Heliyon

A Comparative Study on Text Classification in Bangla and English Using Hybrid Deep Learning Approaches: An Investigation Towards Advancing NLP Research

--Manuscript Draft--

Manuscript Number:	
Article Type:	Original Research Article
Section/Category:	Physical and Applied Sciences
Keywords:	E-commerce; Sentiment Analysis; NLP; Hybrid Deep Learning; LSTM; Bi-LSTM; Conv1D
Manuscript Classifications:	20: Computer Science
Corresponding Author:	Mirajul Islam Daffodil International University BANGLADESH
First Author:	Rajesh Kumar Das
Order of Authors:	Rajesh Kumar Das
	Md. Mahmudul Hasan
	Mirajul Islam
	Sultana Razia
	Mocksidul Hassan
	Sharun Akter Khushbu
Abstract:	This research paper presents an investigation into the efficacy of various machine learning models, including deep learning and combined-based models, for text classification in both English and Bangla languages. The study focuses on sentiment analysis of comments from a popular Bengali e-commerce site, "DARAZ," which comprises both Bangla and translated English reviews. To overcome the challenge of limited resources for Bangla text classification, the paper proposes multiple machine learning approaches, including deep learning and a combined-based model. A modified text set is developed and applied in the preprocessing stage to enhance the performance of the models. The results reveal that the Support Vector Machine and Multi-Class Naive Bayes machine learning models perform better than other models, achieving an accuracy of 80.70% for English text sentiment analysis. For Bangla text, the Support Vector Machine achieves an accuracy of 87.47%. The study also investigates multi-layer deep learning models, including LSTM, Bi-LSTM, and Conv1D, for both languages. The combined multi-layer Conv1D and LSTM-Based Model shows 74.69% accuracy for English and 76.69% accuracy for Bangla sentiment analysis. The LSTM-Based Model demonstrates the best performance in predicting text in both English and Bangla. The paper concludes that the proposed models, with the modified text set, represent significant progress in natural language processing research for text classification, particularly for Bangla.
Suggested Reviewers:	
Opposed Reviewers:	

A Comparative Study on Text Classification in Bangla and English Using Hybrid Deep Learning Approaches: An Investigation Towards Advancing NLP Research

Rajesh Kumar Das[1], Md. Mahmudul Hasan[1], Mirajul Islam[1,2], Sultana Razia[1], Mocksidul Hassan[1] and Sharun Akter Khushbu[1]
[1] Department of Computer Science and Engineering, Daffodil International University, Dhaka 1341, Bangladesh
[2]Faculty of Graduate Studies, Daffodil International University, Dhaka 1341, Bangladesh
Correspondence: Mirajul Islam; merajul15-9627@diu.edu.bd

Abstract:

This research paper presents an investigation into the efficacy of various machine learning models, including deep learning and combined-based models, for text classification in both English and Bangla languages. The study focuses on sentiment analysis of comments from a popular Bengali e-commerce site, "DARAZ," which comprises both Bangla and translated English reviews. To overcome the challenge of limited resources for Bangla text classification, the paper proposes multiple machine learning approaches, including deep learning and a combined-based model. A modified text set is developed and applied in the preprocessing stage to enhance the performance of the models. The results reveal that the Support Vector Machine and Multi-Class Naive Bayes machine learning models perform better than other models, achieving an accuracy of 80.70% for English text sentiment analysis. For Bangla text, the Support Vector Machine achieves an accuracy of 87.47%. The study also investigates multi-layer deep learning models, including LSTM, Bi-LSTM, and Conv1D, for both languages. The combined multi-layer Conv1D and LSTM-Based Model shows 74.69% accuracy for English and 76.69% accuracy for Bangla sentiment analysis. The LSTM-Based Model demonstrates the best performance in predicting text in both English and Bangla. The paper concludes that the proposed models, with the modified text set, represent significant progress in natural language processing research for text classification, particularly for Bangla.

Keywords— E-commerce, Sentiment Analysis, NLP, Hybrid Deep Learning, LSTM, Bi-LSTM, Conv1D

1. INTRODUCTION

E-commerce platforms have emerged as a widely adopted option for companies, providing them with a diverse array of online purchasing and sales possibilities. These platforms enable consumers to make purchases without having to visit a physical store, unlike regular websites which are often used for information gathering [1]. One such e-commerce website, Daraz, is a popular online shopping marketplace in South Asia, including Bangladesh, Pakistan, Sri Lanka, Myanmar, and Nepal. It offers a diverse range of products, with more than 2.5 million items across various categories such as consumer electronics, household goods, beauty, fashion, sports equipment, and groceries. The COVID-19 pandemic has led to a significant increase in online buying as countries issued stay-at-home orders for their citizens. With the majority of retail stores closed and fears of COVID-19 infections, online shopping has become the primary mode for consumers to meet their consumption demands [2].

Bangla is a widely spoken language in Bangladesh and many other countries around the world. As a result, Bangla natural language processing (BNLP) has garnered considerable attention in the field of NLP [3]. Text classification or natural language processing has been one of the earliest problems in NLP [4]. However, despite the existence of a large number of E-commerce sites with comment sections allowing for the expression of opinions in the Bengali language, little research has been conducted on Bangla sentiment analysis, which is concerning. English, on the other hand, is widely used in various industries, including search engines, social media, and customer service, making it a crucial language for NLP support. It is also the primary language of many popular online platforms, such as Google, Facebook, and Wikipedia, generating vast amounts of data in English. Additionally, it is the most commonly used language for scientific publications, making it a focal point for NLP applications in scientific research and knowledge extraction [5,6].

Sentiment analysis is a critical component in assessing opinions on topics such as politics, sports, finances, and product reviews. As humans are subjective, opinions are valuable. E-commerce sites, for instance, are filled with varying perspectives on products, making it essential to detect which statements are positive or negative [7]. Through machine and deep learning algorithms and natural language processing techniques, it is now possible to detect cyberbullying and distinguish between bullying and non-bullying statements [8]. Over the past few years, machine learning has emerged as a potent technique for processing data and computation, enabling smart capabilities across a range of applications [9]. Machine learning algorithms employ statistical, probabilistic, and optimization methods to learn from previous experiences and detect useful patterns in large, complex, and unstructured datasets [10]. The potential uses of these algorithms are numerous, including automatic text categorization [11], breast cancer risk prediction [12], data augmentation in dermatology image recognition using machine learning [13], machine learning for speech processing [14], improving medical diagnosis accuracy with causal machine learning [15], statistical arbitrage in cryptocurrency markets using machine learning [16], and classifying fake news [17], among others. Moreover, deep learning algorithms are becoming increasingly significant in different research areas, including but not limited to binary classification-supported multi-class skin lesion classification [18], and analysis of cellular images [19]. User-generated content (UGC) has emerged as a valuable source for understanding consumers' emotions and experiences regarding online retail services. Text-mining techniques can be employed to identify different aspects of online posts, such as product, retailer promotion, delivery, payment, communication, return/refund, and price [20]. As deep learning technology continues to progress at a rapid pace, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) have emerged as two of the most potent neural networks [21]. CNN models can be built from the infrared spectra without preprocessing the data using hyperparameter adjustment and saliency map [22]. Machine learning and deep learning techniques can be used for text classification, such as Bidirectional Long-Short Term Memory (Bi-LSTM) and CNN models [23], Coordinated CNN-LSTM-Attention (CCLA) models [24], and Tree-Structured Regional CNN-LSTM models [25], for text sentiment classification and dimensional sentiment analysis. Therefore, these techniques can help analyze user-generated content for insights into consumers' perceptions and experiences.

Recently many academic and commercial researchers are currently studying and exploring sentiment analysis [26-29]. In this literature review, we explore various approaches to text sentiment analysis proposed in recent academic research. Liu et al. [30] The text sentiment classification was performed by assessing a bi-LSTM-based structure, incorporating an attention mechanism and convolutional layer. Similarly, in paper [31], a CNN-based Model was presented for sentiment Classification. In Paper [32], a two-layer, bidirectional LSTM network combined with complicated sentiment analysis units was proposed. In contrast, Du et al .[33] suggested that CNN is a viable model for extracting attention from text and conducted a series of experiments. Kim et al. [34] completed a significant number of experiments on one-layer convolutional neural networks. Chatterjee et al. [35] applied two components - content polarity and overall sentiment content - to accurately depict sentiment found in textual data. In the financial domain, Nelson et al. [36] recommended using LSTM networks with technical analysis indicators to predict stock price. In paper [37], sentiment polarity was identified as an important indication of client happiness, enabling businesses to better understand their customers. In terms of classification approaches, Zhou et al.

