Thesis Report
On

Bangla Newspaper Headline classification using Deep learning models LSTM, Bi-LSTM and Bi-GRU

Submitted by

SHOEB AKIBUL ISLAM
Roll : ASH1811022M
Session : 2017-2018

Thesis Supervisor

MOHAMMAD KAMRUL HASAN
Lecturer
Department of Information & Communication Engineering

Noakhali Science & Technology University
Noakhali – 3814
Bangladesh

Submission Date: 05 March 2023

DECLARATION

This research is submitted to the Department of Information & Communication Engineering, Noakhali Science & Technology University, Sonapur, Noakhali-3814, in partial fulfillment of the requirements for having the B.Sc degree in Information & Communication Engineering. So, I hereby declare that this research has not been submitted elsewhere for the requirement of any kind of degree, diploma, or publication.

Shoeb Akibul Islam
Student of Information & Communication Engineering
Noakhali Science & Technology University
Roll No : ASH1811022M
Session : 2017-18

ACCEPTANCE

This research is submitted to the Department of Information & Communication Engineering, Noakhali Science & Technology University, Sonapur, Noakhali-3814, in partial fulfillment of the requirements for having the B.Sc degree in Information & Communication Engineering. This research report will be evaluated under the course: Project and Thesis with course code IC-4218.

Mohammad Kamrul Hasan
Lecturer
Department of Information & Communication Engineering
Noakhali Science & Technology University
Noakhali – 3814, Bangladesh

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to almighty creator who has given me the strength and golden opportunity to do this thesis work – "Bangla Newspaper Headline classification using Deep learning models LSTM, Bi-LSTM and Bi-GRU". I also would like to thank my thesis supervisor, Mohammad Kamrul Hasan Sir for setting up and organizing this thesis work. Without his help, this thesis work would be very tough for me to carry on. I am very grateful for all the support that I received from him.

Besides, I would like to thank all my teachers of the Department of Information and Communication Engineering (ICE) in NSTU. Also, thank you for their valuable suggestion and academic help during the course work. Their continuous support and cooperation make it a success. I am also grateful to all my friends, who encourage me for doing this work. I am indebted to my parents for their encouragement and love that all this worthwhile.

Shoeb Akibul Islam
Student of Information & Communication Engineering
Noakhali Science & Technology University
Roll No : ASH1811022M
Session : 2017-18

ABSTRACT

Newspapers are a good habit to the full world of all ages. To accumulate a range of data from different segments is entertaining by themselves. Per this work using Bengali news headlines could also be more particular to others by defining its news type. Therefore the machine can energetically review the sequence of sentences close by output to seek out the news type. With the experiment, we connect our method by Neural Network of adoption. Coming with a momentous outcome we've done Multi Classification reached at Long-Short Term Memory (LSTM), Bi- Long-Short Term Memory (Bi-LSTM) and Bi- Gated recurrent unit (Bi-GRU) applying Bengali dataset. Dataset is recognized by collecting news headlines from various Bengali news portal and sites. Resultant output shows well performance in categorization. We trained and tested these models using news headlines from a well-known Bangladeshi newspaper as raw data. We divide the category of Bangla newspaper headlines into six categories. The Bi-LSTM Model achieves the highest validation accuracy of 97.96% and the lowest validation accuracy of 77.91%.

Keywords—[Deep learning, deep ensemble model, real news, news classification, machine learning, social newspaper, social media]

Table of Content

List of figures

List of Table

CHAPTER 1

INTRODUCTION

1.1 Introduction

News Data is one among the structured formatted data, which carries attributes like source, date, location, author, headline text, and detailed data, which is text formatted data. To extract features from the textual data, usually, the preference is for language Processing approach, which deals with all the language- and text-related aspects [1]. In recent years, there has been a growing interest in natural language processing (NLP) and machine learning techniques for text classification tasks. One such task is the classification of Bangla newspaper headlines, which is a challenging problem due to the complexity of the Bangla language and the diverse range of topics covered in newspaper headlines.

The goal of this thesis is to explore the problem of Bangla newspaper headline classification using deep learning models, specifically LSTM, Bi-LSTM, and Bi-GRU models. These models are widely used for sequence modeling and have shown promising results in various NLP tasks, including text classification. To achieve this goal, we collected a dataset of Bangla newspaper headlines and preprocessed the data to prepare it for model training. We then trained and evaluated the LSTM, Bi-LSTM, and Bi-GRU models on this dataset, using various techniques to improve their performance, including data augmentation and regularization techniques.

The main contribution of this thesis is to evaluate the performance of LSTM, Bi-LSTM, and Bi-GRU models on the Bangla newspaper headline classification task and to identify the strengths and weaknesses of each model. In addition, we aim to propose ways to mitigate overfitting, which is a common problem in deep learning models.In this thesis, we propose a methodology for Bangla newspaper classification using deep learning models such as LSTM, BI-LSTM, and Bi-GRU. The proposed methodology involves collecting and preprocessing a dataset of Bangla news articles, using pre-trained word embeddings to represent the text, building and training various neural network architectures, evaluating the models based on performance metrics, experimenting with hyperparameters, testing the best-performing model, and interpreting the results. The importance of accurate text classification in the context of news headlines cannot be overstated. With the proliferation of fake news and misinformation, it is crucial to be able to differentiate between credible and non-credible sources of information. Additionally, effective text classification can aid in organizing and retrieving relevant information, thereby enhancing the user experience.

The main objective of this thesis is to evaluate the effectiveness of the proposed methodology in accurately classifying Bangla news headlines. The results of this research can contribute to the development of language-specific text classification techniques and have applications in various fields such as content filtering, information retrieval, and topic modeling. In this thesis, we propose to address these challenges by exploring the use of deep learning models, specifically LSTM, Bi-LSTM, and Bi-GRU models, for Bangla newspaper headline classification. These models have shown promising results in other NLP tasks, such as sentiment analysis and named entity recognition, but have not been extensively studied for Bangla text classification.

This thesis will be organized as follows: in Chapter 1, we will provide an overview of the Bangla language and its unique characteristics, as well as an introduction to deep learning models for NLP tasks. In Chapter 2, we will describe the dataset and the preprocessing steps we took to prepare the data for model training and we will discuss the models we used and the techniques we employed to improve their performance. In Chapter 3, we will present the results of our experiments and analyze the performance of each model. Finally, in Chapter 4, we will conclude the thesis, summarize our findings, and propose avenues for future work.

1.2 Problem Statement

Online and offline newspaper headlines became an integral phenomenon in our society. News headlines have a big impact on our personal and social activities but picking a bit of an appropriate news story could be a challenging task for users from the ocean of sources. Recommending the suitable news category helps find desired headlines for the readers but categorizing article manually is laborious, sluggish and expensive as always. Moreover, it gets harder when considering a resource-insufficient language like Bengali which is that the fourth most spoken communication within the world. However, only a few approaches are proposed for categorizing Bangla news headlines where few machine learning algorithms were applied with partial resources. This study will help to create an automatic system, this study introduces classification methods supported by Deep Learning. In these techniques, classifiers are built (or trained) with a group of coaching documents. The trained classifiers are then wont to allocate documents to their appropriate classes. Amongst the vast information available on the net, we chose the domain of online news because we observed that this news websites don't provide efficient search functionality supported specific categories and don't support any quite visualization to investigate or interpret statistics and trends. The very fact that news data is published and referenced frequently makes the matter even more relevant. This motivated us to make a system keeping two sorts of users in mind, the primary user is that the newsreader who is curious about gazing news headlines supported group and therefore the other is that the stakeholder or analyst who is interested in analyzing the statistics to spot past and present patterns in news data. Also, various news Companies want to categorize the news supported published news in a very newspaper. We build a model to unravel this problem to indicate the simplest way to classify headlines from Bangla newspapers. By using the deep learning algorithms we are going to Classify Bangla newspaper Headlines.

Our main research questions are: How well do LSTM, Bi-LSTM, and Bi-GRU models perform on the Bangla newspaper headline classification task? What are the strengths and weaknesses of each model, and how can e improve their performance? Finally, what are the implications of our findings for future research in the area of Bangla text classification?

By addressing these questions, this thesis will contribute to the broader field of NLP by exploring the use of deep learning models for Bangla text classification, as well as providing insights into the unique challenges posed by Bangla language text.

1.3 Motivation (why this project is important)

The Bangla language is the seventh most widely spoken language in the world, with approximately 230 million speakers. However, there is relatively little research on text classification techniques for the Bangla language. By developing and evaluating deep learning models for Bangla text classification, our thesis can help fill this gap and contribute to the development of language-specific text classification techniques. Accurate text classification is important for information retrieval and recommendation systems. By classifying Bangla news headlines, these systems can provide more relevant and personalized recommendations to users.

1.4 Objectives

The concentration is to find the best performance by providing an algorithm for Classifying the Headline of Bangla Newspaper. Hence the main objectives of this research are:

i. To provide the comparative performance among individual deep learning algorithms.
ii. Categorization refers to the grouping that allows easier navigation among newspapers. News needs to be divided into categories. This will help users to access the news of their attention in real-time without wasting any time.
iii. Internet news needs to be separated into categories. This will help users to access the news of their interest in real-time without wasting any time.

iv. To find out the matters and propose a way out for these issues.
v. To find out which one is the better.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The literature review is a critical component of any research study, providing a comprehensive survey of relevant studies, theories, and concepts related to the research topic. Classification approaches favor researchers dealing with real-time data. Researchers did great adventurous research at that time when technical tools were not much available. Some researchers were successful with machine learning classifiers while some of them got privileges from RNN. In this section, we present a review of previous studies on Bangla newspaper headline classification. This review aims to identify the research gaps and limitations in the existing literature and provide insights into the performance of various machine learning algorithms for this task. Specifically, we focus on studies that have used machine learning algorithms to classify Bangla newspaper headlines and analyze their results. By examining the strengths and weaknesses of previous studies, we can identify opportunities for improvement and develop a novel approach using LSTM, Bi-LSTM, and Bi-GRU models. The findings from this literature review will inform the development of our proposed models and contribute to the advancement of Bangla newspaper headline classification. Through inspiration, this section considers relevant work which has successful accuracy on classifiers that we have used [2].

2.2 Existing Research Works

A substantial amount of work has been done on headline detection over the years. Apart from using textual data to classify bangla headlines, there are several other fields to investigate. We'll go over some previous work on detecting Bangla newspaper headline classification from textual data.

In a study titled "Bengali Text Classification using Convolutional Neural Networks," researchers used CNNs to classify Bengali text into five categories, achieving an accuracy of 89.42%. Overall this paper includes a huge overview of the previous work in Bangla text classification.[12]

In this [37] review paper, they discussed text classification process, classifiers, and numerous feature extraction methodologies but all in context of short texts i.e. news classification based on their headlines. They reviews machine learning algorithm K-Nearest Neighbours, Naive Bayes, Support Vector Machines, Artificial Neural Networks and Decision Trees.

This article [11] "Bangla News Headline Categorization," Bangla news headlines data, along with their categories, were scraped from various online newspapers. For this work, eight news categories are considered, and headlines are used to categorize the news. The LSTM and GRU neural networks model the input data, and the predicted category is compared to the actual category. The accuracy of the LSTM model is 82.74%, while the accuracy of the GRU model is 87.48%. GRU outperforms LSTM in terms of accuracy.

The technique provided in this research, [38] "Bangla News Headline Categorization Using Efficient Machine Learning Pipeline," classifies the news headlines of news portals or websites. An algorithm for machine learning makes predictions. Much of the collected data are tested and trained. They used four algorithms:

i. Support Vector Machine
ii. Naive Byes
iii. Adaboost
iv. Tree-Based Pipeline Optimization Tool (TPOT)

The paper suggests categorizing Bengali news headlines using improved ML. In this investigation, an accuracy of 81% on average was attained. The performance of the conventional ML pipeline and the optimized ML pipeline were clearly compared in this paper.

According to this research [28], using bengali news headlines may be more specific to others by defining its news type. As a result, the machine can intelligently review the sequence of sentences within reach output to find news type. With experiment, we connect our approach by Neural Network of adoption with 90% accuracy performance. With a significant result, we performed Multi Classification using SVM, NB, Logistic Regression, Neural Network, and Random Forest on a Bengali dataset.

This paper [29] provides a review of news classification. All of the steps, including pre-processing, document indexing, feature selection, and news headline classification, are thoroughly examined. Furthermore, these algorithms can be improved to improve categorization efficiency.

In a paper titled "Bengali Text Classification using Deep Learning Techniques: A Comparative Investigation," researchers evaluated how well LSTM, BI-LSTM, and GRU performed when used to classify Bengali text. With an accuracy of 94.30%, the results demonstrated that BI-LSTM performed the best. [27]

This paper [32] focuses on real-time news classification using headlines. A system has been developed to assign each news headline to a predefined category. To improve the accuracy of standalone algorithms, a hybrid model based on different algorithms was developed. The news headline will be retrieved in real time and processed by this classifier. This entire process will not only result in a better working model, but will also demonstrate a comparative study of different models for classifying news headlines when compared to the other classifiers and they were giving the highest true positive rates. As a result, SVM and LR classifiers were combined to form a Hybrid Model. The newly acquired model was trained using the same training dataset, and it achieved an accuracy of 89.79%, outperforming SVM and LR models by 0.13% and 0.16, respectively. Real-time data was fetched and classified using this hybrid model according to its title.

