Sentiment Analysis of Bangla Text Using Deep Learning

BY

Md. Abu Shaleh

ID: 171-15-8748

Md. Nazmul Hasan
ID: 171-15-9557

Ashibur Rahman

ID: 171-15-8763

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

Supervised By
Sheikh Abujar
Senior   Lecturer
Department
of CSE
Daffodil International
University

Co-Supervised By
Ms. Afsara Tasneem Misha
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL

UNIVERSITY DHAKA, BANGLADESH

APPROVAL

This Project/internship titled "**Sentiment Analysis of Bangla Text Using Deep Learning**", submitted by Md. Nazmul Hasan, ID : 171-15-9557, Md. Abu Saleh, ID : 171-15-8748, Ashibur Rahman, ID : 171-15-8763 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 28 January,2021.

BOARD OF EXAMINERS

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

Abdus Sattar   **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Jueal Mia   **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Dewan Md. Farid   **External Examiner**
Associate Professor
Department of Computer Science and Engineering
United International University

i

DECLARATION

We declare that, the project has been completed by us, under the supervision of **Sheikh Abujar, Senior Lecturer and Ms. Afsara Tasneem Misha, Lecturer, Department of C SE,** Daffodil International University. We also notify that this project has not submitted anywhere for the define of any degree program.
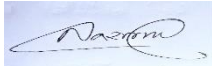
Supervised by:

Sheikh Abujar
Senior Lecturer
Dept. of CSE
Daffodil International University

Co-Supervised by:

Ms Afsara Tasneem Misha
Assistant Professor
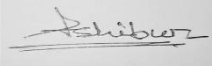Department of CSE
Daffodil International University

Submitted by:

Md. Nazmul Hasan
ID: 171-15-9557
Department of CSE
Daffodil International University

Md. Abu Shaleh
ID: 171-15-8748
Department of CSE
Daffodil International University

Ashibur Rahman
ID: 171-15-8763
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

We want to begin with expressing thankfulness to the Almighty for His blessing to perm it us to culminate effectively of the final year project phase.

We would like to thank our supervisor, **Sheikh Abujar** , for his appropriate direction in completing this research about work on the Sentiment analysis of Bangla Language. His support and direction given us with the certainty to complete the research accurately. He given us with all the related assets and significant data to do this Bengali Language.We also thank **Ms. Afsara Tasneem Misha** , our cosupervisor, for helping us in completing this job. We are earnestly thankful to, our department head, for his important offer assi stance in doing this kind of research about work within the Bengali Language.We would moreover like to thank other workforce individuals and our department's representatives for their Support. We are exceptionally thankful to our DIU NLP and Machine Learning Research Lab for giving us with the assets to complete this research work.At last, we t hank and appreciate our family and well-
wisher for their support and inspiration for completing this research work.

ABSTRACT

Sentiment analysis is a set of instructions to define and derive opinions. Most of the Sentiment Analysis works centered on the English language. Bangla is widely spoken language, approximately 265 million of people uses this language for communication. In pursuance of English language, the huge and standard datasets are not found for Bangla Language. For the lack of standard datasets, the research works for Bangla has been shorten in pursuance of other languages. To recognize the sentiment from Bangla texts is the main target of this research work. In this research work, we have tried to build such a system, which will recognize the sentiments from Bangla texts. It is very challenging to work on such a language which does not have enough standard data. In the end, we have successfully got the sentiments from the Bangla texts. The datasets were collected from social medias and Bangla blogs. Data were collected and trained the data to the system. Lastly, the system shows the expected result which presented into the research. Working with Bangla text is a quite challenging job, there are not enough resources or data sets available in Bangla. Collecting data was a huge challenge for us. We try to collect data from different sources like newspapers, blogs, some social media comments, and status.

TABLE OF CONTENTS

CONTENTS   **PAGE**

CHAPTER

CHAPTER 1: INTRODUCTION   **1-5**

CHAPTER 2: BACKGROUND STUDIES   **6-10**

CHAPTER 3: RESEARCH METHODOLOGY   **11-20**

LIST OF FIGURES

LIST OF ABBREVIATION

NLP   Natural Language Processing

RNN   Recurrent Neural Network

NLTK   Natural Language Tools Kit

LSTM   Long Short Term Memory

CHAPTER 1

Introduction

1.1 Introduction

Bangla is in the 7th position among 100 languages. According to the statistics of 2011 census, about 98% of Bangladesh's population and it too serves as the national dialect of the country Bengali is talked. Bangladeshi individuals are found to communicate with Bangla progressively included in online platforms such as - getting associated to companions and families through social media, define their thoughts on well-known micro-blogging and social organizing locales, sharing suppositions and considerations by implies of comments on online news entries, doing online shopping through online marketplaces and many more. But it's been difficult for all to analyze and supervise the sentiments of individuals. It's very necessary for analyzing the markets and trends for digital connectivity.

Sentiment Analysis is broadly utilized as a machine learning application in numerous ranges, and is known by numerous other terms. Like-supposition extraction, opinion mining, conclusion mining, subjectivity investigation, feeling investigation, survey mining, etc. There are many sentiments analysis research work has done previously in different languages, whereas Bangla Sentiment Analysis is still in a developmental state.

So, sentiments analysis from Bengali texts can be imperative part here for upgrading effectiveness and efficiency. In this time, we have attempted to construct a system on Bengali language which is able to get us the feelings from the writings. Machines can not provide the 100% of accuracy all the time. All the test results are not 100% accurate. But, for the Bengali language, the tests outputs were appeasement.

1.2 Motivation

Social media have played a critical role in the past few years. More than 3.6 billion people were using social media in 2020, according to Statista. And in 2025, up to 4,41 billion are expected to be used. The internet was protected by a vast amount of details. In 2020 36 million people used social media, according to the Data_Reportal. This will contribute richly to the Sentiment Analysis, using these results. Bangladesh is our mother tongue, much like Bangladeshi. We use it to connect and share our emotions. As Bangladesh's feelings analysis still evolve, these data will play a significant role in Emotion Analysis by using these data from a social network. However, we would not miss out on feeling research for Bangla language, if it can be used for multiple reasons, including commercial purposes. If we can continue working with Bengali, so many researchers can benefit from this in our region. In comparison, we have a large number of social media consumers.

Due to the pandemic scenario, even Brands have moved online and emphasize feeling research. Clearly, the way to read people's thoughts on their goods is their preferred way. Focus on social media can allow companies to recognize their consumers in an area that eventually improves their business. Sentiment Research helps brands to answer the exact problems or concerns of their consumers. According to some analysts, sentiment analysis of Twitter data may help to estimate stock market movements. Research reveals that news and social media will profoundly affect the stock market. It was noticed that, only for a short amount of time, reports of the strong general mood referred to a substantial price surge. In the other hand, derogatory information is correlated with a price fall, but has broader implications. In order to accomplish their targets, sentiment assessments should ideally be used by any brand:

1. Target persons to strengthen their programs.

2. Track the thoughts and reactions of customers over time.

3. Determine the type of consumers where a brand feels best.

4. Follow up changes in customer behaviors due to changes in goods.

5. Study the corresponding promoters and detractors.

Of necessity, sentimental research provides an organization with ample essential consumer knowledge. Organizations are now modifying their campaign strategies by responding to them by consumers. Sentiment Research helps businesses to quantify and refine their marketing strategies. Since sentiment analysis helps businesses to easily look at their customers' emotions, they can consider it and fix any challenge that arrives in good time.

There is an imperative part for enhancing quality and efficacy of sentiments research from Bengali texts here.

