SENTIMENT ANALYSIS OF BENGALI TEXTUAL COMMENTS IN FIELD OF SPORTS USING DEEP LEARNING APPROACH

BY

Mumtahina Rahman Tuly

ID: 191-15-12120

AND

Mosa. Rabeya

ID: 191-15-12145

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Saiful Islam
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/internship titled **"Sentiment Analysis of Bengali Textual Comments in Field of Sports Using Deep Learning Approach"**, submitted by Mumtahina Rahman Tuly and Mosa. Rabeya, ID No: FL21D077, Student ID: 191-15-12120 and 191-15-12145 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *23 January, 2023*.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Abdus Sattar, Assistant Professor and Coordinator M.Sc, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Mr. Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

22.01.23

Mr. Saiful Islam
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Mumtahina

Mumtahina Rahman Tuly
ID: 191-15-12120
Department of CSE
Daffodil International University

Rabeya

Mosa. Rabeya
ID: 191-15-12145
Department of CSE
Daffodil International University

©Daffodil International University ii

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully. We really grateful and wish our profound our indebtedness to **Mr. Abdus Sattar**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Deep Learning" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan**, Professor and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

In recent days, people are expressing their emotions, feelings or opinions on various social platforms. In those opinions some are real and some are fake. There are a lot of discussions about sports. When their team wins a match, they celebrate this highly but when a match loses, they criticize, bullying them. And then they express them angrily to different sites, like Facebook pages, Facebook groups etc. This issue may be resolved by using natural language processing (NLP) to analyze the sentiment of the relevant comments. Here we analyze sentiment in various sports related Bangla comments. We collected almost 4061 data from various Facebook pages and groups. After collecting those data, we classified them into five different categories: neutral, happy, sad, positive and negative. We use some preprocessing techniques like removing punctuation, data cleaning, manual validation to prepare our data. In this study, we used three different familiar deep learning models to predict sentiment of our dataset. Here our models are CNN, LSTM and BiLSTM. In these three models CNN with the glove word embedding performed better than other two models, and it is 94.57%. Finally, the CNN model outperforms other models in a way that captures the sentiment of the fans' remarks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Sentiment analysis is the border part of NLP. Recently there have been huge works in this field, in both Bangla and English. There are four types of SA: Fine-grained sentiment, Emotion detection analysis, Aspect based and intent analysis. In this field, used many opinions, reviews, ratings, etc. like Twitter comments, Facebook comments, blogging sites comments, newspaper editor opinions, and other social sites [1]. Text classification is a part of SA that is an automatic system to find out whether a text phase contains objective or opinion-forming content, and it can moreover decide the text or passages sentiment subjectivity and polarity [2]. Using SA for text classification to predict the main feelings or opinions of various people. So, basically Sentiment Analysis is a natural language processing (NLP) technique used to determine whether data is happy, sad, positive, negative or neutral. It is the use of NLP, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information [3]. In our case, we used NLP to identify the different sentiments. The modern world is an age of science & technology. Nowadays, the majority of individuals are social media addicted [4]. In this modern era people can express their opinion using Facebook, Twitter etc. Bengali people are too emotional about the progress and decline of their nation. That's why they always prefer sharing opinions. Cricket is the first weakness of Bengali people. When Bangladesh loses everyone gets very angry, on the other hand when they win there is no limit of joy. There are a lot of people of different minds. They express their opinion in different ways. In Bangladesh perspective there are a lot of people who love Bangladeshi sports like cricket, football, hockey etc. If we notice, we will find that there are a lot of previous works about English text classification. A very few works we'll find with Bengali text [5]. This is one of the reasons to choose Bengali sports comment classification. For this reason, we tried to analyze the sentiment of Bengali people for Bangladesh Sports. We have collected our data from social media [6]. We collected some Bengali comments from many sports related Facebook pages. There were five types of categories of comments, these are positive, negative, happy, sad & neutral. We

use three algorithms for sports text classification. The rest of the paper is structured as: Chapter 2 mentions sports related previous work. Chapter 3 represents the research methodology and experimental design of this paper. The overall result analysis and findings of this paper are mentioned in Chapter 4. And finally, conclusion and future work of this paper presents in Chapter 6.

1.2 Motivation

The method to judge whether a text is neutral, positive, negative, happy and sad is called sentiment classification. For text analysis, Natural Language Processing (NLP) is a part of deep learning incorporated into the Sentiment Analysis (SA) system. In our work we used different pages like – sports related Facebook page and groups, Twitter etc. to rate their comments. SA is important because it has many useful applications based on reviews or opinions mining. Organizations around the world are embracing information mining from social data. The main goal of SA is to find out people's true feelings people living in our society. Everyone expresses their opinion in their own way. we can easily understand their point of view, but some are more confusing. When we collect comments from Facebook page and group, Twitter etc. the sentences produce similar results. We choose opinion to rate based on opinion in our proposed research works for many reasons. We've ever seen more articles on opinion research; There are different types of opinions but Bengali textual comments are not available. Besides, SA is a significant topic for sentiment classification analysis, that’s why we chose this topic [10].

1.3 Expected Outcome

Our project is research based. Accordingly, our main goal is to publish our article in a conference or journal. And we already did this, our paper is accepted in a conference and already published. Our proposed work is under the field of Sentiment Analysis (SA). There are a lot of fields on SA. In our work, we used text classification for people’s comments. The expected result is how editors can give feedback in separate comments. We used deep learning methods to achieve these results. We have to train a machine and respectively the device must learn for any kind of automated system. Then the main version used in the machine backend in conjunction with the website or mobile device application. Commenters can comment on all news independently and these comments we used and categorize what he/she said.

1.4 Report Layout

There are six chapters in total. Each chapter is approached from a different angle, and Each chapter has several sections that are explained in detail. Here is the content of this report:

Chapter 1

Some parts of this chapter include the following - 1.1 Discuss about Introduction part, 1.2 Discussing about the Motivations, 1.3 Expected Outcome, 1.4 Report Layout.

Chapter 2

This chapter we talk about - 2.1 Introduction, 2.2 Related Works and 2.3 Summary of the Research.

Chapter 3

In this chapter, we have described our entire workflow. There is some section in it, 3.1 Introduction, 3.2 Research Subject and Instrumentations, 3.3 Data Collection, 3.4.1 Data Preprocessing, 3.4.2 Tokenization 3.5 Statistical Analysis, 3.6 Proposed Methodology 3.6.1 Word Embedder and Optimizer 3.6.2 Convolutional Neural Network (CNN) 3.6.3 Long-Short-Time Memory (LSTM) 3.6.4 Bidirectional Long Short-Term Memory (BiLSTM) 3.7 Implementation Process and 3.8 Parameter Tuning of Models.

Chapter 4

This chapter includes experience and discussion of the results of this study, 4.1 Introduction, 4.2 Descriptive Analysis, 4.3 Experimental Analysis, 4.3 Discussion.

Chapter 5

In this chapter we discussed social impact in our society, 5.1 Impact on Society, 5.2 Ethical Aspects, 5.3 Sustainability Plan.

Chapter 6

This Section discussed about, 6.1 Summary of The Study, 6.2 Conclusion, 6.3 Implication for Further Work.

CHAPTER 2

Background Studies

2.1 Introduction

There are various text classification works had been done previously. We read it and are highly motivated to work on text classification. We began our research by reading related literature reviews. Here we listed some previous work that we read out.

2.2 Related Works

Mahtab et al. analyzed the sentiment of Bengali people's attitudes about Bangladesh Cricket. Social media platforms were used to get the dataset. Additionally, because their dataset is too limited, they also employ ABSA-based data. They classified the data using the classifier Support Vector Machine and the vectorizer Term Frequency-Inverse Document Frequency (TF-IDF). 10% of the data is used here for a random test set, while 90% is used for a machine learning model. The accuracy of the ABSA-based dataset is 73.490%, and the accuracy of the real-time dataset is 64.597%. Their current goal is to expand the dataset and the three classes they currently have. Additionally, deep learning theory will be used [7].

Wahid et al. The focus of this essay is commentary on Bangla cricket. In this investigation, sentiment analysis was used. Datasets are based on ABSA. They employed an RNN-based Deep Learning version and LSTM for prediction, and the accuracy of the predictions was 95%. The objective now is to increase classes and add more preprocessing processes for accurate NLP models [8].

Faruque et al. This study examined the polarity of Bangla language Facebook posts and the public's perception of Bangla Cricket. They employed the Naive Bayes (NB), Support Vector Machines (SVM), and Logistic Regression machine learning techniques (LR). There were two distinct datasets: one from ABSA dataset and the other from verified Facebook sites. & Used the model to undertake comparative analysis. The LR classifier fared the best, achieving an accuracy of 83.23%. SVC accuracy was 82.24% and NB was 82.04% [9].

Saha et al. mentioned their paper for the three separate models and one of its hybrid models analysis of sports-related remarks. They gathered 3759 sports-related data points from several social media platforms. They divided their dataset into five

categories. Their hybrid CNN-LSTM model outperformed the other two models in accuracy, coming in at 97.45% [10].

Nahar et al. brought social media to Bengali politics and sports news. The social media platforms are where the dataset was gathered. They employed NB, SVM, and NN among other machine learning algorithms. employed the TF-IDF for testing and training as well. 90% of the time is accurate. In comparison to the other methods, the NB classifier fared better. The goal is to employ a CNN-based strategy to address the issue of Bangla's new classification in the future [11].

Mahboob et al. This paper turned into performed if you want to Sentiment Analysis of RSS Feeds on Sports News – A Case Study. They acquire records from on line social media such as- articles, websites, blogs, messages, posts, news channels and the category of textual records into positive, negative and neutral categories. According to voyant-equipment evaluation summary, the badminton RSS feeds corpus has 2,676 overall phrases and 859 precise phrase, cricket RSS feeds corpus has 2,915 overall phrases and 1,071 precise phrase forms, soccer RSS feeds corpus has 2,211 overall phrases and 900 precise phrase forms, hockey RSS feeds corpus has 1,784 overall phrases and 682 precise phrase forms, and tennis RSS feeds corpus has 2,822 overall phrases and 1,068 precise phrase forms. Different feelings labeled as positive and negative were routinely decided the usage of LIWC on line of sports activities RSS (f-test/f-ratio, chi-square) feeds description. They located via way of means of studying the textual content in an outline of RSS feeds that Football has the best positive polarity is 49.1, Hockey has the least positive polarity is 26.7. Cricket has the best negative polarity is 26.8 and Football has the least negative polarity is 16.1[12].

Ljajić et al. presented a study they had written examining the sentiment of textual comments in the field of sport. The Stanford NLP Tool, the TF*IDF approach, and the machine learning method are the three components of the system they proposed. They collect data from social media, Sentiment140, internet forums, and online media. There is a total of 1.194 data points. The algorithm can be divided using the following steps: Pre-processing, Obtaining the characteristics (attributes), which may be written or numerical, the applicable algorithm is used to accomplish classification utilizing previously gathered attributes (Logistic regression, Naive Bayes, Maxent, SVM, etc.).

