CLASSIFYING THE CYBER BULLYING BENGALI WORDS FROM SOCIAL MEDIA COMMENTS USING MACHINE LEARNING ALGORTIHMS

BY

JOY SARKAR
ID:191-15-12979

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

Supervised By

Nishat Sultana
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
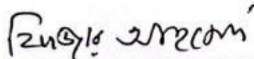
DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/internship titled **"Classifying the cyber bullying Bengali words from social media comments using machine learning algorithm's"**, submitted by Joy Sarkar, ID No:191-15-12979 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 *28-1-2023*.

BOARD OF EXAMINERS

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

Chairman

Dr. Fizar Ahmed
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Taslima Ferdaus Shuva
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md Sazzadur Rahman
Associate Professor
Institute of Information Technology
Jahangirnagar University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Nishat Sultana, Lecturer, Department of CSE** Daffodil International University. We 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:

Name: Nishat Sultana
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Name: Joy Sarkar
ID: -191-15-12979
Department of CSE
Daffodil International University

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 **Nishat Sultana**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of *Machine Learning*" 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.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Professor**, 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

Social media, which provides numerous options for contact, but raises the hazard and frightening circumstances online, particularly for the young people worldwide. The key to effective mitigation is recognition system of potentially hazardous communications. Although much study has been done on English-language material, non-English material, particularly Bangla content, has traditionally been neglected. The use of subscriber data evaluation and machine learning techniques to address these difficulties in comments online or chats shows improved results. Many machines attempting to learn strategies are suggested in English language and literature. However, resolving the problem by using the approaches at hand is rather subject specific, therefore false identification might happen if the materials change from formal English to profanity or snark. Additionally, the political and social behavior of the research group and the linguistic variances between English and non-English material may affect performance. This study investigates how well-known machine learning techniques perform and how accurate they are when applied to Bangla text. Additionally, the influence of subscriber data, such as location, age, gender, the number of likes and comments, etc., is examined for bullying in Bangla. Support Vector Machine (SVM), which has an accuracy rate of 88.0% according to experimental data, is the most effective algorithm for detecting bullying in Bangla. However, when integrating subscriber data with user-posted data, KNN(3- Nearest) obtains the best accuracy with 77.8%, while in the same situation, Support Vector Machine (SVM) reaches 88% performance, which is quite near to the best one. Due to SVM's superior performance in both scenarios, it was decided to use SVM when applying the model to digital networking. As a consequence of the program, a java internet alternative has been created, and according to the reliability of the system, majority of the time the results are comparable to those of a distant analyst.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

TABLES	PAGE NO
Table 1: Sample of annotated data	14
Table 4.2: Algorithm	27

CHAPTER 1

INTRODUCTION

1.1 Introduction

The technology known as Natural Language Processing (NLP) helps computers understand the natural language spoken by people. The use of natural language processing (NLP), a subfield of machine learning, in interactions between machines and people language. The ability to read, decode, grasp, and gain a deep understanding of human languages is a clear objective of NLP. Lexical analysis, syntactic analysis, text analytics, conversation integration, and pragmatic analysis are the typical five processes in the processing of natural language [3]. The process of text processing entails recognizing and examining word structures.

A idiomatic lexicon refers to its vocabulary, or its set of words and expressions. The full collection of texts is broken down into sections, phrases, and words by reading passages. When doing a semantic analysis (parsing), the words in the phrase are examined for syntax and arranged to demonstrate the relationships between the components. The English grammatical analyzer, for instance, ignores the line "I go to my school." The Bengali syntactic analyzer similarly disapproves of the phrase "বষ �ের েরাে দ �ােবনর সৃ�� হয়" Semantic analysis extracts a word's dictionary definition from the content. In those other words, the text is examined for significance. Compositional organization and assignment domains item mapping is used to accomplish this. The semantic analyzer ignores phrases like "hot ice-cream" in English or " "আ�ুল ফু েল কলাগাছ" in Bengali. When a statement's meaning depends on the importance of the sentence that comes before it, this is referred to as discourse integration. It also understands what the immediately next statement means. The information transmitted is again decoded to determine what it actually meant during the pragmatic analysis process. It involves figuring out which linguistic constructs call for accurate experience or knowledge in the real world. Computers can comprehend text, hear voice, decode it, gauge sentiment, and determine which sections

are crucial thanks to the natural language processing (NLP) [4]. Language translation software like Google Translate makes extensive use of NLP [2]. NLP is used by word processing programs like Microsoft Word and Grammarly to verify the syntax of documents [2]. In a similar vein, we utilized NLP to evaluate the tone of the Bengali text. This enables us to determine the extent to which the statement contains a bully expression.