1.3   Rational of the study

Bangladesh's history is really rich. Millions of people today use Bengali as their mother tongue. But Bengali's tools and technologies are not rich like other languages in this modern era. Therefore, infrastructure for that language must be improved. The NLP methods and strategies can solve much of the text-based problems. Analysis of the sentiment in Bengali is a big NLP issue. It includes numerous text strings. It makes people quickly grasp with fluentness and error the importance of anyone's feeling and long email. For other languages, such as English, French, most and significant NLP techniques are already developed. But a few versions were made for the Bengali texts that are not adequate. The Bengali NLP field of study must also be extended. Preprocessing is the biggest barrier to Bengali text. Any Bengali characters and symbols cannot be understood by the engine. The Unicode of these characters or symbols must be used for this issue. For Bengali language, the NLTK library cannot be used.

That is why the methods in Bengali do not function precisely as other languages do. The best way to solve these kinds of problems is through study. So, in the course of this study, we try to demonstrate how Bengali is processed and how we can make a synthesis of

3

abstract text. This helps us to reduce the paper size and to present a fluid shooting time overview.

1.4 Research Questions

- What is Bengali Sentiment analysis?
- How Sentiment analysis works in Bengali?
- What are the benefits of Bengali Sentiment analysis?
- What are the problems of Bengali Sentiment analysis?
- How is Bengali text preprocessed in NLP?
- What are Bengali's future research projects?
- How does the Model Bengal function?

1.5 Expected Output

Since this is research and our goal is to go for Publication after getting a good accuracy, so we have planning to publish no research paper in this related field with the Bengali language. Many people are analyzing the specific topic and trying to find efficient results according to the Bengali language. But as it is complicated to find a polished and fine data set for the Bangla language. As it is difficult to find data set for those who are working with the Bengali language they are not interested to share their data set with everyone, as a result, there is a lot of opportunity for research but lack of data set we cannot do the research works.

If we want to execute those research works we have to create our own dataset some people may create similar kinds of data sit for their work because they don't get any kind of resource publicly. As we're trying to do sentiment analysis in the Bengali language, therefore a machine needs to learn about Bengali language for getting the pattern of the Bengali language. After getting the pattern of the Bengali language then the machine can guess the sentiment from any kind of Bengali sentence. With proper data set machine can be trained easily and properly so that when we want to execute all work it will give us better results. So our expected output will be we will train our machine with proper data set and after training machine should get the accurate sentiment or emotion from any kind of Bengali statements.

CHAPTER 2

Background Studies

2.1 Introduction

Feeling analysis is a phrase that refers to the use of linguistic processing, analysis of texts and machine language to determine a speaker's or writer's attitude towards a particular subject.

In essence, it helps to see how a text reflects positive, negative or neutral emotions. Sentiment analysis is a fantastic way to figure out how people, especially consumers, feel about a certain issue, product or concept. When emotion analyzes were used mostly in written paper papers, they can be traced to the roots of feeling analyses in the 1950s. However, sentiment analysis is used commonly to mine subjective knowledge, such as emails, tweets, blogs, social media, press releases, ratings and feedback from Internet contents. This is accomplished by a number of approaches, including NLP, statistics and machine learning. Organizations then use the information they collect to find potential prospects to help direct their messaging on their target audience. The Obama Administration also forecasts public reaction to its policy announcements using opinion analyses. Attitudes, perceptions and feelings are linked to sentiments. That is, instead of empirical truth, they are subjective experiences. The different forms of emotion analysis are used to classify the emotions found in a given text through various methods and techniques. We strive to find those kinds of feelings. Like: Happy, Sad, Disgust, Anger etc.

6

Figure 2.1.1: Sentiment analysis

The progress of the World Wide Web has contributed to multimedia content in Bengali communicating a sense of content creators from mixed viewpoints. Due to its intrinsic complexity, research on sentiment analysis in Bengali continues to be insufficient and is thriving as a demanding field and drawing researchers at a speedy pace. Researchers spend much time reviewing previously published works in order to understand improvements towards future improvement; one of the most stressful and demanding facets of study in this field. Pre-processing measures along with current research methodologies, data sets and appraisal metrics of outstanding works on sentiment analysis of the Bengali language as our research findings as well. Different Bengal declaration that contains positive, negative and neutral emotions. Sentiment analyzes are used to explore human feelings found in documents. We will demonstrate a deep learning implementation for news headline sentiment analysis. The experiment was conducted in the Bengali Statement dataset, which demonstrates how the adopted method could be extended and tested.

2.2 Related Work

As we are trying to create a sentiment analysis we have learnt a lot of things from the paper called An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques.[1] by R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter and A. K. Das. Try to make an automated system to analyse sentiment from Bangladesh test and they used supervised learning techniques.

Orissa child used romanized Bangla text for the deep recurrent model for sentiment analysis [2] by A. Hassan, M. R. Amin, A. K. A. Azad and N. Mohammed.

Though we used LSTM some research are used RNN technique for sentiment analysis for Bangla text [3]. A researcher uses machine learning process to analyse sentiment and design and empirical framework[4].

We have season long short term memory model who used cricket commentary and cricket related Bangla text sentiment analysis[5]. A researcher named M. Eunus Ali tries to detect multi-level sentiment and emotion from Bangla comments of different YouTube videos.[6] we also see some research works based on in gram sentiment using support vector machine for Bangla text[9].

2.3 Research Summary

In our research, we have introduced the Bengali sentiment analysis. We are planning to use the planning techniques for executing our research work. We have collected some Bengali news, some social media status, social media comments, different Bengali statements, some lines from different books are collected for our research purpose. After taking some data from different resources we have to filter the data. By filtering means removing some stop words, special characters, numbers. There are lots of library functions and NLTK (natural language Toolkit) libraries for or English and other languages. But there are not developed libraries available for the Bengali language. As we are using our data set with the Bengali language so we have to change some libraries according to our research work. There is some research work that existed on Bengali language but those researchers do not give access to their data set or

how they are preprocessing their data. As lack of enough prebuilds library in NLP and NLTK we have to face a lot of challenges. We tried to do our best to purify our data. In our research, we try to introduce a deep learning technique to find out the sentiment from any Bengali statements.

2.4 Challenges

Different Bengal declaration that contains positive, negative and neutral emotions. Sentiment analyzes are used to explore human feelings found in documents. We will demonstrate a deep learning implementation for news headline sentiment analysis. The experiment was conducted in the Bengali Statement dataset, which demonstrates how the adopted method could be extended and tested.

