Mining User Opinions: A Balanced Bangla Sentiment Analysis Dataset for E-Commerce

Article *in* Malaysian Journal of Science and Advanced Technology · November 2023

DOI: 10.56532/mjsat.v3i4.200

Shakib Sadat Shanto
American International University-Bangladesh

Zishan Ahmed
American International University-Bangladesh

MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

Mining User Opinions: A Balanced Bangla Sentiment Analysis Dataset for E-Commerce

Shakib Sadat Shanto[1], Zishan Ahmed [1], and Akinul Islam Jony [*1]

[1] *Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh.*

KEYWORDS

Sentiment Analysis
Bangla language
E-Commerce
Machine Learning
Deep Learning

ARTICLE HISTORY

Received 9 October 2023
Received in revised form 24 November 2023
Accepted 27 November 2023
Available online 27 November 2023

ABSTRACT

The utilization of sentiment analysis has gained significant importance as a valuable method for obtaining meaningful insights from textual data. The research progress in languages such as English and Chinese has been notable. However, there is a noticeable dearth of attention towards creating tools for sentiment analysis in the Bangla language. Currently, datasets are limited for Bangla sentiment analysis, especially balanced datasets capturing both binary and multiclass sentiment for e-commerce applications. This paper introduces a new sentiment analysis dataset from the popular Bangladeshi e-commerce site “Daraz”. The dataset contains 1000 reviews across 5 product categories, with both binary (positive/negative) and multiclass (very positive, positive, negative, very negative) sentiment labels manually annotated by native Bangla speakers. Reviews were collected using an organized process, and labels were assigned based on standardized criteria to ensure accuracy. In addition, a benchmark evaluation of the performance achieved by Machine Learning and Deep Learning algorithms on this dataset is also provided. The new dataset can aid research on multiclass and binary Bangla sentiment analysis utilizing both machine learning, deep learning, and Large Language Models. It can aid e-commerce platforms in analysing nuanced user opinions and emotions from online reviews. The utilization of categorized product reviews also facilitates research in the field of text categorization.

1. INTRODUCTION

Over the past few years, there has been a significant transformation in the field of natural language processing (NLP), combining elements from computer science, artificial intelligence, and linguistics. This transition has resulted in the extensive utilization of statistical methodologies. Sentiment analysis has gained significant attention in Natural Language Processing (NLP) [1]. Sentiment analysis is a process that entails the examination of textual data to ascertain the views and attitudes that are conveyed through it [2]. Various sources can be used for sentiment analysis, encompassing social media comments, news articles, blogs, reviews, and other opinionated texts. The analysis extends beyond identifying positive or negative sentiments and also assigns continuous polarity scores. By using these scores and labels, researchers can compile datasets to train machine learning algorithms for sentiment analysis. The objective of sentiment analysis is to facilitate the comprehension and extraction of subjective information from human language by computers, hence improving the quality of human-computer interaction. While earlier natural language processing (NLP) systems relied on intricate linguistic rules, modern sentiment analysis utilizes statistical and neural network models trained on vast text collections. The utilization of data-driven methodologies has significantly enhanced the precision and adaptability of sentiment analysis in several fields. Nevertheless, the task of effectively analysing informal text genres, such as social media, remains a challenging endeavour.

Sentiment analysis is a natural language processing technique that identifies and extracts thoughts, emotions, and attitudes from subjective data. The objective is determining whether a text expresses positive, negative, or neutral sentiments towards a topic. Sentiment analysis has become an increasingly vital tool for extracting insights from textual content such as social media, surveys, survey responses, etc. It enables enterprises, organizations, and governments to efficiently evaluate public opinion, trends, and preferences on a large scale.

*Corresponding author:
E-mail address: Akinul Islam Jony < akinul@aiub.edu >.
https://doi.org/10.56532/mjsat.v3i4.200
2785-8901/

Bangla, also known as Bengali, is widely spoken as a primary language by a substantial global population of around 200 million. Most of these individuals, approximately 160 million, are native speakers residing in Bangladesh. Over the past several years, there has been a noticeable rise in the participation of Bangladeshi internet users in various online activities, such as expressing their viewpoints on social media platforms, establishing connections with others, providing comments on news articles, and engaging in online shopping. As a result, there has been growing interest in analysing the sentiment and emotion in Bangla user-generated content. Several studies have investigated sentiment analysis in Bangla text [3, 4]. The mentioned studies are motivated by the significance of comprehending public opinion and emotion from various sources such as social media, reviews, forums, and other forms of textual data. The ability to extract sentiment from these sources is essential for both business and social intelligence purposes.