They consult three dictionaries: the LIWC, the MPQA, and the Opinion Lexicon. The results are considerably better when the dictionary developed using this method is complemented with another universal dictionary, such as Opinion Lexicon [13].

Bagić Babac et al. undertook their study in order to undertake a sentiment analysis of social media sport websites' user participation patterns. Sentiment analysis can be used to determine polarity or, in more complicated cases, to identify the aspect that a sentence or an entire document is related to. When dealing with polarity, an algorithm first needs a resource (text) to be evaluated. The algorithm can be broken down into the following steps (based on the proposal [2]): Preprocessing, Getting the features (attributes), which might be textual or numerical, Using previously collected attributes and a suitable classification technique (Logistic regression, Naive Bayes, Maxent, SVM, etc.), classification [14].

2.3 Summary of Research

Nowadays, people express their feelings, or opinions on different social platforms. In these reviews some are real and some are fake. There is a lot of discussion about sports. When their team wins a game, they celebrate loudly but If you lose a match, you will criticize and threaten. And then they express them furiously on various websites, such as Facebook pages, Facebook groups, etc. This problem can be solved using Natural language processing (NLP) to analyze the feelings of the people involved comment [10].

The main goal of this study is to contribute to our efforts in the field of sentiment analysis (SA) through deep learning techniques. In this project we have proposed three different methods. We collected 4061 data from different Facebook pages and groups, Twitter etc. Our dataset is unique because of this dataset built manually [6]. For the dataset, we have created five categories: neutral, happy, sad, positive and negative. Then we preprocessed our dataset. During preprocessing, we removed punctuations, cleaning data, etc. When modeling time, the dataset is split into two parts: one is the training dataset and the other is the test dataset. Once the process is complete, then it's time to training the model. We finally got our expected results for our proposed new models.

CHAPTER 3

Research Methodology

3.1 Introduction

Every proposed work model has some sequential process, so that this work would be more organized. In this work, we showed how our models work step by step.

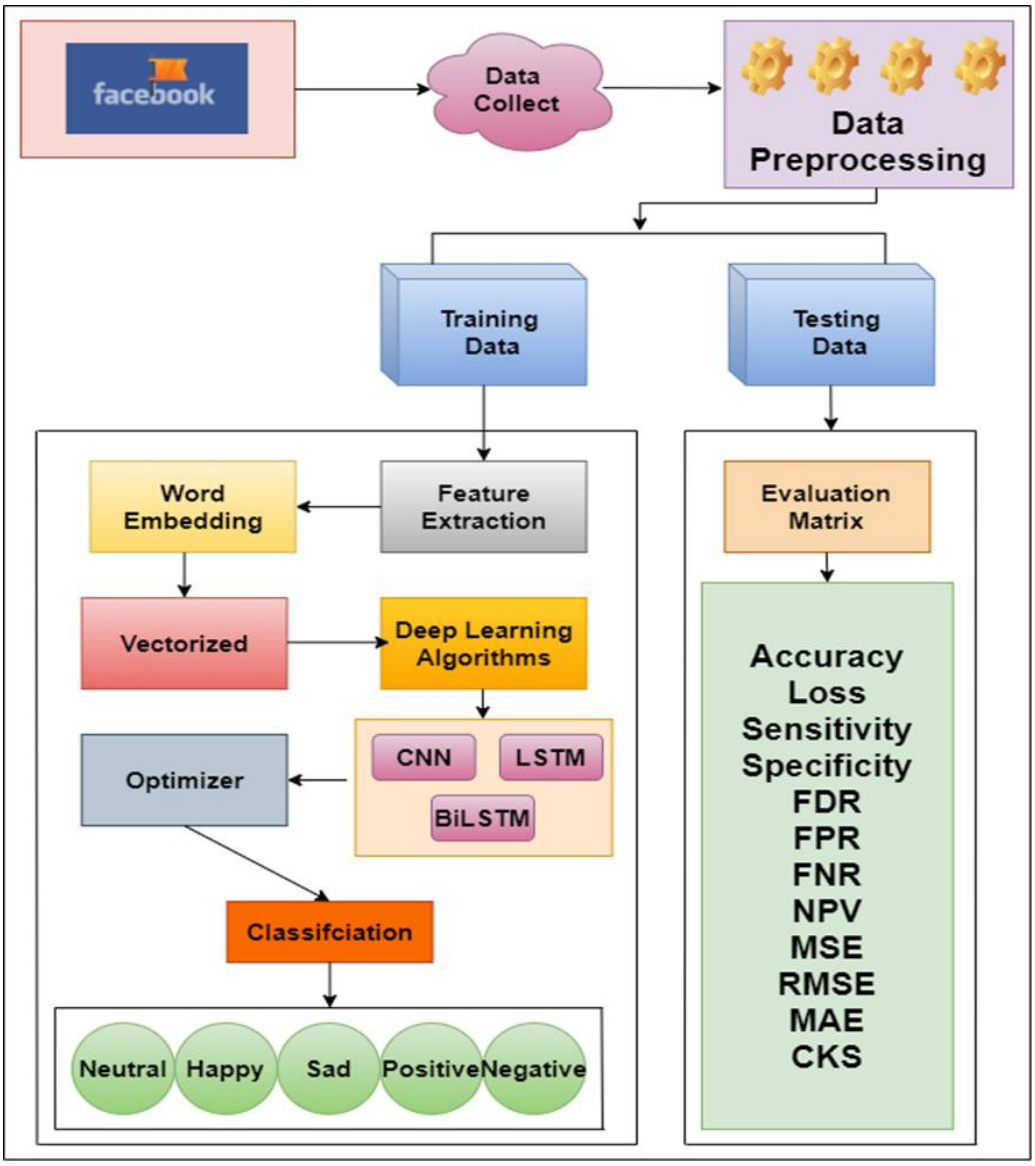


Figure 3.1: Overall Work Procedure.

3.2 Research Subject and Instruments

We proposed the title of our thesis is " Sentiment Analysis of Bengali Textual Comments in Field of Sports Using Deep Learning Approach ". Nowadays, Sentiment analysis is a very important topic in research. In our research method, we used deep learning models because there are certain mathematical functions that we used in our proposed model. It will be extremely difficult for us to work with deep learning model because it requires a high level of skill Configure PC, GPU and other devices. Here is a list of essentials equipment and technology to make our models work-

Software and Hardware:

- Intel Core i3 7th generation with 8GB RAM.
- 1TB HDD
- 256GB SSD
- Google Colab with GPU
- Development Tools:
- Windows 10
- Python 3.7
- 1.15.2 version of TensorFlow Backend
- NumPy
- NLTK
- Panda

3.3 Data Collection

Data is the primary goal of a proposed work. Because without data we cannot start work on a specific topic. A dataset is a collection of some values of a specific topic. We collected our dataset values from some sports related Facebook pages. In these pages, people share their opinions on various sports. We collected football and cricket, hockey and badminton related Bengali comments [10]. We collected almost 4061 data for different Facebook pages. This dataset consists of five categories: neutral, happy, sad, positive and negative. Here we've found 933 neutrals, 788 happy, 732 sad, 753 positive and 855 negative data [6]. Our own build dataset consists of three different columns, one is comments, then tag and the last one is category.

TABLE 3.3: THE DATA TABLE REPRESENTS THE CATEGORY OF SPORTS COMMENTS

Class	Sentence
Neutral	কক্সবাজার জেলা দল চ্যাম্পিয়ন হবে সারা বাংলাদেশে (Cox's Bazar district team will be champion in whole Bangladesh)
Happy	ভাল সবাই মর্নিং ওয়াক করতাছে (Well, everyone is doing morning walk)
Sad	নিজের দেশের একটা খাঁটি ফুটবলার নেই (There is no real football player of his country)
Positive	সব যায়গায় আবাহনী সুবিধা পায় (Abahani benefits everywhere)
Negative	ওনাকে বাংলাদেশের জাতীয় দলে নেওয়া উচিত নাহ (He should not be included in the national team of Bangladesh)

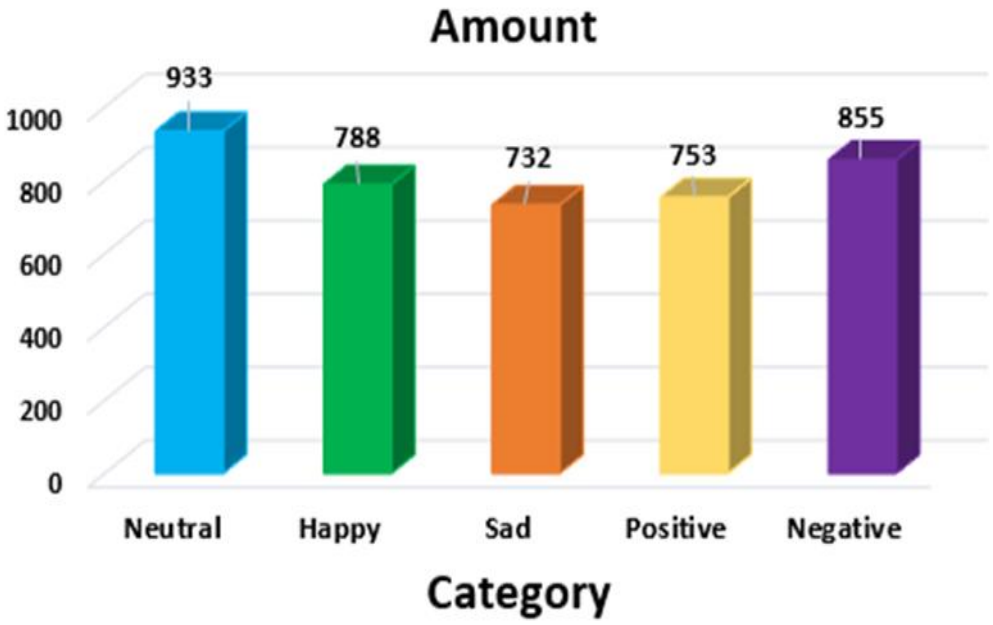


Figure 3.3.1: Amount of category wise dataset.

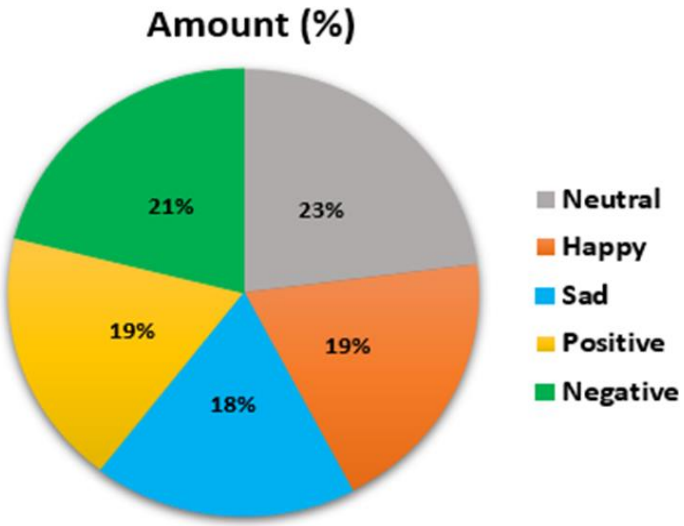


Figure 3.3.2: Percentage of dataset.

