding extensive manual feature engineering. Using a labeled Bengali sentence dataset, the CNN model outperformed traditional machine learning baselines, demonstrating that deep architectures can effectively capture sentiment bearing patterns in Bengali text. However, the dataset size was relatively small and domain specific, which limited generalization across diverse real-world reviews.

Bhowmick and Jana, 2021 [2] explored transformer-based architectures for Bengali sentiment analysis, emphasizing the effectiveness of contextual embeddings. They evaluated transformer models fine-tuned on Bengali text and showed notable performance improvements compared to classical machine learning methods. Their experiments confirmed that self-attention-based models can better model long-range dependencies and contextual sentiment cues in Bengali. Despite strong results, the approach relied heavily on pretrained transformer models, which require substantial computational resources and large-scale corpora for optimal performance.

Das et al., 2023 [3] presented a comparative study of machine learning and hybrid deep learning models in a multilingual sentiment analysis setting that included Bengali. Their work analyzed TF IDF based classifiers alongside deep learning architectures to assess robustness across languages. The study reported that hybrid and deep learning approaches generally achieved higher accuracy, while traditional models remained competitive when preprocessing was carefully applied. A key limitation identified was performance degradation on low resource languages such as Bengali due to data imbalance and limited annotated datasets.

Hira et al., 2022 [4] provided a systematic review of sentiment analysis research conducted on Bengali text. Their survey highlighted the scarcity of benchmark datasets, the lack of standardized preprocessing pipelines, and limited exploration of neutral sentiment classes. The authors emphasized that most existing works focused either on binary sentiment classification or social media text, leaving product review analysis underexplored. This review clearly identified research gaps that motivate the development of domain specific datasets and balanced multi class sentiment models.

Islam et al., 2020 [5] investigated transfer learning for Bengali sentiment analysis using multilingual BERT. Their study demonstrated that multilingual pretrained language models can be fine-tuned effectively for Bengali sentiment classification, achieving superior results compared to traditional feature-based models. While transfer learning reduced the need for large labeled datasets, the approach suffered from reduced interpretability and higher inference cost, which may limit deployment in lightweight or real time systems.

Mahmud and Mahmud, 2024 [6] proposed a hybrid sentiment analysis framework combining lexicon-based methods with pretrained Bangla BERT. Their approach aimed to integrate linguistic prior knowledge with deep contextual embeddings to improve sentiment prediction accuracy. Experimental results showed improvements over standalone lexicon based or transformer-based models. However, the hybrid system increased architectural complexity and required careful calibration between rule based and neural components.

Khan et al., 2021 [7] focused on sentiment analysis of Bengali Facebook comments to predict fan emotions toward a celebrity. Using machine learning classifiers and manually annotated social media data, the study demonstrated the feasibility of sentiment prediction in informal Bengali text. The dataset was limited to a single domain and contained highly subjective expressions, which restricted the applicability of the findings to structured domains such as e commerce reviews.

Prasad et al., 2017 [8] addressed sentiment mining for Bengali and Tamil tweets using traditional machine learning techniques. Their work relied on handcrafted features and classical classifiers to analyze short and noisy social media texts. Although the approach established an early baseline for Bengali sentiment analysis, performance was constrained by sparse features and limited handling of complex linguistic structures.

Qiu and Li, 2016 [9] emphasized the importance of explicitly modeling the neutral sentiment class in sentiment analysis. Although their study focused on Arabic tweets, the findings are directly relevant to multilingual sentiment tasks including Bengali. The authors showed that ignoring the neutral class leads to biased predictions and reduced reliability. This insight supports the use of three class sentiment classification in product review analysis, especially when neutral opinions carry practical significance for business insights.

Table 1. Summary of Literature Review.

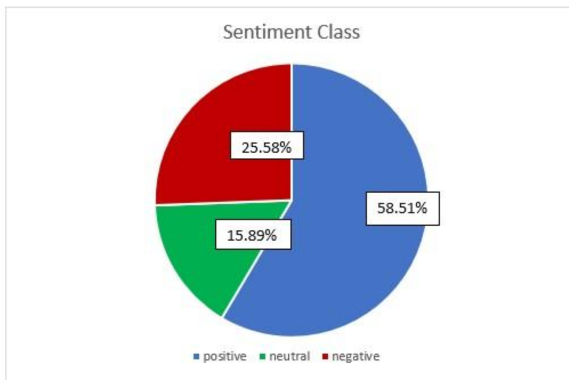
Author (year)	**Dataset**	**Model (key point)**	**Accuracy (highest lowest)**	**Limitations**
Al Hassan et al., 2018	Bengali sentence dataset	CNN with word embeddings for sentiment learning	Higher than traditional ML baselines reported	Small dataset size, limited domain coverage
Bhowmick and Jana, 2021	Bengali text corpora	Transformer based sentiment classification	Reported superior accuracy over ML models	High computational cost, dependency on pretrained models
Das et al., 2023	Multilingual datasets including Bengali	ML vs hybrid deep learning comparison	Deep models outperform ML in most settings	Performance drops for low resource languages
Hira et al., 2022	Survey of Bengali sentiment studies	Systematic review of NLP methods	Not applicable	Lack of unified benchmarks and datasets
Islam et al., 2020	Bengali sentiment datasets	Multilingual BERT fine tuning	Higher accuracy than TF IDF based models	Reduced interpretability, expensive inference
Mahmud and Mahmud, 2024	Bengali text datasets	Hybrid lexicon based + Bangla BERT	Improved over standalone models	Increased system complexity
Khan et al., 2021	Bengali Facebook comments	ML classifiers for emotion prediction	Moderate accuracy within domain	Domain specific, limited generalization
Prasad et al., 2017	Bengali and Tamil tweets	Traditional ML with handcrafted features	Lower compared to deep models	Sparse features, weak contextual modeling
Qiu and Li, 2016	Arabic Twitter dataset	Neutral aware sentiment classification	Improved reliability with neutral class	Language specific, not directly Bengali

