BENGALI SOCIAL MEDIA COMMENTS ANALYSIS TO PREVENT CYBERBULLYING

BY

A. S. M. Sohag Abdullah
ID: 201-15-13845
AND

Sany Debnath
ID: 201-15-14152

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. S. M. Aminul Haque
Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Bengali Social Media Comments Analysis To Prevent Cyberbullying", submitted by **A. S. M. Sohag Abdullah** (ID:201-15-13845) and **Sany Debnath** (ID:201-15-14152) 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 15-07-2024.

BOARD OF EXAMINERS

AHossain 15.07.2024

Dr. Md. Fokhray Hossain (MFH) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

15.7.24

Md. Firoz Hasan (FH) — **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Lamia RukAnsar 15.07.24

Lamia Rukhsara (LR) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

9.7.24

Dr. S. M. Hasan Mahmud (SMH) — **External Examiner**
Assistant Professor
Department of Computer Science
American International University-Bangladesh

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S. M. Aminul Haque , Professor & Associate 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. S. M. Aminul Haque
Professor & Associate Head
Department of CSE
Daffodil International University

Submitted by:

Sohag

A. S. M. Sohag Abdullah
ID: 201-15-13845
Department of CSE
Daffodil International University

Sany

Sany Debnath
ID: 201-15-14152
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Dr. S. M. Aminul Haque**, **Professor & Associate Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*machine learning and artificial intelligence*" 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 heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Professor & Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

One of the significant wonders of modern technology is social networking platforms. This gives a major acceleration to the field of communication. With such good advantages of this technology this has a darker side too. Nowadays people are often seen trolling each other online. There is also a new roasting culture in online platforms which arrived a few years ago, where people are often downgrading other people without any moral consideration. Cyberbullying nowadays has become a trend. People are now just waiting for a new topic to arrive in social media so that they can make fun of each other. So we are doing research on this field so that we can make an automated system which would be able to detect cyberbullying instances on social platforms using textual analysis with the help of machine learning and artificial intelligence. We are using social media comments in Bangla language from Facebook for this research along side we are also taking a dataset from kaggle. Our personal Bangla comments dataset contains 5053 Bengali comments. These comments are initially categorized as bully or non-bully. And the bully comments are further categorized as Sexual, Religious, Threat, and Troll and Political comments. The kaggle dataset also contains Bengali social media comments categorized as Neutral, Political, Sexual, troll and threat. We have run experiments on both the dataset individually and combining them together. We have used both machine learning and deep learning architecture for our experiments. We have achieved highest accuracy of 88% on binary class classification on the kaggle dataset. And highest 65% on multiclass classification on the kaggle dataset as well.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
RF	Random Forest
ETC	Extra Tree Classifier
LR	Logistic Regression
DT	Decision Tree
NB	Naïve Bayes
Bi-LSTM	Bidirectional Long Short-Term Memory
AI	Artificial Intelligence

CHAPTER 1
INTRODUCTION

1.1 Overview

Cyberbullying has now become an uncontrollable modern problem. It has literally become impossible for us to prevent people from engadgeing in this social menace. So it is high time we develop an automated cyberbullying detection system to prevent people from doing this. Many studies has been already done in cyberbullying detection. Most of them include machine learning approaches. We are doing both machine learning approach and deep learning approach. For Machine learning approach we are running experiments using Extra Trees Classifier (ETC), Naïve Bayes (NB), Logistic Regression (LR) and Decision Tree (DT). We are also using stacking classifier to combine three machine learning algorithms to find out how they perform. And also we are checking how deep learning based model such as Bidirectional LSTM works on Bengali social media comments classification.

1.2 Background and Present State

Cyberbullying emerged with the rise of the internet and social media in the late 20th and early 21st centuries. Unlike traditional bullying, which typically occurred in person and could be more easily monitored by adults, cyberbullying introduced new challenges due to its anonymous and instantaneous nature. This anonymity often emboldens perpetrators who might not face immediate consequences for their actions. Early responses to cyberbullying were limited, as awareness of the issue and its impact grew slowly. Platforms initially struggled to develop effective policies and tools to address harassment adequately. Victims often found it difficult to escape the persistent nature of online abuse, as content could be shared widely and remain accessible long after it was posted. Today, cyberbullying remains a significant concern globally, affecting individuals of all ages, though adolescents and teenagers are particularly vulnerable. Social media platforms and technology companies have made strides in combating cyberbullying through various means like policies and reporting, awareness and education etc. Despite these efforts, challenges remain. In summary, while progress has been made in facing cyberbullying, it remains an issue that demands our attention.

1.3 Problem Statement

Cyberbullying is a form of harassment or bullying that occurs through electronic means, typically on the internet or through digital devices. And as there is no stopping people from doing so they are getting encouraged in conducting such activities. Moreover they are now in a situation where there is a virtual competition of conducting cyber bullying. Roasting culture has pushed the virtual world in a state that the people are not thinking about others' mental situation. And women and teenagers are the main victims of it.

1.4 Objectives

As people are going through a lot of trauma in fighting such situations where they are being bullied by other people on a large scale on social platforms, it is high time we take a step towards this social disorder. But making people stop conducting cyberbullying is a challenge in the modern world as we can not identify people who are conducting such activities in real time. And manually stopping people from harassing each other on social platforms is impossible. So automation is essential in this sector. We can achieve this through machine learning and artificial intelligence. Although there are people already doing research about it. Most of the work is in the English language. Only a few works exist in Bengali Language. So we are doing this research on Bengali social media comments to detect cyberbullying instances using machine learning and try to fight against this social disorder to remove toxicity from our society. Our study is using Bangla comments from social media and using them as a learning tool for a system to recognize whether a comment posted on social media is an instance of bully comment or not

1.5 Scope and limitations

Cyberbullying operates within a broad scope, enabled by the widespread reach of digital platforms like social media, messaging apps, and online forums. It transcends geographical boundaries, allowing perpetrators to target victims globally with relative anonymity. Its manifestations are diverse, ranging from harassment and rumor-spreading to more sophisticated tactics like doxxing and impersonation. The psychological impact on victims can be severe, often leading to anxiety, depression, and other mental health issues. Legal and ethical complexities arise due to differing international laws and debates over freedom

of speech versus the harm caused by online behavior.Despite technological advancements and platform policies aimed at prevention, cyberbullying faces inherent limitations. Anonymity online enables perpetrators to evade detection and accountability, complicating efforts to enforce consequences. The sheer volume of digital content challenges platforms' ability to monitor and respond effectively to every instance of abuse. Technological solutions such as AI algorithms for content moderation are not infallible and can struggle with context and nuance, leading to both over- and under-detection of cyberbullying. Underreporting by victims, due to fear of retaliation or a lack of confidence in intervention, obscures the true prevalence of cyberbullying. Educational gaps persist in digital literacy and awareness among both youth and adults, hindering effective prevention strategies. Cultural norms and attitudes towards bullying vary globally, influencing how cyberbullying is perceived and addressed within different communities. Addressing these challenges requires a comprehensive approach involving collaboration between technology companies, policymakers, educators, and communities to foster safer online environments and mitigate the impact of cyberbullying effectively.

1.6 Report Organization

The proposed report is structured with a comprehensive layout to guide readers through the research process, findings, and implications. Each chapter serves a specific purpose, contributing to the overall coherence and depth of the document.

Chapter 1: Introduction

The introductory chapter sets the stage for the research, providing the background for cyberbullying detection. Here we have discussed briefly about our approach to solve the problem of cyberbullying in social media. Also we have discussed the research questions that arises in our research and also about the rational of our study.

Chapter 2: Literature review

In the background chapter, we have briefly summarized all the previous studies outcome. And their limitations and gap. And we have also discussed what we are going to do based on those gaps and discuss how to fill these gaps.

Chapter 3: Methodology

The research methodology chapter describes the approach used to perform the study. It explains the study design, data gathering techniques, model architecture, and the reasoning for selecting various methodology. The chapter also examines ethical concerns and any restrictions that may affect the study.

Chapter 4: Implementation

This Chapter discusses about our web application prototype. Here we have discussed about the design of our application and how our application evaluates given data. And how our application generates output.

Chapter 5: Result and analysis

The societal impact chapter explores the broader implications of the research findings on Cyberbullying detection. It discusses how cyberbullying detection methods can positively influence the mental health of people. Future consideration is also given to improve the outcomes as well

Chapter 6: Impact on society, environment and sustainability

This chapter serves as a summarization of the entire report. It depicts the conclusion of our study and also give the recommendation for future improvement in the field of cyberbullying detection and prevention of cyberbullying on social media platforms.

Chapter 6: Conclusion and future work

This final chapter concludes our research about cyberbullying in Bangla language and suggests us future aspects of this study.

1.7 Summary

This report addresses the pressing issue of cyberbullying, which has become pervasive in the digital age, particularly affecting vulnerable groups like adolescents and women. Highlighting the evolution and challenges of cyberbullying, the study proposes a solution through automated detection systems using both machine learning and deep learning approaches. It focuses on Bengali social media comments, an under-researched area

compared to English, aiming to identify instances of bullying effectively. Despite advancements, cyberbullying's complexities persist due to anonymity, legal variations, and technological limitations in content moderation. The report's structured layout guides readers from introduction to methodology, experimental results, societal impact, and concludes with recommendations for future research, emphasizing the need for collaborative efforts across sectors to mitigate this societal menace effectively.

.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

Before diving into Bengali cyberbullying detection, we need to take some preliminary steps: We need to understand the existing research and methodologies in cyberbullying detection. This will help us understand the gaps of those previous studies. Cyberbully is a kind of bully that occurs on the internet platform via social media platforms. People do this through their keyboards in personal computers, laptops, tablets and mobile phones. For the purpose of our study, we are only relying on textual Bengali data taken from the Facebook comment section of several Facebook posts. Before using our acquired data, we will have to do some preprocessing for better results. Such kind of data preprocessing techniques involve tokenization, stemming, removing stop words etc. And then we need to apply machine learning classification algorithms like ETC, NN, logistic regression, Decision tree etc. and deep learning models like Bi-LSTM to train our model with the dataset. And then validating our model with unknown data will help us with the evaluation of our model.

2.2 Related Works

Here are some related studies regarding cyberbullying:

X. Zhang et al [1] proposed a punctuation based CNN for cyberbullying detection. They used a twitter dataset and a dataset from Formspring.me and applied their model. They achieved 98.9% accuracy in the twitter dataset. And on the 96.8% accuracy on the Fromspring.me dataset. J. Hani et al [2] cyberbullying dataset from Kaggle. They used TFID to extract features from the dataset and used sentiment analysis to find out the polarity of a sentence. They applied two classifiers, SVM (support vector machine) and NN (neural network) . The average n-gram accuracy for SVM is 89.6% and average n-gram accuracy for NN is 91.76%. M. Dadvar et al [3] also used TFID and SVM on WEKA with Myspace dataset. But their limitation was a limited sized dataset. M.Islam et al.[4] also used NLP and machine learning for bullying detection. They used two datasets. One of them contains facebook comments and another one from kaggel. They used SVM for both the dataset and achieved 78.5% accuracy. N. A. Azeez et al [5] used many ML algorithms to detect cyberbullying. For example Naive Bayes classifier, K-Nearest Neighbor (KNN) Algorithm, Logistic Regression Classifier, Decision Tree Classifier,

Random Forest Classifier, Linear Support Vector Classifier, AdaBoost Classifier, Stochastic Gradient Descent Classifier, Bagging Classifier, Ensemble Classifier, Justification for the choice of Multinomial NB, Linear SVC and Logistic Regression for Ensemble . Y. Fang et al [6] designed a sequence model to automatically classify cyberbullying posts in social networks based on deep neural networks. They evaluated their model with other classifiers such as SVM and LR (logistic regression) using precision, recall and F1 score. R. Ghosh et al.[7] researched on Bangla language and applied four classification algorithms Random Forest Classifier, Support Vector Machine,Logistic Regression and Passive Aggressive Classifier. Using n-gram Passive Aggressive Classifiers achieved highest accuracy with 78.1 %. SVM achieved 61.1% accuracy on word level. G. Hussain et al [8] used very root level method to detect cyberbullying in sentences using the weight of bullying words in a sentence. They only used 300 comments for their experiment. M. Alotaibi et al [9] used three advanced deep learning models namely: The transformer block, BiGRU, CNN architecture for cyberbullying detection on social networks. They combined three public dataset and evaluated their proposed model. They achieved 87.99% accuracy. Tokenization or lexical analysis and several python libraries were used for data preprocessing. M. Di Capua et al [10] used an unsupervised method for cyber bullying detection in social networks. They used a public dataset and their own data set for evaluation. On the Formspring.me dataset they achieved 73% accuracy. On the other hand they achieve 69 % accuracy and 72% accuracy on their own dataset which were taken from youtube and twitter. Md. T. Ahmed et al [11] used a hybrid ensemble method using deep learning techniques in Bangla dataset for cyberbullying detection. They achieved an accuracy rate of 85%. They used TF-IDF for feature extraction. L. Cheng et al [12] Hierarchical Attention Networks for Cyberbullying Detection (HANCD) on instagram dataset. For which they used a public dataset. P.-J. Lee et al [13] used python 3 to crawl tweets on Twitter and build the dataset needed for the cyberbullying prediction model. The algorithms they used are,k-Nearest Neighbors (KNN), Support Vector Machine (SVM) and C4.5 Decision Tree (DT). They used two types of variables in their research: i) textual features: variables generated by texts such as text readability, sentiment scores, and vulgar words. ii) User features: the time to last tweet, the total number of tweets of the user, and other records of the user-behavior-related variables. The 10-fold cross-validation is used for evaluations of the three models .No result was given in their paper. Md. T. Ahmed et al

[14] researched both Bangla and Romanized Bangla language to detect cyberbullying. The algorithms used by them are Multinomial naive Bayes, Support vector machine (SVM), Logistic regression, XGBoost. Out of these the Multinomial naive Bayes, Support vector machine (SVM) achieved highest accuracy. For only the Bangla dataset they achieved 76% accuracy. For the Romanized Bangla dataset they achieved 84% accuracy. And with mixed data they acquired 80% accuracy. They took Youtube comment to build their dataset. T. Islam et al[15] applied multinomial Naïve Bayes (MNB), multilayer perceptron (MLP), support vector machine (SVM), decision tree, random forest, stochastic gradient descent (SGD), ridge, perceptron and k-nearest neighbors (k-NN) to determine abusive Bangla text. Out of these classifiers SVM achieved the highest 87% accuracy. For data preprocessing tokenization and stemming was used. And TF-IDF was used for feature extraction. Faisal et al used random forest, SVM, KNN, Naïve Bayes to detect cyberbullying. For binary classification accuracy is 87.91% and for multiclass classification accuracy is 85%. Stop words Removal, Tokenization of String (used a property of tensorflow, which is called 'Tokenizer') and Padded sequence conversion these were used for data preprocessing. M. Dadvar et al [17] experimented with an English dataset with a gender specific approach and achieved higher precision , recall and F-measure. Abdhullah-Al-Mamun et al.[18] used Naive Bayes,J48,SVC,KNN (1- Nearest) ,KNN (3- Nearest) and SVM to detect cyberbullying. Only 2400 Bangla text from social media posts and user specific data were used to evaluate bully detection. Out of these algorithms SVM achieved the highest accuracy. Using only posts, accuracy was 95.40%. And along with user specific data the accuracy was 97.27%. A. Bozyiğit et al [19] achieved 87.6% accuracy without social media features and 90.1% accuracy with social media features. The algorithm used was Adaptive boosting (AdaBoost). A. Akhter et al [20] used Decision Tree, Random Forest, Logistic Regression, Multilayer Perceptron and built a hybrid machine learning model to detect cyberbullying in Bangla social media comments. And they achieved 98.57% accuracy for binary classification and 98.82% accuracy for multiclass classification. The only limitation is they have used a public dataset for their research purpose. And there is no real world application of their findings.