As natural language processing research progresses rapidly with advances in machine learning, deep learning, and large language models (LLM) [5], the Bangladeshi research community continues to explore new directions and agendas for work on the Bangla language. Previous studies have proposed some Bangla datasets, mainly focused on sentiment analysis, as reviewed in the literature. However, there remains a need for balanced Bangla datasets that capture both binary and multiclass sentiment, especially for e-commerce applications. Understanding nuanced user opinions and emotions from online reviews is critical for e-commerce platforms. To help address this gap, this paper introduces a new dataset collected from the popular Bangladeshi e-commerce site Daraz. The dataset is made public for further research and development. The dataset contains both binary and multiclass sentiment labels across reviews from five different product categories. It can facilitate research on sentiment analysis to extract insights from Bangla e-commerce user feedback. Additionally, the categorized product reviews can enable work on Bangla text classification.

The research starts with a comprehensive literature assessment of the approaches to Bengali sentiment analysis and dataset construction that have previously been worked on. The methodology section then describes in detail the construction of the dataset, the criteria for labelling, and the statistical analysis. The benchmark evaluation section afterwards provides results for various sentiment classification models applied to the dataset. The conclusion concludes with a summary of the contributions and recommendations for future research.

2. LITERATURE REVIEW

In recent times, there has been a notable increase in the attention given to sentiment analysis in multiple languages. Sentiment analysis has been conducted using several languages such as English, Chinese, Urdu, Bangla, and several more. Currently, a substantial collection of datasets can be utilized for sentiment research across many languages. The opinion and sentiment analysis in a particular language can be conducted by examining the lexical and syntactical components. There are a lot of datasets available to analyse a sentence, such as Twitter [6, 7], Restaurant [8], News Comment [9], Online shopping [10] and many more.

Although there has been significant advancement in sentiment analysis for languages such as English and Chinese, there has been a relative lack of focus on developing resources for sentiment analysis in Bangla. In contrast to the English language, the availability of extensive and reliable labelled datasets for sentiment analysis in the Bangla language is considerably limited. The absence of available datasets for sentiment analysis in the Bangla language poses a significant obstacle to advancing natural language processing capabilities in this extensively utilized language. Currently, there is a limited availability of datasets for Bangla Sentiment Analysis. Among them, Ali et al. [11] created a dataset named "BanglaSenti", which has 61582 Bangla words with positive, negative, and neutral words. Rahman and Dey [12] offer a dataset that includes two sets of data, one on comments on cricket and the other consisting of restaurant reviews. Chowdhury [13] employed a dataset obtained by querying the Twitter API v1.1 to examine Bangla microblog posts. The posts were afterwards translated into the Bangla language.

Machine learning methodologies are extensively employed across diverse datasets for developing sentiment analysis systems in the Bangla language. Using restaurant reviews, Haque et al. [14] utilized a dataset of 1,500 instances to tackle multiclass classification, achieving an accuracy of 75.58% using Support Vector Machines (SVM). Similarly, in a binary classification setting with a dataset of 1000 instances, Sharif et al. [15] employed Multinomial Naive Bayes (MNB) to attain an accuracy of 80.48%. Regarding e-commerce, a study [16] addressed the issue of multiclass classification using a large dataset consisting of 7905 instances. The study showed significant achievements by employing a K-Nearest Neighbours (KNN) model, which demonstrated an impressive accuracy rate of 96%. In the context of binary classification, the authors [17] used Support Vector Machines (SVM) as their chosen algorithm. They reported an accuracy rate of 88.81% using a dataset consisting of 1020 instances. In the specific field of horoscopes, the study [18] yielded remarkable outcomes by implementing a binary Support Vector Machine (SVM) model. The model achieved an impressive accuracy rate of 98.70% by utilizing a dataset of 6000 instances. In the domain of books, the authors [19] utilized the Random Forests (RF) algorithm to attain a binary classification accuracy of 98.39%. This was accomplished by employing a dataset consisting of 5500 instances. Shifting the focus to social media, the study [20] explored the multiclass classification task using a dataset of 12,628 instances. The researchers employed Logistic Regression (LR) and obtained an accuracy rate of 44%. In film studies, the authors [21] utilized Support Vector Machines (SVM) for binary classification purposes. Their study yielded an accuracy rate of 85.59% using a dataset consisting of 1141 instances.

