

**CYBERBULLYING SENTENCE DETECTION USING MACHINE LEARNING
FOR THE BENGALI LANGUAGE**

BY

**Shamim Hossen
ID: 201-15-14047**

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

Supervised By

Mr. Narayan Ranjan Chakraborty
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Dewan Mamun Raza
Lecturer (Senior Scale)
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

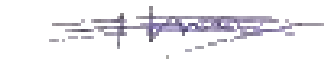
DHAKA, BANGLADESH

19TH JANUARY 2023

APPROVAL

This Project titled “**Cyberbullying Sentence Detection Using Machine Learning for the Bengali Language**”, submitted by **Shamim Hossen**, ID No: **201-15-14047** 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 **19th January 2023**.

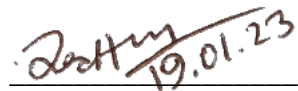
BOARD OF EXAMINERS



Dr. Touhid Bhuiyan
Professor and Head

Chairman

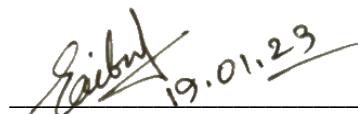
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Md Zahid Hasan
Associate Professor

Internal Examiner

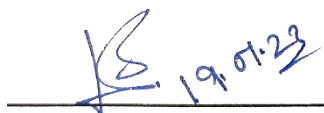
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Mr. Saiful Islam
Assistant Professor

Internal Examiner

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Md. Sazzadur Rahman
Associate Professor

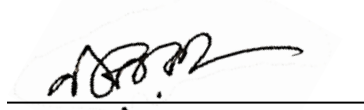
External Examiner

Institute of Information Technology
Jahangirnagar University

DECLARATION

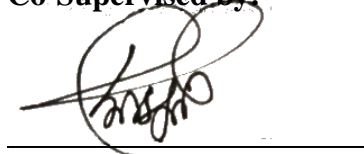
I hereby declare that, this project has been done by me under the supervision of **Mr. Narayan Ranjan Chakraborty, Associate Professor** and co-supervision of **Mr. Dewan Mamun Raza, Lecturer (Senior Scale), Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Mr. Narayan Ranjan Chakraborty
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:



Mr. Dewan Mamun Raza
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:



Shamim Hossen
ID: 201-15-14047
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project successfully.

I really grateful and wish my profound indebtedness to **Supervisor Mr. Narayan Ranjan Chakraborty, Associate Professor** and **Co-Supervisor Mr. Dewan Mamun Raza, Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Machine Learning and 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 stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Mr. Narayan Ranjan Chakraborty**, Associate Professor and **Prof. Dr. Touhid Bhuiyan** Head, Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Cyberbullying has become a significant issue in recent years, particularly among young people, and there is a need for practical tools to detect and prevent it. Cyberbullying is sending offensive, abusive or threatening messages to insult a person. It is more dangerous than traditional bullying because it can occur at any time and from any location and be done anonymously. Social media is getting vast amounts of data every day. We see in current trends that Cyberbullying is a bigger and bigger problem. It is even more severe than regular bullying's on the internet, like Facebook, Twitter, or other internet platforms. Finding a bully is far more challenging. So, we collect a data tool that essentially interacts and communicates with various social media site data using vest technologies like natural language processing and machine learning that automatically detect bullying. The study employs various machine-learning algorithms to develop a model for detecting cyberbullying sentences in Bengali text also discusses the challenges faced in developing a machine-learning model for Bengali cyberbullying detection and the potential solutions. Overall, the study demonstrates the potential of machine learning for detecting cyberbullying in Bengali and contributes to developing practical tools to prevent and combat cyberbullying in the Bengali-speaking community. This paper proposes an approach to detect cyberbullying in Bangla sentences using social media datasets and machine learning techniques, and the evaluation dataset shows that Multinomial Naïve Bayes performs better and achieves an accuracy of 79.49%.

TABLE OF CONTENTS

| CONTENTS | PAGE |
|------------------------------------|-------------|
| Board of examiners | i |
| Declaration | ii |
| Acknowledgements | iii |
| Abstract | iv |
| | |
| CHAPTER | |
| CHAPTER 1: INTRODUCTION | 1-5 |
| 1.1 Introduction | 1 |
| 1.2 Motivation | 1 |
| 1.3 Rationale of the Study | 2 |
| 1.4 Research Questions | 3 |
| 1.5 Expected Output | 3 |
| 1.6 Project Management and Finance | 4 |
| 1.7 Report Layout | 4 |
| | |
| CHAPTER 2: BACKGROUND | 6-11 |
| 2.1 Preliminaries/Terminologies | 6 |
| 2.2 Related Works | 6 |

| | | |
|---|---|--------------|
| 2.3 | Comparative Analysis and Summary | 10 |
| 2.4 | Scope of the Problem | 10 |
| 2.5 | Challenges | 10 |
| CHAPTER 3: RESEARCH METHODOLOGY | | 12-19 |
| 3.1 | Research Subject and Instrumentation | 12 |
| 3.2 | Data Collection Procedure/ Dataset Utilized | 12 |
| 3.3 | Statistical Analysis | 16 |
| 3.4 | Proposed Methodology/Applied Mechanism | 18 |
| 3.5 | Implementation Requirements | 19 |
| CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION | | 20-24 |
| 4.1 | Experimental Setup | 20 |
| 4.2 | Experimental Results & Analysis | 20 |
| 4.3 | Discussion | 24 |
| CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY | | 25-26 |
| 5.1 | Impact on Society | 25 |
| 5.2 | Impact on Environment | 25 |
| 5.3 | Ethical Aspects | 25 |

| | |
|---|--------------|
| 5.4 Sustainability Plan | 26 |
| CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH | 27-28 |
| 6.1 Summary of the Study | 27 |
| 6.2 Conclusions | 27 |
| 6.3 Implication for Further Study | 27 |
| REFERENCES | 29-30 |

LIST OF FIGURES

| FIGURES | PAGE NO |
|--|----------------|
| Figure 3.1: Dataset Summary | 13 |
| Figure 3.2: Separate Stop Word | 14 |
| Figure 3.3: After Data Cleaning | 14 |
| Figure 3.4: Preview of cleaning data | 15 |
| Figure 3.5: Total Dataset | 16 |
| Figure 3.6: Conversion Word Frequency | 17 |
| Figure 3.7: Conversion Character Frequency | 17 |
| Figure 3.8: Dataset Statistics | 18 |
| Figure 3.9: Methodologies Diagram | 19 |
| Figure 4.1: Analysis of Performance | 23 |
| Figure 4.2: Confusion matrix | 24 |

LIST OF TABLES

| TABLES | PAGE NO |
|---|----------------|
| Table 4.1: The performance of Accuracy | 21 |
| Table 4.2: The performance of Precision | 21 |
| Table 4.3: The performance of Recall | 22 |
| Table 4.4: The performance of F1-Score | 22 |