2.3 Comparison between existing works

As we have already seen different studies have different results. Some having higher accuracy then others and some are using different types of algorithm. However the studies are there are always some gaps or limitations. Table 2.3.1 depicts the comparative analysis between some of the studies.

TABLE 2.3.1. COMPARATIVE ANALYSIS WITH PREVIOUS WORK

SL.	Author Name	Used algorithm	Result	Limitations
1	X. Zhang *et al* [1]	pronunciation based convolutional neural network(PCNN)	Twitter dataset : 98.9% Formspring.me : 96.8%	Use of public dataset
2	J. Hani et al[2]	N-gram NN(neural network) N-gram SVM	N-gram NN:91.76% N-gram SVM:89.87%	Use of public dataset
3	N. A. Azeez et al[5]	Naive Bayes,KNN,LR,DT etc	-	This study is also limited to English language tweets and data.
4	R. Ghosh et al[7]	RF,SVM,LR,PAC	PAC: 78.1%	low accuracy
5	M. Alotaibi et al[9]	The transformer block, BiGRU, CNN architecture.	87.99%,	Public dataset
6	M. Di Capua et al[21]	-	FORMSPRING.ME DATASET - 0.73 YOUTUBE DATASET - 0.69 TWITTER DATASET - 0.72	i) Public dataset ii)Low accuracy
7	Md. T. Ahmed et al[11]	Hybrid Ensemble approach	85%	Accuracy can be improved

8	Md. T. Ahmed et al[14]	Multinomial naive Bayes,Support vector machine (SVM) ,Logistic regression,XGBoost	i) Bangla dataset : 76% ii) Romanized Bangla dataset : 84% iii) Mixed dataset : 80%	Accuracy can be improved
9	T. Islam et al [15]	SVM	SVM: 87.0%	Dataset relatively small
10	Md. F. Ahmed et al[16]	NN	i) Binary classification : 87.91% ii) Multi Labeled classification: 85%	Accuracy can be improved
11	Abdhullah-Al-Mamun et al[18]	Naive bayes,J48,SVC,KNN,SVM	i) USING USER'S POSTS ONLY : 95.40% ii) USING POSTS AND USER SPECIFIC DATA: 97.27%	i) Limited to Bangla language ii) Use of public dataset iii) No real world application

2.4 Open Issues

Here are some of the significant challenges encountered during our research:

1. **Dataset Collection:** Acquiring data from social media required manual efforts, posing challenges in identifying cyberbullying comments individually.
2. **Imbalanced Dataset:** The dataset exhibited irregularities and included multiple labels, resulting in imbalance issues that needed careful handling.
3. **Language Barriers:** Modern social media users often employ Romanized Bengali, complicating the process of identifying and analyzing full Bengali comments accurately.

2.5 Summary

This section of the report provides a comprehensive overview of cyberbullying detection research, particularly focusing on Bengali language datasets extracted from Facebook comments. The study outlines essential preliminary steps, including data preprocessing techniques like tokenization and stemming, crucial for enhancing model performance. It draws comparisons with various existing methodologies and algorithms used in cyberbullying detection, highlighting their strengths and limitations. The comparative analysis across studies underscores the variability in results and methodologies, emphasizing the need for tailored approaches in different linguistic and social contexts. Key challenges identified include manual data collection from social media, handling imbalanced datasets, and overcoming language barriers posed by the use of Romanized Bengali. These insights inform the research methodology and pave the way for developing more effective cyberbullying detection systems tailored to Bengali social media environment.

CHAPTER 3
METHODOLOGY

3.1 Overview

The research subject of "Bengali social media comments analysis to prevent cyberbullying" revolves around the application of advanced artificial intelligence techniques in the realm of text classification. Cyberbullying is a harmful social media malpractice where people just for the sake of having some pleasure downgrades another person's dignity. The research delves into the challenges associated with traditional methods of cyberbullying detection, seeking to overcome these limitations through the implication of machine learning and deep learning AI(artificial intelligence) architecture. Here are some detailed information about the used algorithms for our research:

Extra Trees Classifier : Or the short form of extremely randomized tree classifier. It is an ensemble learning method for classification tasks. It is a variant of random forest classifier. It builds multiple decision trees and combines their prediction to improve the accuracy and robustness of a model. Some of the key points about this algorithm is that it has randomization in splitting, bootstrap aggregation and highly parallelizable. This model is available in sckit-learn python library.

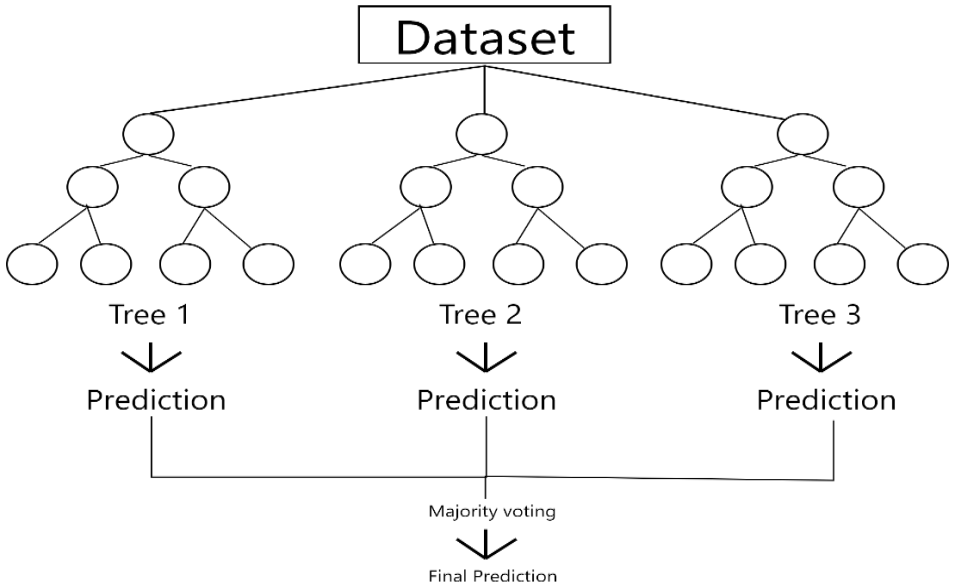

Figure 3.1.1: Extra Trees Classifier architecture

Logistic Regression: It is a statistical method generally used for binary class classification. It is implemented where the output class is between two values. In our case it is whether a text is bully or non-bully. It can also be used multiclass classification. Firstly the textual data is converted to numeric values and then we train the model with the converted values.

Naïve Bayes: Naive Bayes is a powerful and widely used algorithm for classification tasks, particularly in situations where computational resources are limited or when interpretability is important. It serves as a good baseline model and is often used in combination with other algorithms in ensemble methods. There are three kind of Naïve Bayes Gaussian Naive Bayes, Multinomial Naive Bayes, Bernoulli Naive Bayes. In our study we have used Multinomial Naïve Bayes

Decision Tree: Decision Tree is a versatile and widely used supervised learning algorithm for both classification and regression tasks. It builds a tree-like structure where each internal node represents a "decision" based on the value of a feature, each branch represents the outcome of the decision, and each leaf node represents the class label or the target value.

Stacking Classifier: Stacking, also known as stacked generalization, is an ensemble learning technique that combines multiple base classifiers (or regressors) with a meta-classifier (or meta-regressor) to improve predictive performance. It leverages the strengths of different algorithms by using their predictions as input features for a higher-level model. Base classifiers are a set of diverse and heterogeneous classifiers trained on the same dataset but using different algorithms or parameter settings. The meta-classifier, also known as the combiner or blender, takes the predictions of the base classifiers as input features and learns to make the final prediction. In our study we have used Random Forest and Gradient Boosting classifier as our base classifiers and Logistic regression as our meta classifier.

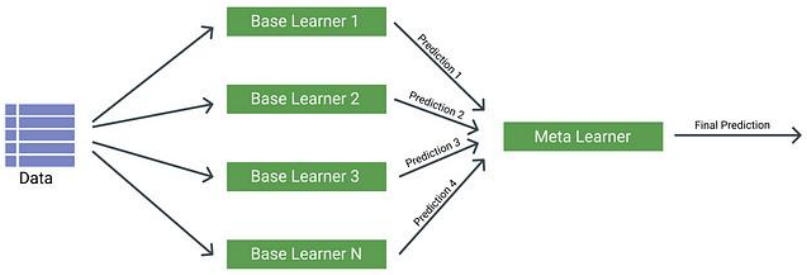

Figure 3.1.2: Stacking Classifier architecture

Bi-LSTM : BI-LSTM stands for Bidirectional Long Short-Term Memory. It's a type of recurrent neural network (RNN) architecture used in natural language processing (NLP) tasks, particularly for sequential data such as text. BI-LSTM networks are an extension of traditional LSTM networks, but they process input sequences in both forward and backward directions.

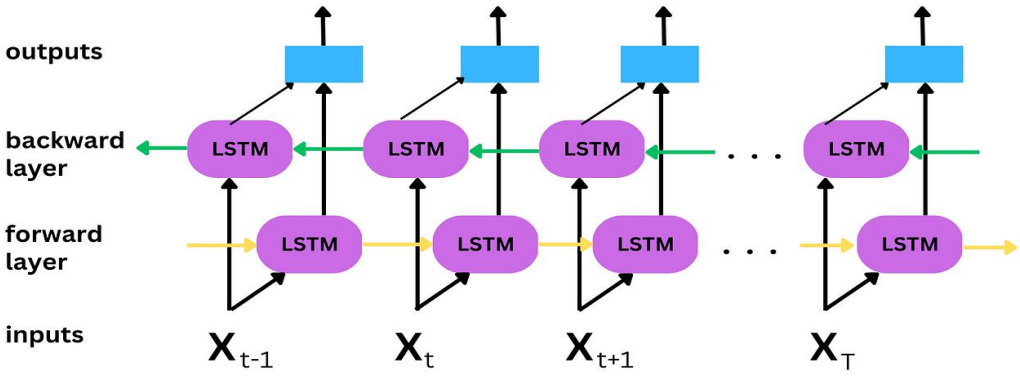

Figure 3.1.2: Bi-LSTM architecture

3.2 Data Collection Procedure

A. Datasets

We have collected various Bangla comments from facebook platform. Most of the comments are taken from public pages like Rafsan shabab, Xefer, Porimoni, Shakib Al hasan, Bangladesh Awami League, Bangladesh nationalist party, Taslima Nasrin, Asad Noor, Madhumita etc. These comments are taken manually and labeled manually. Initially we labeled them into the bully or non-bully category. And then the bully comments are

categorized in troll, political, sexual, religious, threat etc sub categories. There are in total 5053 Bangla comments in our dataset. Where 33.5% is Non-bully comments. The comments which are labeled Threat are 0.9%. 22.9% data are trolling comments.7.9% comments are for sexual comments. 4.3% comments are for political bullying comments. And finally 30.4 % of the comments represent religious bullying comments. We have also collected a public dataset from Kaggle.com. The public dataset also contains social media comments categorized in neutral / non-bully , political, sexual, troll and threat comments. Although there is no religious bullying comments we have enough religious bullying comments in our personal dataset to conduct the experiments. We will be conducting both binary class classification and multiclass classification. To conduct binary class classification we will have to do some preprocessing to the public dataset where we will be adding another column. In which all the neutral / non-bully comments will be labeled 'non-bully' and rest of the comments will be labeled 'bully'. We will be using these two dataset both separately and combinely to conduct our experiments. In the public dataset there are total 6011 comments and each category has 20.0% comments in the dataset.

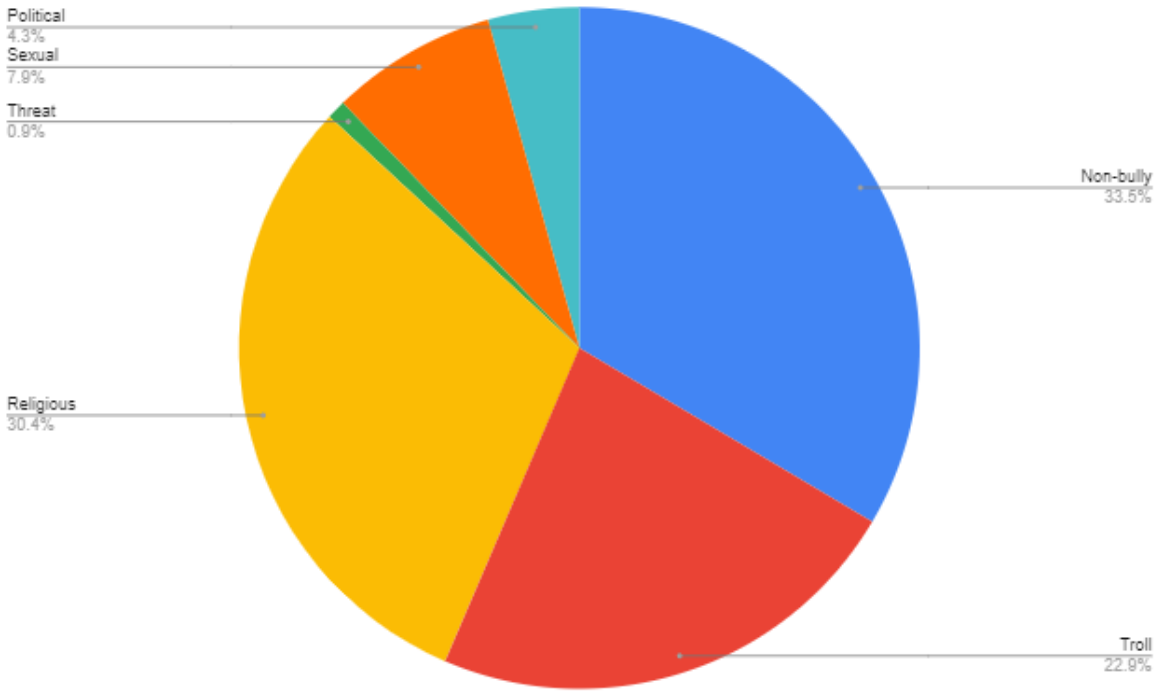

Figure 3.2.1: Data distribution in personal dataset.

