

IDENTIFYING HATE SPEECH IN BANGLISH LANGUAGE

By
Ayesha Siddika
ID: 211-15-14583

Sumaiya Akter
ID: 211-15-14606

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Dr. Md Zahid Hasan
Associate Professor
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

January 12, 2025

APPROVAL

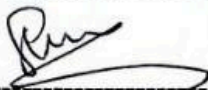
This Project titled “Identifying Hate Speech in Banglish Language”, submitted by Ayesha Siddika, ID No: 211-15-14583 and Sumaiya Akter, ID No: 211-15-14606 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment 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 12 January, 2025.

BOARD OF EXAMINERS



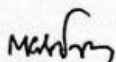
Dr. Sheak Rashed Haider Noori (SRH)
Professor & Head Chairman,
Department of Computer Science and Engineering.
Faculty of Science and Information technology.
Daffodil International University.

Chairman



Sharmin Akter (SNA)
Assistant Professor,
Department of Computer Science and Engineering.
Faculty of Science and Information technology.
Daffodil International University.

Internal Examiner



Mr. Md Mohammad Masum Bakaul (MB)
Sr. Lecturer,
Department of Computer Science and Engineering.
Faculty of Science and Information technology.
Daffodil International University.

Internal Examiner



Dr. Md. Zulfiker Mahmud (ZM)
Professor,
Department of Computer Science and Engineering.
Jagannath University.

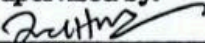
External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Professor & Head, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by: 

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of CSE
Daffodil International University

Co-Supervised by: 

Dr. Md Zahid Hasan
Associate Professor
Department of CSE
Daffodil International University

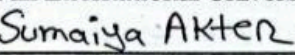
Submitted by: 

Ayesha Siddika

ID: 211-15-14583

Department of CSE

Daffodil International University



Sumaiya Akter

ID: 211-15-14606

Department of CSE

Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratitude to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dr. Sheak Rashed Haider Noori**, Professor & Head, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Natural Language Processing** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

Social media's growth has facilitated the quick spread of both positive and negative content, and hate speech is one of the most harmful forms of online expression. The goal of this study is to detect hate speech in Banglish, a code-mixed language commonly used in social media conversations that blends English and Bangla. The research is about to develop a machine learning and deep learning-based system for identifying hate speech in Banglish in order to get beyond the unique challenges caused by the combination of languages and everyday idioms. A collection of Banglish text from several social media platforms was preprocessed using techniques like tokenization, lemmatization, and normalization. The expansion of words such as "pic" to "picture", "img" to "image" and "u" to "you" The method of lemmatization decreased linguistic variances. We have applied six deep learning models which are LSTM, GRU, Bi-LSTM, Bi-GRU, GRU+LSTM and Bi-LSTM+Bi-GRU. These models' performance was evaluated using confusion matrices, F1-score, accuracy, precision, and recall. The multiclass classification job has an accuracy of 83.93% in "Bi-LSTM+Bi-GRU" model which is a hybrid model in differentiating between seven different classification types. In particular, this study advances automated hate speech detection systems for code-mixed languages like Banglish. The findings indicate that while current models have promise, more study is needed to address problems including data imbalance and the identification of more subtle kinds of hate speech. Future initiatives to improve the precision and robustness of hate speech detection systems across other languages are made possible by this research.

Table of Contents

Approval	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Introduction.....	1
1.2 Motivation.....	2
1.3 Objectives.....	2
1.4 Methodology.....	3
1.5 Project Outcome.....	3
1.6 Organization of the Report.....	4
2 Background	5
2.1 Introduction.....	5
2.2 Literature Review.....	6
2.3 Gap Analysis.....	10
2.4 Summary.....	11
3 Research Methodology	12
3.1 Methodology.....	12
3.1.1 Overview.....	12
3.1.2 Proposed Methodology.....	13
3.2 Detailed Methodology.....	15
3.3 Project Plan.....	16
3.4 Task Allocation.....	16
3.5 Summary.....	17
4 Implementation and Results	18
4.1 Environment Setup.....	18
4.2 Performance Analysis.....	18
4.2.1 Model Description.....	19

4.3	Results and Discussion.....	25
4.4	Summary.....	28
5	Engineering Standards and Design Challenges	29
5.1	Compliance with the Standards.....	29
5.1.1	Software Standards.....	29
5.1.2	Hardware Standards.....	30
5.1.3	Communication Standards.....	30
5.2	Impact on Society, Environment and Sustainability.....	31
5.2.1	Impact on Life.....	31
5.2.2	Impact on Society & Environment.....	32
5.2.3	Ethical Aspects.....	33
5.2.4	Sustainability Plan.....	35
5.3	Project Management and Financial Analysis.....	37
5.4	Complex Engineering Problem.....	39
5.4.1	Complex Problem Solving.....	40
5.4.2	Engineering Activities.....	41
5.5	Summary.....	43
6	Conclusion	44
6.1	Summary.....	44
6.2	Limitation.....	45
6.3	Future Work.....	46
	References	47

List of Figures

3.1	Proposed Methodology diagram	14
4.2.1.1	LSTM model architecture	20
4.2.1.2	GRU model architecture	20
4.2.1.3	Bi-GRU model architecture	21
4.2.1.4	Bi-LSTM model architecture	21
4.2.1.5	GRU+LSTM model	22
4.2.1.6	Bi-LSTM+Bi-GRU model	22
4.3.1	Accuracy and loss graph of LSTM	25
4.3.2	Accuracy and loss graph of GRU	25
4.3.3	Accuracy and loss graph of Bi-LSTM	26
4.3.4	Accuracy and loss graph of Bi-GRU	26
4.3.5	Accuracy and loss graph of GRU+LSTM	26
4.3.6	Accuracy and loss graph of Bi-LSTM+Bi-GRU	27

List of Tables

2.3	Gap analysis of related works	10
4.2.1	Comparative Analysis of Models	22
4.3	Final results of Models	27
5.3.1	Financial analysis	37
5.3.2	Project Management Table	38
5.1	Mapping with complex problem solving.	39
5.2	Mapping with knowledge Profile.	40
5.3	Mapping with complex engineering activities.	42

Chapter 1

Introduction

This chapter will cover all the basic introductions about our research.

1.1 Introduction

In the past few years, figuring out hate speech has become super important, especially in languages that are low in resources. This is true that identifying hate speech in low-resource and mixed language is a critical task in Natural Language Processing (NLP). Let's look at Banglish, which is a mix of Bangla and English. Identifying hate speech here is just like trying to do it in Hinglish, another mixed-code language. There's a big rise in social media users and competing with that the amount of hate comments is also going up. Moreover, the kind of hate speech changes based on time and context. So, it is a continuous task to update the model of identifying hate speech models to prevent hate speech on social media especially in the comment section. A lot of research has been done on finding harmful content in languages we commonly speak. Such as: English, Hindi, Spanish even in Bangla. But hybrid languages, such as Banglish which mixes Bangla words with the English alphabet haven't been studied as much. Banglish is popular in Bangladesh, especially on social media sites. However, identifying hate speech can be really tricky due to its casual style and lack of clear rules as well as low resources. Banglish is still kind of mysterious. That's what got us thinking about this study. A key idea in natural language processing is sentiment analysis. It's super useful for figuring out the feelings behind the words in text. So, our goal is to make online safer and promote better digital interactions by using machine learning specifically for Banglish. Since machine learning can handle the tricky parts of Banglish's informal style, using it is really important for our project. We want to fill a big gap in what people know about the models. This research will propose an effective machine learning model to identify Banglish hate speech.

1.2 Motivation

The growing impact of hate speech in our life is increasing very rapidly which negatively impacts both individuals and society. Banglish is a combination of Bangla and English language which has gained popularity as communication moves online. And especially among young people. However, traditional models find it especially difficult to detect hate speech in Banglish due to its informal nature, frequent code-switching and extensive use of slang. Closing this gap is essential to fostering a more secure and welcoming online community and shielding marginalized groups from the emotional and psychological damage that toxic content can ruin. The necessity to create a sophisticated hate speech detection system designed especially for Banglish is what motivates this project. allowing me to make a significant contribution to the impact of technology on society. This study advances natural language processing (NLP) and establishes the foundation for future research on code-mixed languages by addressing linguistic issues like tokenization, lemmatization and managing imbalanced datasets. This research improves my knowledge of machine learning, deep learning and natural language processing which is very beneficial to me personally. Not only that but also It can save a lot of people who do online business from online bullies.

1.3 Objectives

The main objective of this research is to develop an efficient and accurate Natural Language Processing model that will help to identify hate speech in Banglish (Bangla language written using English letter) in different social media platforms. Our ultimate goal is to identify hate speech in mixed Language and make a contribution to society reducing online based hate comments specially for the working women who face thousands of hate comments on social media and harrassed by the people.

- Examine the difficulties in identifying hate speech in Banglish, which is Bangla written in English. There are a lot of banglish words which are used in different social platforms nowadays.
- For increasing the detection accuracy, construct and assess models .
- For addressing linguistic challenges.
- Make sure to hate speech detection.

- Use the detection system to lessen hate speech online and encourage civil online communication.
- Specifically, we will focus on the business persons who are running their businesses online.

1.4 Methodology

This research technique employs a methodical blueprint of the difficulties of identifying hate speech in the code-mixed language of Banglish. We will be using the dataset from kaggle. Online platforms were used to gather the data in the kaggle, guaranteeing a varied representation of hate speech in Banglish writing. In order to get the data ready for analysis, preprocessing is a must need. To normalize the text and handle common Banglish problems like slang, misspellings and non-standard language, different techniques have to be used here. Such as tokenization, lemmatization, and cleaning will be used. Oversampling techniques will be used to address the class imbalance and ensure that minority classes will be adequately represented. For multiclass classification, six deep learning models will be employed. Among them two models will be hybrid. In order to address the hybrid nature of Banglish and capture contextual meanings, word embeddings will be used. To assess the generalizability of the model, the data will be divided into training and testing subsets in an 80:20 ratio. Metrics like accuracy, precision, recall and F1-score will be used to evaluate the model's performance. Not only that but also they will increase the accuracy for marginalized categories. By establishing an efficient framework for hate speech identification in Banglish, this all-inclusive methodology seeks to improve the field of natural language processing.

1.5 Project Outcome

Six deep learning models that can detect any hate speech in Banglish language with great performance and accuracy in the expected result of our research. One of the seven primarily estimated categories such as: appearance, non-hate, religious, racial, others, sexual and slang will be able to be categorized by the model from any input comments. Not only that but also we will propose a model according to their performances and accuracy.