Cyberbullying, also known as online abusive behavior, is the use of internet communication to threaten a person, usually by sending them messages that are scary or incriminating. The use of obscene and hostile language has considerably expanded in the age of social media and internet networking. These remarks foster a disrespectful atmosphere in online [5]. Cyberbullying wasn't appropriately addressed or ignored in previous years. The professional networking platform's poor user involvement was cited as the cause, and it was advised to screen off or disconnect in the case that one received abusive comments. However, the current state of affairs has radically altered. According to a 2019 survey, 70 of the 100 women who experience online harassment are between the ages of 15 and 25. Bullying and harassment account for 18% of the claims and instances of harassment brought before the sole cybercrime tribunal in the nation [6]. Finding vulgar and insulting language when it occurs on online forums and then reporting these occurrences to Bangladesh in order to track out the real-world perpetrators of such behavior are two important problems in the fight against cyberbullying. A mechanism to intuitively and magnificently separate hostility and instances of online provocation is not integrated into any current online system or world wide web social networks (such as Facebook and Twitter). This important problem wasn't considered to be worth exploring in the past since it wasn't real, but it is now at a hazardous stage. This influence on the digital stage cannot be ignored, which is why it has taken center stage in studies looking at effective solutions. To regulate this movement against online harassment, experts and cybercrime agencies must give it serious thought [5]. Therefore, the aim of this work is to identify objectionable terms and phrases used in Bengali that are deemed to be online harassment on social media platforms.

1.2 Motivation

Nowadays, spending a lot of time on social media sites like Facebook, Instagram, Twitter, and YouTube is one of the most popular digital pastimes. By 2022, there will be about 4.74 billion users of social media, representing a 49% annual growth rate. People use social media for 144 minutes on the average each day. Both the pros and drawbacks of this enormous rise in social media use are enormous. Cyberbullying is on the increase on numerous social media platforms, which is one of the most significant negative effects.

The purposeful and recurrent harming of victims via electronic media is referred to as cyberattacks. More than 80% of kids use social media websites and own cellphones, and 57% of those kids reported having experienced cyberbullying. Additionally, 60% of kids and teens have seen bullying on social media. The freedom to access internet resources is compromised by this terrible experience, which also has severe psychological implications [3]. Cyberbullied individuals have a 1.9 times higher risk of suicide and also suffer from cognitive issues including autism (75%), somatic disorders (70%) and learning difficulties (52%).

The need for cyberbullying avoidance has grown as a result of the rise in cyberbullying. Finding instances of cyberbullying can be a crucial step in its prevention. Cyberbullying-related messages can be avoided from being remarked on if they are identified early enough.

1.3 Rationale of the study

Cyberbullying may be easily found using classification techniques based on machine learning. Machine learning models have demonstrated their effectiveness in prediction and recognition over time. There is a plethora of studies that uses machine continuing to learn forecasting and detection accessible. The use of Bangla on social media is expanding as a result of the acceptance of the Unicode system and the expansion of Internet usage.

Due to a dearth of annotated corpora, lexical definitions, and grammatical analyzers, which necessitate in-depth research from Bangladesh's perspective, relatively few works on Bangla text for social media activity monitoring have been done. Additionally,

implementing the approaches is extremely content sensitive, so if the contents shift from formal English to profanity or humor, false identification may result. Additionally, linguistic disparities between English and non-English material and the affective behavior of the research group may influence quality.

1.4 Research questions

One of today's most crucial issues is identifying cyberbullying.

- Comprehending and understanding current categorization models and methods.
- Evaluating the applicability of current different classifiers to the detection issues facing today's society.
- Machine learning methods comparison of several methods.
- Working on a technology that might identify Cyberbullying from comments alone will assist us in managing several AI-related applications.

1.5 Expected Output

It might be difficult to identify bullies on the Facebook site for a variety of popular languages. This is due to the variety of languages and the various ways that individuals use or write them. In this research, we have concentrated on identifying bully terms for Bengali language on Facebook platform. Our goal is to accurately determine if a user-inputted comment is a bully expression or not. In addition to identifying bullies, we can also say what category and how much it fits into that group. To identify the many types of statements that have been entered, we have created five labels: quasi, sexually, threatening, trolling, and religious. The dataset utilized for this study is current and contains a sizable quantity of data, allowing us to get results with a high degree of precision. Our algorithm required more time to train and occasionally produced false-positive results for relatively lengthy words. This paper's shortcomings can be fixed for future projects.

1.6 Project Management and Finance

There is no such a finance because collecting data and all work done by me. It such a hard work to collect data from social media on Facebook comments.

1.7 Report Layout

Our research idea is based on the fact that Bangladesh is a digital country and every person use internet. In this platform such as Facebook, twitter, YouTube etc.

In Chapter 1, we have already talked about the initiation of the research papers, the objective, the reasons why we are working on this subject and the reasons why we are choosing these particular fruits and vegetables, problem statements, the expected results, the research question, and research methods.

In chapter 2, we'll talk about the history of the study and the relevant tasks that must be completed for this article This part is all about our study summarization. We also talk about the difficulties we have completed our task effectively and evaluate against the concepts we've selected.

In chapter 3, the approach of this research is provided. With the help of an appropriate graphic and table, we will quickly outline our whole coding procedure in this part.

In chapter 4, the experimental findings and discussion are covered. There will be several testing and research models.

Chapter 5 is all about Summary, comparison and analysis, Advantage of our application are covered.

Chapter 6 is all about conclusion and future research.

Chapter 7 Here all the references we used for this research.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

Cyberbullying has increased dramatically over the past ten years, and it doesn't just happen in English; it also happens in other languages. Because cyberbullies thrive in large crowds, it is crucial to recognize cyberbullying in varied linguistic contexts. As a result, the suggested model that we are going to create would aid in identifying cyberbullying materials in Bengali, one of India's more vocal local languages. In this language, very few studies on cyberbullying have been conducted. The cyberbullying approaches utilized in earlier English-language studies will be the ones we employ here. Although effectiveness and performance may vary due to semantic differences between English and non-English content, this analysis assumes the use of machine learning techniques and the consideration of user data for identifying online harassment on Bengali material in order to combat such problems.