[38] proposed a standard CNN and LSTM mixed text categorization approach. Alhawarat et al. [39] developed an Arabic text categorization system using CNNs and TF-IDF feature extraction achieving 98.89% accuracy. In language-specific studies, Chowdhury et al. [40] conducted sentiment analysis on 4000 manually translated Bangla movie reviews achieving 82.42% accuracy with LSTM. In paper [44], a comparative analysis was conducted on traditional SVM and deep learning (LSTM and CNN) algorithms for classifying sentiment in Bangla news comments. Finally, Chakraborty et al. [41] used machine learning to predict feature ratings and diagnostics from text, while Zhang et al. [42] presented a CNN for text classification at the character level that significantly improved accuracy. In paper [43], a unique method using differential privacy was proposed to improve the LSTM model's stock prediction capabilities. Paper [44] performed a comparative analysis for classifying sentiment in Bangla news comments using both traditional SVM and deep learning (LSTM and CNN) algorithms. Paper [45] compared the performance of back-propagation-based neural networks for text classification when compared to other supervised machine learning models. Overall, these studies offer valuable insights into the development and application of various neural network-based models for text sentiment analysis and classification.

After conducting a literature review, it has been found that numerous academic and commercial researchers have focused on developing machine learning and deep learning-based algorithms for analyzing text data in Bangla. However, it has also been observed that the majority of this research has been centered on English text, with relatively less attention being paid to the Bengali language. This knowledge gap in the research on Bengali text data is significant as it limits the potential for advancements in various areas such as sentiment analysis, natural language processing, and text classification. As a result, it is imperative to extend research endeavors towards Bengali text data and to investigate the potential of leveraging machine learning and deep learning methodologies to analyze this language.

The principal contributions of this paper are listed as follows:

1. Build a comprehensive dataset for sentiment analysis in Bangla by collecting reviews in Bangla, along with their English translations. We also performed data cleaning and preprocessing to prepare a suitable Bangla and English comments dataset for analysis.
2. We conducted a comparative analysis between traditional machine learning models, including Logistic Regression, Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor, Support Vector Classifier (SVC), Stochastic Gradient Descent (SGD), and deep learning-based models. To evaluate the performance of each model, we analyzed the values of accuracy, precision, recall, F1-score, and loss metrics of the confusion matrix. Additionally, we determined the maximum, minimum, and average length of Bangla and English text comments. Our analysis provides insights into the effectiveness of both traditional and deep learning-based approaches for text classification in the context of Bangla and English comments.
3. Four unique neural network architectures were designed, namely, Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Convolutional Neural Network with one-dimensional convolutional layers (Conv1D), and a hybrid architecture comprising of Conv1D and LSTM layers (Conv1D-LSTM), to investigate their respective capabilities in modeling sequential data.

The structure of the paper is as follows: Section II presents a description of the materials and methodology employed in this study. The experimental investigation, encompassing the outcomes and performance, is presented in Section III, while Section IV provides a summary of the conclusion of this article.

2. PROPOSED METHODOLOGY

The methodology employed in this study involves several stages, namely data collection, data preparation, model selection, statistical evaluation, and implementation, which are illustrated in Figure 1.

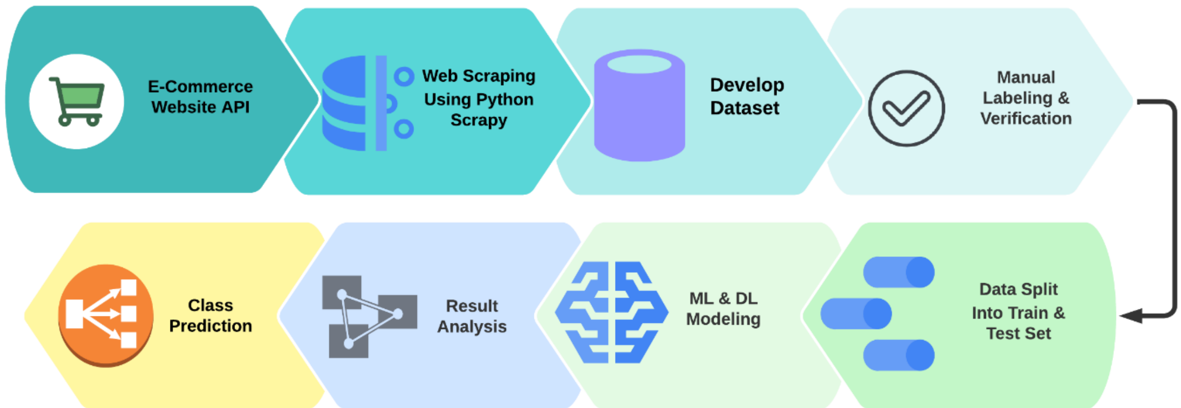


Figure 1. Workflow of Text Classification

2.1.Dataset Preparation

The preparation of a dataset for Natural Language Processing (NLP) typically involves several stages, including data collection and curation, classification into distinct categories, pre-processing and cleaning, as well as splitting the data into training, annotation, balancing, validation, and test sets. This process entails organizing and cleaning the data to make it suitable for use in NLP tasks. The subsequent step involves selecting an appropriate algorithm for the task at hand. The model's performance is evaluated using statistical analysis, and the final step involves deploying the model and monitoring its performance in real-world settings. Notably, the dataset preparation process involves two key stages which are Data Collection and Preprocessing of the Data.

2.2. Data Collection

We have assembled a dataset consisting of 1995 reviews in Bengali, along with their corresponding English translations and class labels. The dataset was collected from the Daraz E-commerce site using web scraping techniques. The reviews in the dataset encompass both positive and negative comments and have been carefully curated specifically for the purposes of this study.

Table 1: Example of collected texts for sentiment analysis in Bangla & English

Bangla Review	English Review (Translated)	Sentiment
আমার জীবনে দেখা সবচাইতে খারাপ পণ্য	The worst product I've ever seen	Negative
ভালো লাগলো আন্তর্জাতিক ব্র্যান্ড গুলো এখন বাংলাদেশেই পাওয়া যাচ্ছে	Good international brands are now available in Bangladesh	Positive
মানুষকে না ঠকিয়ে ভাল প্রোডাক্ট দেওয়ার ব্যবস্থা করেন	Provide good products without deceiving people	Negative
দাম অনুযায়ী এই পণ্য পাওয়া ভাগ্যের ব্যাপার।	It's a matter of luck to get this product by price	Positive
এতো বাজে কমেন্টস দেখে কেনার সাহস হারিয়ে ফেলেছি	I lost the courage to see so many bad comments	Negative

Following data collection, the dataset underwent manual annotation to ensure the accuracy and consistency of class labels. Subsequently, the data was preprocessed to remove irrelevant information, correct errors, and ensure compatibility with the Natural Language Processing (NLP) model. This preprocessing step is essential for improving the quality of the dataset and enhancing the performance of the NLP model during training and evaluation. As depicted in Figure 2, the dataset for English and Bangla text contained a total of 1091 positive comments and 904 negative comments.

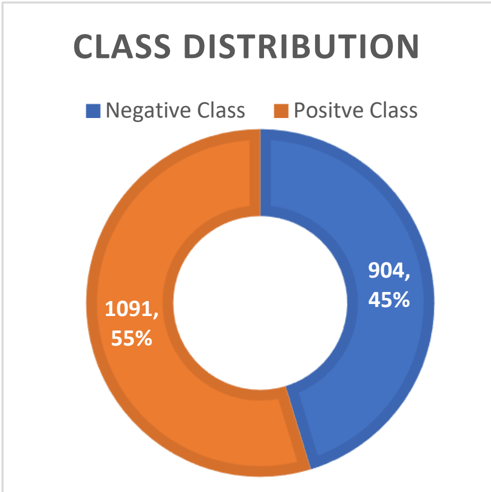


Figure 2: Dataset class distribution overview (English and Bangla both)

2.3. Data Preprocessing

Data preprocessing is a crucial step in preparing data for Natural Language Processing tasks. In NLP, data preprocessing involves the process of cleaning text to remove irrelevant information, correcting errors, tokenizing text into words, phrases, or sentences, reducing words to their base form through stemming or lemmatization, identifying and removing any non-representative data points, and normalizing data to ensure consistency. These steps aid in enhancing the accuracy and efficiency of NLP algorithms. In this study, we present our contribution to preprocessing English and Bangla comments data, as illustrated in Figure 3. The preprocessing steps involved in our study include cleaning the text to remove irrelevant information and errors, tokenizing the text into words and sentences, stemming the words to their base form, identifying and removing any non-representative data points, and normalizing the data to ensure consistency.

