A Comparative Study of Deep Learning Models for Bengali Tech Product Reviews Sentiment Analysis

Janak Mallik
Dept. of C.S.E.
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204106@aust.edu

Md. Maruf Rusafi Arnob
Dept. of C.S.E.
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204088@aust.edu

Faiyaz Ahmed Rabbi
Dept. of C.S.E.
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204070@aust.edu

Intishar Rahman Nafi
Dept. of C.S.E.
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204073@aust.edu

I. Abstract

This research thoroughly investigates how people feel about technology products in Bengali reviews. It specifically looks at using advanced computer programs (deep learning models) to understand the sentiments expressed in these reviews. The study compares different models of BERT to see which one is better at capturing the subtle nuances of feelings in the Bengali language. Additionally, the research includes a detailed record of a new dataset containing Bengali tech reviews. This dataset is important for teaching and testing the sentiment analysis models.

II. Introduction

Sentiment analysis, often termed opinion mining, stands as a pivotal tool in comprehending the emotional tone, attitudes, and opinions expressed within text data. Specifically focusing on Bengali Tech Product, sentiment analysis aims to decipher the sentiments conveyed in these textual expressions. It involves the classification of sentiments into various categories like positive, negative, neutral, or even nuanced emotions like happiness, anger, sadness, and more.

The exponential growth of digital content, particularly in the Bangla language, has led to a wealth of opinions and viewpoints shared across online platforms. However, despite the significance of Bangla as the fourth most spoken language globally, the domain of sentiment analysis in this language remains relatively underexplored. Analyzing sentiments within Bangla news comments holds immense value for media organizations, researchers, and businesses, offering insights into public perception, reception of news, and trends within the Bangla-speaking community.

The process involves employing Natural Language Processing (NLP) techniques and machine learning algorithms to analyze and categorize sentiments expressed within Bangla news comments. This includes data preprocessing, feature extraction, and the application of classification models to discern sentiments accurately. Techniques such as Naïve Bayes, Support Vector Machines (SVM), Recurrent Neural Networks (RNNs), or Transformer-based models like BERT may be utilized to decode the emotional content embedded in these comments.

III. Literature Review

Customer review sentiment analysis has gained a lot of attention and become a fascinating study area. Numerous studies explore the in-depth research carried out by eminent academics on product reviews in different languages.With a focus on the specifics of Bengali language in the context of tech product reviews, The following sections offer a thorough analysis and discussion of various academic thesis.

Atiqur et.al developed a machine learning classifiers for sentiment analysis of **movie reviews**. On their work, they take total 2000 movie reviews for training and testing purposes. Their terms are true positive, false positive, true negative ,false negative for sentiment analysis. The sentiment classification was done by using five types of classifiers, namely, MNB with accuracy of 88.50%, BNB with accuracy of 87.50%, SVM with 87.33%, ME with 60.67% and DT with 80.17% [2].

Omar et.al proposed a model that classifies **online restaurant reviews** into positive and negative sentiment implemented by using Multinomial Naïve Bayes with the maximum accuracy of 80.48% for 6-fold cross validation compared to their test with decision tree and Random forest. They have worked with 1000 restaurant reviews in Bengali for their proposed system [4].

Minhajul et al. developed a system for sentiment analysis of over 1000 Bangla-language online customer feedbacks. Using KNN, Decision Tree, SVM, Random Forest, and Logistic Regression, SVM outperformed with 88.81% accuracy. The proposed system aimed to assist sellers in improving products. Logistic Regression followed with 88.09% accuracy, Random Forest with 85.92%, Decision Tree with 83.03%, and KNN with 80.14% [3].