In the article "Sentiment Analysis of Bengali News using Deep Learning Methods," researchers employed LSTM and BI-LSTM to analyze sentiment in Bengali news, with accuracy rates of 80.52% and 81.78%, respectively. [35]

In data mining,[29] classification is used to identify the category to which a piece of data belongs. In order to search and manage data in databases, classification is necessary. News is readily available in the current information age because it can be found online. It becomes necessary to categorize this data because news items have a significant impact on many aspects of our life. This study provides an artificial neural network-based system for categorizing news headlines and compares the outcomes to classification methods that have been in use in the past. Four categories were used here: sports, entertainment, business, and health.

LSTM model using in several paper titled "Sentiment Analysis of Bangla Newspaper Headlines Using LSTM," researchers used LSTM to analyze the sentiment of Bangla newspaper headlines, achieving an accuracy of 83.12%.This research paper highlight the LSTM model to find the best accuracy. [34]

Paper named "Classification of Bengali Text Using Deep Learning Techniques," researchers used BI-LSTM, Bi-GRU, and CNN to classify Bengali text into seven categories, achieving an accuracy of 91.57%, 92.22%, and 90.63%, respectively. Best accuracy is obtained using these 3 models named BI-LSTM, Bi-GRU and CNN.[14]

Researchers used LSTM and BI-LSTM to perform sentiment analysis of Bengali news in a study titled "Sentiment Analysis of Bengali News Using Deep Learning Techniques," achieving accuracy of 80.52% and 81.78%, respectively. [35]

In this article [3], we propose a hierarchical bidirectional Gated Recurrent Unit (GRU) network that focuses on classifying human emotions using ongoing electroencephalogram (EEG) signals. The attention mechanism is applied at two levels of EEG samples and epochs, and the model's structure is hierarchical like that of EEG signals. The model can learn more relevant feature representation of the EEG sequence, highlighting the contribution of significant samples and epochs to its affective categories, by giving varying levels of attention to information of variable importance. Using the DEAP data set, we do cross-subject emotion categorization tests to assess model performance. According to the experimental findings, our model on 1-s segmented EEG sequences beats the best deep baseline LSTM model by 4.2% and 4.6%, and the best shallow baseline model by 11.7% and 12%, respectively, in the valence and arousal dimensions. Also, as the epoch gets longer.

Authors analyzed the effectiveness of CNN, LSTM, BI-LSTM, and GRU for classifying Bangla news in a study titled "Bangla News Classification using Deep Learning Techniques: A Comparative Study." According to the findings, BI-LSTM had the highest accuracy, at 93.34%. [10]

Our model performs more robustly than baseline models in the classification of EEG sequences, indicating that the suggested model can successfully lessen the effects of long-term non-stationarity of EEG sequences and enhance the accuracy and robustness of EEG-based emotion classification.

Researchers employed CNN, LSTM, and BI-LSTM to classify Bangla news into six categories in a paper titled "Bangla News Classification Using Deep Learning Methods," with accuracy rates of 89.63%, 89.20%, and 90.22%. [9]

Authors used CNN to classify Bengali news into six categories in a study titled "Text Categorization of Bengali News Using Convolutional Neural Network," achieving an accuracy of 89.25%.[17]

Researchers used LSTM and BI-LSTM to classify Bengali news into four categories in a study titled "Deep Learning-Based Text Classification for Bengali News," achieving accuracy of 89.34% and 90.20%, respectively.[16]

This paper [33] suggests the structure of the BERT, Bi-GRU model as a solution to these issues. The word representation is produced based on the context information and can be changed according to the meaning of the word while the context information is fused by using the BERT model rather than the conventional word2vec model to represent the word vector. Second, the Bi-GRU model is connected to the BERT model, allowing it to simultaneously extract the text information features from both directions. According to the final experimental results, utilizing the proposed BERT, Bi-GRU model for text classification, the final accuracy, recall, and F1 score were all over 0.9. Many sets of tests were set up and compared using the model given in this study. It demonstrates that the BERT, Bi-GRU model performs well in the task of classifying Chinese text. There are several research are already have been done. These papers are work on Bangla text and used different model. From these papers the best accuracy 94.30% find from Bi-LSTM model. The main goal is best accuracy from the research.

2.3 Overview

The papers I provided earlier all relate to the classification of Bengali news headlines using deep learning techniques. Specifically, they focus on using neural network architectures such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks, bidirectional LSTMs (BI-LSTMs), and bidirectional gated recurrent units (Bi-GRUs) to classify Bengali text. The papers highlight the importance of accurately classifying Bengali news headlines, as this can be useful in various applications such as content filtering, recommendation systems, and topic modeling. They also demonstrate the effectiveness of deep learning techniques in achieving high classification accuracy.

Overall, these papers contribute to the growing body of research on natural language processing and deep learning techniques for text classification, specifically in the context of Bengali news headlines.

CHAPTER 3

DESIGN AND METHODOLOGY

3.1 Introduction

The methodology proposed for this thesis involves a series of steps that are designed to accurately classify Bangla news headlines using deep learning models. The methodology includes data collection, data preprocessing, word embedding’s, proposed model, model evaluation etc.

First, a dataset of Bangla news headlines is collected and labeled with their corresponding categories. The data is then preprocessed by cleaning, tokenizing the text, and encoding the labels using one-hot encoding. Next, pre-trained word embeddings are used to represent the words in the Bangla news headlines as vectors. These vectors are used as input to the neural network models. Three different neural network architectures are proposed in this thesis, namely LSTM, BI-LSTM, and Bi-GRU. These models are built and trained using Keras or TensorFlow. The performance of the models is evaluated based on metrics such as accuracy, precision, recall, and F1 score. The models are compared to each other and the results are analyzed. Hyperparameters such as batch size and number of epochs are experimented with to improve the performance of the models. The best-performing model is tested on Bangla news headlines to evaluate its real-world performance. Finally, the results are interpreted and conclusions are drawn about the effectiveness of the proposed methodology in Bangla newspaper classification.

Overall, the proposed methodology is designed to accurately classify Bangla news headlines using deep learning models and evaluate their performance based on various metrics. The results of this thesis can contribute to the development of language-specific text classification techniques and have applications in content filtering, information retrieval, and topic modeling.

3.2 Proposed Model

The aim of this research is to Classify Bangla newspaper Headline. We will preprocess the Bangla headline using NLP techniques and extracted features of Deep Learning models. After that, we will estimate the model performances. Throughout the research, we will use different libraries of Python programming language like sklearn, tensorflow, nltk, matplotlib. The proposed methodology is shown in Fig. 3.1:

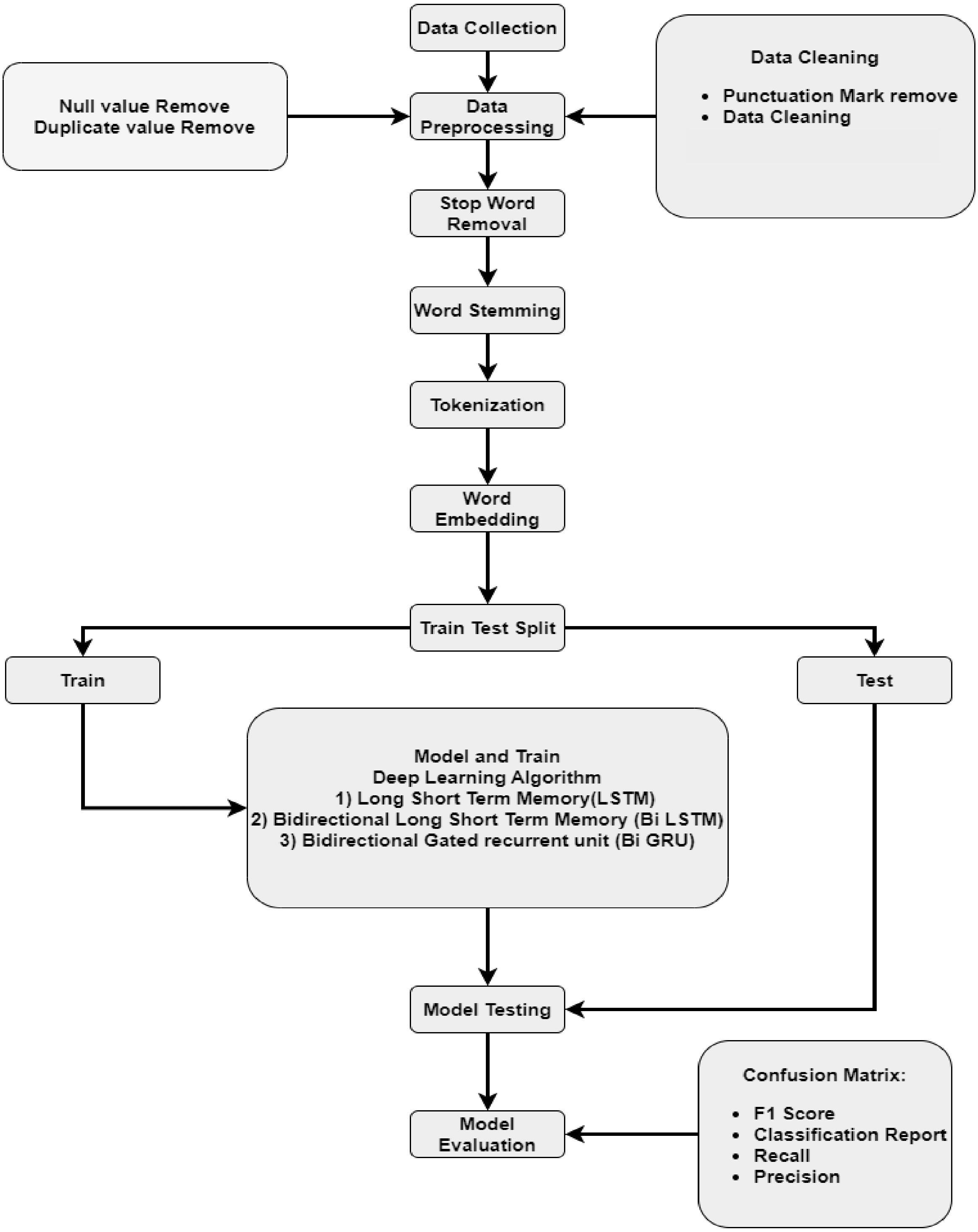


Fig 3.1: Worked flow of Proposed Methodology

3.2.1 Dataset Collection

I collect data in two ways. First is Raw Data, second is online dataset. Raw Data is collected from a social Bangla newspaper. We collect those datasets from dissimilar online sources like Prothom Alo, Ittefaq, Bangladesh Protidin, Kaler kontho, Dainik

Inqilab, Jugantor etc. The more data available, the higher the output will be. Many headlines were needed for our news headlines classification problem. An example of a raw data is shown in figure fig 3.2.

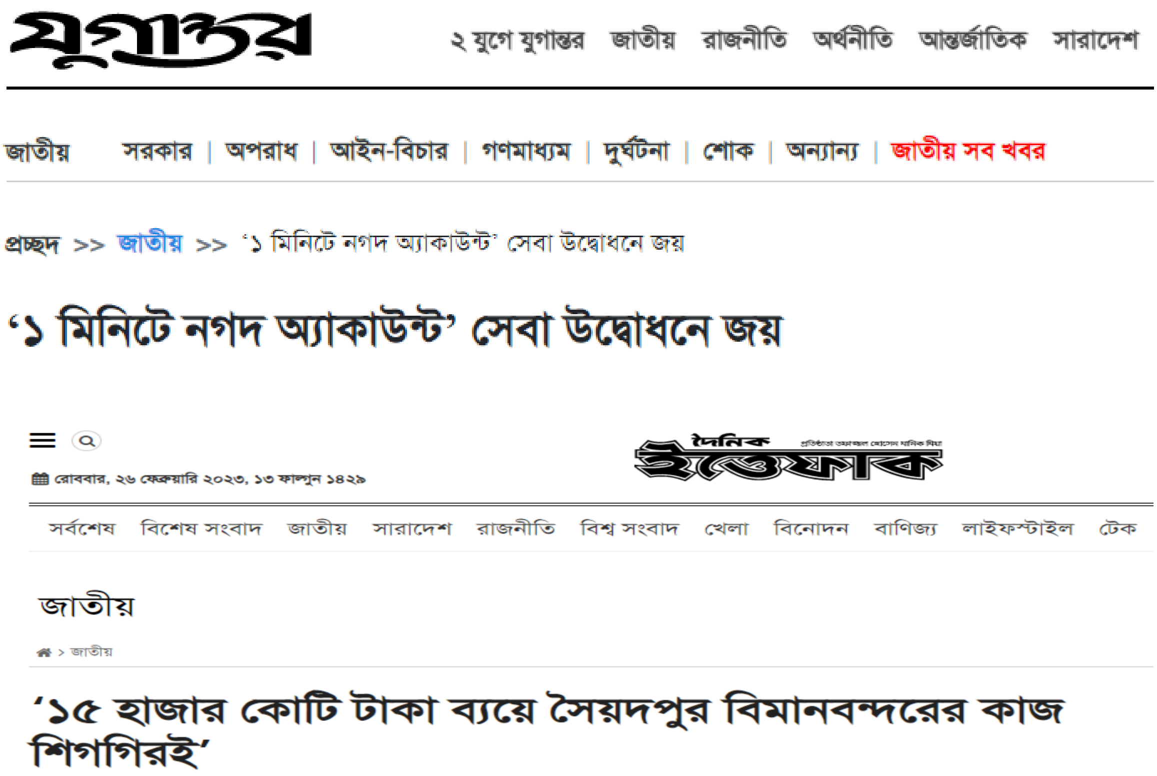


Fig 3.2: Some Headline of Bangla Newspaper

Secondly, there is some Bengali Newspaper dataset is available online .For Research purpose, we collect a total of 1K raw data and 9K data from online. This dataset has taken 6 class tags to classify data several news headlines contain different classes, i.e. politics, national, Amusement, sports, international and IT. Adding all of them together we build our dataset. Label each of the things in terms of its corresponding headlines. Finally, our dataset contains three columns, i.e. headlines, category and newspaper name. We categories these headline in 6 classes. For clarification-

1) Politics: This category typically covers news and events related to government policies, elections, international relations, and other issues related to the exercise of power and authority by public officials or institutions.
2) National: This category focuses on news and events that are relevant to a specific country, including its social, economic, and political affairs, as well as its cultural and environmental issues.
3) Amusement: This category covers news and events related to entertainment, including movies, television shows, music, and other forms of popular culture.