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nowadays, we all use the internet and social media. The cyberbullying effect occurs worldwide, and the number of cases is increasing. The proliferation of technology has led to an increase in cyberbullying, defined as using technology such as cell phones, instant messaging apps, email, and chat rooms, or using websites like Facebook and Twitter to threaten, harass, or abuse someone. Children, in particular, are at risk of cyberbullying as they have access to these technologies at a younger age. Cyberbullying can significantly impact the victim's mental well-being, causing distress, elevated stress levels, depression, and anxiety. Therefore, developing practical tools to prevent and combat cyberbullying in the Bengali-speaking community is essential. Cyberbullying can occur on various digital platforms, such as social media, messaging apps, gaming sites, and mobile devices. Bullying is an abuse of technology's advantages. So, it is Such A Big Problem on the Internet and Social Media. Cyberbullying activity is challenging on the internet and Social Media because online information is too large. So, humans cannot track it. The study employs various machine learning algorithms to develop a model for detecting cyberbullying sentences in Bengali text and compares the results with existing approaches in the literature. We discuss the challenges encountered in creating a machine-learning model for Bengali cyberbullying detection and potential solutions. Despite the growing concern over cyberbullying, there still needs to be more research on the issue in the Bengali language. Therefore, we developed an autonomous machine learning system model to identify cyberbullying and evaluated the outcomes against other machine learning models.

1.2 Motivation

We currently live in a modern era with ongoing technological advancement. We cannot think without the Internet and mobile phones. Everyone can use the Internet, making it

affordable for most people to use. We use it to research information and concepts, which we have proven to be a faster and more efficient form of communication worldwide. However, some people take advantage of this technological advancement. A form of bullying or harassment that occurs online is known as cyberbullying or cyberharassment. Cyberbullying and cyberharassment are both forms of online bullying. With the development of technology and the growth of the digital domain, it has become more common, especially among teenagers. Cyberbullying can negatively impact one's mental well-being, causing distress, elevated stress levels, anxiety, and depression. We see teenagers becoming depressed or committing suicide as a result of cyberbullying. So, it is not a small problem in our society. We recognize that cyberbullying is a prevalent issue in our society. We aim to design a system to identify and prevent cyberbullying by blocking any correspondence considered bullying.

1.3 Rationale of the Study

The lack of research on cyberbullying in the Bengali language highlights the need for further study in this area. It is essential to address and understand the unique challenges and dynamics of cyberbullying within the Bengali-speaking community. This research aims to fill this gap and contribute to the development of practical tools and strategies for preventing and addressing cyberbullying in the Bengali language. We recommend giving more attention to the issue of cyberbullying in the Bengali language, despite the significant research already conducted on this topic in the English language. Our goal is to use machine learning techniques to identify and detect cyberbullying sentences in the Bengali language. The research will provide valuable insights for individuals, educators, and researchers working to combat cyberbullying in the Bengali language and to develop practical tools and strategies for preventing and addressing cyberbullying in the Bengali-speaking community.

1.4 Research Questions

In this research, we presented a model for identifying cyberbullying statements. This research question will be:

- Define the characteristics of cyberbullying participants.
- Does the system predict ultimate performance based on sample data?
- What are the current approaches for detecting cyberbullying in the Bengali language?
- What are the key challenges and limitations in detecting cyberbullying in the Bengali language?
- What are the most effective machine learning algorithms for detecting cyberbullying in the Bengali language?
- What are the most effective machine learning algorithms for detecting cyberbullying in the Bengali language?

These research questions aim to explore the current state of research on cyberbullying detection in the Bengali language, identify the key challenges and limitations in this area, evaluate the effectiveness of different machine learning algorithms for detecting cyberbullying in Bengali, and explore potential solutions to the challenges faced in this area. The research will contribute to developing practical tools and strategies for preventing and addressing cyberbullying in the Bengali language. Cyberbullying refers to using cell phones, instant messaging apps, email, chat rooms, or social media platforms like Facebook and Twitter to harass, threaten, or intimidate someone. This behavior is prevalent among children, particularly as they have access to these technologies at a younger age. Our system predicts ultimate performance based on sample data, and we used seven machine learning models to compare which model gives the best output and performance.

1.5 Expected Output

Following the specifications of our research project, we chose our model and gathered the dataset. Using this collected dataset, we verified that our selected classifier could be applied

to a wide range of situations based on our dataset. Our proposed model achieved an accuracy rate of 79%, which exceeded our expectations. This model's performance paves the way for thinking about how to improve results. The research goal was to determine whether a sentence was bullying or neutral. We used classifiers based on various machine learning models to determine the sentence type. The outcome meets two criteria "Bullying" and "neutral." Each model required 80% of the data to be trained. On different models, we get varying degrees of accuracy. The multi-naive Bayes algorithm achieved the highest accuracy among all the models we tested. The performance of the classifiers was used to assess their decision-making ability. We used several metrics to evaluate the classifier's performance, such as accuracy, precision, recall, and F1 score. We considered overall accuracy as an appropriate standard for a classifier. Knowing the number of correctly classified samples in the test set is essential to determine the classifier's performance.

1.6 Project Management and Finance

Project management for this research will involve planning and organizing the resources required to complete the project. This includes scheduling the tasks to be completed, allocating resources, and monitoring the project's progress. We use Google Colab notebooks for this research project. Google Colab notebooks give a python development environment that executes python code and gives some online RAM and Disk for running the project. It is free and the most popular for machine learning project management. So, we don't need to pay to use this software or web application. Overall, the project management and finance of this research will be closely monitored to ensure that the project is completed within the given time frame and budget while ensuring the quality of the research.

1.7 Report Layout

This research project's report format will be designed to be easily understood and followed. It will be broken down into several sections, each addressing a specific aspect of research. These sections will include an introduction that will provide an overview of the research

topic and the study's goals, as well as information on cyberbullying and the need for further research in this area. A literature review section will thoroughly investigate existing literature on detecting cyberbullying and machine learning in the Bengali language. A methodology section will explain the research design, data collection and preprocessing, and the machine learning algorithms used in the study. A results and discussion section will present the experiments' outcomes, the metrics used to evaluate the model's performance in the existing literature, and the challenges encountered during the research. A section on the research impact will summarize this research's impact and different aspects. A conclusion will summarize the research's main findings and provide suggestions for future work. A references section will list all sources cited in the report. The report will be written in clear and concise language, and all figures and tables will be correctly labeled and captioned to aid understanding. Additionally, the report will be thoroughly proofread and edited to ensure that it is free of errors and conforms to academic standards.

CHAPTER 2 BACKGROUND

2.1 Preliminaries/ Terminologies

Cyberbullying is characterized by individuals intentionally and aggressively conducting actions through electronic communication technologies. Cyberbullying is the act of actively transmitting, publishing, or spreading harmful, negative, or malicious content about another person. It can also include revealing sensitive or personal information about someone else in a way that causes shame or embarrassment. In some cases, cyberbullying can even cross over into illegal or criminal activity. Cyberbullying affects 49% of Bangladeshi school students. Individuals who are victims of cyberbullying may experience a range of adverse effects, such as mental health problems, decreased academic performance, considering dropping out of school, and even suicidal thoughts. According to a study conducted by the Cyber Crime Awareness Foundation, at least 50.27 percent of victims of cybercrime are victims of cyberbullying in various forms, and such crimes are on the rise in Bangladesh. UNICEF calls for coordinated action to prevent bullying and harassment among 32% of online children in Bangladesh. In psychiatrists' opinion, Continual humiliation undermines one's self-esteem and feeds the inferiority complex. To improve the system's accuracy of the model, we utilized 80% of the data for training and 20% of the data for testing. The training dataset dramatically influences the accuracy of the model. We employed specific techniques, such as adjusting the parameters of machine learning models to acquire more precise results. Our study found that the Support Vector Machine had an accuracy rate of approximately 79.45%. The performance of the remaining algorithms is nearly as accurate as possible.

