Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 4, August 2025, pp. 2782∼2792
ISSN: 2302-9285, DOI: 10.11591/eei.v14i4.9122

A machine and DL approach for classifying customer sentiments from online shopping reviews in Bangla text

Md. Arifur Rahman Rejuan[1], Md Assaduzzaman[1], Nafiz Fahad[2], Md. Jakir Hossen[3], Md. Rahmatul Kabir Rasel Sarker[1]
[1]Department of Computer Science Engineering, Faculty of Information Science and Technology, Daffodil International University, Dhaka, Bangladesh
[2]Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia
[3]Center for Advanced Analytics (CAA), COE for Artificial Intelligence, Faculty of Engineering & Technology (FET), Multimedia University, Melaka, Malaysia

Article Info

Article history:

Received Aug 8, 2024
Revised Jan 17, 2025
Accepted Mar 9, 2025

Keywords:

Bangla text
Deep learning techniques
Machine learning techniques
Sentiment analysis
Shopping reviews

ABSTRACT

Due to the widespread availability of the internet all across the world, people prefer shopping online rather than going to a shop. There are various online marketplaces available in Bangladesh, like Daraz, Pickaboo, Rokomari, Othoba, Bikroy, Food Panda, and Robi Shop. With the increasing quantity of customers on online shopping platforms, the number of product reviews also increases with it. Data is classified utilizing machine learning (ML), deep learning (DL), transfer learning, and other data mining algorithms to facilitate the customer's comprehension of the primary subject of the review before making a purchase. Natural language processing techniques are employed to categorize data in any given language for such issues. There are no Bengali shopping review datasets available on online sites. So, we manually collected a dataset of 2,600 reviews. In this paper, reviews are classified into 5 categories (satisfied, very satisfied, not satisfied, fairly satisfied, and satisfied but delivery problem). DL (long short-term memory (LSTM) and convolutional neural network (CNN)) and ML (support vector machine (SVM), random forest (RF), gradient boosting (GB), and extreme gradient boosting (XGBoost)) model have been applied. Among the DL models, CNN has the best accuracy (91.27%), and the RF classifier provides the highest accuracy (84.39%) out of all the ML models.

Corresponding Author:

Md. Jakir Hossen
Center for Advanced Analytics (CAA), COE for Artificial Intelligence
Faculty of Engineering & Technology (FET), Multimedia University
75450 Melaka, Malaysia
Email: jakir.hossen@mmu.edu.my

1. INTRODUCTION

The internet has become a crucial part of human life, with people increasingly relying on it for information and services. Online e-commerce marketing and transactions are expanding globally due to high-speed internet accessibility. As of January 2018, there were around 80.83 million internet users in Bangladesh, with 30 million using social media [1]. Many people are selling products online instead of physically visiting the market, and customers are also shopping online. A new dimension to online product shopping has been added, with individual reviews written on each product. These reviews can include product condition, delivery time,

and more. Sentiment analysis, a machine learning (ML) technique, is becoming increasingly essential for e-commerce business. Although many studies have applied ML techniques for sentiment analysis, few have focused specifically on classifying Bengali-language reviews in the context of e-commerce. ML technologies facilitate the learning of emotion recognition without human intervention by using textual emotion samples as training data. Customer reviews can be utilized in various applications, including social networking platforms, in terms of identifying sentiment [2]. ML has gained attention for its potential impact on e-commerce operations, but it remains unclear how to optimize neural networks for massive data recognition and ML. However, some believe that over the next five years, ML will significantly affect radiology and e-commerce [3].

Like other developed countries of the world, online shopping sites and online shopping tendencies are increasing day by day in Bangladesh. In this way, as time goes by, the e-commerce sites are increasing and the quantity of new products on various e-commerce websites or apps is increasing. However, categorizing these Bengali e-commerce site reviews is not done well, meaning that there is little work to be done to determine which reviews fall into which categories. Our research addresses these gaps by offering a new dataset and classification approach for Bengali-language e-commerce reviews, contributing to both ML and sentiment analysis fields.