2.2 Related works

In the area of naturally occurring language processing, significant effort has been made to introduce various methods for handling text data. In several works of literature, the polarity of textual information was estimated [7, 27]. Accordingly, they separated sentiment classification into three sections: document level, phrase level, and entity and aspect level. A score was determined using the favorable and unfavorable dictionaries after data cleaning, compression, and stemmed. We can determine from the end result whether the input sentence was favorable, negative, or neutral. They were effective in giving the sentences supplied as input favorable, negative, or value 0, which led to a final review of the content. Based on the focal phrase and circumstances, a better baseline method for sentiment classification was presented [8]. They learned what to pay particular attention to in a phrase and how to manage with the word's dynamic emotion. Another research suggested an effective system for assessing social comments and determining whether they

were hostile or not using NLP and algorithms for machine learning [9]. The key element of their final conceptual model that aids in the detection of social media-based cyberbullying media is an efficient classifier. In this example, SVM Classifier and Logistic Regression Machinery outperformed Logistic Regression and Random Forest Classifier trained on the feature stacking. Due to their widespread usage in a variety of contexts and languages, many offensive terms and phrases are left out of dictionaries. In the literature, two novel hypotheses for feature extraction were presented that may be helpful in differentiating cyberbullying [10]. They put together a model that anticipated comments classified as bullying or not. Standardization, background subtraction from baseline data, feature extraction from new data, feature selection, and classifying are the associated processes. They create feature vectors for conventional feature extraction by using N-grams, enumeration, and TF-IDF scores. The final result is the probability that the comment is disparaging against the members. Results indicate that their hypothesis increases accuracy by 4% and could be employed to separate comments directed at peers.

The efficiency of sentiment utilizing CNNs and SVM was confirmed by the effective improvement in the accuracy of sentiment categorization from text when compared to traditional CNN [11]. automated, uncontrolled, and tractor trailer approaches are used to categorize emotions. SVM served as an automatic emotional classifier, CNN served as an automatic image converter, and pre-trained phrase vector served as the input. The average accuracy of the combination CNN-SVM model was shown to be much greater than that of the SVM and CNN alone. Another study [12] attempted to create word representations depending on lexicons to forecast text emotion. The construction of a sentiment lexicon involved the use of language model, followed by the assessment of the dipole moment of the sentiment analysis from a set of data of comment threads, and eventually the utilization of naive Bayes to categorize the characterized features on a sizable dataset and user reviews from the app store. The vocabulary created using word2vec has greater and simpler accuracy and recall rates than the lexicon created by PMI. Due to the fact that many of the sentimental terms used in the responses were uncommon, the average recall was quite low.

To determine whether model works better, a recent study employed two distinct types of datasets to identify hate speech using RNN [14]. Dataset-A has a little quantity of data, but

Dataset-B contains more than three times as much data. From a broad perspective, using Dataset-B may result in improved performance. According to their findings, the results were satisfactory while utilizing a small dataset. And when more data was used for their tests, good results may be produced by employing deep convolutional neural network. Additionally, they used some data taken from the raw data at a ratio of 10% to 90% as testing data, and discovered that Regression Analysis performed better generally than SVM. The performance of TF-IDF was more apparent while employing SVM. According to the most current experimental development, using BIRNN could lead to better outcomes.

To examine the opinions of many merchants and forecast the common opinion in foreign currency markets, content analysis combining Nave Bayes and a lexicon-based algorithm was utilized [19]. Based from the results, they were able to reach a 90% accuracy rate. Another research [21] classified tweets' opinions about a certain product as good, negative, or neutral.

Text Blob, which is based on Stanford Natural Language Toolkit, was used to analyze the text information, and Colonnaded Bayes, which classifies the text, was employed (NLTK).

A feed-forward neural network was given the training data, and its output layer decided the ultimate orientation. The accuracy of the findings is then calculated using a discriminant function, and it is determined to be between 79 and 87%. In order to exclude undesired URLs, tags, and stop words, research was done after data from Twitter was collected using Glove word embedding. Later, the words were labeled using POS tagging. After that, the likelihood was evaluated based on the polarity using the MSP framework for essential piece sentiment analysis. Depending on the outcomes, the accuracy was 74.66%, demonstrating that the MSP model improves accuracy in comparison to formerly employed neural networks models. A thorough investigation focused on using sentiment classification on a mixed population of multiple languages. For categorization, they employed datasets from Telugu movie reviews and a Bengali-English collection that were both run through a single layer of CNN. They achieved a 73.2% accuracy in the mixed Bangla dataset using aggregating and washout batch normalization, and a 51.3% efficiency in the Telugu movie reviewing dataset.

2.3 Comparative Analysis and Summary

We attempted to use text classification techniques to identify cyberbullying in Bengali. Even though we have used a variety of text-based different classifiers, including SVM, Logistic Regression, Passive Aggressive, and Random Forest, other machine learning models or techniques, like CNN and even Natural Language Processing, may be used in the coming years for the input data we have been working with.

2.4 Scope of the Problem

Bullying is not a recent concept, and as the significant communication tools switched to digital platforms, cyberbullying emerged on its own. Social media, such as Facebook, Twitter, and others, clearly allow users to connect with anybody else, at any time.