Another difficult work is the text data to forecast the feeling of the text. It is still difficult to deal with the Bengali text. For Bengali text processing the NLTK Library is open. Thus, in the preparation stage, the text as an input of a model needs to be prepared with raw coding. Assume the Unicode of each punctuation and delete it by raw coding is required when extracting the punctuation of the text.
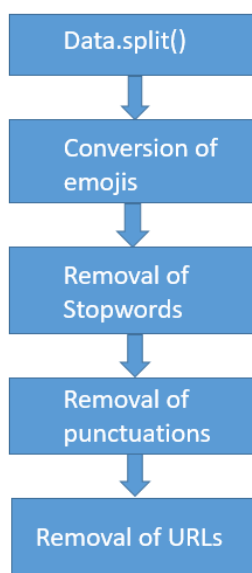
An additional challenge is avoiding word elimination. In other languages, such as English, a built-in library can be used to remove text words. But to capture the word type for the Bengali language online, place them in a text file and erase words from the text using that file. Another difficulty in this analysis is a wide vocabulary. If the data collection includes a large amount of data, it will provide a large vocabulary and help to forecast the feeling accurately.
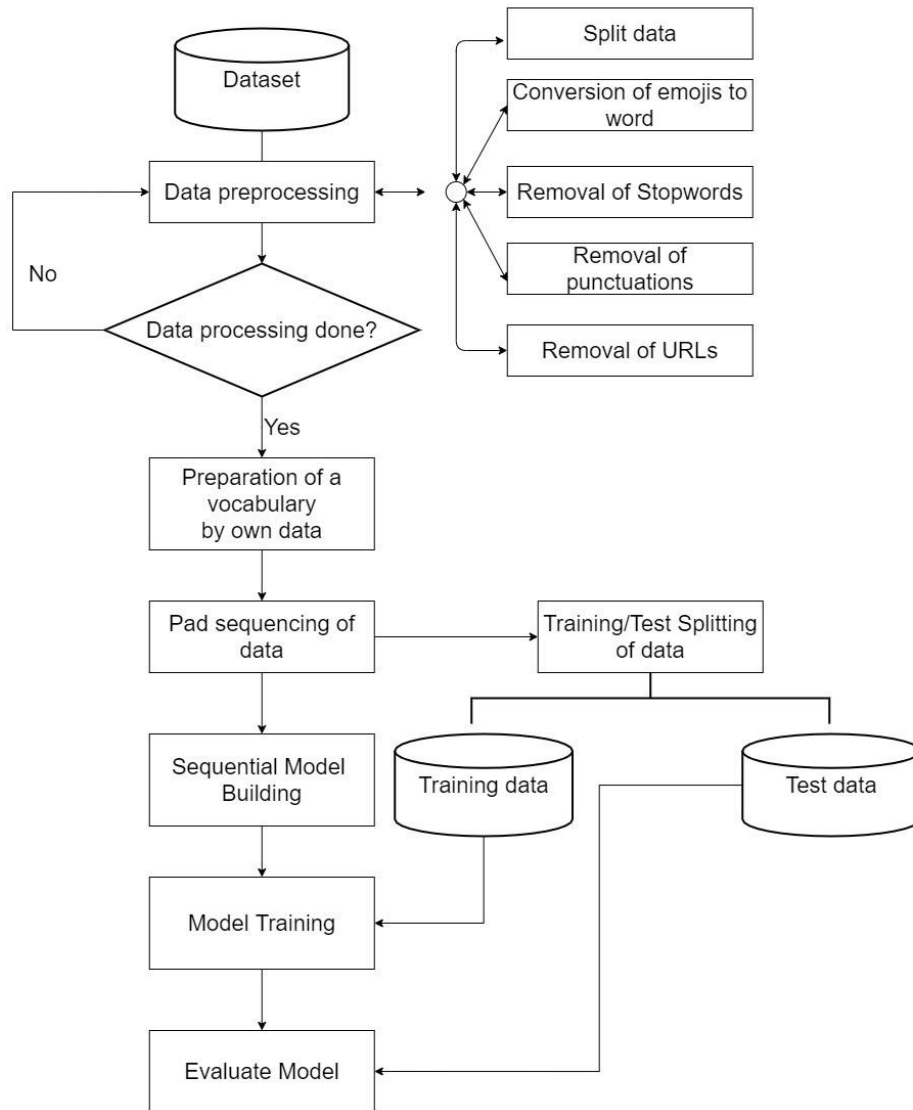
CHAPTER 3

Research Methodology

3.1 Introduction

In this segment, we will address the whole technique of the research work. Each analysis has a particular method of solving. The technique component involves the introduction of both methods. Here you can analyze the application with a concise explanation of each specific aspect of the approach using the model. We used a deep learning model in our research to examine our Bengali sentiment. Deep learning algorithms are used according to the type of analysis they contain. LSTM is used to solve text-related problems in deep learning. Any model of deep learning requires a good data set to find an effective outcome. The dataset algorithm must be compiled and pre-processed before it is implemented. Each part of the technique is addressed separately in detail. Provided that all portions are followed until the analysis is done. Better clarification of the technique would improve the productivity of the job and offer the nobility. In this part, all the steps of the technique are briefly discussed. The workflow of the entire research work is given below, which provides a brief summary of the overall research work.

```
┌─────────────────────┐
│     Data.split()    │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Conversion of    │
│       emojis        │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│     Removal of      │
│      Stopwords      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│     Removal of      │
│     punctuations    │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   Removal of URLs   │
└─────────────────────┘
```

Figure 3.1 Workflow for Bangla sentiment analysis

3.2 Research Subject and Instrumentation

In the first place, we have discussed the process of Sentiment Analysis in Bengali with the conceptual and theoretical process. Deep learning model needs high configuration pc with GPU and others instrument. Now a list is given below of the required instrument for this model.

Hardware and Software:

- Intel Core i3 7th generation
- 1 TB HDD
- 8GB RAM

Development Tools:

- Windows 10
- Python 3.8
- Tensorflow Backend Engine
- NLTK
- BNLP/BNLTK
- Pandas
- Keras
- Numpy

3.3   Data collection and Data preprocessing

We have collected power Bengali statement data for research purpose. All data collected from the different type of social media comment, some news portal, some social media status, different Bengali blogs. As we know that different researchers working on Bengali statements but they do not share their own data set. As they do not share their data set everyone had to make their own dataset for this it takes time to do any kind of research on Bengali statements. As we have collected raw data with various types of special characters and emojis to we need to preprocess our data set. There is a diagram given below, on the diagram step by step data preprocessing systems are explained.

3.3.a Data Split

After collecting data, we have to split our data set, some long sentences should be processed and minimized. We collect different Bengali statements then we give labels like which Bengali statement is representing what kind of sentiment or emotion. Splitting data set is also helpful to create vocabulary that will be helpful for NLP. This vocabulary is more relevant with word embedding.

3.3.b Conversion of Emojis:

As previously we have told that we have collected data from different sources like social media comments and status, different kind of books and blogs. So, people sometimes use emojis to express their emotions. Sometimes people use emojis between their lines or use only emojis to express any kind of emotions. So, emoji conversion is a major part for our research work. Because if we cannot convert those emojis or some emojis may not Express the emotions properly so we have to remove them for convert them to corresponding sentiments. If someone is showing happy sad for Angry kind of emoji. That can be converted for our research. But if anyone is using different type of emojis we should remove them because not all the sentiments can be covered in one research.

Fig 3.3.2: Conversion of Emojis

3.3.c Removal of punctuations

The removal of punctuations in Bangla language is very important for our research. As weird getting different kind of statements write different social media comments and status so removing punctuation marks is mandatory. There are almost similar kind of punctuation marks in Bengali language like English. Apart from that there are very few special punctuation marks which is dedicated to Bengali languauge so we have to remove that kind of punctuation mark to purify our data. Removing punctuation is a very important part of data preprocessing. Is the punctuation marks are not removed when we train our machine or our model it will not predict accurately. We are using deep learning so we try to train as a pattern so punctuation is important.

3.3.d Stop word remove

Remove stop word text is a popular NLP operation. The key use of the stop word is to delete the redundant word from the language. Build-in the NLTK library to remove the word stop from the English text. But there is no library available for Bengali to remove the word stop. So, first of all, we're gathering all the stop word from online in the Bengali language. There is a cumulative number of 393 stop words, and then we inserted it in a register for further use. We have already discussed about the shortage of Bengali language data sets. So, stopping stop words have to be done by us. And the stop words of Bengali language are different from English. So, we cannot use any kind of Pre build libraries to remove the stop words.