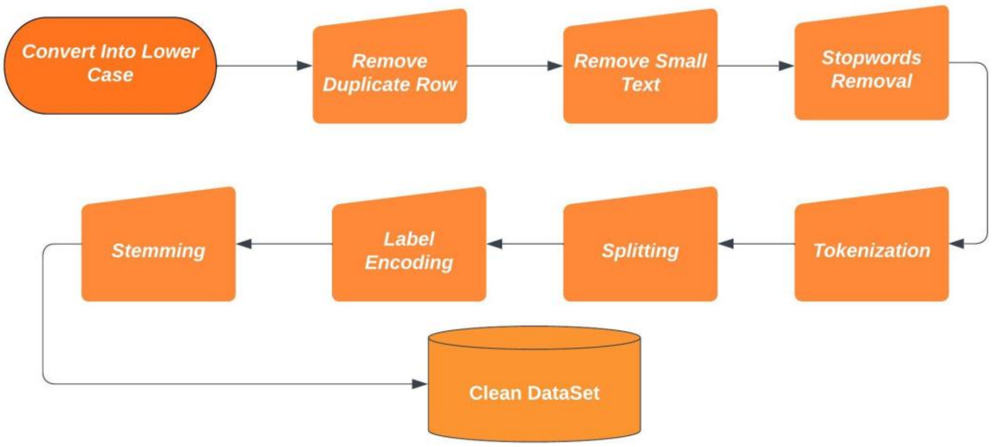


Figure 3: Data preprocessing steps

2.3.1. Convert Into Lower Case

Text normalization is a common preprocessing step in sentiment analysis for English text, whereby all characters are converted to lowercase. However, for Bangla text, this process is unnecessary. In our study, we converted all words to lowercase for the English language to simplify the data and facilitate processing and sentiment analysis by machine learning models. Capitalization variations in sentiment analysis datasets can hinder accurate sentiment identification by algorithms. Lowercasing eliminates such variations and generates a more consistent dataset. Furthermore, reducing the number of unique tokens through this process leads to increased computational efficiency in analysis and model training, resulting in more effective sentiment analysis. By reducing the variations in text and improving the consistency of the data, preprocessing steps such as lowercasing enable machine learning models to more accurately identify and classify sentiment.

2.3.2. Removing Duplicate Rows

The removal of duplicate rows is a crucial step in data preprocessing for NLP tasks since they can lead to inaccuracies in models and increased processing time. In this study, we removed duplicate rows from the English and Bangla datasets. This process involves identifying the duplicate rows, selecting a strategy for removing them, such as keeping the first instance or the one with the highest confidence score and implementing the strategy using a programming language or tool. By eliminating duplicate rows, we aimed to enhance the accuracy of NLP models and reduce processing time. This step contributes to the quality of the dataset and the overall effectiveness of sentiment analysis. Therefore, we removed duplicate rows for both English and Bangla datasets to ensure the reliability of the results obtained from our analysis.

2.3.3. Removing Small Texts

Filtering small texts, or texts that are below a certain length is an important step in preprocessing data for NLP tasks. Such texts, including individual words or brief sentences, may not carry enough information and thus can be removed to improve the quality of the dataset. The process involves determining the minimum length of texts to be kept in the dataset, identifying texts below the threshold, and utilizing a programming language or software tool to remove them. This approach has the potential to enhance the precision of NLP models, and the quality of the dataset can be enhanced. In our study, we determined the length of the sentence for both English and Bangla. We then provided the minimum length of texts to be kept in the dataset. After cleaning small texts, we removed one small conversation for English sentences and one small conversation for Bangla sentences.

2.3.4. Stopwords Removal

Stopwords removal is a frequently used preprocessing technique in NLP tasks. Stopwords are words such as "the," "a," and "and" that do not carry much meaning and can be removed to simplify the dataset and speed up processing. The process involves identifying stop words, determining a strategy for removal such as removing all or only the most frequently occurring, and using a programming language or software tool to execute the strategy. This helps to make the dataset smaller and processing faster, potentially enhancing the accuracy of NLP models. For the English dataset, we cleaned words like - {“about”, “don’t”, “across”, “after”, “again”, “your”, “me”, “not”} etc. For Bangla dataset, we cleaned word like - {“এই”, “কোন”, “আমি”, “আপনার”, “যা”, “যে”, “মনে”, “করি”} etc.

2.3.5. Tokenization

NLP tokenization is an essential process in natural language processing (NLP) that involves dividing text into individual words or phrases known as tokens. The tokenization process is a critical step in preprocessing data for NLP tasks such as sentiment analysis, text classification, and machine translation. Tokenization helps to standardize the text data by breaking it down into smaller units that are easier to process and analyze. This process often involves removing punctuation, special characters, and other non-text elements from the data, which helps improve the accuracy and effectiveness of NLP models. Tokenization can be done using various techniques, including rule-based methods, Both statistical methods and deep learning-based techniques are employed. The method selected is dependent on the particular NLP task, at hand, The decision regarding the choice of method is influenced by the nature of the language being analyzed and the characteristics of the text data. Overall, NLP tokenization is an essential aspect of text preprocessing that plays a crucial role in improving the accuracy and efficiency of NLP models.

2.3.6. Punctuation, Special Character and Number Removal

Punctuation like; ।, . , ? !, etc. and special character like @, #, $, %, ^, &, *, etc. and numbers that are not important for sentiment analysis are removed from the whole Dataset.

2.3.7. Splitting

Splitting is a fundamental technique in Natural Language Processing (NLP) that involves dividing text into smaller, meaningful units for analysis. It is a critical preprocessing step in NLP tasks as it enables the processing of text data by machine learning models. Text splitting can be performed at various levels, such as the word, sentence, or paragraph level. At the word level, the text is divided into individual words, also known as tokens, by a process called tokenization.

2.3.8. Categorical Encoding

Categorical encoding is a process in NLP where categorical variables are converted into numerical values that can be used as inputs for machine learning models. Categorical variables are those that have a limited number of possible values, such as words in a vocabulary or items on a menu. There are two types of categorical encoding entitled label encoding and embedding. Label encoding is a way of converting categorical data into numerical format by assigning integer values to each unique category. In our study, label encoding provides a means of transforming categorical data, which cannot be processed directly, into a numerical form. Label encoding assigns each category a unique integer value. The integer values are assigned without implying any ranking or order between the categories. On the other hand, Embedding is a method applied in NLP and machine learning to convert data into dense, low-dimensional vectors. For NLP, words or phrases are transformed into word embeddings, which are high-dimensional vectors that depict their semantic significance and the relationships between words. Word embedding methods rely on the principle that words with similar meanings tend to be used in similar contexts. For example, in our English comment dataset, “good” and “best” would likely be found in similar contexts and thus have similar meanings. For example, in our Bangla comment dataset, "ইয়ারফোন", "হেডসেট" and "হেডফোন" would likely be found in similar contexts and thus have similar meanings also. The word embeddings of these words would be alike, reflecting the relationships between them. Embedding enhances NLP model performance by encapsulating word meaning and word relationships in a dense, low-dimensional format.

2.3.9. Stemming

Stemming is a technique used in NLP to simplify words to their base form. The objective of stemming is to reduce the complexity of the text dataset and make analyzing it easier. For English texts, for example, words like "recommended”, "recommends", and "recommendation" would be transformed to "recommend." For Bangla texts, for example, words like “করেছিলাম”, “করছে” would be transformed to “করেছি”. The process of stemming involves the following steps: extracting the words from the text, utilizing a stemming algorithm, and saving the stemmed words in a new data structure for further analysis. Although stemming can help improve the accuracy of NLP models, it can also lead to a loss of information and reduced interpretability.

2.4. Dataset Summary

Our summary entails various details such as the total number of comments, the distribution of data across different classes, the number of words, the number of unique words, most frequent words, average length, maximum length, and a minimum length of comments. Understanding the characteristics of the dataset is crucial for selecting appropriate models and techniques for an NLP task and comprehending and evaluating the results of model evaluations. Figures 4 and 5 depict the length-frequency distribution for English and Bangla texts for the length of the word, providing a clear visual representation. Furthermore, Figures 6 and 7 provide a clear visual representation of the length-frequency distribution for English and Bangla texts for the length of the character.

