# Team Error Point at BLP-2023 Task 2: A Comparative Exploration of Hybrid Deep Learning and Machine Learning Approach for Advanced Sentiment Analysis Techniques

**Rajesh Kumar Das, Kabid Yeiad, Moshfiqur Rahman Ajmain, Jannatul Maowa, Mirajul Islam, Sharun Akter Khushbu**
Department of Computer Science and Engineering
Daffodil International University, Dhaka, Bangladesh
{rajesh15-13032, yeiad15-14440, moshfiqur15-14090, jannatul15-14095, merajul15-9627, sharun.cse}@diu.edu.bd

## Abstract

This paper presents a thorough and extensive investigation into the diverse models and techniques utilized for sentiment analysis. What sets this research apart is the deliberate and purposeful incorporation of data augmentation techniques with the goal of improving the efficacy of sentiment analysis in the Bangla language. We systematically explore various approaches, including preprocessing techniques, advanced models like Long Short-Term Memory (LSTM) and LSTM-CNN (Convolutional Neural Network) Combine, and traditional machine learning models such as Logistic Regression, Decision Tree, Random Forest, Multi-Naive Bayes, Support Vector Machine, and Stochastic Gradient Descent. Our study highlights the substantial impact of data augmentation on enhancing model accuracy and understanding Bangla sentiment nuances. Additionally, we emphasize the LSTM model's ability to capture long-range correlations in Bangla text. Our system scored 0.4129 and ranked 27th among the participants.

## 1 Introduction

Sentiment analysis, the process of extracting emotional information from textual data, has witnessed significant advancements in recent years. Our participation in the Sentiment Analysis Shared Task-2 at the BLP Workshop during EMNLP 2023 underscores our progress in Bangla Language Processing (BLP) and sentiment analysis (Hasan et al., 2023a). This study arises from the critical need to address sentiment expression issues specific to Bangla, a language with distinct linguistic nuances. Additionally, with the proliferation of Bangla content online, effective sentiment analysis tools are invaluable for applications ranging from social media monitoring to customer feedback analysis. (Jahan et al., 2021) Pronoun Replacement-Based Special Tagging System (PRS-TS) highlights context-specific language, improving Bangla sentiment analysis. The use of a Broad Multitask Transformer Network (BMT-Net) showed that multitask learning works in sentiment analysis (Zhang et al., 2022). (Zhang and Qian, 2020) Convolution over Hierarchical Syntactic and Lexical Graphs revealed ways to use syntactic and lexical information for aspect-level sentiment analysis. (Zhang et al., 2020) Convolutional multi-head self-attention on memory improved aspect sentiment categorization. The fusion strategy by (Zhou et al., 2020) for hate speech detection and the augmentation of BERT representations with context-aware embedding demonstrate contextual embeddings potential in sentiment analysis (Li and et al., 2020). (Hosain Sumit et al., 2018) Bangla Sentiment Analysis uses word embeddings to adapt to different languages. Long Short-Term Memory (LSTM) networks in hardware-accelerated sentiment analysis have also expanded this field (Wen and et al., 2021). Twitter is a popular social media tool for sentiment research. (Sigirci et al., 2020) use of heterogeneous multi-layer network representation and embedding shows new ways to look at unstructured textual data.

Our comprehensive study uses conventional preprocessing methods, advanced models like Long Short-Term Memory (LSTM) and LSTM-CNN Combine, and traditional machine learning models like Logistic Regression, Decision Tree, Random Forest, Multi-Naive Bayes, Support vector machine (SVM), and Stochastic gradient descent (SGD). Deliberate data augmentation is a hallmark of our study. Strategic augmentation has improved our dataset and sentiment analysis approaches, demonstrating data augmentation's ability to improve model accuracy and illuminate Bangla sentiment expression. We analyse LSTM and LSTM-CNN models with and without data augmentation as our main focus. We use dataset partition, performance evaluation criteria, and extensive per-class analysis in our experiments. The following discussion emphasises data augmentation's importance for model



*Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 331–335
December 7, 2023 

efficacy. Comparing LSTM models to combined LSTM-CNN models shows that the former captures long-range correlations in Bangla text better, advancing Bangla sentiment analysis research. [1] final implementation with an anonymous GitHub link[2].