2.2 Related Works

They proposed an approach to finding out cyberbullying action using trained and supervised machine learning. They achieved a neural network that performs 92.8% and 90.3 SVM accuracy and got their labeled dataset from Kaggle. [1]

After analysis, the Bully framework was introduced. This collaborative study used the multi-modal method on social media data to specifically identify cyberbullying text and initially reformulate multi-model social networking information as a different system and finds out node embedding representations upon it. Working with Instagram and Vine's datasets shows that this framework is superior to the progressive cyberbullying detection models. [2] The study aimed to evaluate the use of video games as tools by analyzing real-world multi-modal social media datasets from Instagram and Vine. It looked at the mechanics and types of games, the users, the studies conducted, the number of users, and the accessibility of these tools to understand how society benefits from the potential of video games. The results reveal a diverse range of video games being utilized for different purposes, one of them being countering cyberbullying. Most of those games are not presently available. [3] The authors outlined the steps and techniques used in identifying cyberbullying in their study. They discussed using various feature selection methods and machine learning techniques in combination with gathering data to detect cyberbullying and feature engineering processes to predict cyberbullying behaviors. [4] They proposed a new way for Multilingual Cyberbullying identifying model to find out about cyberbullying but two different Indian languages that are Hindi and Marathi. They developed a prototype for these two languages and got the language's F1 scores up to 96% and 97% accuracy by doing experiments using many datasets. [5] The authors proposed using a blend of text and image-based features to detect various forms of cyberbullying by utilizing a deep learning model with a pre-trained VGG-16 network. They improved the system's performance using a genetic algorithm and achieved a 78% F1 score with the same dataset. [6] Their paper proposes a new technique for detecting cyberbullying on a deep neural network, which depends on contrast to contrast. An artificial Neural Network is implemented for the best outcomes. They discuss several topics like System Flow, Word Embedding, Architecture, Input, Embedding, Dropout, Convolutional layer, and many more. [7] They propose specific designs for cyberbullying detection with early identification techniques and two different features. Using an extensive experiment with a real-world dataset, they improved and achieved up to 42%.for baseline detection. [8] They proposed a technique based on multiple SMPs and performed extensive evaluation using Form spring, twitter.

They collected data from actual world dataset posts and achieved a 78.5% score accuracy. [9] They proposed a new model training scheme for cyberbullying detection by adopting general limitations and validating approaches using various data sets. They successfully reduce some unintentional biases without sacrificing the model's quality. They utilized machine learning techniques that are neutral, clear, and moral in their approach. [10] They aimed to detect cyberbullying detection on the importance of social media. It gives more successful prediction performance by machine learning algorithms. [11] They compare a mixture of lexicon-based techniques and machine learning algorithms that can enhance the precision of detecting cyberbullying. The study also revealed that the Support Vector Machine (SVM) algorithm effectively identified cyberbullying with the most incredible precision. [12] They present a study that aims to detect cyberbullying on social media platforms using a hybrid model that combines machine learning and natural language processing techniques. They collected a dataset of tweets in Spanish and labeled them as either cyberbullying or non-cyberbullying. They then used natural language processing techniques such as word embeddings and sentiment analysis. The hybrid model achieved an accuracy of 90.7% in detecting cyberbullying tweets, which is higher than the models that only used either machine learning or natural language processing techniques. [13] They surveyed the various approaches used for detecting cyberbullying messages and natural language processing techniques used to detect cyberbullying. The study also looks at the advantages and limitations of each approach. The article concludes that a combination of machine learning and natural language processing techniques can effectively detect cyberbullying messages with high accuracy. [14] Authors propose a method for detecting cyberbullying in online social networks using AI models, specifically machine learning and deep learning. They use a feature selection technique called chaotic salp swarm optimization (CSSO) to extract relevant features from OSN data, a stacked autoencoder model for classification, and the mayfly optimization (MFO) algorithm for parameter tuning. They conduct experiments to evaluate the performance of their proposed method and claim that it is superior to other techniques. [15] They explore the issue of cyberbullying on social media and the need for intelligent systems to detect potentially harmful messages. It uses linear support vector machines with a rich feature set and

investigates which information sources are most useful for the task. The results reveal promising results for detecting cyberbullying-related posts with an F1 score of 64% and 61% for English and Dutch, respectively, outperforming baseline systems. [16] The study aims to create a method for identifying and preventing cyberbullying on social media by using Natural Language Processing and other machine-learning techniques. The research team used a dataset from Kaggle, including Twitter data that had been labeled to train the algorithm. They found that the logistic regression model had the best performance and accuracy compared to other methods such as SVM, Random Forest, Naive-Bayes, and Xgboost. The article concludes that the model they developed is an effective solution for detecting cyberbullying on social media sites. [17] They aim to address the issue of cyberbullying on social media platforms, mainly focusing on sarcasm. The authors review previous research on the topic and propose their approach using a Support Vector Machine (SVM) classifier. The results of their study show that the SVM classifier performed better than other classifiers in detecting cyberbullying with sarcasm. [18] They focus on developing an effective detection technique for handling online harassment in Bengali, the seventh most spoken language in the world. The study used 44,001 user comments from popular public Facebook pages, classified into five categories: Non-bully, Sexual, Threat, Troll, and Religious. Evaluation of the recommended models' efficacy showed that the binary classification model had an accuracy of 87.91%, while the multiclass classification model using an ensemble technique after the neural network had an accuracy of 85%. This study demonstrates the need for more investigation into cyberbullying detection in various languages. [19] They focus on identifying and recognizing abusive language in Bengali using three Steps. Firstly, a vast dataset is gathered from Facebook and YouTube, and several supervised machine learning classifiers are applied to find the most optimal results. The second step is creating a balanced dataset by randomly reducing the majority class. [20] Finally, the final experiment is carried out by using a Bengali statement on the dataset. The outcome indicates that Support Vector Machine with the full dataset had the highest accuracy of 88%.

2.3 Comparative Analysis and Summary

The objective of this study aimed to develop a detection system through various machine learning models that can extract sentiment from Bangla sentences and detect bullying sentences. Technology based on machine learning has its own set of issues and restrictions. It has undergone significant improvement compared to the original edition of various models. Seven machine learning models were employed in this research project, including Random Forest, Decision Tree, Multinomial Naive Bayes, Stochastic Gradient Descent, Support Vector Machine, Logistic Regression, and K-Nearest Neighbors. These models implement through python programming language to get the outcome or result with our Bangla language dataset.