Social media has replaced traditional social gatherings as the main venue for social contact nowadays. It also makes it possible to start new relationships in an anonymous manner. Interpersonal media's transparency makes it more likely that kids and teens will encounter harmful scenarios, such as recruiting or sexually inappropriate behavior, signs of sadness and suicide thoughts, and cyberbullying. Social media is a simple technique for bullying to targeted their victims outside of the playground because users may use it at any time (24/7) and can remain private. The loss can be reduced if the main keywords is identified early. Due to linguistic variations and reliance on several social media factors (such as age, number of likes, number of comments, etc.), manual or current keyword-based recognition methods are insufficient for Bangla text. Consequently, there is a chance to develop an application that solves this issue and satisfies the criteria for machine learning-based social media bullying detection on Bangla text.

2.5 Challenges

Programs for social media networking and online talking offer a platform for anybody to share the information, skill, and grow their network, but a small number of users abuse this opportunity by threatening other users with harassment assaults, which makes it problematic to utilize these platforms. Finding someone manually who publishes abusive remarks predicated on racism and sexism takes a lot of work in today's online world. This

research uses several machine learning algorithms on the Video Datasets to determine if a particular opinion is a bullying comment or not.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

Our research subject is Find the cyber bullying Bengali words from social media comments using machine learning algorithms. The suggested cyberbullying detection system's development process was separated into two phases:

Phase 1: Identifying the optimal machine learning method for Bangla Text after analysis and evaluation

Phase 2: Using the top-performing model found in Phase 1, developing the software system for an efficient cyberbullying detection technique.

This stage involves setting up an immersive experience for assessing the most well-known algorithms for machine learning in the research using accuracy, precision, recall, and F-1 score. The information from social media set will be retrieved in this experimental setting, followed by data cleaning, model training, and model building for several machine learning algorithms. To select the optimum machine learning method for the Bangla and English datasets, the model was validated and assessed in the final step.

3.2 Data Collection Procedure/Dataset Utilized

One can assess outcomes and react to applicable questions by gathering and analyzing information on specific characteristics in a current framework. The goal of data collection is to amass convincing proof that will allow analysis to produce answers that are convincing and believable to the problems at hand. So, the procedure of collecting, categorizing, and 22 evaluating technique that detects for study following acknowledged required for proper is known as representative information gathering. Using the information, they have obtained, a critical evaluation of the material. The initial and most important phase of any study research is often data collection. Assuring that accurate data is gathered for quantitative tests so that scientific conclusions may be informed by

information is the most important goal of data collecting. The essential component for creating predictive results for cyberbullying is a proper dataset. This section describes the creation of two corpora, one in Bangla and the other in English, both of which contain social media posts.

According to Statista's research of the largest and most used online communities globally by active users (July 2019), which used data from the Global Web Index panel, Facebook has the most active members (in millions). It controls the vast majority of the global market share with more over 2 billion active users. Second place goes to Google's YouTube, followed closely by Facebook-owned WhatsApp and Messenger.

Figure3.2 ranks Twitter as the 11th most successful social networks in the world with more than 330 million active users
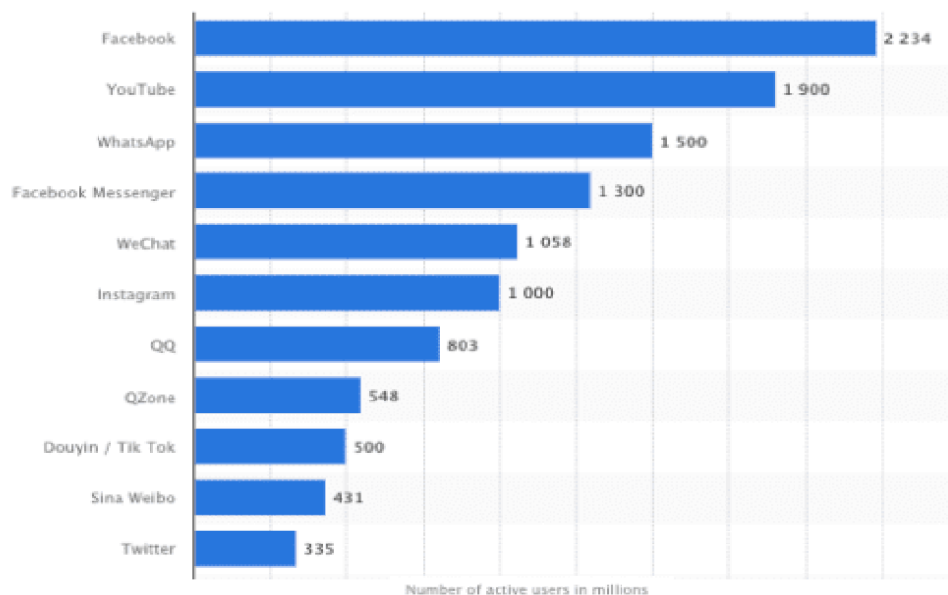

Fig 3.2 No. of Active user in million in Jul-2019[9]

The illustrated demarcation of individuals using social media in Bangladesh is shown in Figure-3.2.