The usage of deep learning approaches is also getting very popular while classifying sentiments in the Bangla language because they perform better compared to traditional machine learning approaches. The categorizing of comments in a binary manner is explored in the study [22]. By utilizing a distinctive combination of Bangla-BERT and LSTM, the research attained a noteworthy accuracy rate of 94.15%. In the realm of binary comment classification, the authors [23] employed Convolutional Neural Networks (CNN) to achieve an impressive accuracy rate of 99.87%. In the context of restaurant reviews, the study [24] employed Bidirectional Long Short-Term Memory (BiLSTM) networks, which yielded a noteworthy accuracy rate of 91.35%. In the field of news commentary, the authors [25] employed LSTM networks, achieving a notable accuracy rate of 94%. Incorporating multiclass classification with news comments, the BERT+GRU technique was employed in the study [26]. They obtained an accuracy of 71% in the binary classification setting and 60% in the multiclass scenario. In the study [27], Long Short-Term Memory (LSTM) networks were utilized to classify online food reviews. The results of this investigation demonstrated a significant accuracy rate of 90.86%. In the context of multiclass settings, the problem of abusive remarks was examined by researchers [28]. They employed BERT, a language model, to tackle this issue and achieved a commendable accuracy of 88%, demonstrating its effectiveness. Finally, a study conducted by the authors [29] investigated the application of Skip-gram for multiclass comment classification. The results of this study indicated that an accuracy rate of 75% was achieved.

Based on the comprehensive literature review, it becomes apparent that there is a pressing requirement for a publicly accessible and adequately diverse Bangla dataset for sentiment analysis, particularly within the expanding e-commerce sector, which is currently experiencing significant growth. Utilizing a carefully constructed and well-balanced dataset can facilitate the advancement of sentiment analysis systems by incorporating both machine learning and deep learning models. Furthermore, a dearth of datasets addresses the diverse range of multiclass emotions in the Bangla language. This study aims to fill the existing gaps by offering a comprehensive e-commerce dataset that effectively incorporates binary and multiclass sentiments in the Bangla language. The dataset also includes five distinct categories of product types corresponding to each review, enabling the classification of product types.

3. METHODOLOGY

The methodology to construct our dataset is illustrated in Fig. 1. Source selection commenced the methodology for creating the dataset, followed by data collection and product type labelling. Following this, sentiment labelling criteria were established, implementing both binary and multiclass sentiment labelling. In conclusion, statistical analysis was done to assess sentiment distribution and product type representation.

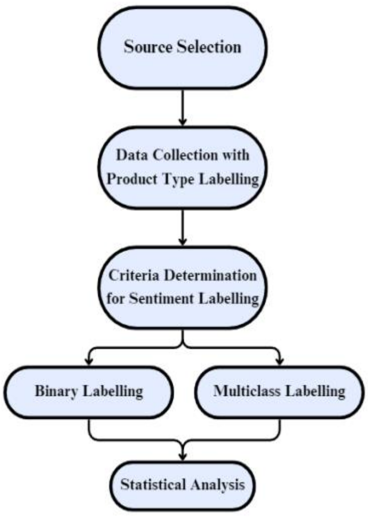

Fig. 1. Methodology for Creating Dataset

3.1 Dataset Source Selection

The e-commerce industry in Bangladesh has grown tremendously in recent years, resulting in the introduction of several popular online platforms catering to domestic consumers. Among the available e-commerce platforms, "Daraz" appears as the first form and is most widely used in the interior country.

There are hundreds of dedicated e-commerce websites in Bangladesh. If one is looking for a place to find genuine Bengali comments regarding a wide range of themes and products, Daraz is the place to go. Providers and service seekers in Bangladesh may trust the platform because of its sizable user base, extensive selection, emphasis on local markets, and built-in review system.

Daraz is well-known for offering an extensive selection of items across several categories. It's easy to find genuine comments made in Bangla because of the site's emphasis on local participation, encouraging users to use their native language. The review system implemented by Daraz allows customers to offer their comments on their purchases while also allowing users to designate ratings to products and furnish comprehensive comments. To uphold the trustworthiness of consumer feedback, Daraz employs a verification process wherein customer purchases are authenticated prior to enabling them to post comments. This practice aids in guaranteeing the authenticity and reliability of reviews.

3.2 Data Collection

Using an organized manual process, Bangla comments were gathered from various product types on “Daraz”, emphasizing those expressing certain sentiments. The process began by going to the “Daraz” website and searching the various product pages for comments written exclusively in Bangla.

Organized criteria were implemented to the comment collection method to provide a dataset representative of authenticity and quality. The reviews that avoided using spammy or repetitious wording were given more weight. Opinions from a variety of sentiments were collected. Balance was ensured so no one sentiment type got extensive priority. Moreover, extra weight was given to comments that expressed themselves in a clear and thorough way. A commitment to inclusion led to seeking comments on a wide variety of products, including those in the areas of clothing, electronics, household stuff, fashion and beauty products.