## 2 Literature Review

Recent studies in sentiment analysis, particularly in Bangla Language Processing (BLP), have catalysed the field (Hasan et al., 2023b). A key aspect of this progress lies in the development of specialised techniques for Bangla sentiment analysis. (Ritu et al., 2018) showed how word embeddings can be used in different linguistic settings. Another study by (Rahman et al., 2020) looked into more complex models, specifically how to group opinions in Bangla sentences. Considering structural aspects in sentiment analysis, (Tuhin et al., 2019) engineered an automated system for sentiment analysis from Bangla text using supervised learning techniques. (Abdalla and Özyurt, 2021) underscored the flexibility of deep learning techniques through a comprehensive sentiment analysis spanning various domains. Innovative methodologies are exemplified by (Zhu et al., 2018) bi-directional LSTM-CNN model, placing emphasis on fine-grained sentiment information extraction. (Wang et al., 2020) introduced an emotion-semantic-enhanced bidirectional LSTM with a multi-head attention mechanism for microblog sentiment analysis, showcasing the potential of attention mechanisms.(Luan and Lin, 2019) demonstrated the effectiveness of convolutional and recurrent neural network models for sentiment analysis tasks. (Hasan et al., 2023a) comparative study on modeling approaches for Bangla Sentiment Analysis yielded valuable insights. Moreover, (Islam et al., 2021) introduced SentNoB, a valuable resource for scrutinizing sentiment in informal and noisy textual data. Finally, (Zhou et al., 2016) integrated bidirectional LSTM with two-dimensional max pooling, showcasing the potential of amalgamating techniques for sentiment analysis tasks.

[1]https://github.com/blp-workshop/blp_task2#leaderboard

[2]https://anonymous.4open.science/r/EMNLP_2023_BLP_Workshop_Task2-46AE

## 3 Data and Methodology

Within the section, we provide a comprehensive overview of the data sources utilized and the rigorous research methodologies employed, ensuring transparency and credibility in our approach.

### 3.1 Dataset Description

Our study utilized the dataset sourced from BLP-2023 Task 2 (Hasan et al., 2023b) with the objective of discerning the sentiment expressed within textual content. This task involves the classification of sentiment into three categories: positive, negative, or neutral, thereby presenting a multi-class classification challenge. In Table 1, we present an overview of the dataset distribution used for experimentation in this shared task.

Table 1: Data splits and distributions of Shared Task-2

| Class Label | Train | Dev | Test | Total |
|---|---|---|---|---|
| Negative | 15767 | 1753 | 3338 | 20858 |
| Positive | 12364 | 1388 | 2092 | 15844 |
| Neutral | 7135 | 793 | 1277 | 9205 |
| Total | 35266 | 3934 | 6707 | 45907 |

Table 2: Dataset Split for Machine Learning Algorithms with and without Augmentation

| Data Augmentation | Training Set Size | Testing Set Size | Total Dataset Size |
|---|---|---|---|
| No | 20472 | 5118 | 25590 |
| Yes | 31379 | 7845 | 39224 |

Table 3: Dataset Split for Deep Learning Models with and without Data Augmentation

| Data Augmentation | No | Yes |
|---|---|---|
| Training Set Size | 16,377 | 19,433 |
| Testing Set Size | 5,118 | 6,073 |
| Validation Set Size | 4,095 | 4,859 |
| Total Dataset Size | 25,590 | 30,365 |

Table 2 presents the dataset partitioning for machine learning algorithms, highlighting distinctions between augmented and non-augmented data subsets. It offers a clear overview of the experimental design for model evaluation.