3.3.e Removing URLs

As we have already told that we are collecting data from different kind of social media comments and status, so someone Mein share different kind of URLs in their Bengali statements. So, it is quite possible that data can be trained with different kind of URLs if we do not remove it. So, for our research purpose removing URL is mandatory. As we have already removed all the stop words and punctuation marks, if we remove the URLs also it

will give us cleaner look which will be very helpful for train our data set to any kind of model using deep learning techniques. So, removing URL can give us more processed the data and it may show us good accuracy.

3.3.f Purified text & summary

After following all the steps of data preprocessing like data split, conversion of emojis, removing stop words, removing punctuation marks finally the data we get is pea processed or purified data. These preprocess data will look clean than the original data. The original data may contain different kind of things but after pre-processing in gets a clean look which is easier to be trained for machines. So, the preprocess data is actually what we are giving to the machine to be trained. And the machine will be trained with the pattern to recognize any kind of statements with all the things.

Before preprocessing	After preprocessing
বানানে এতো ভুল। খুবই সাধারণ শব্দগুলোতে এরকম ভুল বানান করা কাম্য নয়।	বানানে এতো ভুল খুবই শব্দগুলোতে এরকম ভুল বানান কাম্য
আরে শালার 😏 বদমাইশ।তুই এই জঘন্য কাজ কেন করলি?	আরে শালার রাগান্বিত বদমাইশ তুই জঘন্য করলি
অপ্রাসংগিক, তবুও খুব পছন্দের একটা পেইজ বলে জানতে চাচ্ছি এরকম টা হওয়ার কি কারণ থাকতে পারে, বা এরকমটা সত্য কিনা। https://www.facebook.com/photo.php?fbid=10153832472312269&set=a.10 150251565832269.374678.669192268&type=3	অপ্রাসংগিক তবুও পছন্দের একটা পেইজ চাচ্ছি এরকম টা হওয়ার থাকতে এরকমটা সত্য কিনা
যদি আপনি আপনার কাজকে স্যালুট করেন , তবে আপনাকে আর কাউকে স্যালুট করতে হবে না। আল্লাহ আপনাদের মঙ্গল করুক। আপনাদের বাবা মা কে ধন্যবাদ আপনাদের মত সুসন্তান কে এই পরিত্যক্ত বাংলায় উপহার হিসেবে দেয়ার জন্য।	কাজকে স্যালুট আপনাকে স্যালুট আল্লাহ আপনাদের মঙ্গল করুক আপনাদের বাবা মা ধন্যবাদ আপনাদের মত সুসন্তান পরিত্যক্ত বাংলায় উপহার হিসেবে দেয়ার
রহিম সাহেব হঠাত করে মারা গেলেন ।	রহিম সাহেব হঠাত মারা গেলেন
সে আমাকে মারার হুমকি দিয়েছে।আমি কি করব বুঝছি নাহ।	মারার হুমকি দিয়েছে করব বুঝছি নাহ
আমি পরিক্ষায় প্রথম স্থান অর্জন করেছি।আমি আনন্দে দিশেহারা।	পরিক্ষায় স্থান অর্জন করেছি আনন্দে দিশেহারা

Fig 3.3.3: Before and After Pre-processing data

3.4 Statistical Analysis

1. We use small numbers of data. These data have 2 subsections such as Text and Sentiment. A short view of our dataset is given below.

Text	Sentiment
বানানে এতো ভুল। খুবই সাধারণ শব্দগুলোতে এরকম ভুল বানান করা কাম্য নয়।	disgust
আরে শালার 😡 বদমাইশ।তুই এই জঘন্য কাজ কেন করলি?	angry
অপ্রাসংগিক, তবুও খুব পছন্দের একটা পেইজ বলে জানতে চাচ্ছি এরকম টা হওয়ার কি কারণ থাকতে পারে, বা এরকমটা সত্য কিনা। https://www.facebook.com/photo.php?fbid=10153832472312269&set=a.10150251565832269.374678.669192268&type=3	surprise
যদি আপনি আপনার কাজকে স্যালুট করেন , তবে আপনাকে আর কাউকে স্যালুট করতে হবে না। আল্লাহ আপনাদের মঙ্গল করুক। আপনাদের বাবা মা কে ধন্যবাদ আপনাদের মত সুসন্তান কে এই পরিত্যক্ত বাংলায় উপহার হিসেবে দেয়ার জন্য।	happy
রহিম সাহেব হঠাত করে মারা গেলেন ।	sad
সে আমাকে মারার হুমকি দিয়েছে।আমি কি করব বুঝেছি নাহ।	fear
আমি পরিক্ষায় প্রথম স্থান অর্জন করেছি।আমি আনন্দে দিশেহারা।	happy

Fig 3.4: Dataset Preview

2. The total size of the vocabulary is 50000.
3. The total number of the unique word of the dataset is 3671.
4. 94% of the word we used in our model.

5. Dataset is saved in excel file which extension is .xlsx.

3.5 Implementation Requirements

3.5.a. Preparation of a Vocabulary

As we already know that the dataset of Bengali language is limited. So, there is not enough vocabulary available. So, we have to create our own vocabulary set. First we need to take the Bengali words from a data set and with Bengali words we have to give a numeric value or we can say a we created or own vocabulary because machine cannot understand any word for the meaning of those words. So, what we have to do in we have to generate and do some indexing of the Bengali words and use does indexing for different deep learning techniques. Creating our own vocabulary is one of our huge challenge. As we have already processed our data so our data have more polished and cleaner look for machine to prepare vocabulary.

3.5.b. Pad sequencing of data

The concept of pad Sequencing is to convert all the sequences to the same length. In deep learning technique pad Sequencing is a major issue. Let us consider a situation like if my data set have different lengths of statements, sometimes it need decrease the accuracy while training the machine. So, if we do pad sequence, we can create all the sequence of data with same length.

3.5.c. Sequential model:

Deep learning is an increasingly common machine learning subset. Deep learning models are developed using neural networks. The neural network takes inputs, which are then deal with in unseen layers using weights that are changed during preparation. Then the model is heaving out the estimate. The weights are modified in order to identify trends in order to allow better forecasts. The consumer does not need to decide what patterns to search for—the neural network learns on its own.

Compiling the model is based on two constraints: optimization and failure. The optimizer monitors the learning proportion. We're using our optimizer. A good optimizer to be used in certain situations. The Optimizer changes the learning rate over the course of study. The learning rate determines how easily the optimal weights for the model are determined. A lower learning rate will lead to more accurate masses, but the time it takes to measure the weights would be longer.

3.5.d. Evaluation of model:

We have done our data preprocessing, we have created our sequential model, now we have to fragmented our data set into train set and test set. After piercing the data set then the machine will learn from the train set. From the information pleat from the train set it will be applied in the test set to get the exact outcome. So, the model will be used to train and test the data which are pre-processed.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

Sentiment analysis for Bengali statements a complex problem in natural language processing. The machine can take any Bangla text then it will realize what kind of sentiment it is expressing. Finding accurate emotion is quite difficult. The calculation of probability is quite hard enough. The machine can give the maximum probability on the basis of its training. After the preprocessing data, it needs to be trained on any kind of model for the learning purpose of a machine. For training, we have used some backend engine in the deep learning model. In the experiment, we used TensorFlow 1.15 for a backend engine. For Train, we need some parameters to define. The parameters are epochs, batch size, number of layers, etc.   And those parameters are dependent. Reducing the loss of train time is very important. One of the important things of research is reducing the loss of training time. So if we can reduce the training time it will show the accuracy faster that will be most optimized for research work. We used jupyter notebook for writing our co code, and we need a very good GPU surface in our machine that will reduce our training time. If we use some parameters to our model and our machine is powerful enough it works properly.