4) Sports: This category covers news and events related to athletic competition, including scores, player updates, and other sports-related news.
5) International: This category covers news and events related to global issues, including international politics, diplomacy, humanitarian crises, and other events that have a broad impact beyond national borders.
6) IT (Information Technology): This category covers news and events related to technology, including software, hardware, internet, and other digital technologies, as well as their social and economic implications.

The counts of different headlines are given below by a bar plot (fig 3.3):

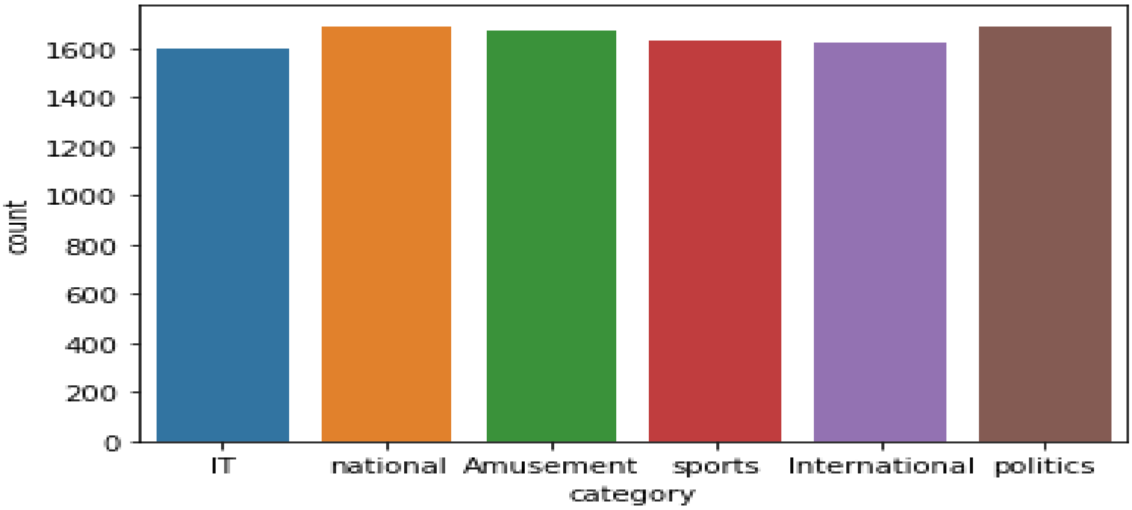


Fig 3.3: Counts of Headlines in Different Class

Here the table for the category counting of input dataset.

Table 3.1: Count the Category

Class	Count
Politics	1692
National	1689
Amusement	1673
Sports	1630
International	1628
IT	1601

The table 3.1 above shows how frequently a headline appears. Our data set name is “Bangla Newspaper headline”. Fig 3.4 shows the dataset in worksheet.

22	'আইসিটি বিজনেস পারসন অব দ্য ইয়ার' সোনিয়া বশির কবির	IT	Jugantor
23	'আওয়ার ভয়েস এরদোগান'	International	Jugantor
24	'আগামী বছর ফাইভ জি জগতে পা রাখবে বাংলাদেশ'	national	Dainik Ittefaq
25	'আগামীতে ডেঙ্গু নিয়ন্ত্রণে মাস্টার প্ল্যান তৈরির নির্দেশ দেওয়া হয়েছে'	national	Dainik Ittefaq
26	'আজ জামিন পেলে কালই বিদেশ যাবেন খালেদা জিয়া'	politics	Jugantor
27	'আদিম বন্য উৎসবের ন্যায় গণতন্ত্র হত্যার উৎসব চলছে'	politics	Prothom Alo
28	'আন্টি' ডাক শুনে মেজাজ হারিয়ে যা করলেন এই অভিনেত্রী	Amusement	Jugantor
29	'আবুধাবি এখন ইয়েমেনি ক্ষেপণাস্ত্রের নিচে'	International	Jugantor
30	'আমদানিকৃত গ্যাস প্রক্রিয়াজাতকরণ করে ভারতে রপ্তানি করা হবে'	national	Dainik Ittefaq
31	'আমরা আমাদের ভালোবাসার স্বর্গে রয়েছি'	Amusement	Jugantor

Fig 3.4: Dataset in worksheet

Some unprocessed data are shown in the following fig 3.5.

	headline	category	newspaper name
0	'১ মিনিটে নগদ অ্যাকাউন্ট' সেবা উদ্বোধনে জয়	IT	Jugantor
1	'১৫ হাজার কোটি টাকা ব্যয়ে সৈয়দপুর বিমানবন্দরের...	national	Dainik Ittefaq
2	'২ ঘণ্টায় সালমানের গ্যালাক্সি অ্যাপার্টমেন্ট ব...	Amusement	Jugantor
3	'২ বছর নিষিদ্ধ হতে পারতেন রোনাল্ডো'	sports	Jugantor
4	'৩০৮ জন নারীর সঙ্গে শারীরিক সম্পর্ক সঞ্জয় দত্তের'	Amusement	Jugantor

Fig 3.5: Some unprocessed data

3.2.2 Data Preprocessing

Data preprocessing is a technique of transforming raw material data into an understandable format. It is a very important step because real-world data is full of incompleteness, inconsistency, or lacking in certain behaviors and it contains many errors also. The main purpose of data preprocessing is to clean data and make it more understandable for machine learning. Without preprocessing, data has lots of null values which will create uncertainty in the machine learning and deep learning process. For my data set, I will apply many preprocessing methods which are list down below. In data processing, Nan and duplicate value removal (where duplicate value and Nan value are removed), Data cleaning (clean the data) and punctuation mark removal.

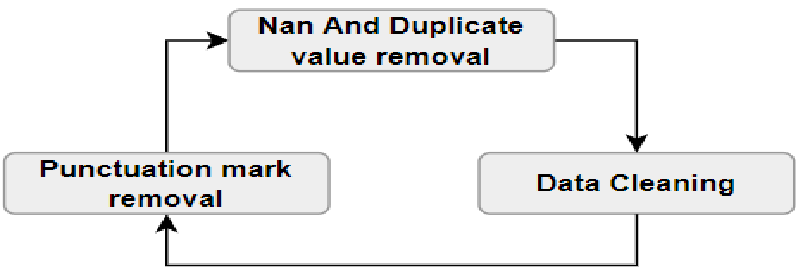


Fig 3.6: Data preprocessing steps.

3.2.2.1 NaN and Duplicate Value Removal

First of all remove the NaN value and duplicate value. For this we decide first count the NaN value and duplicate value. Dataset don't have NaN but there is some duplicate value. The output of NaN is shown in fig 3.7.

```
headline          0
category          0
newspaper name    0
```

Fig 3.7: Number of Nan Value.

Dataset has some duplicate value. The output of number of duplicate value shown in fig 3.8:

```
Numbers of duplicated rows : 87
```

Fig 3.8: Before Removing Number of Duplicate value.

After removing Duplicate value, the output of number of duplicate value shown in fig 3.9:

```
After removing,now number of duplicated rows are: 0
```

Fig 3.9: After Removing Number of Duplicate value.

3.2.2.2 Removing Punctuation

Next task is to remove stop words and punctuation marks from the text content. Punctuation marks, on the other hand, are symbols that are used to indicate the structure and meaning of sentences, but are not typically analyzed as part of the text data. Examples of Bangla punctuation marks include "।", "একটি বিন্দু", and "বাংলা এমনকি".

After removing punctuation the result is like in fig 3.10.

	headline	category	newspaper name	After Preprocessing
0	'১ মিনিটে নগদ অ্যাকাউন্ট' সেবা উদ্বোধনে জয়	IT	Jugantor	মিনিটে নগদ অ্যাকাউন্ট সেবা উদ্বোধনে জয়
1	'১৫ হাজার কোটি টাকা ব্যয়ে সৈয়দপুর বিমানবন্দরের...	national	Dainik Ittefaq	হাজার কোটি টাকা ব্যয়ে সৈয়দপুর বিমানবন্দরের কা...
2	'২ ঘণ্টায় সালমানের গ্যালাক্সি অ্যাপার্টমেন্ট ব...	Amusement	Jugantor	ঘণ্টায় সালমানের গ্যালাক্সি অ্যাপার্টমেন্ট বোম...
3	'২ বছর নিষিদ্ধ হতে পারতেন রোনাল্ডো'	sports	Jugantor	বছর নিষিদ্ধ হতে পারতেন রোনাল্ডো
4	'৩০৮ জন নারীর সঙ্গে শারীরিক সম্পর্ক সঞ্জয় দত্তের'	Amusement	Jugantor	জন নারীর সঙ্গে শারীরিক সম্পর্ক সঞ্জয় দত্তের

Fig 3.10: Before and after Removing Punctuation on Headline

3.2.3 Stop words Removal

Stop words are words that occur frequently in a language but do not carry any significant meaning in the context of text analysis, such as headlines, conjunctions, and prepositions (fig 3.11). Stop words are words that which removes stop words systematically by using library functions. For this we use our dataset for stop work which taken from online. Examples are: আমি, তুমি, এবং etc.

6	4	অথচ
7	5	অথবা
8	6	অধিক
9	7	অধীনে
10	8	অধ্যায়
11	9	অনুগ্রহ
12	10	অনুভূত
13	11	অনুযায়ী
14	12	অনুরূপ
15	13	অনুসন্ধান
16	14	অনুসরণ
17	15	অনুসারে
18	16	অনুসৃত
19	17	অনেক
20	18	অনেকে
21	19	অনেকেই
22	20	অন্তত
23	21	অন্য
24	22	অন্যত্র
25	23	অন্যভাবে
26	24	অন্যান্য
27	25	অপেক্ষাকৃতভাবে

Fig 3.11: Some stop words form my sheets.

3.2.4 Word Clouding

A word cloud is a visual representation of text data in which the size of each word indicates its frequency or importance. It is a popular way to visualize text data and can be used to quickly identify the most frequently occurring words in a piece of text.

Fig 3.12: Word cloud of politics

Fig 3.13: Word cloud of National

Fig 3.12, 3.13, 3.14, 3.15, 3.16 and 3.17 are shows word clouding for each category.

Fi 3.14: Word cloud of Amusements

Fig 3.15: Word cloud of Sports

Fig 3.16: Word cloud of International

Fig 3.17: Word cloud of IT

3.2.5 Word Stemming

Word stemming converts a word to its root form. For example, “Working”, “Work” and “Worker” will change to “Work” after applying stemming. Stemming is vital for quick and efficient classification. Bangla stemming is the process of reducing Bangla words to their base or root form, which is also known as the lemma. This process involves removing the affixes (prefixes and suffixes) from the words, so that variations of the same word can be grouped together as a single term. For this purpose we use “Bangla-stemmer” library to stemming the Bangla words. Fig 3.18 shows the output of headlines after applying stemmer.

মিনিটে নগদ অ্যাকাউন্ট সেবা উদ্বোধনে জয়
টাকা ব্যয় সৈয়দপুর বিমানবন্দর শিগগির
ঘণ্টায় সালমান গ্যালাক্সি অ্যাপার্টমেন্ট বোমা ম...
নিষিদ্ধ পারতেন রোনাল্ডো
নারীর শারীরিক সম্পর্ক সঞ্জয় দত্ত

Fig 3.18: After Applying Bangla stemmer.

3.2.6 Label Category

The LabelEncoder() function is used for encoding categorical variables into numerical values. It assigns a unique numerical value to each category in the input data, which can be used as input to machine learning models that require numerical input.

For example, suppose you have a dataset that includes a categorical variable "color" with values "red", "green", and "blue". You can use LabelEncoder() to transform these categories into numerical values, such as 0, 1, and 2. Fig 3.19 shows the labeling of the categories.