3.4.1 Data Preprocessing

Data preprocessing techniques are the most valuable part of a dataset. We know that dataset is not appropriate at all because every dataset has some missing, noisy, errors values. Preprocessing techniques is the process to fix this type of errors, missing or noisy data. After all, we used the stop word removal techniques to remove sentences stop word [15]. Some sentences were incomplete; we complete those sentences using a manual validation process. We used the tokenization process to divide a sentence into a token.

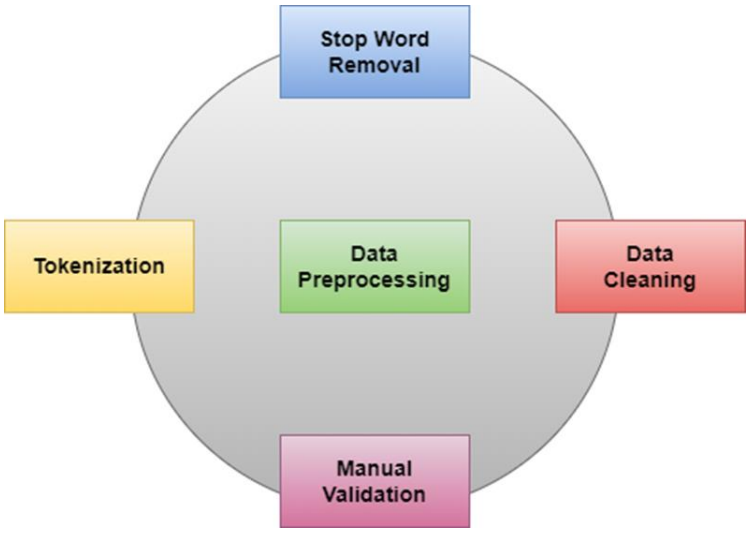


Figure 3.4.1: Preprocessing techniques.

3.4.2 Tokenization

Tokenization is one of the initial steps of NLP. It means dividing a raw sentence into small tokens or chunks of words [16]. When a sentence is split into words it's called word tokenization. when we applied to data security, it is the process of replacing a piece of sensitive data with a non-sensitive equivalent, known as a token, that has no intrinsic or realizable meaning or value. Token is a reference that refers to sensitive data through the tokenization system.

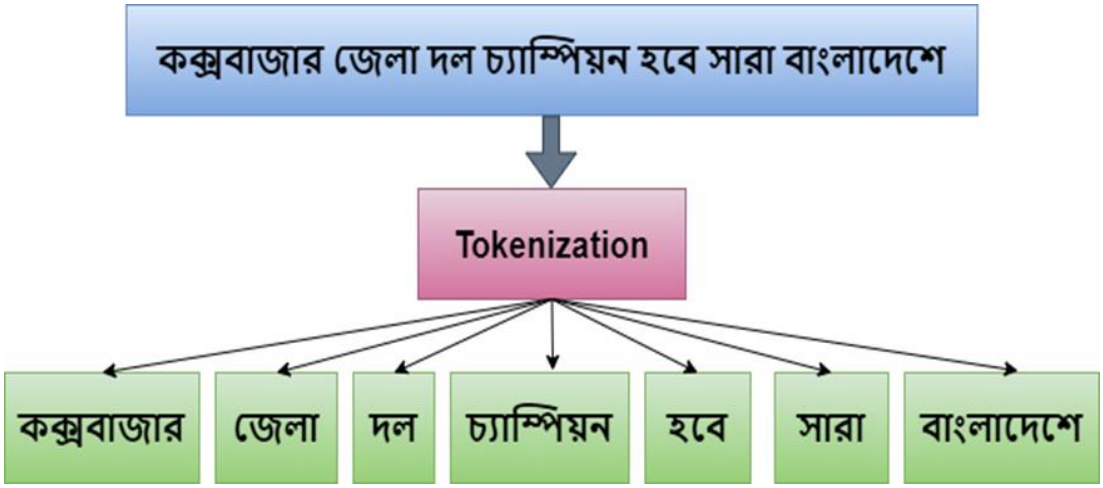


Figure 3.4.2: Tokenization Process.

3.5 Statistical Analysis

In this work, we collected 4061 datasets from different Facebook groups and pages etc. There are many sports related news like football, cricket, hockey, badminton, etc. manually entered. There are 933 neutral data, 753 positive, 855 negative and 788 happy, 732 sad. Right here we used 3249 samples for training and 812 samples for validation.

3.6 Methodology

Each work is executed sequentially and there is one work procedures to complete the work. In deep learning, there are different types of models, and these models have different types of motives. Convolution Neural Networks (CNN), Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) will be very convenient for our proposed work.

3.6.1 Word Embedder and Optimizer

The mathematical representation of words or documents is called word embedding. We can basically use it to convert words to numbers. Embedded Word Used when we need numbers to train a Deep learning technique. In this article, we have used a combination of the words glove for the algorithm.

Glove:

Glove is a word vector demo created by a Unsupervised learning algorithm. Gloves are another word integration of an implementation using a concurrent occurrence matrix. This model takes into account global properties in a database [21].

Adam Optimizer:

The Adam optimizer is basically a combination of AdarGrade and RMSProp. RMSProp uses squares derivative, while Adargrader uses first-order derivative or the normal derivative (gt) (gt2) [10].

3.6.2 Convolutional Neural Networks (CNN)

CNNs seem to be multi - layer perceptron's that are capable of identifying characteristics from text data and other complicated characteristics from data. The main part of CNN is the convolution layer and the pooling layer. Recently, CNN highlighted the impact of the uprising on the taxonomy of the NLP field. For our model, we firstly used the embedding layer with the help of spatial dropout value 0.2 then using the convolution one dimensional (Conv1D). Here we used filters size 32, kernel size 5 and the activation is relied [17]. After adding Conv1D, we used GlobalMaxPooling1D and added a layer called dense with activation SoftMax. Then we added an optimizer for a well-trained model. Adam optimizer was employed to improve the model's accuracy. In the end, we trained our CNN model using 256 batches and 25 epochs.

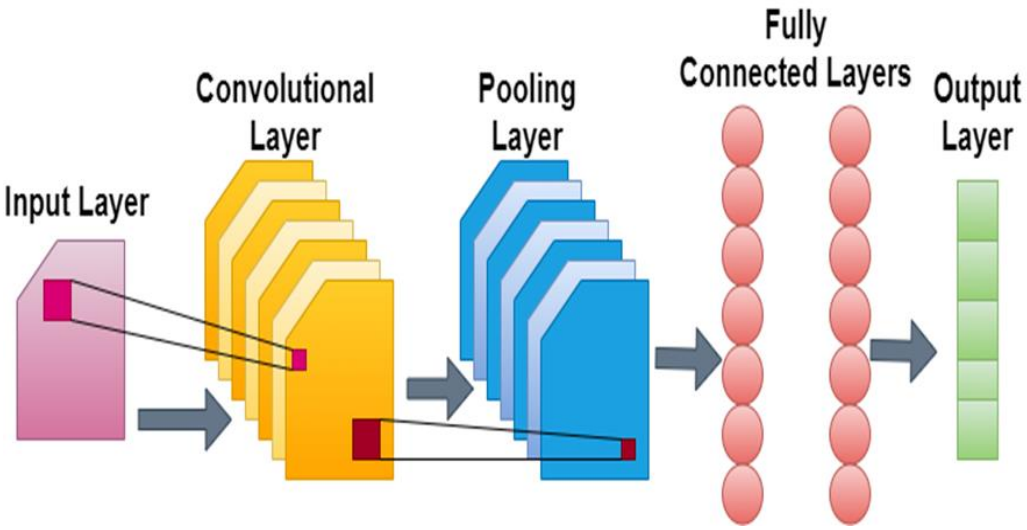


Figure 3.6.2: Architecture of CNN Model [18].

3.6.3 Long Short-Term Memory (LSTM)

LSTM refers to the analogy that standard RNNs have both "long-term" and "short-term" memory. The weights and biases of connections within the community are extruded to match the training episode. This is analogous to how physiological changes in synaptic strength store long-term memory. Activation patterns in the network are extruded to match the time steps. This is similar to how momentary changes in motorized firing patterns store short-term memory in the brain. The LSTM architecture aims to provide short-term memory, hence "long-short-term memory", for RNNs that can take thousands of time-steps [19].

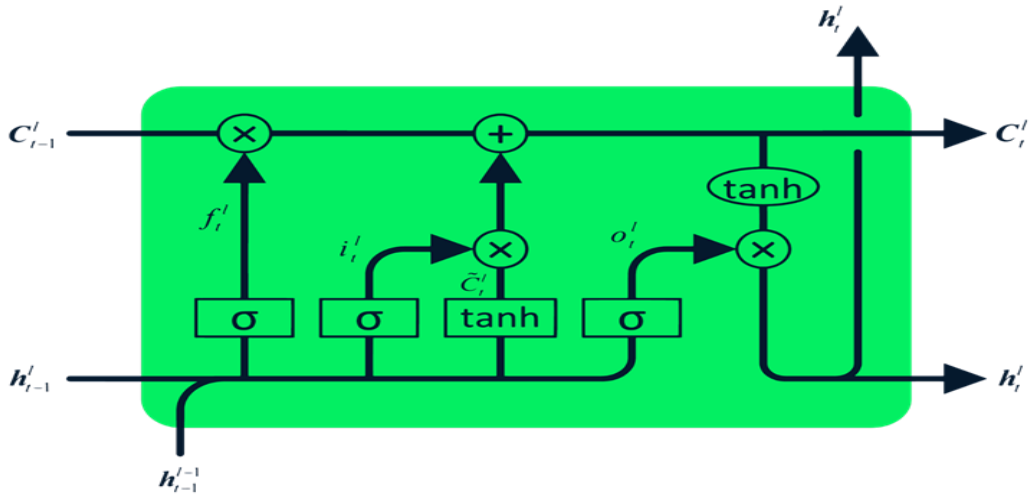


Figure 3.6.3: Architecture of LSTM Model [27].

3.6.4 Bidirectional Long Short-Term Memory (BiLSTM)

A recurrent neural network called Bidirectional LSTM (BiLSTM) is mostly utilized for processing natural language. In contrast to conventional LSTMs, inputs flow in both directions, and data from both sides can be utilized. Additionally, it is an effective tool for simulating word and phrase dependencies in both directions of the sequence. There are two LSTMs in it. One accepts input going forward, and the other accepts input going backward [20]. By using BiLSTM, the network's access to information is effectively increased, and algorithms have better context (for example, knowing which words immediately follow and precede the words in a sentence). The outputs from the two LSTM layers are then combined in a variety of methods, including average, sum, multiplication, and concatenation [21].