1.6 Organization of the Report

Several chapters are included in this report to give the readers a clear and detailed knowledge about the study. Here we can see that in chapter 1, this report is giving the basic knowledge about our study which is the foundation of our report. Chapter 2 will provide all the background studies about this topic. Not only that but also it will also cover the gap analysis of our study. In chapter 3, we will show our methodology step by step with a detailed diagram. Then in chapter 4, we will discuss our implemented result and the performance of our model. In chapter 5, we will discuss the engineering standards and design challenges. Here we will discuss the impact of our study not only in the society but also in our life and after that we will also write down our justification about complex engineering with our work. And finally we will discuss the conclusion and future work of our study.

Chapter 2

Background

This chapter will represent the whole background of our research on our topic which is “Hate Speech Detection in Banglish Language”.

2.1 Introduction

In any research, background means the part where the context and fundamental information are given about a particular topic. It also explains the broader area of research and the current state of any topic. It also summarizes the findings, studies and theories of any research topic. Not only that but also it interprets the gap of any research topic. In our topic, Over the previous years, researchers have worked with English Language, recently working on one multilanguage like Bangla, Hindi, Tamil etc. They have even done research on mixed languages like tamil+English, Hindi+English, Telugu+English but there are a few works on Banglish Language. Studies in Bangla and Additional Languages has several papers from 2010 to 2024. And they have accuracy from 57% to 90%. All the recent studies shared some neat ideas about different machine learning (ML) and deep learning (DL) models. They really paid special attention to three methods like zero-shot, one-shot and few-shot learning. These methods are very important because sometimes we don't have enough labeled data available. Collecting that data can be a long process and take lots of effort as well. They tackle the problem of scarce labeled data well. The study examined how effective zero-shot, one-shot, and few-shot learning could be at identifying hate speech in Hinglish comments on YouTube. This relates to the issues faced in Banglish too. It showed that transfer learning can really help the models figure out comments without needing huge amounts of labeled data. In fact, zero-shot learning had an accuracy rate of 54%. That's not bad at all when dealing with smaller datasets. As per our topic, there is no research paper about hate speech detection in multiclass classification approach. This research will make a contribution in detecting hate speech in seven different categories.

2.2 Literature Review

Here is the literature reviews are given below:

Over the previous years, researchers have worked with English Language, recently working on one multilanguage like Bangla, Hindi ,Tamil etc. They have even done research on mixed languages like tamil+English, Hindi+English, Telugu+English but there are a few works on Banglish Language. Studies in Bangla and Additional Languages has several papers from 2010 to 2024. And they have accuracy from 57% to 90%. All the recent studies shared some neat ideas about different machine learning (ML) and deep learning (DL) models. They really paid special attention to three methods: zero-shot, one-shot, & few-shot learning. These methods are very important because sometimes we don't have enough labeled data available. Collecting that data can be a long process and take lots of effort as well. Large models like Bidirectional Auto-Regressive Transformers (BART) or GPT-3 have proven to be really helpful. They tackle the problem of scarce labeled data well. The study examined how effective zero-shot, one-shot, and few-shot learning could be at identifying hate speech in Hinglish comments on YouTube. This relates to the issues faced in Banglish too. It showed that transfer learning can really help the models figure out comments without needing huge amounts of labeled data. In fact, zero-shot learning had an accuracy rate of 54%. That's not bad at all when dealing with smaller datasets.

Islam et al., [1] explored the necessity of effective detection and mitigation of online hate speech. The autores focused on a hybrid language combining Bangla and English. They used 5000 Banglish comments of different categories to differentiate between hate speech and non hate speech which is a binary approach. They approached several models including Support Vector Machine (SVM), Logistic Regression, Random Forest, advanced deep learning architecture and Innovative hybrid models(CNN+BiLSTM). However, For feature extraction they applied word embedding, TF-IDF and Bag of Words. Through this advancement the model reached an accuracy of 90% , precision, recall and F1 score.

Zamir et al., [2] used BERT multilingual model to detect hate speech in Telugu code-mixed text with moderate performance, average F1 score of 0.6151. They highlight that the main difficulties are posed by code-switching itself and the dearth of data rather than language pairs. Since this works on Telugu their observations on multilingualism and code mixing can be extended for Banglish also.

Sharan et al., [3] explored spam detection using several machine learning approaches, methods to detect spam and identify hate speech in text, using algorithms like Naive Bayes, K Nearest Neighbor (KNN) , Logistic regression , Decision trees. Their methodology emphasized the use of Terms Frequency Inverse Document Frequency (TD_IDF) as a feature extraction technique and incorporated lemmatization during preprocessing to improve text normalization. The result shows a comparison of performance metrics for different classifiers applied to datasets of hate speech and spam messages. Their model achieved a remarkable accuracy of 96% in spam detection, omit earlier models that relied on Simpler Vectorization Techniques. This outcome highlighted the importance of robust feature extraction preprocessing steps in enhancing performance of machine learning models for text classification.

Romim et al., [4] explored abusive comment detection in Bengali, employing six algorithms, with SVM achieving the highest accuracy of 85.7%. Their study emphasized the scarcity of Bengali-specific datasets and the complexities of contextual analysis, making it directly relevant to Banglish hate speech detection.

Chan et al., [5] examined the challenges of detecting LGBTQIA+ hate speech across multiple languages. Their findings reveal the limitations of machine translation in preserving cultural nuances, emphasizing the need for language-specific approaches.

Hashmi et al., [6] focused on hate speech detection in Norwegian, leveraging FAST-RNN and multilingual transformers. Their work highlights the difficulties of low-resource languages, which are relevant to the Banglish linguistic context.

Dhanya et al., [7] proposed a hybrid neural network for Malayalam-English code-mixed hate speech detection, achieving 96.4% accuracy. Their innovative approach combining GloVe embeddings and domain-specific preprocessing offers insights for Banglish datasets.

Sharma et al., [8] developed the THAR dataset for detecting religion-targeted hate speech in Hindi-English code-mixed text. Using transformer-based models like MuRIL, they demonstrated the importance of high-quality datasets and multi-class classification for hate speech detection.

Ghosal et al., [9] introduced an emoji-powered hate speech detection model for Bengali,

utilizing extended fuzzy SVM and mBERT text embeddings. Their study improved detection accuracy by 9.11%, offering potential enhancements for Banglish models.

Yadav et al., [10] used weakly annotated data to detect Hinglish hate speech through transfer learning. Their study on few-shot and zero-shot learning can be beneficial for resource starved Banglish datasets.

Yadav et al., [11] investigated explainable machine learning techniques for code-mixed hate speech detection and achieved a maximum accuracy of 67.61% with Bernoulli Naive Bayes. Their emphasis on interpretability aligns with the need for transparency in Banglish detection systems.

Romim et al., [12] presented a large dataset for Bengali hate speech detection, achieving 87.5% accuracy with SVM. Their annotated dataset and baseline evaluation serve as foundational resources for Banglish detection.

Chava et al., [13] explored Telugu-English hate speech detection using BERT and HateBERT models. Their transliteration techniques and findings underscore the importance of preprocessing in multilingual scenarios.

Kabir et al., [14] introduced G-BERT, which obtains an accuracy of 95.56% for Bengali hate speech detection through BERT and GRU integration, that is also a versatile architecture for Banglish.

Naznin et al., [15] used sentiment analysis along with hate speech detection to handle language ambiguity and provide a hierarchical framework for understanding the role of emotions in Banglish datasets.

Ndabula et al., [16] adopted SVM and Random Forest for Nigerian code-mixed languages with much emphasis on preprocessing techniques therefore this method will be useful for Banglish.

Kabir et al., [17] used XML-Roberta for multilingual hate speech detection and show that the findings of cross-lingual models are also applicable in the Banglish scenarios.

Saroar et al., [18] introduced BanglaHateBERT, a retrained BERT model, achieving 94.3%

accuracy for abusive language detection in Bengali. Their domain-specific model provides a benchmark for Banglish adaptations.

Das et al., [19] developed an attention-based recurrent neural network for Bengali hate speech detection, achieving 77% accuracy. Their use of LSTM, GRU, and attention mechanisms addresses challenges in classifying complex hate speech. They also emphasize on fine-grained labeling which is kind of similar as Banglish hate speech is very thin grained.

Davidson et al. [20] tried to differentiate between hate speech and offensive language proposing context-aware models. They also emphasize on fine-grained labeling which is kind of similar to how Banglish hate speech is very thin-grained.

Sai et al., [21] focused on hate speech detection in Telugu-English language which is a mixed language of Telugu and English whose main motivation was to solve the challenges posted by multilingual social media. They collected a dataset of 4500 where 4000 were used for training data set and 500 for testing. After transliteration and translation as important preprocessing steps it implement several model like BERT, XLM-RoBERTa and HateBERT. Among them BERT shows the best performance with accuracy of 69% and 0.6565 F1 score. This approach suggested the need of more context -aware models in Banglish Language hate speech detection.

Mim et al., [22] explored hate speech detection on social media using BERT models. They emphasize on low resource languages like Bangla. Nevertheless, they use multilingual models like XML-Robota which perform better than monolingual models. It suggests cross-linguistic approaches could be more helpful for Banglish hate speech. The paper highlights the rising importance of transformer models in handling multilingual humiliating content identification and removal.

2.3 Gap Analysis

Here is the gap analysis table given below:

Table 2.3 gap analysis

Aspect	Related Work	Gap Identified	Actions to Address the Gap
Language of Focus	Most studies have focused on Bangla, English, Hindi, Tamil.	Limited studies in code mixed languages particularly in Banglish language.	Take dataset and apply different models to get a good performance
Linguistic Complexity	One related work with Banglish language address basic code-mixing challenges	less effective data process methods.	Implement advanced preprocessing techniques like tokenization, lemmatization etc.
Model Architecture	Studies use traditional models like Naive Bayes, random forest and one hybrid model CNN+BLSTM.	used only three models.	apply more models.
Classification Scope	Has two classifications.	For this topic, there are just two classifications.	have seven classifications
Performance Evaluation	Accuracy is 90% for binary classification	No work till now	will propose a model with good accuracy .

Table 2.3 includes the analyzed gap of our research topic in past years.

2.4 Summary

There are very few papers about hate speech detection in multi class approaches in any language till now. In these papers, they have worked with different models like: multinomial Naive Bayes, RNN, Random Forest, CNN-LSTM, mBERT, BERT, Bi-LSTM and Bi-GRU etc. But there are a lot of papers about identifying hate speech in different languages by using a binary approach. In these papers, they have used different traditional models and also hybrid models to get their desirable accuracy. After doing all the background research and analyzing all the information, we can conclude that there is no existing research paper on our topic in the multiclass approach.