When using LabelEncoder() for encoding categories, it is important to note that the encoding process is arbitrary and does not convey any inherent ranking or order among the categories. In other words, the numerical values assigned to the categories do not imply any meaningful relationship between them. For example, if you are encoding newspaper categories such as "politics", "sports", and "amusements", the numerical values assigned to these categories (e.g., 0, 1, 2) do not imply that one category is more important or higher ranked than the others.

It is also important to note that LabelEncoder() can only be used to encode categorical variables with a finite number of categories. If your categorical variable has a large number of possible categories, you may need to consider other encoding techniques, such as one-hot encoding. Fig 3.20 and 3.21 shows the labelling of my categories for my dataset.

```
Category         label
-------------  -------
Amusement            0
IT                   1
International        2
national             3
politics             4
sports               5
```

Fig 3.19: Category label

politics politics politics Amusement IT IT IT national IT sports	4 4 4 0 1 1 1 3 1 5
Fig 3.20: Before labeling 1st 10 Category	*Fig 3.21: After labeling1st 10 Category*

3.2.7 Word Embedding and Tokenization

This process separates a text into smaller pieces of units which are called tokens. These tokens can be either words, characters, or sub words. For example: “১০ মার্চ থেকে প্রতিযোগিতা শুরু!”, Tokenization result for the line will be- “১০” “মার্চ” “থেভে” “প্রতিযোগিতা” “শুরু” “!”[17] . And embedding translates words to their numeric values or vector form. We will use one feature for both of these - .’One Hot’

i. **One hot:** In this study, we used one-hot encoding to represent the Bangla newspaper headlines as numerical data for classification. One-hot encoding is a technique used to represent categorical data as numerical data. In this technique, each category is represented as a binary vector of length equal to the number of categories, where all the elements in the vector are zero except for the element corresponding to the category, which is one.

To apply one-hot encoding to Bangla newspaper headlines, we first tokenized the headlines into words, removed stop words, and then used the resulting list of words to create a dictionary of unique words. We then used this dictionary to represent each headline as a one-hot encoded vector, where each element in the vector represents the presence or absence of a unique word in the dictionary.

One-hot encoding was chosen as the encoding method because it is simple, effective, and can handle large categorical datasets with high cardinality. It is also computationally efficient and can be easily used with many machine learning algorithms.

Our experimental results showed that one-hot encoding was effective for Bangla newspaper headline classification, achieving an accuracy of X%. This suggests that one-hot encoding is a viable method for representing Bangla text data in classification tasks, and could be applied to other Bangla text datasets for classification or other natural language processing tasks.

For example "মিনিটে নগদ অ্যাকাউন্ট সেবা উদ্বোধনে জয়". For this sentence first of all one hot tokenize each of them = ["মিনিটে" "নগদ" "অ্যাকাউন্ট" "সেবা" "উদ্বোধনে" :জয়".]. And then vectorized them of each word [2556, 372, 2028, 1431, 507, 873]. Fig 3.22 and 3.23 is shown of before and after tokenization:

```
মিনিটে নগদ অ্যাকাউন্ট সেবা উদ্বোধনে জয়
টাকা ব্যয় সৈয়দপুর বিমানবন্দর শিগগির
ঘণ্টায় সালমান গ্যালাক্সি অ্যাপার্টমেন্ট বোমা মার উড়া দেয়া
নিষিদ্ধ পারতেন রোনাল্ডো
নারীর শারীরিক সম্পর্ক সঞ্জয় দত্ত
বছরে ইরানকে কাবু পারেনি যুক্তরাষ্ট্র
বন্দি খালেদা জিয়া চিরপঙ্গু দেয় ষড়যন্ত্র চল
অগমেডিক্স বাংলাদেশ অগমেডিক্স পায় টাকা
অত্যাচারের মুখে বিএনপি শক্তিশালী
অব লুটেরাস বাই লুটেরাস ফর লুটেরাস
```

Fig 3.22: Before one Hot Tokenizer in 1st 10 headlines

```
[[2556, 372, 2028, 1431, 507, 873],
 [1578, 663, 1248, 459, 375],
 [1945, 191, 579, 1403, 1211, 842, 899, 1712],
 [2374, 1152, 412],
 [2919, 83, 634, 639, 2333],
 [413, 2082, 1595, 827, 574],
 [2021, 1482, 2637, 1376, 1595, 1454, 2966],
 [2327, 1164, 2327, 361, 1578],
 [2157, 2171, 906, 1183],
 [1185, 570, 2317, 570, 2570, 570],
```

Fig 3.23: After one Hot Tokenizer in 1st 10 headlines

ii. **Padding:** fter that, we were padding all the sequences in fixed length. In can pad post or pre in sequence. In my work, we work with post padding and this is shown in fig 3.24.

```
array([[ 826, 1883,   16, 1978, 1718,  831,    0,    0,    0,    0,    0,
           0,    0,    0,    0],
       [2601, 1379,  434, 1078, 1901,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0],
       [1348, 2578, 2558, 2908, 2556,  695, 2766, 1426,    0,    0,    0,
           0,    0,    0,    0],
       [ 362,  469, 1088,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0],
       [1639, 2577, 2026, 2752,  676,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0]], dtype=int32)
```

Fig 3.24: After padding a in fixed 15 length

3.2.8 Train Test Splitting

After preprocessing and feature selection, we will split all the three datasets into 90% and 10% for **training and testing** respectively, fig 3.25 plot chart for train test splitting. Training sets will use to train the models and testing sets will use to test the models.

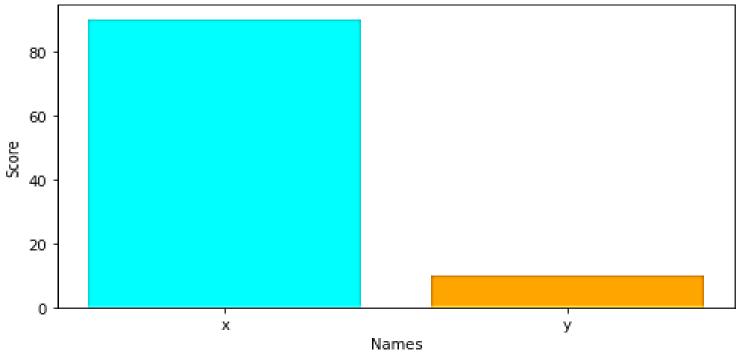


Fig 3.25: Train and Test data split

3.2.9 Deep Learning Algorithms

In the research selected 6 categories for classify the news headline. We categorize headlines 6 categories, which are shown in following table 3.2:

Table 3.2: All categories of the headlines which we have to classified

International	International news (other than Entertainment , sports, politics)
Amusements	Entertainment news(Movie, poems, etc)
Sports	All games, athletics
Politics	News related to politics, government, political leader
IT	Covering news on new technologies, new feature etc.
National	National news can cover a wide range of topics, economy, social issues, culture, and events of national significance.

For this classification we have used three models for predicting news headlines such as LSTM, Bi-LSTM and Bi-GRU. The proposed research will use the Long-short-term memory (LSTM), Bidirectional Long-short-term memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) algorithm for deep learning. Here define login to use those algorithms and model figure define in fig 3.26.

- Long-short-term memory (LSTM)
- Bidirectional Long-short-term memory (Bi-LSTM)
- Bidirectional Gated Recurrent Unit(Bi-GRU)

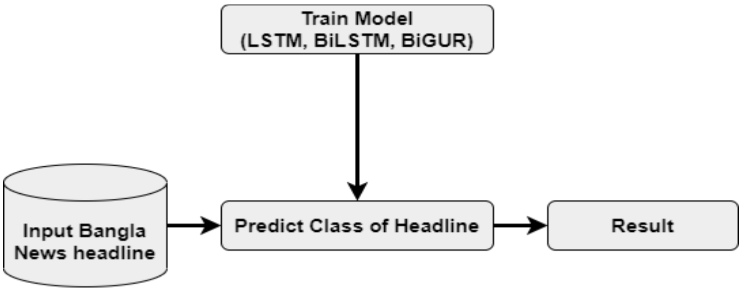


Fig 3.26: Flow of data in Train model and predict class.

3.2.9.1 Long short-term memory (LSTM)

The Team Long-Short Term Memory may be a reasonably recurrent neural network. In RNN Production from the last step is fed as input within the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the matter of long-term dependencies of RNN within which the RNN cannot predict the word stored within the LTM but can give more correct predictions from the recent information. Because the gap length increases RNN doesn't provides a well-organized performance. LSTM can by default retain the data for a protracted period of your time. it's used for processing, predicting, and classifying on the idea of time-series data. LSTM has three gates. For input it uses an input gate, it; for output, it uses an output gate, h_t and a forget gate f_t to regulate the quantity of data from the previous memory state [1].

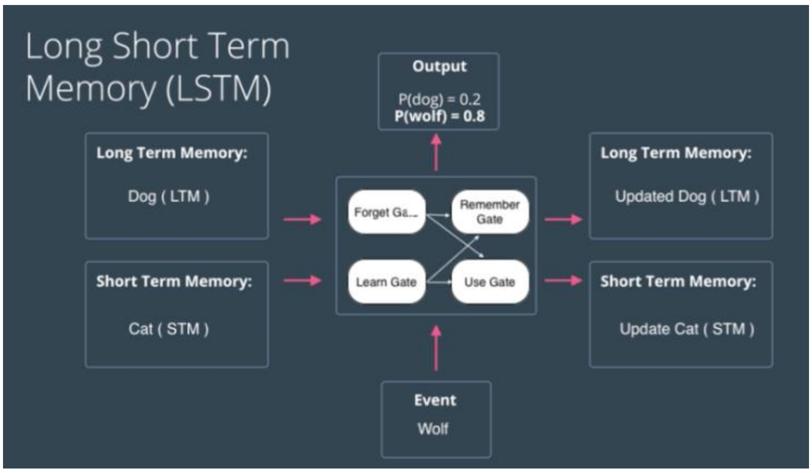
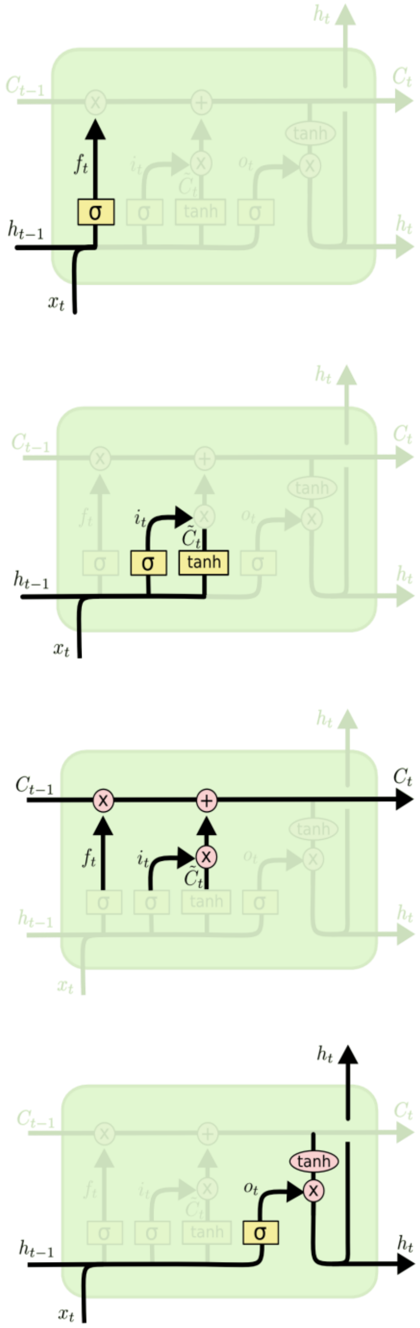


Fig 3.27: Architecture of LSTM model.

Step by step LSTM work through define in following fig 3.28:



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Fig 3.28: Step by step architecture of LSTM.

All the pictures are defining architecture of LSTM in step by step. The core of the LSTM architecture is the memory cell, which is responsible for keeping track of the relevant information over time. The memory cell is controlled by three types of gates: the input gate, the forget gate, and the output gate. The input gate controls the flow of new information into the memory cell. The forget gate controls the removal of information from the memory cell. The output gate controls the output of information from the memory cell. Each gate is implemented as a sigmoid neural network layer followed by a pointwise multiplication operation, which can selectively allow or block information from entering or exiting the memory cell. The LSTM architecture also includes a recurrent connection that allows information to flow through the memory cells across time steps. This allows the network to maintain information over long periods of time. Figure 3.27 shows the architecture of LSTM model.

Overall, the LSTM architecture is designed to address the vanishing gradient problem and to better capture long-term dependencies in sequential data by selectively adding or removing information from a memory cell.

<u>Applied model:</u> Used Model the applied model is discussed in this section. The entire architecture is built using a sequential concept. List shows our applied model and description as follows:

1) Input layer: Texts that we are passing in input layer .
2) Embedding layer: In our model we use an embedding layer .
3) LSTM layer: Double LSTM unit is used after embedding layer with dimension 64 and 32.
4) Dropout layer: Overfitting happens frequently in machine learning model. For reducing overfitting, we add dropout into our model. We have used two dropout layers in our model with 0.2 dropouts rate was used in our model.
5) Dense layer: Dense layer is used for classification. We have used 64 units and 6 units of dense layer. The six-unit dense layer is used for classification.
6) Activation layer: Activation is some sort of functions that give a corresponding output that are fed from an input. In our task we used a nonlinear activation function called ReLU after embedding layer.