According to a report by gs.statcounter.com, Facebook is the most widely used social networking platform in Bangladesh. Twitter follows YouTube in terms of Bangladeshi users with 1.32%.
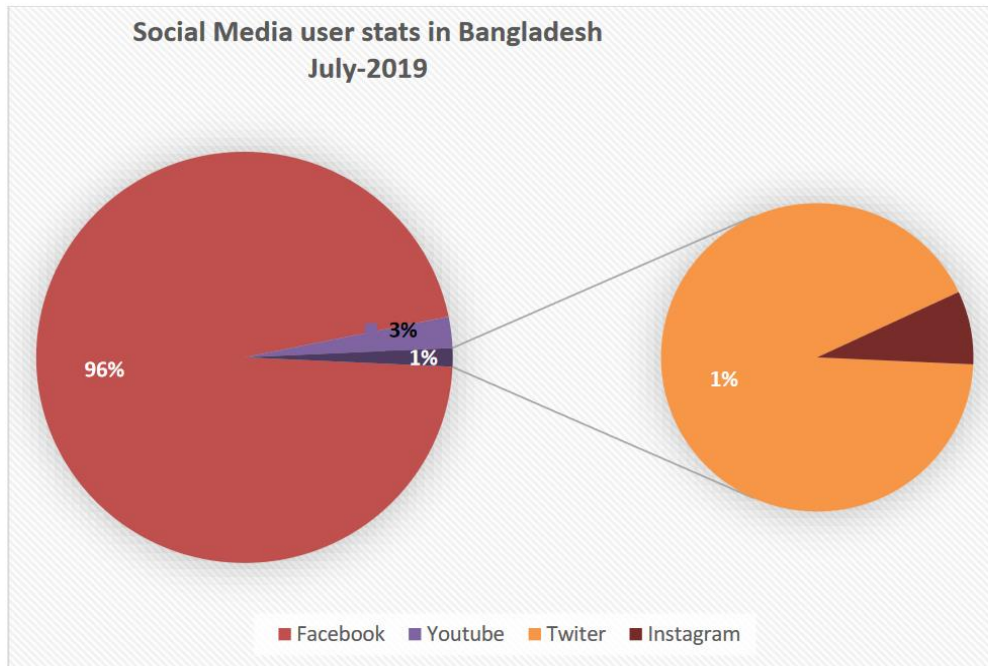
Figure 3.3: Social media user in Bangladesh-July, 2019[6]

Figure 3.3: This picture taken by google. According to the abovementioned data, Facebook is the most widely used social network both internationally and in Bangladesh. Twitter and YouTube also play a big part in the online world. Due to this reality, the main sources of social networking media accounts used were Facebook, YouTube, and Twitter.

By gathering information from many social networking sites, like Facebook, Twitter, YouTube, etc., two corpora were created. Through the use of the Google video Perspective Discussions Analysis software API (developers.google.com/apis-explorer), Twitter Api Server (developer.twitter.com), and Instagram Graph API (developers.facebook.com), the information was gathered by browsing public media broadcasts and conversations.

For the purpose of crawling and retrieving social media platforms, a Java application has been created. For the purpose of gathering examples of Bangla and English material, the system employed web scraping on well-known social media platforms including Facebook, YouTube, and Twitter. Twenty thousand public status updates in English and fifteen

thousand in Bangla, as well as certain emoticon and user-related data, have all been collected using the Facebook Graph API. Additionally, the user's Twitter demography data has been gathered via the Twitter REST API. Last but not least, a human observer has been used to classify each item as having been bullied or not. Additionally, the bag of words approach was used to create the classification model.

Table 1 Sample of annotated data

	Data	Situation	spam
0	থামবেল ঠিকক্রে বানা	cyberbullying	1.0
1	মাদারচোদ রানু খানকিমাগী তোর গোটা গুষ্টি চুদী খ...	cyberbullying	1.0
2	তুই বেশি কথা বলিস এজন্য তোর এই অবস্তা	cyberbullying	1.0
3	এগুলো সুনলে মাথা খারাপ হয়ে যায়	cyberbullying	1.0
4	ছিঃ ছিঃ লজ্জা করে না রানু মন্ডল একদম ফালতু পায...	cyberbullying	1.0
...	...	...	...
3994	প্রাচীন পল্লির এই ছিল রুপ	Non cyberbulling	0.0
3995	শহর থেকে গ্রামের বাড়ি চলেছি একটি জরুরি কাজে	Non cyberbulling	0.0
3996	আদিকাল থেকেই পল্লিগ্রাম ছিল দেশের প্রাণকেন্দ্র	Non cyberbulling	0.0
3997	বাংলাদেশের গ্রাম বললেই সবুজ-শ্যামলে জড়াজড়ি ক...	Non cyberbulling	0.0
3998	চারিদিকে এক রকম ধু-ধু প্রান্তর	Non cyberbulling	0.0

3.3 Statistical Analysis

We commenced the preparation by purging our datasets of any duplicate information. The following step involved removing all numerals, emotions, capitalization, links, user tags, unified resource location (URL), extended words, and user mentions from all three datasets. In order to get accurate findings, several of the sentences that combined Bangla and Romanized Bangla were taken out of the dataset. Additionally, because the messages were primarily in the language of the country, we did not do any stop part of speech tagging or stemmed on our information.

The database used in this research consisted opinions from the engagement section under posts made by celebrities on the Fb page, including actresses, singers, legislators, as well as other media personalities [29]. There were a total of 4125 reviews. 68.1% of comments are targeted against female victims, while 31.9% are targeted at victims of rape, according to our data. Additionally, 5.98% of remarks are directed towards politicians, and 21.31% are directed at victims who are social influencers. The dataset's graph is described below.