2.4 Scope of the Problem

Cyberbullying detection model goal is to recognize the cyberbullying text and be aware of its meaning. Also, it highlights various forms of verbal aggressiveness and cyberbullying in texts and their intensity. The prevalence of cyberbullying is rising. The harmful information that can affect students' physical and mental well-being in the present and future can be permanently recorded due to its digital nature. The primary goal of the problem is to detect the presence of cyberbullying in online activities like a post, comments, and chat conversations that are meant to reduce the bullying behavior in online conversations or activities on social media Sites. Also, to prevent teenagers from getting depressed or committing suicide due to cyberbullying activities.

2.5 Challenges

Due to the lack of resources and datasets in Bangla, Data collection took much work to acquire. We were given the challenge to gather our data because of this. Manual data collecting is not as simple. We are manually gathering information from posts and comments on several social media sites, such as Twitter, Facebook, Quora, Reddit, and YouTube. In this situation, we maintain some procedures for data collection to get our data.

- Gathering accurate chat data
- Collecting reliable conversation data
- The gathering of Bengali stop words
- Discovering appropriate Python library packages
- Increasing the model's accuracy

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The research subject of this thesis is the detection of cyberbullying in Bengali language text using machine learning techniques. The different models can accurately classify Bengali language text samples as cyberbullying or non-cyberbullying. In order to achieve this goal, various machine learning techniques and tools were used as the instrumentation for this research. Acceptable methodologies require the implementation of research methods. It aids the researcher in determining which strategy is best suited to their requirements. Every research project follows a similar path to achieve the desired result, which is the research goal. Many words must be classified in order to accomplish this.

- What information should be gathered?
- How do you know the information you've gathered is correct?
- How should each data set be organized?
- What should the labels for each data point be?

3.2 Data Collection Procedure/ Dataset Utilized

Google spreadsheet is a free and powerful sheet for data storage. We use google datasheet and Microsoft Excel to store data in the dataset. The dataset was then exported into an.xlsx file and dumped into a few columns. After that, two columns were created, one named "Result" as the target column and the other "Comment" as a feature column. It has already been mentioned that there are two sorts of comments: "Bullying" and "Neutral." Each exchange of words or sentences is the beginning of the dialogue, and some words in a sentence do not express their meaning but aid in the formation of a sentence. Data was collected manually from random social media sites like Facebook, Twitter, YouTube, Reddit posts, and comment sections. Around two thousand plus sizes of the dataset were collected. After the data collection, data was labelled and verified by three different people.

There are two kinds of labelling, bullying, and Natural, in the dataset. Bangla sentence dataset summary and preview as Figure 3.1 given below

| | Comment | Result |
|------|--------------------------|----------|
| 0 | তোমাকে আজকে সুন্দর লাগছে | Neutral |
| 1 | হালায় কি কামড়া করলো | Bullying |
| 2 | ভালা হয়ে যাও মাসুদ | Neutral |
| 3 | ছেলে একটু স্বাস্থ্যবান | Neutral |
| 4 | তুই তো আফ্রিকান | Bullying |
| ... | ... | ... |
| 2364 | কাটা লাগা উই মা উই মা | Bullying |
| 2365 | গাঞ্জাখুরের মত লাগে | Bullying |
| 2366 | ওয়াও লুকিং সো গর্জিয়াস | Neutral |
| 2367 | ভালো নেই আমাদের সিলেট | Neutral |
| 2368 | ওই হাতির বাচ্চা | Bullying |

2369 rows × 2 columns

Figure 3.1: Dataset Summary

In the first step of preprocessing, we remove punctuation. Stop words must be identified and, for every language, deleted in natural language processing that is given Figure 3.2, For our study, we gathered Bangla-stopped words and cleaned up our data by removing them. The Bangla language contains approximately 526 stop words. As an example: “অতএব”, “অথচ”, “এ”, “এখন” etc. The data cleaning Figure 3.3 and the preprocessing procedure are given below:

```

def process_conversations(Conversation):
    stp = open('bangla_stopwords.txt', 'r', encoding="utf8").read().split()
    result = Conversation.split()
    Conversation = [word.strip() for word in result if word not in stp ]
    Conversation = " ".join(Conversation)
    Conversation = re.sub('[^\u0980-\u09FF]', ' ', str(Conversation))
    return Conversation

```

Figure 3.2: Separate Stop Word

```

Original:
মেয়ের তো এক হাত বাকা
Cleaned:
এক হাত বাকা
Detected as:-- Bullying

Original:
ওরে দেখতে নাইজিরিয়ানদের মত লাগে
Cleaned:
ওরে নাইজিরিয়ানদের মত লাগে
Detected as:-- Bullying

Original:
আমি তোমাকে ভালোবাসি
Cleaned:
ভালোবাসি
Detected as:-- Neutral

```

Figure 3.3: After Data Cleaning

After removing Bangla-stopped words and punctuation from the original sentence or conversion, we get our root sentence or phrase from this cleaned data. The train machine learning model detects bullying or non-bullying sentences based on the root sentence or comment.

| | Comment | Result | cleaned | length | no_char |
|------|--------------------------|----------|--------------------------|--------|---------|
| 0 | তোমাকে আজকে সুন্দর লাগছে | Neutral | সুন্দর | 1 | 6 |
| 1 | হালায় কি কামড়া করলো | Bullying | হালায় | 1 | 5 |
| 2 | ভালা হয়ে যাও মাসুদ | Neutral | ভালা যাও মাসুদ | 3 | 14 |
| 3 | ছেলে একটু স্বাস্থবান | Neutral | স্বাস্থবান | 1 | 10 |
| 4 | তুই তো আফ্রিকান | Bullying | আফ্রিকান | 1 | 8 |
| ... | ... | ... | ... | ... | ... |
| 2357 | কাটা লাগা উই মা উই মা | Bullying | কাটা লাগা উই উই | 4 | 19 |
| 2358 | গাঞ্জাখুরের মত লাগে | Bullying | গাঞ্জাখুরের | 1 | 8 |
| 2359 | ওয়াও লুকিং সো গর্জিয়াস | Neutral | ওয়াও লুকিং সো গর্জিয়াস | 4 | 15 |
| 2360 | ভালো নেই আমাদের সিলেট | Neutral | ভালো সিলেট | 2 | 8 |
| 2361 | ওই হাতির বাচ্চা | Bullying | হাতির বাচ্চা | 2 | 26 |

2362 rows × 5 columns

Figure 3.4: Preview of cleaning data

Figure 3.4 shows the total Bangla sentence or comment with the cleaned word, number of characters, number of lengths, and two kinds of correction: bullying or neutral comment. Data cleaning is a crucial step in the process of using machine learning for the detection of cyberbullying in Bengali language text. This step involves removing irrelevant information and ensuring consistency in the text samples. Overall, the data-cleaning process for this thesis involved removing extraneous information, reducing words to their base form, and eliminating duplicates and irrelevant samples to improve the performance of the machine-learning models. These techniques were applied to the dataset of Bengali language text samples used for this thesis to prepare the data for the machine learning models.

