Combating Online Harassment: A Machine Learning Approach for Detecting Abusive Bangla Text

Nusrat Jahan Tandra

ID : 191-15-12383

Supervised By

Raja Tariqul Hasan Tusher

Assistant Professor

Department of Computer Science & Engineering

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

Dhaka, Bangladesh

January 2022

APPROVAL

This Project/internship titled "**Combating Online Harassment: A Machine Learning Approach for Detecting Abusive Bangla Text**", submitted by Name : **Nusrat Jahan Tandra**, Project **ID No :FL1D462,** Student **ID : 191-15-12383** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment 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 date 28-1-2023

BOARD OF EXAMINERS

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

Nazmun Nessa Moon — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Zakia Sultana — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Shamim H Ripon — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that, this project has been done by me under the supervision of **Raja Tariqul Hasan Tusher,Assistant Professor, Department of Computer Science &**

Engineering, Daffodil International University. I 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:

Raja Tariqul Hasan Tusher
Assistant Professor
Department of Computer Science & Engineering
Daffodil International University

Submitted by

Tandra
Nusrat Jahan Tandra

ID: 191-15-12383

Department of Computer Science & Engineering

Daffodil International University

©Daffodil International University iii

ACKNOWLEDGEMENT

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

We really grateful and wish our profound our indebtedness to **Raja Tariqul Hasan Tusher Assistant Professor**, Department of Computer Science & Engineering, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Prof. Dr. Touhid Bhuiyan and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Online harassment such as trolling, threatening, and bullying becomes a serious issue regarding our country's present situation. Millions of people from this country are connected through the internet. Regarding some major social platforms like Facebook, Youtube and tiktlak, abusive content creates enormous public issues including fear. Numerous research has been conducted considering this issues in different languages. however, there were very few research has published on abusive Bangla text detection. In our experiment, we tried to extract abusive comments from Bangla text using several classification algorithms such as (Logistic Regression, Multinomial Naive Bayes,Decision Tree, Support vector Machine ,Random Forest,KNN). The data, we used in this research, was collected from different online social media, forum, Bangla slang books and direct speech. The data set consist 2999 sentences in which 1710 sentences are abusive and the rest of the data are non abusive. In the pre processing part, we have categorized the data into two polarities abusive noted as 1 and non abusive as 0. The data was labeled manually. All the special characters and symbols was removed from the raw data. TF-IDF was used to extract feature from the data. After applying all the algorithms,the experiment has shown that SVM and LR both performed well. they both attained 84% accuracy in which SVM achieved 0.84 precision along with 0.91 recall

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

The Internet is an open-access platform where millions of people can gain access easily. With the popularity of the internet in Bangladesh, more and more people are gaining access to various social platforms and websites. This rapid open access to these platforms creates the issue of online harassment which includes cyberbullying, trolling, threatening which is very concerning compared to the present situation. In our country especially women and minority groups are the top victims of trolling and threatening To prevent online harassment and threat many works have been done previously in Different language. Due to the fact that very few works have been done on the detection of abusive text in social networks in Bangla than in English, we have decided to conduct this relevant study.

In our proposed model we have conducted machine learning approach to accomplished this study. We gathered 2999 Bengali data from different online sources. We mainly focused on Facebook Youtube and TikTok because these platforms has a vast community of Bengali people, from the data of internet we came to know that Facebook has more than 45 million Bengali user and counting, in terms of youtube the number is 34million and counting. After collecting our desired data set, we preprocessed the data set to make an appropriate for Machine learning algorithm.We have labeled our data set as abusive and none abusive denoted respectively 1 and 0. We have used TF-IDF which was required to convert the text in numerical values so that our algorithm can easily detect the polarity of any given data Throughout the whole study, we have used 6 major supervised classification algorithms on our labeled data set to find out positive and negative comments accurately. Our proposed model can have a significant positive impact on reducing abusive content. As our model can detect abusive and none abusive text and also can address them it can help individuals from the threat.

1.2 Motivation

This paper's intention is to address abusive behavior in social media. Humans are social beings, they can not stay alone. But day by day they are becoming dependent on the social network for their communication via messages, chats, and comments. In Bangladesh on November 2022 facebook stats is 92.46%, Youtube 4.39%, Twitter 1.08%, Linkedin 1.02%, Instagram 0.59%, and Pinterest 0.33%. Users share their sentiments and their opinions by posting, readers acquire their opposite opinion by reacting or commenting. As social media is an open platform, one can easily compose offensive, abusive, hateful, insulting, vulgar, attacking, Gross, and Filthy comments. This abusive behavior becomes a research topic for preventing cybercrime such as online harassment, cyberbullying, grooming, sexual abuse, and blackmailing which are becoming the main concern in Bangladesh nowadays. So, the main motive of this study is to compile a unique data set with those toxic comments from social media and develop a classifier model which will detect, whether the comments are abusive or non-abusive rapidly and efficiently.