In the modern research field, several new methodologies for text analysis have been deployed. In a study article, Bitto *et al.* [4] performed sentiment analysis on feedback from clients from Bangladeshi food delivery services using ML and deep learning (DL) methods. In comparison to other models such as convolutional neural network (CNN) and extreme gradient boosting (XGBoost), the long short-term memory (LSTM) algorithm yielded the greatest accuracy of 91.07%. A sentiment analysis dataset from the e-commerce site Daraz, with an emphasis on reviews written in Bangla, was presented by Shanto *et al.* [5]. A number of ML and DL algorithms were benchmarked in their study, and Bangla-bidirectional encoder representations from transformers (BERT) produced the top results, scoring 94.5% on binary classification tests. Akter *et al.* [6] have proposed a research paper mainly reviewing customer reviews from Bengali e-commerce websites Daraz using the five-ML techniques and the K-nearest neighbors (KNN) algorithm is the best method with an accuracy of 96.25%. Khan *et al.* [7] conducted sentiment analysis on Bengali text to detect depression-related posts using document-level classification. They employed various ML algorithms, with Multinomial Naive Bayes achieving the highest accuracy 86.67%. The study emphasizes the challenges of Bengali text preprocessing and the effectiveness of supervised learning in low-resource language sentiment analysis, particularly for identifying ”happy” or ”sad” expressions linked to mental health. Archana and Thambusamy [2] built several models in their study work based on the remarks of online purchasing consumers utilizing artificial intelligence, i.e., ML methods. The support vector machines (SVM) model is the most effective method proposed by the researchers. Bharathi *et al.* [3] explored in their research work primarily three ML strategies to figure out the sentiment based on online feedback of Amazon items. From them, the SVM algorithm gave the highest result. Ullah *et al.* [8] developed a research paper mainly analyzing any type of Amazon product review and product quality measures using some DL techniques. Among all algorithms, the self-developed quantized lightweight enhanced BERT (QLeBERT) algorithm gave the highest result of 91%. In the work put forth by Moon *et al.* [9] proposed mainly analyzing customer's online shopping satisfaction level using ML techniques. 4 thousand data points were collected to analyze customer satisfaction. Among all algorithms, the highest results were obtained using the Apiorir algorithm. Barua *et al.* [10] explored a research paper that mainly analyzed customers' online shopping and online in-store shopping satisfaction using ML techniques. Among the said algorithms, the highest results were obtained using the gradient boosting (GB) machine.

Our main contribution to this work is collecting datasets from various e-commerce websites and apps, which had not been done before in the context of Bengali sentiment analysis. Then all data is labeled manually with five different types of sentiment by reading each data point. Datasets have an issue with imbalanced data points in each class. As a preprocessing technique, the RandomOverSampler method has been used to solve the issue. Additionally, to check that our suggested model operates best, we performed an ablation investigation. This enabled us to rigorously analyse the influence of various components of the model, resulting to a more refined and efficient method to sentiment classification in Bengali-language review.

This research paper is structured as follows: section 2 describes the method, covering the data collection, preprocessing methodologies, and model architectures employed. Section 3 presents the experimental setup and outcome analysis, where the training environment and model performance measures are provided. Section 4 discusses the ablation research, examining the contribution of separate components to the model's performance and a comparative study with past studies and model deployment, where we deploy our model in

a web application. Section 5 presents the discussion on conclusion, which explains the larger implications of our results and future research initiatives.

2. METHOD

The main objective of our research is to further improve the online shopping experience by classifying the review, which is written in Bengali text. The working process structure is shown in Figure 1. The project begins with the data collection phase. Completing data collection data are labeled and applied best matching data processing and split the dataset with 70% for training, 20% for testing, and 10% for validation. Several DL (LSTM and CNN) and ML (SVM, RF, GB, and XGB) models are implemented and chose the best-performed model based on performance analysis and ablation studies. The primary objective of the research is to discover the most effective algorithm for sentiment classification in Bangla text.