- Define Epochs = 30
- Define Batch size = 32
- Set the RNN cell size = 256
- Set the number of neuron layer = 64
- Keep the probability rate is = 0.58 to 0.65

4.2 Experimental Results

We know no machine can give us 100% output. Similarly, we can train our model and change some parameters to get more accuracy and train properly. Most of the output is nearly enough to the actual output. But sometimes it gives wrong results. But the maximum number of responses is recording to the statements. Like it can predict the right emotion or sentiment according to the statement. We have used 30 epochs the loss of data became reduced. We are checking the output more and more. We use TensorFlow and matplotlib to generate some graphical view to check how is the accuracy and other things are going on. The sequential model is good enough to give us good accuracy. We will give some graphical view and some photos so that anyone can relate how our model and deep learning technique is working.

As we are taking 6 classes for our sentiment analysis which are: happy, sad, surprise, angry, disgust, fear. Our machine is predicting these 6 classes from the data set which we provided Bangla statements or Bangla text from different sources. It will analyze the Bangla text and show the corresponding statement for emotion.

There are some photos given below which is showing our research work in a short form. These photos actually represent or summarize our work.

```
1    93
3    88
0    77
4    64
2    47
5    40
Name: Sentiment, dtype: int64
```
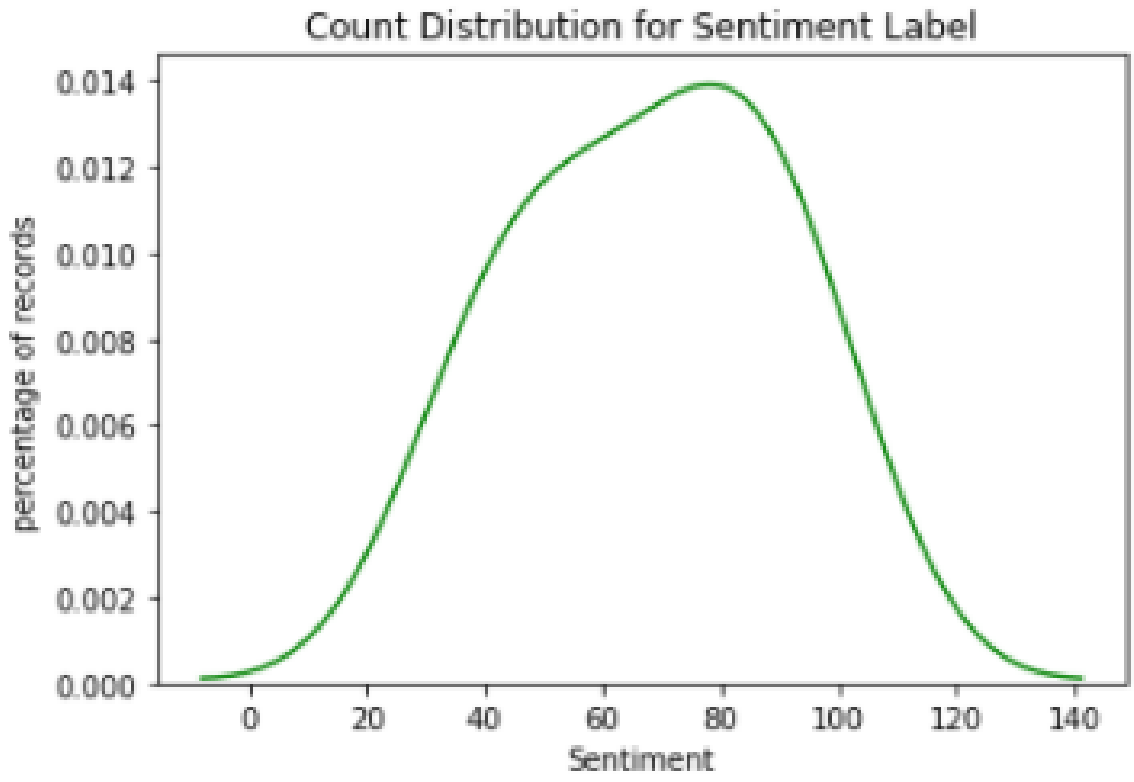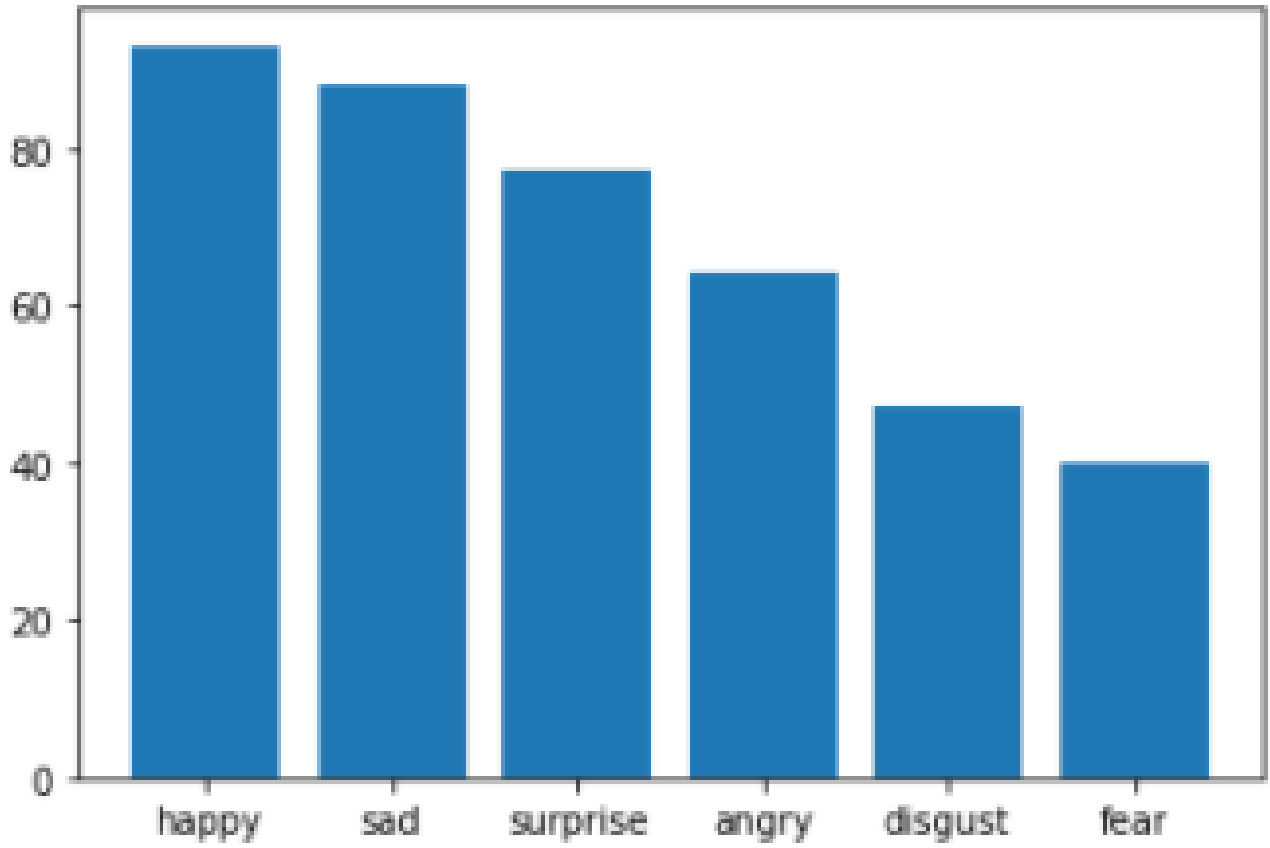