3.3 Statistical Analysis

The dataset for this thesis was collected from various online sources and consisted of Bengali language text samples labelled as either cyberbullying or non-cyberbullying. We used 80% of the dataset's data to train our model and 20% to test it. To divide the dataset, the train-test split function was used. Also given below is Figure 3.5,

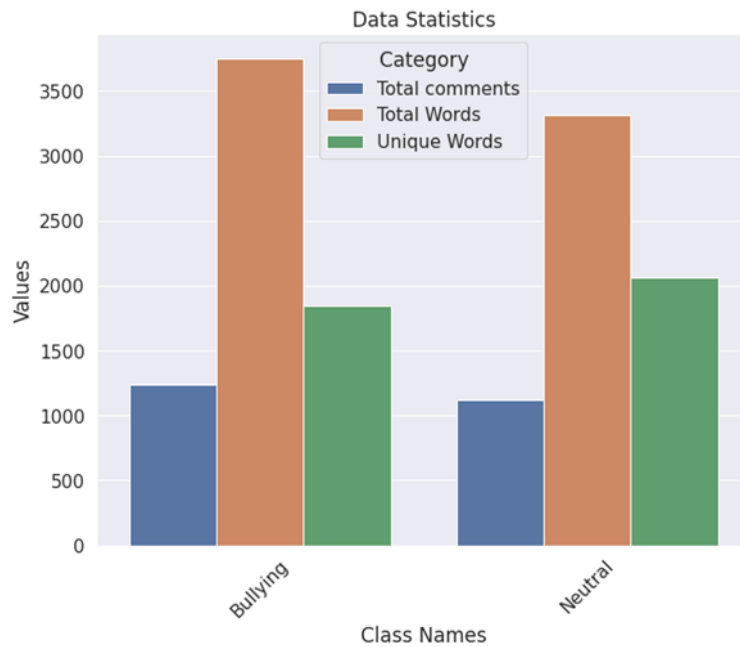


Figure 3.5: Total Dataset

Which describes the dataset's characteristics, such as the number of total Bengali text samples labelled. Bullying or neutral comments, the number of words, and unique words present in the dataset. The statistics of Word Frequency Figure 3.6 and Character Frequency Figure 3.7 of conversion are given below.

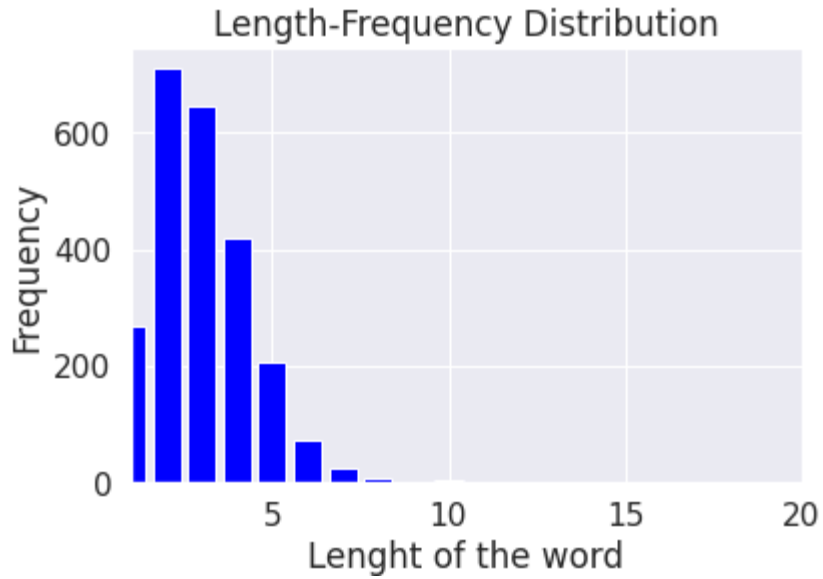


Figure 3.6: Conversion Word Frequency

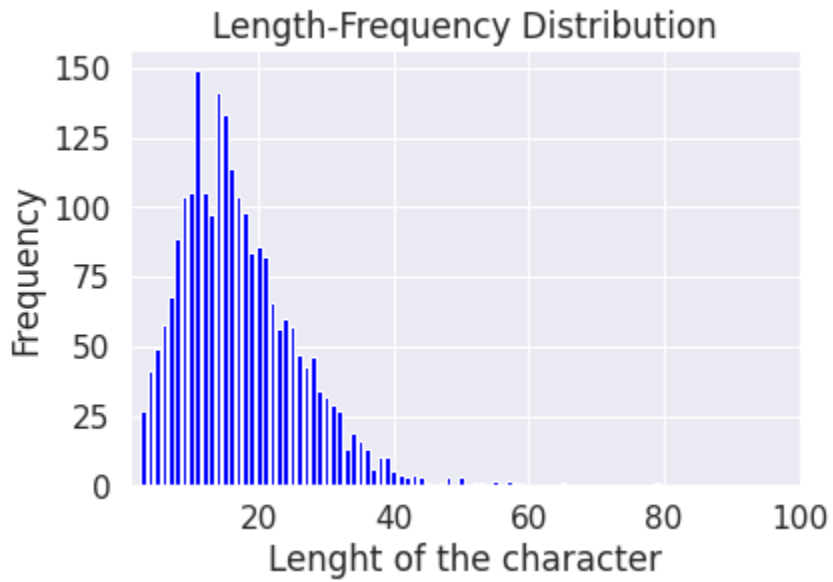


Figure 3.7: Conversion Character Frequency

The dataset was preprocessed to remove irrelevant information and ensure the text samples' consistency. The entire dataset statistics are given below in Figure 3.8, Here, approaches for supervised machine learning were utilized.

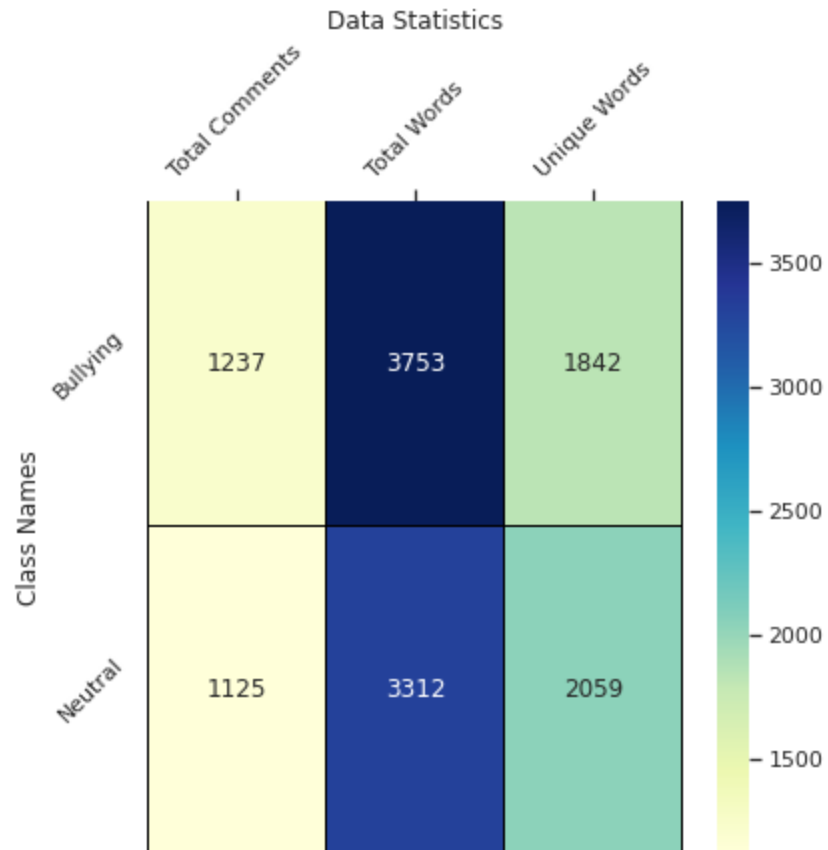


Figure 3.8: Dataset Statistics

To evaluate the accuracy of our datasets, we used classifier-based techniques: Decision Tree, Logistic Regression, Random Forest, KNN, SVM, Multinomial Nave Bayes, and SGD. We used machine-learning approaches to fulfill our goal and compare their accuracy and output result.