Chapter 3

Research Methodology

This Chapter is the detailed explanation of the methodology of our research

3.1 Methodology

To address the difficulties presented by this code-mixed language, a systematic approach is used in the methodology for detecting hate speech in Banglish. The first step is data collecting which involves gathering Banglish text from kaggle to guarantee a varied representation of hate speech categories like slang, appearance, racial, religion, sexual, non hate and others. Words like "pic" are changed to "picture" for uniformity using preprocessing techniques like tokenization and lemmatization which are used to handle spelling variants, slang, and mixed grammar. In order to rectify class disparities and guarantee equitable coverage of all forms of hate speech, the dataset is then balanced using strategies. For classification, six deep learning models are used, utilizing embeddings to capture the softness of Banglish text's context. One hundred epochs are used to train the models. The dataset consists of 5,000 Banglish comments used for hate speech detection. Out of these, 2,164 comments are classified as non-hate speech, while 2,836 are identified as hate speech. The hate speech comments are divided into several categories, including Sexual (928), Slang (575), Racial (135), Religious (198), Appearance (410), and Other (590) [23].

3.1.1 Overview

The research focuses on proposing a suitable model for detecting hate speech in banglish languages in seven categories. We have collected our data from Kaggle and preprocessed it by following some steps to increase the accuracy. After preprocessing the dataset, we have applied different models, such as traditional and hybrid models. For applying the models, we have run at least 100 epochs to get our desired accuracy and performance.

3.1.2 Proposed Methodology

Our proposed methodology for identifying hate speech in Banglish (Bangla language written in English alphabet) is a structured flow, which starts from raw text input as a sentence to the final classification output. This approach is a hybrid Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) architecture for robust hate speech detection.

The process starts with the input Banglish text, which includes social media comments or posts. These texts often mix Bengali and English, containing informal language and different forms. To get the data ready for model training, it goes through preprocessing. This step has several parts, such as Noise Removal, which gets rid of unnecessary characters, symbols, and emojis; Stopword Removal, which takes out common but unhelpful words; and Tokenization, where the text is cut into smaller pieces, like words or subwords. This preprocessing makes sure the data is clean and organized for analysis. The tokenized text is then changed into numerical forms by an Embedding Layer. This layer connects each token to a dense vector with a fixed size, capturing the meanings and links between words. For instance, words like "hate" and "anger" would have similar vectors, showing their related meanings. These vectors are sent into the first key part of the hybrid structure, the Bidirectional Long Short-Term Memory (BiLSTM) Layer. The BiLSTM handles the sequences in both forward and backward ways, capturing long-term links and context within the text. This bidirectional approach is important for grasping the nuanced meaning of Banglish phrases, where word meaning often relies on surrounding words. The result from the BiLSTM layer is improved by the Bidirectional Gated Recurrent Unit (BiGRU) layer, which boosts the model's ability to capture short-term links and patterns in both directions. The features that have been processed go to a dense layer that uses Softmax activation. This layer gives a probability spread over set hate speech categories: appearance, not hate, others, racial, religious, sexual, and slang. The category with the highest chance is chosen as the final result. This result shows if the input text has hate speech and puts it in the right category. The model that mixes BiLSTM and BiGRU, shown in the flow chart, uses the best parts of both types to fully understand the text data. This method allows for the correct classification of hate speech in Banglish, tackling the language issues of this mixed language with good trust and steadiness. The model is fitted to a balanced dataset for 100 epochs with 64 batch sizes, while 20% of the training set was held for validation to avoid overfitting the other 80% and to enhance the generalization of the model. There is no information where the model was trained.

This model is tested against a different test set, and the result metrics were used: loss, accuracy, a full classification report consisting of precision, recall, and f1-score, and a confusion matrix, which was used for class-wise observations. The BiGRU-based method makes sure the model stays focused on the main topic while managing informal, poorly written, or non-standard text typical of Banglish. Using bidirectional processing helps the model see relationships between words, allowing it to catch both the overall context and nearby connections in the text. This ability makes it a capable solution for finding complex kinds of hate speech in Banglish. Additionally, combining BiGRU with BiLSTM improves the model's skill in handling various datasets, allowing it to work well for analyzing multilingual and code-mixed texts. Here is the design of the methodology is given below:

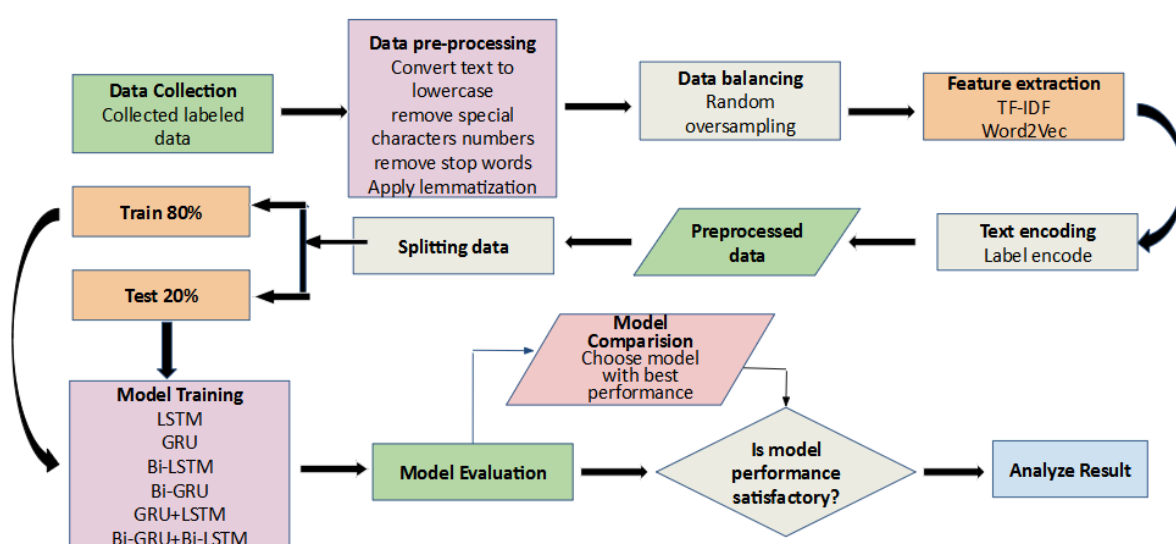


Figure 3.1: Proposed Methodology diagram

In fig. 3.1, the flow diagram of the proposed hate speech detection methodology in the Banglish language depicts a systematic flow from the raw text to classification output. The process Changes through a required step: Input Banglish Text, containing code-mixed, informal and multilingual Comments or Posts from social media. The first stage is Preprocessing, in which the text is cleaned and prepared through steps like noise removal, stopword removal, and tokenization. For example, noise removal removes unnecessary characters, symbols, and emojis, while stopword removal removes frequent words that do not add useful context. After, The text is divided into smaller units for further analysis and this process is known as tokenization, which yields structured and tokenized data. Word embeddings for the tokenized data is performed by the Embedding Layer, which derives dense numerical vectors that encapsulate semantic relations among the words. Semantically related

words, such as “hate” or “dislike” will share closer vectors in your embedding space. These embedding vectors are given to the Bidirectional Long Short-Term Memory (BiLSTM) Layer, which goes through the sequences in both forward and backward directions. This two-way approach enables the model to learn long-range expectancies and contextual connections within the text. In the aforementioned example of Banglish, the meaning of any word heavily relies on the words preceding it and succeeding it, which can be accomplished by a BiLSTM. The last BiGRU layer captures the most recent features and refines information by learning short-range dependencies in the output of the previous BiLSTM layer in both input and output directions. BiLSTM and BiGRU layers comprehensively record the semantic information of the text, which enables the model to identify complex features of hate speech. The transformed data is fed into a dense layer followed by Softmax Activation, ending up with a probability output over the seven labels: Appearance, Not Hate, Others, Racial, Religious, Sexual, Slang. The model finally outputs the class with the highest probability as classification results.

The final output is the hate speech category, which helps to detect the type of hate speech in Banglish text accurately. BiLSTM is effective in recognizing long-range dependencies, and BiGRU focuses on extracting the contextual information needed for our dataset; thus, combining both techniques provides a robust architectural combination. The data preprocessing step is essential for such a systematic pipeline to handle the linguistic complexities of Banglish and the proposed simple frameworks can classify the multilingual text data effectively.

3.2 Detailed Methodology

Our datasheet contains hate speech and classes. We choose our data set with relevant features. We titled our data set Banglish Hate Speech . We have already started to collect data from different social media including Facebook, instagram,youtube,Tiktok etc.We have collected almost 1000 data but those are not enough to train the model. However, We have used a dataset containing 5000 data on our relevant topic from Kaggle. We have taken a number of procedures to preprocess our data. First, we changed the uppercase to lowercase in our data. First, we took this action to guarantee data uniformity. This will assist us in avoiding errors that may occur while using the same word in several contexts. For instance, the words "This" and "this" have the same meaning. However, it will count as two separate words if we don't change it from uppercase to lowercase. Then we remove special characters from all the comments. Special characters are “ !@#\$\$%^&()_+==;',,: ”. This step makes sure that unwanted characters will not interfere in this process. After that, we have removed the stopwords from

our data. Stopwords are those words which carry little meaning or no meaning in a sentence. Stop words are “ a, the , at, for, with,from,on, to in “ etc. We are removing it to reduce the noise and focus on meaningful words. Then we have used lemmatization in our dataset to make it more accurate to get better accuracy. For example, we have replaced the word “amr” by the word “amar”. In this way, we have replaced almost 10,000+ words with the more accurate word. As we are working with the Banglish (Bangla language written in English language), we have used a special library which is “NLTK library.” Here, firstly, we have defined a list of stopwords for removing those words from the data. We have split the text into words then we have filtered out those words which are present in the stopwords list. These preprocessing steps are applied throughout all the comments one by one. Then the clean data was stored in a column for further analysis. TF-IDF(Term Frequency-Inverse Document Frequency) was applied to this preprocessed data for text vectorization. We have used this dataset to make it suitable for deep learning models. After preprocessing the dataset, the categorization process makes use of six deep learning models. The model's efficacy and accuracy in multiclass hate speech detection are assessed after it has been trained over 100 epochs.

3.3 Project Plan

To guarantee careful development and assessment, the project strategy for hate speech identification in Banglish takes a methodical approach. Data gathering is the main focus of the first phase, which leverages a Kaggle dataset to ensure a varied representation of various hate speech categories such as racial, religious, and slang-based material. After that, this data will go through preprocessing, which includes lemmatization and tokenization, to handle informal language and standardize the text. To address class imbalances and guarantee that every category is fairly represented, the dataset will be balanced using random oversampling in the following stage. The preprocessed data will be divided into training and testing sets, and classification will be done using six deep learning models. To capture the subtleties, the models will be trained by 100 epochs.