Fig4.2.1a : Label Distribution
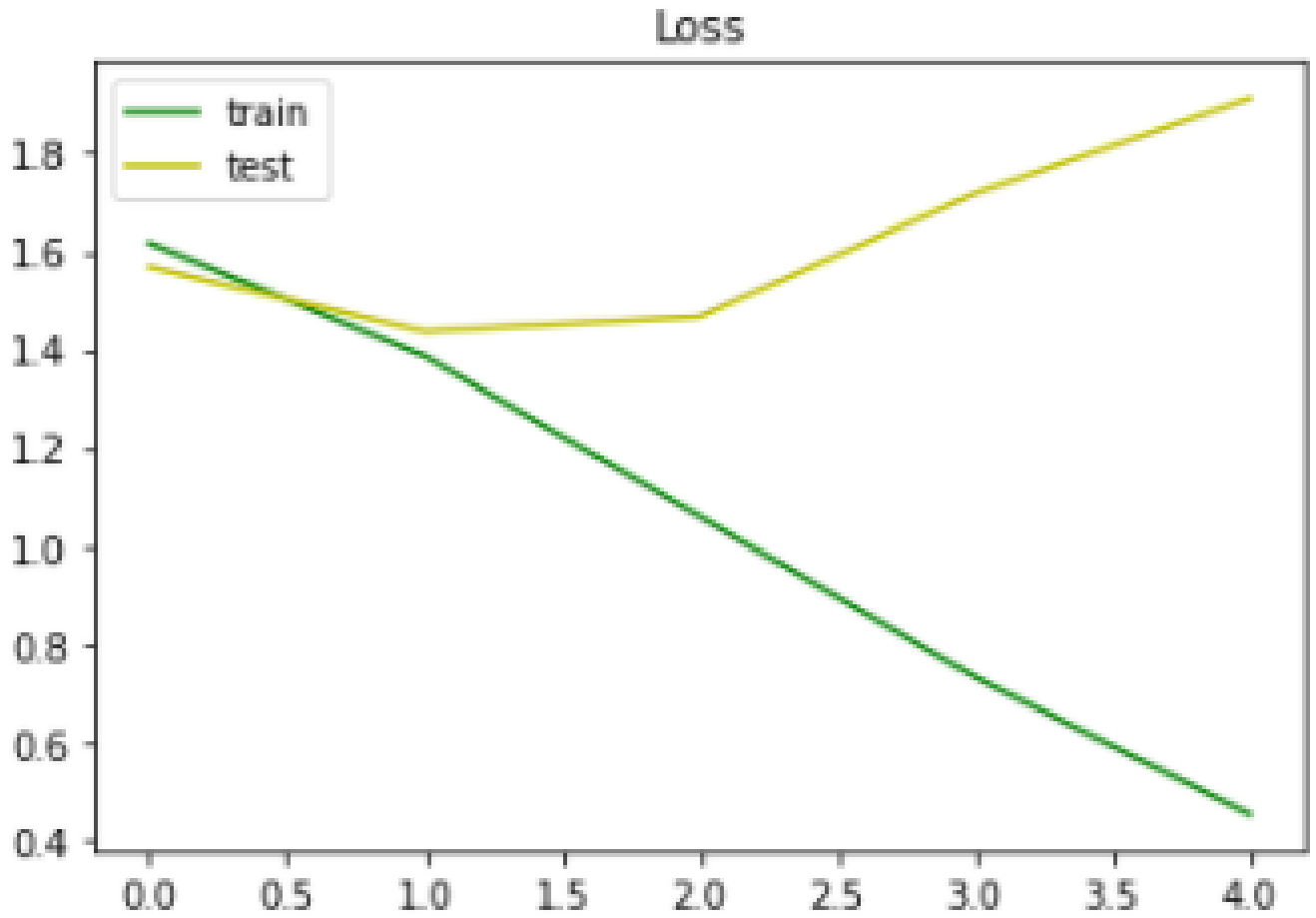
Fig 4.2.1b : Label distribution bar graph

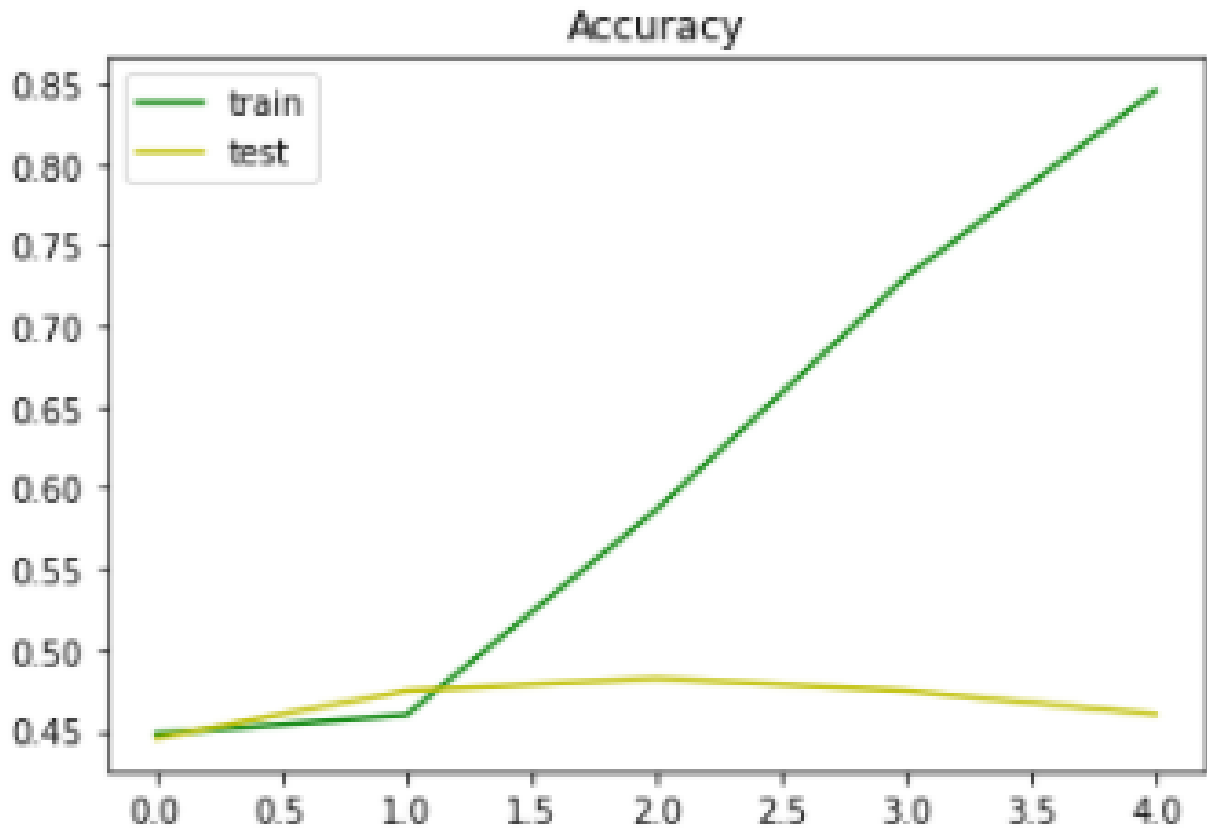Fig 4.2.2 : Loss graph of train and test set

Fig 4.2.3 : Accuracy graph for train and test set

The graph of loss and accuracy of train and test set actually showing how much train data is accurately trained and tested, and how much data is lost why train and test set was running.

```
Model: "sequential_1"
_____
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 250, 100)          5000000

lstm_1 (LSTM)                (None, 64)                42240

dense_2 (Dense)              (None, 256)               16640

dropout_1 (Dropout)          (None, 256)               0

dense_3 (Dense)              (None, 6)                 1542
=================================================================
Total params: 5,060,422
Trainable params: 5,060,422
Non-trainable params: 0
_____
None
```

Fig 4.2.4 : Model build

This photo is showing in the model in which we are using LSTM in the deep learning technique. Where we can see that we used some layers of embedding, LSTM, dense, dropout. These layers are used in long short-term memory to build our model and execute the model.