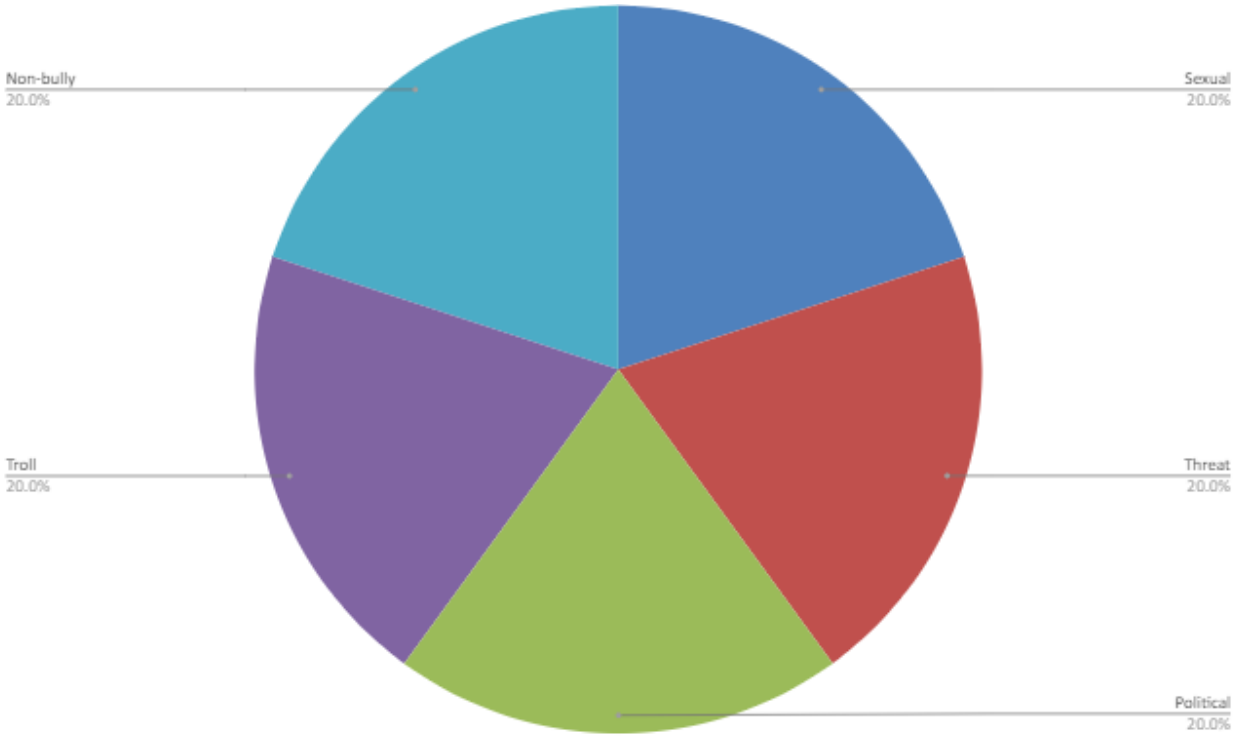

Figure 3.2.2: Data distribution in public dataset.

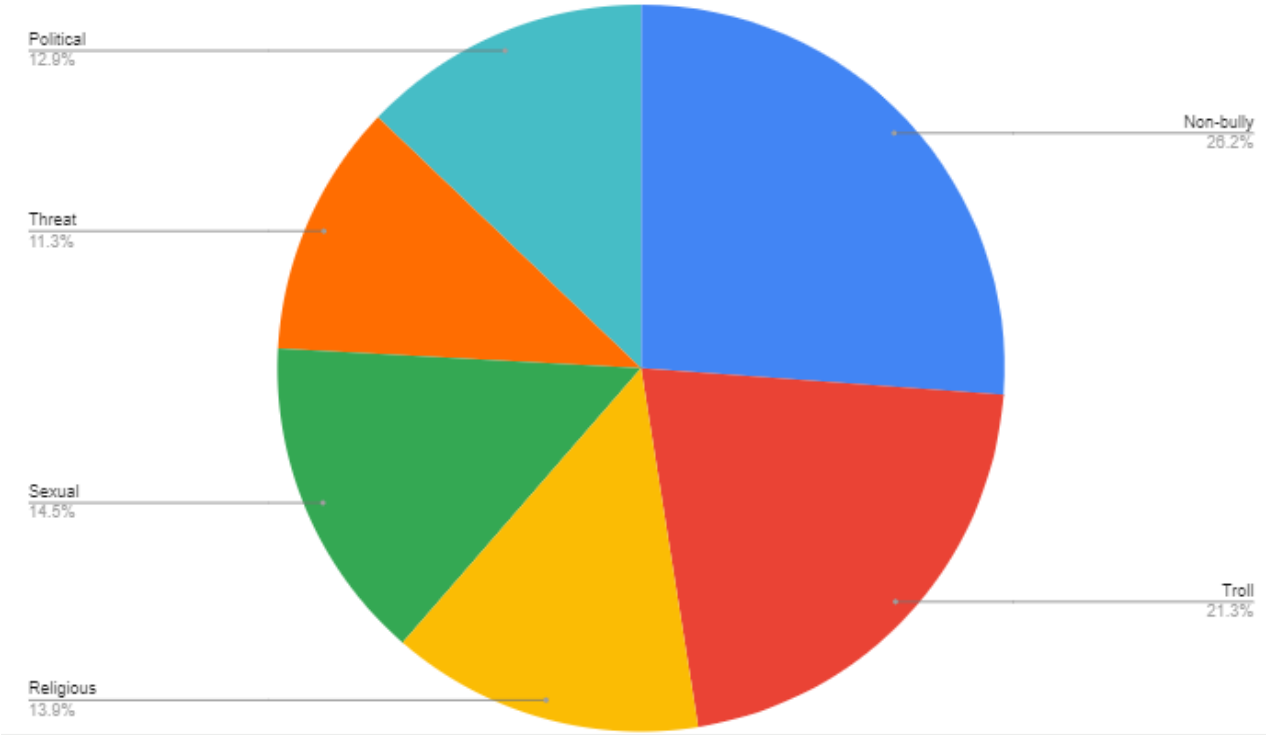

Figure 3.2.3: Data distribution in mixed dataset.

As we can see in the mixed dataset 26.2% comments are non-bully,12.9% comments are political bully,11.3% comments are threat, 14.5% comments are sexual,13.9% comments are religious bully and 21.3% comments are troll. Here is how the datasets look like:

TABLE 3.2.1: Dataset visualization

Comments	Category	Sub-category
বিষয়টা ভীষণ সেনসেটিভ এবং একান্তই ব্যক্তিগত। দিনশেষে ভালো থাকাটাই বড় কথা। আপনাদের দু'জনের জন্যই শুভকামনা রইলো।	Non-bully	Non-bully
সেলিব্রিটি লাইফ এমনই।। কিছু হলেই মানুষ তিল কে তাল বানাবে।।।আজকাল হাজার হাজার ডিভোর্স হচ্ছে সে খবর ক জন জানে?	Non-bully	Non-bully
মেডিয়া জগতের মানুষ গুলাই বাজারে কুত্তার মত তাদের এক তরকারি ভালো লাগেনা।	Bully	Troll
ভাই যে যাই বলুক, আপনার জীবন আপনার সিদ্ধান্ত।	Non-bully	Non-bully
এরে তোর কোহলে জাটার বাড়ি জাটার বাড়ি। তুই জাহান্নামি জাহান্নামি জাহান্নামি 😂	Bully	Religious
বিষয়টা ভীষণ সেনসেটিভ এবং একান্তই ব্যক্তিগত। দিনশেষে ভালো থাকাটাই বড় কথা। আপনাদের দু'জনের জন্যই শুভকামনা রইলো।	Non-bully	Non-bully

3.3 Statistical Analysis

The statistical analysis conducted in this research on "Bengali social media comments analysis to prevent cyberbullying" plays a vital role in deriving meaningful insights from the experimental results. In our experiments we looked for patterns in the data. It showed

that different types of comments contains different types of words. Although these findings was not use in model training but its gives us a better understanding about the datasets. In case of bully and non-bully comments the frequency of words in different cases are :

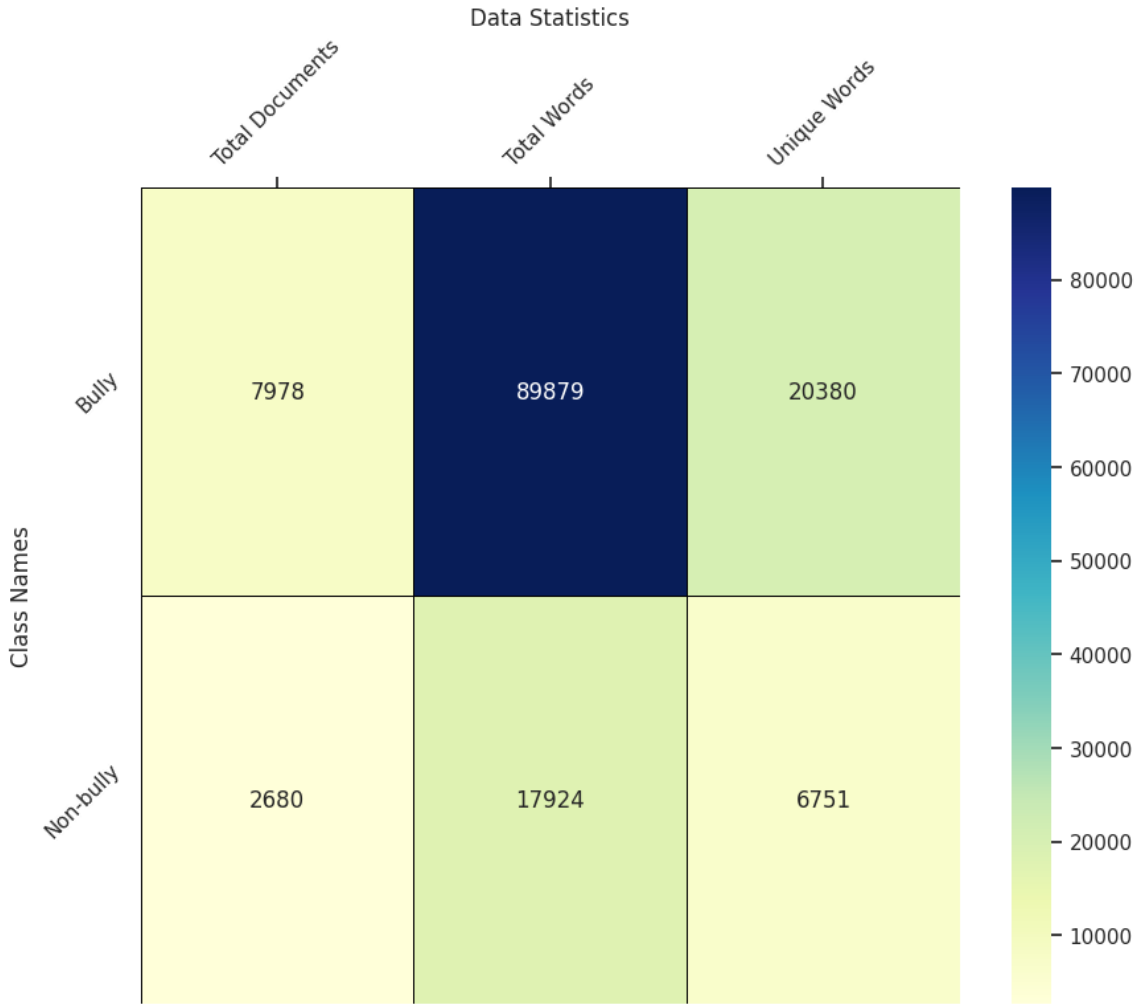

Figure 3.3.1: Word frequency in mixed dataset.

3.4 Proposed Methodology

As discussed before we are conducting our research in three different way. 1) Using Machine Learning model, 2) Using Stacking Classifier, 3) Using Deep Learning. We will firstly take combine take the datasets and then preprocess them. After that we will split the dataset. In case of machine learning and stacking classifier we will split the dataset into two parts, train and test set. And in deep learning we will split the dataset into three parts, train data, validate data and testing data. After training the models with the train data on each dataset, we will evaluate our models using the test data. Here are the three approach

we are going to use in cyberbullying detection. Here are the visual representations of the three different approaches:

Machine Learning Model:

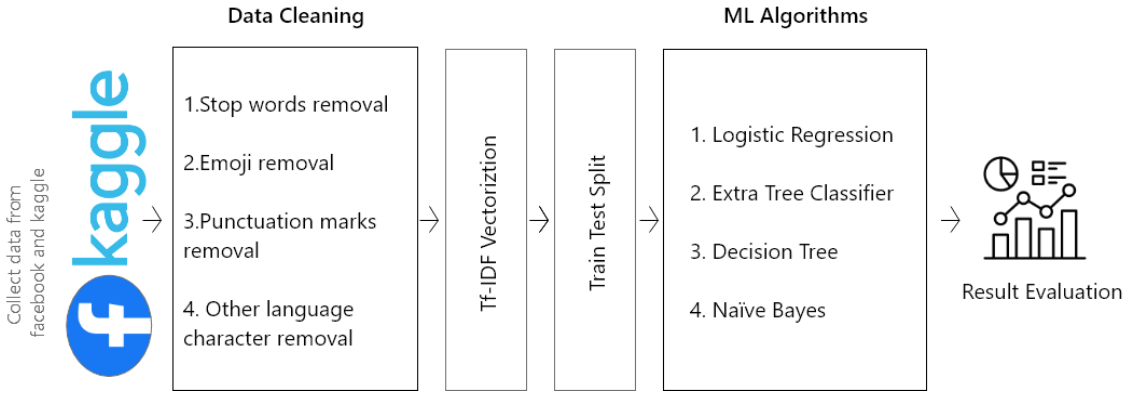

Figure 3.4.1: Proposed Methodology for machine learning model

As here we can see that in case of the machine learning model we are firstly collecting the raw data and preprocessing it. After that we are splitting the dataset into train and test set. The preprocessing techniques involved are stop words removal,emoji removal, punctuation marks removal, other language character removal and tfidf vectorization. After that we train the model with the training data and then we test the model with the test data. Then we evaluate the models performance with accuracy, precision, recall, confusion matrix and F1-scores for binary class classification. But for multiclass data we are only measuring the performance of the model using accuracy and confusion matrix.