1.3 Rationale of the Study

The majority of individuals of our country are engaged with various social media where they express their feelings and opinion. sometimes opinion becomes an argument and when it comes to this way they started bullying and attacking others personally which then create a messy situation. Sometimes people used double-meaning word to abuse others, they talk in a complex manner that is not identifiable easily. the use of abusive language on social platforms may lead us to unexpected fear and harm. Many research has been done based on abusive text detection using English language so that there are millions of data available and ones can easily implement their system using these data so that it it becomes easier to tackle the above-mentioned problem however there is a lot of scopes to study in Bengali abusive text detection. We have gathered our data set manually from different sources including Bangla slang books. when we took a look at these books we found numerous rare Bangla slang for example "তেহিসরা মাগী,ছইল কত্তা ছুদানির পুয়া,ঘারি মারানির পুলা,জটের মাল" as we have used a unique data set, we are hoping to get a better outcome

1.4 Project Management and Finance

From the beginning of this project, we maintained a strict time schedule for a particular task. We make up this project in different phases. First of all, we elected our topic and title after that we selected some particular platforms for data collection. Following data collection, we have done our coding portion for this project.

There is no need for financial support for this project as we have done everything on our own but in the future, if we want to implement this project into a website or other kinds of system we may need financial support.

1.5 Objective

1. To construct machine learning algorithm which will be able to detect abusive text in Bengali language.
2. To promote a positive online environment
3. To understand the common pattern of Bengali abusive and none abusive text
4. To improve user experiences in online platforms

1.6 Expected Outcome

In this research, our first and foremost goal is to design a ML model that can accurately identify and classify abusive and none abusive Bangla text with a high degree of precision. Throughout this study, we are trying to get the best accurate outcome by using a supervised machine learning algorithm. People can express their anger and hatred with a double-meaning Bangla word which is very complex and would be difficult to identify. Regarding the issue we collected our desired data manually so that our model can generate the best possible outcome. We expect to gain comparatively better result.

1.7 Report Layout

We organized our study into 6 sections in order to demonstrate the research findings in an accessible manner for the readers. The report is structured to provide a comprehensive overview of the research process and results, making it an efficient resource for those who are eagerly interested in this area of study.

Table 1.1 Report Layout

Chapter	Description
1	This segment gives an introduction to the study.It begins with the research topic and the motivation behind the study then explains the rationale for the study also discusses the desired outcome. At the end of this phase the layout of the report is outlined, which provides a roadmap for the rest of the study
2	This Particular phase contains the background study of the research paper which also include various machine learning techniques and Previous works related to this investigation .this chapter also conducts a comparative analysis to highlight the challenges and limitations of the study.
3	In Chapter 3, the research methodology is described in detail. This includes a detailed explanation of the algorithmic specifics for each of the algorithms used. This chapter serves a comprehensive overview of the technical aspects of the research and how they were implemented
4	This segment represents the result analysis of each and every machine learning algorithm which were used in this research paper that also includes Experimental environment, discussion
5	Chapter 5 highlights the ethical aspects and sustainability of this research paper which includes impact of this research to our society
6	This section will discuss the possible future direction of this study and how it can be expanded in the near future. It also suggests some recommendations regarding the study at the end it concludes by summarizing the key findings. Overall offers a comprehensive summary of the entire study.

CHAPTER 2

LITERATURE REVIEW

2.1 Preliminaries

In this research, we sought to address the lack of research on abusive text detection in Bengali language by developing supervised machine learning algorithms to accurately detect abusive comments in this language. While there have been numerous studies on abusive text detection in other languages, there is a shortage of research on this topic specific to Bengali language. Given the prevalence of abusive language on social media platforms in Bangladesh, addressing this issue is of great importance. However, collecting and annotating data for this study was a challenge due to a lack of resources. Through our research, we hope to contribute to a better understanding of abusive language in Bengali language and to the development of effective algorithms and techniques for detecting it.

2.2 Related Works

A growing body of research has been done on using machine learning to find abusive text in different languages. In this project, we have focused different approaches including the use of natural language processing techniques, lexical analysis, and supervised learning algorithms. In this section, we will be examining some of the key studies that have been conducted in this area and also discuss how they have developed their tools and techniques for detecting abusive text

Sarker, Manash, et al. [1] Proposed supervised machine learning techniques, along with N-grams, Gated Recurrent Unit model to detect abusive Bengali text, and they got highest accuracy of 80.51 % in MNB.

Emon, Estiak Ahmed, among others [2] have implemented a machine learning and deep learning model where they used 4 supervised ML algorithms and 2 deep learning-based algorithms named ANN, RNN to detect multiple types of abusive text in Bangla. They compared the results of the two models and discovered that the RNN based on beep learning performed better. Banik, Nayan, et al. [3] worked on abusive Bangla toxic comment classification based on five supervised machine learning algorithms and CNN,

LSTM. They used word frequency for feature extraction where Both of their deep learning-based algorithms got the best accuracy of 94.13% and 95.30%

In order to detect abusive text, Awal, Md Abdul, et al. [4] applied the Naive Bayes model. The model's effectiveness was assessed in their study using cross-validation. Direct and Google translation are used to translate English into Bangla, and "Youtube.com" is used to gather the English dataset.

Hussain, Md Gulzar, et al. [5]Also works on this relative are. From their conducted study we came to know that they utilized a manually developed algorithm that has two distinctive parts one is for training another is for classifying comments as abusive and non-abusive. Bags of word model introduced to extract feature. Data collection had been done manually which consist of 300 Bengali comments.

Lucky, Effat Ara Easmin, et al. [6] recently worked on sentiment analysis on child abusive pubic comments where they introduced six major machine learning algorithms such as MNB, RF, DT, SVC, KNN, LR, and two deep learning algorithms namely RNN, BRT. By comparing two distinctive area they find out BERT outperformed and from supervised algorithms MNB gave the best result.