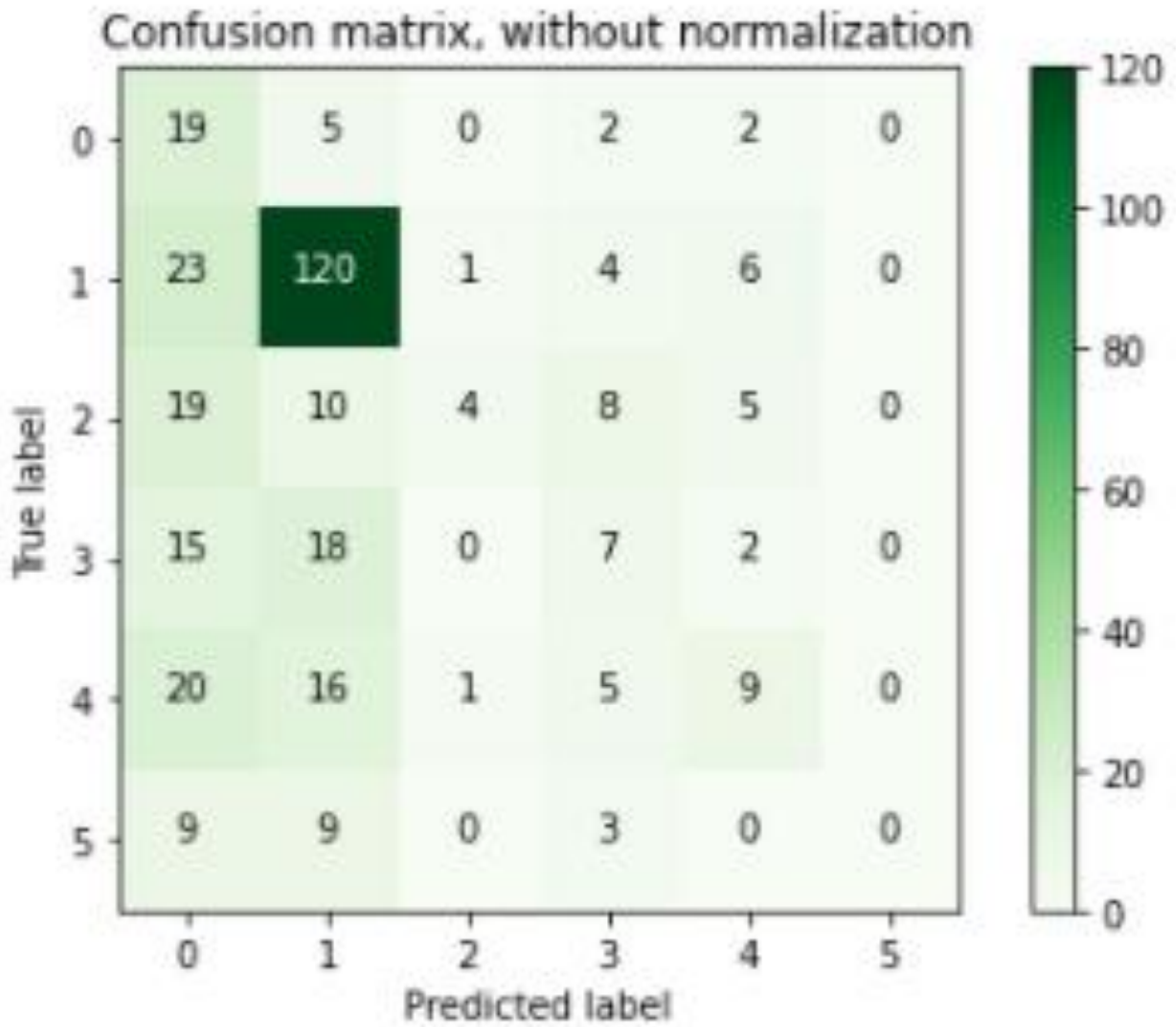|	precision	recall	f1-score	support
Class 0	0.18	0.68	0.29	28
Class 1	0.67	0.78	0.72	154
Class 2	0.67	0.09	0.15	46
Class 3	0.24	0.17	0.20	42
Class 4	0.38	0.18	0.24	51
Class 5	0.00	0.00	0.00	21
accuracy			0.46	342
macro avg	0.36	0.31	0.27	342
weighted avg	0.49	0.46	0.43	342