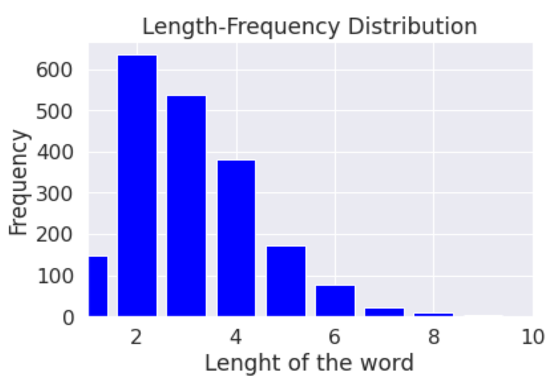


Figure 4. Length of the word for English

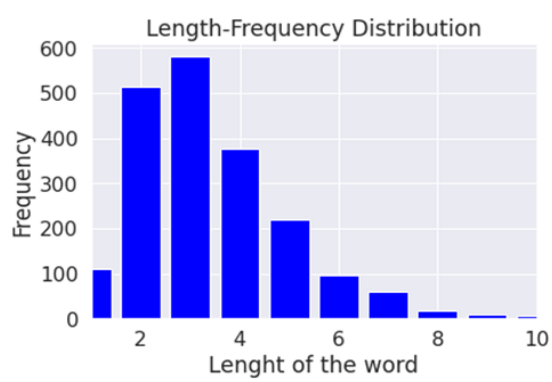


Figure 5. Length of the word for Bangla

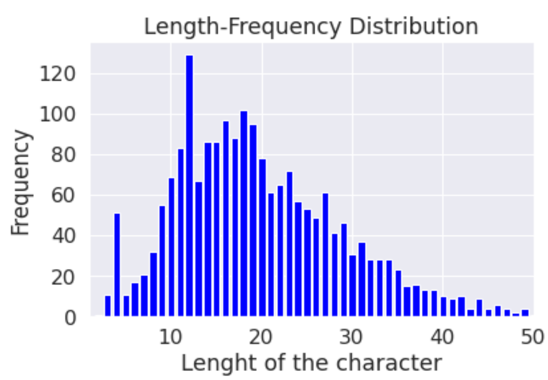


Figure 6. Length of character for English texts

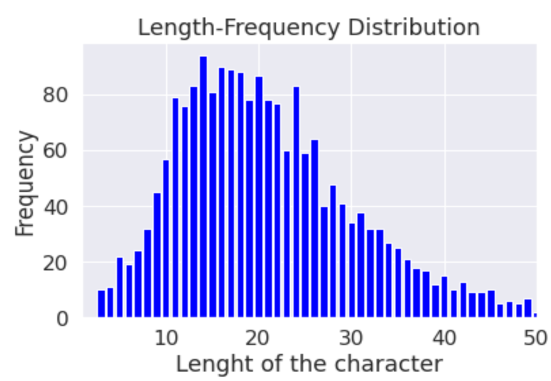


Figure 7. Length of character for Bangla texts

2.5. Machine Learning Algorithms and Statistical Analysis

The dataset underwent division utilizing the train-test split technique, whereby the larger portion of the data (80%) was designated for training the model, with the remaining 20% utilized for testing. This procedure was applied to both English and Bangla texts. The approach is commonly employed in machine learning to evaluate the performance of models on unseen data. The training data is employed to train the model, while the test data is used to measure the model's accuracy in predicting outcomes for new data. Such an approach aids in detecting overfitting, which is a common problem in machine learning, and enables the model to generalize well to unseen data. In total, 1995 reviews were included in this study, of which 1596 were allocated for training, and the remaining 399 reviews were used for testing. The study employed several supervised machine learning algorithms, including Support Vector Machine, Multinomial Naive Bayes, K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, and Stochastic Gradient.

2.5.1. Feature Extraction

In the field of natural language processing, machine learning techniques are used for accomplishing diverse goals. One such technique is tokenization, which involves breaking down phrases into individual word components. These components, whether common or unique, are then analyzed for specific characteristics. Another crucial technique is TF-IDF, a numerical metric that evaluates the importance of specific terms within a text. This approach has been widely utilized by reputable publications in multiple languages and has been proven effective. Our study was inspired by these successful methods, and we have discovered that our learning algorithms attain high accuracy when employing them.

2.5.2. Data Vectorization or Distribution

The Count Vectorizer, a beneficial tool offered by the Python Scikit-learn library, can transform a sentence into a vector by considering the frequency of each word throughout the text. The size of the n-grams used can be specified using the ngram_range parameter. For example, a value of 1, 1 would result in unigrams (n-grams made up of a single word), while a value of 1-3 would result in n-grams made up of one to three words.

- Unigram: By passing a value of n=1 to the n-grams function, unigrams or 1-grams can be produced, and the word frequency of the words can also be calculated.
- Bigram: By passing a value of n=2 to the n-grams function, bigrams or 2-grams can be produced, and the word frequency of the words can also be calculated.
- Trigram: By passing a value of n=3 to the n-grams function, trigrams or 3-grams can be produced, and the word frequency of the words can also be calculated.

Table 2. Example of n-gram distribution for English

Sentence	Uni-gram	Bi-gram	Tri-grams
This is the best phone in the budget.	('the', 2), ('this', 1), ('is', 1)	('this is’, 1), ('is the', 1), ('the best', 1)	('this is the', 1), ('is the best’, 1), ('the best phone', 1)

Table 3. Example of n-gram distribution for Bangla

Sentence	Uni-gram	Bi-gram	Tri-grams
আমি মনে করি আমি আমার টাকা অপচয়(I think I wasted my money)	('আম', 3), ('মন', 1), ('কর', 1)	('আম মন', 1), ('মন কর', 1), ('কর আম', 1)	('আম মন কর', 1), ('মন কর আম', 1), ('কর আম আম', 1)

2.6. DNN-Based Models

DNN stands for Deep Neural Network, it's a type of artificial neural network that is composed of multiple layers, called hidden layers, in addition to input and output layers. These hidden layers allow the network to learn and represent complex patterns and relationships in the data, making

DNNs particularly useful for tasks such as image and speech recognition, natural language processing, and other tasks that involve high-dimensional data. In the present study, we designed and implemented four unique neural network models, namely Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Convolutional Neural Network with one-dimensional convolutional layers (Conv1D), and a hybrid architecture comprising Conv1D and LSTM layers (Conv1D-LSTM).

Table 4. Experimental setup of four Deep Neural Network (DNN) models.

Model Name	Embedding Layer	Conv1D Layer	MaxPooling1D Layer	LSTM Layer	Bi-LSTM Layer	Fully Connected Layer	Dropout Layer	Classification Layer	Batch Size	Epoch
LSTM Based Model	64	N/A	N/A	Layer: 1 Unit: 64	N/A	Layer: 1 Unit: 256	Layer: 1 (40%)	Softmax	32	150
Bi-LSTM Based Model	60	Layer: 1 Unit: 100	Layer: 2 Pooling Size: 2	N/A	Layer: 1 Unit: 100 Layer: 2 Unit: 200	Layer: 3 Unit: 16	Layer: 1 (30%)	Softmax	64	50
Conv1D Based Model (CNN)	64	Layer: 1 Unit: 50	N/A	N/A	N/A	Layer: 1 Unit: 100	Layer: 1 (20%)	Softmax	32	150
Combine Conv1D & LSTM Based Model	128	Layer: 1 Unit: 50	Layer: 1 Pooling Size: 2	Layer: 1 Unit: 64(L1) 64(L2)	N/A	Layer: 1 Unit: 28	Layer: 1 (20%)	Softmax	64	150

2.6.1. LSTM (Long Short-Term Memory)

This is a type of DNN that is particularly well-suited for sequential data, such as time series data, speech, and text. LSTM networks are composed of LSTM cells, which are designed to remember information for a long period of time and to selectively forget irrelevant information. This allows LSTM networks to learn patterns in the data that span multiple time steps, making them useful for tasks such as language modeling and speech recognition.

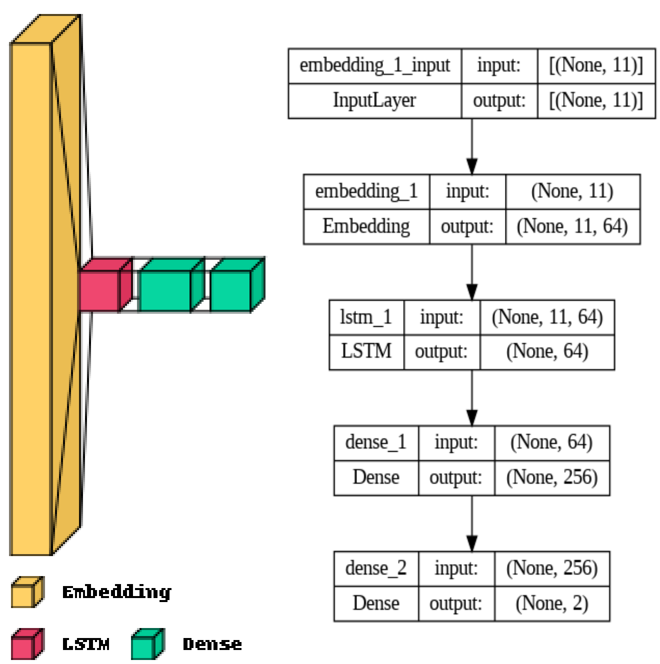


Figure 8. LSTM-based Model

2.6.2. Bi-LSTM (Bidirectional Long Short-Term Memory)

Bi-LSTM is a recurrent neural network that uses two separate LSTM models - one for processing the input sequence forward, and another for processing it backward. Unlike standard LSTM models that process input sequences in one direction, Bi-LSTM models consider both past and future context. This quality makes them a suitable choice for tasks like speech recognition, natural language processing, and image captioning that need a complete understanding of the input sequence. Bi-LSTM models have proven to be effective for many sequence modeling tasks and are widely used in deep learning.