Akhter et al. [8] Focused on different supervised ML algorithms while working on social media bullying detection. In their monograph, they collected data by using java tools from online platforms. Bangla and English datasets were conducted throughout the whole study. They have also utilized 10-fold cross validation in both bangla and English data set to separate non bullied text from none-bullied so that they could get more precise result. In both language, they got a remarkable result in SVM they that is above 95%

Romim, Nauros, et al. [9] To address unusual speech detection in the Bengali language, they collected 30000 user data from crowdsourcing. They divided their data set into seven different classes and have done multi-classification by using a combination of deep learning and machine learning approaches. Word2Vec, FastTest, and BengFastText had been used to extract features.Islam, Md Manowarul, et al. [11] Working on Cyberbullying detection. Throughout the study, they performed multiclassification to detect and categorize Bengali abusive text. A combination of ML and GRU has been

introduced for the classification of user comments. TF-IDF, character N-gram feature has been used to extract feature.

2.3 Comparative Analysis and Summary

As we worked on a unique data set, which we managed to collect and labeled manually we believed that our model can give a better outcome. Much research has been raised to tackle the issue of abusive text. But most of them used annotated data sets rather than creating their own. From some previous works we came to know that they have utilized variety of supervised and unsupervised ML algorithms, but we used supervised learning because this type of algorithm can give the best outcome based on unique data set. We have also gathered some rare slang by interviewing some native Bengali speakers which might help to get more precise result.

2.4 Scope of the Problem

Abusive language detection in Bangla texts involves identifying and flagging texts that use language to harm, intimidate, or demean individuals or groups. This is a significant problem as abusive language can cause serious harm, including mental health issues, promoting violence and discrimination, and contributing to a negative online environment. To solve this issue we can use our proposed model, which can help identify and take action against abusive texts by deleting them or blocking the users who posted them. We believe that our proposed model will be able to detect abusive text in our native Bengali language. Which can promote a healthy and safe environment for everyone.

2.5 Challenges

We gathered our data manually from different online platforms and forums where people mostly engaged. While we were collecting data, we faced some issues such as some sentences were very confusing either it can be abusive or none abusive for instance “বয়স তো মাশাআল্লাহ্ ভালই কিন্তু পিচ্ছি পিচ্ছি লাগে,ভালো হয়ে জা,সামনে তো হিমালায় পাহাড়” at the same time Different people and cultures may have different opinions on what is considered abusive, and it can be difficult to create a comprehensive and objective definition of abusive language. As there is not much work available in terms of the Bengali language we felt a lack of related resources, though it was not very easy to search and collect thus data.

CHAPTER 3

RESEARCH METHODOLOGY

3.1. Research Subject

Our title of the topic is “Combating online harassment” that involves machine learning techniques to identify and classify abusive text in the Bangla language. Most of the people of our country are engaged with various online platform such as Facebook, Tik Tok , You tube, Instagram etc. Through these platform online harassment are being occurring continuously. In these platforms,there is no restriction about commenting and posting in Bangla language so that and people can post whatever they want ,that might abuse someone or terrify targeted individuals Online harassment such as cyber bullying, threatening,hate speech is a serious issue regarding our country’s present situation that can have significant negative impacts on the mental health and well-being of individuals. Therefore, our main aim is to remove vulgarity from our Bengali communities through the above-mentioned platforms in this case machine learning algorithms can help to automatically detect abusive text in online platforms, such as social media or online forums, and provide alerts to users or moderators to take appropriate action. To tackle this problem we implemented six major supervised machine learning algorithms to detect and classify user comments as abusive and non-abusive. As our main aim is to find out whether a comment is abusive or none-abusive, we focused on binary classification

3.2 Instrument

To conduct our research in this area we have collected 2999 Bangla comments from v arious online platform such as Facebook ,You tube , Tik Tok as well as Instagram. To evaluate and justify these comments we have applied. Six major ML classification algorithms to asses whether the given data is abusive or not , as we are doing binary classification so that the supervised algorithm has the best match

3.3 Data Collection Procedure

As more and more people in Bangladesh acquire access to the internet and social media platforms, online harassment is becoming a bigger issue. Cyberbullying, stalking, and online threats are just a few of the various forms it can take . It becomes a major problem of our country a lot of work to done to prevent this type of crime. That's why ,we decided to do this abusive text detection research regarding Bengali language. Unfortunately There has been less research done in Bangla about the detection of abusive text in social networks, compared to how much has been done in English therefore data set is not much available to do this . Therefore we determined to collect data manually from online platforms .we have collected our data from various social platforms such as Facebook, Instagram, Youtube, Tik Talk and some other Bengali slang books named "Bangla Gali Obhidhan" We have also used Youtube comment scrapping python tools to extract user Bengali comments for this study we have managed to collect 2999 Bengali text including both abusive and none-abusive text among them 1710 sentences are abusive and 1289 sentences are none-abusive text. Throughout this data collection process, we have encountered some difficulties in finding appropriate sentences to use for our analysis. Despite these challenges, we believe that our research on abusive text detection in Bengali language will be valuable in helping to address this important issue in Bangladesh.