3.4 Task Allocation

The two team members split up the work for the project on detecting hate speech in Banglish. The first steps, which include obtaining and preparing the dataset from Kaggle, are the responsibility of Member 1. To handle slang and informal language, this involves tokenizing and lemmatizing the dataset to make sure the text is clear and consistent for model training. Additionally, Member 1 will train the six deep learning models over 100 epochs. In addition to

documenting the experimental setup and technique and helping to write the final report sections pertaining to experimental results and discussions, we will optimize the model's performance by modifying hyperparameters. However, Member 2's job is to use Random Oversampling to balance the dataset so that all hate speech categories are equally represented. Metrics like accuracy and loss will be used to assess the model, and visualizations like confusion matrices will be used to examine the findings. In addition, Member 2 will compose the report's background and methodology parts and perform a literature evaluation of relevant works in hate speech identification. The two team members will work together to complete the report and resolve any issues that arise during the implementation stage.

3.5 Summary

For getting better accuracy and performance, we have applied six deep learning models. We have worked with 5000 data from the kaggle dataset. For getting better performance, data preprocessing is a must. So, we preprocessed our dataset by using different preprocessing methods. From these models, we will find out the proposed model which will give the output of seven categories which are slang, sexual, non-hate, appearance, racial, others and religious. We have also calculated the precision, recall, F1-score and support to evaluate the performance of the models.

Chapter 4

Implementation and Results

This chapter will provide the accuracy, performance and final results of the models and also prove why our proposed model is better than other models.

4.1 Environment Setup

To manage the intricate procedures of data preprocessing, model training and evaluation, the research environment setup needed a strong computational infrastructure. Specifically, the Bidirectional GRU-based architecture was used to speed up the training of deep learning models on a machine that has a GPU. An NVIDIA GPU with CUDA capabilities was part of this configuration, allowing for the effective handling of big datasets and computationally demanding workloads. Python was chosen as the main programming language in the software environment because of its many machine learning and natural language processing packages. Important libraries included scikit-learn for preprocessing and evaluation, pandas and numpy for data manipulation and analysis and TensorFlow and Keras for constructing and training the deep learning model. In addition, Seaborn and Matplotlib were employed to visualize model performance and data distributions. To control dependencies and guarantee compatibility across various programs, a virtual environment was developed. Google Collab was used to run the program, Jupyter Notebook which provides an interactive platform for iterative experimentation and debugging, was used to process and analyze the dataset. Cloud storage was added to the entire arrangement in order to handle big datasets and backup learned models. This all-inclusive environment guaranteed efficient code execution, efficient resource use and the capacity to carry out thorough experimentation.

4.2 Performance Analysis

To analyze the performance of our model, we calculated various matrices such as accuracy, precision, recall and F1-score. For testing our models and increasing our performances, we have applied six models and also calculated the matrices. Here is a brief description of them

given below:

Accuracy: A typical evaluation metric in deep learning and machine learning is accuracy, which quantifies the frequency with which a model correctly predicts the target labels relative to the total number of predictions. The following formula is used to compute it:

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

Precision: A metric for evaluating a model's accuracy in making positive predictions is called precision. Stated differently, it indicates the percentage of real positive examples among all occurrences that were anticipated to be positive. The following formula is used to compute it:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall: The evaluation parameter known as recall, which is sometimes referred to as sensitivity or true positive rate, assesses a model's capacity to accurately identify each and every pertinent positive case in a dataset. Stated differently, it indicates the proportion of real positive examples that the model accurately detected. The following formula is used to compute it:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F1-score: An indicator that balances precision and recall by combining the two into a single value is the F1 score. The F1 score, which is computed as follows:

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

4.2.1 Model Description

Now we will give a brief explanation about our applied six models which are LSTM, GRU, Bi-LSTM, Bi-GRU, GRU + LSTM and Bi-LSTM + Bi-GRU. The explanation is given below:

LSTM (Long Short-Term Memory) : A particular kind of recurrent neural network called LSTM is made to handle and forecast sequential input. By employing memory cells to capture

long-term relationships, it gets around the vanishing gradient issue that plagues conventional RNNs. In tasks where comprehending context across lengthy periods is crucial, such as time series forecasting, natural language processing, and speech recognition, LSTM performs exceptionally well. Here figure 4.2.1.1 represents the LSTM model architecture.

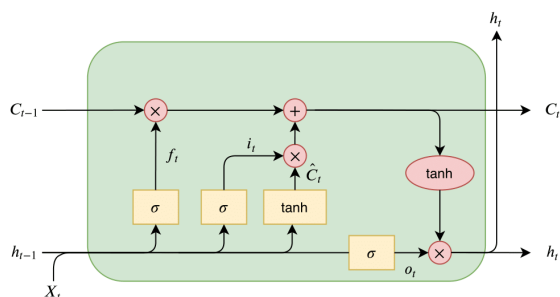


figure 4.2.1.1: LSTM model architecture

GRU(Gated Recurrent Unit): One kind of recurrent neural network that handles sequential data is called a GRU (Gated Recurrent Unit). By merging the input and forget gates into a single update gate, it streamlines LSTM and preserves performance. GRU is effective at identifying relationships in time series and sequence applications, such as speech recognition, machine translation, and natural language processing. Figure 4.2.1.2 represents the architecture of the GRU model.

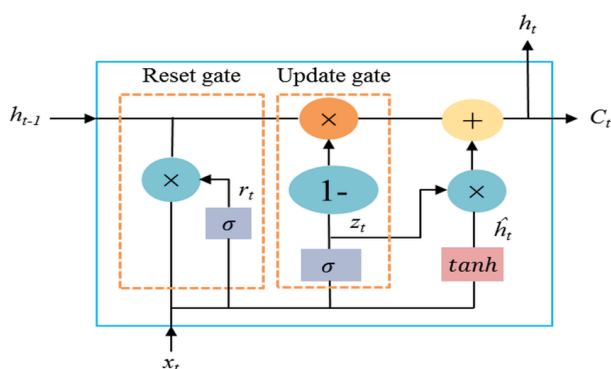


Figure 4.2.1.2: GRU model architecture

Bi-GRU (Bidirectional Gated Recurrent Unit): The GRU model is extended by the Bi-GRU which can process data both forward and backward. It is more effective for activities where comprehension of both previous and succeeding material is essential because it captures context from both past and future sequences. Speech recognition, natural language processing, and sequence prediction tasks all make extensive use of Bi-GRU. Here figure 4.2.1.3 represents the working process of the Bi-GRU model.

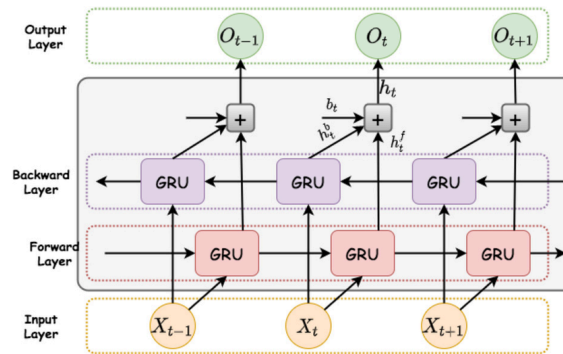


Figure 4.2.1.3: Bi-GRU model architecture

Bi-LSTM (Bidirectional Long Short-Term Memory): The LSTM model is extended by Bi-LSTM which can process data both forward and backward. It improves performance in activities where comprehending context from both ends is crucial by catching information from both past and future sequences. Bi-LSTM is frequently utilized for time series prediction, speech recognition, and natural language processing tasks. Figure 4.2.1.4 shows the working process of the model.

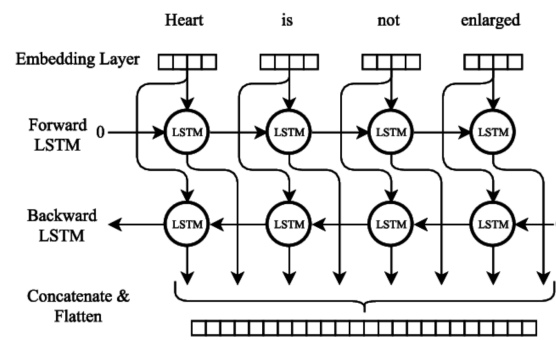


Figure 4.2.1.4: Bi-LSTM model architecture

GRU + LSTM: The advantages of both Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) networks are combined in the hybrid model known as GRU + LSTM. It makes use of both the strong long-term memory capacities of LSTM and the ease of use and effectiveness of GRU. For applications like time series prediction, natural language processing, and speech recognition, this combination enhances performance in identifying both short-term and long-term dependencies in sequential data. Here, figure 4.2.1.5 shows the GRU+LSTM model which is a hybrid model.

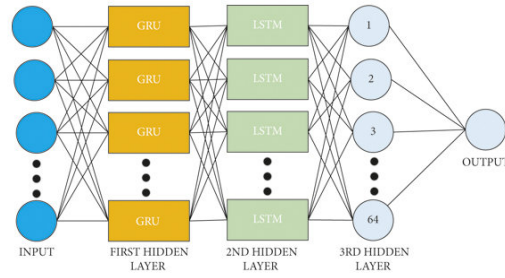


Figure 4.2.1.5: GRU+LSTM model

Bi-LSTM+Bi-GRU: The bidirectional characteristics of both Long Short-Term Memory (Bi-LSTM) and Gated Recurrent Unit (Bi-GRU) networks are combined in the hybrid model known as Bi-LSTM + Bi-GRU. This paradigm captures dependencies from both past and future contexts by processing sequential data both forward and backward. For tasks like natural language processing, machine translation, speech recognition, and time series forecasting, LSTM's capacity to retain long-term dependencies and GRU's effective gating mechanism work together to improve performance. Figure 4.2.1.6 represents the Bi-LSTM+Bi-GRU model here.