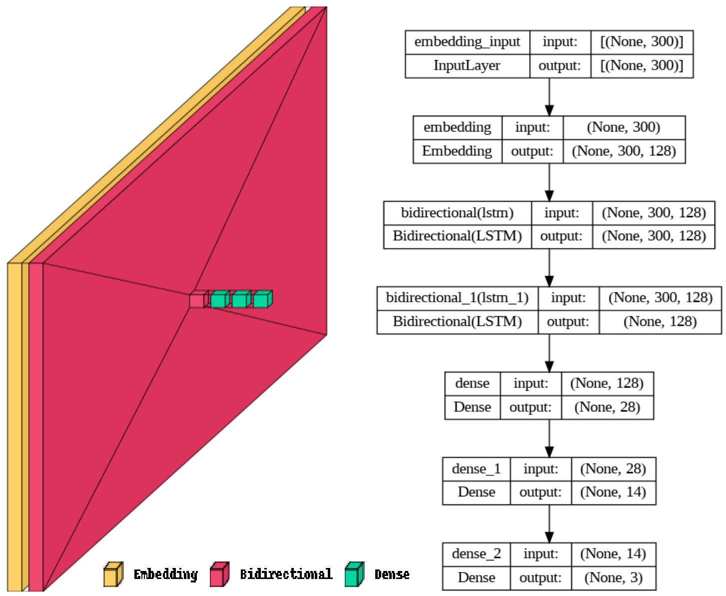


Figure 9. Bi-LSTM Based Model

2.6.3. CNN (Convolutional Neural Network)

Convolutional Neural Networks (CNNs) are a type of neural network commonly used in image recognition and computer vision tasks. They consist of multiple layers of interconnected nodes, where each node performs a convolution operation on a subset of the input data. The output of each layer is then fed into the next layer, allowing the network to learn increasingly complex features of the input data. The convolution operation involves sliding a filter over the input data and computing the dot product between the filter weights and the corresponding input values. This operation is typically followed by a non-linear activation function, such as the Rectified Linear Unit (ReLU), to introduce non-linearity into the network.

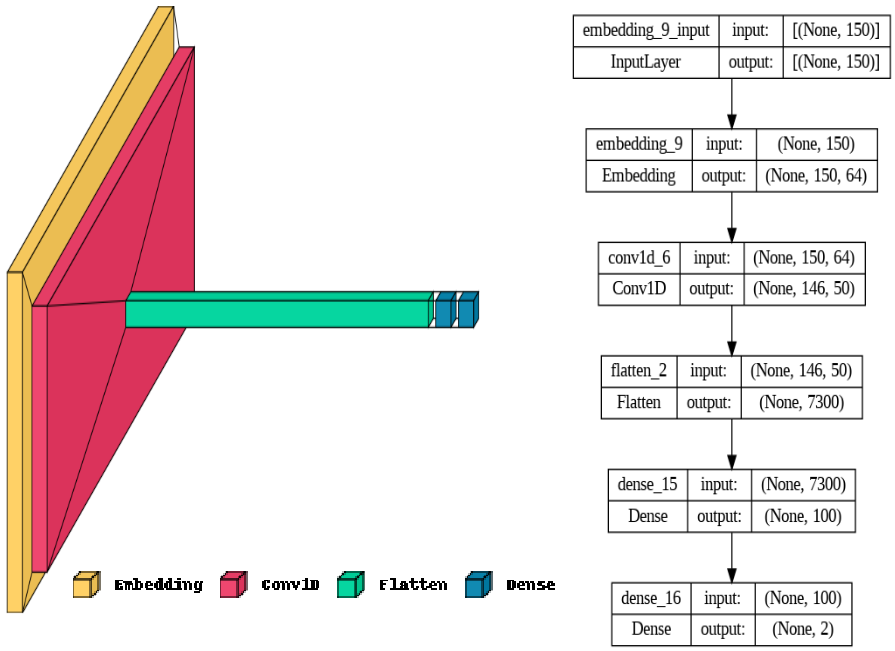


Figure 10. Conv1D-based model (CNN)

2.6.4. Hybrid Conv1D-LSTM

Both CNNs and LSTMs are often used in combination with other types of DNNs to improve performance on various tasks. The architecture mentioned is commonly employed in tasks that involve analyzing sequential data, such as natural language processing and speech recognition. The Conv1D layers in this model perform feature extraction on the input sequence, capturing local patterns and relationships within the data. The LSTM layers, on the other hand, are capable of capturing long-term dependencies in sequential data by incorporating a memory component that can selectively retain or discard information over time. Overall, the Hybrid Conv1D-LSTM model is a powerful architecture for sequential data analysis tasks that require both local and global feature extraction and modeling.

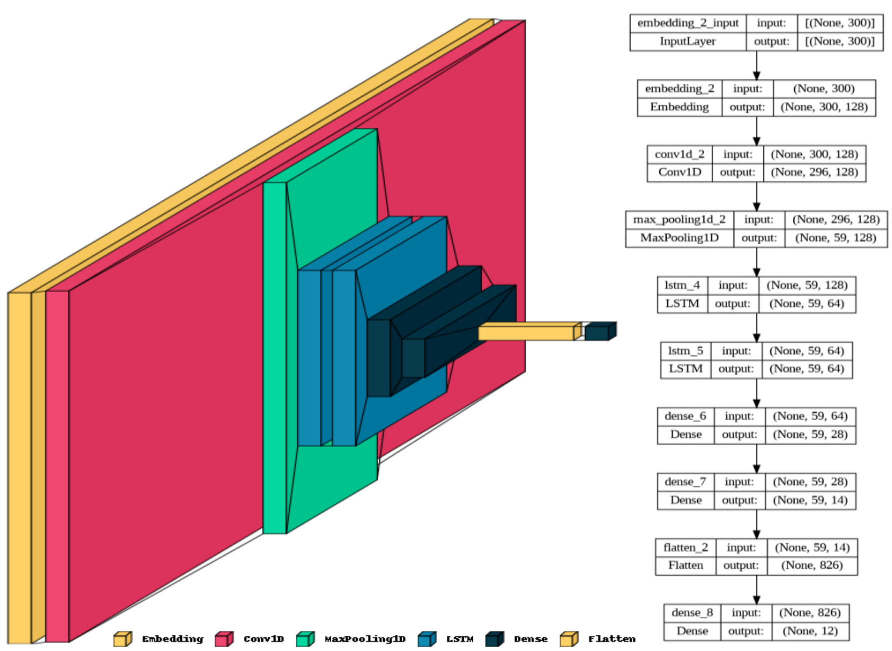


Figure 11. Combined Conv1D and LSTM-based model

2.7. Experimental Setup

The present experiment was conducted utilizing the Google Colab platform for the purpose of training conventional machine learning as well as deep learning models. This platform offers unrestricted access to high-performance Graphics Processing Units (GPUs) with minimal setup requirements. In traditional machine learning models, a train and test set are utilized to evaluate the model's performance. In contrast, for deep learning models, a train, test, and validation set are commonly employed to assess the model's generalizability. The incorporation of a validation set in deep learning models enables the monitoring of the model's performance during the training phase, preventing overfitting, and enhancing the model's generalization ability. The validation set is used to assess the model's performance on new and unseen data, thereby providing a means of determining the model's ability to generalize beyond the training data. Figure 12 facilitates a comprehensive understanding of the model evaluation process.