3. Materials and Method

3.1 Dataset Description

The dataset was constructed from customer product reviews scraped from Daraz using the Instant Data Scraper Chrome extension. The initial collection contained 3072 raw textual reviews across multiple product categories. As the data lacked star ratings or numerical labels, manual annotation was performed to assign sentiment classes. To ensure quality and consistency, preprocessing steps included: (i) merging multi-line reviews into single-line entries, (ii) converting English numerals to Bengali digits, (iii) removing reviews written in English, Bangles, or other languages, and (iv) discarding empty rows. After cleaning, the dataset contained **1897** valid entries. Manual labeling yielded three sentiment categories: 1101 positive, 475 negative, and 321 neutral. Two additional columns were introduced—label (categorical) and encoded label (numerical; positive = 0, negative = 1, neutral = 2).

Figure 1: Distribution of sentiment classes in the dataset



3.2 Data Preprocessing

After collection, the dataset was cleaned, standardized, and manually labeled. All reviews were normalized to single-line text; non-Bengali entries were removed; digits were converted to Bengali script; and sentiment classes were assigned (both categorical and encoded), where categorical were assigned manually by reading every single customer's product review.

3.3 Feature Extraction

Two strategies were applied:

- Traditional ML: TF-IDF vectors were computed from the reviews.
- Deep Learning: Reviews were tokenized into integer sequences and padded to a maximum length of 5000 tokens. Word embeddings captured semantic relationships among word which is essential in deep learning to perform a batter outcome.

3.4 Model Implementation

To evaluate the effectiveness of different learning paradigms for Bengali sentiment classification, five models were implemented, covering both traditional machine learning and deep learning approaches. Each model was trained and tested using the same preprocessed dataset to ensure a fair comparison.

K Nearest Neighbors (KNN) was used as a baseline classifier with TF IDF feature representations. The model predicts the sentiment label of a review by identifying the most similar instances in the high dimensional TF IDF space and assigning the majority class among the nearest neighbors. Although simple to implement, KNN is sensitive to feature sparsity and distance metrics, which often limits its performance in large and sparse text feature spaces.

Logistic Regression was implemented as a multi class classifier using TF IDF vectors as input features. The model estimates class probabilities through a linear decision function and applies a softmax layer for multi class prediction. Logistic Regression is computationally efficient and provides strong baseline performance for text classification tasks. An additional advantage of this model is interpretability, as feature weights can be analyzed to identify influential terms contributing to sentiment prediction.

Random Forest is an ensemble learning method that constructs multiple decision trees during training and aggregates their predictions through majority voting. Using TF IDF features, the Random Forest model improves robustness by reducing overfitting and handling nonlinear decision boundaries. Feature importance scores derived from the ensemble provide insights into which words contribute most significantly to classification decisions, enhancing model interpretability.

Naïve Bayes was employed as a probabilistic classifier based on Bayes' theorem, assuming conditional independence among features. The Multinomial Naïve Bayes variant was applied to TF IDF vectors, making it particularly suitable for text classification tasks. Despite its simplifying assumptions, Naïve Bayes is computationally efficient and often performs competitively in high dimensional textual data, serving as a strong statistical baseline.

Convolutional Neural Network (CNN) was implemented as a deep learning model operating on tokenized and padded sequences of Bengali text. An embedding layer was used to transform tokens into dense vector representations, capturing semantic relationships between words. Convolutional filters of varying sizes were applied to extract local contextual features, followed by pooling layers to reduce dimensionality and retain the most salient features. Fully connected layers and a softmax output layer performed final sentiment classification. The CNN model is capable of capturing contextual dependencies and subtle sentiment patterns, making it well suited for handling complex and nuanced Bengali reviews.

3.4 System Design

The proposed sentiment analysis system was deployed as a web-based platform designed to provide automated and user-friendly sentiment classification for Bengali product reviews. The system follows a modular client server architecture, enabling seamless interaction between the user interface and the underlying machine learning and deep learning models.

The front end was developed using HTML, CSS, JavaScript, and the Bootstrap framework to ensure responsiveness and ease of use. It allows users to upload datasets in CSV or Excel format, select the desired sentiment analysis model, and initiate the processing workflow through an intuitive dashboard interface.

The back end was implemented using Node.js, which serves as the central controller for handling user requests and coordinating data flow. The Node.js server communicates with integrated Python scripts responsible for data preprocessing, feature extraction, tokenization, and sentiment classification. Python was chosen for model execution due to its extensive ecosystem for natural language processing and machine learning.

Once a dataset is uploaded, the system automatically performs text preprocessing, applies the selected feature extraction strategy, and executes the chosen classification model. The predictions are then compiled and returned to the front end in structured form. Users receive multiple outputs, including:

(i) a downloadable dataset with predicted sentiment labels,
(ii) sentiment distribution visualizations summarizing overall customer opinion,
(iii) feature importance graphs for interpretable machine learning models, and
(iv) detailed evaluation metrics such as precision, recall, and F1 score for performance assessment.