An illustration of our data set is given below :

Table 3.1 the cleaned data of our data set

Comments	Remarks
ইসলাম হলো চোদার ধর্ম	1
খানকির পোলাগোরে জুতাপটি করা দরকার	1
হাত কেটে দিবো তর	1
জারজের বাচ্চা জারজ	1
আমি রোমিন আপুর জন্য সব সময় দোয়া করি সত্য কথা বলার জন্য ।	0
তোরে এক রাতের জন্য ভাড়ায় নিবো মাগী	1
ভাইয়া আপনার কন্ঠ অনেক সুন্দর আপনার জন্য দোয়া রইলো আমার	0
তর হেডা দেখাইলে ও কেউ সিডিউস হইব না মাতারি	1
আল্লাহ তাকে হিদায়েত দান করুক ।	0

3.4 Statistical Analysis

The total data that we have gathered for our research will be illustrated in the following figure which also includes abusive and none-abusive data

Table 3.2 Size of the data set

Total Data	2999
Abusive Data	1710
None-Abusive Data	1289

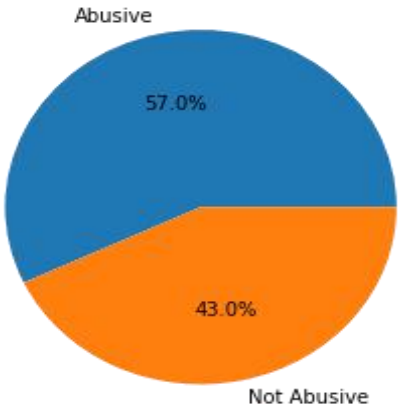
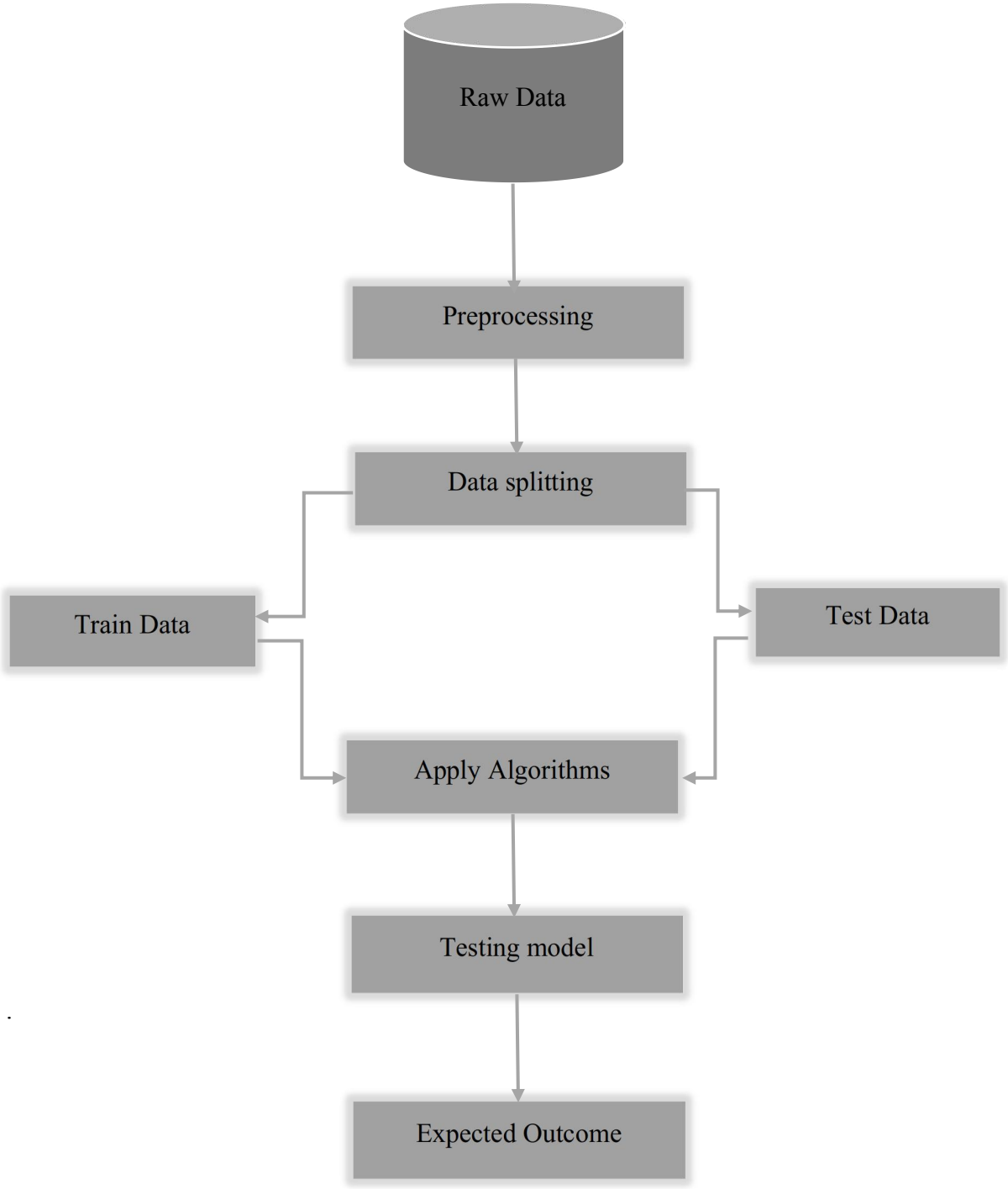


Figure 3.1 Graphical representation of our data set

3.5 Proposed Methodology

Our proposed methodology for detecting abusive text in Bengali language involved collecting data from various online platforms and Bengali slang books. After collecting raw comments, we pre-processed the collected data and manually labeled each piece of text as either containing abusive language or not. We also split the labeled data into a training set and a testing set, and used the training set to train a machine learning model to detect abusive text. Our main motive was to follow the supervised learning approach and apply a classification algorithm for detecting abusive text in Bengali language. We experimented with different model architectures to find the best model. We then tested the trained model on the testing data to evaluate its performance, and refined the model as needed based on the results. This combination of manual data labeling and machine learning techniques allowed us to create a robust model for detecting abusive text in Bengali language.