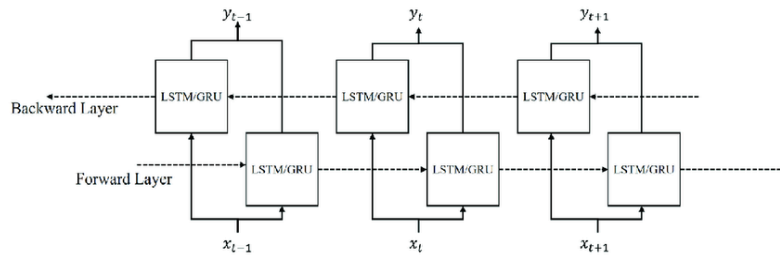


Figure 4.2.1.6: Bi-LSTM+Bi-GRU model

Now we will discuss the performance of matrices for our models. The given table 4.2.1 is the comparison between our applied six models:

Table 4.2.1 Comparative Analysis of Models

Model Applied	class name	Precision	Recall	F1-score	Support
LSTM	Class 0	0.99	0.38	0.55	403
	Class 1	0.88	0.09	0.16	390

	Class 2	1.00	0.06	0.12	403
	Class 3	0.97	0.87	0.92	409
	Class 4	0.96	0.92	0.94	400
	Class 5	0.96	0.12	0.21	403
	Class 6	0.87	0.66	0.75	392
GRU	Class 0	0.91	0.68	0.78	403
	Class 1	0.76	0.31	0.44	390
	Class 2	0.93	0.27	0.42	403
	Class 3	0.95	0.94	0.95	409
	Class 4	0.97	0.99	0.98	400
	Class 5	0.92	0.18	0.30	403
	Class 6	0.83	0.74	0.79	392
Bi-LSTM	Class 0	0.93	0.41	0.57	403
	Class 1	0.83	0.09	0.16	390
	Class 2	1.00	0.05	0.09	403
	Class 3	0.98	0.88	0.93	409
	Class 4	0.97	0.89	0.93	400
	Class 5	1.00	0.01	0.02	403
	Class 6	0.87	0.50	0.63	392
Bi-GRU	Class 0	0.88	0.54	0.67	403
	Class 1	0.74	0.30	0.43	390
	Class 2	1.00	0.02	0.05	403

	Class 3	0.97	0.92	0.95	409
	Class 4	0.94	0.98	0.96	400
	Class 5	0.91	0.31	0.46	403
	Class 6	0.85	0.78	0.81	392
GRU + LSTM	Class 0	0.84	0.75	0.80	403
	Class 1	0.62	0.32	0.42	390
	Class 2	0.80	0.67	0.73	403
	Class 3	0.96	0.88	0.92	409
	Class 4	0.97	0.96	0.97	400
	Class 5	0.90	0.65	0.76	403
	Class 6	0.79	0.83	0.81	392
(proposed model) Bi-LSTM + Bi-GRU	Class 0	0.85	0.91	0.88	403
	Class 1	0.65	0.57	0.61	390
	Class 2	0.82	0.76	0.79	403
	Class 3	0.96	1.00	0.98	409
	Class 4	0.94	0.98	0.96	400
	Class 5	0.76	0.73	0.74	403
	Class 6	0.90	0.90	0.90	392

Here table 4.2.1 includes all the matrices performances for six models.

Our one related paper also applied different machine learning models like SVM, Random forest, LSTM, GRU, logistic regression, CNN+BLSTM etc. Among all the models, their best model or proposed model was “CNN+BLSTM” which had an accuracy of 90%. As per our

research, we have applied six models and our best performing model is “Bi-LSTM+Bi-GRU” which has an accuracy of 83%. So we can say that as we have applied deep learning models for our dataset and will have the output in seven classes, our work is all new. Additionally we can say that our research work is all unique and fresh.

4.3 Results and Discussion

This part is for discussing the result including the accuracy and loss for each model applied to get the detailed idea about each model. Here all models accuracy and loss graph are provided down below:

LSTM:

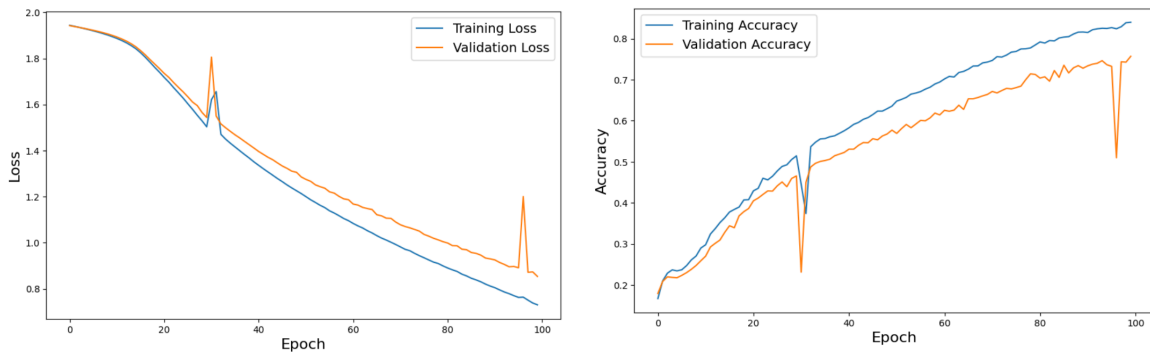


Figure 4.3.1: Accuracy and loss graph of LSTM

GRU:

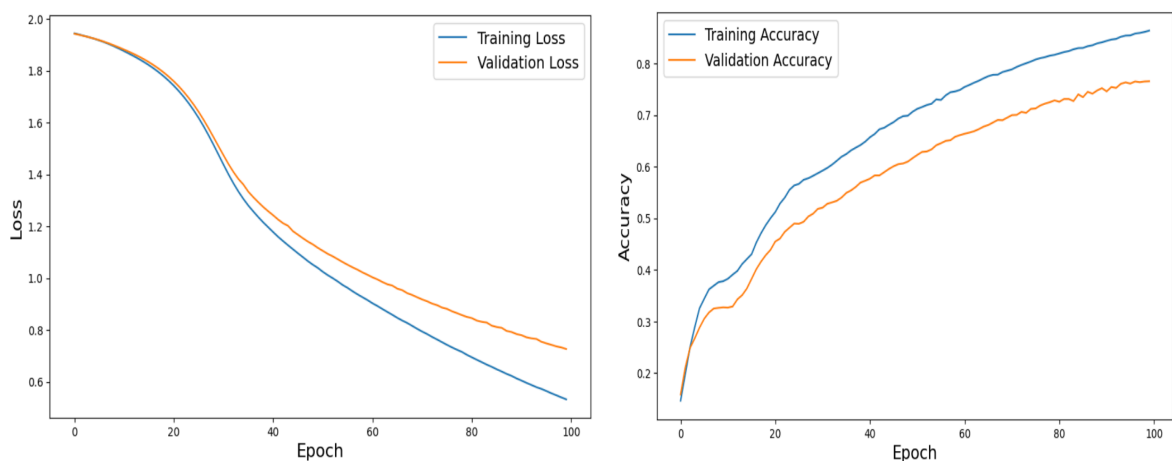


Figure 4.3.2: Accuracy and loss graph of GRU

Bi-LSTM:



Figure 4.3.3: Accuracy and loss graph of Bi-LSTM

Bi-GRU:

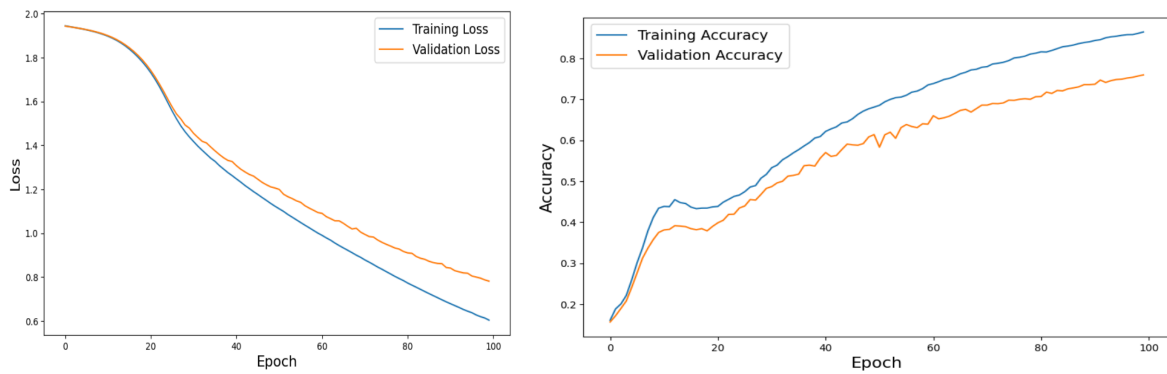


Figure 4.3.4: Accuracy and loss graph of Bi-GRU

GRU+LSTM:



Figure 4.3.5: Accuracy and loss graph of GRU+LSTM

Bi-LSTM+Bi-GRU:



Figure 4.3.6: Accuracy and loss graph of Bi-LSTM+Bi-GRU

In the above figures we can see the accuracy and loss graph of six different models which we have applied in our research. Figure 4.3.1 represents the graph of the LSTM model. Then figure 4.3.2, figure 4.3.3, figure 4.3.4, figure 4.3.5 and figure 4.3.6 represent the GRU, Bi-LSTM, Bi-GRU, GRU+LSTM and Bi-LSTM+Bi-GRU models graph of the accuracy and loss respectfully.

Now the given table is the final results of all the models we have applied in our work.

Table 4.3 Final results of Models

Model	Loss	Accuracy
LSTM	0.8365	76.75
GRU	0.7179	77.57
Bi-LSTM	0.8994	76.21
Bi-GRU	0.7770	76.43
GRU + LSTM	0.6723	78.68
Bi-LSTM + Bi-GRU	0.6578	83.93

Here table 4.3 represents the final accuracies and losses of our applied models.

At the end of applying all six models one by one, we can come to the decision that The highest accuracy is 83.93% in the “Bi-LSTM+Bi-GRU” model which is a hybrid model. And the lowest

loss is also in this model. After analyzing all six models' performances, accuracies and loss we can conclude that our best model is the “Bi-LSTM+Bi-GRU” model.

4.4 Summary

In this section we have discussed our model applying environment, performance and final result. After analyzing six models performances and outputs we can say that our proposed model for detecting hate speech in banglish language in seven categories is “Bi-LSTM+Bi-GRU” model.

Chapter 5

Engineering Standards and Design Challenges

This chapter bears the challenges and impacts of our work.

5.1 Compliance with the Standards

This study follows the following technical and communication standards including software standards, hardware standards, and communicational standards. Each standard compares with the alternative that fits best for this study.

5.1.1 Software Standards

In this research, we used Python 3.8.0 to write the whole code part, following PEP 8 (Python enhancement proposals) guidelines to keep the code clear and clean. In the coding part, we used libraries such as TensorFlow, scikit-learn, and Pandas to handle tasks like data augmentation, preprocessing, training, and evaluating the model. NLP methods are used for removing stop words and tokenizing text that follows common rules. Multiclass classification for hate speech types has been done using standard supervised learning techniques. In this research initially we used six deep learning models. The model's performance has been checked by precision, recall, accuracy and F1-score to make it work well. The dataset has been tested and balanced using a method called RandomOverSampler from imblearn. so that all classes are represented bias-free. Comments are written to make the code more readable and reuse it. All the codes are tacked together with github for collaboration and working in teams. It controls the version.