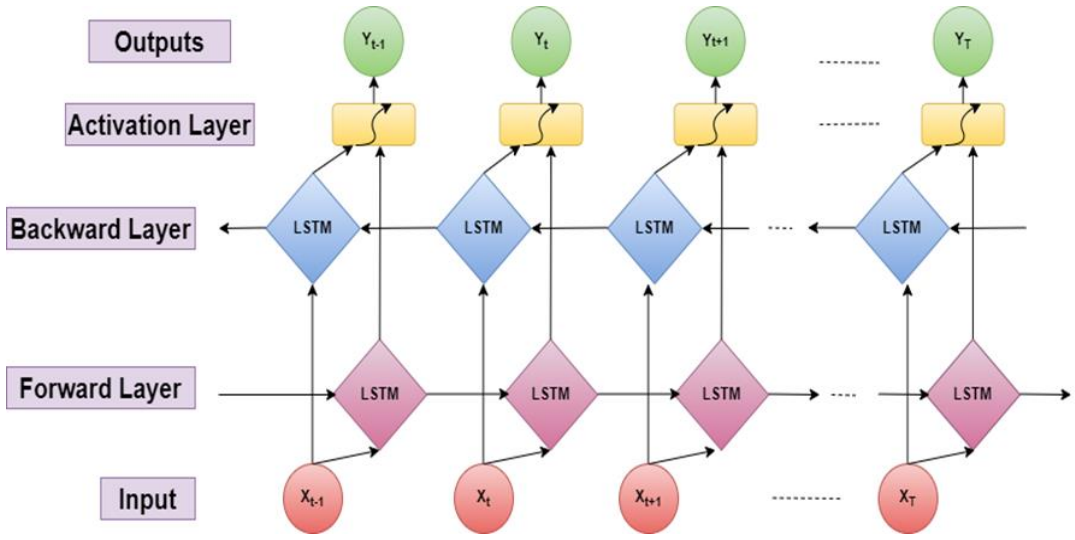


Figure 3.6.4: Architecture of BiLSTM Model [22].

3.7 Implementation Necessity

In our proposed work, we used three models. These models are very useful for sentiment analysis in Bengali language. Our dataset is mainly the size of the chunks collected in different logs. These models work as sentence by sentence through layers of wisdom. These sentences turn into words, then we got our final output.

3.8 Parameter Tuning of Models

Parameter tuning means how much the model's batch size, epoch size and maximum length.

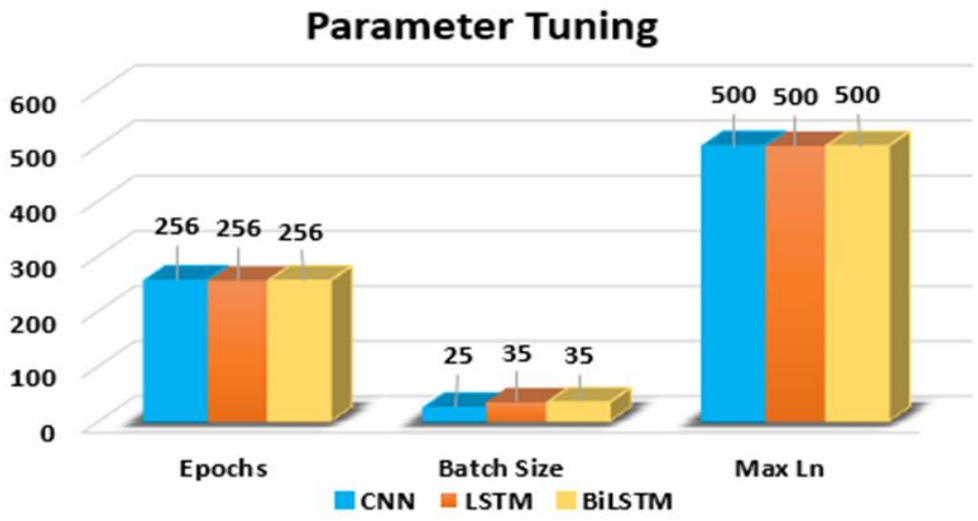


Figure 3.8: Graphical representation of Parameter Tuning.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

We would like to take a detailed look at the different algorithms used in this project before moving on to this section chapter. Now, in this chapter, we will talk about the results of different algorithms. Before speaking, we used word embedding technique that is Glove [21]. We provided the models are CNN, LSTM and BiLSTM. In this section, we discuss the results of our models and other resulting analyses. To minimize optimization and model loss, we used Adam optimizer. It can send us a good accuracy after a food model train. In these three optimizer, the Adam optimizer is important for the training time of other optimizers. Adam optimizer is important because it computes all the parameters. For Deep Learning algorithm model training, a well-configured computer or laptop is essential. A GPU is required to train the dataset. Firstly, we used a live PC to train our proposed model. Therefore, running the model takes one for a long time and the results obtained are not consistent with the models that we propose. So, for that project, we used Google Colab to train our models. It provides users with a free GPU a service [10].

4.2 Descriptive Analysis

Accuracy:

An indicator of the model's performance across all classes is accuracy. It is helpful when all classes are equally significant. Accuracy is one of the metrics that is most frequently utilized when performing classification [23]. It is expressed as the ratio of correctly classified (TP+TN) and (TP+TN+FP+FN).

Precision:

The accuracy is calculated as the ratio of Positive samples that were correctly classified to all samples that were classified as Positive (True or False). The precision evaluates how accurately the model categorizes positive samples.

Precision= TP/(TP+FP)

The denominator will rise and the precision will fall if the model makes many false positive classifications or few true positive classifications. Contrarily, accuracy is

greater when: The model produces accurate Positive classifications (maximize True Positive). Less erroneous positive classifications are made by the model (minimal False Positive). Imagine a person who is respected by others and who is taken seriously when he makes predictions. Like this individual, the precision [28]. If the model forecasts a sample as positive with a high level of precision, you can believe it. Thus, precision enables us to determine how correctly the model predicts that the sample is positive.

Recall:

The percentage of positive samples that were correctly identified as positive to all positive samples is used to calculate recall. The model's capacity to identify positive samples is measured by recall. More positive samples are found the higher the recall.

Recall= TP/(TP+FN)

Recall the classification of positive samples. This is unrelated to the classification given to negative samples. for accuracy. The recall is 100% if the model properly classifies all positive samples as positive, even if all negative samples are misclassified as positive [28].

F1-Score:

The F1 score, which is the harmonic mean of precision and recall, provides insight into the relationship between these two measurements. When recall and precision are equal, it is maximized. But there are dangers in this. F1 score interpretability is poor. That implies that we are unsure of the classifier's preference for precision or recall. As a result, it can be used with other evaluation measures to provide a fuller view of the outcomes. In reality, trying to increase model precision reduces recall, and the opposite is also true. Both trends are represented by a single value in the F1-score [28].

4.3 Experimental Analysis

TABLE 4.3.1. MODELS ACCURACY

Algorithms	Accuracy (Validation)	Accuracy (Train)
CNN	94.57%	82.20%
LSTM	90.58%	77.47%
BiLSTM	88.77%	77.25%

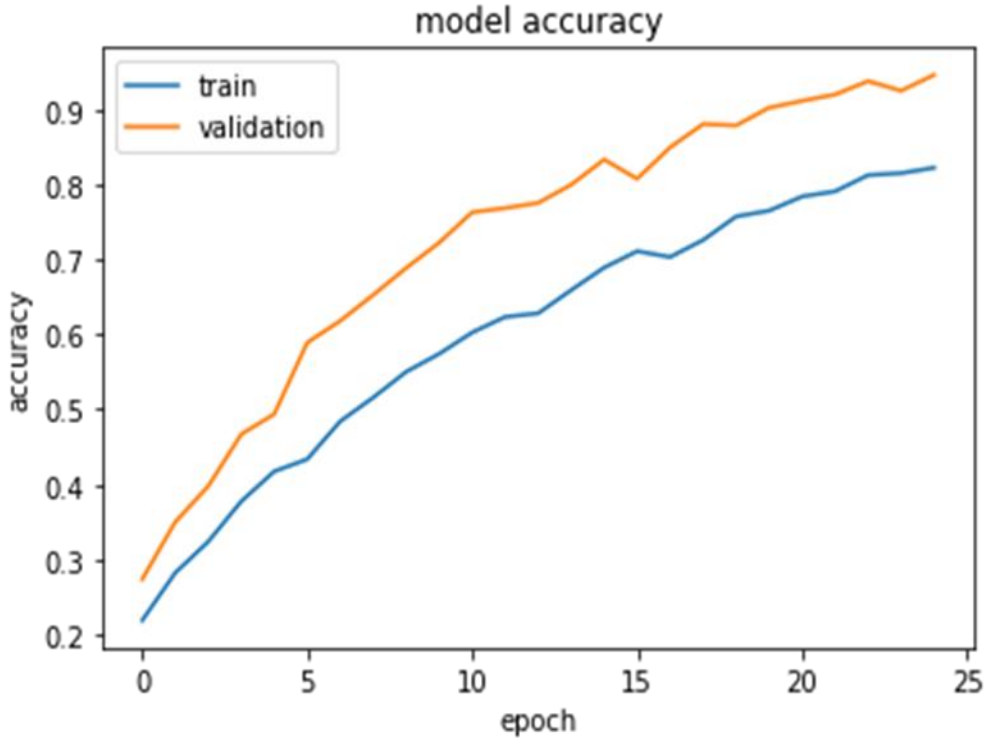


Figure 4.3.1: Training and Validation Accuracy graph of CNN Model.