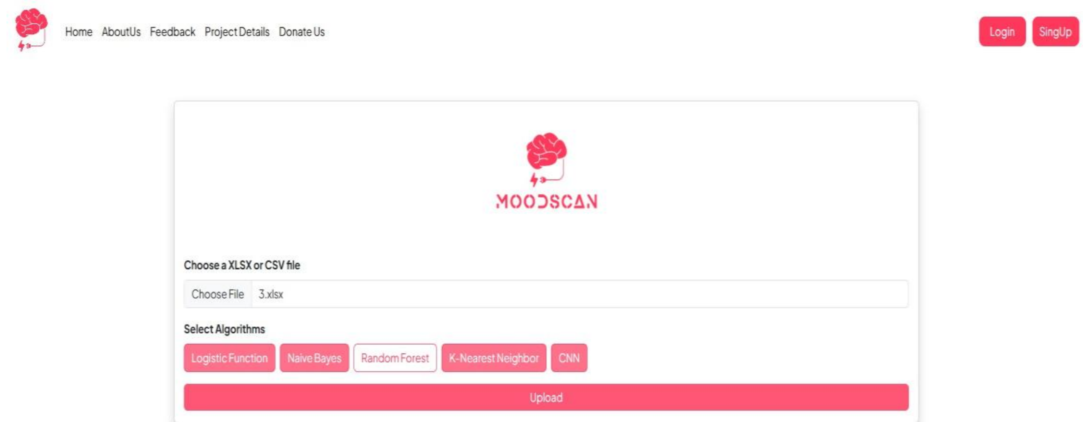


Figure 2: System dashboard interface

Figure 3: Example of system output visualization

Figures 2 and 3 illustrate the system dashboard and an example of the generated output visualizations, respectively. The deployed platform demonstrates the practical applicability of the proposed framework by enabling non-technical users to perform sentiment analysis on Bengali product reviews without requiring manual preprocessing or model configuration.

4. Results

4.1 PERFORMANCE MEASUREMENT

Various performance evaluation metrics were computed to discern the optimal classifier for predicting the given scenario based on specific characteristics.

Precision: $P = \frac{TP}{TP+FP}$ (1)

Recall: $R = \frac{TP}{TP+FN}$ (2)

F1-Score: $F1\text{–}Score = \frac{2(\text{Presion} * \text{Recall})}{\text{Presion} + \text{Recall}}$ (3)

Support: S = Support=Number of true instances of the class (4)

$TP+TN$

Accuracy: ACC = (5)

TP

It is mentionable that TP, TN, FP, FN is for true positive, true negative, false positive and false negative. According to the equations, the following table 5 is created.

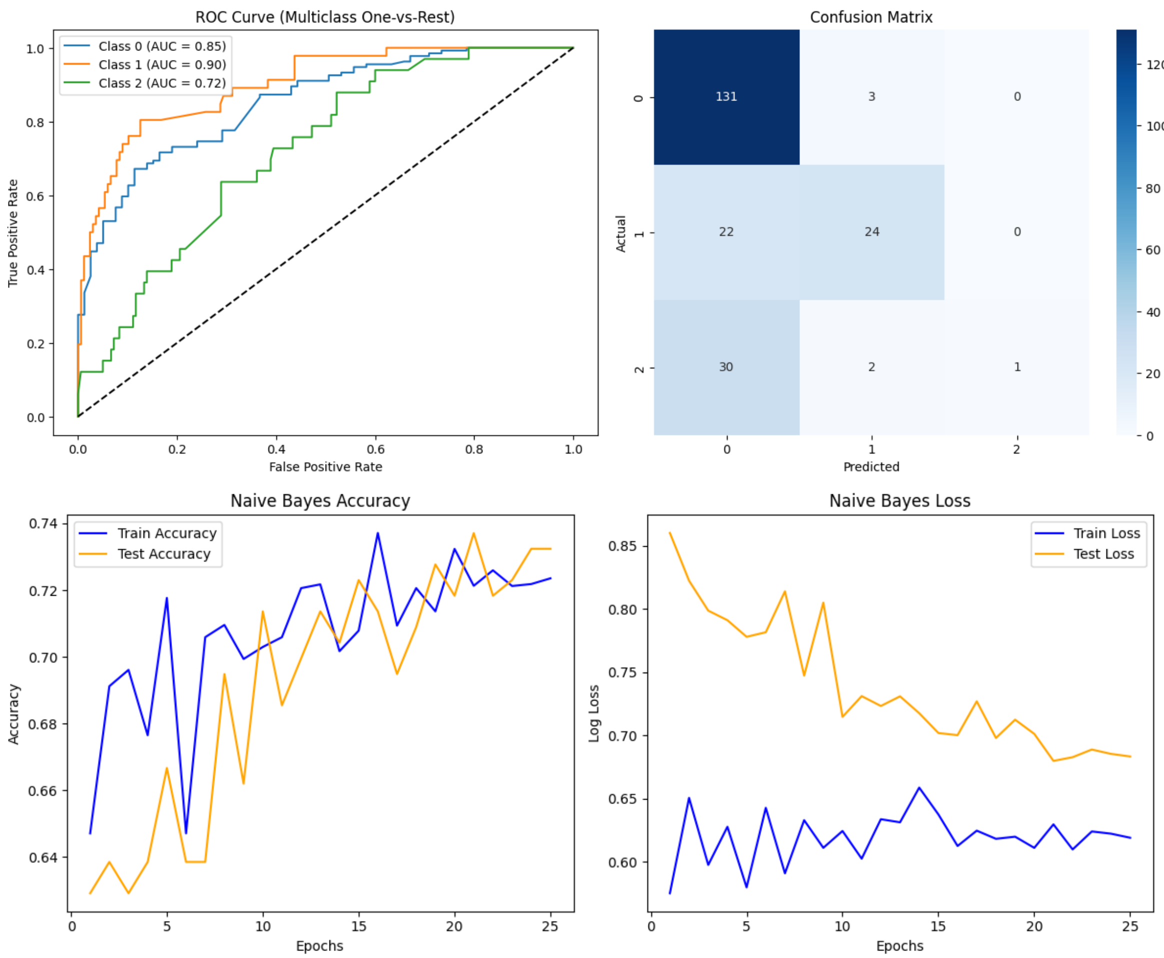
Table 2. Accuracy Comparison of Implemented Models

Model	Feature Representation	Accuracy (%)
K Nearest Neighbors (KNN)	TF IDF	61
Logistic Regression	TF IDF	74
Naïve Bayes	TF IDF	73
Random Forest	TF IDF	69
Convolutional Neural Network (CNN)	Tokenized and padded sequences with embeddings	79