Stacking Classifier Model:

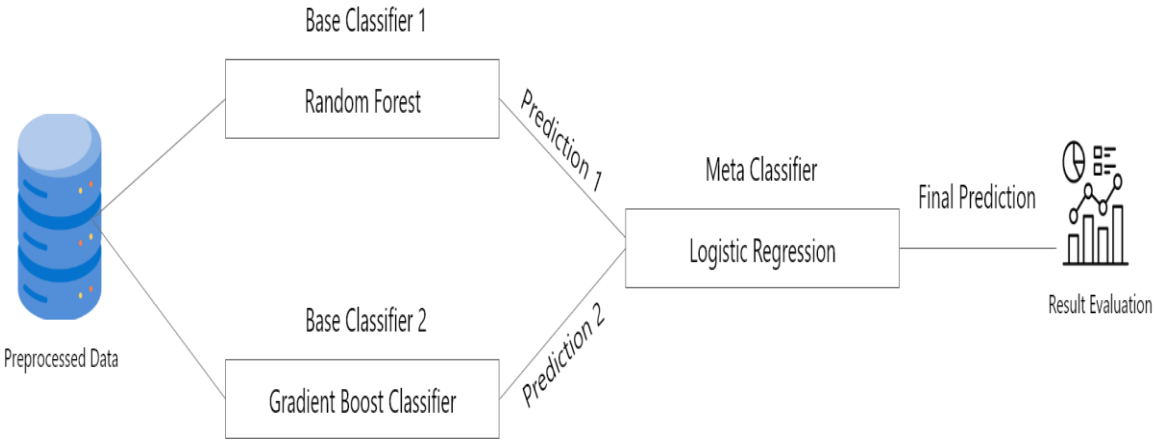

Figure 3.4.2: Proposed Methodology for Stacking classifier model

As shown in the picture most of the tasks in stacking classifier is same as machine learning model. But in case of the AI model we are actually merging three machine learning algorithm to make a stacking classifier. We are using random forest(RF) and Gradient Boost classifier as base classifiers and Logistic regression as the meta classifier. In this case we are also using accuracy, precision, recall, confusion matrix and F1-score to evaluate the performance of the model in binary classification. And in case of multiclass classification we are only using accuracy and confusion matrix for the evaluation.

Bi-LSTM (deep learning) Model:

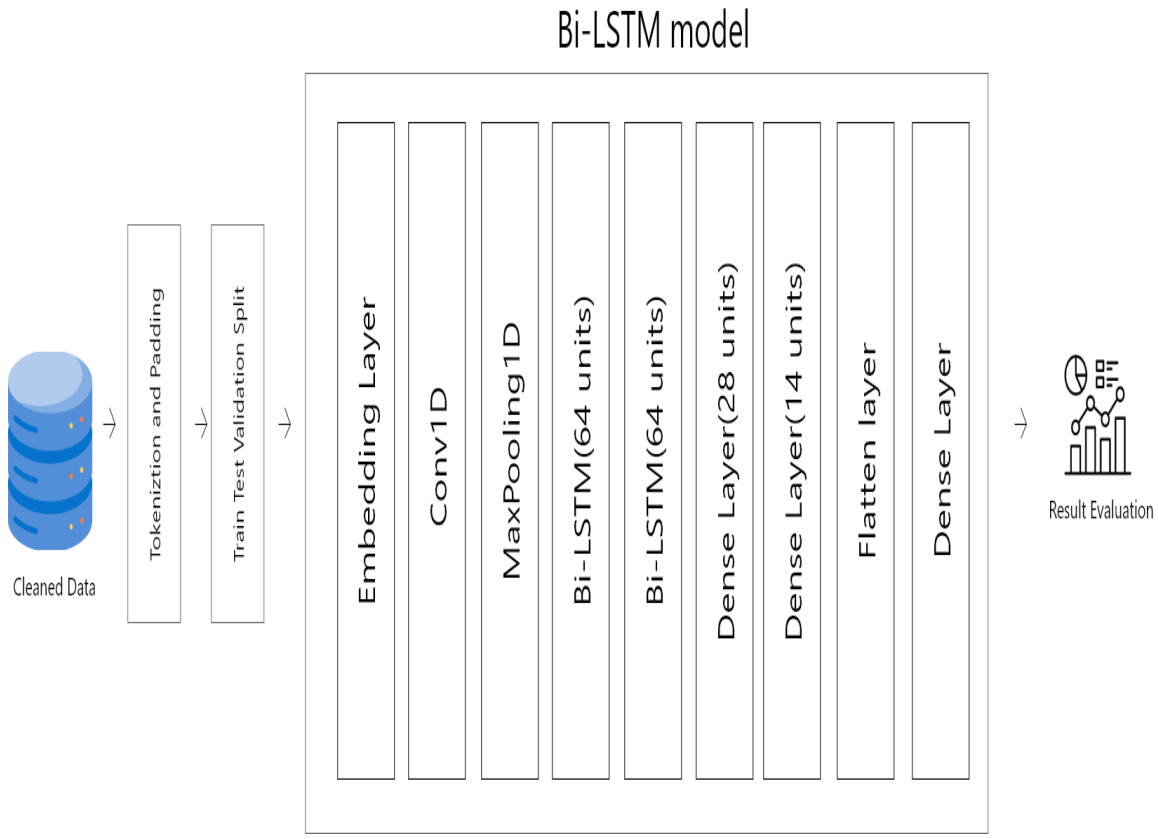

Figure 3.4.3: Proposed Methodology for Deep Learning model

And in case of Bi-LSTM model we are doing the same task upto text preprocessing. But here we are not using tfidf vectorization. Rather we are tokenizing the texts and making it a sequence of integer numbers. These numbers will represent individual values of the words used in the texts. And after preprocessing we are splitting the dataset into three parts: train, test and validation data. The train and validation data will be used to train and validate the model. And the test data will be used evaluate the accuracy and confusion matrix of the model. Here we are evaluating the models performance using training and validation accuracy, training and validation loss and confusion matrix for both binary and multiclass classification. The Bi-LSTM model is consist of one embedding layer, one conv1D layer, one maxpooling1D layer, two Bi-LSTM layer with 64 neurons, two dense layer with 28 and 14 units , one flatten layer and finally anther dense layer to provide the final output. This is a sequential model. Which means that the layers are stacked on top of each other one by one and one layers output is the following layers input.

Text Preprocessing: Here we are taking a text as input and iterating through the whole text to look for the characters which are not present in the range of Bangla Unicode characters value. If we are able to find such characters we are deleting that characters. In such way characters like emoji, numeric values, punctuation marks, English alphabets, web urls etc. are being filtered out of the Bangla texts.And we are also removing the stopwords as well. The stopwords list is taken from a kaggle dataset. Here is how the cleaned documents would look like before and after text preprocessing:

TABLE 3.4.1: Dataset before and after cleaning

Raw Texts	Cleaned Texts
আপনার জন্ম প্রক্রিয়ার সময় আপনার মায়ের ভিতর কি ...	জন্ম প্রক্রিয়ার সময় মায়ের ভিতর বাবা হুমায়ুন কব...
ওই হালার পুত এখন কি মদ খাওয়ার সময় রাতের বেলা...	হালার পুত মদ খাওয়ার রাতের বেলা মদ খাই দিনের ব...
আল্লাহ আপনাকে সুস্থ রাখুক। ভাইরাস থেকে দেশের ম...	আল্লাহ আপনাকে সুস্থ রাখুক ভাইরাস দেশের রাব্বুল...
আল্লাহ আপনার উপর রহমত বর্ষণ করুন, আমিন।	আল্লাহ রহমত বর্ষণ করুন আমিন

3.5 Hardware and Software Requirements

As we are using both machine learning and deep learning approach we have some hardware and software requirements based on the different approaches. Here are the hardware and software requirements of our proposed methodologies:

Hardware Requirements:

- **CPU:** A modern multi-core processor (e.g., Intel Core i5 or higher, AMD Ryzen 5 or higher) is sufficient for training and using the model on small to medium-sized datasets. However, for larger datasets or more complex models, a more powerful

CPU or GPU may be beneficial..

- **Graphics Processing Unit (GPU):** A GPU is essential for accelerating the training of the Bi-LSTM model, significantly reducing the time required for model development.
- **RAM:** Sufficient RAM to hold our dataset in memory during training and prediction. The exact amount of RAM required depends on the size of our dataset and the memory requirements of the models being used. As a general rule, having at least 8 GB of RAM is recommended, but larger datasets or more complex models may require more RAM.

Software Requirements:

- **Python:** We need Python installed on our system, as TensorFlow and Keras are Python libraries.We have to ensure that we have Python version 3.x installed.
- **TensorFlow:** We have to install TensorFlow, the deep learning framework, which provides an interface to work with neural networks.
- **Keras:** This will come with TensorFlow. So we won't need to install it separately
- **Streamlit:** Streamlit is an open-source Python library that allows developers to create interactive web applications for machine learning and data science projects quickly.

Model Training and Evaluation:

- **Metrics for evaluation:** Definition and utilization of appropriate evaluation metrics, such as confusion matrix, accuracy, precision, recall, and F1 score, to assess the performance of the models in classification tasks.

Ethical Considerations:

- **User privacy:** For the purpose of user privacy we haven't use the user information of those who have commented on the social media posts.

By meeting these implementation requirements, we would be able to fluently finalize our research.

3.6 Project Management and Financial Analysis

There are many items we can take into consideration for doing the study on cyberbullying detection. These would require proper management and we would also need some financial help to conduct this study. Here it is depicted shortly:

TABLE 3.6.1. FINANCES FOR THE STUDY

SL	Components	Estimated Cost (BDT)
1	Hardware/Infrastructure	95,000-110,000
2	Software and Tools	2,000-3,000
3	Data Collection and Processing	2,500-3,000
4.	Documentation and Report Writing	500-1,000
5	Miscellaneous (e.g., licenses, tools)	500-1,000
	Total	100,500-118,000

These are the market prices of the instruments given based on February 2024. Prices may vary depending on the market.

3.7 Summary

In this section, the research focuses on leveraging advanced artificial intelligence techniques to analyze Bengali social media comments for cyberbullying detection. Various machine learning and deep learning models are explored, including Extra Trees Classifier, Logistic Regression, Naïve Bayes, Decision Tree, Stacking Classifier, and Bi-LSTM. Data collection involved manually labeling comments from public Facebook pages and utilizing

a Kaggle dataset, totaling 11,064 comments across different categories like troll, political, sexual, religious, and threat. The study employs both binary and multiclass classification approaches, emphasizing preprocessing steps such as stop word removal and character filtering. Hardware requirements include a modern CPU and GPU for efficient model training, while software requirements encompass Python, TensorFlow, Keras, and Streamlit for application development. Ethical considerations ensure user privacy protection, underscoring the ethical framework guiding the research. Financial estimates outline costs for hardware, software, data processing, documentation, and miscellaneous expenses, essential for managing the project effectively.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

We have implemented our system using Streamlit. Streamlit is an open-source Python library that allows developers to create interactive web applications for machine learning and data science projects quickly. It simplifies the process of building and sharing these applications by providing a straightforward interface where developers can write Python scripts to visualize data, create forms, and incorporate machine learning models seamlessly.

With Streamlit, developers can transform their Python scripts into web apps with minimal effort, using intuitive commands to add widgets, charts, and text to the user interface. This makes it accessible even to those without extensive web development experience, focusing instead on the functionality and interactivity of the application.

4.2 Prototype Design

We have used our best-suited model for our web application prototype. Then we have used that model and feed it to our Streamlit backend. And Streamlit has automatically created a web application interface for us. Using this application interface we would be able to input Bengali comments and check whether the text is bully or non-bully comment or not. Here is the design of our prototype:

Figure 4.2.1: Prototype design using Streamlit

4.3 System Testing

Here is a depiction of our prototype system :

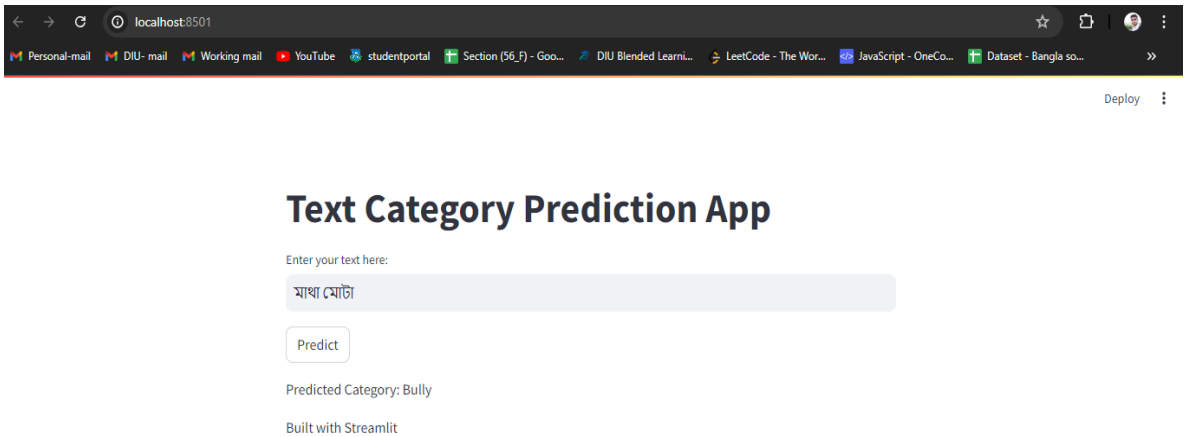

Figure 4.2.2: Streamlit application interface predicting bully comments

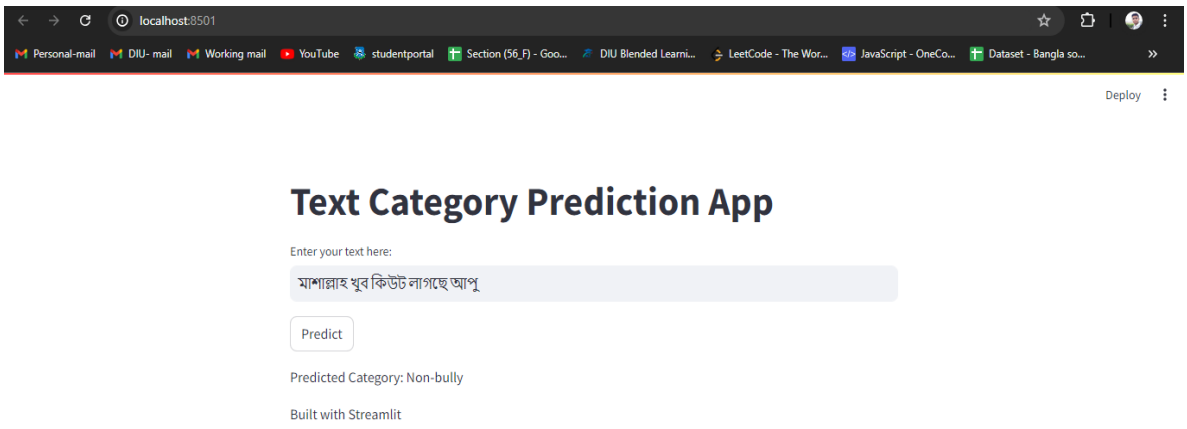

Figure 4.2.3: Streamlit application interface predicting non-bully comments

Here We can see that Streamlit has generated an automated web application interface for us. On Figure 4.2.2 we can see that a comment like "মাথা মোটা" is categorized as "Bully" and on Figure 4.2.3 the comment "মাশাল্লাহ খুব কিউট লাগছে আপু" is categorized as Non-bully comment.