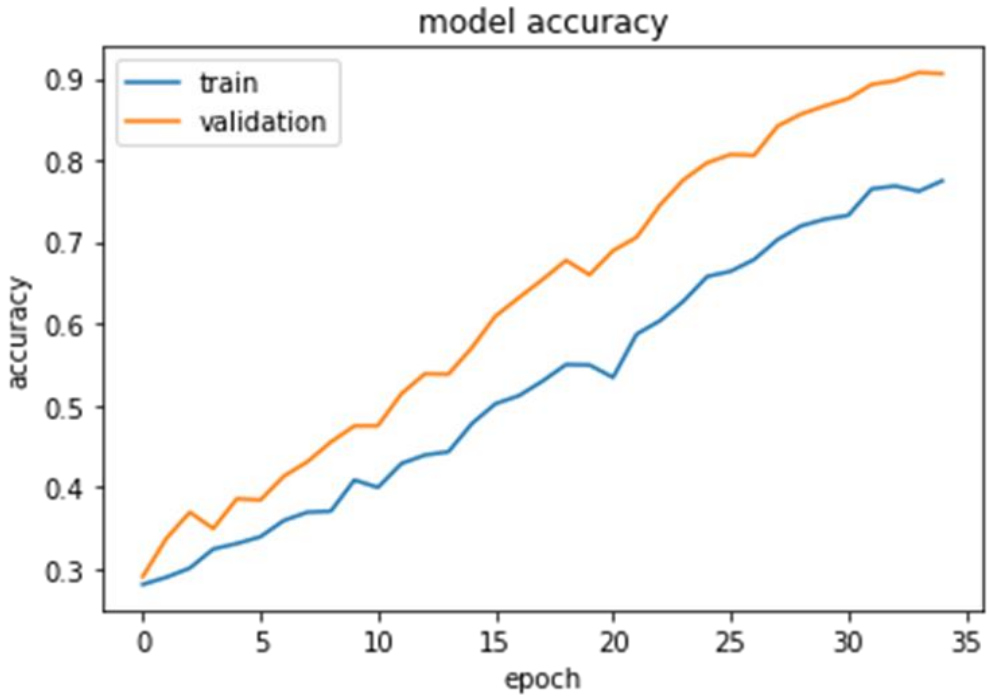


Figure 4.3.2: Training and Validation Accuracy graph of LSTM Model.

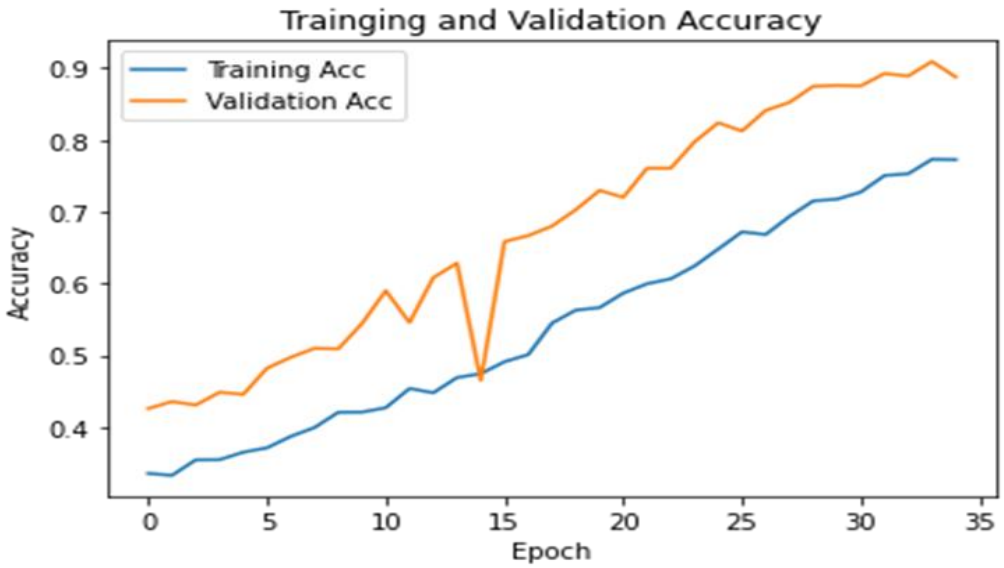


Figure 4.3.3: Training and Validation Accuracy graph of BiLSTM Model.

TABLE 4.3.2. MODELS LOSS

Algorithms	Loss (Validation)	Loss (Train)
CNN	53.38%	72.04%
LSTM	37.37%	62.49%
BiLSTM	43.07%	62.19%

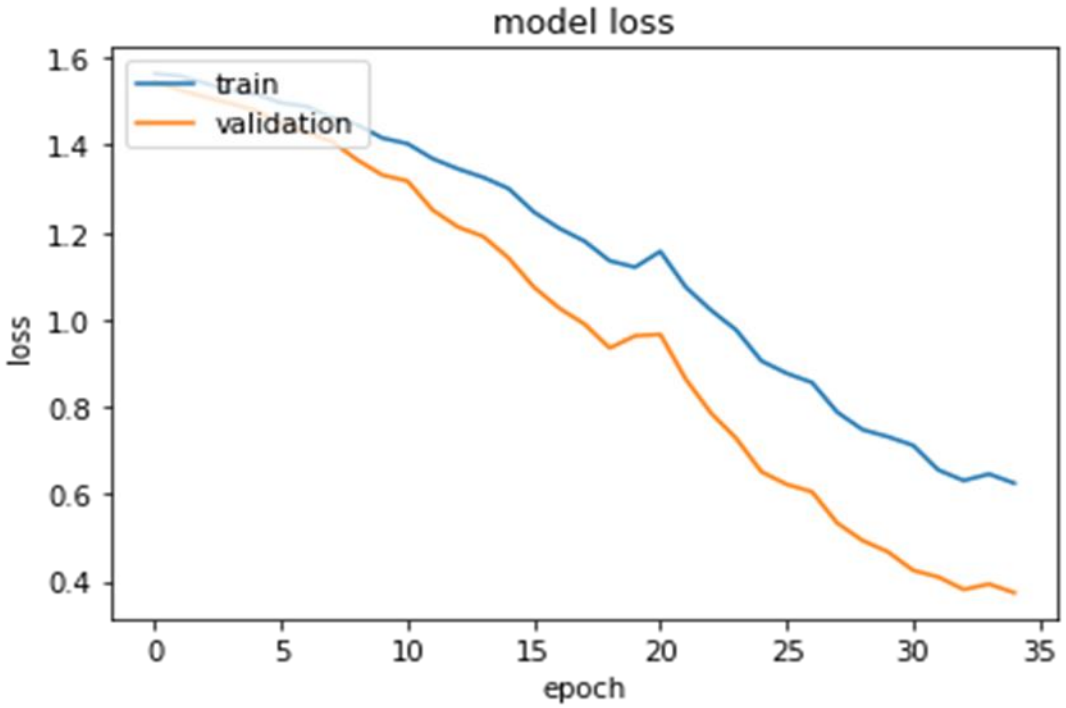


Figure 4.3.4: Training and Validation Loss graph of CNN Model.

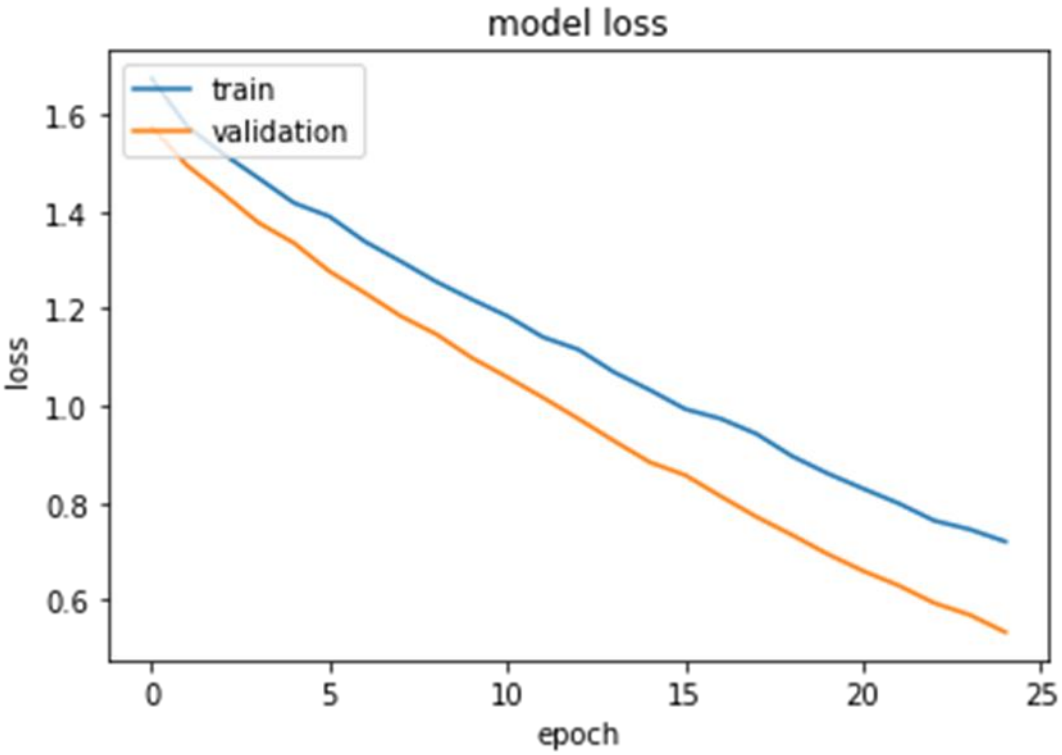


Figure 4.3.5: Training and Validation Loss graph of LSTM Model.

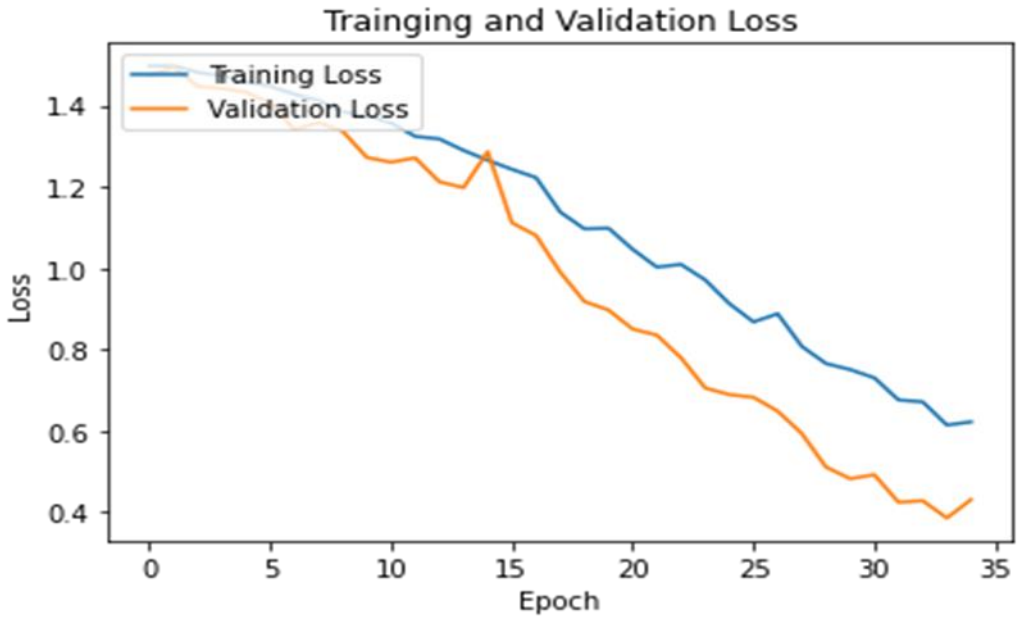


Figure 4.3.6: Training and Validation Loss graph of BiLSTM Model.