Non-bully: Statements that aren't meant to personally attack someone are generally considered non-bully. For example, এদের কষ্টের কথা শুনছেন ভাল কথা তার পর কিছু টাকা দিয়ে আসতে পারতেন, কিছু না করে

Sexual: This category includes remarks that promote sexual animosity or harass someone in a sexual manner.

For instance: পার্নারশীপ আছে নাকি

Threat: Comments posted by people that make threats to hurt or kill somebody person fall under the threat category. For instance: থুর চোতমারানী মাগী তুই জীতিয় বেইমান"

Troll: This category refers to remarks made by users in an offensive manner to ridicule another individual. For instance: "এই অপদাথ � মিহলার চেহারা দেখলই যেন বিম আেস।".
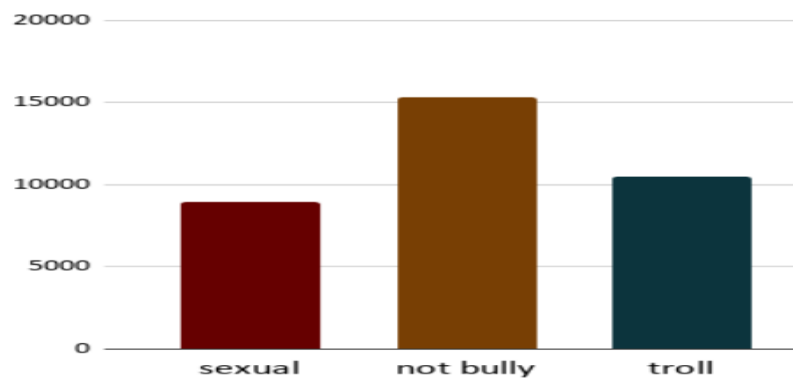

Fig 3.3 Graphical representation

3.4 Proposed Methodology/Applied Mechanism

3.4.1 Machine learning classifiers

Many people use machine learning classifiers to forecast categorical variables. Today, a variety of intelligence systems that facilitate decision-making are built using machine learning. Unmonitored, supervised, and reinforcement learning are all included under the umbrella term of "machine learning." Four unsupervised machine learning classifiers were utilized in this study.

3.4.2 Sklearn. Naive_Bayes. Multinomialnb

Multivariate regression models using naive Bayes classifier. With identified, the multivariate regression Naive Bayes classifier is appropriate (e.g., word counts for text classification). Usually, the standard deviation requires decimal characteristic values. In truth, though, component counts like tf-idf might also work.

3.4.3 Sklearn.Metrics.Confusion_Matrix

To assess a classification's effectiveness, construct a discriminant function. A multiple regression is by necessity one in which the number of observations that are both known to be in a subgroup and projected to be in a group, C I j, are equal. The effectiveness of a classification method is frequently calculated in machine learning using the discriminant function.

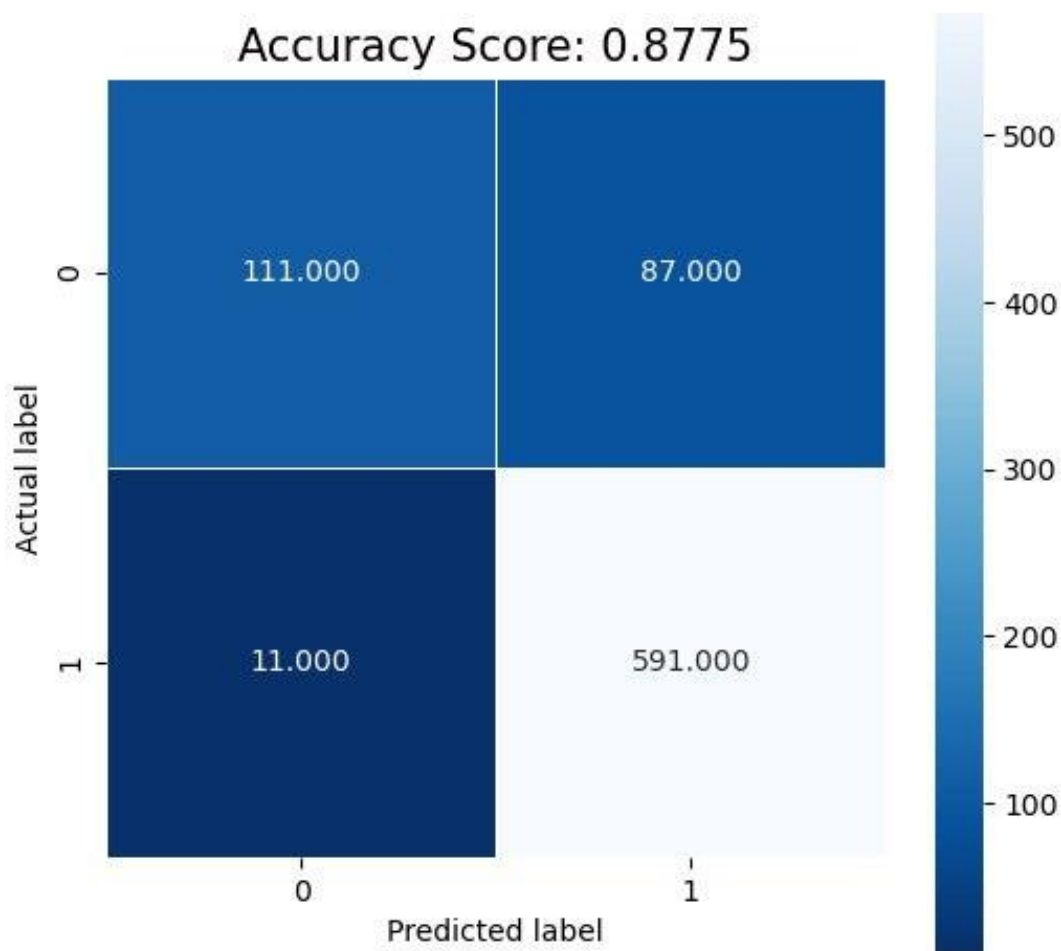|	precision	recall	f1-score	support
0.0	0.96	0.44	0.61	198
1.0	0.84	0.99	0.91	602
accuracy			0.86	800
macro avg	0.90	0.72	0.76	800
weighted avg	0.87	0.86	0.84	800