Once a suitable comment was identified, the Bangla comment was copied and pasted into an Excel spreadsheet in the designated column titled 'Comment'. Concurrently, the respective product type was included with each comment in the 'Product_Type' column.

Table 1 shows a few samples of our collected comments with their corresponding product types. For the benefit of readers unfamiliar with Bengali, we've included an English translation of the comments in the Table.

Table 1. Samples of Comments with their Product type

Comment	English Translation	Product Type
আইলাইনার টা ভালো না। রং বেশি গাঢ় না। আর এটা ওয়াটার প্রুফ না।	The eyeliner is not good. The color is not too dark. And it is not waterproof.	Beauty Products
হেডফোনটা ১ মাস ও যায় নি, ভালো না	The headphone did not last for even 1 month, not good	Electronics
এই জুতো জোড়া অবশ্যই ভালো মানের। তাই আমার পক্ষ থেকে 5★	This pair of shoes is definitely of good quality. So, 5★ from me	Fashion
শাড়ি কালার কোন পরিবর্তন হয়নি একদম সেম ই পেয়েছি যেমনটা চেয়েছিলাম	There was no change in the color of the saree, I took it exactly as I wanted	Clothing

3.3 Data Labelling

A critical component of the data collection procedure for our research undertaking was the careful labelling of comments according to their distinct sentiments. We, the authors, all of whom are native Bengali speakers with profound knowledge of the language's complexities and cultural implications, independently dedicated themselves to this undertaking. Our expertise provided us with a distinct benefit in determining the contextual meaning concealed within the comments, thereby enabling us to designate sentiments precisely and accurately. By leveraging the combined knowledge as native speakers, it was ensured that the sentiments assigned to the comments were not only linguistically accurate but also closely corresponded with the details and characteristics of the Bengali language.

The dataset was initially labelled according to the multiclass labelling format. The annotators implemented the criteria specified in Table 2 as an important guide throughout the process of labelling comments. These criteria played a crucial role in guaranteeing the precise categorization of every comment based on its unique sentiment. Our objective in implementing this standard set of criteria was to enhance the level of accuracy in our manual labelling procedure.

Using this method, the meaning of each comment could be dissected, the underlying emotions identified, and appropriate categories assigned. Precise oversight was kept, and the reviewers' intended feelings and ideas were captured by manually labelling the dataset. The labels of each comment were added in the 'Comment_Type' column. Table 3 shows a few samples of multiclass labelled comments. English translations were included alongside the comment samples in the Table to facilitate comprehension for non-Bengali speakers.

The "Positive" and "Very Positive" labels were merged into one, simply called "Positive," in an effort to achieve a binary classification. At the same time, the "Negative" and "Very Negative" labels were combined into a single one, simply called "Negative." Since the dataset was created with the primary goal of differentiating between positive and negative feelings, this intentional consolidation simplified the classification work and closely aligned with this goal.

Table 2. Comment Labelling Criteria

Comment Type	Criteria	Description
Positive	Comments in this category expressed moderate levels of contentment, praise, and gratitude.	Comments that are classified as "Positive" often convey a little enthusiastic expression of satisfaction, approbation, or gratitude. In general, they may show a good attitude without being too enthusiastic.
Very Positive	The majority of the comments in this section were really enthusiastic, positive, and complimentary .	Comments labelled as "Very Positive" displayed an exceptionally high degree of optimism, usually associated with too generous compliments, strong encouragement, and enthusiastic language. Comments like this usually indicate a great deal of appreciation or support.
Negative	Comments in this category expressed moderate levels of displeasure, rage, and disappointment	Comments labelled as "Negative" revealed a wide range of degrees of disapproval, irritation, and other negative emotions. None of the remarks were very nasty, but they did convey a sense of discontent or disapproval.
Very Negative	Comments in this section were heavily critical, slangy, aggressive, or condemning.	Comments with an extremely negative tone, such as expletives, slang, or statements of condemnation, were filed under the "Very Negative" label. These comments often conveyed extreme discontent, rage, or harsh disapproval.

Table 3. Labelling samples for Multiclass classification

Comment	English Translation	Comment Type
স্মার্ট ওয়াচ দাম হিসেবে খারাপ না।	The smart watch is not bad for the price	Positive
খুবই ভালো ফ্যানটা। অনেক ধন্যবাদ সেলার ভাইকে এত ভালো একটা প্রোডাক্ট দেওয়ার জন্য।	Very good fan. Thank you very much seller brother for providing such a good product.	Very Positive
আশা করছিলাম আরো ভালো কিছুর। স্মেল বেশীক্ষন থাকেনা। কোনোরকম কভার ছিলনা	I was hoping for something better. The smell does not last long. There was no cover	Negative
এই দামে স্ক্রাব খুবই ছোট, খুবই দুঃখজনক! তাছাড়া পরিমান টাও পর্যন্ত লেখা নেই পন্যের গায়ে। হতাশ!	The scrub is so small at this price, so sad! Moreover, the quantity is not even written on the product. disappointed!	Very Negative

A separate Excel sheet was created for binary labelling where the 'Comment_Type' column has all the comments in binary labels. Here, as mentioned earlier, all the comments were labelled either as 'Positive' or 'Negative'. Table 4 shows a few samples of binary labelled comments. English translations of the comments are added here as well.