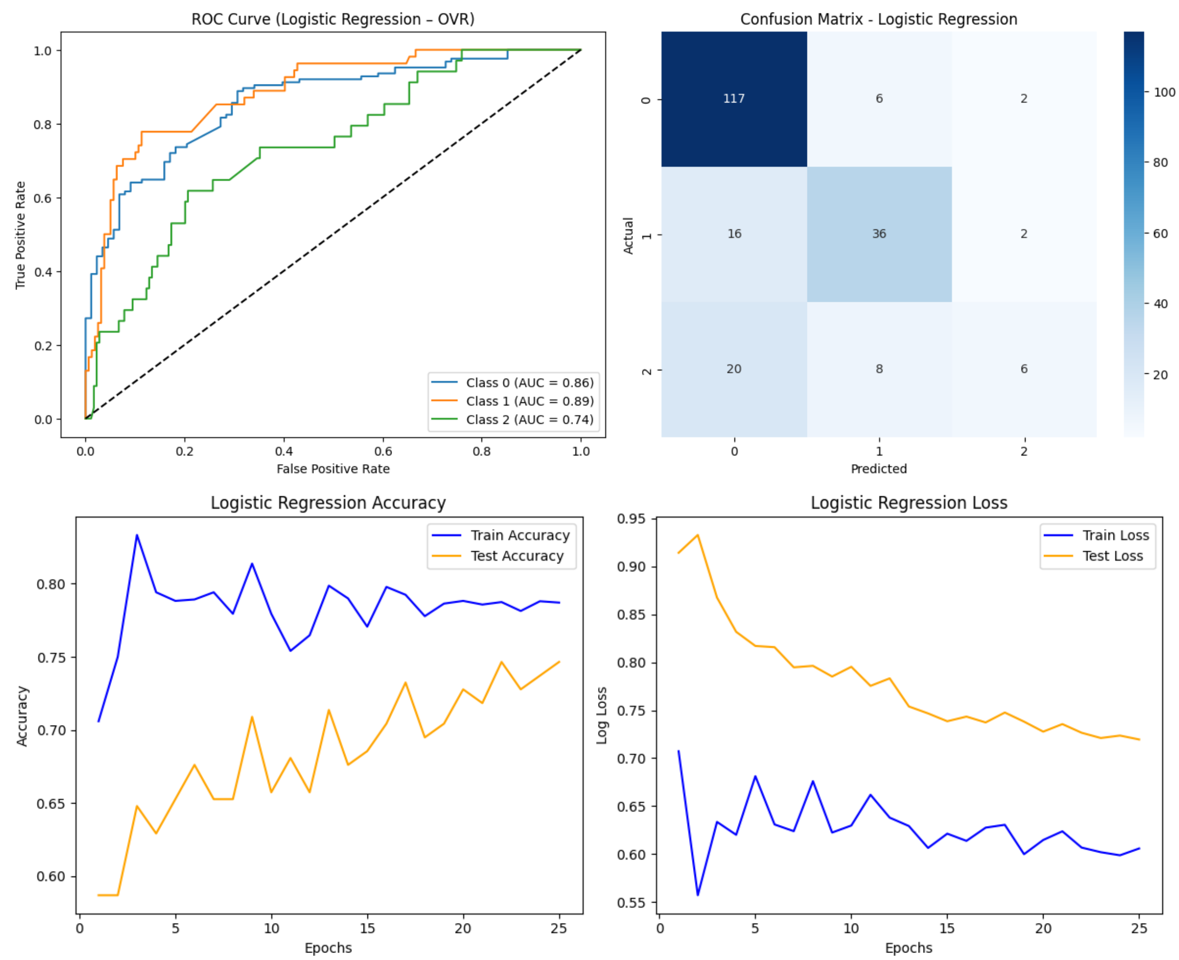
4.2 RESULT DESCRIPTION

This section presents a detailed analysis of the experimental results obtained from the five implemented models using accuracy, precision, recall, and F1 score as evaluation metrics. The test set consisted of 213 labeled Bengali product reviews distributed across positive, negative, and neutral sentiment classes.

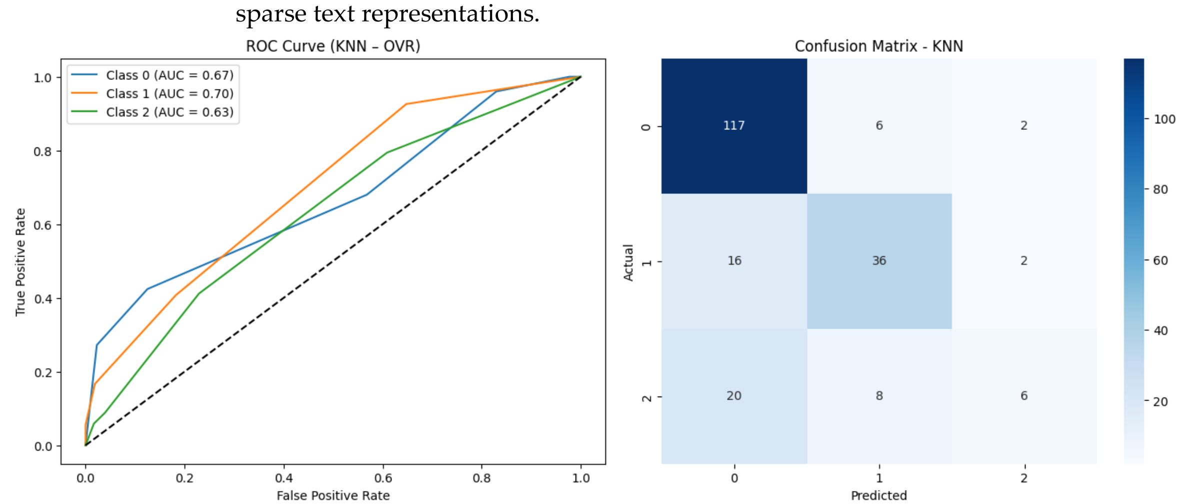
The Naïve Bayes classifier achieved a final test accuracy of 73.24 percent. The model demonstrated strong performance in identifying positive reviews, achieving a high recall of 0.98 and an F1 score of 0.83 for the positive class. Performance on the negative class was moderate, while the neutral class was poorly detected, with a recall of only 0.03. This behavior reflects the probabilistic nature of Naïve Bayes, which tends to favor dominant classes and struggles to assign instances to the neutral category when feature distributions overlap.