TABLE 4.3.3: PRECISION, RECALL AND F1 SCORE RESULT

Algorithms	Categories	Precision (%)	Recall (%)	F1 Score (%)
CNN	Neutral	0.96	0.90	0.93
	Happy	0.93	0.93	0.93
	Sad	0.94	0.96	0.95
	Positive	0.97	0.95	0.96
	Negative	0.94	0.99	0.96
LSTM	Neutral	0.80	0.91	0.85
	Happy	0.90	0.95	0.92
	Sad	0.93	0.90	0.92
	Positive	0.93	0.91	0.92
	Negative	0.95	0.85	0.90
BiLSTM	Neutral	0.81	0.93	0.87
	Happy	0.98	0.87	0.92
	Sad	0.91	0.88	0.90
	Positive	0.92	0.85	0.88
	Negative	0.80	0.93	0.86

Macro Average:

The macro average is the arithmetic mean of each class with respect to precision, memory and f1 score. We use the average of macro score when we need to treat all classes equally to evaluate the classifier's overall performance against the most common class labels. The use of Macro point F1 score is that it gives equal weight to all data points [28]. Then F1 Micro gives equal weights to the whole class and it is not affected by deviations in the distribution of the class log loss, it penalizes small deviations in the class.

Weighted average:

A procedure known as a weighted average account for the varied significance of the numbers in a dataset. Each number in the data set is multiplied by a predefined weight before the final computation is made when calculating a weighted average [28].

TABLE 4.3.4: MACRO AND WEIGHTED AVERAGE RESULT

Algorithms	Macro/Weighted Average	Precision (%)	Recall (%)	F1 Score (%)
CNN	**Macro Average**	0.95	0.94	0.95
	Weighted Average	0.95	0.95	0.95
LSTM	**Macro Average**	0.90	0.90	0.90
	Weighted Average	0.91	0.91	0.91
BiLSTM	**Macro Average**	0.89	0.89	0.89
	Weighted Average	0.89	0.89	0.89

A classification algorithm's effectiveness can be determined using a confusion matrix. It displays both actual values and counts of divined values. The "TN" output stands for True Negative and reflects the number of perfectly negative samples [24]. Additionally, "TP" stands for True Positive and denotes the quantity of positively classified samples that are true. The terms "FP" and "FN" stand for false positive and false negative values, respectively, and refer to the number of real negative samples that were mistakenly categorized as positive samples.

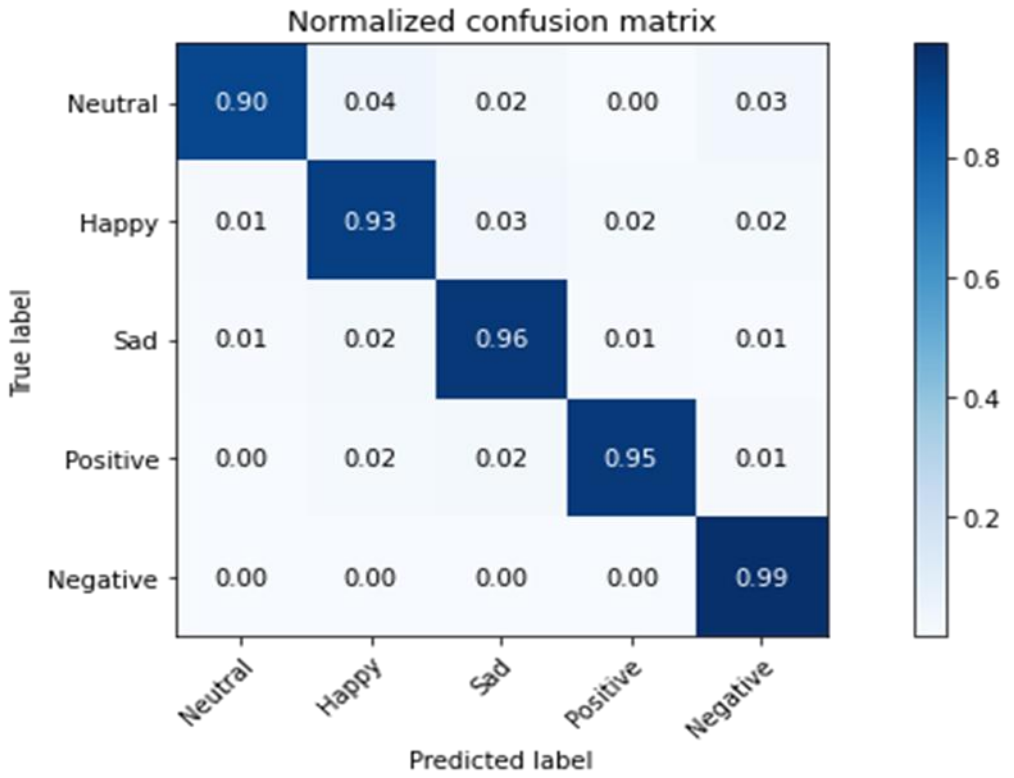


Figure 4.3.7: Confusion Matrix of CNN Model.

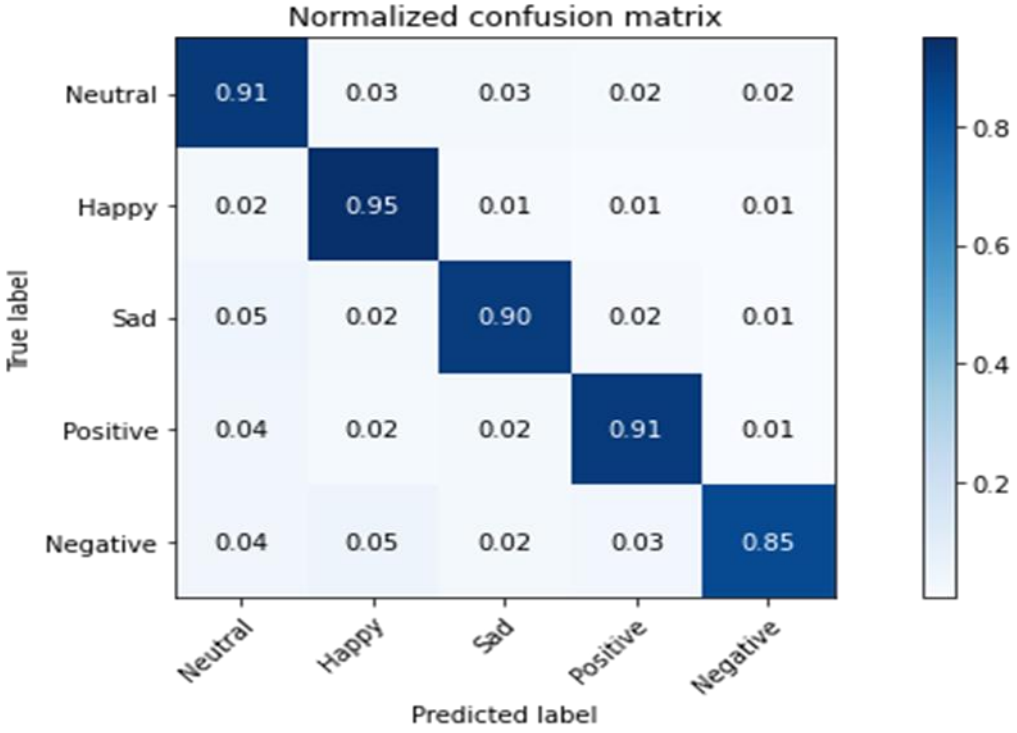


Figure 4.3.8: Confusion Matrix of LSTM Model.

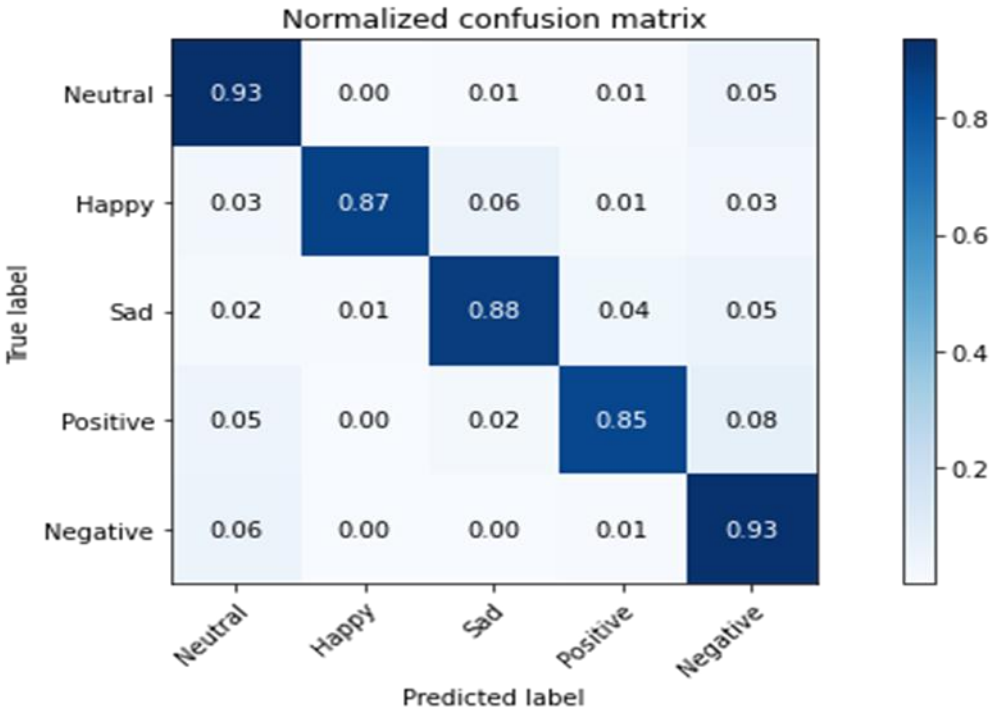


Figure 4.3.9: Confusion Matrix of BiLSTM Model.