Table 3 shows a complete distribution of the deep learning dataset, separating augmented and non-augmented data segments. The academic setting relies on it to explain the experimental framework, especially for data augmentation. Figure 1 presents a word cloud representation for three sentiment categories: positive, negative, and neutral.


Figure 1: Word Cloud

### 3.2 Preprocessing

The BLP-2023 Task 2 dataset comprises two main components: the Multiplatform Bangla Sentiment (MUBASE) and SentNob datasets. The SentNob dataset encompasses public comments from various domains, including politics, education, and agriculture, sourced from news articles and videos. Meanwhile, the MUBASE dataset is a cross-platform compilation containing content from both Facebook and Twitter posts, all meticulously annotated to indicate sentiment polarity. As part of our preprocessing steps, we performed duplicate removal, filtered by text length, removed punctuation, links, emojis, non-character elements, and eliminated stopwords. We excluded very short or extremely long texts to focus on those that provide meaningful insights. Short texts might lack context, while overly long ones could introduce noise. In the process of removing stopwords, we systematically eliminate common, non-informative words to enhance the text's focus on meaningful content.

### 3.3 Algorithms

In our classification experiments, we employed a dual approach, encompassing both deep learning models and traditional machine learning algorithms like logistic regression (Nick and Campbell, 2007), decision trees (Kotsiantis, 2013), random forests (Rigatti, 2017), multi-naive bayes (Rish, 2001), SVM (Yang et al., 2012), and SGD (Chauhan et al., 2013). Specifically, within the domain of deep learning, we utilized the Long Short-Term Memory (LSTM) (Yu et al., 2019) model as well as a hybrid model combining LSTM and the Convolutional Neural Network (CNN) architecture (Li et al., 2021). This comprehensive approach allowed us to harness the strengths of both traditional and state-of-the-art methodologies, enhancing the depth and breadth of our analytical exploration.

### 3.4 Experimental Setup

In order to train the traditional models, we commenced by transforming the preprocessed data into TF-IDF vectors, integrating weighted n-grams, encompassing unigrams, bigrams, and trigrams. This approach was adopted to harness contextual information effectively. To address class imbalance, we implemented an up-sampling technique specifically focused on the neutral class within the merged dataset. We have used the train_test_split method from scikit-learn to organize the data for machine learning. This method divides the data into two parts: one for training (80%) and one for testing (20%). The parameters were selected to optimize model performance and ensure robustness in our deep learning-based classification approach listed in Table 6.

## 4 Results and Discussion

In this section, we present the outcomes of our experiments and engage in a comprehensive analysis of the findings.

Table 4: Performance scores for ML Models (With Augmentation)

| Model Name | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| Logistic Regression | 71.88 | 72.52 | 71.88 | 71.50 |
| Decision Tree | 65.29 | 64.79 | 65.29 | 64.67 |
| Random Forest | 72.36 | 73.36 | 72.36 | 71.79 |
| Multi. Naive Bayes | 71.22 | 72.51 | 71.22 | 70.83 |
| SVM | 75.02 | 75.26 | 75.02 | 74.85 |
| SGD | 60.84 | 65.69 | 60.84 | 59.34 |

Table 5: Performance scores for ML Models (Without Augmentation)

| Model Name | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| Logistic Regression | 64.20 | 66.81 | 64.20 | 59.55 |
| Decision Tree | 55.84 | 55.91 | 55.84 | 55.87 |
| Random Forest | 61.65 | 60.36 | 61.65 | 59.74 |
| Multi. Naive Bayes | 62.84 | 62.97 | 62.84 | 62.89 |
| SVM | 65.89 | 66.03 | 65.89 | 62.30 |
| SGD | 59.44 | 69.29 | 59.44 | 52.47 |

Table 4 displays machine learning model scores with data augmentation. SVM excels with 75.02% accuracy, showcasing its prowess in handling large datasets, clear separation, and noise robustness for