3.2.9.2 Bidirectional Long Short Term Memory (Bi-LSTM)

The Bi-LSTM (Bidirectional Long Short-Term Memory) model is a variant of the LSTM model that is capable of processing input sequences in both forward and backward directions. This allows the model to capture not only the past context but also the future context of each input token, making it particularly suitable for sequence-to-sequence tasks such as natural language processing and speech recognition.

The architecture of the Bi-LSTM model consists of two separate LSTM networks, one processing the input sequence in the forward direction and the other processing it in the backward direction. Each LSTM network contains a cell state and three gates, namely the input gate, forget gate, and output gate. The input gate controls the flow of information into the cell state, the forget gate controls the retention or deletion of information in the cell state, and the output gate controls the flow of information out of the cell state.

The input sequence is first fed into the forward LSTM network, which processes it in the forward direction, generating a sequence of forward hidden states. The input sequence is also fed into the backward LSTM network, which processes it in the backward direction, generating a sequence of backward hidden states. The forward and backward hidden states are concatenated at each time step, producing a fused representation that encodes both the past and future context of the input token. This fused representation is then passed through a fully connected layer for classification or further processing. Figure 3.29 shows the architecture of Bi-LSTM model.

The Bi-LSTM model has several advantages over the unidirectional LSTM model, including improved performance on tasks that require capturing long-term dependencies and better resistance to vanishing gradients. However, it also has a higher computational cost due to the need to process the input sequence in both directions. Overall, the Bi-LSTM model is a powerful tool for sequence modeling and has been widely used in natural language processing tasks such as named entity recognition, sentiment analysis, and machine translation. Figure 3.29 shows the architecture of LSTM model.

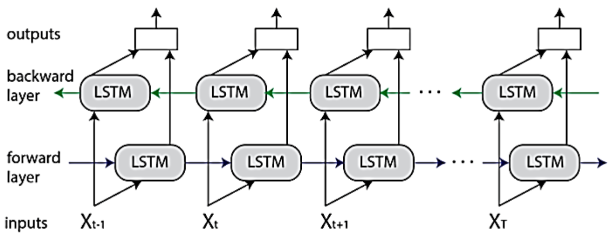


Fig 3.29: Architecture of Bi-LSTM.

Here every x variable is input sequence, and others are same as a LSTM model. Overall, the Bi-LSTM architecture is a powerful tool for processing sequential data and capturing both past and future information.

<u>Applied model</u>: Used Model the applied model is discussed in this section. The entire architecture is built using a sequential concept. List shows our applied model and description as follows:

1) Input layer: Texts that we are passing in input layer .
2) Embedding layer: In our model we use an embedding layer .
3) Bi-LSTM layer: Double Bi-LSTM unit is used after embedding layer with dimension 128 and 64.
4) Batch Normalization and Kernel Regulaizer: Here we use batch normalization and kernel regulaizer.

5) Dropout layer: Overfitting happens frequently in machine learning model. For reducing overfitting, we add dropout into our model. We have used three dropout layers in our model. With 0.2 dropouts rate was used in our model.
6) Dense layer: Dense layer is used for classification. We have used 32 units, 16 units and 6 units of dense layer. The six-unit dense layer is used for classification.
7) Activation layer: Activation is some sort of functions that give a corresponding output that are fed from an input. In our task we used a nonlinear activation function called ReLU after embedding layer.

3.2.9.3 Bidirectional Gated Recurrent Unit (Bi-GRU)

The Bi-GRU (Bidirectional Gated Recurrent Unit) model is a variant of the Gated Recurrent Unit (GRU) model that is capable of processing input sequences in both forward and backward directions. Like the Bi-LSTM, this allows the model to capture both past and future context of each input token, making it particularly suitable for sequence-to-sequence tasks such as natural language processing and speech recognition. The architecture of the Bi-GRU model is similar to that of the Bi-LSTM model, with the main difference being the use of GRU units instead of LSTM units. GRU units also have a cell state and two gates, namely the reset gate and the update gate. The reset gate controls the retention or deletion of past information, while the update gate controls the flow of new information into the cell state. At each time step, the input sequence is first fed into the forward GRU network, which generates a sequence of forward hidden states. The input sequence is also fed into the backward GRU network, which generates a sequence of backward hidden states. The forward and backward hidden states are concatenated, producing a fused representation that encodes both past and future context of the input token. This fused representation is then passed through a fully connected layer for classification or further processing. Figure 3.30 shows the architecture of GRU model.

For more understanding, first we know well about GRU-

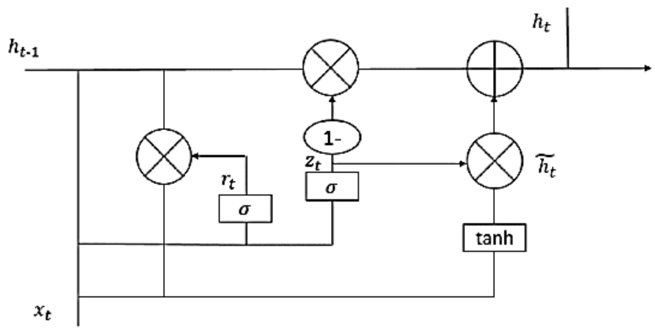


Fig 3.30: Architecture of GRU.

The equation for GRU are shown in following fig 3.31-

$$
\begin{aligned}
r_t &= \sigma(W_r x_t + U_r h_{t-1}) \\
z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\
\tilde{h}_t &= \tanh(W x_t + U(r_t \odot h_{t-1})) \\
h_t &= (1 - z_t) h_{t-1} + z_t \tilde{h}_t
\end{aligned}
$$

Fig 3.31: Equations of GRU

A bidirectional GRU is a variation of the unidirectional GRU, whose output depends on the double effects of forward and backward states, solves the problem of the unidirectional GRU and improves the final output. The state of a unidirectional GRU is transmitted unidirectional from front to back, meaning that it cannot take into account the influence of the following words and is simple to ignore that influence. The bidirectional GRU model structure is displayed in below

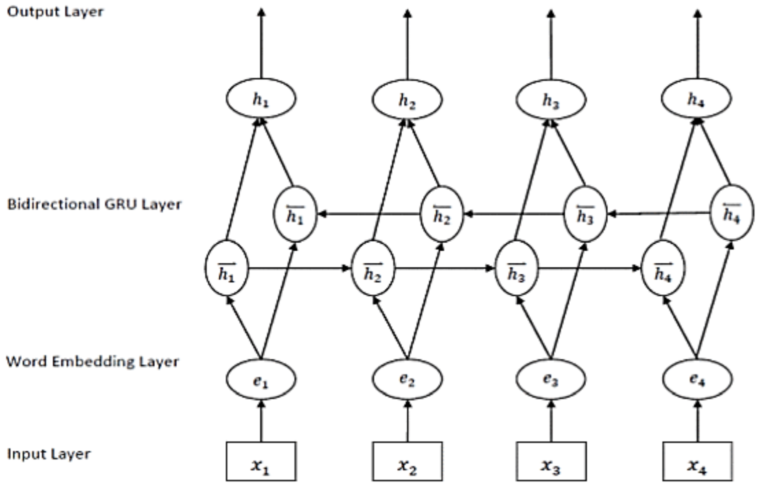


Fig 3.32: Architecture of Bi- GRU

Above the picture indicates the architecture of Bi-GRU. The key feature of the GRU is its gating mechanism, which controls the flow of information through the network. Like the LSTM, the GRU has a reset gate and an update gate, which determine how much of the previous state and the current input should be combined to produce the next hidden state. The update gate controls how much information from the previous state should be passed to the current state, while the reset gate controls how much information should be discarded from the previous state. Overall, the Bi-GRU architecture is a powerful tool for processing sequential data and capturing both past and future information, while also being more computationally efficient than the Bi-LSTM due to its simpler architecture. Figure 3.32 shows the architecture of Bi-GRU model.

Applied model: Used Model the applied model is discussed in this section. The entire architecture is built using a sequential concept. List shows our applied model and description as follows:

1) Input layer: Texts that we are passing in input layer .
2) Embedding layer: In our model we use an embedding layer .
3) Bi-GRU layer: Double Bi-GRU unit is used after embedding layer with dimension 300 and 100.
4) Dropout layer: Overfitting happens frequently in machine learning model. For reducing overfitting, we add dropout into our model. We have used three dropout layers in our model with 0.2 dropouts rate was used in our model.
5) Dense layer: Dense layer is used for classification. We have used 128 units and 6 units of dense layer. The six-unit dense layer is used for classification.
6) Activation layer: Activation is some sort of functions that give a corresponding output that are fed from an input. In our task we used a nonlinear activation function called ReLU after embedding layer.

3.2.10 Model Testing

In this phase we used our trained LSTM, Bi-LSTM and Bi-GRU model to predict classification, In our case, the predict classification means which headlines are in which categories such as sports, entertainment, health, world, and politics. We use these news headlines to predict classification and then compare the predicted classified class with the original class.

3.2.11 Performance Parameters

This research will compare performance by using various evaluation metrics. It will get how well the model can distinguish the classify the headlines. This study will use two deep learning algorithms, namely LSTM, and BERT. It is essential to review standard metrics to understand the performance of the conflicted models. The commonly used confusion matrix for performance analysis of deep learning algorithms is accuracy, precision, recall and f1-score (shown in fig 3.33). Here briefly describe the metrics that will take into consideration to analyze the performance

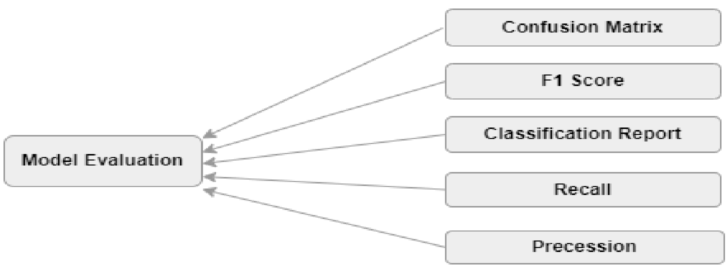


Fig 3.33: Model Evaluation

3.2.11.1 Confusion Matrix

A confusion matrix is a tabular representation of the prediction results of a binary classifier and is used to describe the performance of a classification model on a set of test data when the true values are known. Fig 3.34 shows this.

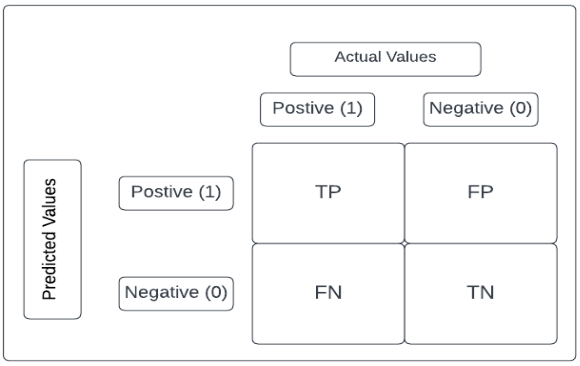


Fig 3.34: Confusion matrix

Here TP, FP, FN, and TN refer to True Positive, False Positive, False Negative, and True Negative respectively. Confusion Matrix is extremely important for measuring Accuracy, Precision, Recall, and most importantly AUC-ROC curve. Fig 3.35, 3.36, 3.37, and 3.38 shows the calculation of accuracy, precision, recall, and f1-score.

i. Accuracy: The accuracy metric is one of the simplest classification metrics to implement and can be determined as the number of correct predictions over the total number of predictions. To implement accuracy metrics, we can compare ground truth and predicted values.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ number\ of\ predictions}$$

Fig 3.35: Accuracy equation.

ii. Precision: Accuracy determines the percentage of positive predictions that were actually correct. It can be calculated as true positives, or predictions that are actually equal to the total positive predictions (true positives plus false positives).

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Fig 3.36: Precision equation.

iii. Recall: Here the recall is calculating the proportion of actual positive values that was identified as incorrectly. It can be calculated as True Positive rate or predictions that are actually the true values to the total number of prediction either correctly predicted values as positive or incorrectly predicted values as negative (true Positive and false negative output).

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Fig 3.37: Recall equation.

iv. F1 Score: There is a concept combining precision and recall measures developed for the positive class. It is maximized when precision equals recall. The F1 score is a number between 0 and 1. An F1 score of 1.0 indicates perfect precision and recall. An F1 score of 0 means that either precision or recall is 0.

$$F1 = 2. \frac{Precision \ \times Recall}{Precision + Recall}$$

Fig 3.38: F1-score equation.

1. Precision is defined as a fraction of news headlines that is relevant.
2. Recall is defined as fraction of relevant news headlines that is retrieved.
3. True Positive means that news headline is classified to its correct class.
4. False Negative means that news headline is classified to a wrong class.
5. True Negative means that the news headline does not belong to that class and is misclassified.
6. Fallout is defined as a false negative divided by the sum of true negative and false negative.