4.4 Summary

The implementation of our system using Streamlit has streamlined the process of creating interactive web applications for analyzing Bengali social media comments. Streamlit, an open-source Python library, facilitated rapid development by enabling us to convert our best-suited machine learning model into a user-friendly web interface effortlessly. The prototype design Figure 4.2.1 showcases how Streamlit integrates with our model, allowing users to input Bengali comments and promptly classifying them as bully or non-bully Figures 4.2.2 and 4.2.3. This automated web application interface not only enhances accessibility for users but also demonstrates the effectiveness of our model in real-time comment analysis.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

As we have already discussed we have three different kinds of approach to detect Bangla cyberbullying instances. We have two datasets. One is our personal dataset and one is taken from kaggle. We have run our experiments on both the dataset individually. And also we have combined the datasets and ran our experiments on more time. For evaluation of the models we have used some metrics like accuracy, precision, recall, F1-score and confusion matrix. Here are some short description about these :

Accuracy: This metric measures the proportion of correctly classified instances among the total instances. It's calculated as the ratio of the number of correct predictions to the total number of predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: Precision measures the proportion of true positive predictions (correctly identified instances of a class) to the total number of instances predicted as belonging to that class. It's calculated as:

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall : Recall measures the proportion of true positive predictions to the total number of actual instances of a class. It's calculated as :

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-score: F1-score is the harmonic mean of precision and recall. It provides a single score that balances both precision and recall. It's calculated as 2 * (precision * recall) / (precision + recall).

Where:

- TP (True Positives) is the number of instances correctly predicted as positive.
- TN (True Negatives) is the number of instances correctly predicted as negative.
- FP (False Positives) is the number of instances incorrectly predicted as positive.
- FN (False Negatives) is the number of instances incorrectly predicted as negative.

Confusion Matrix: A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. It presents a summary of the performance of the classification algorithm by displaying the counts of true positive, true negative, false positive, and false negative predictions.

5.2 Experimental Results

Binary class classification using Machine Learning Model:

TABLE 5.2.1: Binary classification result on personal dataset using ml model

Algorithm	Accuracy	Precision	Recall	F1-score
Extra Trees Classifier	69.42%	58.68%	40.69%	48.05%
Random Forest	69.32%	57.24%	46.41%	51.27%
Naïve Bayes	69.12%	55.49%	56.44%	55.97%
Decision Tree	42.92%	37.56%	97.42%	54.22%

As shown in Table 5.2.1 , 5.2.2 and 5.2.3 the maximum accuracy achieved using our machine learning model on personal dataset in 69.42% using ETC. And in case of mixed dataset the highest accuracy achieved is 74.63% using ETC and on the kaggle dataset 78.83% using DT accuracy was achieved.

TABLE 5.2.2: Binary classification result on mixed dataset using ml model

Algorithm	Accuracy	Precision	Recall	F1-score
Extra Trees Classifier	74.63%	52.88%	39.68%	45.34%
Random Forest	73.05%	50.69%	44.02%	47.12%
Naïve Bayes	72.24%	48.27%	55.80%	51.76%
Decision Tree	35.84%	28.74%	96.18%	44.25%

TABLE 5.2.3: Binary classification result on Kaggle dataset using ml model

Algorithm	Accuracy	Precision	Recall	F1-score
Extra Trees Classifier	75.85%	43.39%	52.46%	47.5%
Random Forest	73.89%	40.77%	56.14%	47.24%
Naïve Bayes	75.76%	44.08%	61.06%	51.20%
Decision Tree	78.83%	48.14%	21.31%	29.55%

Binary class classification using Stacking classifier Model:

TABLE 5.2.4: Binary classification result on all the datasets using stacking classifier model

Dataset	Accuracy	Precision	Recall	F1-score
Personal dataset	69.62%	56.60%	54.13%	51.87%
Mixed dataset	74.71%	55.20%	50.09%	45.84%
Kaggle dataset	74.65%	39.11%	45.50%	54.38%

Here using stacking classifier model the highest accuracy was 74.71% achieved in mixed dataset. Rest of the achieved results are shown in Table 5.2.4. And the confusion matrix on the three dataset for stacking classifier model are shown in Figure 5.2.1,5.2.2 & 5.2.3. In the confusion matrix 0 stands for ‘bully’ and 1 stands for ‘non-bully’ labeled comments.

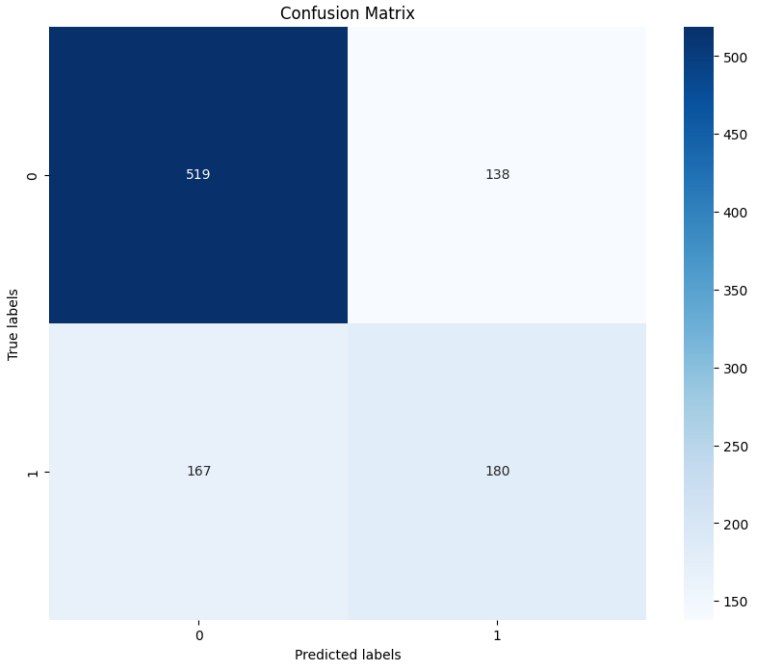

Figure 5.2.1: Confusion matrix using stacking classifier on personal dataset

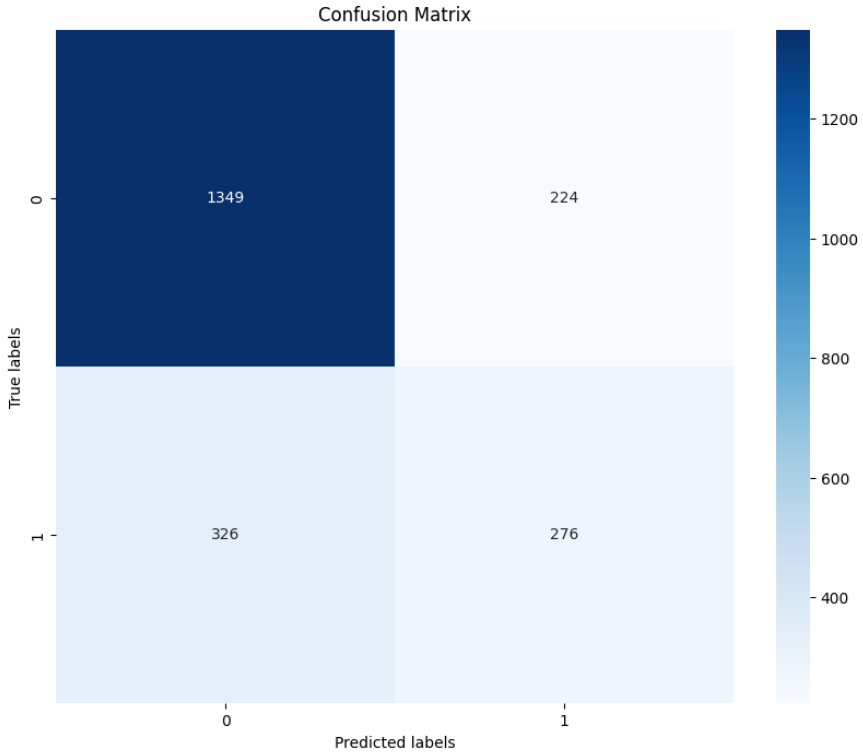

Figure 5.2.2: Confusion matrix using stacking classifier on mixed dataset

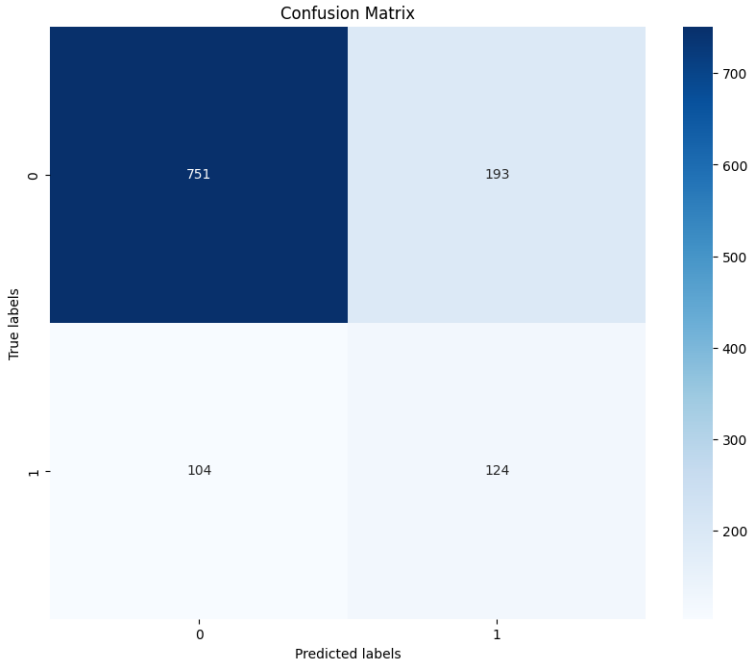

Figure 5.2.3: Confusion matrix using stacking classifier on Kaggle dataset

Binary class classification using Bi-LSTM Model:

TABLE 5.2.5: Binary classification result on all the datasets using Bi-LSTM model

Dataset	Accuracy	Precision	Recall	F1-score
Personal dataset	76.08%	63.63%	62.02%	62.82%
Mixed dataset	81.71%	70.18%	54.06%	61.07%
Kaggle dataset	88.64%	64.65%	67.36%	65.97%

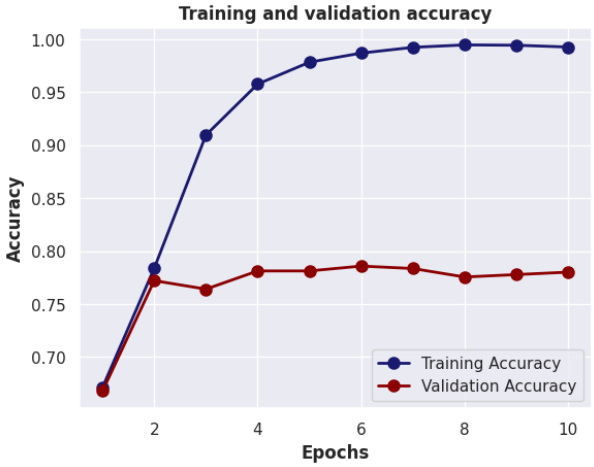

Figure 5.2.4: Training and validation accuracy on personal dataset using Bi-LSTM

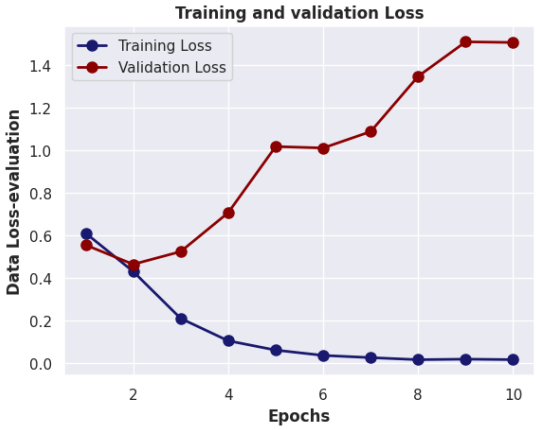

Figure 5.2.5: Training and validation loss on personal dataset using Bi-LSTM

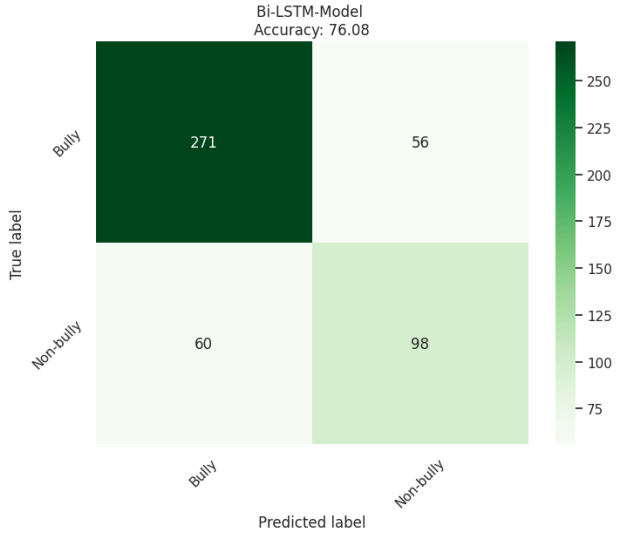

Figure 5.2.6: Confusion matrix of personal dataset using Bi-LSTM in binary classification