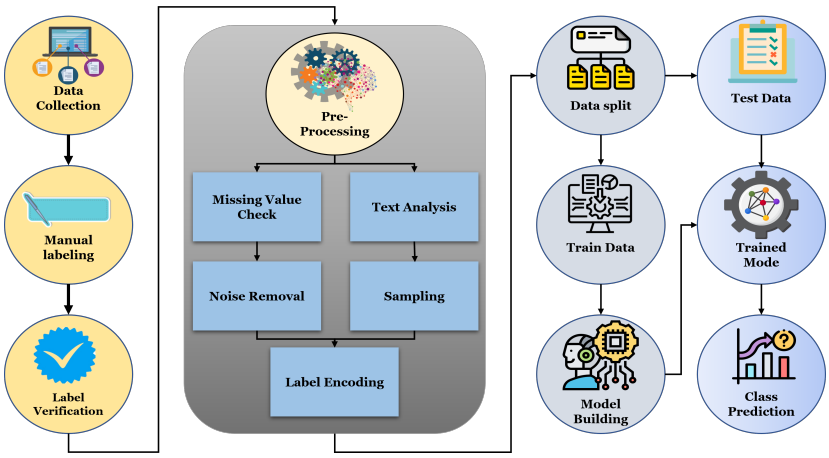

Figure 1. Proposed method of the sentiment analysis

2.1. Describe the dataset

The dataset was gathered manually from social media marketing and a variety of e-commerce websites. After visiting several websites and applications, such as Pikabo, Gorer Bazar, Facebook, Instagram, and Daraz, we gathered feedback from customers in Bangla text about their purchasing experiences. All the collected reviews were separated into specific classes. The main dataset used in this paper consists of four columns: platform, product, reviews, and category. The dataset has a total of 5 marketplaces and 13 different product reviews. There are five categories under the sentiment category: satisfied, very satisfied, not satisfied, fairly satisfied, and satisfied but delivery problem. The collection has 2,603 titles in total. About 2,040 comments have been made on Daraz Marketplace. There are around 452 comments on the Facebook marketplace. There are around 25 comments on the Picabo marketplace. There are around 4 comments on the Ghorer Bazar marketplace.

2.2. Preprocessing

This study applies preprocessing techniques, such as null value elimination, removing unwanted noise (word, emoji, and stop-word), tokenization, and sampling.

2.2.1. Drop null value

Dealing with missing or null values is a crucial preprocessing step in ML, since several ML algorithms are unable to handle missing input. There are several methods that may be used to handle missing values in the dataset. Eliminating the null values in between might be a laborious process. This technique is used to eliminate rows or objects that have null values.

2.2.2. Removing unwanted noise (word, emoji, and stop-word)

An essential phase in the field of natural language processing (NLP) is text preprocessing. The process involves performing data cleaning and transformation on text data to ensure its compatibility with ML algo-

rithms [11]. The particular preprocessing step we use involves eliminating Bengali numbers, English words, emojis, and stop-words. This process makes the dataset more usable for our learning. Figure 2 shows sample data before and after removing noise.

No	Original Text	Label	Cleaned Text
1	অসাধারণ একটি লেন্স। ম্যাক্রো ফটোগ্রাফির জন্য নিয়েছি, অন্য গুলো খুলেও দেখিনি। মাত্র 100 টাকায় 1000 টাকার জিনিস পেয়ে গিয়েছি। [A great lens. Took it for macro photography, didn't even open the others. I got 1000 rupees worth of only 100 rupees.]	খুবই সন্তুষ্ট [Very Satisfied]	অসাধারণ একটি লেন্স ম্যাক্রো ফটোগ্রাফির জন্য নিয়েছি অন্য গুলো খুলেও দেখিনি মাত্র টাকায় টাকার জিনিস পেয়ে গিয়েছি [A wonderful lens Took for macro photography have not opened the others got rupees worth of only rupees]
2	দারাজ ও সেলার কে অসংখ্য ধন্যবাদ এবং খুব তাড়াতাড়ি প্রডাক্টটি হাতে পেয়েছি। 😊 😊 [Many thanks to Daraz and seller and received the product very quickly. 😊 😊]	খুবই সন্তুষ্ট [Very Satisfied]	দারাজ সেলার কে অসংখ্য ধন্যবাদ খুব তাড়াতাড়ি প্রডাক্টটি হাতে পেয়েছি [Many thanks to Daraz seller received the product very quickly]

Figure 2. Sample clanned text

2.2.3. Tokenization

Tokenization refers to the procedure of segmenting text into distinct words or tokens in the fields of ML and NLP. Each token generally corresponds to a word or a sub-word unit, including a punctuation mark, symbol, or portion of a word. Tokenization is a fundamental NLP preprocessing method that turns unstructured text input into a ML algorithm-acceptable structure [12].

2.2.4. Sampling

One of the hardest problems in ML and DL approaches is learning a system using unbalanced data. Numerous approaches have been put out in an attempt to address this issue. In the field of NLP, selecting a subset of data from a larger dataset is known as ”sampling” (NLP). In the data balancing preprocessing, there are two sampling inclinations: reduce the number of data points or duplicate data points from minority classes. One is known as under-sampling, and the other is oversampling [13]. Oversampling achieves the feat of equivalent distribution for all sentiment classifications [8]. In large datasets, it's best to use under-sampling, but in medium datasets, using under-sampling makes the dataset a small dataset. So, in this case, oversampling is best for this dataset. In Figure 3, we can see that we have an issue with data unbalancing. Therefore, we use the oversampling strategy in our dataset to balance the data points. After applying oversampling, in Figure 3 we can show all classes become balanced, containing 932 data points in each class.