Naveed et.al evaluated RNN variations for predicting the sentiment of Amazon.com mobile phone reviews, using five algorithms (RNN, LSTM-RNN, GLSTM-RNN, GRU-RNN, UG-RNN) and three word embedding methods (Glove, word2vec, FastText). Addressing the challenge of text classification, the authors processed and balanced the dataset. GLSTM-RNN with FastText achieved the highest accuracy (93.75%) for unbalanced datasets, while LSTM-RNN with FastText performed best for balanced datasets (88.39%). The study surpasses previous results, achieving a maximum accuracy of 92.75% with the CNN algorithm [5].

Abu et.al developed a model for distinguishing positive and negative Bengali food reviews using a dataset of 1,400 reviews from popular food delivery service and Facebook pages. The dataset was split into an 80% training set and a 20% testing set. Four machine learning algorithms (Multi-Nominal Naive Bayes, Random Forest, XGBoost, Decision Tree) and three deep learning algorithms (CNN, LSTM, RNN) were employed. The XGBoost model achieved the highest accuracy among machine learning techniques at 89.64%, while the LSTM model led among deep learning methods with an accuracy of 91.07% [1].

IV. DATASETS

This thesis is about creating a big collection of reviews about tech products in Bengali. The process starts by gathering texts from Daraz. The collection covers a wide range of products such as routers, power banks, smartphones, laptops, adapters, neckband, bluetooth speaker, monitor, adapters, headphones, gaming mouses, mechanical keyboards etc and more, making it useful for various tech-related topics.

To enhance the dataset's suitability for language understanding in computer programs, a meticulous annotation process is implemented. Two independent annotators, referred to as Person1 and Person2, individually annotate the entire dataset without access to each other's assessments. Annotations are structured as Positive data (1), Negative data (0), and Neutral data (2) to capture the sentiment nuances expressed in the reviews.

The merging of annotations involves a comparative analysis, highlighting disparities between Person1 and Person2's annotations. In cases of mismatched annotations, a third annotator, Person3, intervenes to provide conclusive decisions based on their judgment. The final merged annotated dataset is derived from Person3's annotations, ensuring a unified and consistent sentiment labeling for each review. This meticulous process aims to enhance the dataset's reliability and utility for training deep learning models.

The final dataset contains around 3100 annotated reviews, providing a substantial and diverse collection of Bengali tech-related opinions. Because it includes different types of products, researchers can use it to study specific areas or compare different tech domains.

This Bengali Tech Review Dataset is valuable for computer programs that analyze language. It can be used for tasks like figuring out sentiments in reviews, understanding opinions, and extracting product features. Additionally, it helps in creating and testing computer models that are specifically designed to understand the Bengali language when it comes to tech reviews.

V. METHODOLOGY

In this study, our goal was to analyze the sentiments conveyed in Bengali tech product reviews using a method known as text sentiment analysis. We introduced a detailed process to enhance the accuracy of sentiment analysis, making use of the Bengali BERT model. We began with an annotated dataset, meaning it already contained information about sentiments. To better understand the text, we utilized BERT Base for creating word representations.

Subsequently, we delved into the classification phase, employing various classifiers including K-Nearest Neighbors (KNN), Decision Tree (DT), Linear Support Vector Machine (SVM), Random Forest, and Gradient Boosting. These classifiers were tasked with categorizing the reviews based on their sentiments.

To evaluate the performance of our models, we generated a classification report. This report included metrics such as precision, recall, F1-score, and support, offering insights into how effectively the computer could discern positive or negative sentiments within the reviews.

Set A focuses on preserving the core meaning of the text. We use code to get rid of unnecessary stuff like punctuation, numbers, stopwords, emojis, and English words. This helps us capture the detailed meaning of the text, which is vital for accurate sentiment analysis.

Set C strikes a balance between cleaning up the text and keeping important details. We take out punctuation, numbers, emojis and English words keeping the stopwords. This way, we maintain relevant information while cutting down on unnecessary clutter.

Set D is all about capturing sentiment expressed in English. We cut down on emojis, punctuation, numbers, and stopwords, but we keep the English words. This set gives a lot of importance to understanding sentiment through the language used, making sure the multi-linguistic context is taken into account.