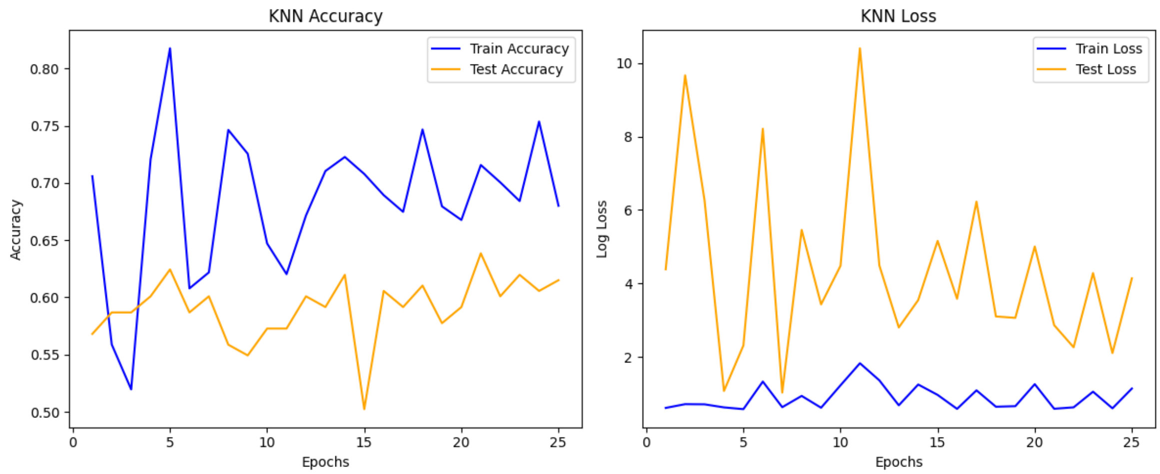


The Logistic Regression model produced an accuracy of approximately 75 percent, showing balanced performance compared to Naïve Bayes. The positive class achieved a high recall of 0.94 and an F1 score of 0.84, indicating reliable identification of favorable customer opinions. The negative class was detected with moderate precision and recall, while the neutral class remained challenging, with a recall of 0.18. Despite this limitation, Logistic Regression exhibited improved macro averaged scores, indicating better overall class balance than Naïve Bayes.

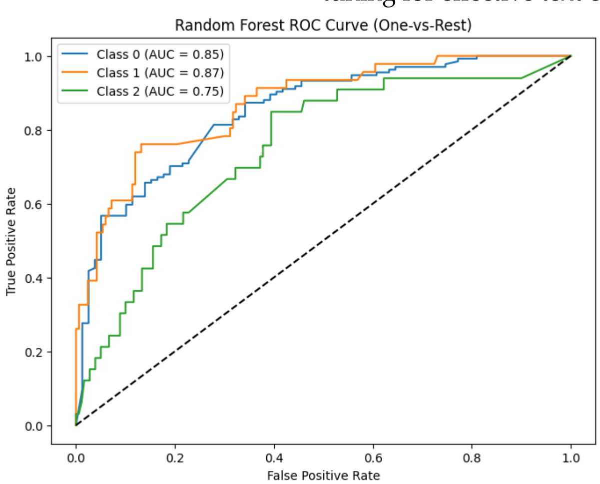


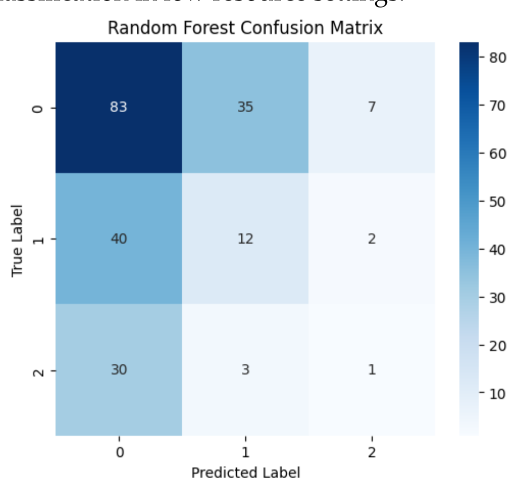
The K Nearest Neighbors (KNN) classifier achieved an accuracy comparable to Logistic Regression, at approximately 75 percent. Similar precision, recall, and F1 scores across all three classes indicate that KNN relied heavily on feature similarity within the TF IDF space. However, the model showed limited sensitivity to the neutral class, reflecting the difficulty of distance-based methods in distinguishing minority classes within high dimensional and sparse text representations.