Sensitivity:

Take sentiment analysis on Bengali sports comments as an example. Sensitivity describes how well a test can identify different remark types, such as happy, sad, neutral, negative and positive [25]. The sensitivity of the test is the percentage of comments in which categories, and it is used to detect a condition. Mathematically, this can be expressed as:

Sensitivity = number of true positive / (number of true positive + number of false negative)

Specificity:

The test's ability to be neutrally uncategorized is related to its specificity. A test's specificity is determined by how many people are actually negative for the illness [26]. Mathematically, this can also be written as:

Specificity = number of true negative /(number of true negative + number of false positive)

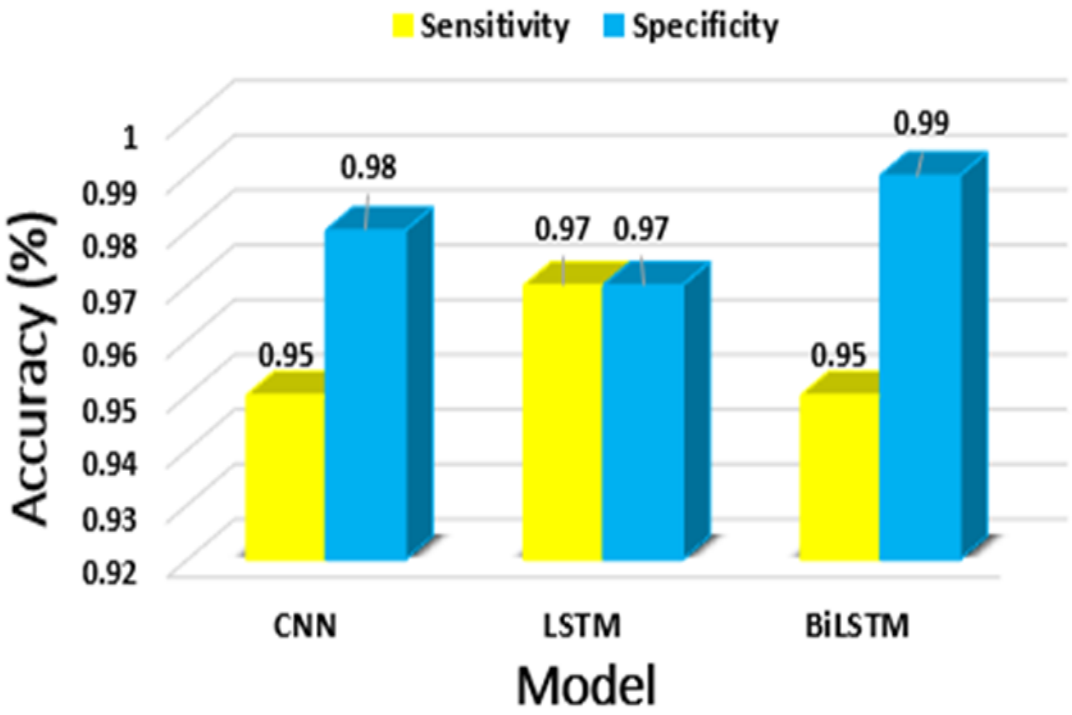


Figure 4.3.10: Sensitivity and Specificity result of our models.

TABLE 4.3.5: CONFUSION MEASURE MATRIX RESULT

Algorithms	FPR	FNR	NPV	FDR	MAE	MSE	RMSE	CKS
CNN	0.029	0.022	0.986	0.047	0.155	0.047	0.218	0.931
LSTM	0.018	0.041	0.974	0.029	0.101	0.035	0.187	0.881
BiLSTM	0.002	0.045	0.968	0.003	0.111	0.417	0.204	0.858

We examined the three models' accuracy in figure 14. In this instance, the CNN model had the highest accuracy (94.57%).

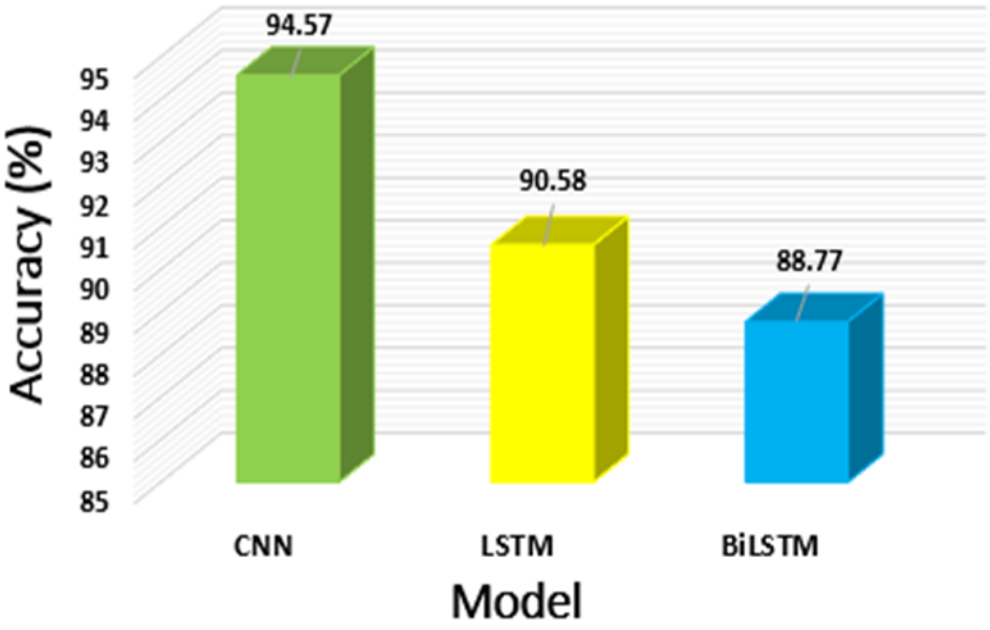


Figure 4.3.11: Accuracy comparing of three models.

The effectiveness of a model is determined by how closely the projected and observed results correlate. The following table shows the accuracy of our model's forecasts for the inputs.

TABLE 4.3.6: PREDICTED TESTING OUTCOME

Algorithms	Sentence	Actual Class	Predicted Class
CNN	কক্সবাজার জেলা দল চ্যাম্পিয়ন হবে সারা বাংলাদেশে	Neutral	Neutral
	ভাল সবাই মর্নিং ওয়াক করতাছে	Happy	Happy
	সব যায়গায় আবাহনী সুবিধা পায়	Sad	Sad
	নিজের দেশের একটা খাঁটি ফুটবলার নেই	Positive	Positive
	ওনাকে বাংলাদেশের জাতীয় দলে নেওয়া উচিত নাহ	Negative	Negative
LSTM	কক্সবাজার জেলা দল চ্যাম্পিয়ন হবে সারা বাংলাদেশে	Neutral	Neutral
	ভাল সবাই মর্নিং ওয়াক করতাছে	Happy	Happy
	সব যায়গায় আবাহনী সুবিধা পায়	Sad	Sad
	নিজের দেশের একটা খাঁটি ফুটবলার নেই	Positive	Positive
	ওনাকে বাংলাদেশের জাতীয় দলে নেওয়া উচিত নাহ	Negative	Negative

Algorithms	Sentence	Actual Class	Predicted Class
BiLSTM	কক্সবাজার জেলা দল চ্যাম্পিয়ন হবে সারা বাংলাদেশে	Neutral	Positive
	ভাল সবাই মর্নিং ওয়াক করতাছে	Happy	Happy
	সব যায়গায় আবাহনী সুবিধা পায়	Sad	Sad
	নিজের দেশের একটা খাঁটি ফুটবলার নেই	Positive	Positive
	ওনাকে বাংলাদেশের জাতীয় দলে নেওয়া উচিত নাহ	Negative	Negative

4.4 Discussion

We proposed three models in our article. We used deep learning algorithms to our proposed models that is Bidirectional Long-Short-Term Memory (BiLSTM), Long-Short-Term memory (LSTM) and Convolutional Neural Network (CNN) with word embedding technique and generated the proposed model. Without the optimizer, the results are very poor and we worry about that. But when we use the optimizer, the results are very good and we are satisfied. Adam Optimizer obtained the best results of all methods. We measure all kinds of loss functions in the process learning phase and wait for all iterations to complete before calculating final loss a function. We separated the data into testing and training groups before starting to train the model.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Our discoveries will have a significant impact on society. In this study, we worked with opinion classification using deep learning method. Now is the information age technology, and information technology has penetrated every aspect of our society. Our methods have a certain impact on society. Every day we talk to each other but we can't understand their true feelings. Our models can help with this type of text classification.

5.2 Ethical Aspects

Ethical considerations in research are a set of principles that guide research designs and practices. Scientists and researchers must always adhere to a certain code of conduct when collecting data from people. The goals of human research often include understanding real-life phenomena, researching effective treatments, studying behaviors, and improving lives in other ways.

Our models or types of work do not violate any basic human rights or ethical secrecy. We have collected data from different Twitter, Facebook pages and groups and all the Facebook group and pages still available in online. Therefore, the data we cannot be used to determine or hurt someone. We did not do any work or collect any data by hurting or threaten people in the performance of our duties. Because our work is data-driven, we have Taken extra precautions when collecting and storing data. When we do our job, we did not claim the work of any other entity or individual as our own. While working, we used our computer. We didn't use anyone else's device and we didn't steal other people's information or data. We conduct our research now honesty, respect for the law, integrity, lawfulness and transparency [29].

5.3 Sustainability Plan

Our main goal is to use deep learning algorithms to classify opinions. Lots of company or organizational changes can be made in Naive Bayes using our program. Only certain datasets will work with our model. So to promote and sustain this future work we will need a large number of datasets related to different Bengali sentences. In future, when

the necessary Bengali dataset is improved, this model can be applied to educational, military, industrial and commercial sectors, in which others. Our proposed new model will help in the future for other large data sets or other data sets.

CHAPTER 6

Conclusion and Future Work

6.1 Summary of the Study

For this project, our entire work focused on sentiment analysis. We did our job to classifies comments of social media commentators using deep learning algorithms. Our work is very useful to categorize comments. We obtained good results for all the models in the dataset we built. From the time we started collecting data to the completion of the project, we took three months. To accomplish this, we have to go through a series of step by step. The entire job step is now summarized and listed below-

Step 1:

Commenter's opinion datasets were collected from various Facebook group and pages, Twitter etc.

Step 2:

Store all data in an Excel.xlsx file.

Step 3:

Split the dataset.

Step 4:

Data set preprocessing.

Step 5:

Use Word Embedding.

Step 6:

Use LSTM, BiLSTM and CNN model classes.

Step 7:

Model test and train.

Step 8:

Control output or results check.

We accomplished our task by following these steps in order. In this model will help us Categorize editor's comments and experiment further with other projects.

6.2 Conclusion

Text mining and sentiment analysis have increased in research area day by day. This area has many obstacles as it includes language processing. It has many varieties, users can benefit from its results, for example, check news, slideshows, answer reader questions or users do. The essential experiences of opinion are communication on the web, especially from social networks is essential for some organizations and institutions, whether it's product reviews, the open mind or emotions that they share their opinion with financial professionals.

In this study, we mainly worked on Bengali people's comments of various sports, like cricket, football, hockey, badminton etc. In this analysis, we used different models to extract sentence sentiment. There are a lot of sports that exist in Bangladesh. And Bangladeshi people love that. When a player plays badly then some of the fans are bullying this player. We collect this type of bullying or others comments and classify them by manually. Three well-known deep learning models—CNN, LSTM, and BiLSTM are employed. A total of 4061 data points were gathered and divided into five categories. The CNN model's accuracy in these three models is 94.57%, which is higher than the accuracy of the other two models.

6.3 Implementation for Further Work

We discovered some limitations in doing this, such as the fact that we disagree with closed domains and our dataset is not enough. Any kind of model designed for Upcoming improvements, as we all know. Because any kind of experimental study is an endless method are improving day by day. After completing this study, we will have to extend the model. In this project, we use three models with Word Embedding. For further work, we will use more data to get better accuracy and also use some other deep learning algorithms or hybrid algorithms for better work. Further We will do this work by comparing various models, like ML and DL.