5.1.2 Hardware Standards

This research work was conducted by a personal computer, whose hardware specifications are:

- Processor: Intel(R) Core(TM) i5-10400 CPU @ 2.90 GHz
- RAM: 16.0 GB
- System Type: 64-bit Operating System, x64-based Processor
- Operating System: Windows 10 Pro, Version 22H2
- OS Build: 19045.5247
- Experience Pack: Windows Feature Experience Pack 1000.19060.1000.0

5.1.3 Communication Standards

In this research, the following communication standards are used to ensure team collaboration, tracking the version of code, consistency and clarity of all the tasks throughout the study:

- All the code as well as every task that has been done throughout this study are well documented. Each line is explained with comments for making the code more readable and for future use. Variable names are named to be readable.
- All the resources, such as files, datasets, and results, were shared among teammates using Google Drive as well as Google Docs to ensure accessibility to seamless teamwork.
- Communication was maintained with the supervisor, co-supervisor, and teammate through email and telegram. Whenever any important updates need to be shared, we do it through Google Meet.
- Academic communication was maintained by following referencing guidelines for any external resources and datasets. IEEE citation style was followed all over the report.
- Research findings were represented effectively through clear visualization, tables, and graphs.

5.2 Impact on Society, Environment and Sustainability

In context of identifying hate speech in banglish language using deep learning model like bidirectional LSTM, GRU, Bi-GRU, Bi-LSTM, GRU+LSTM and Bi-GRU+Bi-LSTM. This research contains significant importance on society, environment and sustainability.

5.2.1 Impact on Life

The research to detect hate speech in the Banglish language has tremendous societal impact, especially in relation to creating a safer online communicative environment. Banglish is extensively used in Bangladesh and its diaspora and can commonly be found on popular social media platforms like Facebook, Instagram and YouTube. However, these platforms have also turned into the hotbed for spreading hate speech, cyberbullying and inappropriate content which creates an unhealthy environment that can harm individuals as well as communities. To the best of our knowledge, this is the first paper that intends to work on a model to detect and categorize hate speech in Banglish i.e., promoting inclusiveness and respect. This research has profound ethical implications as it tries to manage recognizing and reducing hate from social media platforms in Banglish. Building machine learning and deep learning models towards identifying hateful comments raises several ethical issues especially related to privacy, bias, abuse prone etc. The societal implications will resonate with protecting potentially targeted vulnerable groups such as women, minority people or people from different religious or ethnic backgrounds whose target-specific hate speeches can be tracked more holistically because of our proposed model's ability to generalise across racial/religious/sexual/appearance-related slanders thereby reducing victims' emotional stress and psychological trauma. Also at broader scope, as this model could serve as a plugin or extension for social media platforms it will help real time monitoring of malicious posts spreading through their network which can discourage the misuse of freedom of expression against hate speech users unnecessarily but positively encourage good practices among these users. In a broader sense further digital literacy will be fostered to gain knowledge about responsible use of cyberspace. Moreover, law-making agencies can also employ patterns identified from this research to tackle increased rates of cyber harassment suitably. Finally our contribution towards NLP (natural language processing) will concentrate on augmenting linguistic peculiarities associated with Banglish. thus setting up a stage using low resource language this curiosity-driven advancement reflects that future works relating regional/hybrid languages. The research aimed at identifying hate speech in the Banglish language has considerable societal significance, especially regarding

the creation of safer online communication environments. Banglish, a popular mode of expression in Bangladesh and its diaspora, is widely found on social media platforms, especially on Facebook, Instagram, and YouTube. But these platforms frequently serve as vehicles for the dissemination of hate speech, cyberbullying and inappropriate content, resulting in a harmful atmosphere that can adversely affect individuals and communities. By creating a model specifically designed to detect and categorize hate speech in Banglish, this research directly confronts this issue, helping to cultivate a more inclusive and respectful online space. The societal branching of this research encompasses the protection of vulnerable groups frequently targeted by hate speech, including women, minorities, and those from diverse religious or ethnic backgrounds. The capacity to recognize hate speech across various categorizations such as racial, religious, sexual, and appearance-based scandals promises comprehensive tracking, thus mitigating the emotional and psychological distress experienced by victims. Additionally, the model's potential to function as a plugin or extension for social media platforms facilitates real-time intervention against the spread of harmful content, deterring hate speech offenders and encouraging responsible online conduct. On a larger scale, this research cherishes digital literacy and social awareness by underscoring the significance of respectful communication. It also aids policymakers and law enforcement in recognizing patterns of hate speech, which supports the development of more effective strategies and measures to stop online harassment. Furthermore, this study advances the field of natural language processing (NLP) by tackling the specific linguistic complexities associated with Banglish, setting the foundation for future advancements in language technology that cater to regional and hybrid languages.

5.2.2 Impact on Society & Environment

Although the immediate environmental effects of this research may seem minor, the suggested model plays an important role in a more sustainable digital ecosystem. The wave in online communication and the dramatic rise in digital content production have resulted in substantial energy usage, particularly in the hosting, monitoring, and management of extensive user-generated material. By automating hate speech detection in Banglish with an efficient machine learning model, this research lowers the reliance on human moderators who often require significant resources and tend to be inefficient. Automated moderation can reduce the need of high computing resources and improve the content management process which in turn may cause reduced energy consumption compared to manual moderation systems. Such efficient mechanisms can help minimize the carbon footprint associated with social media and

digital services. Furthermore, by encouraging positive online behavior and curbing the dissemination of hate speech, this research contributes to a healthier online community, lessening the digital clutter from harmful content that could otherwise necessitate extra resources for investigation and resolution. Additionally, the social multiplier effect of reducing hate-speech may also help environmental awareness initiatives. Negativity-free online spaces act as more powerful mediums for sustainability promotion, climate advocacy and environmental education. This research indirectly assists the causes against pressing issues of the environment by making sure these spaces remain positive and inclusive. In conclusion, the environmental impact arises from the efficient utilization of digital resources, decreased energy needs for content moderation, and the promotion of online environments that more effectively support environmental awareness efforts. While the direct benefits for the environment may be subtle, the indirect gains highlight the capacity of technology to enhance sustainability.

5.2.3 Ethical Aspects

Users should be aware of how their data is used and provide consent. Privacy should not be violated, users' personal data should not be inadvertently exposed due to dataset leaks or model outputs, and user data should be anonymized to prevent it from being abused by third parties.

- **Privacy and Data Protection:** The model depends on extensive datasets of user-generated content which may contain sensitive or private information. It is essential to ensure that the data utilized in this research is anonymized and complies with strict privacy regulations. Users should be aware of how their data is used and provide consent. Privacy should not be violated. User's personal data should not be inadvertently exposed due to dataset leaks or model outputs and user data should be anonymized to prevent it from being abused by third parties.
- **Bias and Fairness:** There is a major ethical concern on the possible embedding of biases in the models, as well as perpetuating them. Since the model is trained using social media data, it might inherit societal biases, including those related to gender, race or culture. Such biases could lead to the unfair treatment of specific groups by either overrepresented or underrepresented certain forms of hate speech. To address this, we need to ensure that we have a balanced and representative dataset and use bias detection mechanisms as part of the model testing. It is important to have ongoing

model performance tests to make sure there isn't any bias with regard to a certain group, where content disproportionately gets flagged from already marginalized groups.

- **Transparency and Accountability:** We should be clear how the models are being used if they are driving decisions about removal of content or you know, moderation in our online services. Social media platforms and online services need to be transparent about how automated systems like this are employed in content moderation. Clear guidelines should be provided to users, defining what qualifies as hate speech and how determined decisions are made. Furthermore, the employment of machine learning in content moderation raises concerns about accountability. If the model erroneously categorizes content, users may encounter repercussions such as post removals or account suspensions. Thus, it is important to give users the ability to contest a decision made automatically by an algorithm and correct any mistakes caused by it as well.
- **Misuse Prevention:** Also, potential misuse of the model can be considered as ethical issues. Although our model helps in preventing harmful content, malicious people can use similar technology for the purpose of Censorship or silencing dissenting opinions which may not qualify as "hate speech". So, we have to make sure that this technology isn't used for suppressing free speech or censoring valid political, social or cultural discussions. Clearly defining ethics and establishing oversight mechanisms can help in preventing this kind of misuse of the model.
- **Responsibility and Regulation:** As this model would be deployed on platforms reaching billions of people, it is important to define clear lines of responsibility. This ethical responsibility falls not only on the developers creating and training the model but also on the organizations that choose to deploy it. Public accountability and transparency by means of regulation will guarantee that AI-driven moderation systems are not misused and are used for the benefit of a greater part of humanity. Committing post-hoc, through an invoice document or otherwise, to those principles during the entire lifecycle of creation deployment not only helps as a signal for increasing trust with external entities, but drives internal priorities.