.

Figure 3.2 overall Methodology

3.6 Pre-processing

Data pre-processing is the fundamental and most important part to find out valuable insight from raw data.It is a method of raw data transformation. Raw data may have many irrelevant and missing values in this part data cleaning take place so that we can get meaningful insights for our research. It is also regarded as the topmost important step in terms of machine learning process because it can have significant impact on the performance of overall machine learning approaches. To conduct this project we gathered our desired data form various online sources, for instance, Facebook, Youtube, Tik Talk, and Bengali slang books , we collected our data manually but we also used youtube comment scrapper tool to extract Youtube comments. After collecting data, we stored thus in several Google sheet as partial data set then we encountered unnecessary and noisy data from our datasets. we have removed unnecessary special character

(@,!#$%^&*></':;' +_~`) also eliminated special symbols, emojis, English word and punctuation from every single comment. We have also separated irrelevant sentences from our required sentences and eliminated them so that we might get a better outcome. In our dataset we have two columns one is marked as comments another is remarks, we have labeled our data set and in the remarks column, we have denoted every possible abusive text as 1 and none-abusive text as 0. After cleaning and labeling all the data we have merged every partial data set into one final dataset and saved as csv format.

	comments	remarks
2954	আপনাকে অনেক অনেক অভিনন্দন প্রিয় ভাই আমার	0
1263	মাশাল্লাহ, আপু আপনার কণ্ঠটা অনেক সুন্দর	0
2700	এইগুলাকে সফট পর্ন বলে	1
1271	কাপড় খুলে একটা ভিডিও বানাও ভালো লাগবে তাহলে	1
905	ওর বোদায় বালু ডালো আর সাবোল মারো	1
...	...	...
1452	সেক্স করবেন আপু আমার সাথে	1
51	তোমার এগুলো ঢেকে রাখতে পারো না। এখন তো এগুলো অ...	1
546	মহান আল্লাহ পাক তাদেরকে হেদায়েত দানকরুন	0
2938	আপনার মত এক জন রাজনৈতিক নেতা আজ সংসদ দরকার	0
2803	তর মারে জিগাইস তর বাপ কয়ডা	1

2999 rows × 2 columns

3.3 Illustration of an insightful data set

3.7 Feature Extraction

After pre-processing the desired data set it requires to extract features. It is a process of identifying relevant and informative to use in machine learning algorithms.There are several algorithms that could be used for feature extraction such as Bag of words, Term Frequency Inverse Document Frequency, and Word Embeddings. Among them, Bag of words works by compiling a vocabulary of words that appear most frequently in the data set. in the data set, and representing each document as a numerical vector indicating the presence or absence of each word in the vocabulary, Term Frequency Inverse Document Frequency is a sub-variant of Bag of words which works by considering the rarity of a word across a corpus of documents. It can capture the importance of these specific words or phrases in identifying abusive language. which might be appropriate for supervised learning model. In our data set, we have employed TF IDF for extracting features

3.8 Training the Model

Training a machine learning model involves optimizing the model's parameters to fit the training data as accurately as possible. This is typically done by minimizing a loss function, which measures the difference between the model's predictions and the true labels or targets in the training data. For our research, we split our data set into a test set and a train set. We used 20% of the data for testing and the remaining 80% for training. To optimize the performance of our model, we applied six different supervised classification algorithms: Logistic Regression, Random Forest Tree, Decision Tree, Multinomial Naive Bayes, Support Vector Machine, and K-neighbors classifier. The results of these algorithms were then compared to determine the most effective approach for our specific data set and task. It is important to carefully evaluate the performance of different algorithms and select the one that provides the best results for the specific problem at hand.

3.9 Algorithms

Throughout this research, we have used several classification algorithms to detect and classify Bengali abusive and non-abusive text. We carefully selected and pre-processed our data, and used a range of supervised learning algorithms to train our model and make predictions. A brief summary of outperformed algorithm is given below

3.10 Support Vector Machine

Support Vector Machine (SVM) is a very useful technique for detecting abusive text in the Bengali language. It operates by locating the hyperplane in a high-dimensional space that most effectively distinguishes abusive from the non-abusive text. One of the key strengths of SVM is their ability to handle high-dimensional and non-linearly separable data, making them particularly effective for tasks such as detecting abusive text in large datasets. In addition, they are known for their robustness and ability to generalize well to unseen data. However, it is important to carefully tune the hyperparameters of an SVM model in order to achieve optimal performance in detecting abusive text in Bengali language, as they can be computationally intensive. Despite the potential for increased training time and resources, SVM have proven to be a reliable choice for many abusive text detection tasks.