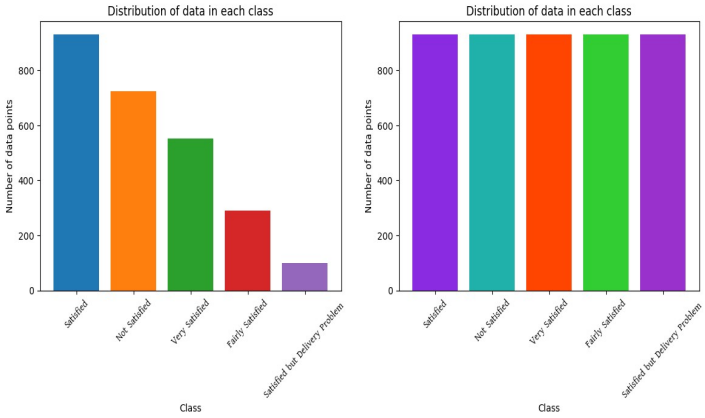

Figure 3. Before and after sampling

2.3. Model selection and description

The study explores the techniques for bangla text sentiment analysis taken by several machine and DL algorithms. The models include RF, GB, XGBoost, SVM, CNN, and LSTM. The preprocessed dataset was used to train each model, and cross-validation was used to adjust the hyperparameters. These models were selected for their proved effectiveness in text classification problems. CNN was utilised for its efficiency in collecting

spatial correlations among words, and LSTM was utilised for its efficacy in managing sequential data. We did an ablation study to highlight the implications of various components of our models. We painstakingly deleted crucial properties and compared the ensuing performance against the complete model. This permitted the discovery of the variables that most substantially boosted model performance, resulting in the optimum design of our suggested model.

2.3.1. Convolutional neural network

A CNN is a ML algorithm that processes sequential data like text. It starts with an embedding layer, convolutional layers, and filters to capture local patterns and features. The output layer produces predictions, with dropout layers to prevent overfitting. CNN models are widely used due to their computational efficiency and ability to operate on various devices [14]. They are used in various businesses for tasks like document analysis, picture recognition, object identification, climate comprehension, and facial recognition. Figure 4 shows the working process architecture of CNN.

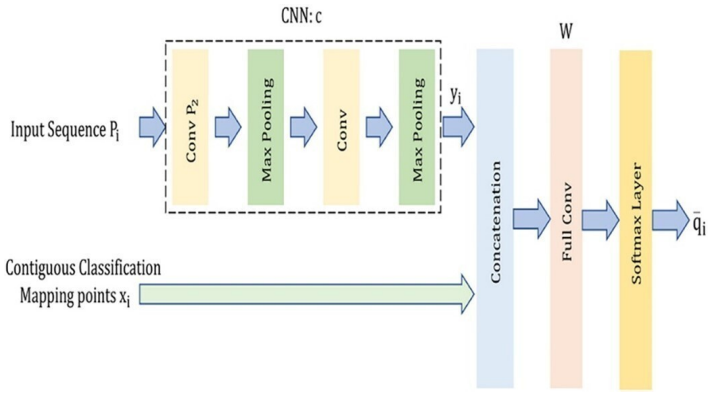

Figure 4. Working process architecture of CNN [15]

2.3.2. Long short-term memory

The recurrent neural network (RNN) LSTM network addresses gradient issues in conventional RNNs. Its relative sensitivity with gap length makes it useful for text categorization tasks. To train LSTMS, each word in the text is taught as a time step [16]. Tokenization and sound embedding methods transform text into numerical displays, which are then sent to the LSTM. The LSTM processes every word in the sequence simultaneously, producing hidden states for each syllable [17]. The output is passed through a softMax activation function and a fully connected layer to produce a class score.

2.3.3. Support vector machine

SVMs are widely used in ML for applications such as handwriting recognition, web pages, email categorization, face identification, intrusion detection, and gene classification. They are a supervised learning technique that can handle regression and classification on both linear and non-linear data [18]. SVMs are particularly effective in categorization situations and aim to find a hyperplane in an n-dimensional space that cleanly classifies data points [19].

2.3.4. Extreme gadient boosting classifier

XGBoost classifier is a robust ML algorithm known for its efficiency and accuracy in classification tasks. It uses a gradient-boosting framework, regularization techniques, parallel processing, and customizable objective functions [20]. Its popularity is due to its high predictive performance, scalability, and versatility in handling diverse datasets, making it a popular choice for competitions and real-world applications.

2.3.5. Gradient boosting

The GB approach is a powerful ML technique used to reduce bias error in models [21]. This method, unlike AdaBoost, does not allow for a base estimator but instead uses a fixed decision stump as the basis estimator. Similar to AdaBoost, the n-estimator can be modified, but the default value is 100 if the number is omitted. This approach is particularly effective in addressing variance and bias errors in ML systems.