3.4 Proposed Methodology/Applied Mechanism

The proposed methodology detects cyberbullying sentence detection using machine learning technic. First, we take the dataset as input. The data will be preprocessed with stop word and removal to remove any irrelevant information and to ensure consistency in the text samples. After that, choose the best algorithm for the machine learning model and select that model to train the dataset. Then feature extraction and training data set will be

applied to some classifiers like the Random Forest, SVM, etc. The model was evaluated using the testing dataset to measure its performance in detecting cyberbullying in Bengali language text. The evaluation metrics used were accuracy, precision, recall and F1 score. At last, test that model with a real-life example. So, we get our result or output for identifying cyberbullying data in Bengali sentences. The entire methodology Figure 3.9 is given below as a diagram.

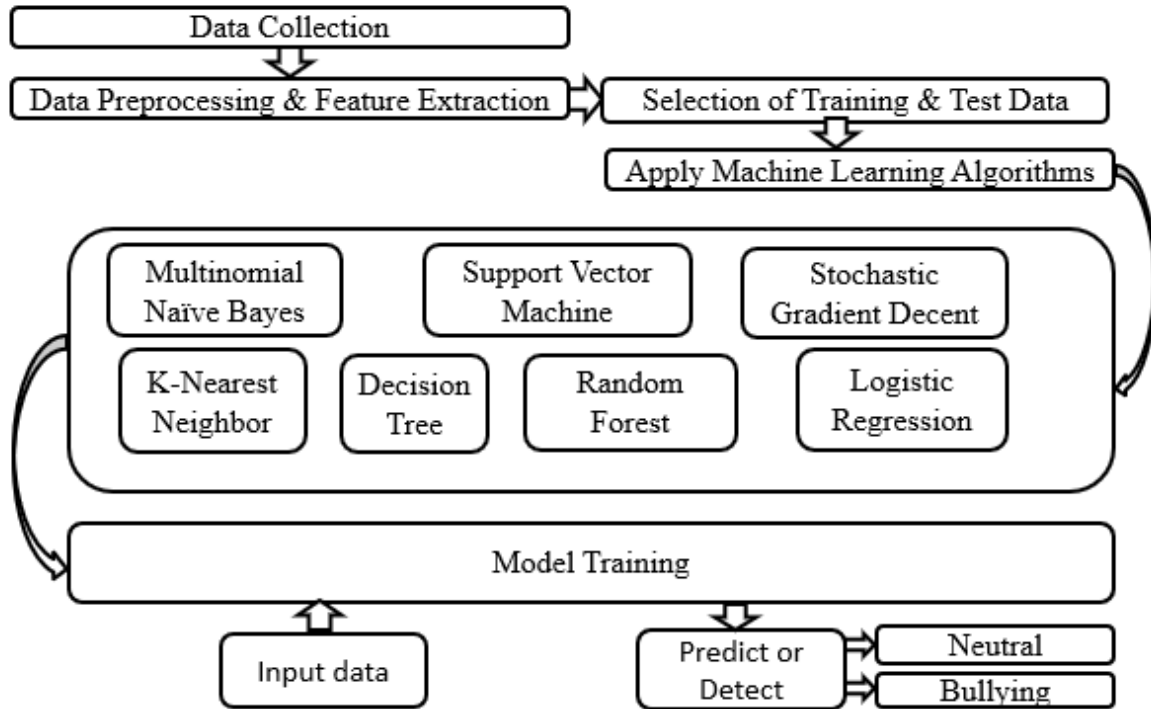


Figure 3.9: Methodologies Diagram

3.5 Implementation Requirements

Python is used as a programming language for implementation. Moreover, google Colab notebooks to execute python code. Also, we used Anaconda and some other IDE for this implementation. Python 3.11 is the most recent version. It is a programming language with a high level of abstraction. The majority of researchers use it to conduct their research. Google Colab is an open-source and free Python programming language distributor. We use it for online collaboration. Anaconda is a scientific computing environment. Anaconda is an offline Python and R programming language distribution. We used Jupyter Notebook for faster execution.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

To build our model and run the scripts, we need some data. The setup procedure is as follows: Collecting conversation data is the first and most crucial step in extracting sentiment from chats. We needed help gathering the necessary information. We worked hard to make the dataset more relevant and subjective to our objectives. For this purpose, we obtain conversation sentence data from various popular social media sites, post comments, and chat sections. All conversations are possible in communication systems based on a data resource. We collect them as phrases and save them in a .xlsx file, detecting the sentence type (bullying and neutral). The data is ready to use after the preparation procedure. This was the end of the line, and we began making additional preparations.

4.2 Experimental Results & Analysis

The experimental results were obtained by training and testing the model on a dataset of Bengali sentences labelled as cyberbullying or non-cyberbullying. The model's performance was evaluated using several metrics such as accuracy, precision, recall, and F1-score. The dataset used in the experiments consisted of a total of 2369 sentences, with 1238 sentences labelled as cyberbullying and 1131 sentences labelled as non-cyberbullying. Training and testing sets were created from the dataset, with 80% of the data used for training and 20% used for testing. The model was trained using various machine learning models, Naive Bayes, Random Forest, and Logistic Regression. Table 4.1, Table 4.2, Table 4.3 and Table 4.4 is given accuracy, precision, recall and f1-score of the performance of the machine learning model and algorithms. The best performance was achieved using Multinomial Naïve Bayes, which reached an accuracy of 79.49%, precision of 76.73%, recall of 75.61%, and F1-score of 76.17%.

Table 4.1: The performance of Accuracy

| Classifier | Accuracy |
|-------------------------|-----------------|
| Logistic Regression | 76.53 |
| Decision Tree | 73.36 |
| Random Forest | 69.77 |
| Multinomial Naïve Bayes | 79.49 |
| KNN | 75.05 |
| SVM | 76.53 |
| SGD | 77.17 |

We find different amounts of accuracy on individual models. Between seven models, Multi. Naive Bayes works effectively and has the best accuracy. Table 4.1 is given the accuracy of the classifier. Among these machine learning classifiers, Multinomial Naïve Bayes provides the highest accuracy.