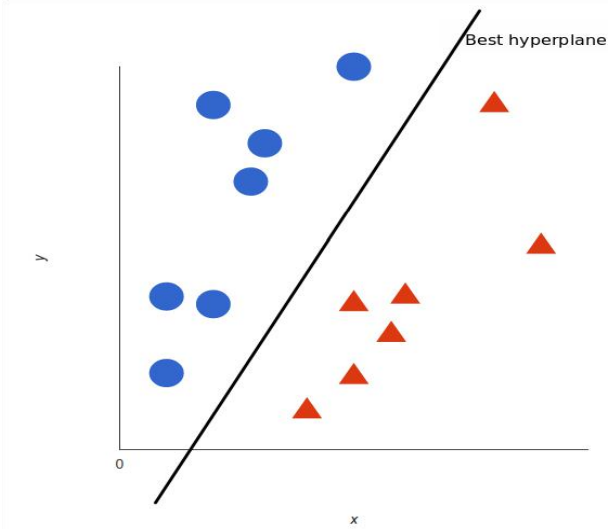


Figure 3.4 Support Vector Machine

3.11 Logistic Regression

Logistic regression is a widely-used machine learning technique for classification tasks. It is a statistical model that estimates the probability of a binary outcome - an outcome with only two possible values - based on the values of predictor variables. For example, logistic regression could be used to predict the probability that a customer will churn

based on their past behavior or the probability that a patient has a certain disease based on their medical history. This model is considered to be linear, meaning that it makes predictions based on a linear combination of the input variables. Logistic regression is often used in situations where the outcome is binary and the predictor variables are continuous or categorical. It is a straightforward and effective method for making predictions in these scenarios, and is often used as a baseline model for comparison with more complex machine learning models.

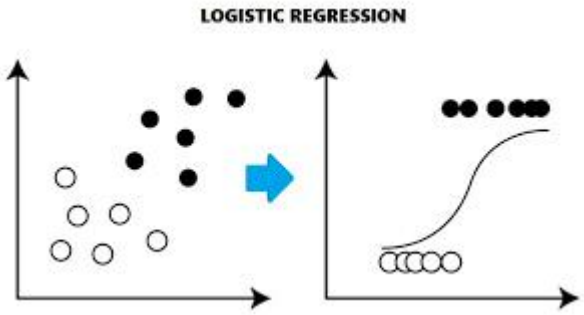


Figure 3.5 Logistic Regression

3.5 Implementation Requirement

To carry out this study we have utilized a variety of tools and online platforms which are given below

1. Google Co Laboratory
2. Python
3. MS Excel sheet
4. Google Sheet
5. Notepad
6. You tube comment extractor tool
7. Several online platforms

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

For the evaluation of this study, we gathered in total of 2999 Bangla data. We targeted mostly visited websites, social media, and forums to extract data. The main focus have been given on Facebook, Youtube, Tiktok and a Bengali slang book named “Gali Obhidan” the data set consist of 1710 abusive sentence and 1289 none abusive sentences. We saved our data set in Google drive. Data labeling and cleaning procedure are done manually then we created our proposed model with 80% of the data in the train set and 20% of the data in the test set.

4.1 Experimental Results & Analysis

We have utilized 6 major ML algorithms to examine the test data set. We observed that logistic regression and SVM got the highest accuracy compared to the other four algorithms. Multinomial Naive Bayes and Random forest tree also performed very well but could not overtake SVM and Logistic regression. Logistic regression attained an accuracy of 84.83% at the same time SVM also got the same accuracy which is also 84%. Random forest tree and MNB has a relatively high accuracy of 82.16% and 82.33%. On the contrary, the decision tree model has a lower accuracy of 77.66% and following this Kneighbors classifier got the lowest accuracy which is 72.16% from the scenario it can be said that DST and KNN could not perform as per our expectation. Here,we can see that Logistic Regression and Support vector machine was very effective considering our data set, this is because these algorithms are well suited for binary classification and can generate precise result from unique data set. These algorithms can also handle high-dimensional data effectively.DST algorithm did not suited our data set because this algorithm requires large data set, even larger than our data set.