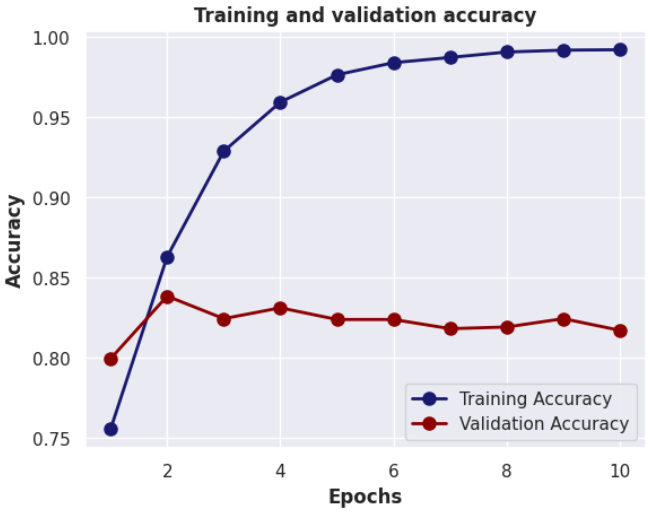

Figure 5.2.7: Training and validation accuracy on mixed dataset using Bi-LSTM

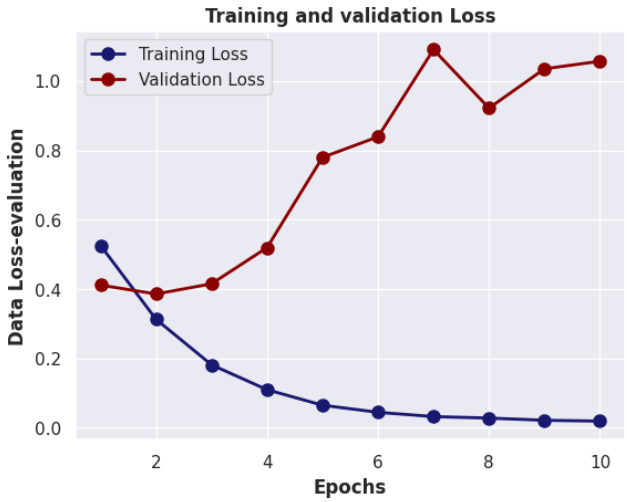

Figure 5.2.8: Training and validation loss on mixed dataset using Bi-LSTM

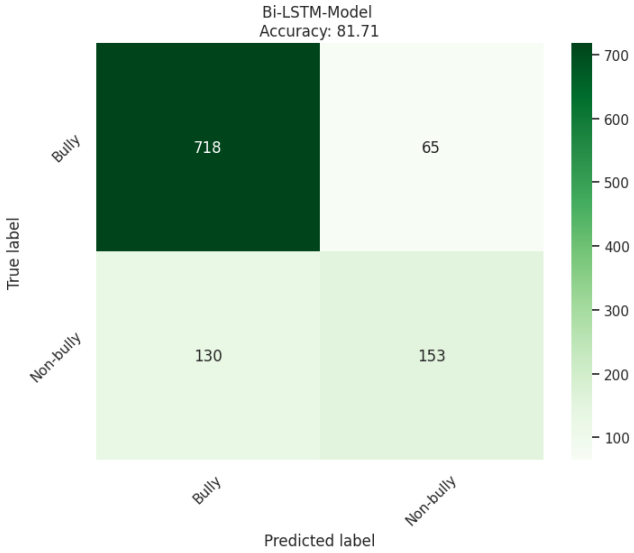

Figure 5.2.9: Confusion matrix of mixed dataset using Bi-LSTM in binary classification

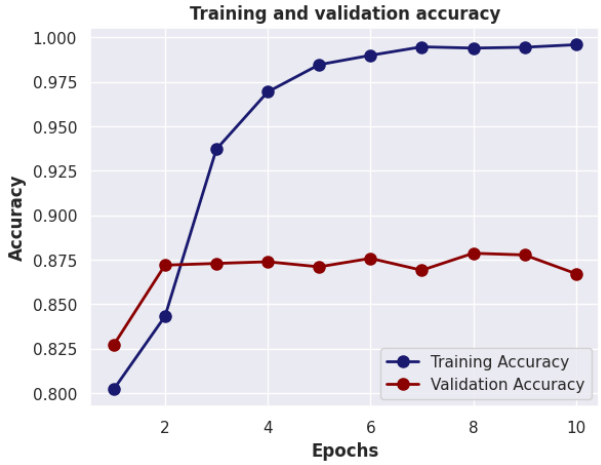

Figure 5.2.10: Training and validation accuracy on Kaggle dataset using Bi-LSTM

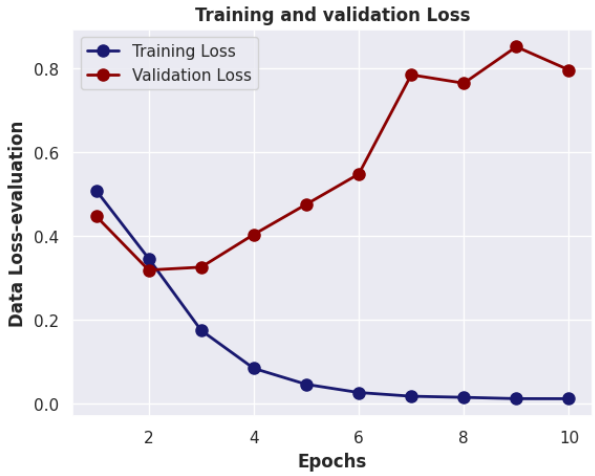

Figure 5.2.11: Training and validation loss on Kaggle dataset using Bi-LSTM

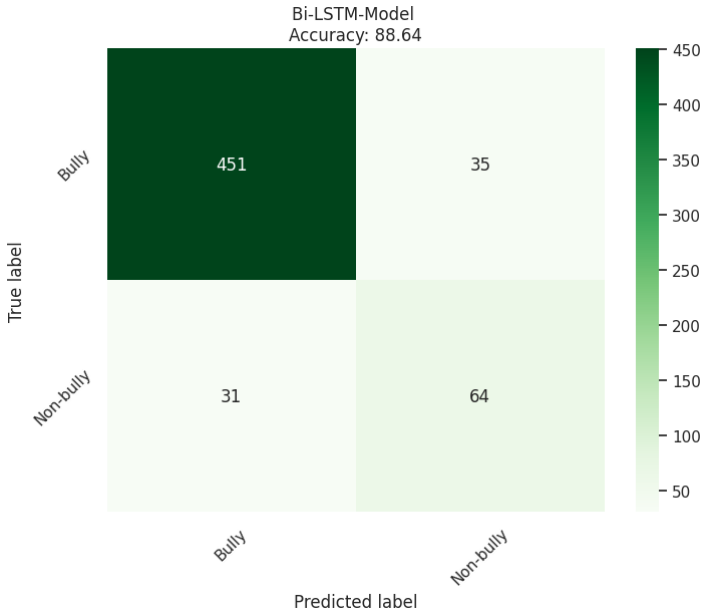

Figure 5.2.12: Confusion matrix of Kaggle dataset using Bi-LSTM in binary classification

We have achieved the best results so far on binary classification using our deep learning model. The highest accuracy was achieved on the kaggle dataset. 88.64% accuracy was achieved from this model. Table 5.2.5 shows the full results on the three dataset for our deep learning model for binary classification. And Figure 5.2.4 , 5.2.7 & 5.2.10 shows the training and validation accuracy curve for the three dataset. Figure 5.2.5, 5.2.8 & 5.2.11 shows the training and validation loss curve of the three dataset. And Figure 5.2.6 , 5.2.9 and 5.2.12 shows us the confusion matrix for our model.

Multiclass classification using Machine Learning Model:

TABLE 5.2.6: Multiclass classification result on personal dataset using ml model

Algorithm	Accuracy
Extra Trees Classifier	46.51%
Random Forest	45.11%
Naïve Bayes	45.71%
Decision Tree	18.32%

TABLE 5.2.7: Multiclass classification result on mixed dataset using ml model

Algorithm	Accuracy
Extra Trees Classifier	41.45%
Random Forest	42.14%
Naïve Bayes	43.10%
Decision Tree	29.27%

TABLE 5.2.8: Multiclass classification result on Kaggle dataset using ml model

Algorithm	Accuracy
Extra Trees Classifier	45.22%
Random Forest	45.30%
Naïve Bayes	47.01%
Decision Tree	29.52%

Table 5.2.6 , 5.2.7 & 5.2.8 shows us the achieved accuracies using machine learning models for multiclass classification. The highest accuracy was recorded for the kaggle dataset using NB algorithm which was 47.01%.

Multiclass classification using Stacking Classifier Model:

TABLE 5.2.9: Multi class classification result on all the datasets using stacking classifier model

Dataset	Accuracy
Personal dataset	47.50%
Mixed dataset	42.50%
Kaggle dataset	46.58%

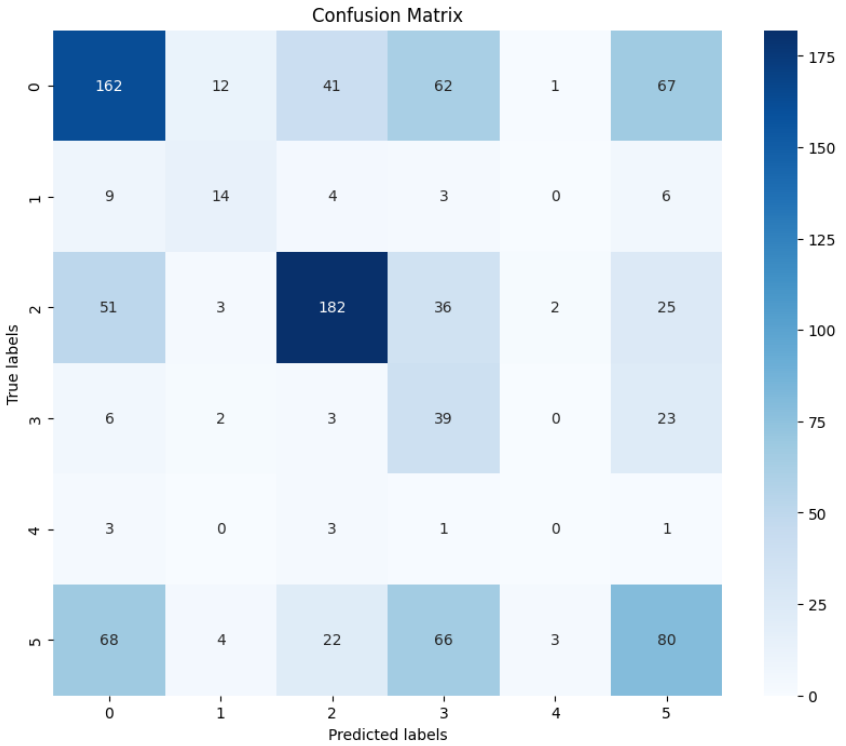

Figure 5.2.13: Confusion matrix of personal dataset using Stacking classifier in multi class classification

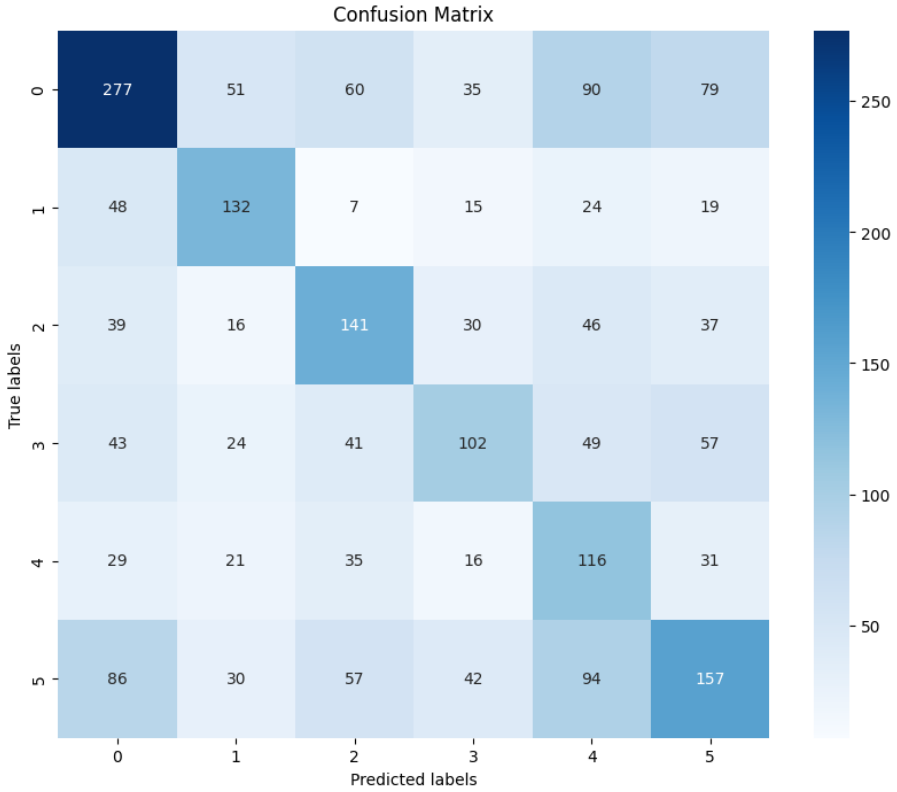

Figure 5.2.14: Confusion matrix of mixed dataset using Stacking classifier in multi class classification

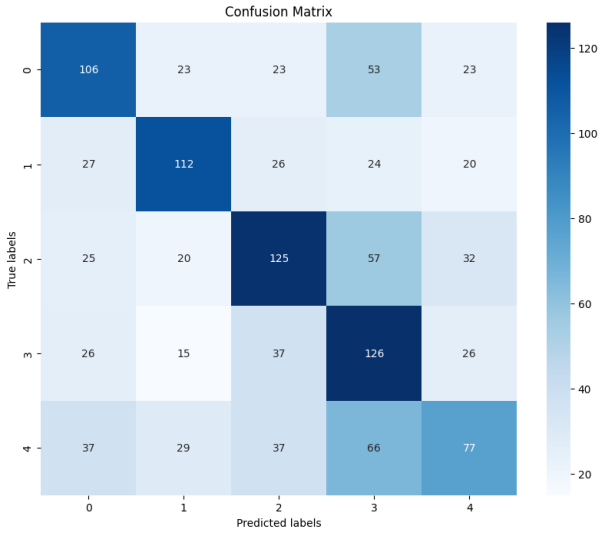

Figure 5.2.15: Confusion matrix of Kaggle dataset using Stacking classifier in multi class classification

For the stacking classifier model the highest accuracy achieved was 47.50% for our personal dataset. Table 5.2.9 shows us the full result for our stacking classifier model for our three individual dataset. Figure 5.2.13,5.2.14 & 5.2.15 shows us the confusion matrix for personal dataset, mixed dataset and kaggle dataset. Here 0,1,2,3,4 and 5 stands for 'Non-bully', 'political', 'religious', 'sexual' ,'threat' and 'troll' comments accordingly. And as the kaggle dataset has a class less than the other two dataset, for the Figure 4.2.15 the 0,1,2,3,4 stands for 'Non-bully', 'political', 'sexual' ,'threat' and 'troll comments.