Table 4. Labelling samples for Binary classification

Comment	English Translation	Comment Type
ছোট ছোট কেরী পোকা হাঁটছিল এবং দুর্গন্ধের জন্য মুখে দেওয়া যাচ্ছিল না।	Small caterpillars were walking and could not be given to the mouth because of the stench.	Negative
যা মনে করেছিলাম তার চাইতে অনেক গুনে ভালো খেতে অনেক স্বাদ	Much better than what I thought, so many flavors to eat	Positive
কাপড় খুবই পাতলা আর নরমাল। ডেলিভারি চার্জ সহ দাম অনেক বেশিইই জিনিস হিসাবে।	The fabric is very thin and normal. The price including delivery charges is very high as the item.	Negative
কম্বলটা এই দামে সেরা বলা ছাড়া উপায় নাই। যথেষ্ট বরো সাইজ আর অনেক গরম হয়। এই কম্বলের উপরে একটা কাথা নিলে অনেক বেশি শীত ও পারি দেয়া যায়।	The blanket is the best at this price. Good enough bore size and gets very hot If you take a blanket on top of this blanket, you can get a lot of warmth and comfort I couldn't imagine getting such a beautiful thing at such a low price. Thanks to the seller.	Positive

3.4 Statistical Analysis of the Dataset

After preparation of the dataset, some statistics were analysed to visualize the distribution of the dataset.

The analysis of the pie charts in Fig. 2 and Fig. 3 shows that the dataset has a nearly even distribution of 1000 comments between positive and negative sentiments in both the multiclass and binary classification settings. The previously mentioned balance might provide benefits in machine learning model training, as it minimizes the possible challenges resulting from class imbalance.

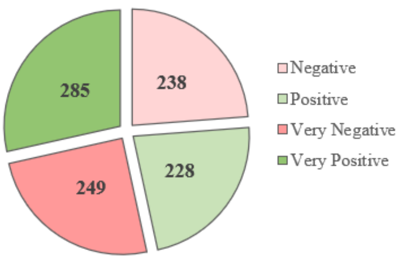

Fig. 2. Dataset Distribution on Multiclass Labelling

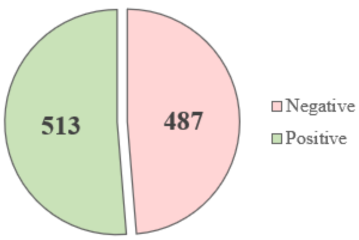

Fig. 3. Dataset Distribution on Binary Labelling

The pie chart in Fig. 4 shows how comments for several categories of products, such as Clothing, Electronics, Household Stuff, Fashion and Beauty Products, are distributed. It is observed that the number of comments is fairly distributed among these different product categories. This balance shows that the dataset represents a broad range of product categories, making it useful for exploring consumer sentiment in various fields.

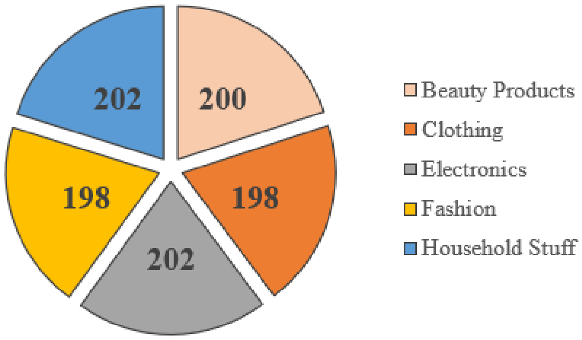

Fig. 4. Dataset Distribution according to Product type

4. BENCHMARK EVALUATIONS ON THE DATASET

The dataset was applied to both machine learning [30] and deep learning models [31] to evaluate the classification accuracy using several evaluation metrics.