Fig 4.2.5: precision_recall_f1_support

Fig 4.2.6: confusion matrix before normalized

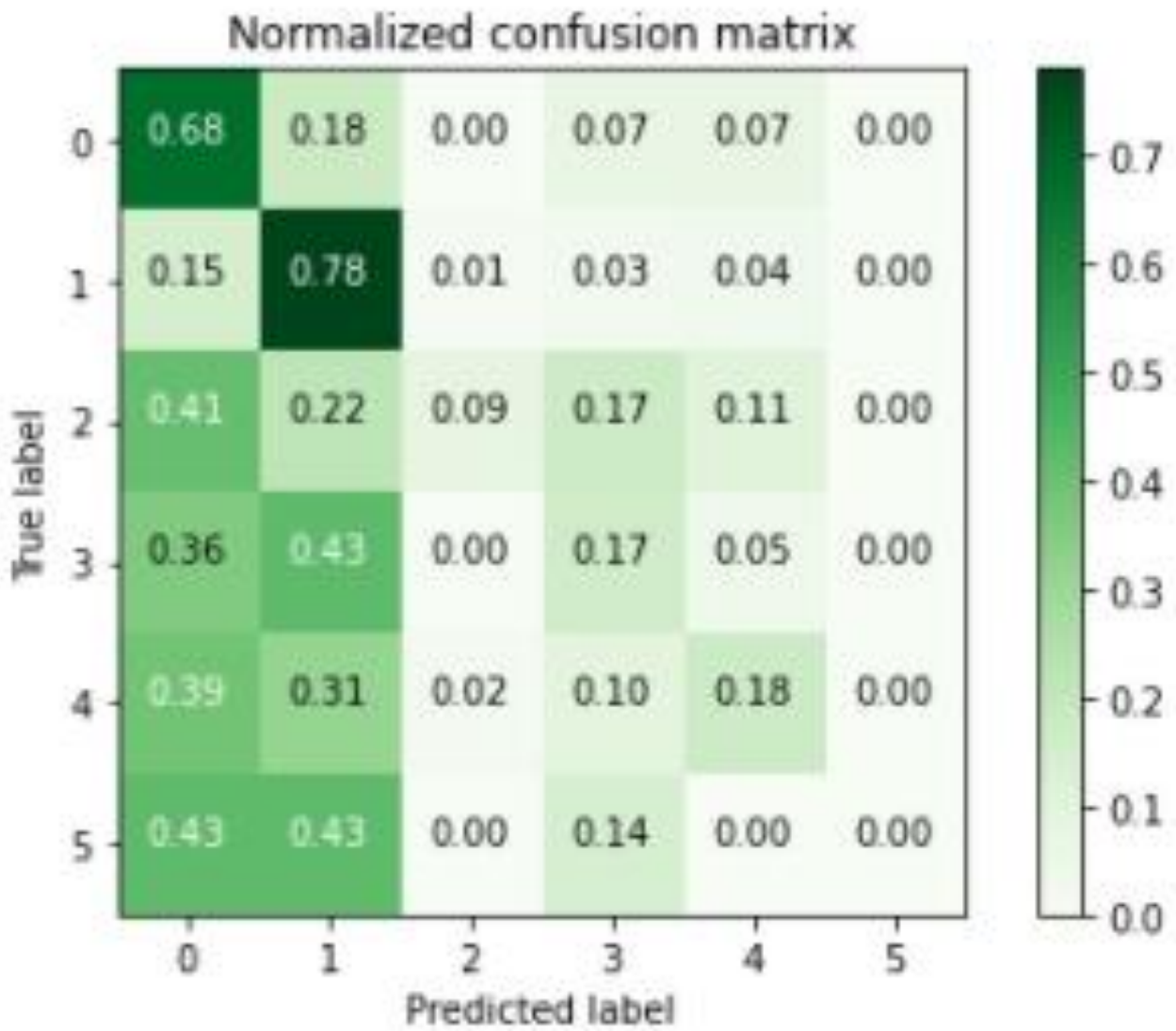Fig 4.2.7: normalized confusion matrix

4.3 Descriptive Analysis

Before making the Bengali text sentiment analysis we have tried something in English also because there are a lot of data set available in English so we tried that and get compress to make Bengali language sentiment analysis research work. Our model is giving us acquired good results according to our dataset. But as we have already discussed that Bengali text data sets are not available for all, so we have made our own data set. So there may be some options to improve our data set we will get a higher accuracy rate if we can polish our data save more and more. First, when we tried to use LSTM it was showing more data loss. So we tried to reduce the data loss. We also change some parameters to get a more accurate result. so We are using almost 80 to 90% data for training and 10 to 20% data for testing.

4.4 Summary

So, we can see here what previously we are talking about in those above points is to experiment and show how our model is working. An overall there are some opportunities to increase the accuracy of all work we are working on it. And we tried to reduce the training time and increase the response time of the machine.

CHAPTER 5

Conclusion and Future Work

5.1 Summary of the Study

Our whole project involves the Bengali NLP. In this project, we have developed a deep learning model for the Bengali Text Sentiment Analysis. This is very useful in analyzing thoughts and feelings. This project has been finished within 2 months. The entire project is split into sections. The complete description of the project is offered step by step below.

Step 1: Data collection form various

way

Step 2: Label all the collected data

Step 3: Data preprocessing

Step 4: Vocabulary creation

Step 5: Vocabulary count

Step 6: Load Dataset

Step 7: Pad sequencing

Step 8: Use LSTM

Step 9: Build sequential model

Step 10: Train model

Step 11: Check the result analysis the response of the machine

This model will help our Bengali NLP research group develop a fully-dependent Sentiment Analyzer and more research into the Bengali text. We will therefore discuss the future work and the end of this research report.

32

© Daffodil International University

5.2 Conclusion and Future work:

As we know there are limited resources of Bangla Natural Language Processing so we are trying to improve and develop the research area. Bengali text vocabulary and input model should be improved. We are trying to build and enrich the Bengali text data sets which will improve power research work and help other researchers to do their works. We already told that we have taken data from different resources it actually helps the machine to train with different Bangla text and recognize the pattern which will increase the accuracy of our work. We have a plan to increase the number of classes and emotions which will help us to you recognize any kind of sentiment or emotions which anyone wants to share. You can also train different types of emoji corresponding to their emotions. So, if anyone uses emojis that will also help us to train the model with emojis.

33

References

[1] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter and A. K. Das, "An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 360-364, doi: 10.1109/CCOMS.2019.8821658.

[2] A. Hassan, M. R. Amin, A. K. A. Azad and N. Mohammed, "Sentiment analysis on bangla and romanized bangla text using deep recurrent models," 2016 International Workshop on Computational Intelligence (IWCI), Dhaka, 2016, pp. 51-56, doi: 10.1109/IWCI.2016.7860338.

[3] M. S. Haydar, M. Al Helal and S. A. Hossain, "Sentiment Extraction From Bangla Text : A Character Level Supervised Recurrent Neural Network Approach," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465606.

[4] N. Tabassum and M. I. Khan, "Design an Empirical Framework for Sentiment Analysis from Bangla Text using Machine Learning," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ECACE.2019.8679347.

[5] 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), Sylhet, Bangladesh, 2019, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201500.

[6] N. Irtiza Tripto and M. Eunus Ali, "Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554875.

[7] A. K. Paul and P. C. Shill, "Sentiment mining from Bangla data using mutual information," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4, doi: 10.1109/ICECTE.2016.7879569.

[8] M. Asimuzzaman, P. D. Nath, F. Hossain, A. Hossain and R. M. Rahman, "Sentiment analysis of bangla microblogs using adaptive neuro fuzzy system," 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, 2017, pp. 1631-1638, doi: 10.1109/FSKD.2017.8393010.

[9] S. Abu Taher, K. Afsana Akhter and K. M. Azharul Hasan, "N-Gram Based Sentiment Mining for Bangla Text Using Support Vector Machine," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-5, doi: 10.1109/ICBSLP.2018.8554716.

[10] A. Rahman and M. S. Hossen, "Sentiment Analysis on Movie Review Data Using Machine Learning Approach," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201470.

[11] M. G. Hussain, T. A. Mahmud and W. Akthar, "An Approach to Detect Abusive Bangla Text," 2018 International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, 2018, pp. 1-5, doi: 10.1109/CIET.2018.8660863.

[12] O. Sharif, M. M. Hoque and E. Hossain, "Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934655.

15% SIMILARITY INDEX   15% INTERNET SOURCES   1% PUBLICATIONS   % STUDENT PAPERS

PRIMARY SOURCES

1   dspace.daffodilvarsity.edu.bd:8080
Internet Source   14%

2   clubs.ewubd.edu
Internet Source   1%

3   Nayan Banik, Md. Hasan Hafizur Rahman, Shima Chakraborty, Hanif Seddiqui, Muhammad Anwarul Azim. "Survey on Text-Based Sentiment Analysis of Bengali Language", 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019
Publication   <1%

4   Md. Tazimul Hoque, Ashraful Islam, Eshtiak Ahmed, Khondaker A. Mamun, Mohammad Nurul Huda. "Analyzing Performance of Different Machine Learning Approaches With Doc2vec for Classifying Sentiment of Bengali Natural Language", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019
Publication   <1%

36

5 hdl.handle.net
Internet Source

<1%

6 Hasmot Ali, Md. Fahad Hossain, Shaon Bhatta Shuvo, Ahmed Al Marouf. "BanglaSenti: A Dataset of Bangla Words for Sentiment Analysis", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020
Publication

<1%

| Exclude quotes | On | | Exclude matches | < 10 words |
| Exclude bibliography | On | | | |