2.3.6. Random forest algorithm

The supervised learning method known as RF is composed of decision trees. This approach is mostly used in the fields of classification and regression. The method chooses many sample types, averages the regression, and classifies the samples with the highest vote to create a decision tree. In addition, averages are employed in dataset analysis to improve forecast precision [22].

3. RESULT ANALYSIS

The performance of multiple ML and DL models was examined, with CNN appearing as the most effective model for Bengali sentiment categorization, obtaining the greatest accuracy of 91.27%.

3.1. Accuracy

The standard or known value is called accuracy. It speaks about the measured value's proximity [15].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

3.2. Precision

Precision is the degree of accuracy that exists between at least two or more measurements. Accuracy is not dependent. Precision is defined as a positive prediction value. It is a minor part of all the situations [23].

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3.3. Recall

It is the degree of sensitivity to recall. Additionally, several relevant instances were acquired. You may think of it as a likelihood [24].

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

3.4. F1-measure

The F-score, also known as the F1-measure, quantifies the accuracy of a model on a given dataset. The F-score is frequently employed to assess the performance of information retrieval systems and ML models, particularly in the field of NLP [25].

$$F1\text{-}Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

3.5. Performance analysis

This section presents the study on ML and DL methods, employing four ML approaches and two DL algorithms in this study. The DL methods used are CNN and LSTM. The ML approaches include SVM, RF, GB algorithm, and XGBoost. Table 1 shows that the accuracy of CNN is 91.27%, while LSTM is 88.84%. The SVM has an accuracy of 74.5%, while RF has an accuracy of 84.39%. The GB algorithm has an accuracy of 76.37%, and XGBoost has an accuracy of 81.17%.

Table 1. Models performance analysis

Models	Accuracy	Precision	Recall	F1-measure	Findings
CNN	91.27	91.13	91.27	91.17	Highest accuracy
LSTM	88.84	88.62	88.84	88.59	Lowest accuracy
SVM	74.5	74	75	74	Lowest accuracy
RF alassifier	84.39	84	84	84	Lowest accuracy
GB algorithm	76.37	76	76	76	Lowest accuracy
XGBoost classifier	81.17	81	81	81	Lowest accuracy

3.6. Proposed model result analysis

Completing ablation studies, we have built our model with these hyperparameters, embedding dim=400, filters=128, kernel size=5, dropout rate=0.2, activation=relu and last layer SoftMax, batch size =32, loss=sparse-categorical-crossentrophy, optimizer=adam, and learning rate=0.001. After applying all of this hyperparameter my model performs 91.27% accuracy. In the below section, we can see the visualization of the confusion matrix and graph of receiver operating characteristic (ROC) curve graphs of the fine-tuned model. The visualization graph of the confusion matrix and ROC curve in Figure 5 also shows that our model has given a good performance.

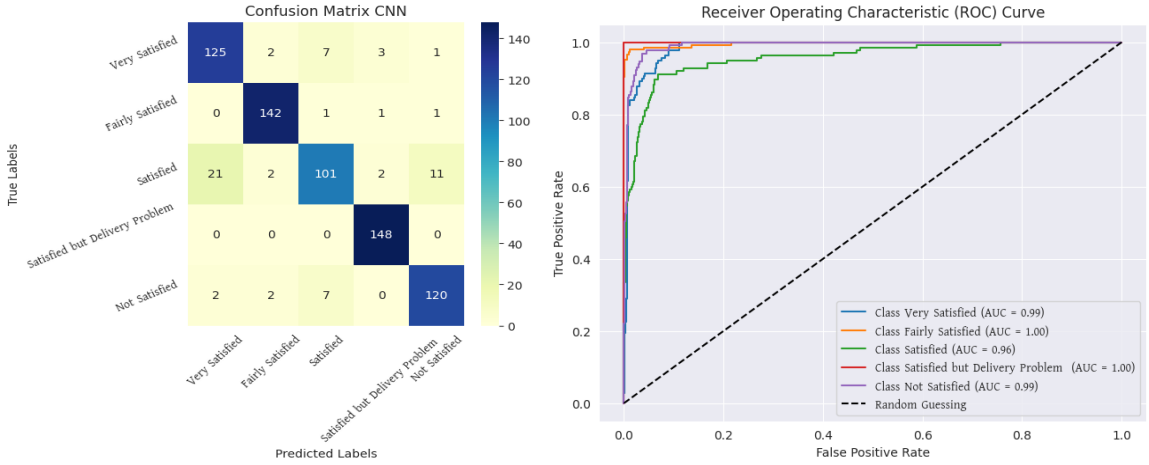

Figure 5. Proposed model confusion matrix and ROC curve

4. ABLATION STUDY, COMPARATIVE ANALYSIS, AND MODEL DEPLOYMENT ON WEB APPLICATION

Ablation studies are a technique used to assess the impact of various components or parameters of a machine and DL system on its performance. This study employs ablation studies to explore the impact of various architectures, layers, and hyperparameters on performance. This study performed an ablation study and tried different parameters to get the best-fitting model. After completing the ablation study chose the best performed parameter for the proposed model. Tables 2 to 5 show all the performances of different parameters.