The proposed preprocessing workflow offers a nuanced and multifaceted approach to text sentiment analysis using Bengal BERT. Each set within the workflow is tailored to address specific aspects of the data, ensuring the preservation of relevant information while minimizing noise. Through comparative experiments, the effectiveness of each set can be evaluated, providing valuable insights into the optimal preprocessing strategy for sentiment analysis with Bengal BERT in diverse linguistic contexts.

Set A
code after reducing - punctuation, numbers, stopwords, emojis, English words

Set C
code after reducing - punctuation, numbers, emojis, English words
keeping - stopwords

Set D
code after reducing -punctuation, numbers, stopwords, emojis
keeping - English words

VI. Experiment Results

In the experiment analysis, various machine learning classifiers were employed to assess the efficacy of sentiment analysis on Bengali tech product reviews. The performance metrics, including accuracy and classification reports, were evaluated for each classifier. The results are summarized below:

Set A
code after reducing - punctuation, numbers, stopwords, emojis, English words

A. K-Nearest Neighbors (KNN) Classifier

- **Accuracy:** 0.70
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.21	0.14	0.17	195
1	0.76	0.91	0.83	905
2	0.00	0.00	0.00	115

- **Macro avg:** 0.32 (Precision), 0.35 (Recall), 0.33 (F1-Score)
- **Weighted avg:** 0.60 (Precision), 0.70 (Recall), 0.64 (F1-Score)

B. Decision Tree (DT) Classifier

- **Accuracy:** 0.53
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.21	0.32	0.25	195
1	0.78	0.62	0.69	905
2	0.10	0.17	0.13	115

- **Macro avg:** 0.36 (Precision), 0.37 (Recall), 0.36 (F1-Score)
- **Weighted avg:** 0.62 (Precision), 0.53 (Recall), 0.57 (F1-Score)

C. Linear Support Vector Machine (SVM) Classifier

- **Accuracy:** 0.65
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.24	0.30	0.27	195
1	0.78	0.80	0.79	905
2	0.13	0.04	0.07	115

- **Macro avg:** 0.38 (Precision), 0.38 (Recall), 0.38 (F1-Score)
- **Weighted avg:** 0.63 (Precision), 0.65 (Recall), 0.64 (F1-Score)

D. Random Forest Classifier

- **Accuracy:** 0.75
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.62	0.04	0.08	195
1	0.75	0.99	0.85	905
2	0.00	0.00	0.00	115

- **Macro avg:** 0.45 (Precision), 0.34 (Recall), 0.31 (F1-Score)
- **Weighted avg:** 0.66 (Precision), 0.75 (Recall), 0.65 (F1-Score)

E. Gradient Boosting Classifier

- **Accuracy:** 0.60
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.21	0.33	0.26	195
1	0.76	0.73	0.74	905
2	0.06	0.03	0.04	115

- **Macro avg:** 0.35 (Precision), 0.36 (Recall), 0.35 (F1-Score)
- **Weighted avg:** 0.61 (Precision), 0.60 (Recall), 0.60 (F1-Score)

Set C
code after reducing - punctuation, numbers, emojis, English words
keeping - stopwords

F. K-Nearest Neighbors (KNN) Classifier

- **Accuracy:** 0.73
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.26	0.06	0.09	195
1	0.75	0.97	0.85	905
2	0.00	0.00	0.00	115

- **Macro avg:** 0.34 (Precision), 0.34 (Recall), 0.31 (F1-Score)
- **Weighted avg:** 0.60 (Precision), 0.73 (Recall), 0.65 (F1-Score)

G. Decision Tree (DT) Classifier

- **Accuracy:** 0.57
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.18	0.24	0.21	195
1	0.76	0.70	0.73	905
2	0.08	0.09	0.08	115

- **Macro avg:** 0.34 (Precision), 0.34 (Recall), 0.34 (F1-Score)
- **Weighted avg:** 0.60 (Precision), 0.57 (Recall), 0.58 (F1-Score)