Multiclass classification using Bi-LSTM Model:

For our deep learning model we have achieved the highest accuracy of 64.89% using our kaggle dataset. This model worked better in comparison to other models. Table 5.2.10 shows the results of our deep learning model for our three datasets. Figure 5.2.16 , 5.2.19 & 5.2.22 shows us the training and validation accuracy curve. Figure 5.2.17, 5.2.20 & 5.2.23 shows us the training and validation loss curve and Figure 5.2.18 , 5.2.21 & 5.2.24 shows the confusion matrix for the three datasets

TABLE 5.2.10: Multi class classification result on all the datasets using Deep learning model

Dataset	**Accuracy**
Personal dataset	60.00%
Mixed dataset	61.91%
Kaggle dataset	64.89%

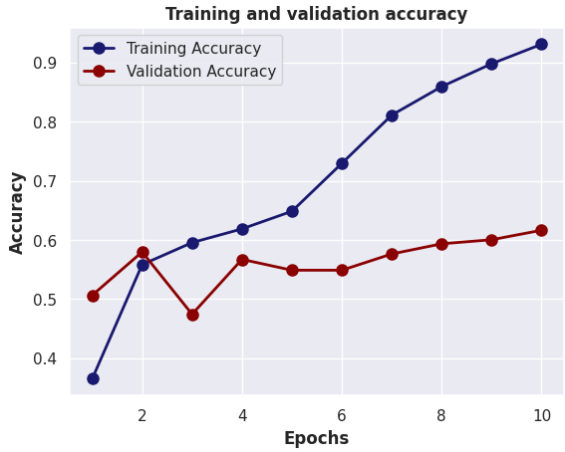

Figure 5.2.16: Training and validation accuracy on personal dataset using Bi-LSTM for multi class classification

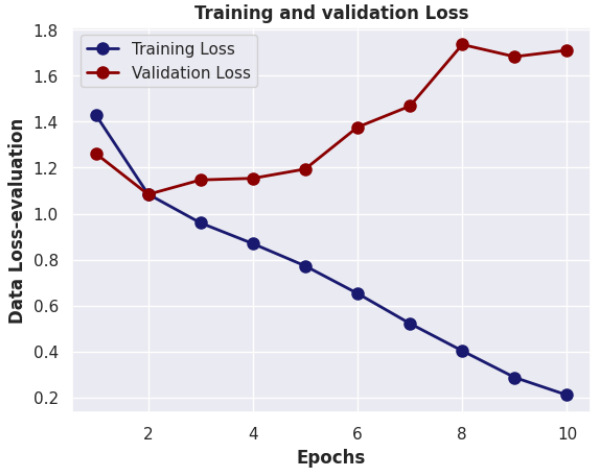

Figure 5.2.17: Training and validation loss on personal dataset using Bi-LSTM for multi class classification

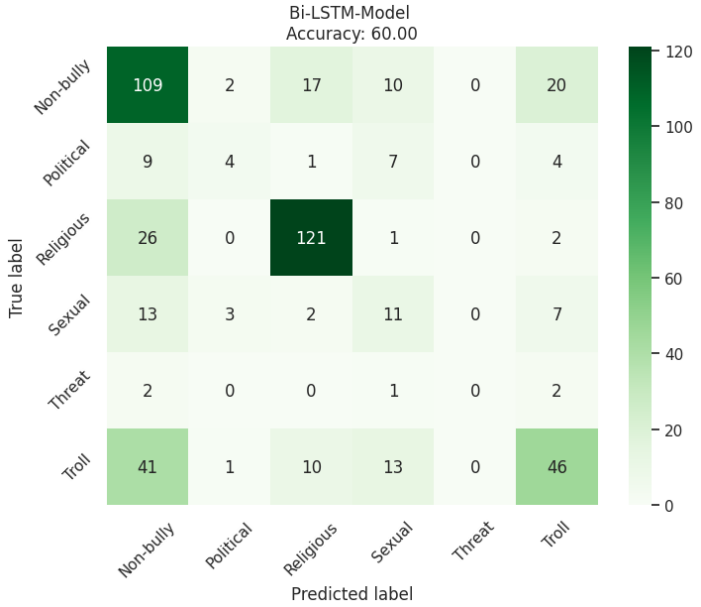

Figure 5.2.18: Confusion matrix of personal dataset using Bi-LSTM in multiclass classification

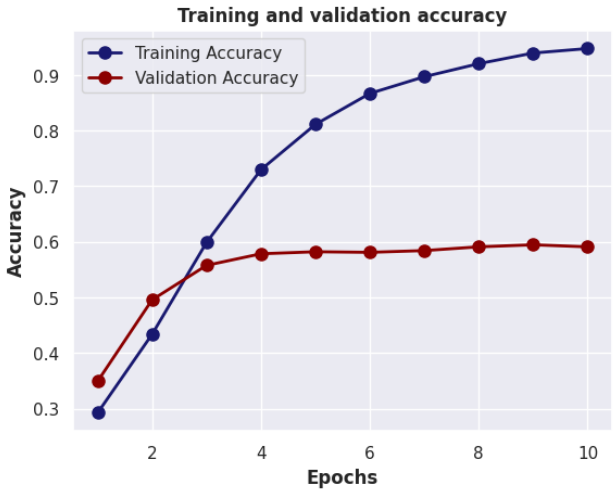

Figure 5.2.19: Training and validation accuracy on mixed dataset using Bi-LSTM for multi class classification

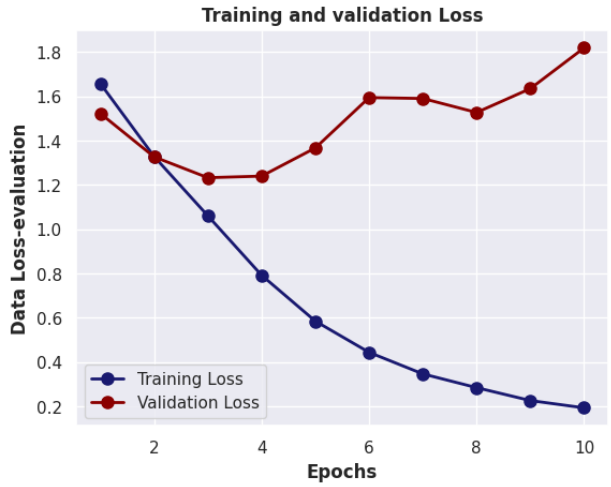

Figure 5.2.20: Training and validation loss on mixed dataset using Bi-LSTM for multi class classification

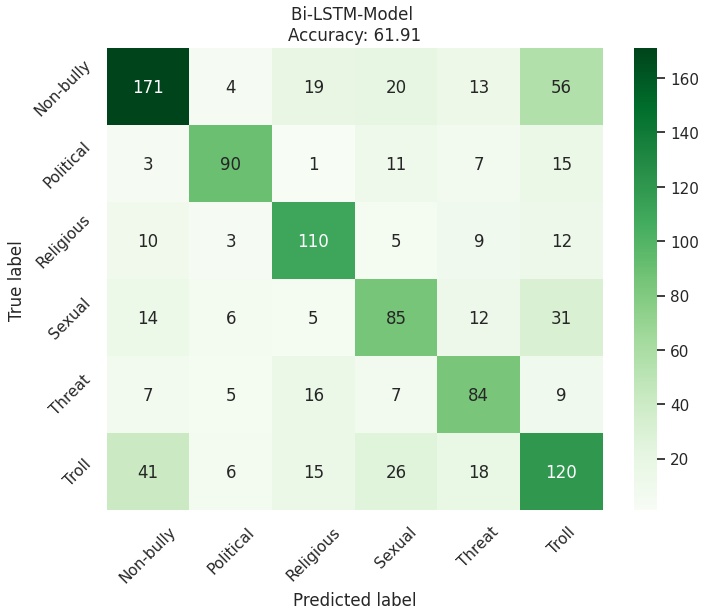

Figure 5.2.21: Confusion matrix of mixed dataset using Bi-LSTM in multiclass classification

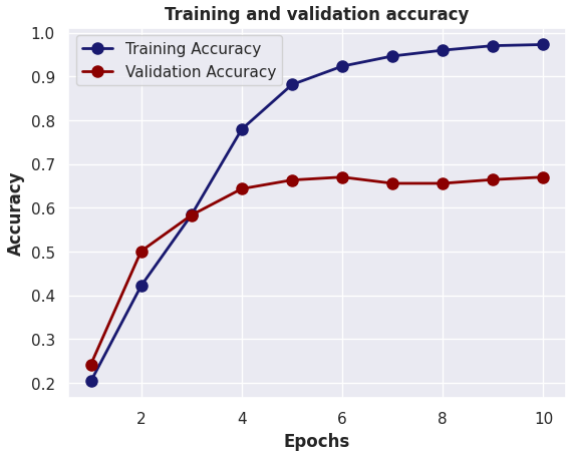

Figure 5.2.22: Training and validation accuracy on Kaggle dataset using Bi-LSTM for multi class classification

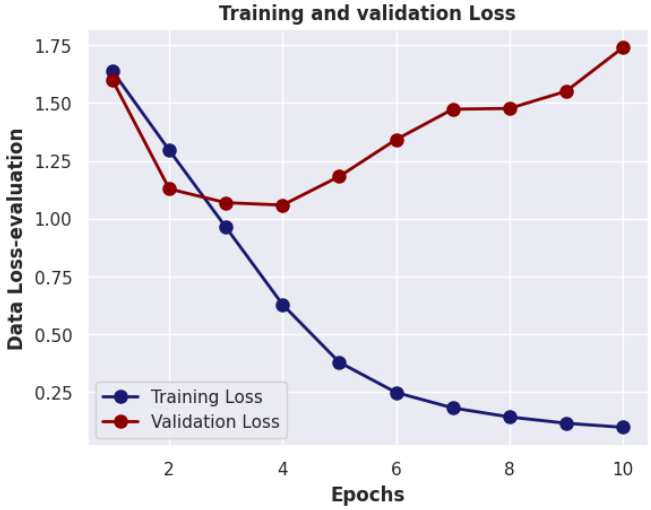

Figure 5.2.23: Training and validation loss on Kaggle dataset using Bi-LSTM for multi class classification

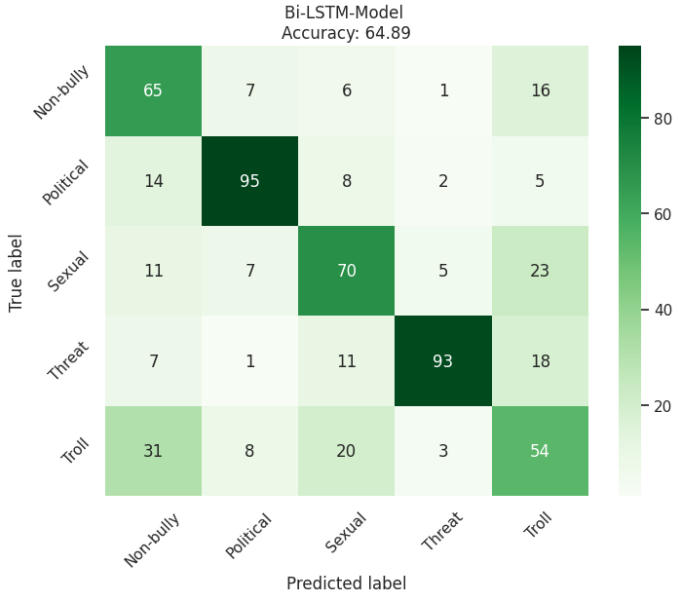

Figure 5.2.24: Confusion matrix of kaggle dataset using Bi-LSTM in multiclass classification

5.3 Performance Analysis

As we can see that among the three approaches applied in our study the deep learning approach performed better in both binary class classification and multiclass classification. In our literature review we have also observed that deep learning models work better in natural language processing in comparison to machine learning algorithms. We have achieved the highest results on the kaggle dataset as the comments there are evenly distributed among the different classes. But we have also gained satisfactory results on our personal dataset and also in case of mixed dataset using our deep learning model. As the results found in case of multiclass classification are significantly lower in comparison to binary class classification we will be using the model trained for binary class classification on the kaggle dataset for our cyberbullying detection prototype app. Another reason for using this model is that it achieved the highest accuracy among all the other datasets. For the kaggle dataset the accuracy for binary class classification was almost 89%.

5.4 Summary

In our study on Bangla cyberbullying detection, we explored three distinct approaches using machine learning and deep learning models across different datasets. For binary classification, machine learning models like Extra Trees Classifier and Naïve Bayes yielded accuracies up to 78.83% on the Kaggle dataset, while our best-performing deep learning model, Bi-LSTM, achieved 88.64% accuracy on the same dataset. This deep learning approach consistently outperformed traditional machine learning methods, showcasing its effectiveness in natural language processing tasks. However, multiclass classification results were notably lower across all models and datasets. The Bi-LSTM model stood out with 64.89% accuracy on the Kaggle dataset for multiclass classification. Given these findings, we opted to use the Bi-LSTM model trained on the Kaggle dataset for our cyberbullying detection prototype due to its superior performance and robust accuracy of nearly 89% in binary classification. This choice ensures our application leverages the most effective model for accurately identifying bully and non-bully comments in Bengali social media contexts.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

Research focused on cyberbullying detection has profoundly influenced individuals' lives by providing effective tools to identify and combat online harassment. By developing advanced algorithms and technologies, researchers have enabled platforms and users to detect bullying behavior swiftly, thereby mitigating its detrimental impact on victims' mental health and social well-being. Early detection allows for prompt intervention and support, fostering safer online environments and promoting positive digital interactions. Moreover, such research has raised awareness about the prevalence and seriousness of cyberbullying, encouraging proactive measures from policymakers, educators, and technology companies to address this pervasive issue. Overall, these efforts contribute to a more secure and supportive online community, empowering individuals to navigate digital spaces with confidence and resilience.

6.2 Impact on Society & Environment

The introduction of a cyberbullying detection system holds promise for society on multiple fronts. Firstly, it offers a means to mitigate the often devastating effects of cyberbullying. By swiftly identifying and addressing instances of online harassment, the system can help reduce the psychological distress and harm suffered by victims. Moreover, in educational settings, where cyberbullying among students is prevalent, such a system could empower teachers and administrators to intervene proactively, fostering safer and more supportive school environments. Beyond immediate effects, the system's presence could contribute to broader societal well-being by promoting responsible digital citizenship and encouraging respectful online interactions. However, it's crucial to navigate potential legal and privacy implications carefully, ensuring that the system operates ethically and respects users' rights. Ultimately, by striking a balance between effective detection and privacy protection, a cyberbullying detection system has the potential to significantly enhance online safety and mental health outcomes, thereby fostering a healthier digital society.