Table 2. Experiment analysis by changing batch size

Test no.	Batch size	Epoch×training time	Accuracy	Finding
1	32	10*5	91.27	Best performance
2	64	10*2	90.41	Drop accuracy slightly
3	128	10*2	90.51	Drop accuracy slightly

Table 3. Experiment analysis by changing pooling layer

Test no.	Pooling layer	Epoch×training time	Accuracy	Finding
1	GlobalMaxPooling1D	10*3	91.27	Best performance
2	MaxPooling1D	10*3	88.98	Drop accuracy slightly
3	GlobalAveragePooling1D	10*2	87.41	Drop accuracy slightly

Table 4. Experiment analysis by changing optimizer

Test no.	Optimizer	Epoch×training time	Accuracy	Finding
1	Adam	10*3	91.27	Best performance
2	Nadam	10*3	89.27	Drop accuracy slightly
3	RMSprop	10*2	90.99	Drop accuracy slightly

Table 5. Test analysis by changing learning rate

Test no.	Learning rate	Epoch×training time	Accuracy	Finding
1	0.001	10*3	91.27	Best performance
2	0.0001	10*2	85.84	Drop accuracy slightly
3	0.01	10*2	86.41	Drop accuracy slightly

In the comparative study, some of the authors of my related work are conducted. Among them, in Table 6, we can see that one study applied both ML and DL and others only used ML or DL. Furthermore, our proposed model gets height accuracy with good precision, recall, and F1-measure from them which is 91.27%.

Table 6. Comparative analysis

Authors name	Year	Proposed method	Accuracy(%)	ML and DL	Remarks
Ullah *et al.* [8]	2023	QLeBERT	0.91	No	Good in performance
Moon *et al.* [9]	2021	Apiorir algorithm	0.88	No	Good in performance
Bitto *et al.* [4]	2023	LSTM	0.91	Yes	Good in performance
Bharathi *et al.* [3]	2022	SVM	0.90	No	Good in performance
Proposed model		Fine-tuned CNN	0.9127	Yes	Best performance

In this work for real-life deployment, we have built Python apps using model performance files which are saved in the time of model training. Here HTML is used for front end design and Python for back end design. Figure 6 shows a sample prediction by our developed applications.

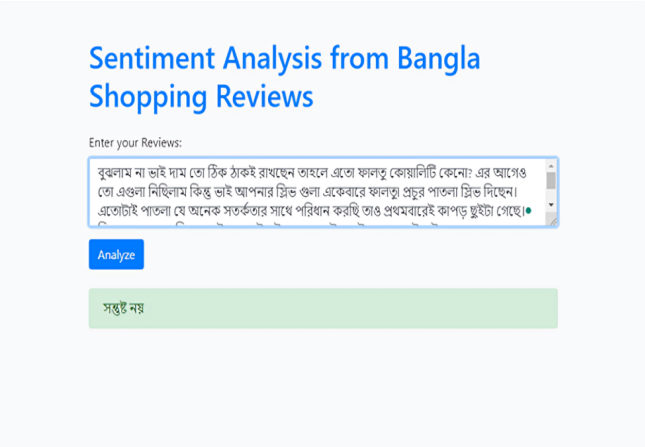

Figure 6. Model deployment with sample

5. CONCLUSION

The explosive growth of e-commerce has led to an increase of customer reviews, which remain largely untapped due to the lack of effective classifying systems, especially for languages like Bengali. Our study targets this gap by creating a solid method for recognising Bengali-language reviews from different online shopping sites. By leveraging both ML and DL models, we have proven that review categorization can be achieved with high accuracy, enabling better usage of customer feedback for e-commerce platforms. In this study, we categorized reviews into five different mood groups, offering a framework that can help businesses and buyers alike to better understand product feedback. Among the models tried, the CNN model beat others, getting a classification accuracy of 91.27%, showing its usefulness in handling text-based mood classification tasks. Our results have important implications for the field of NLP and e-commerce, especially in underrepresented languages like Bengali. Future work could explore increasing the dataset and improving the models to handle more complex sentiment groups or international datasets. Additionally, real-time uses of these classification systems could be combined into e-commerce platforms, giving instant feedback analysis to businesses and helping customers make informed purchase decisions. By continuing to build and improve these methods, we hope to add usefully to the wider field of data science and NLP, especially in Bengali-language study.