Table 4.2: The performance of Precision

| Classifier | Precision |
|-------------------------|------------------|
| Logistic Regression | 69.11 |
| Decision Tree | 62.87 |
| Random Forest | 59.57 |
| Multinomial Naïve Bayes | 76.73 |
| KNN | 74.03 |
| SVM | 68.65 |
| SGD | 69.80 |

The results demonstrate that the proposed model can effectively detect cyberbullying sentences in the Bengali language. The model's high accuracy, precision, recall, and F1 score indicate that the model can correctly identify cyberbullying sentences while minimizing the number of false positives and false negatives.

Table 4.3: The performance of Recall

| Classifier | Recall |
|-------------------------|---------------|
| Logistic Regression | 82.93 |
| Decision Tree | 94.15 |
| Random Forest | 94.15 |
| Multinomial Naïve Bayes | 75.61 |
| KNN | 65.37 |
| SVM | 84.39 |
| SGD | 83.41 |

Table 4.4: The performance of F1-Score

| Classifier | F1-Score |
|-------------------------|-----------------|
| Logistic Regression | 75.39 |
| Decision Tree | 75.39 |
| Random Forest | 72.97 |
| Multinomial Naïve Bayes | 76.17 |
| KNN | 69.43 |
| SVM | 75.71 |
| SGD | 76.00 |

Different results may be obtained depending on the dataset and algorithms used based on a specific dataset and machine learning algorithms. More tests with various parameters and methodologies can enhance the model's performance. A thorough examination of the efficiency and effectiveness of multiple algorithms was conducted to determine the optimal method for completing the task at hand. The graph of performance analysis on different algorithms is shown in Figure 4.1.

The performance analysis of different algorithms revealed distinct strengths and weaknesses among the various approaches, ultimately selecting the most suitable algorithm for achieving the desired outcome.

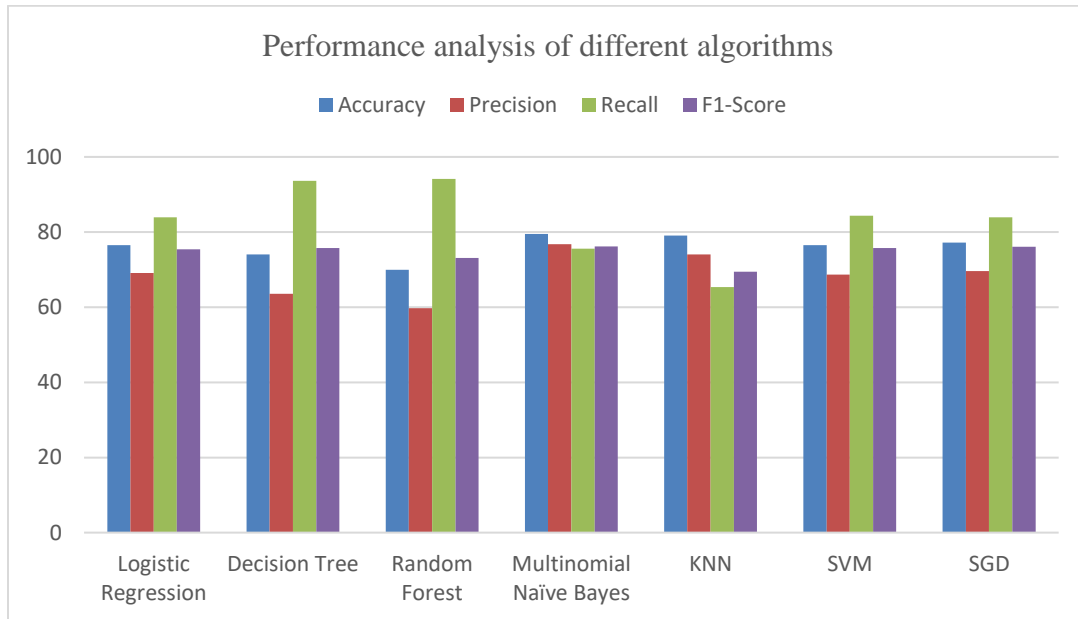


Figure 4.1: Analysis of Performance

The confusion matrix for the Bengali language cyberbullying sentence detection using machine learning showcased the precision and recall rates of the algorithm, providing valuable insight into its performance and identifying areas for improvement. The confusion matrix was used to evaluate the model's performance which helped to fine-tune the model and improve its accuracy. It also helped to identify the specific types of errors that the model was making, such as misclassifying cyberbullying sentences as non-cyberbullying sentences or vice versa, which helped to focus the efforts on addressing these issues and improving the overall performance of the model. Furthermore, the confusion matrix provided a comprehensive understanding of the model's performance across different classes of cyberbullying sentences, enabling the team to optimize the algorithm for better performance on specific types of cyberbullying sentences and ultimately improve the detection accuracy of the model.

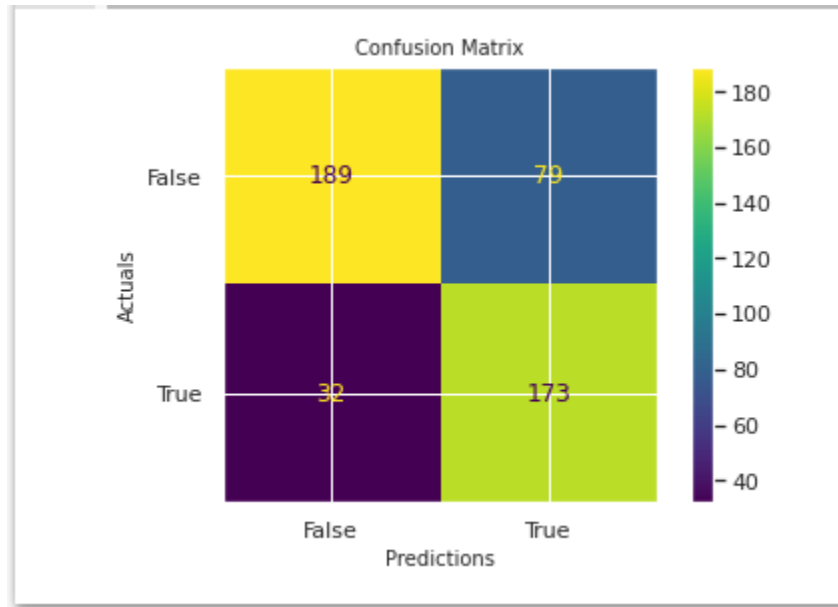


Figure 4.2: Confusion matrix

In conclusion, the proposed cyberbullying sentence detection model for the Bengali language using machine learning techniques has shown promising results and has the potential to be used in practical applications to detect and prevent cyberbullying in the Bengali language.

4.3 Discussion

Our model and dataset are being updated to fulfill our requirements. By making this adjustment, we can ensure that our classifier is appropriate for various applications using our dataset. This proposed approach accurately performed 79 percent prediction performance, which was a successful result. The performance of this model paves the way for thinking about improving results.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Bullying online is harmful to both bullies and victims and can result in behavioral and mental problems. Cyberbullying could occur accidentally, on purpose, or in a lighthearted way among friends. As a result, a person's daily life is badly impacted, and depression can result in suicide.