H. Linear Support Vector Machine (SVM) Classifier

- **Accuracy:** 0.68
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.24	0.18	0.21	195
1	0.76	0.87	0.81	905
2	0.06	0.02	0.03	115

- **Macro avg:** 0.35 (Precision), 0.36 (Recall), 0.35 (F1-Score)

- **Weighted avg:** 0.61 (Precision), 0.68 (Recall), 0.64 (F1-Score)

I. Random Forest Classifier

- **Accuracy:** 0.74
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.50	0.03	0.05	195
1	0.75	0.99	0.85	905
2	0.00	0.00	0.00	115

- **Macro avg:** 0.42 (Precision), 0.34 (Recall), 0.30 (F1-Score)
- **Weighted avg:** 0.64 (Precision), 0.74 (Recall), 0.64 (F1-Score)

Set D
code after reducing -punctuation, numbers, stopwords, emojis keeping - English words

J. K-Nearest Neighbors (KNN) Classifier

- **Accuracy:** 0.67
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.19	0.17	0.18	195
1	0.75	0.87	0.81	905
2	0.00	0.00	0.00	115

- **Macro avg:** 0.32 (Precision), 0.35 (Recall), 0.33 (F1-Score)
- **Weighted avg:** 0.59 (Precision), 0.67 (Recall), 0.63 (F1-Score)

K. Decision Tree (DT) Classifier

- **Accuracy:** 0.60
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.19	0.21	0.20	195
1	0.75	0.75	0.75	905
2	0.10	0.08	0.09	115

- **Macro avg:** 0.35 (Precision), 0.35 (Recall), 0.35 (F1-Score)
- **Weighted avg:** 0.60 (Precision), 0.60 (Recall), 0.60 (F1-Score)

L. Linear Support Vector Machine (SVM) Classifier

- **Accuracy:** 0.64
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.23	0.27	0.25	195
1	0.77	0.80	0.78	905
2	0.11	0.04	0.06	115

- **Macro avg:** 0.37 (Precision), 0.37 (Recall), 0.37 (F1-Score)
- **Weighted avg:** 0.62 (Precision), 0.64 (Recall), 0.63 (F1-Score)

M. Random Forest Classifier

- **Accuracy:** 0.73
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.15	0.02	0.04	195
1	0.75	0.98	0.84	905
2	0.00	0.00	0.00	115

- **Macro avg:** 0.30 (Precision), 0.33 (Recall), 0.29 (F1-Score)
- **Weighted avg:** 0.58 (Precision), 0.73 (Recall), 0.64 (F1-Score)

N. Gradient Boosting Classifier

- **Accuracy:** 0.65
- **Classification Report:**

	Precision	Recall	F1-Score	Support
0	0.18	0.15	0.16	195
1	0.75	0.84	0.79	905
2	0.10	0.03	0.05	115

- **Macro avg:** 0.34 (Precision), 0.34 (Recall), 0.34 (F1-Score)
- **Weighted avg:** 0.59 (Precision), 0.65 (Recall), 0.62 (F1-Score)

These results provide insights into the comparative performance of different classifiers, helping to identify the most suitable model for sentiment analysis on Bengali tech product reviews.

VII. Result Analysis

All these result give almost close numbers of accuracy with the highest one for code after reducing all (punctuation, numbers, stopwords, emojis, English words) for Random Forest Classifier with the overall accuracy 75%.

- **Set A**
  K-Nearest Neighbors (KNN) Classifier: Achieved an precision of 70%, with notable precision for class 1 (76%) but lower performance for classes 0 and 2.
  Decision Tree (DT) Classifier: Attained an accuracy of 53%, with relatively balanced precision but lower recall and F1-scores for all classes.
  Linear Support Vector Machine (SVM) Classifier: Showed an accuracy of 65%, with better performance for class 1 but lower precision and recall for classes 0 and 2.
  Random Forest Classifier: Outperformed other classifiers with an accuracy of 75%, displaying high precision and recall for class 1 but challenges in recognizing classes 0 and 2.
  Gradient Boosting Classifier: Achieved an precision of 60%, with class 1 exhibiting the best precision and recall.
- **Set C**
  K-Nearest Neighbors (KNN) Classifier: Improved precision to 73%, with an emphasis on class 1.