A variety of classical machine learning models were selected for benchmark evaluation on the dataset, including both linear models like Logistic Regression, Naive Bayes, SGD, as well as non-linear models like SVM, Random Forest and KNN. The main goal was to cover a diverse set of algorithms spanning different underlying principles to provide a robust analysis. Simpler linear models can act as baselines while advanced non-linear techniques may achieve greater performance. Regarding deep learning models, Bangla-BERT, a Transformers-based model, and transfer learning-based models like LSTM and GRU were chosen. Based on the previous research and their demonstrated effectiveness for Bangla sentiment analysis, the models for both machine learning and deep learning were also chosen for benchmarking.

The machine learning models were implemented in Python using the Scikit-Learn library. For feature extraction from the text, both count vectorization and TF-IDF vectorization methods were applied. The models were evaluated using standard metrics like accuracy, precision, recall and F1-score.

The deep learning models like LSTM, GRU and Bangla-BERT were chosen due to their recent success in sentiment analysis tasks. They can capture semantic relationships in text through their neural architectures. Transfer learning with pretrained models like BERT allows leveraging knowledge from large external corpora. The deep learning models were implemented using PyTorch and HuggingFace libraries. They were trained for 15 epochs with early stopping based on validation loss. Key hyperparameters tuned include hidden layer dimensions, learning rate, dropout rate. Evaluation was done similar to the ML models above.

From Table 5 and Table 6, the accuracies of different ML models can be seen for both multiclass and binary classification approaches. Logistic Regression (LR) achieved the highest accuracy of 82.64%, along with precision, recall and F1-scores with the Bigram-CountVec feature extraction method for multiclass classification. Moreover, for binary classification, Random Forest (RF) achieved the highest accuracy of 94.44%, along with good precision, recall and F1-Scores when the Unigram TF-IDF feature extraction method was used.

Table 5. Results of machine learning models in Multiclass classified dataset

Machine Learning Model	Accuracy	Precision	Recall	F1-Score
LR	**82.64%**	**81.65**	**82.70**	**81.66**
Decision Tree	68.06%	67.62	65.55	65.73
Random Forest	75.69%	75.48	76.39	74.27
MNB	71.53%	71.40	70.63	68.77
KNN	60.42%	61.46	60.49	58.47
Linear SVM	77.08%	77.04	78.15	75.34
RBF SVM	76.39%	75.71	77.10	74.65
SGD	79.17%	76.92	78.67	77.26

Table 6. Results of machine learning models in Binary class classified dataset

Machine Learning Model	Accuracy	Precision	Recall	F1-Score
LR	88.89%	89.46	89.16	88.88
Decision Tree	86.11%	86.26	86.26	86.11
Random Forest	**94.44%**	**94.42**	**94.49**	**94.44**
MNB	92.36%	92.48	92.26	92.33
KNN	84.03%	84.22	83.86	83.93
Linear SVM	88.89%	89.46	89.16	88.88
RBF SVM	90.97%	91.22	91.16	90.97
SGD	93.06%	93.03	93.10	93.05

Table 7 and Table 8 show the multiclass and binary classification accuracy of several Deep learning models. In addition to high precision, recall, and F1-scores, Bangla-BERT's accuracy for multiclass classification was 88.78%. In addition to having the highest accuracy of 94.5%, with precision, recall, and F1-scores, Bangla-BERT also excelled in binary classification.

Table 7. Results of Deep learning models in Multiclass classified dataset

Deep Learning Model	Accuracy	Precision	Recall	F1-Score
GRU	84.7	87.6	84.8	86.1
LSTM	85.0	82.0	83.7	85.5
Bangla-BERT	**88.78**	**88.78**	**88.68**	**89.77**

Table 8. Results of Deep learning models in Binary classified dataset

Deep Learning Model	Accuracy	Precision	Recall	F1-Score
GRU	90.97	91.06	91.10	90.97
LSTM	91.67	92.01	91.88	91.67
Bangla-BERT	**94.5**	**94.42**	**94.49**	**94.44**

The benchmark evaluations reveal several insightful patterns regarding the sentiment classification performance. Overall, the deep learning models achieve higher accuracy than traditional machine learning approaches.

Among the machine learning models, Random Forest reaches the top accuracy of 94.44% for binary classification with the TF-IDF text vectorization. The Logistic Regression model attains an accuracy of 82.64% for the more challenging multiclass case with count vectorization.

On the other hand, the deep learning methods demonstrate superior accuracy due to their representation learning capabilities. The LSTM and GRU recurrent models which specialize in sequential data, provide results around 85-90% accuracy owing to memorizing long-term contextual dependencies. Finally, using the pretrained Bangla-BERT language model leads to state-of-the-art accuracy of 88.78% and 94.5% for multi-class and binary classification respectively. BERT's bidirectional transformer encoding captures both semantics and context most effectively.