FUNDING INFORMATION

The authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Md. Arifur Rahman Rejuan	✓	✓	✓	✓	✓		✓	✓	✓	✓			✓	
Md. Assaduzzaman	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Nafiz Fahad		✓					✓			✓			✓	✓
Md. Jakir Hossen						✓				✓		✓	✓	✓
Md. Rahmatul Kabir Rasel Sarker			✓		✓			✓		✓				✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] DATAREPORTAL, "Digital 2022: Bangladesh," [Online]. Available: https://datareportal.com/reports/digital-2022-bangladesh. (Date accessed: Dec. 25, 2023).

[2] M. Archana and V. Thambusamy, "Impact of customer reviews on purchase based data using sentiment analysis with machine learning algorithm," *Journal of Fundamental and Comparative Research*, vol. 8, no. 2, pp. 57–67, 2022.

[3] R. Bharathi, R. Bhavani, and R. Priya, "Twitter text sentiment analysis of Amazon unlocked mobile reviews using Supervised Learning Techniques," *Indian Journal of Computer Science and Engineering*, vol. 13, no. 4, pp. 1242–1253, 2022, doi: 10.21817/indjcse/2022/v13i4/221304100.

[4] A. K. Bitto, M. H. I. Bijoy, M. S. Arman, I. Mahmud, A. Das, and J. Majumder, "Sentiment analysis from Bangladeshi food delivery startup based on user reviews using machine learning and deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 2282–2291, Aug. 2023, doi: 10.11591/eei.v12i4.4135.

[5] S. S. Shanto, Z. Ahmed, and A. I. Jony, "Mining User Opinions: A Balanced Bangla Sentiment Analysis Dataset for E-Commerce," *Malaysian Journal of Science and Advanced Technology*, pp. 272–279, 2023, doi: 10.56532/mjsat.v3i4.200.

[6] T. Akter, M. Begum, and R. Mustafa, "Bengali sentiment analysis of e-commerce product reviews using K-nearest neighbors," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2021, pp. 40–44, doi: 10.1109/icict4sd50815.2021.9396910.

[7] Md. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar, and S. A. Hossain, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning," in *11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–7, doi: 10.1109/ICCCNT49239.2020.9225511.

[8] A. Ullah *et al*., "Understanding quality of products from customers' attitude using advanced machine learning methods," *Computers*, vol. 12, no. 3, pp. 49–63, 2023, doi: 10.3390/computers12030049.

[9] N. N. Moon, I. M. Talha, and I. Salehin, "An advanced intelligence system in Customer Online Shopping Behavior and Satisfaction Analysis," *Current Research in Behavioral Sciences*, vol. 2, pp. 100051–100057, 2021, doi: 10.1016/j.crbeha.2021.100051.

[10] L. Barua *et al*., "Modeling household online shopping demand in the U.S.: A machine learning approach and comparative investigation between 2009 and 2017," *Transportation*, vol. 50, no. 2, pp. 437–476, 2021, doi: 10.1007/s11116-021-10250-z.

[11] J. Kaur and R. S. Sohal, "Noise Estimation and Removal in Natural Language Processing," *Handbook of Vibroacoustics, Noise and Harshness*, pp. 1–25, 2023, doi: 10.1007/978-981-99-4638-9_38-1.