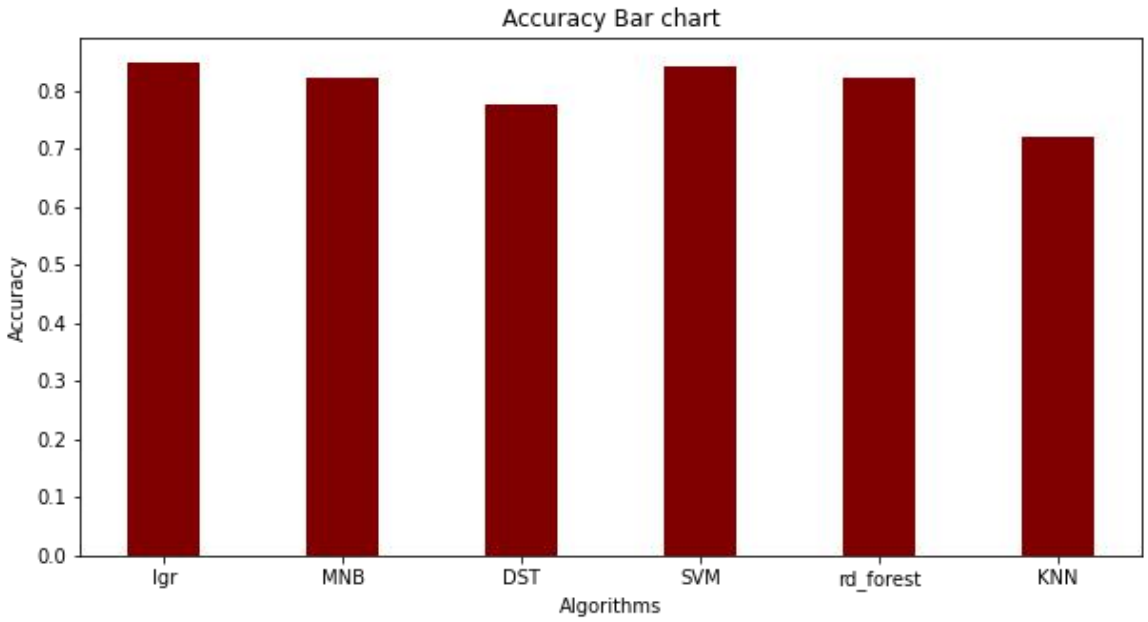


Figure 4.1.1Illsutrating Accuracy of algorithms

Though SVM and Logistic gave the best outcome but its not enough to evaluate the study as the there may have some biased data so that we have also measured precision, recall and f1 scores.

Table 4.1 Evaluation measurement

Precision	Recall	F1 score
0.84	0.91	0.87

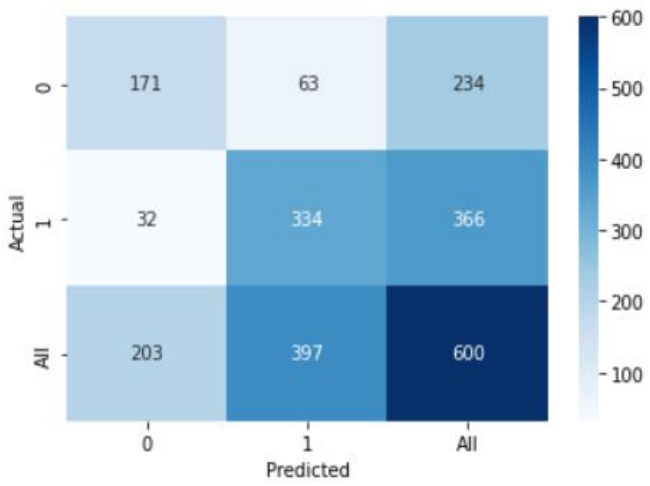


Figure 4.2 Confusion Matrix

4.1 Discussion

The results of our experiment showed that logistic regression and SVM were the most effective algorithms for detecting abusive text in Bengali language, while decision tree and kneighbors had lower accuracies. These findings are consistent with previous research, which has also found that logistic regression and SVM are well-suited for binary classification tasks and can generate accurate results with unique data sets.To overcome some drawbacks we need to do further study in future.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Online harassment is a serious issue in any country including Bangladesh. People can easily harass, and threaten people through online by being anonymous. Which may affect the victim very badly. It is particularly prevalent on social media platforms, where it can be difficult to control and regulate. It is often targeted at vulnerable groups, such as women and minorities, and can have a disproportionate impact on these groups. In this case, we think that automated abusive text detection has the potential to make a positive impact on Bangladesh's online communities. We believe that by identifying and flagging abusive text this approach will build a hospitable and safe online environment in our country. In addition, the development and implementation of this type of work can contribute to the growth and advancement of the field of artificial intelligence in Bangladesh

5.2 Impact on the Environment

No direct environmental effects are caused by our study. pact however but if we consider technological terms there may have some impact for instance, to train and test data set requires some computational power which may indicate energy consumption which is also ignorable.

5.3 Ethical Aspects

We gathered our information from numerous social media sites with a sizable Bengali presence. We have kept in mind about user’s privacy while collecting thus data that is why we collected our data anonymously. No user should have a concern about their safety. Therefore, there was no privacy issue with data collection at the same time we carefully considered the source and diversity of training data and took necessary steps to minimize potential biases in our outcome.

5.3 Sustainability

To ensure the viability of our research project we need to continually improve and update the machine learning algorithms and other technologies that have been used. This may involve incorporating new techniques such as deep learning or artificial intelligence.

In the upcoming future, there are much more possibilities for getting more precise result by using updated and new technology. As detecting Bengali abusive text and cutting them off from our online community is the main motive of this study,we believe that this project can be used in

many different online sectors in our country. It can be apply any website where there are Bengali opinions or comments exist. It can also be applicable to Facebook, Twitter, Youtube, Tiktalk, and so on. If this project can be applied to a website nobody can harass or threaten over that website or platform because this automation will detect abusive comments so that the admin panel will be able to take necessary steps against the particular abuser

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

We utilized approximately 3000 Bengali data from different online sources and labeled them manually and also maintained a standard proportion of abusive and none abusive sentences after that we applied six ML-supervised algorithms and checked for accuracy. Logistic regression and SVM gave us the best outcome whereas MultinomialNB and Random Forest tree also gave us a satisfactory result. Also measured precision, recall, and f1 score for the best of the study. For the nature of our data set SVM and Logistic regression performed outstandingly.

6.2 Limitations

- Limitation of Data
- Context-dependent definitions
- Textual complexity
- Bangla is an Evolving language

6.2 Conclusions

In this study, we proposed a methodology for detecting abusive text in the Bengali language using a combination of manual data labeling and ML techniques. By collecting and labeling a data set of Bengali text, we aimed to train a ML model that could accurately detect abusive language and help to create a safer and more positive online experience for all users. We tested the performance of several machine learning algorithms on this data set, including logistic regression, support vector machine, random forest tree, decision tree, MNB, and kneighbors.