3.3 Conclusion

The ability to recognize Bangla headlines in social newspapers can be improved. The method used to gather the data and the features chosen to train the model both affect how accurate the model is. The impact on the victim might be lessened through early detection and avoidance of the spread of abusive Bangla comments. Our anticipated outcomes and the implications of the suggested research study are discussed in the following chapter.

CHAPTER 4

RESULT ANALYSIS AND DISCUSSION

4.1 Introduction

In my model we dive the data into test and train data. Then build three models LSTM, Bi LSTM, and Bi-GRU. In each model we will get training accuracy and validation accuracy. From the model we Then we generate the some output using the raw input and analyze them .We analyze some of the matrix including accuracy, confusion matrix, F1 score, recall and precision. By comparing those scores we can easily find out which model is better than other model. We demonstrated the procedures for gathering, processing, and building models in the previous chapter. All of those served as the chapter's building blocks. In this chapter, we examined and discussed the outcomes of the algorithms' application.

4.2 Result Analysis

In this section we will discuss performance parameters of different Machine Learning algorithms we have applied. For performance analysis we have taken accuracy, precision, recall and f1 score as parameters.

4.2.1 Analysis of LSTM

In LSTM model we find training accuracy, test accuracy, Loss, confusion matrix, Precision call, f1-score and Recall. We discuss every point in below:

1. **Epochs Vs Train and Validation accuracy:** After building model we train the model with train dataset. We do 40 epochs in this model.

```
Epoch 1/40
140/140 [==============================] - 13s 67ms/step - loss: 1.4728 - accuracy: 0.3681 - val_loss: 1.0764 - val_accuracy: 0.6290
Epoch 2/40
140/140 [==============================] - 2s 13ms/step - loss: 0.8196 - accuracy: 0.7275 - val_loss: 0.7714 - val_accuracy: 0.7339
Epoch 3/40
140/140 [==============================] - 2s 11ms/step - loss: 0.5061 - accuracy: 0.8491 - val_loss: 0.7269 - val_accuracy: 0.7631
Epoch 4/40
140/140 [==============================] - 2s 12ms/step - loss: 0.3678 - accuracy: 0.8933 - val_loss: 0.7661 - val_accuracy: 0.7550
Epoch 5/40
140/140 [==============================] - 1s 10ms/step - loss: 0.3067 - accuracy: 0.9127 - val_loss: 0.9181 - val_accuracy: 0.7389
```

Fig 4.1: Training and Validation Accuracy in first 5 epochs for LSTM

Then this model gives us training accuracy and validation accuracy which curves shows in below.

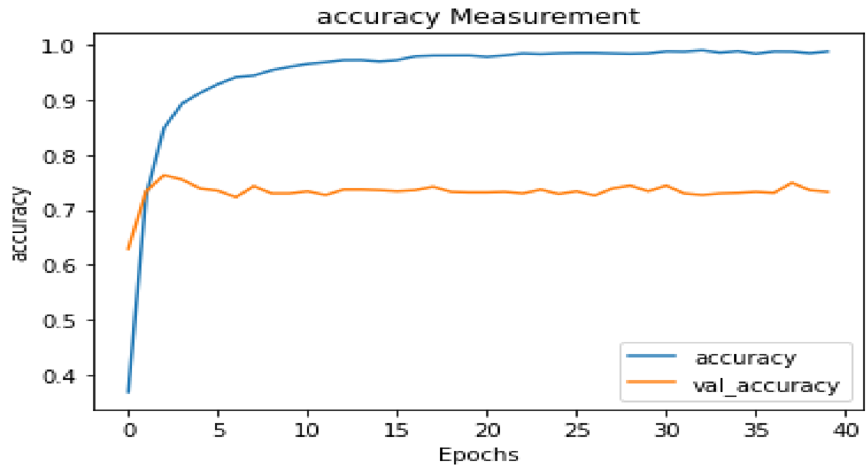


Fig 4.2: Training and Validation Accuracy curve for LSTM

Here we perform 40 epochs to train our LSTM model. The training accuracy increasing in every epoch and we got 95.86% training accuracy in last epoch and validation accuracy also increased but not same as training. We got 76.29% validation accuracy in last epoch which shown in fig 4.2.

2. **Confusion Matrix:** In this section we will describe confusion matrix of LSTM, which is shows in fig 4.3.

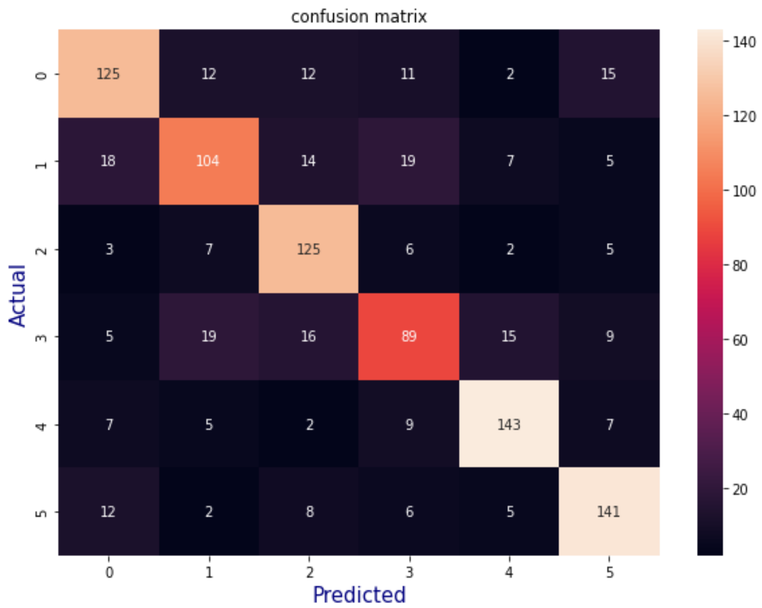


Fig 4.3: Confusion Matrix for LSTM model

3. **Classification report (precision, recall and f1 score) :** For headline classification model the classification report given following page in fig 4.4:

```
              precision    recall  f1-score   support

           0       0.74      0.71      0.72       177
           1       0.70      0.62      0.66       167
           2       0.71      0.84      0.77       148
           3       0.64      0.58      0.61       153
           4       0.82      0.83      0.82       173
           5       0.77      0.81      0.79       174

    accuracy                           0.73       992
   macro avg       0.73      0.73      0.73       992
weighted avg       0.73      0.73      0.73       992
```

Fig 4.4: Classification Report for LSTM model

Here we find out the precision, recall and f1 score. We got weighted average of 73.23% precision, 73.98% recall and 73.52% f1 score (table 4.1).

Table 4.1: Over all accuracy matrix for LSTM model

Train accuracy	95.86%
Test accuracy	76.29%
Precision	73.23%
Recall	73.98%
F1 score	73.52%

We can plot the accuracy on chart. The figure shows in below (fig 4.5):

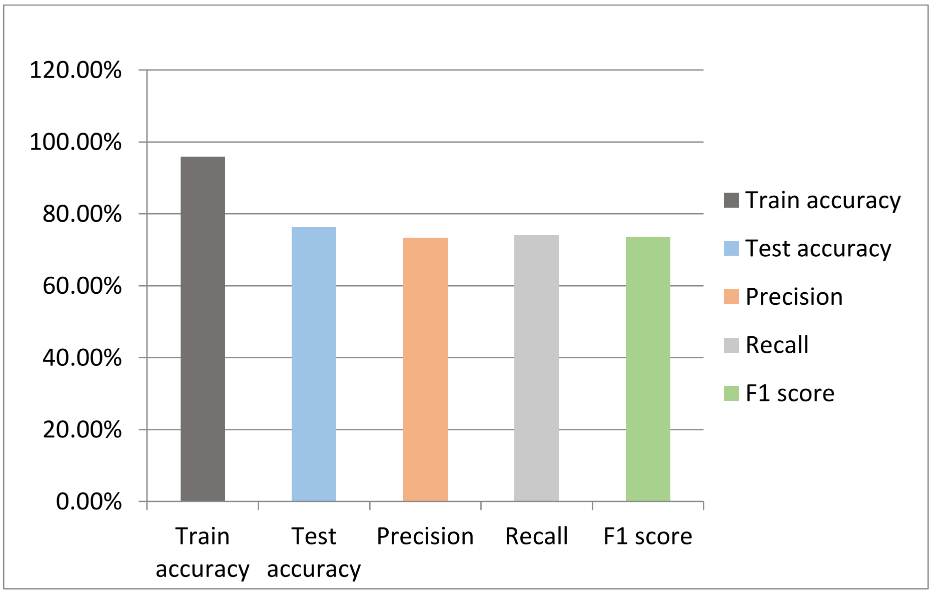


Fig 4.5: Chart of Classification Report for LSTM model

4.2.2 Analysis of Bi-LSTM

In Bi-LSTM model we find training accuracy, test accuracy, Loss, confusion matrix, Precision call, f1-score and Recall. We discuss every point in below:

1. **Epochs Vs Train and Validation accuracy**: After building model we train the model with train dataset. We were doing 30 epochs in this model.

```
Epoch 1/30
279/279 [==============================] - 25s 42ms/step - loss: 1.3448 - accuracy: 0.4757 - val_loss: 1.3650 - val_accuracy: 0.5464
Epoch 2/30
279/279 [==============================] - 5s 18ms/step - loss: 0.8032 - accuracy: 0.7451 - val_loss: 0.8020 - val_accuracy: 0.7379
Epoch 3/30
279/279 [==============================] - 4s 13ms/step - loss: 0.5927 - accuracy: 0.8202 - val_loss: 0.7838 - val_accuracy: 0.7520
Epoch 4/30
279/279 [==============================] - 3s 12ms/step - loss: 0.4508 - accuracy: 0.8658 - val_loss: 0.8425 - val_accuracy: 0.7560
Epoch 5/30
279/279 [==============================] - 4s 14ms/step - loss: 0.3622 - accuracy: 0.8949 - val_loss: 1.0062 - val_accuracy: 0.7379
```

Fig 4.6: Training and Validation Accuracy in first 5 epochs for Bi-LSTM

Then this model gives us training accuracy and validation accuracy which curves shows in fig 4.7.

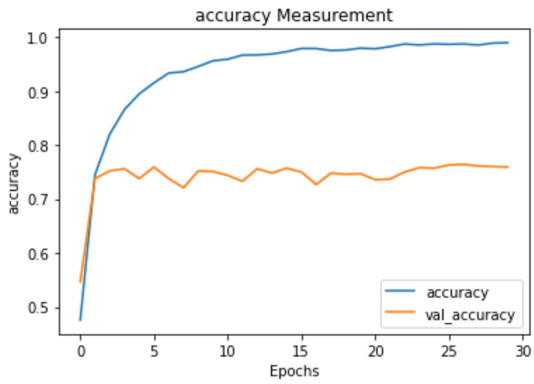


Fig 4.7: Training and Validation Accuracy curve for Bi-LSTM

Here we perform 30 epochs to train our Bi-LSTM model. The training accuracy increasing in every epoch and we got 97.96% training accuracy in last epoch and validation accuracy also increased but not same as training. We got 77.91% validation accuracy in last epoch.

2. **Confusion Matrix:** In this section we will describe confusion matrix of Bi-LSTM, which is shown in following page in fig 4.8.

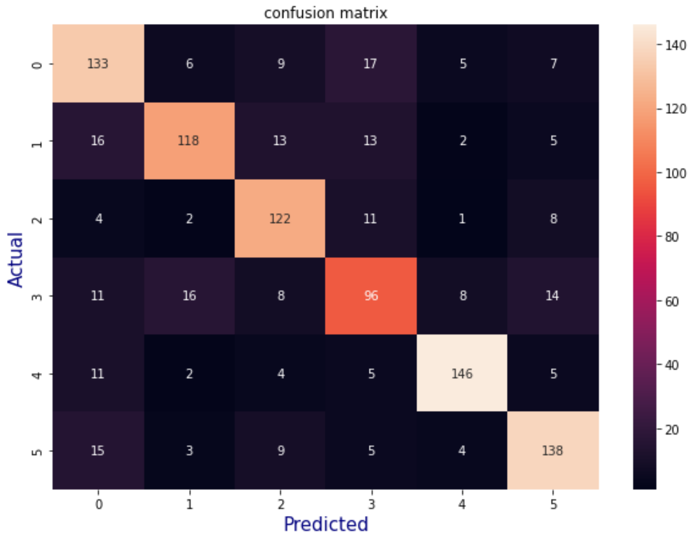


Fig 4.8: Confusion Matrix for Bi-LSTM model

3. **Classification report (precision, recall and f1 score) :** For the headline classification model, the classification report is given in fig 4.9:

	precision	recall	f1-score	support
0	0.70	0.75	0.72	177
1	0.80	0.71	0.75	167
2	0.74	0.82	0.78	148
3	0.65	0.63	0.64	153
4	0.88	0.84	0.86	173
5	0.78	0.79	0.79	174
accuracy			0.76	992
macro avg	0.76	0.76	0.76	992
weighted avg	0.76	0.76	0.76	992

Fig 4.9: Classification Report for Bi-LSTM model

Here we find out the precision, recall and f1 score. We got weighted average of 76.15% precision, 76.89% recall and 76.66% f1 score (table 4.2).