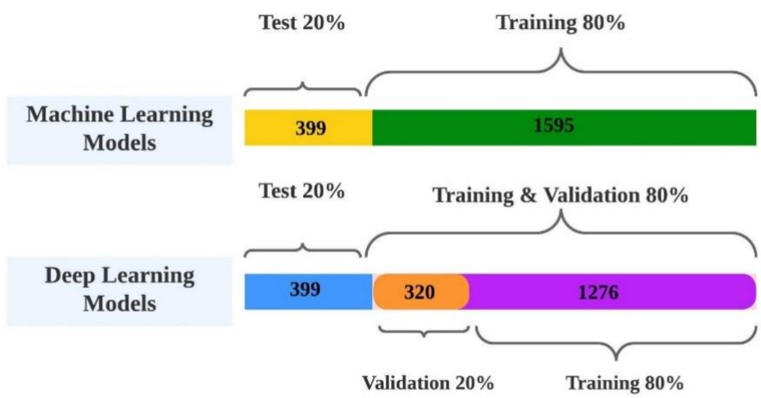


Figure 12. Data Distribution

3. RESULT AND DISCUSSION

The Results and Discussion section provides an overview of the experimental findings and their interpretation. This section presents a detailed analysis of the data, Metrics like accuracy, precision, recall, and F1-score are examples of performance evaluation measures that are frequently used. It also discusses the impact of the different parameters and hyperparameters on the model's performance. Additionally, the section compares the results obtained from the traditional machine learning models with those obtained from the deep learning models. The findings are discussed in light of the current literature, highlighting the strengths and weaknesses of the proposed approach. Overall, the Results and Discussion section provides a comprehensive analysis of the experimental outcomes, demonstrating the efficacy and limitations of the proposed models and providing insights for future research in the field.

Table. 5. Performance score of ML models for English text

Model Name	Accuracy	Precision	Recall	F1 Score
Logistic Regression	79.95%	77.69%	87.85%	82.46%
Decision Tree	78.45%	82.65%	75.70%	79.02%
Random Forest	78.70%	82.09%	77.10%	79.52%
Multi Naïve Bayes	80.70%	80.18%	85.05%	82.54%

KNN	77.69%	77.29%	82.71%	79.91%
SVM	80.70%	82.16%	81.78%	81.97%
SGD	78.95%	80.95%	79.44%	80.19%

Table. 6. Performance score of ML model for Bangla text

Model Name	Accuracy	Precision	Recall	F1 Score
Logistic Regression	84.96%	80.83%	93.27%	86.61%
Decision Tree	76.69%	74.68%	83.65%	78.91%
Random Forest	79.20%	72.89%	95.67%	82.74%
Multi Naïve Bayes	84.46%	82.59%	88.94%	86.65%
KNN	75.19%	69.96%	91.83%	79.42%
SVM	87.47%	86.92%	89.42%	88.15%
SGD	84.71%	83.26%	88.46%	85.78%

The Logistic Regression Classifier, a supervised method used to establish a relationship between dependent and independent variables, demonstrated an accuracy of 79.95% and 84.96% for English and Bangla texts, respectively. The Multi Naïve Bayes Classifier yielded an accuracy of 80.70% and 84.46% for English and Bangla texts, respectively. The Decision Tree Classifier, commonly employed for classification and regression problems, achieved a prediction accuracy of 78.45% and 76.69% for English and Bangla texts, respectively. Furthermore, the Random Forest Classifier, comprising multiple decision trees, proved effective in handling high-dimensional data and delivered an accuracy of 78.70% and 79.20% for English and Bangla texts, respectively. The K-Nearest Neighbor (KNN) Classifier, utilized for regression and text processing, classifies new data based on its similarity to existing data and assigns it to the most similar category. Our KNN model yielded an accuracy of 77.69% and 75.19% for English and Bangla texts, respectively. The Support Vector Machine (SVM) Classifier, a powerful algorithm that produces accurate results even with minimal training data, classifies data by drawing a hyperplane that separates the categories. Our dataset achieved an accuracy of 80.70% and 87.47% for English and Bangla texts, respectively, utilizing the SVM method. Finally, the Stochastic Gradient Descent (SGD) Classifier, an optimization procedure commonly used in machine learning, enabled the identification of model parameters that best matched the expected and actual outputs. Our method using SGD achieved an accuracy of 78.95% and 84.71% for English and Bangla texts, respectively.

The study evaluated the performance of various machine learning algorithms, including Logistic Regression, Multi Naïve Bayes, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD), in classifying English and Bangla texts. The results showed that SVM was the most accurate method for Bangla texts. However, all algorithms yielded relatively high accuracy rates for both languages, indicating their suitability for text classification tasks.

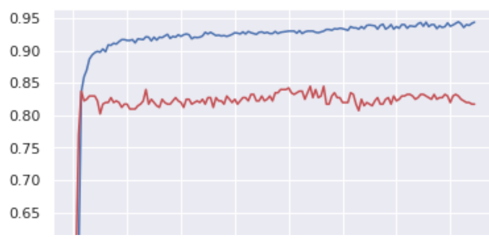
Table. 7. Performance score of DL models for English text

Model Name	Accuracy	Precision	Recall	F1 Score
LSTM Based Model	81.70%	77.30%	82.18%	79.67%
Bi-LSTM Based Model	77.44%	77.60%	74.35%	75.94%
Conv1D Based Model	68.42%	68.03%	82.06%	74.39%
Conv1D–LSTM Based Model	74.69%	71.84%	77.49%	74.56%

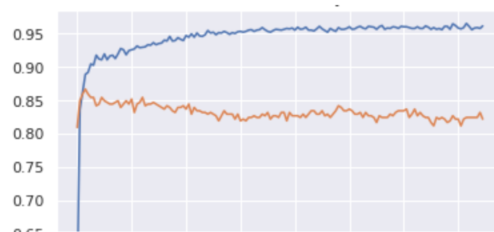
Table. 8. Performance score of DL models for Bangla text

Model Name	Accuracy	Precision	Recall	F1 Score
LSTM Based Model	82.21%	77.25%	83.91%	80.44%
Bi-LSTM Based Model	78.70%	78.80%	75.92%	77.33%
Conv1D Based Model	67.92%	67.06%	78.97%	72.53%
Conv1D–LSTM Based Model	76.69%	76.34%	74.35%	75.33%

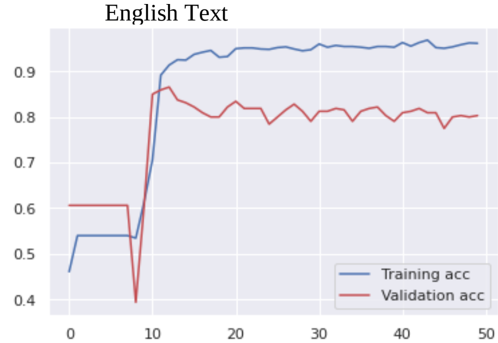
The performance of deep learning models in natural language processing tasks varies across different languages and contexts. This research project assessed the accuracy of four deep learning models in classifying text data from two distinct languages, namely English and Bangla. Specifically, we compared the performance of LSTM-based, bi-LSTM-based, Conv1D-based, and Conv1D-LSTM-based models listed in Table7 & 8. The results of our study indicated that the LSTM-based model demonstrated the highest level of accuracy in classifying text data from both English and Bangla languages, with 81.70% and 82.21%, respectively. The bi-LSTM-based model also performed well, with accuracy scores of 77.44% and 78.70% for English and Bangla, respectively. The Conv1D-based model had the lowest accuracy scores, with 68.42% and 67.92% for English and Bangla, respectively. The Conv1D-LSTM-based model had intermediate accuracy scores, with 74.69% and 76.69% for English and Bangla, respectively. Overall, our findings suggest that the LSTM-based and bi-LSTM-based models are effective for text classification in both English and Bangla, while the Conv1D-based model may not be the best choice for this task. Future studies can further explore the performance of these models in other languages and contexts.



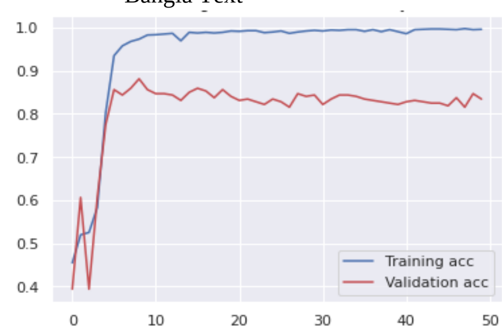
(b) ACC curve of LSTM-based network for English Text



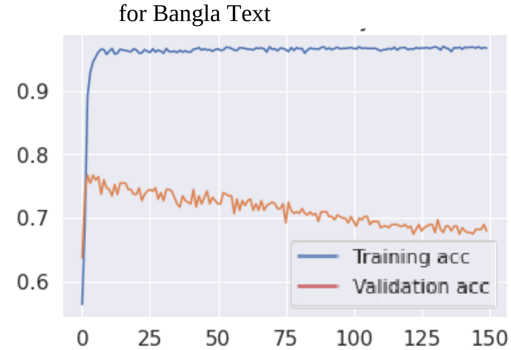
(a) ACC curve of LSTM-based network for Bangla Text