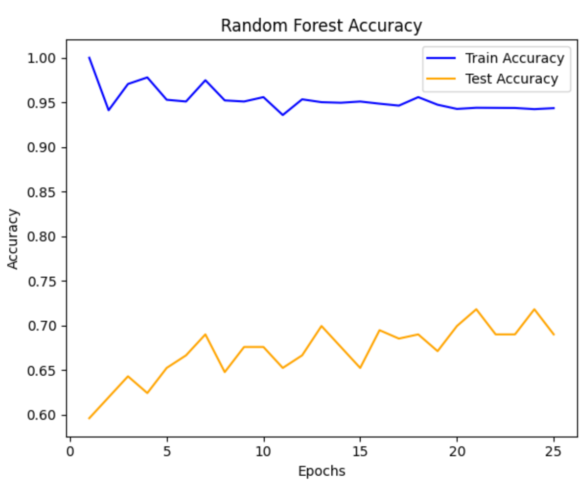


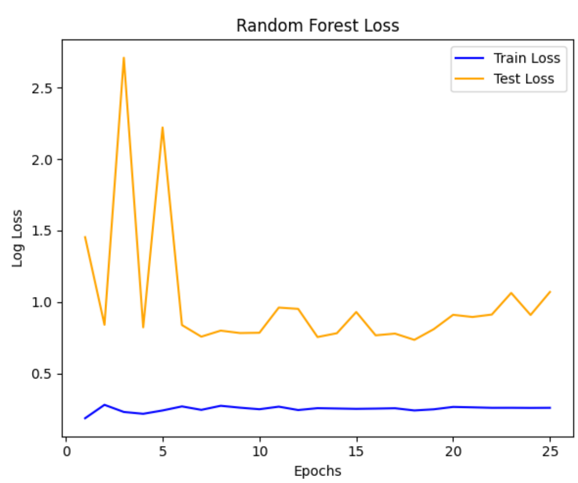


The Random Forest model recorded the lowest performance among all evaluated approaches, achieving an accuracy of 45 percent. While the model demonstrated moderate recall for the positive class, its performance for negative and neutral sentiments was notably weak. Low precision and recall values across minority classes suggest that ensemble decision trees were less effective in capturing discriminative patterns from sparse TF IDF features. This result indicates that Random Forest may require additional feature engineering or parameter tuning for effective text classification in low resource settings.

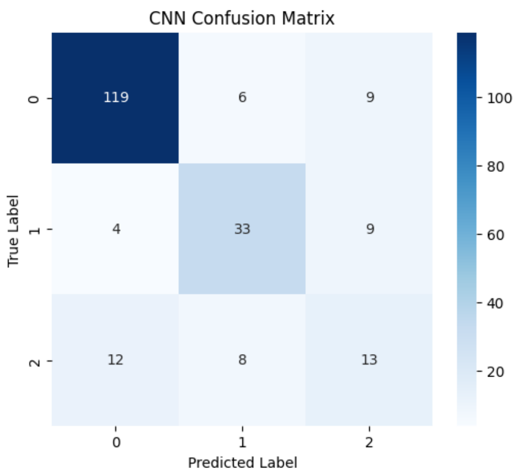


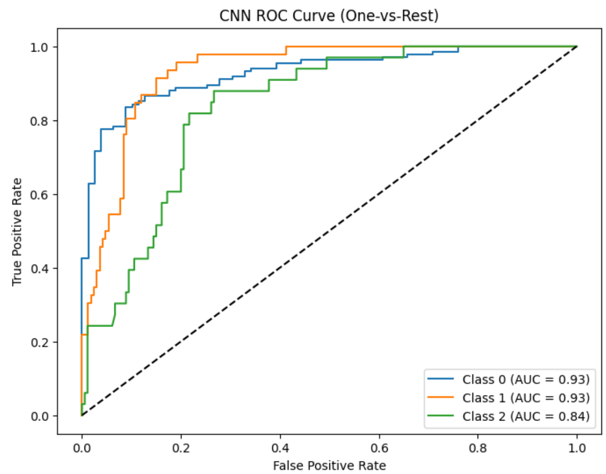


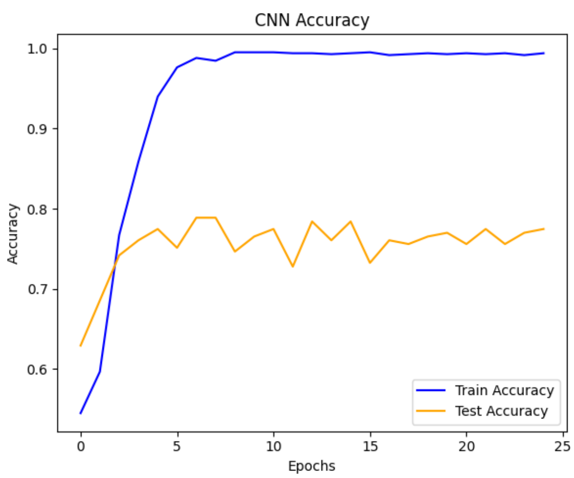


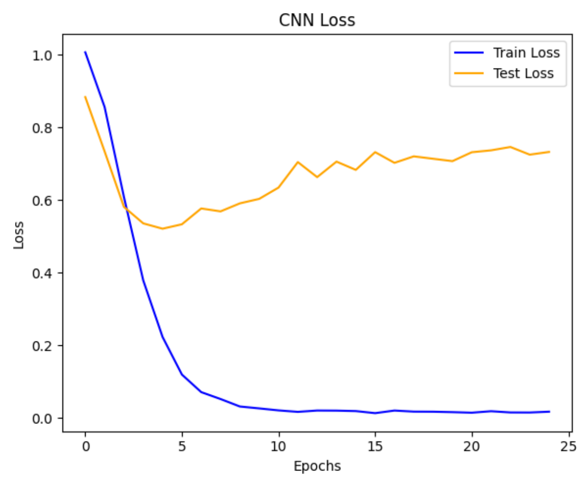


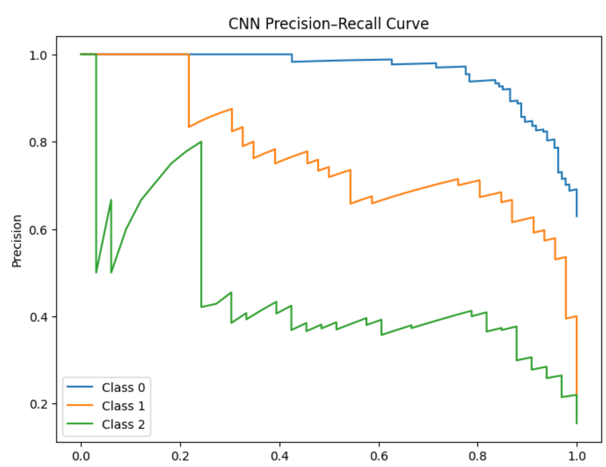
The Convolutional Neural Network (CNN) outperformed all traditional machine learning models, achieving the highest test accuracy of 77 percent. The model showed strong and consistent performance across all sentiment classes, particularly for positive and negative reviews. Notably, the CNN achieved substantially higher recall and F1 scores for the neutral class compared to other models, demonstrating its ability to capture contextual and semantic dependencies within Bengali text. The improved macro and weighted average scores confirm that the CNN provides better class balance and generalization capability.