Fig 3.4.3 classification report

Fig 3.4.3 Confusion matrix

3.4.4 Multinomial naive Bayes

An NLP favorite for Bayesian training is the multiple linear regression naive Bayes approach. The Bayes hypothesis is used by the program to predict [24]. It determines the value with the highest likelihood utilizing a particular sample's review and adjust and produces (4)

$$P(c|x) = P(x|c) * P(c) / P(x)$$

3.4.5 Logistic regression

Logistic regression is a popular predictor for classification model. Logistic regression employs a sigmoid activation function to categorize data. Every real number around 0 and

1 may be converted using the function [26]. In (5), the sigmoid activation function is displayed.

$S(z) = 1/1 + e^{-z}$

The method transforms the numbers it returns into 0 or 1. A threshold value is established to do this. Numbers over the barrier are categorized as class 1, while values below it is categorizedasclass0.

3.4.6 Random Forest classifier

The randomized forest classifier is a method for communal tree-based categorization. The random forest classification algorithm consists of a set of selection trees selected from a small portion of the trained data that is picked at randomly. It integrates the opinions from various evaluation trees to establish the testing materials evaluated and the results.

	precision	recall	f1-score	support
0.0	0.96	0.44	0.61	198
1.0	0.84	0.99	0.91	602
accuracy			0.86	800
macro avg	0.90	0.72	0.76	800
weighted avg	0.87	0.86	0.84	800

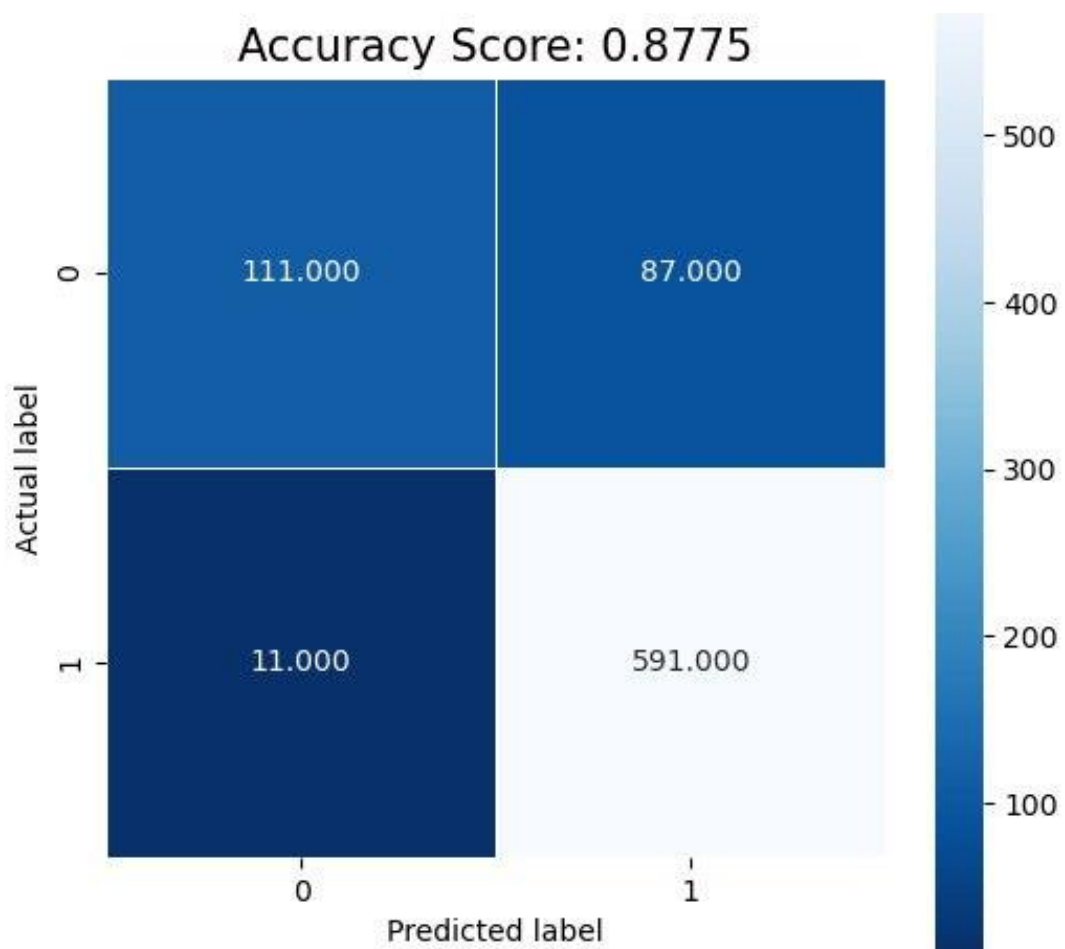Fig 3.4.6 Random Forest classifier report