(c) ACC curve of Bi-LSTM-based network for English Text

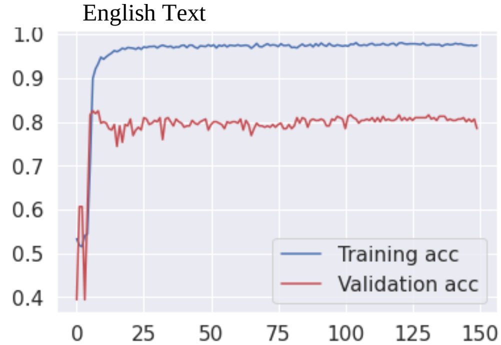


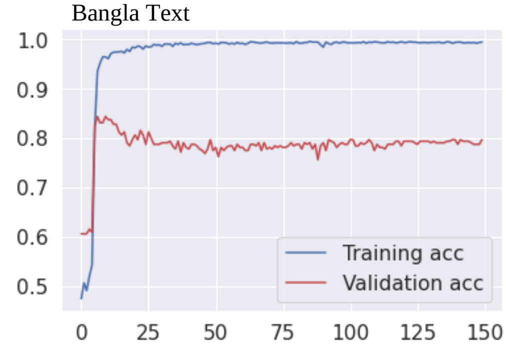
(d) ACC curve of Bi-LSTM-based network for Bangla Text



(f) ACC curve of Conv1D-Based network for English Text

(e) ACC curve of Conv1D-Based network for Bangla Text



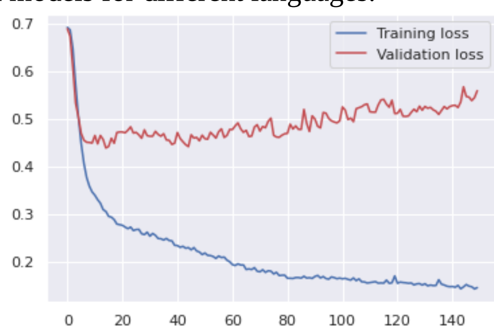


(h) ACC curve of Conv1D-LSTM Based network for English Text

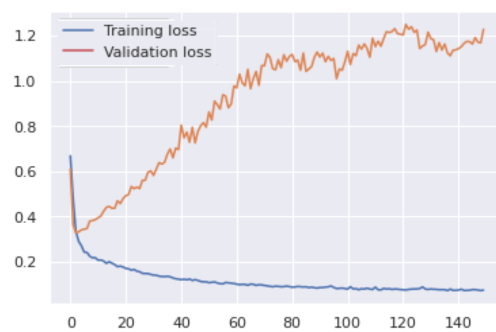
(g) ACC curve of Conv1D-LSTM Based network for Bangla Text

Figure 13. (a–h) Graphical illustration of epochs versus accuracy of each network for training and validation dataset.

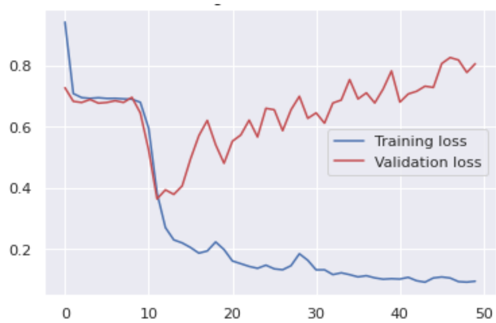
Graphical comparisons of four distinct models, including the LSTM-Based Model, Bi-LSTM Based Model, Conv1D Based Model, and combined Conv1D-LSTM Based Model, for English and Bangla text classification are presented in Figure 13(a-h). The relationship between the accuracy of each model on both training and validation datasets and the number of training epochs is illustrated in the figure. The results indicate that the accuracy of the models generally increases with the number of training epochs until it reaches a plateau, indicating that additional epochs do not significantly improve the accuracy of the models. Additionally, the figure shows that the rate of accuracy improvement varies across the different models and languages. Furthermore, the figure also highlights the general trend of lower accuracy on the validation dataset compared to the training dataset, indicating a possible overfitting of the models to the training data. Its provides valuable insights into the performance of different neural network models for Bangla and English text classification. The results underscore the importance of carefully selecting appropriate models and balancing the number of training epochs to achieve optimal accuracy on both training and validation datasets. The findings of this study can inform the development of more accurate and robust text classification models for different languages.



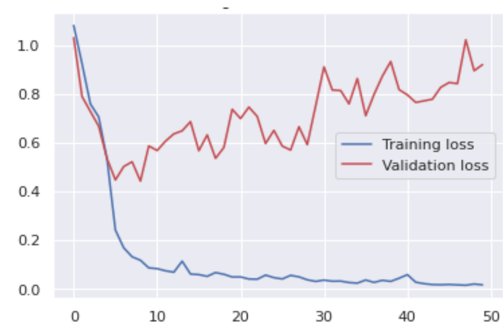
(a) Loss curve of LSTM-based network for English Text



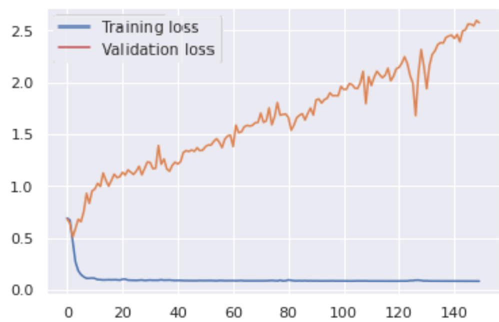
(b) Loss curve of LSTM-based network for Bangla Text



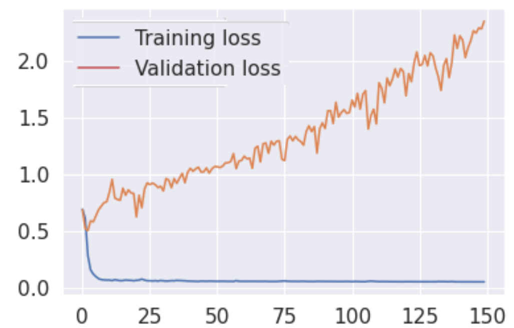
(d) Loss curve of Bi-LSTM-based network for English Text



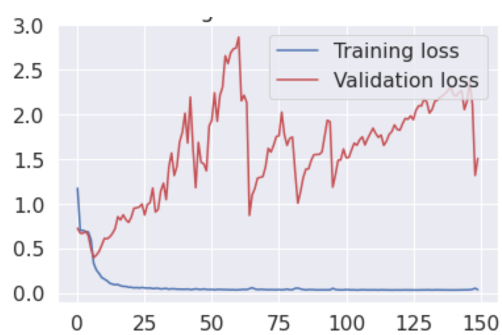
(c) Loss curve of Bi-LSTM-based network for Bangla Text



(f) Loss curve of Conv1D-Based network for English Text



(e) Loss curve of Conv1D-Based network for Bangla Text



(h) Loss curve of Conv1D-LSTM Based network for English Text

(g) Loss curve of Conv1D-LSTM Based network for Bangla Text

Figure 14. (a–h) Graphical illustration of epochs versus loss of each network for training and validation dataset.

Figure 14 presents a graphical illustration of the relationship between the number of training epochs and the loss of four different neural network models for both the training and validation datasets. The figure shows how the loss of the models changes with the number of epochs during the training process. Specifically, the figure indicates that the loss of the models generally decreases with the number of training epochs until it reaches a plateau, indicating that additional epochs do not significantly improve the loss of the models. Additionally, the figure reveals that the loss of the models on the validation dataset is generally higher than that on the training dataset, which suggests the presence of overfitting. This observation highlights the importance of using appropriate validation strategies to evaluate the generalization ability of the models.

4. CONCLUSION

In this paper, we presented an overview of machine learning and deep learning models applied to sentiment analysis in e-commerce data. We proposed a novel model based on multi-level dilated convolution, which can extract both semantic-unit-level information and word-level information. Our experiments compared seven machine learning techniques for sentiment analysis in English and Bangla texts and generated Conv1D multi-layer deep learning models and Conv1D-LSTM combined models for text classification. Our findings indicate that the suggested model is superior to the other techniques in regards to both accuracy and efficiency. However, there is still room for improvement, especially if the dataset is increased for both English and Bangla texts. We believe that LSTM-based models and combinations of Conv1D-LSTM models may lead to even better results. Looking to the future, we plan to collect additional data and explore other innovative machine learning and deep learning algorithms to develop more accurate and efficient models for sentiment analysis in e-commerce data. Our research provides a foundation for further studies in this field, which may have significant implications for enhancing the quality of sentiment analysis in various applications.