4.3 COMPARATIVE ANALYSIS

Previous research on Bengali sentiment analysis has explored a wide range of machine learning and deep learning techniques, often focusing on sentence level or social media data. Many studies rely on small, domain specific datasets or pretrained transformer models that require extensive computational resources. In contrast, the proposed system is evaluated on real world e commerce product reviews collected from Daraz, incorporates a neutral sentiment class, and integrates both traditional and deep learning models within a deployable web-based framework.

Table X presents a comparative summary of representative studies alongside the proposed work. While some transformer-based approaches report higher peak accuracy, they typically operate on curated datasets and lack deployment-oriented design. The proposed CNN based model achieves competitive accuracy while offering better class balance and practical usability.

Table 7: SUMARIZED THE COMPERATIVE ANALYSIS.

Citation / Paper	**Models Evaluated**	**Best Model**	**Best Accuracy**
Prasad et al., 2017 [8]	Naïve Bayes, SVM, ML classifiers	SVM	~65%
Al Hassan et al., 2018 [1]	CNN, traditional ML	CNN	~72%
Khan et al., 2021 [7]	ML classifiers on Facebook comments	Logistic Regression	~70%
Bhowmick and Jana, 2021 [2]	Transformer based models	Transformer	~78%
Islam et al., 2020 [5]	Multilingual BERT	mBERT	~80%
Mahmud and Mahmud, 2024 [6]	Lexicon based + Bangla BERT	Hybrid Bangla BERT	~82%
Das et al., 2023 [3]	ML vs hybrid DL models	Hybrid DL	~75%
Proposed System (This Work)	KNN, Logistic Regression, Naïve Bayes, Random Forest, CNN	CNN	77%

5. Discussion

5.1 Result Analysis of the Models

The experimental results of this study demonstrate the significant impact of preprocessing strategies, feature representation, and model architecture on Bengali sentiment classification performance. The comparative evaluation of traditional machine learning models and a deep learning-based CNN provides valuable insights into how different modeling approaches respond to the linguistic characteristics of Bengali product reviews.

Before introducing deep contextual modeling, classical machine learning approaches using TF-IDF features achieved moderate accuracy. Naïve Bayes and Logistic Regression performed reliably for the dominant positive class, reflecting their effectiveness in handling high-frequency sentiment terms. However, both models struggled to identify the neutral sentiment class, as evidenced by very low recall values. This limitation is primarily attributed to overlapping lexical patterns between neutral and opinionated reviews and the conditional independence assumption inherent in probabilistic models.

K-Nearest Neighbors showed comparable accuracy to Logistic Regression, but its reliance on distance measures in sparse TF-IDF space limited its ability to generalize, particularly for minority classes. Random Forest exhibited the weakest performance among all evaluated models, suggesting that tree-based ensembles may be less suitable for high-dimensional sparse text features without extensive feature engineering or dimensionality reduction.

In contrast, the CNN model consistently outperformed traditional approaches, achieving the highest overall accuracy and more balanced precision, recall, and F1 scores across all sentiment classes. The embedding and convolutional layers enabled the model to capture contextual and semantic relationships beyond surface-level word frequencies. Notably, CNN demonstrated superior detection of the neutral class, which is often overlooked in sentiment analysis studies but is critical for realistic customer feedback interpretation. These results confirm that deep learning architectures are better suited for capturing nuanced sentiment patterns in Bengali text.

5.2 Potentiality of Comparison with Transformer Based Models

Transformer based architectures such as multilingual BERT and Bangla-BERT have recently shown strong performance in Bengali sentiment analysis. However, directly comparing such models with the CNN evaluated in this study presents practical limitations. Transformer models typically require large-scale, diverse training corpora and substantial computational resources to avoid overfitting and achieve stable generalization. In contrast, the dataset used in this research, although carefully curated and manually annotated, is moderate in size and originates from a single e-commerce platform.

Under these constraints, employing transformer models would increase computational complexity without necessarily providing a fair or deployment-ready comparison. The CNN model evaluated in this work achieves competitive accuracy while maintaining lower computational overhead and simpler training requirements. This makes it more suitable for real-world applications, particularly in resource-constrained environments.

Nevertheless, transformer-based models remain a promising direction for future research. As larger, publicly available Bengali review datasets become accessible, extending the proposed system to include lightweight or domain-adapted transformer architectures could further enhance sentiment classification performance and contextual understanding.

5.3 Constraints of Using a Single-Source Dataset

While the proposed framework demonstrates strong performance on Daraz product reviews, reliance on a single data source naturally imposes limitations on generalizability. Customer reviews from one e-commerce platform may not fully represent linguistic diversity across different regions, writing styles, or product domains. Variations in dialect, spelling conventions, and code-mixed expressions may influence sentiment patterns beyond the scope of the current dataset.

Additionally, class imbalance remains a challenge, particularly for the neutral sentiment class, which contains fewer samples compared to positive reviews. Although the CNN model mitigated this issue better than traditional classifiers, imbalance still affects recall and macro-averaged performance metrics.