The deep learning algorithms outshine machine learning in accuracy and other evaluation metrics. The performance gap is greater in the multiclass setting as increased detail introduces more complexity. The results validate the strength of deep learning for sentiment analysis especially with low-resource languages like Bangla where annotated datasets are scarce.

5. CONCLUSION

Specifically focusing on the e-commerce sector, this research set out to compile a brand-new Bangla-language sentiment analysis dataset. Insufficient publicly accessible datasets with both binary and multiclass sentiment labelling prompted this study. A dataset of 1,000 Bangla comments was assembled from the popular e-commerce portal Daraz, covering 5 distinct product categories through manual gathering and annotation by local speakers. Extensive measures were taken to guarantee fairness and precision.

There are many useful insights to be gained from this new dataset. To begin with, it paves the way for studies of sentiment analysis in Bangla, which will develop natural language processing for this widely spoken language. Machine and deep learning models may be effectively trained due to the almost even distribution of positive and negative sentiment across classes. Second, going beyond simple positive/negative polarity, the use of binary and multiclass labelling yields richer emotional insights. This can help e-commerce sites grasp sophisticated customer feedback. Thirdly, the dataset reflects a realistic e-commerce environment since it covers a wide range of products. Finally, a study into text categorization may be made possible thanks to the sorted product reviews.

Multiple machine-learning models, including logistic regression, random forest, and naive Bayes, achieved accuracy rates of around 80% in sentiment classification in the benchmark assessments. Also, deep learning methods, such as LSTM and BERT, achieved good accuracies. This points to the dataset's ability to be used to train high-quality models. Improving performance with cutting-edge deep learning systems might be the subject of future study. Emotion categorization systems might make use of multiclass labels.

A Bangla dataset designed for e-commerce sentiment analysis has been given in this study. In addition to helping online companies, it makes a significant contribution to the development of Bangla natural language processing. The dataset might be improved in the future by including more reviews, more product kinds, and neutral polarity. Aspect-based sentiment analysis may be possible if the labels are enriched with information such as review titles and star ratings. Associating visuals with critiques opens the door to multimodal analysis. This dataset is a starting point for developing better methods of sentiment analysis for the Bangla language.

Data Availability Statement: The proposed dataset can be found on https://github.com/shakib-sadat/Bangla-E-commerce-Dataset.

REFERENCES

[1] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," Multimedia tools and applications, vol. 82, no. 3, pp. 3713-3744, 2023.

[2] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams engineering journal, vol. 5, no. 4, pp. 1093-1113, 2014.

[3] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter, and A. K. Das, "An automated system of sentiment analysis from Bangla text using supervised learning techniques," in 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019: IEEE, pp. 360-364.

[4] T. Al Mahmud, S. Sultana, T. I. Chowdhury, and F. R. Anando, "A New Approach to Analysis of Public Sentiment on Padma Bridge in Bangla Text," in 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), 2022: IEEE, pp. 1-6.

[5] M. A. Hasan, S. Das, A. Anjum, F. Alam, A. Anjum, A. Sarker, and S. R. H. Noori, "Zero-and Few-Shot Prompting with LLMs: A Comparative Study with Fine-tuned Models for Bangla Sentiment Analysis," arXiv preprint arXiv:2308.10783, 2023.

[6] B. Alharbi, H. Alamro, M. Alshehri, Z. Khayyat, M. Kalkatawi, I. I. Jaber, and X. Zhang, "ASAD: A twitter-based benchmark arabic sentiment analysis dataset," arXiv preprint arXiv:2011.00578, 2020.

[7] K. Sailunaz and R. Alhajj, "Emotion and sentiment analysis from Twitter text," Journal of Computational Science, vol. 36, p. 101003, 2019.

[8] K. Zahoor, N. Z. Bawany, and S. Hamid, "Sentiment analysis and classification of restaurant reviews using machine learning," in 2020 21st International Arab Conference on Information Technology (ACIT), 2020: IEEE, pp. 1-6.

[9] H. Rahab, A. Zitouni, and M. Djoudi, "SANA: Sentiment analysis on newspapers comments in Algeria," Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 7, pp. 899-907, 2021.

[10] A. Alrehili and K. Albalawi, "Sentiment analysis of customer reviews using ensemble method," in 2019 International conference on computer and information sciences (ICCIS), 2019: IEEE, pp. 1-6.

[11] H. Ali, M. F. Hossain, S. B. Shuvo, and A. Al Marouf, "Banglasenti: A dataset of bangla words for sentiment analysis," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020: IEEE, pp. 1-4.

[12] M. A. Rahman and E. Kumar Dey, "Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation," Data, vol. 3, no. 2, p. 15, 2018.