References:

1. Sarvjeet Kaur Chatrath, G.S. Batra, Yogesh Chaba, “Handling consumer vulnerability in e-commerce product images using machine learning”, HELIYON VOLUME 8, ISSUE 9, E10743, SEPTEMBER 01, 2022.
2. Balakrishnan, V., Shi, Z., Law, C.L. et al. A deep learning approach in predicting products’ sentiment ratings: a comparative analysis. J Supercomput 78, 7206–7226 (2022).
3. Bhowmik, Nitish Ranjan, Mohammad Arifuzzaman, and M. Rubaiyat Hossain Mondal. "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms." Array 13 (2022): 100123.
4. Alam, Tanvirul, Akib Khan, and Firoj Alam. "Bangla text classification using transformers." arXiv preprint arXiv:2011.04446 (2020).
5. Dreisbach, Caitlin, Theresa A. Koleck, Philip E. Bourne, and Suzanne Bakken. "A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data." International journal of medical informatics 125 (2019): 37-46.
6. Ferreira-Mello, Rafael, Máverick André, Anderson Pinheiro, Evandro Costa, and Cristobal Romero. "Text mining in education." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 9, no. 6 (2019): e1332.
7. Moqsadur Rahman, Summit Haque, Zillur Rahman Saurav, “Identifying and Categorizing Opinions Expressed in Bangla Sentences using Deep Learning Technique”, International Journal of Computer Applications (0975 – 8887), Volume 176 – No. 17, April 2020.
8. M. T. Ahmed, M. Rahman, S. Nur, A. Islam, and D. Das, “Deployment of machine learning and deep learning algorithms in detecting cyberbullying in Bangla and romanized bangla text: A comparative study,” 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2021.
9. I. H. Sarker, M. H. Furhad, and R. Nowrozy, “Ai-Driven Cybersecurity: An overview, security intelligence modeling and research directions,” SN Computer Science, vol. 2, no. 3, 2021.
10. Mitchell, Tom M., and Tom M. Mitchell. Machine learning. Vol. 1. No. 9. New York: McGraw-hill, 1997.
11. Sebastiani, Fabrizio. "Machine learning in automated text categorization." ACM computing surveys (CSUR) 34.1 (2002): 1-47.

12. Ming, C., Viassolo, V., Probst-Hensch, N. et al. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Res 21, 75 (2019).
13. Aggarwal, 1st Lt Pushkar. "Data augmentation in dermatology image recognition using machine learning." Skin Research and Technology 25.6 (2019): 815-820.
14. Jung, Young Hoon, et al. "Flexible piezoelectric acoustic sensors and machine learning for speech processing." Advanced Materials 32.35 (2020): 1904020.
15. Richens, Jonathan G., Ciarán M. Lee, and Saurabh Johri. "Improving the accuracy of medical diagnosis with causal machine learning." Nature communications 11.1 (2020): 1-9.
16. Fischer, Thomas Günter, Christopher Krauss, and Alexander Deinert. "Statistical arbitrage in cryptocurrency markets." Journal of Risk and Financial Management 12.1 (2019): 31.
17. Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PK, Khan WZ. An ensemble machine learning approach through effective feature extraction to classify fake news. Future Generation Computer Systems. 2021 Apr 1;117:47-58.
18. Balazs Harangi, Agnes Baran, Andras Hajdu, “Assisted deep learning framework for multi-class skin lesion classification considering a binary classification support”, Biomedical Signal Processing and Control, Volume 62, 2020, 102041,ISSN 1746-8094.
19. Moen, Erick, Dylan Bannon, Takamasa Kudo, William Graf, Markus Covert, and David Van Valen. "Deep learning for cellular image analysis." Nature methods 16, no. 12 (2019): 1233-1246.
20. Jia-Jhou Wu, Sue-Ting Chang, Exploring customer sentiment regarding online retail services: A topic-based approach, Journal of Retailing and Consumer Services, Volume 55, 2020, 102145, ISSN 0969-6989.
21. Y. Luan and S. Lin, “Research on text classification based on CNN and LSTM,” 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2019.
22. Daniela C.S.Z. Ribeiro, Habib Asseiss Neto, Juliana S. Lima, Débora C.S. de Assis, Kelly M. Keller, Sérgio V.A. Campos, Daniel A. Oliveira, Leorges M. Fonseca, Determination of the lactose content in low-lactose milk using Fourier-transform infrared spectroscopy (FTIR) and convolutional neural network, Heliyon, Volume 9, Issue 1, 2023.
23. Jang, Beakcheol, Myeonghwi Kim, Gaspard Harerimana, Sang-ug Kang, and Jong Wook Kim. "Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism." Applied Sciences 10, no. 17 (2020): 5841.
24. Zhang, Y.; Zheng, J.; Jiang, Y.; Huang, G.; Chen, R. A Text Sentiment Classification Modeling Method Based on Coordinated CNN-LSTM-Attention Model. Chin. J. Electron. 2019, 28, 120–126.
25. Wang, Jin, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang. "Tree-structured regional CNN-LSTM model for dimensional sentiment analysis." IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2019): 581-591.
26. Manguri, Kamaran H., Rebaz N. Ramadhan, and Pshko R. Mohammed Amin. "Twitter sentiment analysis on worldwide COVID-19 outbreaks." Kurdistan Journal of Applied Research (2020): 54-65.
27. Ardianto, Rian, Tri Rivanie, Yuris Alkhalifi, Fitra Septia Nugraha, and Windu Gata. "Sentiment analysis on E-sports for education curriculum using naive Bayes and support vector machine." Jurnal Ilmu Komputer dan Informasi 13, no. 2 (2020): 109-122.
28. Alamoodi, Abdullah Hussein, Bilal Bahaa Zaidan, Aws Alaa Zaidan, Osamah Shihab Albahri, Khalid Ibrahim Mohammed, Rami Qays Malik, Esam Motashar Almahdi et al. "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review." Expert systems with applications 167 (2021): 114155.

29. Rezaeinia, Seyed Mahdi, Rouhollah Rahmani, Ali Ghodsi, and Hadi Veisi. "Sentiment analysis based on improved pre-trained word embeddings." Expert Systems with Applications 117 (2019): 139-147.
30. Liu, G., Guo, J., 2019. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing 337, 325–338.
31. Dos Santos, C., Gatti, M., 2014. August). Deep convolutional neural networks for sentiment analysis of short texts. In: In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp.69–78.
32. K. S. Tai, R. Socher, and C. D. Manning, "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks," ACL-IJCNLP 2015 - 53rd Annu. Meet. Assoc. Comput. Linguist. 7th Int. Jt. Conf. Nat. Lang. Process. Asian Fed. Nat. Lang. Process. Proc. Conf., vol. 1, pp. 1556–1566, Feb. 2015.
33. J. Du, L. Gui, R. Xu, and Y. He, "A Convolutional Attention Model for Text Classification," in National CCF Conference on Natural Language Processing and Chinese Computing, 2017: Springer, pp. 183-195.
34. Y. Kim, "Convolutional neural networks for sentence classification," Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.
35. Chatterjee, S. (2020). Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach. International Journal of Hospitality Management, 85, Article 102356.
36. Nelson, D., Pereira, A., Oliveira, R.: Stock market's price movement prediction with LSTM neural networks. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 1419–1426(2017).
37. E. Park, "Motivations for customer revisit behavior in online review comments: Analyzing the role of user experience using big data approaches," J. Retail. Consum. Serv., vol. 51, no. May, pp. 14–18, 2019.
38. C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM Neural Network for Text Classification," arXiv:1511.08630 2015.
39. M. Alhawarat and A. O. Aseeri, "A Superior Arabic Text Categorization Deep Model (SATCDM)," vol. 8, pp. 24653–24661, 2020.
40. R. R. Chowdhury, M. S. Hossain, S. Hossain, and K. Andersson, "Analyzing Sentiment of Movie Reviews in Bangla by Applying Machine Learning Techniques," pp. 27–28, 2019.
41. Chakraborty, I., Kim, M., & Sudhir, K. (2021). Attribute sentiment scoring with online text reviews: Accounting for language structure and attribute self-selection. Cowles Foundation Discussion Paper, No. 2176R2.
42. X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in Advances in neural information processing systems, 2015, pp. 649-657.
43. Li, X., Li, Y., Yang, H., Yang, L., Liu, X.: DP-LSTM: differential privacy-inspired LSTM for stock prediction using financial news. arXiv:1912.10806 (2019).
44. S. Shovon and S. Haque, "Data Set For Sentiment Analysis On Bengali News Comments And Its Baseline Evaluation," pp. 27–28, 2019.
45. M. Rahman, S. Haque, and Z. Rahman, "Identifying and Categorizing Opinions Expressed in Bangla Sentences using Deep Learning Technique," Int. J. Comput. Appl., vol. 176, no. 17, pp. 13–17, 2020.