Decision Tree (DT) Classifier: Showed an accuracy of 57%, with balanced precision and recall.

Linear Support Vector Machine (SVM) Classifier: Attained 68% precision, with improved precision and recall for class 1.

Random Forest Classifier: Maintained a high accuracy of 74%, with class 1 dominance but challenges in recognizing classes 0 and 2.

Gradient Boosting Classifier: Achieved an accuracy of 60%, similar to Set A.

- **Set D**

  K-Nearest Neighbors (KNN) Classifier: Displayed a 67% precision, with improved recall for class 1.

  Decision Tree (DT) Classifier: Achieved an accuracy of 60%, with balanced precision and recall.

  Linear Support Vector Machine (SVM) Classifier: Attained 64% precision, with class 1 exhibiting the best precision and recall.

  Random Forest Classifier: Maintained a high accuracy of 73%, with challenges in recognizing class 0.

  Gradient Boosting Classifier: Achieved an accuracy of 65%, similar to Set A and Set C.

- **Analysis**
  - Random Forest consistently outperformed other classifiers across all sets, emphasizing its robustness.
  - The models struggled with class imbalances and exhibited challenges in recognizing minority classes.
  - The choice of features in each set influenced the performance, highlighting the importance of feature selection in sentiment analysis.
  - Further fine-tuning, ensemble methods, and exploration of advanced architectures may enhance model performance.
  - Increased dataset size and diversity could potentially improve the models' ability to generalize.

VIII. Conclusion And Future Works

In this study, we successfully developed a Bengali text review dataset comprising 3000 (60% for training instances and employed BERT base for embedding generation. Leveraging these embeddings, we implemented basic machine learning models for sentiment analysis, classifying the texts into three distinct classes. Our model achieved a highest accuracy of 75%, indicating the effectiveness of the chosen approach in capturing sentiment patterns within Bengali text. Working exclusively with a dataset of 3000 instances, where 60% was allocated for training, suggests that an increase in the dataset size could potentially lead to improved accuracy.

In order to further improve the performance of our Bengali sentiment analysis model based on BERT, fine-tuning strategies can be implemented. This involves customizing the pre-trained BERT model on a domain-specific corpus or employing transfer learning techniques to enhance its understanding of the intricacies of sentiment expression in Bengali. Additionally, enhancing the dataset's size and diversity through data augmentation techniques or the collection of a more expansive range of Bengali text samples can contribute to increased model robustness and generalization.

References

[1] Abu Kowshir Bitto, Md Hasan Imam Bijoy, Md Shohel Arman, Imran Mahmud, Aka Das, and Joy Majumder. Sentiment analysis from bangladeshi food delivery startup based on user reviews using machine learning and deep learning. *Bulletin of Electrical Engineering and Informatics*, 12(4):2282–2291, 2023.

[2] Atiqur Rahman and Md. Sharif Hossen. Sentiment analysis on movie review data using machine learning approach. In *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–4, 2019.

[3] Minhajul Abedin Shafin, Md. Mehedi Hasan, Md. Rejaul Alam, Mosaddek Ali Mithu, Arafat Ulllah Nur, and Md. Omar 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)*, pages 1–5, 2020.

[4] Omar Sharif, Mohammed Moshiul Hoque, and Eftekhar 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)*, pages 1–6, 2019.

[5] Naveed Sultan. Sentiment analysis of amazon product reviews using supervised machine learning techniques. *Knowledge Engineering and Data Science*, 5(1):101–108, 2022.