The deployment of a cyberbullying detection system can also have environmental implications, albeit indirectly. By fostering safer online environments and promoting positive digital interactions, such a system could contribute to a reduction in stress-related behaviors associated with cyberbullying. Research suggests that exposure to online harassment can lead to increased stress levels and negative emotional responses, which may in turn impact individuals' offline behaviors and well-being. By mitigating the harmful effects of cyberbullying, the system could potentially alleviate some of the environmental stressors associated with mental health issues, such as anxiety and depression. Additionally, by creating safer digital spaces, the system may encourage individuals to engage more positively and productively online, leading to a healthier digital ecosystem overall. However, the specific environmental impact of a cyberbullying detection system would likely be secondary to its primary social and psychological benefits. Nonetheless, by fostering a more supportive online environment, the system could indirectly contribute to improved mental well-being, potentially leading to broader positive effects on society and the environment alike.

6.3 Ethical Aspects

The implementation of a cyberbullying detection system raises several ethical considerations that warrant careful examination. Firstly, there is the issue of privacy and data protection. While the system's purpose is to identify and address instances of cyberbullying, it requires access to potentially sensitive personal data, such as online communications and interactions. Ensuring that the system operates within the bounds of privacy laws and respects individuals' rights to privacy is paramount.

Another ethical concern is the potential for false positives and false negatives. A detection system may inadvertently flag innocent interactions as cyberbullying or fail to detect harmful behavior, leading to unjust consequences for individuals. Striking a balance between the system's sensitivity and specificity to minimize these errors is essential to maintain fairness and accuracy.

Additionally, there is the question of accountability and transparency. It is crucial for the developers and operators of the system to be transparent about its functioning, including the algorithms used for detection and the criteria for labeling behavior as cyberbullying. Moreover, mechanisms should be in place to hold responsible parties accountable for any

misuse or unintended consequences of the system.

Furthermore, there is the risk of over-reliance on technology to address complex social issues like cyberbullying. While a detection system can be a valuable tool, it should not be viewed as a panacea. It is essential to complement technological solutions with educational initiatives, mental health support services, and community interventions to address the root causes of cyberbullying comprehensively.

Lastly, there is the ethical imperative to consider the broader societal implications of deploying such a system. This includes potential stigmatization of individuals flagged by the system, the impact on freedom of expression and online discourse, and the distribution of resources towards technological solutions at the expense of other pressing social issues.

In summary, while a cyberbullying detection system holds promise for addressing online harassment, it is crucial to navigate the ethical landscape thoughtfully, ensuring that the system upholds principles of privacy, fairness, transparency, accountability, and societal well-being.

6.4 Sustainability Plan

Creating a sustainable cyberbullying detection system entails a multifaceted approach that balances technological innovation with ethical considerations and user empowerment. Continuous improvement lies at the core, with a commitment to ongoing research and development to enhance accuracy and effectiveness while adapting to evolving cyberbullying tactics. Privacy protection is paramount, demanding robust measures to safeguard user data and uphold ethical guidelines that prioritize transparency, fairness, and accountability. User education and awareness campaigns play a crucial role in empowering individuals to recognize and report cyberbullying incidents, fostering a culture of responsible online behavior. Collaboration with stakeholders such as schools, community organizations, and mental health professionals enhances the system's impact and reach, while integration with support services ensures seamless access to assistance for those affected. Sustainability models, including diverse funding sources and scalable infrastructure, ensure the system's long-term viability. Regular evaluation and adaptation to changing threats further reinforce its effectiveness, enabling it to make a lasting impact in promoting online safety and well-being.

6.5 Summary

The comprehensive approach to cyberbullying detection outlined covers its profound impact on individuals' lives, society, and ethical considerations. Research has enabled effective tools to mitigate online harassment, promoting safer digital interactions and raising awareness. Societal benefits include fostering supportive environments in schools and promoting responsible digital citizenship. Ethical aspects emphasize privacy protection, fairness in detection, transparency, and accountability. Sustainability planning involves continuous innovation, user empowerment through education, collaboration with stakeholders, and robust funding models. Overall, the system aims to enhance online safety and well-being while navigating complex ethical challenges and ensuring long-term effectiveness and societal benefit.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

As we have discussed, cyberbullying poses a significant threat to our social fabric and has profound implications for mental health. Therefore, automating the detection of cyberbullying instances on social media platforms is imperative. We have developed a system capable of automatically identifying instances of cyberbullying in Bengali language content. Throughout our study, we encountered various types of cyberbullying comments on social media, including political, religious, troll, and threatening remarks. Our dataset initially suffered from imbalance, which we addressed by incorporating a publicly available dataset. Our observations indicate that our deep learning model outperformed traditional machine learning models in this context. Therefore, we recommend exploring the realm of deep learning within artificial intelligence for natural language processing tasks such as cyberbullying detection. Our findings strongly suggest that employing deep learning models is preferable for effective cyberbullying detection.

7.2 Further Suggested Works

Based on our findings, it is evident that further exploration is necessary in the field of automatic cyberbullying detection. Currently, our study focuses exclusively on detecting cyberbullying in the Bangla language. However, expanding our research to encompass other languages would significantly enhance our efforts. Developing a system capable of adapting to all languages worldwide would greatly augment its utility and effectiveness.Since our study does not encompass all areas of artificial intelligence, exploring additional techniques is essential for effectively addressing the problem of cyberbullying detection.

7.3 Limitations

The most important limitation of our research is that our study is only limited to Bengali language only and another significant limitation is that the performance of our our model is also low in comparison to some other previous researches.

REFERENCES

[1] X. Zhang *et al*., "Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network," *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, USA, 2016, pp. 740-745, doi: 10.1109/ICMLA.2016.0132.

[2]. J. Hani, N. Mohamed, A. E. Mostafa, Z. Emad, E. Amer, and A. Mohammed, "Social Media Cyberbullying Detection using Machine Learning," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, Jan. 2019, doi: 10.14569/ijacsa.2019.0100587.

[3] M. Dadvar and F. de Jong, "Cyberbullying detection," Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion, 2012, doi: https://doi.org/10.1145/2187980.2187995.

[4]. M. Islam, Md. A. Uddin, L. Islam, A. Akter, S. Sharmin, and U. K. Acharjee, "Cyberbullying detection on social networks using machine learning approaches," *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Dec. 2020, doi: 10.1109/csde50874.2020.9411601.

[5] N. A. Azeez, S. O. Idiakose, C. J. Onyema, and C. Van Der Vyver, "Cyberbullying detection in social Networks: Artificial Intelligence approach," *Journal of Cyber Security and Mobility*, Jun. 2021, doi: 10.13052/jcsm2245-1439.1046.

[6]. Y. Fang, S. Yang, B. Zhao, and C. Huang, "Cyberbullying Detection in Social Networks Using Bi-GRU with Self-Attention Mechanism," *Information*, vol. 12, no. 4, p. 171, Apr. 2021, doi: 10.3390/info12040171.

[7]. R. Ghosh, S. Nowal, and G. Manju, "Social Media Cyberbullying Detection using Machine Learning in Bengali Language," *International Journal of Engineering Research and Technology*, vol. 10, no. 5, May 2021, [Online]. Available: https://www.ijert.org/research/social-media-cyberbullying-detection-using-machine-learning-in-bengali-language-IJERTV10IS050083.pdf

[8]. G. Hussain, T. A. Mahmud, and W. Akthar, "An Approach to Detect Abusive Bangla Text," *An Approach to Detect Abusive Bangla Text*, Dec. 2018, doi: 10.1109/ciet.2018.8660863.

[9]. M. Alotaibi, B. Alotaibi, and A. Razaque, "A multichannel deep learning framework for cyberbullying detection on social media," *Electronics*, vol. 10, no. 21, p. 2664, Oct. 2021, doi: 10.3390/electronics10212664.

[10] M. Di Capua, E. Di Nardo and A. Petrosino, "Unsupervised cyber bullying detection in social networks," *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, 2016, pp. 432-437, doi: 10.1109/ICPR.2016.7899672.

[11] Md. T. Ahmed, A. S. Urmi, M. Rahman, A. Z. M. T. Islam, D. Das, and Md. G. Rashed, "Cyberbullying Detection Based on Hybrid Ensemble Method using Deep Learning Technique in Bangla Dataset," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, Jan. 2023, doi: 10.14569/ijacsa.2023.0140958.

[12]. L. Cheng, R. Guo, Y. N. Silva, D. L. Hall, and H. Liu, "Hierarchical attention networks for cyberbullying detection on the Instagram social network," in *Society for Industrial and Applied Mathematics eBooks*, 2019, pp. 235–243. doi: 10.1137/1.9781611975673.27.

[13]. P.-J. Lee, Y.-H. Hu, K. Chen, J. M. Tarn, and L. Cheng, "Cyberbullying detection on social network services," *Cyberbullying Detection on Social Network Services*, p. 61, Jan. 2018, [Online]. Available: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1060&context=pacis2018

[14]. Md. T. Ahmed, M. Rahman, S. Nur, A. Z. M. T. Islam, and D. Das, "Natural language processing and machine learning based cyberbullying detection for Bangla and Romanized Bangla texts," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 20, no. 1, p. 89, Feb. 2021, doi: 10.12928/telkomnika.v20i1.18630.

[15]. T. Islam, N. Ahmed, and S. Latif, "An evolutionary approach to comparative analysis of detecting Bangla abusive text," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2163–2169, Aug. 2021, doi: 10.11591/eei.v10i4.3107.

.

[16]. Md. F. Ahmed, Z. Mahmud, Z. T. Biash, A. A. N. Ryen, A. Hossain, and F. B. Ashraf, "Cyberbullying Detection Using Deep Neural Network from Social Media Comments in Bangla Language," *arXiv (Cornell University)*, Jun. 2021, doi: 10.48550/arxiv.2106.04506.

[17]. M. Dadvar, F. De Jong, R. Ordelman, and D. Trieschnigg, "Improved cyberbullying detection using gender information," *Improved Cyberbullying Detection Using Gender Information*, pp. 23–25, Feb. 2012, [Online]. Available: https://dolf.trieschnigg.nl/publications/DIR.2012.dadvar.pdf

[18]. Abdhullah-Al-Mamun and S. Akhter, "Social media bullying detection using machine learning on Bangla text," *Social Media Bullying Detection Using Machine Learning on Bangla Text*, Dec. 2018, doi: 10.1109/icece.2018.8636797.

[19]. A. Bozyiğit, S. Utku, and E. Nasıbov, "Cyberbullying detection: Utilizing social media features," *Expert Systems With Applications*, vol. 179, p. 115001, Oct. 2021, doi: 10.1016/j.eswa.2021.115001.

[20] A. Akhter, U. K. Acharjee, Md. A. Talukder, Md. M. Islam, and Md. A. Uddin, "A robust hybrid machine learning model for Bengali cyber bullying detection in social media," *Natural Language Processing Journal*, vol. 4, p. 100027, Sep. 2023, doi: 10.1016/j.nlp.2023.100027.

[21] M. Di Capua, E. Di Nardo, and A. Petrosino, "Unsupervised cyber bullying detection in social networks," *Unsupervised Cyber Bullying Detection in Social Networks*, Dec. 2016, doi: 10.1109/icpr.2016.7899672.

[22] Moshiur Rahman Faisal. (2022). Bengali Cyber Bullying Dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/4229698

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
BENGALI SOCIAL MEDIA COMMENTS ANALYSIS TO PREVENT CYBERBULLYING

Student ID: 201-15-13845,201-15-14152

CO Description for FYDP

CO	CO Descriptions	PO
Phase –I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **Cyberbullying detection** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase –II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project demonstrates **fundamental engineering (K3)** principles by employing deep neural networks, machine learning techniques, and diverse RNN architectures for natural language processing and classification tasks. The project demonstrates **specialist knowledge (K4)** by conducting text classification with advanced machine learning algorithms like logistic regression and deep learning models like Bi-LSTM networks.	**Page no:** [18-22] **Section:** [3.4] **Page no:** [12-14] **Section:** [3.1]
				The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by employing machine learning and deep learning models.	**Page no:** [18-22] **Section:** [3.4] **Page no:** [12-14]

					Section: [3.1]
				This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies, to advance cyberbullying detection using machine learning and deep learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [6-10] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** as we had to collect social media comments from Facebook and as it is a large social media platform we had to face many challenges during data collection	**Page no:** [14-17] **Section:** [3.2]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by comparing experimental outcomes, highlighting our deep learning model(Bi-LSTM) as the chosen significant solution for cyberbullying detection among other discussed approaches.	**Page no:** [30-42] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting individual and mental health which indicates **EP-4**.	**Page no:** [6-8] **Section:** [2.2]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A

6.	**EP6: Extends of stakeholders involved and conflicting**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	Our project solves many small problems like data collection ,data preprocessing, model building etc. ultimately solving a big problem which is discussed in our report which ensures the attainment of **EP-7**.	**Page no:** [14-22] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to cyberbullying detection.	**Page no:** [22-24] **Section:** [3.5]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		Cyberbullying detection fosters safer online spaces, safeguarding mental health by identifying and addressing harmful behaviors promptly. It promotes positive digital interactions, reducing societal discord and enhancing community well-being. By tackling cyberbullying, detection efforts contribute to a healthier online environment, supporting a culture of respect and empathy..	**Page no:** [44-47] **Section:** [6.1, 6.2, 6.3,6.4]

5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in cyberbullying detection on Bengali language through machine learning and deep learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [6-10] **Section:** [2.1,2.2, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [3] **Section:** [1.6]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing individuals privacy and not disclosing personal information of users which comply with **CO9**.	**Page no:** [45-46] **Section:** [6.3]
3	CO10	Yes	We collected our dataset and build machine learning and deep learning models and these tasks were divided among our team members which ensures the attainments of **CO10**	**Page no:** [14-24] **Section:** [3.2,3.3,3.4]
4	CO12	Yes	The project demonstrates its commitment to ongoing learning and adaptation in today's fast-paced technological environment through extensive data gathering, rigorous statistical analysis, careful methodology refinement, and detailed examination of experimental outcomes. This dedication underscores its effort to remain current and improve strategies in response to contemporary challenges which attains **CO12**.	**Page no:** [14-24, 30-42] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

Bengali social media comments analysis to prevent cyberbullying

ORIGINALITY REPORT

15%	12%	9%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	core.ac.uk Internet Source	1%
3	academic-accelerator.com Internet Source	1%
4	Submitted to Liverpool John Moores University Student Paper	<1%
5	Submitted to Daffodil International University Student Paper	<1%
6	Submitted to University of Huddersfield Student Paper	<1%
7	medium.com Internet Source	<1%
8	ojs3.unpatti.ac.id Internet Source	<1%