Table 6: Experimental setup for both DL models

| Model | Data Augmentation | Embedding Dimension | Input Length | Vocabulary Size | Number of Classes | Batch Size | Number of Epochs |
|---|---|---|---|---|---|---|---|
| **LSTM** | No | 128 | 300 | 5,000 | 3 | 64 | 50 |
| **LSTM** | Yes | 128 | 300 | 5,000 | 3 | 64 | 50 |
| **LSTM-CNN** | No | 128 | 300 | 5,000 | 3 | 64 | 50 |
| **LSTM-CNN** | Yes | 128 | 300 | 5,000 | 3 | 64 | 50 |

Table 7: Performance scores for Deep Learning Models

| Model | Augmentation | Class | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) |
|---|---|---|---|---|---|---|
| LSTM | With | Positive | 70.94 | 64.45 | 67.54 | 68.43 |
| | | Negative | 70.52 | 78.24 | 74.18 | |
| | | Neutral | 63.04 | 63.07 | 63.06 | |
| LSTM-CNN | With | Positive | 67.85 | 64.79 | 66.29 | 67.59 |
| | | Negative | 71.88 | 77.16 | 74.43 | |
| | | Neutral | 62.25 | 60.97 | 61.60 | |
| LSTM | Without | Positive | 65.91 | 65.88 | 65.89 | 58.89 |
| | | Negative | 36.88 | 30.64 | 33.47 | |
| | | Neutral | 59.22 | 64.22 | 61.62 | |
| LSTM-CNN | Without | Positive | 64.01 | 67.90 | 65.90 | 57.74 |
| | | Negative | 34.28 | 37.67 | 35.93 | |
| | | Neutral | 62.94 | 55.03 | 58.72 | |

sentiment analysis. In contrast, SGD underperforms at 60.84% accuracy, indicating challenges with complex datasets or potential tuning requirements. Table 5 displays machine learning model performance metrics without data augmentation. SVM leads with 65.89% accuracy, validating its effectiveness in sentiment classification. In contrast, SGD underperforms with 59.44% accuracy, suggesting difficulties in handling dataset complexity without data augmentation. Table 7 summarizes deep learning model performance. "With Augmentation," LSTM excels in positive sentiment accuracy at 68.43%, and LSTM-CNN leads with 67.59% in negative sentiment accuracy. "Without Augmentation," LSTM's positive accuracy drops to 58.89%, and LSTM-CNN achieves 57.74% in negative sentiment, showing data augmentation's benefit.

## 5 Conclusion

This research offers a comprehensive examination of sentiment analysis in Bangla. It explores various models and techniques, traditional and advanced, with and without data augmentation. While not specifying accuracy rates, data augmentation notably boosts model effectiveness. Our study underscores the importance of addressing Bangla's unique challenges in sentiment analysis and the role of data augmentation. Comparative analysis between LSTM and LSTM-CNN models reveals LSTM's proficiency in capturing long-range correlations in Bangla text. These findings advance Bangla sentiment analysis and lay the groundwork for future research in this field.

## References


G. Abdalla and F. Özyurt. 2021. Sentiment analysis of fast food companies with deep learning models. *The Computer Journal*, 64(3):383–390.

H. Chauhan, V. Kumar, S. Pundir, and E.S. Pilli. 2013. A comparative study of classification techniques for intrusion detection. In *2013 International Symposium on Computational and Business Intelligence*, pages 40–43. IEEE.

M. A. Hasan, S. Das, A. Anjum, F. Alam, A. Anjum, A. Sarker, and S. R. H. Noori. 2023a. Zero- and few-shot prompting with llms: A comparative study with fine-tuned models for bangla sentiment analysis. *arXiv [Cs.CL]*. Http://arxiv.org/abs/2308.10783.

Md. Arid Hasan, Firoj Alam, Anika Anjum, Shudipta Das, and Afiyat Anjum. 2023b. Blp-2023 task 2: Sentiment analysis. In *Proceedings of the 1st International Workshop on Bangla Language Processing (BLP-2023)*, Singapore. Association for Computational Linguistics.