References:

[1] P. Lavanya and E. Sasikala, "Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, pp. 603-609, doi: 10.1109/ICSPC51351.2021.9451752.

[2] J. Chai and A. Li, "Deep Learning in Natural Language Processing: A State-of-the-Art Survey," 2019 International Conference on Machine Learning and Cybernetics (ICMLC), 2019, pp. 1-6, doi: 10.1109/ICMLC48188.2019.8949185.

[3] M. H. Abd El-Jawad, R. Hodhod and Y. M. K. Omar, "Sentiment Analysis of Social Media Networks Using Machine Learning," 2018 14th International Computer Engineering Conference (ICENCO), 2018, pp. 174-176, doi: 10.1109/ICENCO.2018.8636124.

[4] M. K. Mahbub, M. A. M. Miah, S. M. S. Islam, S. Sorna, S. Hossain and M. Biswas, "Best Eleven Forecast for Bangladesh Cricket Team with Machine Learning Techniques," 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2021, pp. 1-6, doi: 10.1109/ICEEICT53905.2021.9667862.

[5] S. Alam, M. A. U. Haque and A. Rahman, "Bengali Text Categorization Based on Deep Hybrid CNN–LSTM Network with Word Embedding," 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET), 2022, pp. 577-582, doi: 10.1109/ICISET54810.2022.9775913.

[6] Bangladesh Football Federation Facebook Group, available <<https://www.facebook.com/bff.football/>>, last accessed on 30 August, 2022.S.

[7] Arafin Mahtab, N. Islam and M. Mahfuzur Rahaman, "Sentiment Analysis on Bangladesh Cricket with Support Vector Machine," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018, pp. 1-4, doi: 10.1109/ICBSLP.2018.8554585.

[8] M. F. Wahid, M. J. Hasan and M. S. Alom, "Cricket Sentiment Analysis from Bangla Text Using Recurrent Neural Network with Long Short-Term Memory Model," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), 2019, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201500.

[9] M. A. Faruque, S. Rahman, P. Chakraborty, T.Choudhury, J. Um, T. P. Singh,“Ascertaining polarity of public opinions on Bangladesh cricket using machine learning techniques”, Spatial Information Research. https://doi.org/10.1007/s41324-021-00403-8

[10] U. Saha, M. S. Mahmud, S. Shimu, S. N. Eva, I. A. Asif and S. A. Khushbu, "The Corporeality of Infotainment on Fans Feedback Towards Sports Comment Employing Convolutional Long-Short Term Neural Network," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1081-1086, doi: 10.1109/ICOEI53556.2022.9777216.

[11] L. Nahar, Z. Sultana, N. Jahan and U. Jannat, "Filtering Bengali Political and Sports News of Social Media from Textual Information," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934605.

[12] K. Mahboob, F. Ali and H. Nizami, "Sentiment Analysis of RSS Feeds on Sports News – A Case Study", International Journal of Information Technology and Computer Science. December 2019. DOI: 10.5815/ijitcs.2019.12.02

[13] A. Ljajić, E. Ljajić, P. Spalević, B. Arsić and D. Vučković, "Sentiment analysis of textual comments in the field of sport", ERK. September 2015.

[14] M. B. Babac, and V. Podobnik, "A sentiment analysis of who participates, how and why, at social media sport websites: How differently men and women write about football", Online Information Review, Vol. 40 No. 6, pp. 814-833. https://doi.org/10.1108/OIR-02-2016-0050

[15] H. S. Obaid, S. A. Dheyab and S. S. Sabry, "The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning," 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), 2019, pp. 279-283, doi: 10.1109/IEMECONX.2019.8877011.

[16] P. Khowrurk and R. Kongkachandra, "A Machine Learning Approach for the Classification of Methamphetamine Dealers on Twitter in Thailand," 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2020, pp. 1-5, doi: 10.1109/iSAI-NLP51646.2020.9376817.

[17] X. Ouyang, P. Zhou, C. H. Li and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015, pp. 2359-2364, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.349.

[18] CNN Images, available <<https://www.davidsbatista.net/assets/images/2018-03-31-mylenet_cnn.png>>, last accessed on 15 September, 2022.

[19] D. Li and J. Qian, "Text sentiment analysis based on long short-term memory," 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), 2016, pp. 471-475, doi: 10.1109/CCI.2016.7778967.

[20] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[21] U. Saha, M. S. Mahmud, A. Chakrobortty, M. T. Akter, M. R. Islam and A. A. Marouf, "Sentiment Classification in Bengali News Comments using a hybrid approach with Glove," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 01-08, doi: 10.1109/ICOEI53556.2022.9777096.

[22] BiLSTM Architecture, available <<https://miro.medium.com/max/766/0*ZsTT3zzTNGF6OsR.jpg>>, last accessed on 15 September, 2022.

[23] E. Dufourq and B. A. Bassett, "Text compression for sentiment analysis via evolutionary algorithms," 2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech), 2017, pp. 116-121, doi: 10.1109/RoboMech.2017.8261133.

[24] M. S. Mahmud, A. Jaman Bonny, U. Saha, M. Jahan, Z. F. Tuna and A. Al Marouf, "Sentiment Analysis from User-Generated Reviews of Ride-Sharing Mobile Applications," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), 2022, pp. 738-744, doi: 10.1109/ICCMC53470.2022.9753947.

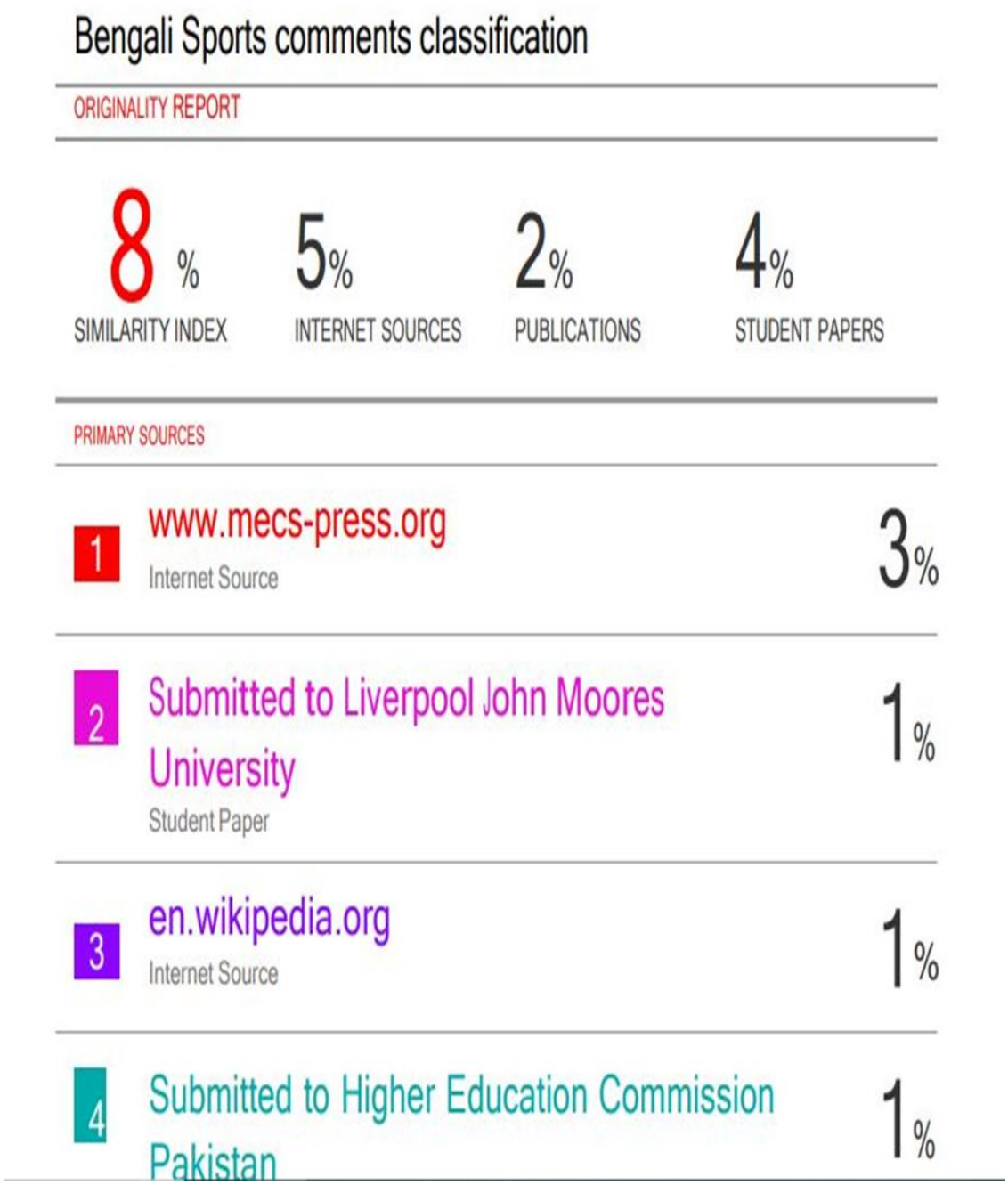
[25] D. O'Loughlin et al., "Sensitivity and Specificity Estimation Using Patient-Specific Microwave Imaging in Diverse Experimental Breast Phantoms," in IEEE Transactions on Medical Imaging, vol. 38, no. 1, pp. 303-311, Jan. 2019, doi: 10.1109/TMI.2018.2864150.

[26] U. Saha, M. S. Mahmud, M. Keya, E. A. E. Lucky, S. A. Khushbu, S. R. H. Noori and M. M. Syed, "Exploring Public Attitude Towards Children by Leveraging Emoji to Track Out Sentiment Using DistilBERT a Fine-Tuned Model", Third International Conference on Image Processing and Capsule Networks. ICIPCN 2022. https://doi.org/10.1007/978-3-031-12413-6_26

[27] Learn about Wikipedia in LSTM, available at <<https://en.wikipedia.org/w/index.php?title=Special:CiteThisPage&page=Long_short-term_memory&id=1134018038&wpFormIdentifier=titleform>> last accessed on 06-06-2022 at 12:00 PM.

[28] Learn about Precision, Recall, f1 score, macro average definition, available at <<https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f>> last accessed on 15-07-2022 at 9:00 PM.

[29] Learn about Ethical aspects, available at <<https://www.scribbr.com/methodology/research-ethics/#:~:text=What%20are%20ethical%20considerations%20in,for%20harm%2C%20and%20results%20communication.>> last accessed on 06-08-2022 at 11:00 PM.



Bengali Sports comments classification

ORIGINALITY REPORT

8%	5%	2%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mecs-press.org Internet Source	3%
2	Submitted to Liverpool John Moores University Student Paper	1%
3	en.wikipedia.org Internet Source	1%
4	Submitted to Higher Education Commission Pakistan	1%