[13] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in Bangla microblog posts," in 2014 International Conference on Informatics, Electronics & Vision (ICIEV), 2014: IEEE, pp. 1-6.

[14] F. Haque, M. M. H. Manik, and M. Hashem, "Opinion mining from bangla and phonetic bangla reviews using vectorization methods," in 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 2019: IEEE, pp. 1-6.

[15] O. Sharif, M. M. Hoque, and E. Hossain, "Sentiment analysis of Bengali texts on online restaurant reviews using multinomial Naïve Bayes," in 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT), 2019: IEEE, pp. 1-6.

[16] M. 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: IEEE, pp. 40-44.

[17] M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur, and M. O. Faruk, "Product review sentiment analysis by using nlp and machine learning in bangla language," in 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020: IEEE, pp. 1-5.

[18] T. Ghosal, S. K. Das, and S. Bhattacharjee, "Sentiment analysis on (Bengali horoscope) corpus," in 2015 Annual IEEE India Conference (INDICON), 2015: IEEE, pp. 1-6.

[19] M. E. Khatun and T. Rabeya, "A Machine Learning Approach for Sentiment Analysis of Book Reviews in Bangla Language," in 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022: IEEE, pp. 1178-1182.

[20] T. Ahmed, S. F. Mukta, T. Al Mahmud, S. Al Hasan, and M. G. Hussain, "Bangla Text Emotion Classification using LR, MNB and MLP with TF-IDF & CountVectorizer," in 2022 26th International Computer Science and Engineering Conference (ICSEC), 2022: IEEE, pp. 275-280.

[21] M. Hassan, S. Shakil, N. N. Moon, M. M. Islam, R. A. Hossain, A. Mariam, and F. N. Nur, "Sentiment analysis on Bangla conversation using machine learning approach," International Journal of Electrical and Computer Engineering (IJECE), vol. 12, no. 5, pp. 5562-5572, 2022.

[22] N. J. Prottasha, A. A. Sami, M. Kowsher, S. A. Murad, A. K. Bairagi, M. Masud, and M. Baz, "Transfer learning for sentiment analysis using BERT based supervised fine-tuning," Sensors, vol. 22, no. 11, p. 4157, 2022.

[23] M. H. Alam, M.-M. Rahoman, and M. A. K. Azad, "Sentiment analysis for Bangla sentences using convolutional neural network," in 2017 20th International Conference of Computer and Information Technology (ICCIT), 2017: IEEE, pp. 1-6.

[24] E. Hossain, O. Sharif, M. M. Hoque, and I. H. Sarker, "Sentilstm: a deep learning approach for sentiment analysis of restaurant reviews," in International Conference on Hybrid Intelligent Systems, 2020: Springer, pp. 193-203.

[25] A. Ahmed and M. A. Yousuf, "Sentiment analysis on Bangla text using long short-term memory (LSTM) recurrent neural network," in Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020, 2020: Springer, pp. 181-192.

[26] K. I. Islam, M. S. Islam, and M. R. Amin, "Sentiment analysis in Bengali via transfer learning using multi-lingual BERT," in 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020: IEEE, pp. 1-5.

[27] M. I. H. Junaid, F. Hossain, U. S. Upal, A. Tameem, A. Kashim, and A. Fahmin, "Bangla food review sentimental analysis using machine learning," in 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022: IEEE, pp. 0347-0353.

[28] E. A. E. Lucky, M. M. H. Sany, M. Keya, S. A. Khushbu, and S. R. H. Noori, "An attention on sentiment analysis of child abusive public comments towards bangla text and ml," in 2021 12th international conference on computing communication and networking technologies (ICCCNT), 2021: IEEE, pp. 1-6.

[29] M. Rahman, M. R. A. Talukder, L. A. Setu, and A. K. Das, "A dynamic strategy for classifying sentiment from Bengali text by utilizing Word2vector model," Journal of Information Technology Research (JITR), vol. 15, no. 1, pp. 1-17, 2022.

[30] S. S. Shanto , Z. Ahmed, N. Hossain, A. Roy, and A. I. Jony , "Binary vs. Multiclass Sentiment Classification for Bangla E-commerce Product Reviews: A Comparative Analysis of Machine Learning Models," International Journal of Information Engineering and Electronic Business (IJIEEB), vol. 15, no. 6, 2023.

[31] Z. Ahmed, S. S. Shanto, and A. I. Jony , "Advancement in Bangla Sentiment Analysis: A Comparative Study of Transformer-Based and Transfer Learning Models for E-commerce Sentiment Classification," Journal of Information Systems Engineering and Business Intelligence, vol. 9, no. 2, Oct. 2023.