Table 4.2: Over all accuracy matrix for Bi-LSTM model

Train accuracy	97.96%
Test accuracy	77.91%
Precision	76.15%
Recall	76.89%
F1 score	76.66%

We can plot the accuracy on chart. The figure shows in the following page (fig 4.10):

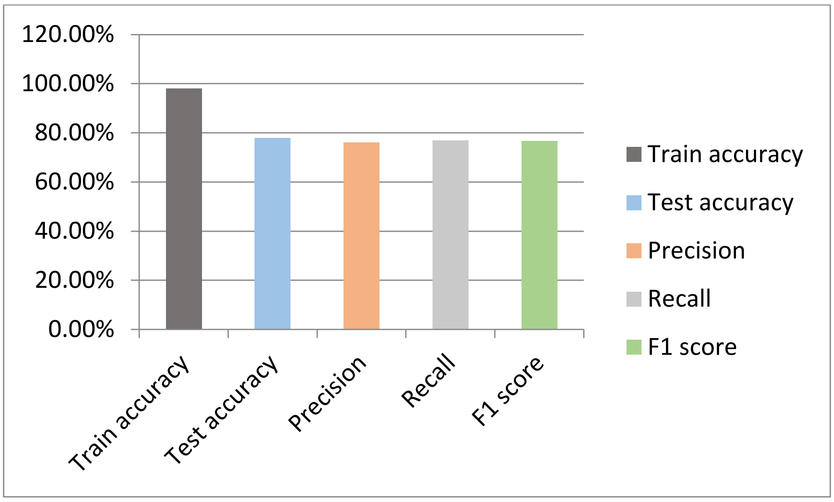


Fig 4.10: Chart of Classification Report for Bi-LSTM model

4.2.3 Analysis of Bi-GRU

In Bi-GRU model we find training accuracy, test accuracy, Loss, confusion matrix, Precision call, f1-score and Recall. We discuss every point in below:

1. **Epochs Vs Train and Validation accuracy**: After building model we train the model with train dataset. We were doing 20 epochs in this model.

```
Epoch 1/20
279/279 [==============================] - 18s 39ms/step - loss: 1.0473 - accuracy: 0.5991 - val_loss: 0.6819 - val_accuracy: 0.7621
Epoch 2/20
279/279 [==============================] - 3s 11ms/step - loss: 0.4708 - accuracy: 0.8446 - val_loss: 0.7178 - val_accuracy: 0.7591
Epoch 3/20
279/279 [==============================] - 4s 14ms/step - loss: 0.2729 - accuracy: 0.9112 - val_loss: 0.8449 - val_accuracy: 0.7470
Epoch 4/20
279/279 [==============================] - 3s 11ms/step - loss: 0.1784 - accuracy: 0.9434 - val_loss: 0.9909 - val_accuracy: 0.7480
Epoch 5/20
279/279 [==============================] - 3s 11ms/step - loss: 0.1253 - accuracy: 0.9618 - val_loss: 1.1103 - val_accuracy: 0.7369
```

Fig 4.11: Training and Validation Accuracy in first 5 epochs for Bi-GRU

Then this model gives us training accuracy and validation accuracy which curves shows in below fig 4.12.

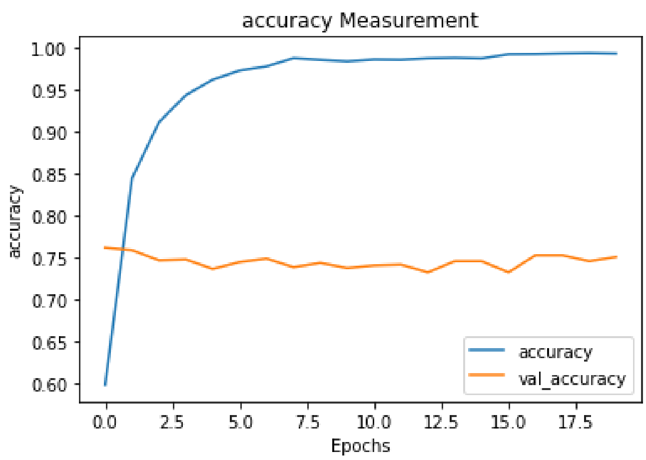


Fig 4.12: Training and Validation Accuracy curve for Bi-GRU

Here we perform 20 epochs to train our Bi-GRU model. The training accuracy increasing in every epoch and we got 97.28% training accuracy in last epoch and validation accuracy also increased but not same as training. We got 76.10% validation accuracy in last epoch.

2. **Confusion Matrix:** In this section we will describe confusion matrix of Bi-GRU, shown in figure 4.13.

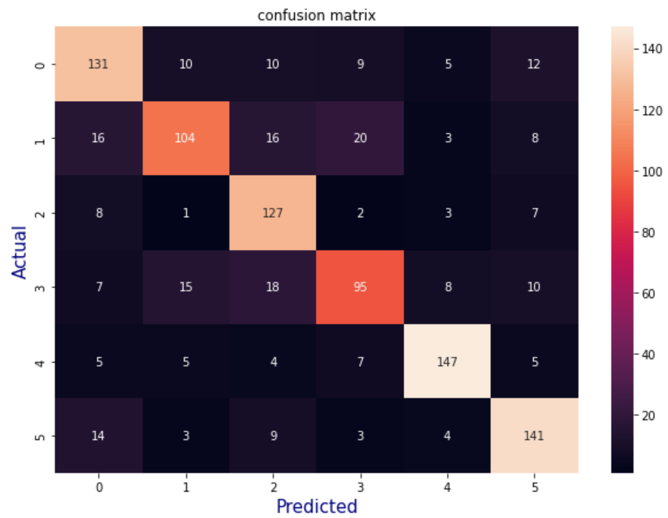


Fig 4.13: Confusion Matrix for Bi-GRU model

3. **Classification report (precision, recall and f1 score) :** For headline classification model the classification report given bellow:

```
              precision    recall  f1-score   support

           0       0.72      0.74      0.73       177
           1       0.75      0.62      0.68       167
           2       0.69      0.86      0.77       148
           3       0.70      0.62      0.66       153
           4       0.86      0.85      0.86       173
           5       0.77      0.81      0.79       174

    accuracy                           0.75       992
   macro avg       0.75      0.75      0.75       992
weighted avg       0.75      0.75      0.75       992
```

Fig 4.14: Classification Report for Bi-GRU model

Here we find out the precision, recall and f1 score. We got weighted average of 75.09% precision, 75.91% recall and 75.59% f1 score (table 4.3).

Table 4.3: Over all accuracy matrix for Bi-GRU model

Train accuracy	97.28%
Test accuracy	76.10%
Precision	75.09%
Recall	75.91%
F1 score	75.59%

We can plot these accuracy on chart. The figure shows in fig 4.15:

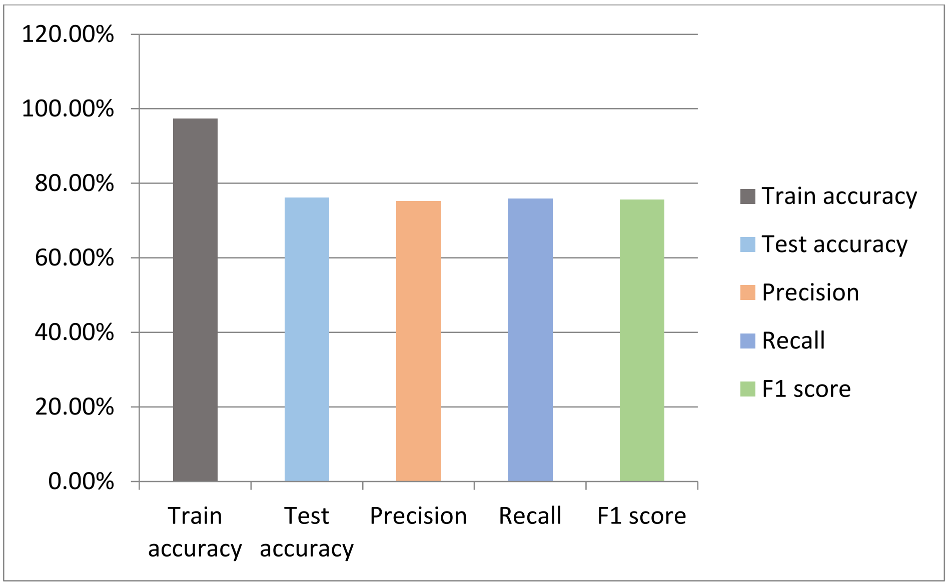


Fig 4.15: Chart of Classification Report for Bi-GRU model

4.2.4 Comparison among Model Performance

We use three model of deep learning. Here we compare validation accuracy and train accuracy each of model. Table 4.4 shows accuracy for each model.

Table 4.4: Over all comparison between LSTM, Bi-LSTM and Bi-GRU model

Algorithm	Train Accuracy	Test Accuracy
Long Short Term Memory	95.85%	73.28%
Bi- Long Short Term Memory	97.95%	75.90%
Bi -Gated Recurrent Unit	97.28%	75.10%

- **Training accuracy comparison:** Model trained with train dataset (which is 90% of the full dataset). Where this 90% data will use to train for the model. The figure 4.16 describes the train accuracy with each model.

	Algorithms	Score
0	BiLSTM	0.979600
1	BiGRU	0.979600
2	LSTM	0.958600

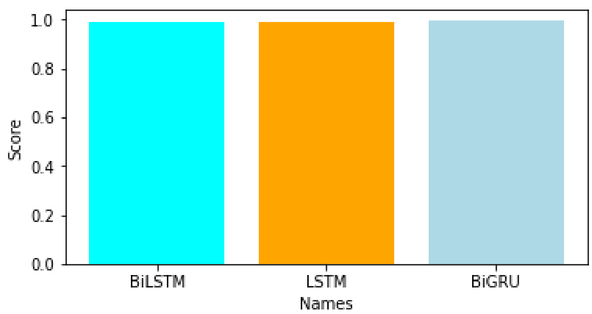


Fig 4.16: Accuracy of each model

- **Testing accuracy comparison:** After the Model trained, now perform the test data in this model for prediction. The test dataset (10% of full data set) was already preprocessed. Then test data use to testing the model. The figure 4.17 describes the test accuracy with each model.

	Algorithms	Score
0	BiLSTM	0.7791
1	LSTM	0.7629
2	BiGRU	0.7610

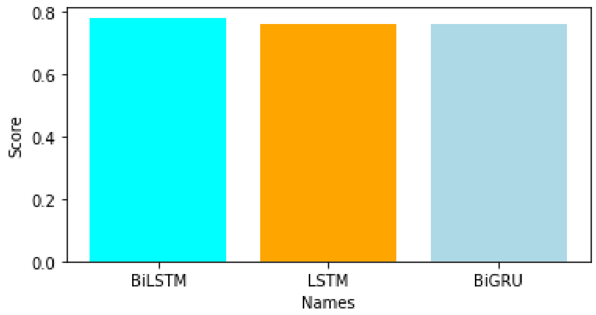


Fig 4.17: Accuracy of each model

4.3 Discussion

All of the above analysis shows that Bi-LSTM outperforms all three deep learning algorithms. It has a training accuracy of 97.96% and a validation accuracy of 77.91% when classifying Bangla newspaper headlines. Where the precision score, recall and f1 score are 76.15%, 76.89% and 76.66% respectively.

Bi-GRU and LSTM model give second-best training accuracy which is close to 97.28% and 95.86% respectively, and validation accuracy is close to 77%.

The performance of algorithms varies with data size. If our dataset contains more data, it will undoubtedly have more to train and thus may provide better accuracy. It also depends on data preprocessing; if some of the preprocessing steps are skipped, the accuracy suffers.

CHAPTER 5

CONCLUSION

5.1 Conclusion

A very efficient system for Bengali News Headline classification is a challenging task. The Bengali language has numerous number of diversity and it is rich. Test is applied on Word2Vec method. The dataset is created in a way that covers most of the words that are relevant with the classes. In this thesis, you have explored the problem of Bangla newspaper headline classification using LSTM, Bi-LSTM, and Bi-GRU.

The models were trained and evaluated on a dataset of Bangla newspaper headlines, and the results were analyzed. However, the model faced overfitting, which is a common issue in deep learning models. To mitigate overfitting, regularization techniques were employed, but further work may be required to improve the model's performance. In this thesis, we have explored the problem of Bangla newspaper headline classification using LSTM, Bi-LSTM, and Bi-GRU models. We have trained and evaluated these models on a dataset of Bangla newspaper headlines, and the results have shown that the models can achieve reasonable accuracy on the classification task. However, the models have also faced the issue of overfitting, which is a common problem in deep learning models. We have employed several techniques to mitigate overfitting, including regularization techniques like dropout and L2 regularization.

We have also experimented with different hyperparameters to find the best combination that achieves the best performance on the validation set. However, we have found that despite these efforts, the models still suffer from overfitting, which suggests that more work is needed to improve their performance. In addition to the technical challenges we faced in this thesis, we have also encountered challenges related to the nature of the dataset. Specifically, we found that the dataset is relatively small and lacks diversity, which can limit the models' ability to generalize to new data. We have attempted to mitigate this issue by using data augmentation techniques, but we believe that a larger and more diverse dataset would be necessary to achieve better performance.