5.2.4 Sustainability Plan

The sustainability of the project, both in terms of technical maintenance and societal impact is important to ensure its long term success. A sustainable model involves ensuring that the solution does not stop functioning or improving over time and it does not negatively affect the community or environment. The sustainability strategy for this initiative will emphasize several critical areas: ongoing technical maintenance, scalability, ethical considerations and continuous enhancement driven by feedback and new technologies. To maintain the relevance and effectiveness of the hate speech detection model, it will be necessary to implement regular updates and improvements. With time, new forms of hate speech may arise which may require the model to get adapted again. We will perform periodic retraining using updated broad and inclusive data with newer samples of Banglish hate speech. This requires a transparent process for data collection and labelling new instances along with a strong feedback loop to help the model learn from mistakes. Moreover, working closely with linguists and social scientists will aid in recognizing emerging linguistic patterns in hate speech. Implementing the model on large scale platforms such as social media will involve constant monitoring, maintenance and technical support. This entails ensuring that server structures can accommodate increased traffic and that the model performs optimally for real-time operations with low latency. Additionally, it should be able to scale and easily plug into other platforms than social media . The code, model and deployment process will be kept in a public version-controlled repository, along with all the documentation necessary for developers and contributors to use the model in different contexts, ethics and social impact. A sustainable model operates transparently, fairly, and inclusively. Periodic audits on bias, fairness, privacy should be conducted to ensure that unintended harms are not imposed on users or communities. Partnerships with external organizations as well as experts on online ethics can help the model stay up-to-date with social norms and regulations when these change over time. A sustainable model also does not infringe upon user's rights to free speech or disproportionately affect underrepresented groups. Besides, We will actively seek feedback from users, moderators and other stakeholders to make the system better and suited to real world problems. We will create feedback loops via user surveys, comments section on the hate-speech report form and looking at moderation data to get insights about the model performance as well as user experience. Engaging with the community regularly will help us surface those categories of posts which could be misclassified by the model so that we can improve it iteratively. Also it is expected that over time, models can go stale as content on the Internet evolves reflecting changing social norms. Frequent interactions will help us identify

any such situations too. Similar to other AI applications, environmental costs associated with maintaining large models need to be minimized when doing research in this domain. To train deep learning based models effectively and run high-performance servers, significant computational resources including energy are needed resulting in greenhouse gas emissions related to carbon footprint borne by them. Therefore, for efficient use of these limited computational resources while training or deployment we aim at making the lifecycle of models developed as part of this toolkit as much energy efficient as possible through methods like model pruning and quantization etc. achieving energy vs accuracy trade-offs without compromising much on performances would also be targeted. Additionally, cloud services which run on renewable power sources would be utilized wherever possible for hosting the same. For long term viability it is important that partnerships are formed with both academic institutions where research advancements around ML can flow into these deployments but also industry collaborators who can guide in terms real-world deployment issues becomes critical from perspective of short-term funding but more importantly over a longer term infrastructure upkeep issues apart from collaborations around tackling new challenges/features etc., An essential aspect of the model's long-term sustainability is enhancing public understanding of hate speech and the role of automated systems in content moderation. Educational initiatives can inform users about the functioning of the system, how to report content moderation issues, and how they can help foster safer online environments. The academic partnership should not stop only at partnering during infra development but also schools and universities to help them generate awareness about responsible online behavior as well as ethical usage of AI in context of hate content. The financial viability of all aspects above depends on being able to effect technology improvements in the ML/DL space, while maintaining ethics for public good. This should go hand-in-hand with continued support from partners.

5.3 Project Management and Financial Analysis

Table 5.3.1: Financial analysis

SL.	Field of cost	Cost
1.	Research Writing course	7000
2.	python course	6000
3.	related paper purchase	3960
4.	Internet bills	1500
5.	Utility	500
Total cost		18960

Here table 5.3.1 represents the financial details for doing our full work.

Table 5.3 outlines the budgets required for the study. Here, we used a personal computer, so there was no cost here. This study also required no paid software and used all the open source libraries. But, for this study, I have completed a course on writing thesis papers held by Research Help Bangladesh, which cost 7000 tk and a course on Python conducted by Phitron, which cost 6000 taka. These two courses are to sharpen my knowledge in research and analysis. It helps me a lot in this study. Besides, we need to purchase paid related research work for our background study. It cost 3960 tk. The overall cost is 18960 tk throughout the research work.

Now we will see our full time and work management through the given table 5.3.2 below:

Table 5.3.2: Project Management Table

Tasks	Weeks																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Data Preprocessing																		
Data augmentation and lemmatization																		
Applying different models																		
Writing the whole report																		

Estimated Work Period	
Actual Work Period	

Table 5.3.2 project management of our tasks of 17 weeks. We have completed our task according to our estimated time period.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Now we will discuss the mapping with complex problem solving, knowledge Profile and complex engineering activities.

Table 5.1: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analyses	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake- holder Involvement	EP7 Interdependence
✓		✓	✓			

This Table 5.1 maps to the complex problem solving with our research.

EP1: Dept of Knowledge

Rationale: For justifying the EP1, my research topic covers different field's complexity of a hybrid language. For applying different preprocessing methods like tokenization, lemmatization, vectorization, handling of stop words, handling of special characters and text cleaning, we had to go through the depth of the knowledge about these techniques. Not only that but also we have applied six models to get our desirable performance and accuracy. They are all deep learning models here. For applying the non-hybrid models like LSTM, GRU, Bi-LSTM and Bi-GRU, we had to learn about them very clearly and deeply. However, we have also had to gather knowledge from different sources and fields to do our work properly.

EP1: Dept of Analysis

Rationale: For justifying EP3, We had to go through different types of analysis. We had to know about nearly everything related to our topic and also analyzed the gap of other research. Here we have also analyzed each model and its effectiveness to any dataset. By applying six

models, we have analyzed one by one model's accuracy, loss, precision, recall, F1-score and support. we had to analyze the whole dataset to split it into training and testing parts. Not only that but also for increasing each model's accuracy we had to analyze 10000 plus data from the dataset.

EP1: Dept of Familiarity of Issues

Rationale: To justify EP4, which is the familiarity of issues, we can say that there are a lot of issues in our society. Among them, issues faced in social media platforms are huge. The “familiarity of issues” is immediately addressed by my topic which tackles a challenging and urgent natural language processing issue. Banglish, which is a combination of Bangla and English, is very relevant in today's world because it is frequently used in digital communication, particularly on social media platforms. Specialized techniques for detecting hate speech are necessary due to the particular difficulties presented by Banglish, including the frequent usage of slang, code-switching, misspellings, and non-standard syntax. This paper explores these difficulties and offers insightful information about how machine learning approaches might be used to solve them. By examining these intricacies, the study makes a substantial contribution to the expanding corpus of knowledge in social media monitoring and linguistic analysis, guaranteeing that it stays extremely relevant to contemporary concerns in technology and society.

Mapping with Knowledge Profile for EP1

This table 5.2 is designed to map the EP1 to the Knowledge Profile.

Table 5.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓		✓	✓	✓

K3(Engineering Fundamentals)

Engineering fundamentals like mathematics, physics, different statistical methods, machines and dynamics are playing a vital and main role in our research. For doing our research we had

to go through the fundamental information of these topics. Using basic data science and deep learning methods, my research combines key ideas from machine learning, natural language processing and text categorization. Based on mathematical, probabilistic, and computational theory concepts, we have used preprocessing techniques like lemmatization and tokenization along with statistical approaches and algorithms like GRU, BiGRU, and LSTM. It covers the K3 part.

K5(Engineering Design)

In our methodology part we had to draw the diagram of our workflow. In order to identify hate speech in Banglish, our research uses engineering design principles to address practical issues in the field of social media content moderation by utilizing deep learning models and optimizing them for performance measures such as accuracy, precision, recall, and F1-score. It covers the K5 part.

K6(Engineering Practice)

Our study sticks to engineering usually by applying theoretical ideas and techniques to a real-world issue. The application of technical expertise to solve societal issues is demonstrated by the usage of machine learning algorithms in practical situations like hate speech identification. Furthermore, my work uses industry-standard technologies such as Scikit-learn and TensorFlow to put these techniques into practice. And it covers the K6 part.

K8(Research Literature)

Our study adds to the body of knowledge in the fields of text classification, hate speech detection and multilingual NLP. We can find gaps in the existing literature on Banglish, where hybrid language processing is understudied, by examining earlier research. Thus, our study advances the field's knowledge of multilingual hate speech identification and offers fresh perspectives. For doing our research work, we had to do a lot of literature review by reading at least 50 plus papers. Among them we have only found one paper about our related topic which is closer to our work. That is why we can say that we have also covered K8 in this part.

5.4.2 Engineering Activities

Our research is involved with multiple complex engineering activities. These mappings are playing an important role to highlight the impacts and contribution of our research to solve

any immediate issues.

Table 5.3: Mapping with complex engineering activities.

EA1 Range of re- sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
	✓	✓	✓	

Here table 5.3 is the mapping of complex engineering activities.

EA2(Level of Interaction)

Rationale: For justifying EA2, our research will interact with human and machine learning systems. As we are working with Banglish language, Our research will interact with real world complexities and combining cultural idioms, slang and code-switching which increases its adaptability to a variety of user-generated material. Our research also combines the linguistic, computer science and social science fields together. It shows a high level of interdisciplinary interaction by analyzing the sociolinguistic components of hate speech, developing algorithms for processing hybrid languages, and assessing the effects on society. Our research will interact with social media platforms such as facebook, you-tube, instagram, twitter and others, possibly influencing the content moderation policies and user experiences. Then it will showcase the interaction with broader systems and stakeholders. In this way it covers the EA2 part.

EA3(Innovation)

Rationale: For justifying EA3, Our research is playing a very important part in innovation. As there is no paper about our particular topic, our research is the first step to explore and innovate new knowledge for other people in this topic. Our study tackles a linguistic area that is frequently ignored in natural language processing by concentrating on Banglish, a unique combination of Bangla and English. In contrast to previous methods, this invention involves developing specialized preprocessing procedures for handling code-switching, slang, and non-standard grammar. Binary classification (hate vs. not hate) is the main focus of many hate speech detection systems. Our multiclass method like Appearance, Not Hate, Others, Racial,

Religious, Sexual, and Slang gives hate speech detection more detail and provides light on certain hate speech categories. This degree of specificity can enhance content moderation and intervention tactics. Reliable performance and technical innovation are ensured by the employment of advanced deep learning models such as LSTM, Bi-LSTM, Bi-GRU and GRU and hybrid architectures such as GRU + LSTM and Bi-LSTM + Bi-GRU. By combining the advantages of various architectures, hybrid models enhance the system's capacity to manage intricate Banglish linguistic patterns. The results of our thesis can be used to improve multilingual hate speech detectors, improve moderation tools and aid in the comprehension of cultural sensitivities in Banglish conversation. Not only that but also regarding the identification of hate speech in Banglish, there is only one recognized work. Our work establishes a standard for Banglish language processing and makes a substantial contribution to this emerging discipline.

EA4(Consequences for society and environment)

Rationale: For justifying EA4, Our study, which focuses on hate speech identification in Banglish, fills a crucial need in the field of online toxicity moderating in multilingual and code-mixed environments. By reducing the psychological effects of damaging information and encouraging diversity and moral communication, this novel approach creates safer online environments. Our research empowers marginalized groups and promotes polite conversation by addressing hate speech relating to sexual orientation, race, religion, and appearance, which benefits various communities. Additionally, it opens the door for more developments in AI systems designed for multilingual cultures. But there are difficulties. Concerns regarding freedom of expression, misclassification and restriction may arise from automated systems. Errors may result in unexpected limitations or unreported misuse. Even though hybrid model optimization reduces computational load, complex system training can still have an adverse effect on the environment because of its high energy consumption. To guarantee that our research optimizes societal advantages while successfully resolving any negatives, it will be crucial to strike a balance between ethical use, fairness and sustainability.

5.5 Summary

This whole chapter covers a lot of information, impacts and proves our research topic. Our research's impact on our society, environment and life is positive. We have discussed the financial analysis and management of our research. Not only that but also we have proved the EP's, EA's and K's by giving the proper reasons and explanations.