[12] S. Sunny, S. Pinky, S. Jalal, M. Kayser, M. Wadud, and N. Mansoor, "Bangla E-Commerce Sentiment Analysis Optimization Using Tokenization and TF-IDF," in *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*, Dhaka, Bangladesh, 2024, pp. 1-6, doi: 10.1109/iCACCESS61735.2024.10499476.
[13] M. A. Hasan, J. Tajrin, S. A. Chowdhury, and F. Alam, "Sentiment Classification in Bangla Textual Content: A Comparative Study," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2020, pp. 1–6, doi: 10.1109/ICCIT51783.2020.9392681.
[14] X. Chen, Y. Wang, and Q. Liu, "Visual and textual sentiment analysis using deep fusion convolutional neural networks," in *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 2017, pp. 1557–1561, doi: 10.1109/ICIP.2017.8296543.
[15] S. K. Prabhakar, H. Rajaguru, K. So, and Dong-Ok Won, "A Framework for Text Classification Using Evolutionary Contiguous Convolutional Neural Network and Swarm Based Deep Neural Network," *Frontiers Computing Neurosciences*, vol. 16, 2022, doi: 10.3389/fncom.2022.900885.
[16] D. Li and J. Qian, "Text sentiment analysis based on long short-term memory," in *2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI)*, Wuhan, China, 2016, pp. 471–475, doi: 10.1109/CCI.2016.7778967.
[17] G. V. Houdt, C. Mosquera, and G Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, pp. 5929–5955, 2020, doi: 10.1007/s10462-020-09838-1.
[18] K. -P. Lin and M. -S. Chen, "On the Design and Analysis of the Privacy-Preserving SVM Classifier," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1704–1717, Nov. 2011, doi: 10.1109/TKDE.2010.193.
[19] Y. Yang, J. Li, and Y. Yang, "The research of the fast SVM classifier method," in *2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu, China, 2015, pp. 121–124, doi: 10.1109/ICCWAMTIP.2015.7493959.
[20] Simplilearn, "What is XGBoost? An Introduction to XGBoost Algorithm in Machine Learning," [Online]. Available: https://www.simplilearn.com/what-is-xgboost-algorithm-in-machine-learning-article#: :text=XGBoost. (Date accessed: Dec. 25, 2023).
[21] C. Bentéjac, A. Csörgő, and G. A. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, pp. 1937–1967, 2021, doi: 10.1007/s10462-020-09896-5.
[22] S. J. Rigatti, "Random Forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017, doi: /10.17849/insm-47-01-31-39.1.
[23] W. M. Stallings and G. M. Gillmore, "A NOTE ON "ACCURACY" AND "PRECISION," *Journal of Educational Measurement*, vol. 8, no. 2, pp. 127–129, Jun. 1971, doi: 10.1111/j.1745-3984.1971.tb00916.x.
[24] B. Carterette, "Precision and Recall," *Encyclopedia of Database Systems*, pp 2126–2127, 2009, doi: 10.1007/978-0-387-39940-9_5050
[25] D. J. Hand, P. Christen, and N. Kirielle, "F*: an interpretable transformation of the F-measure," *Machine Learning*, vol. 110, no. 3, pp. 451–456, 2021, doi: 10.1007/s10994-021-05964-1.

BIOGRAPHIES OF AUTHORS

Md. Arifur Rahman Rejuan is a graduate teaching assistant at Daffodil International University, Dhaka, Bangladesh, where he completed his B.Sc. in Computer Science and Engineering. His expertise lies in machine learning, deep learning, data mining, and natural language processing, with a strong passion for advancing the field of artificial intelligence and its real-world applications. For any collaboration possibilities or talks involving his areas of expertise. He can be contacted at email: arifur15-13810@diu.edu.bd.

Md Assaduzzaman is a Senior Lecturer in the Department of Computer Science and Engineering, at Daffodil International University, Dhaka, Bangladesh. He has an M.Sc. in Computer Science and Engineering and a Bachelor's degree in Computer Science and Engineering. He has extensive expertise in both teaching and research. His current research interests are machine learning and deep learning. He has published many research papers in several international peer-reviewed journals and conferences. He can be contacted at email: assaduzzaman.cse@diu.edu.bd.

Nafiz Fahad received B.Sc. degree in Computer Science from American International University-Bangladesh (AIUB) in August 2023. Currently, he works as a Graduate Research assistant at the Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia. His main research interests are in the fields of machine learning, data mining, computer science, subfields of AI, AI explainability, large language models, and data science. He can be contacted at email: fahadnafiz1@gmail.com.

Md. Jakir Hossen graduated in Mechanical Engineering from the Dhaka University of Engineering and Technology (Former Name is BIT Dhaka,1997), a Higher Diploma in Computer Science and Engineering (1999), a Masters degree in Communication and Network Engineering from Universiti Putra Malaysia (2003) and Ph.D. degree in Smart Technology and Robotic Engineering at Universiti Putra Malaysia (2012). He is currently an Associate Professor at the Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia. His research interests are in the areas of artificial intelligence, data analytics, and applications in anomaly detection, robotics, and control. He can be contacted at email: jakir.hossen@mmu.edu.my.

Md. Rahmatul Kabir Rasel Sarker is a Lecturer in the Department of Computer Science and Engineering at Daffodil International University (DIU), Bangladesh. He holds an M.Sc. in Computer Science from Jahangirnagar University and a B.Sc. in Computer Science and Engineering from DIU. His research interests include deep learning, natural language processing, and human-computer interaction, with publications on topics such as plant disease detection and emotion recognition from music lyrics. He has also completed professional training in mobile application development and .NET technologies. He can be contacted at email: raselsarker.cse@diu.edu.bd.