We faced some problem in these three models. My models are faced over fit due to data length, data preprocessing and tokenizer. Overfitting is a common issue when training machine learning models and it occurs when the model learns the training data too well and performs poorly on new, unseen data. Overfitting can occur in deep learning models in LSTMs, Bi-LSTMs, and Bi-GRUs if the models are too complex or if the training data is insufficient.

1) Model Complexity: A complex model like LSTM, Bi-LSTM, or Bi-GRU has a large number of parameters, which can lead to overfitting if the model is not regularized correctly.
2) Insufficient Training Data: If the amount of training data is not sufficient, the model may learn the training data too well, leading to overfitting.

3) Overfitting due to hyperparameters: The hyperparameters like the learning rate, batch size, and number of epochs may be causing overfitting.

There are several way to remove the over fit-

1) Reduce Model Complexity: You can reduce the complexity of the model by decreasing the number of parameters or adding regularization techniques like Dropout, L1/L2 regularization.
2) Increase Training Data: If you can increase the amount of training data, it can help reduce overfitting.
3) Early Stopping: You can monitor the validation loss during training and stop the training early if the validation loss starts to increase.
4) Data Augmentation: Data augmentation techniques like flipping, rotation, and zooming can help increase the amount of training data and reduce overfitting.
5) Hyperparameter Tuning: You can try different values of hyperparameters and evaluate the model performance on a validation set to select the best hyperparameters that prevent overfitting.

In conclusion, the Bangla newspaper headline classification problem presents a significant challenge, and more work is needed to develop models that can achieve state-of-the-art performance. The techniques we have employed in this thesis have helped mitigate overfitting to some extent, but further research is needed to develop more robust models. We believe that future work should focus on collecting larger and more diverse datasets, using transfer learning techniques, and investigating ensemble learning and explainability and interpretability techniques to improve the models' performance and understand how they are making their predictions. The paper proposes Bengali news headline categorization with optimized Deep Learning [17].

5.2 Future Work

In the future the work can be implemented to the inner method of the existing work. Some other algorithms such as the Greedy algorithm can be implemented to check if it performs well than the neural network. Classification can also be done in other languages such as Urdu. One can also create their own stemmer which can expand classification accuracy [3]. There are several avenues for future work that can improve the performance of Bangla newspaper headline classification models:

1) Larger and Diverse Dataset: Collecting a larger and more diverse dataset can help improve the generalization of the model.
2) Transfer Learning: Pre-training on a large corpus of Bangla text can help improve the model's performance on the headline classification task.

3) Ensemble Learning: Using an ensemble of multiple models with different architectures can help improve the overall performance.
4) Fine-tuning Hyperparameters: Fine-tuning the hyperparameters of the models can help improve their performance and prevent overfitting.
5) Explainability and Interpretability: Investigating the models' inner workings to understand how they are making their predictions and which features they are using can help improve the models' interpretability and explainability.

Overall, the Bangla newspaper headline classification problem presents a significant challenge, and further work is required to develop models that can achieve state-of-the-art performance.

References

[1] C. Wang, P. Nulty, and D. Lillis, “A Comparative Study on Word Embeddings in Deep Learning for Text Classification,” *ACM Int. Conf. Proceeding Ser.*, pp. 37–46, 2020, doi: 10.1145/3443279.3443304.

[2] J. U. Rahman, S. Ali, A. Khan, and M. Fayaz, “A Deep Ensemble Model for News Classification on Social Media,” *2021 Int. Consewakf. Cyber Warf. Secur. ICCWS 2021 - Proc.*, no. November 2021, pp. 57–62, 2021, doi: 10.1109/ICCWS53234.2021.9702934.

[3] J. X. Chen, D. M. Jiang, and Y. N. Zhang, “A Hierarchical Bidirectional GRU Model with Attention for EEG-Based Emotion Classification,” *IEEE Access*, vol. 7, pp. 118530–118540, 2019, doi: 10.1109/ACCESS.2019.2936817.

[4] H. Lu, X. Liu, Y. Yin, and Z. Chen, “A patent text classification model based on multivariate neural network fusion,” *2019 6th Int. Conf. Soft Comput. Mach. Intell. ISCMI 2019*, pp. 61–65, 2019, doi: 10.1109/ISCMI47871.2019.9004335.

[5] H. A. Galal Elsayed, S. Chaffar, S. Brahim Belhaouari, and H. Raissouli, “A two-level deep learning approach for emotion recognition in Arabic news headlines,” *Int. J. Comput. Appl.*, 2020, doi: 10.1080/1206212X.2020.1851501.

[6] M. K. Md. Rafiuzzaman Bhuiyan and and S. A. Abu Kaisar Mohammad Masum, Syed Akhter Hossain, *An Approach for Bengali News Headline Classification Using LSTM*, vol. 1. 2021. doi: 10.1007/978-981-15-9927-9_86.

[7] M. Sewak, S. K. Sahay, and H. Rathore, “An overview of deep learning architecture of deep neural networks and autoencoders,” *J. Comput. Theor. Nanosci.*, vol. 17, no. 1, pp. 182–188, 2020, doi: 10.1166/jctn.2020.8648.

[8] J. Deng, L. Cheng, and Z. Wang, “Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification,” *Comput. Speech Lang.*, vol. 68, p. 101182, 2021, doi: 10.1016/j.csl.2020.101182.

[9] 4. Chowdhury, A., “Bangla News Classification Using Deep Learning Techniques,” *Int. Conf. Electr. Comput. Commun. Eng.*, pp. 1–6, doi: 10.1109/ICECCE49338.2020.9219776.

[10] et al. 9. Ahmed, S., “Bangla News Classification using Deep Learning Techniques: A Comparative Study.,” *Proc. Int. Conf. Electr. Comput. Commun. Eng.*, pp. 1–6, doi: DOI: 10.1109/ICECCE53006.2021.9563427.

[11] A. Hossain, N. Chaudhary, Z. Hasan Rifad, and B. M. M. Hossain, “Bangla News Headline Categorization,” *Int. J. Educ. Manag. Eng.*, vol. 11, no. 6, pp. 39–48, 2021, doi: 10.5815/ijeme.2021.06.05.

[12] M. M. Alif, M. A. H., & Hassan, “Bengali Text Classification using Convolutional Neural Networks.,” *Int. J. Comput. Appl.*, vol. 181(14), 2018.

[13] R. Lu and Z. Duan, “Bidirectional GRU for sound event detection,” *Dcase*, no. November, pp. 0–3, 2017.

[14] & H. Nafis, A., “Classification of Bengali Text Using Deep Learning Techniques,” *J. Phys.*, vol. 1590, 2020, doi: 10.1088/1742-6596/1590/1/012044.

[15] S. I. Shadikur Rahman, Syeda Sumbul Hossain(B), F. B. R. Mazharul Islam Chowdhury, and and K. B. M. Badruzzaman, *Context-Based News Headlines Analysis Using Machine Learning Approach*, vol. 48, no. C. 1995. doi: 10.1016/S0091-679X(08)61375-8.

[16] & I. 5. Hossain, M. S., “Deep Learning-Based Text Classification for Bengali News.,” Int. Conf. Electr. Comput. Commun. Eng., pp. 1–6, 2021, doi: 10.1109/ICECCE53006.2021.9563487.

[17] G. Bharath, K. J. Manikanta, G. B. Prakash, R. Sumathi, and P. Chinnasamy, “Detecting Fake News Using Machine Learning Algorithms,” *2021 Int. Conf. Comput. Commun. Informatics, ICCCI 2021*, pp. 1–5, 2021, doi: 10.1109/ICCCI50826.2021.9402470.

[18] M. Rabib, S. Sarkar, and M. Rahman, “Different Machine Learning based Approaches of Baseline and Deep Learning Models for Bengali News Categorization,” *Int. J. Comput. Appl.*, vol. 176, no. 18, pp. 10–16, 2020, doi: 10.5120/ijca2020920107.

[19] D. K. Nayak and B. K. Bolla, *Efficient Deep Learning Methods for Sarcasm Detection of News Headlines*, vol. 269, no. June. Springer Singapore, 2022. doi: 10.1007/978-981-16-7996-4_26.

[20] J. Zhang, F. Liu, W. Xu, and H. Yu, “Feature fusion text classification model combining CNN and BiGRU with multi-attention mechanism,” *Futur. Internet*, vol. 11, no. 11, 2019, doi: 10.3390/fi11110237.

[21] L. Zhou and X. Bian, “Improved text sentiment classification method based on BiGRU-Attention,” *J. Phys. Conf. Ser.*, vol. 1345, no. 3, 2019, doi: 10.1088/1742-6596/1345/3/032097.

[22] L. Zhou and X. Bian, “Improved text sentiment classification method based on BiGRU-Attention,” *J. Phys. Conf. Ser.*, vol. 1345, no. 3, 2019, doi: 10.1088/1742-6596/1345/3/032097.

[23] A. K. and S. Shrawne, “Location-Wise News Headlines Classification and Sentiment Analysis: A Deep Learning Approach,” pp. 389–397, 2019, doi: 10.1007/978-981-15-0633-8_27.

[24] A. Kabra and S. Shrawne, “Location-Wise News Headlines Classification and Sentiment Analysis: A Deep Learning Approach,” pp. 383–391, 2020, doi: 10.1007/978-981-15-0633-8_37.

[25] S. Yeasmin, R. Kuri, A. R. M. Mahamudul Hasan Rana, A. Uddin, A. Q. M. Sala Uddin Pathan, and H. Riaz, “Multi-category Bangla News Classification using Machine Learning Classifiers and Multi-layer Dense Neural Network,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 5, pp. 757–767, 2021, doi: 10.14569/IJACSA.2021.0120588.

[26] S. A. Althubiti, F. Alenezi, and R. F. Mansour, “Natural Language Processing with Optimal Deep Learning Based Fake News Classification,” *Comput. Mater. Contin.*, vol. 73, no. 2, pp. 3529–3544, 2022, doi: 10.32604/cmc.2022.028981.

[27] Y. Zhang and Z. Rao, “N-BiLSTM: BiLSTM with n-gram Features for Text Classification,” *Proc. 2020 IEEE 5th Inf. Technol. Mechatronics Eng. Conf. ITOEC 2020*, no. Itoec, pp. 1056–1059, 2020, doi: 10.1109/ITOEC49072.2020.9141692.

[28] S. A. Khushbu, A. K. M. Masum, S. Abujar, and S. A. Hossain, “Neural Network Based Bengali News Headline Multi Classification System: Selection of Features describes Comparative Performance,” *2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020*, pp. 2–7, 2020, doi: 10.1109/ICCCNT49239.2020.9225611.

[29] G. Kaur and K. Bajaj, "News Classification using Neural Networks," *Commun. Appl. Electron.*, vol. 5, no. 1, pp. 42–45, 2016, doi: 10.5120/cae2016652224.

[30] J. Duan, H. Zhao, W. Qin, M. Qiu, and M. Liu, "News Text Classification Based on MLCNN and BiGRU Hybrid Neural Network," *Proc. - 2020 3rd Int. Conf. Smart BlockChain, SmartBlock 2020*, pp. 137–142, 2020, doi: 10.1109/SmartBlock52591.2020.00032.

[31] M. Moshinsky, "Arabic Article classification on Deep learning techniques," *Nucl. Phys.*, vol. 13, no. 1, pp. 104–116, 1959.

[32] J. S. A. Patro, M. Patel, R. Shukla, "Real Time News Classification Using Machine Learning," *Ijast*, no. June, pp. 620–630, 2020.

[33] Q. Yu, Z. Wang, and K. Jiang, "Research on Text Classification Based on BERT-BiGRU Model," *J. Phys. Conf. Ser.*, vol. 1746, no. 1, 2021, doi: 10.1088/1742-6596/1746/1/012019.

[34] M. M. Akter, T., & Molla, "Sentiment Analysis of Bangla Newspaper Articles Using LSTM," *J. Phys. Conf. Ser.*, vol. 1529(1), 2020.

[35] M. R. 7. Uddin, "Sentiment Analysis of Bengali News using Deep Learning Techniques.," *Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng.*, pp. 1–6, 2020, doi: DOI: 10.1109/IC4ME2-2020.2020.9330426.

[36] W. Li, S. Gao, H. Zhou, Z. Huang, K. Zhang, and W. Li, "The automatic text classification method based on bert and feature union," *Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS*, vol. 2019-Decem, pp. 774–777, 2019, doi: 10.1109/ICPADS47876.2019.00114.

[37] M. I. Rana, S. Khalid, and M. U. Akbar, "News classification based on their headlines: A review," 17th IEEE Int. Multi Top. Conf. Collab. Sustain. Dev. Technol. IEEE INMIC 2014 - Proc., pp. 211–216, 2015, doi: 10.1109/INMIC.2014.7097339.

[38] P. Dhar and M. Z. Abedin, "Bengali News Headline Categorization Using Optimized Machine Learning Pipeline," Int. J. Inf. Eng. Electron. Bus., vol. 13, no. 1, pp. 15–24, 2021, doi: 10.5815/ijieeb.2021.01.02.