5.2 Impact on Environment

A significant issue, cyberbullying has an impact on not just the young victims but also their families, the bully, and innocent people. We must recognize the issues currently confronting today's youth, as it is our responsibility to recognize the environmental elements that might affect a person's well-being. It is becoming more prevalent among adolescents and is now observed in young children. Bullying has many causes and causes, including social environment determinants such as interpersonal, school community, or even policies.

All these can lead to a child becoming a bully or victims failing to report it. Even though everyone is aware that cyberbullying exists and is a significant issue for people, particularly teenagers, many are unaware of how widespread it is. More than half of teens who use social media have witnessed cyberbullying. The majority of them ignored the behavior. Cyberbullying victims are more likely to experience despair, anxiety, and poor self-esteem due to the bullying.

5.3 Ethical Aspects

A collection of values regarding what is ethically right or bad is referred to as ethics. Everyone is unique, and there is no general agreement on where to draw the line between morality and immorality. Sending, distributing, or disseminating unfavorable, damaging,

or misleading information about another person is referred to as cyberbullying. It can also involve revealing private or sensitive facts about another person in a way that makes them feel embarrassed or degraded. Cyberbullying can sometimes take the shape of criminal activity. It is unethical behavior to abuse someone who disturbs other people. Cyberbullying is an issue that regularly impacts youngsters' and students' daily life in around the whole world.

5.4 Sustainability Plan

Cyberbullying detection is challenging, but it is not impossible. We can implement a bullying detection model for online social media. We can educate everyone about cyberbullying. When everybody is aware of cyberbullying, the number of cyberbullying cases can decrease. Also, Government can make rules, regulations, and policies to prevent cyberbullying.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

This research uses machine learning techniques and approaches to detect online cyberbullying and neutral sentences in the Bangla conversation. There are many libraries and resources in English for detecting bullying, whereas there needs to be more research in the Bangla language. We randomly collected our dataset of around two thousand Bangla sentences from Facebook, Twitter, YouTube, Quora, and Reddit. There were two different labels for the dataset bullying and neutral. As a machine learning approach for cyberbullying detection Random Forest, Decision Tree, Logistic Regression, Multinomial Nave Bayes, SVM, KNN, and SGD were used.

6.2 Conclusions

The detection of cyberbullying and neutral sentences from Bangla conversations or sentence data is the expected outcome of this research project. Sentence detection in the Bangla language is challenging. Though, working with some constraints due to a lack of resources took much work. We tried to overcome these challenges. The communication industry is becoming more convenient as technology advances. However, we must acknowledge the progress we have made in ensuring that the vast amount of data is under our control.

6.3 Implication for Further Study

This study's objective is to create a system for assessing and detecting the scope of working with Bangla sentence data. We compared our results with various machine learning models. We detect two categories in our dataset, but on a larger scale, it can detect more things, such as threats, trolls, slang, religion, hate, sexual harassment, speech, and so on. We can use a deep learning strategy for the Big dataset to improve efficiency.

However, the amount of training data restricts the ability to identify patterns of cyberbullying. Therefore, too many cyberbullying datasets are needed in order to improve performance.

REFERENCE

- [1] J. Hani, M. Nashaat, M. Ahmed, 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, pp. 703–707, 2019, doi: 10.14569/ijacsa.2019.0100587.
- [2] L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, "XBully: Cyberbullying detection within a multi-modal context," in *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, Jan. 2019, pp. 339–347. doi: 10.1145/3289600.3291037.
- [3] A. Calvo-Morata, C. Alonso-Fernández, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review," *Comput Educ*, vol. 157, Nov. 2020, doi: 10.1016/j.compedu.2020.103958.
- [4] M. A. Al-Garadi *et al.*, "Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges," *IEEE Access*, vol. 7, pp. 70701–70718, 2019, doi: 10.1109/ACCESS.2019.2918354.
- [5] South Dakota State University, IEEE Region 4, and Institute of Electrical and Electronics Engineers, *2019 IEEE International Conference on Electro Information Technology (EIT)*.
- [6] K. Kumari and J. P. Singh, "Identification of cyberbullying on multi-modal social media posts using genetic algorithm," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 2, Feb. 2021, doi: 10.1002/ett.3907.
- [7] Institute of Electrical and Electronics Engineers. Madras Section and Institute of Electrical and Electronics Engineers, *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*.
- [8] M. F. López-Vizcaíno, F. J. Nóvoa, V. Carneiro, and F. Casheda, "Early detection of cyberbullying on social media networks," *Future Generation Computer Systems*, vol. 118, pp. 219–229, May 2021, doi: 10.1016/j.future.2021.01.006.
- [9] S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10772 LNCS, pp. 141–153. doi: 10.1007/978-3-319-76941-7_11.
- [10] O. Gencoglu, "Cyberbullying Detection with Fairness Constraints," *IEEE Internet Comput*, vol. 25, no. 1, pp. 20–29, Jan. 2021, doi: 10.1109/MIC.2020.3032461.
- [11] A. Bozyiğit, S. Utku, and E. Nasibov, "Cyberbullying detection: Utilizing social media features," *Expert Syst Appl*, vol. 179, Oct. 2021, doi: 10.1016/j.eswa.2021.115001.
- [12] A. Muneer and S. M. Fati, 'A comparative analysis of machine learning techniques for cyberbullying detection on twitter', *Future Internet*, vol. 12, pp. 1–21, 11 2020.
- [13] C. Raj *et al.*, 'Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques', 2021.

- [14] I. Alanazi and J. Alves-Foss, 'Cyber Bullying and Machine Learning: A Survey', *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 18, pp. 1–8, 2020.
- [15] A. Al-Marghilani, 'Artificial Intelligence-Enabled Cyberbullying-Free Online Social Networks in Smart Cities', *International Journal of Computational Intelligence Systems*, vol. 15. Atlantis Press, 12 2022.
- [16] C. Van Hee et al., 'Automatic detection of cyberbullying in social media text', *PLoS ONE*, vol. 13. Public Library of Science, 10 2018.
- [17] M. H. U. Rahman, 'Cyberbullying Detection using Natural Language Processing', *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, pp. 5241–5248, 5 2022.
- [18] A. Ali and A. M. Syed, 'Cyberbullying Detection Using Machine Learning', 2020.
- [19] F. Ahmed et al., 'Cyberbullying Detection Using Deep Neural Network from Social Media Comments in Bangla Language', 6 2021.
- [20] 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, pp. 2163–2169, 8 2021.

Final Test

ORIGINALITY REPORT

24%
SIMILARITY INDEX

20%
INTERNET SOURCES

9%
PUBLICATIONS

14%
STUDENT PAPERS

PRIMARY SOURCES

1 dspace.daffodilvarsity.edu.bd:8080 **9%**
Internet Source

2 Submitted to Daffodil International University **4%**
Student Paper

3 Submitted to Morgan Park High School **2%**
Student Paper

4 www.researchgate.net **1%**
Internet Source

5 www.mdpi.com **<1%**
Internet Source

6 Submitted to BRAC University **<1%**
Student Paper

7 thesai.org **<1%**
Internet Source

8 John Batani, Elliot Mbunge, Benhildah Muchemwa, Goabaone Gaobotse et al. "Chapter 46 A Review of Deep Learning Models for Detecting Cyberbullying on Social