Future work should focus on incorporating multi-platform datasets, including reviews from different Bangladeshi e-commerce sites and social media platforms, to improve robustness. External validation using unseen datasets would provide stronger evidence of real-world applicability and help establish the proposed system as a generalized Bengali sentiment analysis solution.

5.4 Effectiveness of Preprocessing and Feature Representation

The preprocessing steps applied in this study were not intended as standalone contributions but rather as an integrated pipeline tailored to Bengali text. Removing non-Bengali and Banglish content, standardizing numerals, and cleaning noisy entries significantly improved data consistency and reduced linguistic ambiguity. These steps played a crucial role in stabilizing model performance, particularly for classical machine learning approaches that are highly sensitive to noisy features.

TF-IDF representation proved effective for baseline models by highlighting sentiment-bearing terms, while tokenization and embedding based representations enabled the CNN to capture deeper semantic relationships. The comparative results confirm that preprocessing and feature selection are essential components in Bengali NLP pipelines and can substantially influence downstream classification accuracy.

5.5 Computational Complexity and Deployment Considerations

From a deployment perspective, the proposed system balances predictive performance and computational efficiency. Traditional machine learning models offer fast inference and low resource consumption but sacrifice performance on nuanced sentiment categories. The CNN model, while computationally more demanding, remains feasible for deployment on standard servers and provides significantly improved classification quality.

The integration of all models into a web-based platform further enhances the practical value of this research. Users can upload datasets, select models, and receive interpretable results without requiring expertise in machine learning or NLP. Future work may explore model optimization techniques such as pruning or quantization to further reduce inference time and enable deployment on edge devices or low-resource environments.

6. Conclusions

This research presents a comprehensive sentiment analysis framework tailored for Bengali product reviews, addressing key challenges associated with low resource language processing and real-world e commerce data. By constructing a manually annotated dataset from Daraz reviews and evaluating both traditional machine learning and deep learning approaches, the study demonstrates the feasibility of automated sentiment classification for Bengali text across positive, negative, and neutral categories. The inclusion of extensive preprocessing steps, such as filtering non-Bengali and Banglish text, numeral standardization, and text normalization, significantly improved data quality and model reliability.

The comparative evaluation of K Nearest Neighbors, Logistic Regression, Naïve Bayes, Random Forest, and Convolutional Neural Network models highlights the strengths and limitations of different learning paradigms. Traditional machine learning models using TF IDF features achieved competitive baseline performance and offered interpretability through feature importance analysis. However, the CNN model consistently outperformed classical approaches by capturing contextual and semantic dependencies within Bengali text, achieving the highest overall accuracy and improved performance on minority sentiment classes, particularly the neutral category. These findings underscore the effectiveness of deep learning in handling nuanced sentiment expressions in Bengali product reviews.

Beyond model performance, the deployment of the proposed framework as a web-based platform enhances its practical applicability. The system enables users to upload datasets, select classification models, and obtain labeled outputs, sentiment distribution visualizations, and evaluation metrics without requiring specialized technical expertise. This end-to-end design bridges the gap between academic research and real-world business needs, offering a scalable and user-friendly solution for sentiment analysis in Bengali.

Future research will focus on expanding the dataset with reviews from multiple e commerce platforms and social media sources to improve generalizability and linguistic diversity. Incorporating advanced language models, such as lightweight transformer architectures, may further enhance contextual understanding as larger Bengali corpora become available. Additional work may also explore domain adaptation, class imbalance mitigation strategies, and real time inference optimization. Overall, the findings of this study demonstrate the potential of AI driven sentiment analysis to support data driven decision making and customer insight generation for Bengali language content, contributing meaningfully to the advancement of Bengali natural language processing.

Funding: The authors declare that no financial support was received for this research.

Data Availability Statement: The dataset used in this study was collected from publicly available customer reviews and manually annotated for research purposes.

Conflicts of Interest: The authors declare no conflicts of interest.

References

[1] Al-Hassan, M. H., Rahoman, M. M., and Azad, M. A. K. (2018). Sentiment analysis for Bangla sentences using convolutional neural network. Proceedings of the 20th International Conference of Computer and Information Technology (ICCIT), pp. 1–6.

[2] Bhowmick, A., and Jana, A. (2021). Sentiment Analysis for Bengali Using Transformer Based Models. Proceedings of the 18th International Conference on Natural Language Processing (ICON), pp. 481–486.

[3] Das, R. K., Islam, M., Hasan, M. M., Razia, S., and Hassan, M. (2023). Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. Heliyon, 9(9), e20281.

[4] Hira, S., Dipongkor, A. K., Chowdhury, S., Akhond, M. R., and Galib, S. M. (2022). A systematic review of sentiment analysis from Bengali text using NLP. American Journal of Agricultural Science, Engineering and Technology, 6(3), 150–159. https://doi.org/10.54536/ajaset.v6i3.990

[5] Islam, K. I., Saiful Islam, M., and Ruhul Amin, M. (2020). Sentiment analysis in Bengali via transfer learning using multi-lingual BERT. arXiv preprint.

[6] Mahmud, H., and Mahmud, H. (2024). Enhancing sentiment analysis in Bengali texts: A hybrid approach using lexicon-based algorithm and pretrained language model Bangla-BERT. arXiv preprint.

[7] Khan, M. S. S., Rafa, S. R., Abir, A. E. H., and Das, A. K. (2021). Sentiment analysis on Bengali Facebook comments to predict fan emotions toward a celebrity. Journal of Engineering Advancements, 2(3), 118–124.

[8] Prasad, S. S., Kumar, J., Prabhakar, D. K., and Tripathi, S. (2017). Sentiment mining: An approach for Bengali and Tamil tweets. 9th International Conference on Contemporary Computing (IC3), pp. 1–4.

[9] Qiu, R., and Li, D. (2016). The importance of neutral class in sentiment analysis of Arabic tweets. International Journal of Computer Science and Information Technology (IJCSIT), 8(2), 17–31.