6.3 Implication for Further Study

1. Enrich the data set by adding more data
2. Plan to add phonetic Bangla data set to our model
3. Build a platform
4. Implement an application that can check for abusive content on any online platform
5. Apply Deep learning algorithm

REFERENCES

[1] Sarker, Manash, et al. "A Machine Learning Approach to Classify Anti-social Bengali Comments on Social Media." *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*. IEEE, 2022.

[2] Emon, Estiak Ahmed, et al. "A deep learning approach to detect abusive bengali text." *2019 7th International Conference on Smart Computing & Communications (ICSCC)*. IEEE, 2019.

[3] Banik, Nayan, and Md Hasan Hafizur Rahman. "Toxicity detection on bengali social media comments using supervised models." *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE, 2019.

[4] Awal, Md Abdul, Md Shamimur Rahman, and Jakaria Rabbi. "Detecting abusive comments in discussion threads using naïve bayes." *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. IEEE, 2018.

[5] Hussain, Md Gulzar, and Tamim Al Mahmud. "A technique for perceiving abusive bangla comments." *Green University of Bangladesh Journal of Science and Engineering* (2019): 11-18.

[6] Lucky, Effat Ara Easmin, et al. "An Attention on Sentiment Analysis of Child Abusive Public Comments Towards Bangla Text and ML." *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2021.

[7] Hussain, Md Gulzar, Tamim Al Mahmud, and Waheda Akthar. "An approach to detect abusive bangla text." *2018 International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE, 2018.

[8] Akhter, Shahin. "Social media bullying detection using machine learning on Bangla text." *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*. IEEE, 2018.

[9] Romim, Nauros, et al. "Hate speech detection in the bengali language: A dataset and its baseline evaluation." *Proceedings of International Joint Conference on Advances in Computational Intelligence*. Springer, Singapore, 2021.

[10] Islam, Md Manowarul, et al. "Cyberbullying detection on social networks using machine learning approaches." *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. IEEE, 2020.

[11] Ishmam, Alvi Md, and Sadia Sharmin. "Hateful speech detection in public facebook pages for the bengali language." *2019 18th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2019.

[12] Eshan, Shahnoor C., and Mohammad S. Hasan. "An application of machine learning to detect abusive bengali text." *2017 20th International conference of computer and information technology (ICCIT)*. IEEE, 2017.

[13] Perera, Andrea, and Pumudu Fernando. "Accurate cyberbullying detection and prevention on social media." *Procedia Computer Science* 181 (2021): 605-611.

[14] Ali, Aaminah, and Adeel M. Syed."Cyberbullying detection using machine learning." *Pakistan Journal of Engineering and Technology* 3.2 (2020): 45-50.

[15] Reynolds, Kelly, April Kontostathis, and Lynne Edwards. "Using machine learning to detect cyberbullying." *2011 10th International Conference on Machine learning and applications and workshops*. Vol. 2. IEEE, 2011.

[16] https://www.ebookmela.co.in/download/%E0%A6%97%E0%A6%BE%E0%A6%B2%E0%A6%BF-%E0%A6%85%E0%A6%AD%E0%A6%BF%E0%A6%A7%E0%A6%BE%E0%A6%A8-%E0%A6%B8%E0%A6%AE%E0%A7%8D%E0%A6%AA%E0%A6%BE%E0%A6%A6%E0%A6%95-%E0%A6%86%E0%A6%AC%E0%A6%A6%E0%A7%81

[17] https://www.google.com/search?q=support+vector+machine&source=lnms&tbm=isch&sa=X&ved=2ahUKEwi596acw-T8AhW0SmwGHYXvAtYQ_AUoAXoECAEQAw&biw=1536&bih=732&dpr=1.25#imgrc=Eb8sR_htyu8OeM

Document Viewer

Turnitin Originality Report

Processed on: 02-Jan-2023 21:08 +06
ID: 1987975004
Word Count: 6917
Submitted: 1

Tndra By Tndra Tndra

Similarity Index	Similarity by Source	
21%	Internet Sources:	19%
	Publications:	11%
	Student Papers:	14%

exclude quoted | exclude bibliography | exclude small matches | mode: quickview (classic) report | print | refresh | download

3% match (Internet from 21-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

2% match (Internet from 21-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

2% match (student papers from 05-Jan-2022)
Submitted to Daffodil International University on 2022-01-05

1% match (Internet from 21-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

1% match (Internet from 19-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

1% match (M. F. Mridha, Md. Anwar Hussen Wadud, Md. Abdul Hamid, Muhammad Mostafa Monowar, M. Abdullah-Al-Wadud, Atif Alamri. "L-Boost: Identifying Offensive Texts from Social Media Post in Bengali", IEEE Access, 2021)
M. F. Mridha, Md. Anwar Hussen Wadud, Md. Abdul Hamid, Muhammad Mostafa Monowar, M. Abdullah-Al-Wadud, Atif Alamri, "L-Boost: Identifying Offensive Texts from Social Media Post in Bengali", IEEE Access, 2021

1% match (student papers from 03-Dec-2022)
Submitted to Indira Gandhi Delhi Technical University for Women on 2022-12-03

<1% match (Internet from 11-Dec-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 20-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 21-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 25-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 21-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 04-Jun-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 21-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 21-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 17-Nov-2021)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 20-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 03-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 04-Apr-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 29-Sep-2022)
http://dspace.daffodilvarsity.edu.bd:8080

<1% match (Internet from 21-Nov-2022)