Chapter 6

Conclusion

This chapter is all about the conclusion of our research. And also the limitations and future works are discussed here.

6.1 Summary

The purpose of this study is to design a hybrid model based on machine learning techniques that can identify hate speech in Banglish, a language of using the English alphabet to write Bangla. The increase of hate speech on social media sites also makes the task difficult of preserving a safe and friendly cyber environment. Because the Banglish language is different in many aspects, the conventional approaches generally do not perform well, and so, an appropriate approach is must. Besides non-hate category, the proposed model applied a bidirectional GRU-based neural network architecture to separate different types of hate Banglish texts, includes the categories of appearance, racial, religious, sexual, slang, and others. In order to prepare the datasets to train a deep learning model, a preprocessing pipeline process applied including text tokenizing, padding, and stop-word removing. To improve the model's performance and decrease the class imbalance issue, there used oversampling methods for the training dataset. Besides, Data augmentation including constructing a large dictionary of the word to its base form was created by us. It was about a dictionary of 10000 words. Full form of any abbreviation was written as well as the base form of a word. The methods applied in the research simplify the achievement of better outcomes, especially the accuracy of the classification, which was significantly improved by the application of advanced techniques applied to categorical cross-entropy loss and embedding layers. Performance measures obtained through application of confusion matrix and classification report evaluation indicators greatly suggest that the model can detect hate speech with great precision after subjecting it to wide-ranging training and assessment procedures. The research has also highlighted ethical concerns with respect to privacy, fairness and bias in the design and usage of the model. Society is also concerned about encouraging content moderation on digital platforms, making the Internet safer, as well as reducing the psychotraumatic impact of hate speech. The study also implied environmental

and sustainability aspects as they focused on modelling training that optimizes computer efficiency and considers renewable energy sources. In conclusion, our research addresses a crucial gap regarding the detection of hate speech in Banglish and presents a workable, ethical, and scalable approach to curb hate speech on the internet. The proposed automated recognition system may act as independent hardware, as a search engine or as a browser plug-in or as a script in the code when a website is built, potentially revolutionizing social media moderation and computational linguistics. The research also admitted social and sustainability by maximizing computing efficiency and renewable energy sources during the time of model training.

6.2 Limitation

In spite of achieving good results in this research, there are some limitations that should be highlighted. First, the dataset, even though being balanced by oversampling techniques, may not represent well the diversity and small variation of real world data Banglish hate speech. Applying RandomOverSampler for balancing the class would likely increase an overfitting problem since the actual balanced data is different from what is simulated in the training process. Moreover, the dataset also strongly biases certain specific types of hate speech forms that could miss new types or slight language differences from hidden categories of hate speech in unseen data. The second limitation comes from traditional deep learning models dependency e.g LSTM, GRU and their hybrid varieties.. These models perform well but take a long time to train with large datasets as well as restriction with real time application. Also capturing long term dependencies and context variation in Banglish often cause difficulties for these models are such natural complexity due to codo-mixing bi-langauge structure. The more recent Transformer-based architectures e.g BERT or DistilBERT were not employed into our study but they could achieve better performance by adopting contextual relationship and meaning more appropriate with representing Banglish text. This study suffered several limitations for classifying minority classes such as Sexual and Others which resulted in lower precision and recall performance compared to the dominating classes. It reflects that it is difficult to train effective models for recognizing different types of hate speech by models, using training examples of rarely occurring text sequences in dataset. Also, some preprocessing steps like removal of stop words or tokenization may either remove contextually important information which could have had an impact on overall classification accuracy. In addition, developed models were also trained and evaluated on a static dataset. We cannot ensure whether the proposed approach can be useful in practice real time such as

live social media data. The Banglish language is very dynamic and changing too fast following culture and trends than other languages. So, a static model could fail to provide desirable results. In addition, the models were trained and evaluated on a static dataset, so it is still uncertain that the obtained results can be transferred to real-time data (live social media data). Banglish is a dynamic language which modifies very quickly, denoting cultural and social fashion. In this case, the static models could be less effective with due course of time. Thus future works should attempt to incorporate continuous learning mechanisms to cater these variations and it would also be beneficial if such models should be validated in large scale more dynamic datasets from different sources like facebook, instagram, snapchat, youtube, X etc. Securities issues of using these sources also play a vital role for getting authorization for looking into those corpora. It will also be beneficial if such similar studies should concentrate towards some particular domains e.g. Business women, entrepreneur, public figure, teacher etc.

6.3 Future Work

In future we will apply more traditional, deep and hybrid models to increase the performance and accuracy. And we will try to increase our accuracy from 83.93% to nearly 90%. Not only that but we also want to take our work further by writing a journal about our topic. We will build a real time web page. Not only that but also we will apply our own created model to this web page. We will break the limitations for difficult models and also reduce our data scarcity as we know that the more we will increase our dataset, the more we can achieve higher accuracy.

References

1. M. H. Islam, K. Farzana, I. Khalil, S. Ara, M. R. A. Shazid, and M. H. K. Mehedi, “Unmasking Toxicity: A Comprehensive Analysis of Hate Speech Detection in Banglish,” Department of Computer Science and Engineering, Bangladesh University of Business and Technology, and Brac University, Dhaka, Bangladesh, 2024.
2. M. Zamir, S. Tash, Z. Ahani, A. Gelbukh, and G. Sidorov, “Lidoma@DravidianLangTech 2024: Identifying Hate Speech in Telugu Code-Mixed: A BERT Multilingual,” 2024.
3. T. Sharan, A. Dhaka, and D. Jain, “Spam Detection and Hate Speech Identification,” 2024.
4. N. Romim, M. Ahmed, H. Talukder, and M. Islam, “Detection of Abusive Bengali Comments for Mixed Social Media Data Using Machine Learning,” 2024.
5. F. Chan, D. Nguyen, and A. Joshi, “Is Hate Lost in Translation? Evaluation of Multilingual LGBTQIA+ Hate Speech Detection,” 2024.
6. E. Hashmi and S. Y. Yayilgan, “Multi-class Hate Speech Detection in Norwegian Using FAST-RNN and Multilingual Transformers,” 2024.
7. L. Dhanya and K. Balakrishnan, “Integrating Hybrid Neural Networks and Domain-Specific Embeddings for Detecting Hate Content in Code-Mixed Social Media Comments,” 2024.
8. D. Sharma, A. Singh, and V. K. Singh, “THAR: Targeted Hate Speech Against Religion: A High-Quality Hindi-English Code-Mixed Dataset with the Application of Deep Learning Models for Automatic Detection,” 2024.
9. S. Ghosal, A. Jain, D. K. Tayal, V. Menon, and A. Kumar, “Inculcating Context for Emoji Powered Bengali Hate Speech Detection Using Extended Fuzzy SVM and Text Embedding Models,” 2024.

10. S. Yadav, A. Kaushik, and K. McDaid, "Leveraging Weakly Annotated Data for Hate Speech Detection in Code-Mixed Hinglish: A Feasibility-Driven Transfer Learning Approach with Large Language Models," 2024.
11. S. Yadav, A. Kaushik, and K. McDaid, "Hate Speech is Not Free Speech: Explainable Machine Learning for Hate Speech Detection in Code-Mixed Languages," presented at the School of Informatics and Creative Art, Dundalk Institute of Technology, Dundalk, Ireland, Sep. 2023.
12. N. Romim, M. Ahmed, H. Talukder, and M. Islam, "Hate Speech Detection in the Bengali Language: A Dataset and Its Baseline Evaluation," 2024.
13. S. Chava, S. R. Kumar, S. Saumya, and S. Biradar, "IITDWD_SVC@DravidianLangTech-2024: Breaking Language Barriers; Hate Speech Detection in Telugu-English Code-Mixed Text," 2024.
14. J. A. Kabir, K. Kabir, N. Shamme, M. Mridha, Y. Watanobe, and M. Islam, "G-BERT: An Efficient Method for Identifying Hate Speech in Bengali Texts on Social Media," 2024.
15. F. Naznin, M. Rahman, and S. Alve, "Hierarchical Sentiment Analysis Framework for Hate Speech Detection: Implementing Binary and Multiclass Classification Strategy," 2024.
16. J. Ndabula, O. Olanrewaju, and F. Echobu, "Detection of Hate Speech Code-Mix Involving English and Other Nigerian Languages," 2024.
17. J. Kabir, M. Oussalah, and A. Singhal, "Cross-Linguistic Offensive Language Detection: BERT-Based Analysis of Bengali, Assamese, & Bodo Conversational Hateful Content from Social Media," 2024.
18. M. Saroar, M. Haque, N. Arhab, and M. Oussalah, "BanglaHateBERT: BERT for Abusive Language Detection in Bengali," 2024.
19. A. K. Das, A. A. Asif, A. Paul, M. Hossain, and A. Das, "Bangla Hate Speech Detection on Social Media Using Attention-Based Recurrent Neural Network," Department of

Computer Science Engineering, East West University, Dhaka, Bangladesh, 2024.

20. T. Davidson, D. Warmley, M. Macy, and I. Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," 2024.
21. C. Sai, R. Kumar, S. Saumya, and S. Biradar, "IIITDWD_SVC@ DravidianLangTech-2024: Breaking Language Barriers; Hate Speech Detection in Telugu-English Code-Mixed Text," in Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, Mar. 2024, pp. 119–123.
22. J. K. Mim, M. Oussalah, and A. Singhal, "Cross-Linguistic Offensive Language Detection: BERT-Based Analysis of Bengali, Assamese, & Bodo Conversational Hateful Content from Social Media," arXiv preprint, arXiv:2312.10528, Dec. 2023.
23. <https://www.kaggle.com/datasets/tahbibmanzoor/bengali-in-english-letter-hate-speech-dataset> (Access: 22 April 2024).

IDENTIFYING HATE SPEECH IN BANGLISH LANGUAGE Final Year Design Project

ORIGINALITY REPORT

8%

SIMILARITY INDEX

4%

INTERNET SOURCES

6%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Universidad Internacional de la Rioja

Student Paper

1%

2

fastercapital.com

Internet Source

1%

3

Submitted to United International University

Student Paper

<1%

4

Anil Kumar Dubey, Vikash Yadav, Munesh Chandra Trivedi. "Computational Intelligence in the Industry 4.0", CRC Press, 2024

Publication

<1%

5

www.mdpi.com

Internet Source

<1%

6

Mauajama Firdaus, Hitesh Golchha, Asif Ekbal, Pushpak Bhattacharyya. "A Deep Multi-task Model for Dialogue Act Classification, Intent Detection and Slot Filling", Cognitive Computation, 2020

Publication

<1%