S. Hosain Sumit, M. Zakir Hossan, T. Al Muntasir, and T. Sourov. 2018. Exploring word embedding for bangla sentiment analysis. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–5.

K. I. Islam, S. Kar, M. S. Islam, and M. R. Amin. 2021. Sentnob: A dataset for analysing sentiment on noisy bangla texts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3265–3271.

Busrat Jahan, Md. Ismail Emon, Sharmin Milu, Mohammad Hossain, and S. S. Mahtab. 2021. A pronoun replacement-based special tagging system for bengali language processing (blp). In *Proceedings of the Conference*, page 80.

S.B. Kotsiantis. 2013. Decision trees: a recent overview. *Artificial Intelligence Review*, 39:261–283.

X. Li and et al. 2020. Enhancing bert representation with context-aware embedding for aspect-based sentiment analysis. *IEEE Access*, 8:46868–46876.

Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. 2021. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*.

Y. Luan and S. Lin. 2019. Research on text classification based on cnn and lstm. In *2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*.

T.G. Nick and K.M. Campbell. 2007. Logistic regression. In *Topics in biostatistics*, pages 273–301. Springer.

Moqsadur Rahman, Summit Haque, and Zillur Rahman Saurav. 2020. Identifying and categorizing opinions expressed in bangla sentences using deep learning technique. *International Journal of Computer Applications (0975 – 8887)*, 176(17).

S.J. Rigatti. 2017. Random forest. *Journal of Insurance Medicine*, 47(1):31–39.

I. Rish. 2001. An empirical study of the naive bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, volume 3, pages 41–46.

Z.S. Ritu, N. Nowshin, M.M.H. Nahid, and S. Ismail. 2018. Performance analysis of different word embedding models on bangla language. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–5.

I.O. Sigirci, H. Özgür, A. Oluk, H. Uz, E. Çetiner, H.U. Oktay, and K. Erdemir. 2020. Sentiment analysis of turkish reviews on google play store. In *2020 5th International Conference on Computer Science and Engineering (UBMK)*, pages 314–315.

R.A. Tuhin, B.K. Paul, F. Nawrine, M. Akter, and A.K. Das. 2019. 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)*, pages 360–364.

S. Wang, Y. Zhu, W. Gao, M. Cao, and M. Li. 2020. Emotion-semantic-enhanced bidirectional lstm with multi-head attention mechanism for microblog sentiment analysis. *Information*, 11(5):280.

S. Wen and et al. 2021. Memristive lstm network for sentiment analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(3):1794–1804.

Y. Yang, J. Wang, and Y. Yang. 2012. Improving svm classifier with prior knowledge in microcalcification detection. In *2012 19th IEEE International Conference on Image Processing*, pages 2837–2840. IEEE.

Y. Yu, X. Si, C. Hu, and J. Zhang. 2019. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–70.

Mi Zhang and Tieyun Qian. 2020. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In *Proceedings of the Conference*, pages 3540–3549.

T. Zhang, X. Gong, and C. L. P. Chen. 2022. Bmt-net: Broad multitask transformer network for sentiment analysis. *IEEE Transactions on Cybernetics*, 52(7):6232–6243.

Y. Zhang, B. Xu, and T. Zhao. 2020. Convolutional multi-head self-attention on memory for aspect sentiment classification. *IEEE/CAA Journal of Automatica Sinica*, 7(4):1038–1044.

P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu. 2016. Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. *arXiv preprint arXiv:1611.06639*.

Y. Zhou, Y. Yang, H. Liu, X. Liu, and N. Savage. 2020. Deep learning based fusion approach for hate speech detection. *IEEE Access*, 8:128923–128929.

Y. Zhu, X. Gao, W. Zhang, S. Liu, and Y. Zhang. 2018. A bi-directional lstm-cnn model with attention for aspect-level text classification. *Future Internet*, 10(12):116.