Fig 3.4.6 Confusion matrix

3.4.7 Decision Tree Classifier

A class called Decision Tree Classifier may do number of co categorization on a database. The classification will forecast the class with the smallest index among such classifications if several classes have the same and highest likelihood.

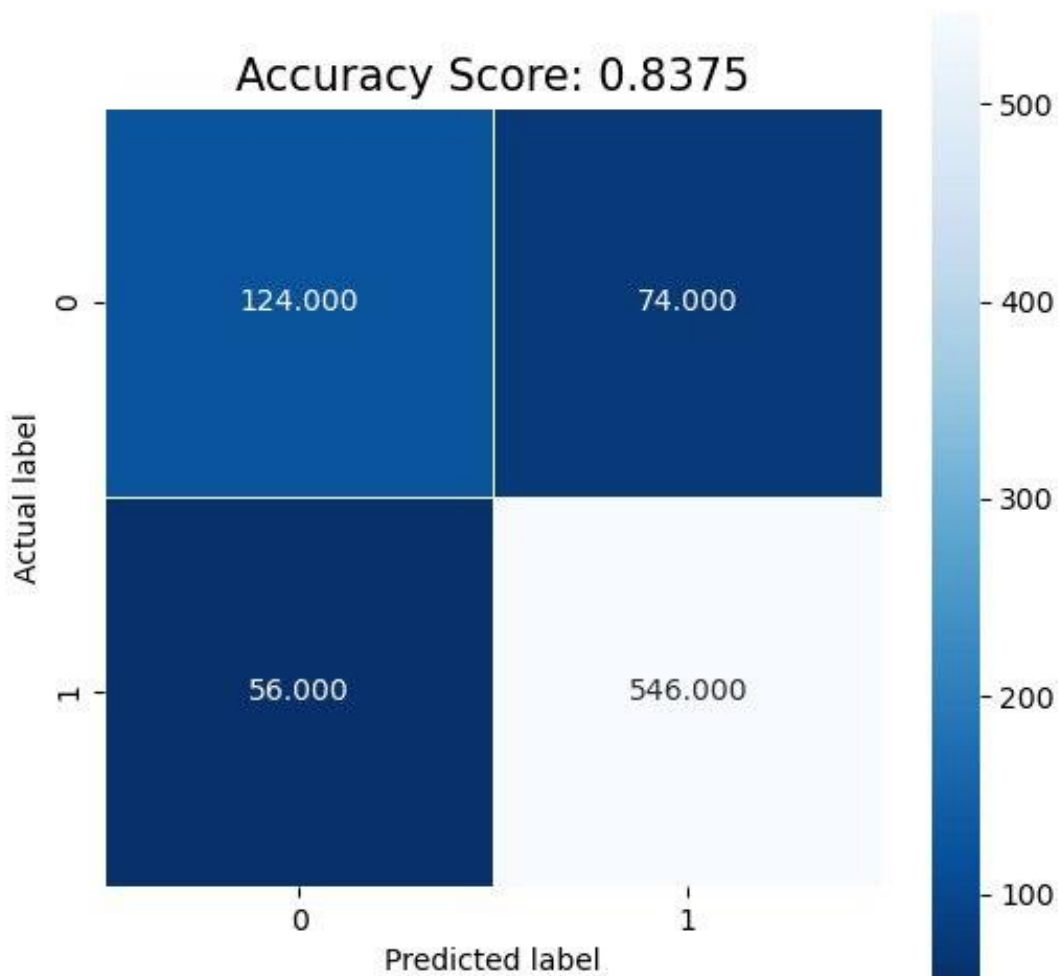|	precision	recall	f1-score	support
0.0	0.69	0.63	0.66	198
1.0	0.88	0.91	0.89	602
accuracy			0.84	800
macro avg	0.78	0.77	0.77	800
weighted avg	0.83	0.84	0.83	800

Fig 3.4.7 Decision Tree Classifier report

Fig 3.4.7 Decision Tree Classifier confusion matrix

3.4.8 SVM

Support Vector Separator, a method of supervised learning, is commonly used exclusively for classifiers. By translating the pieces of data to a strong environment and then identifying the optimal higher - dimensional feature environment, SVC divides the array into different groups.

Through the application of kernel functions, SVM may be utilized to resolve non-linear issues. For particular, data points can be mapped onto a higher dimensional space using the well-known RBF (radial basis function) foundation, allowing the information points linearly separable. Following the mapping of the sample points, SVM will locate the ideal hyperplane in the newly created environment that may classify the data points into two groups.

	precision	recall	f1-score	support
0.0	0.69	0.63	0.66	198
1.0	0.88	0.91	0.89	602
accuracy			0.84	800
macro avg	0.78	0.77	0.77	800
weighted avg	0.83	0.84	0.83	800

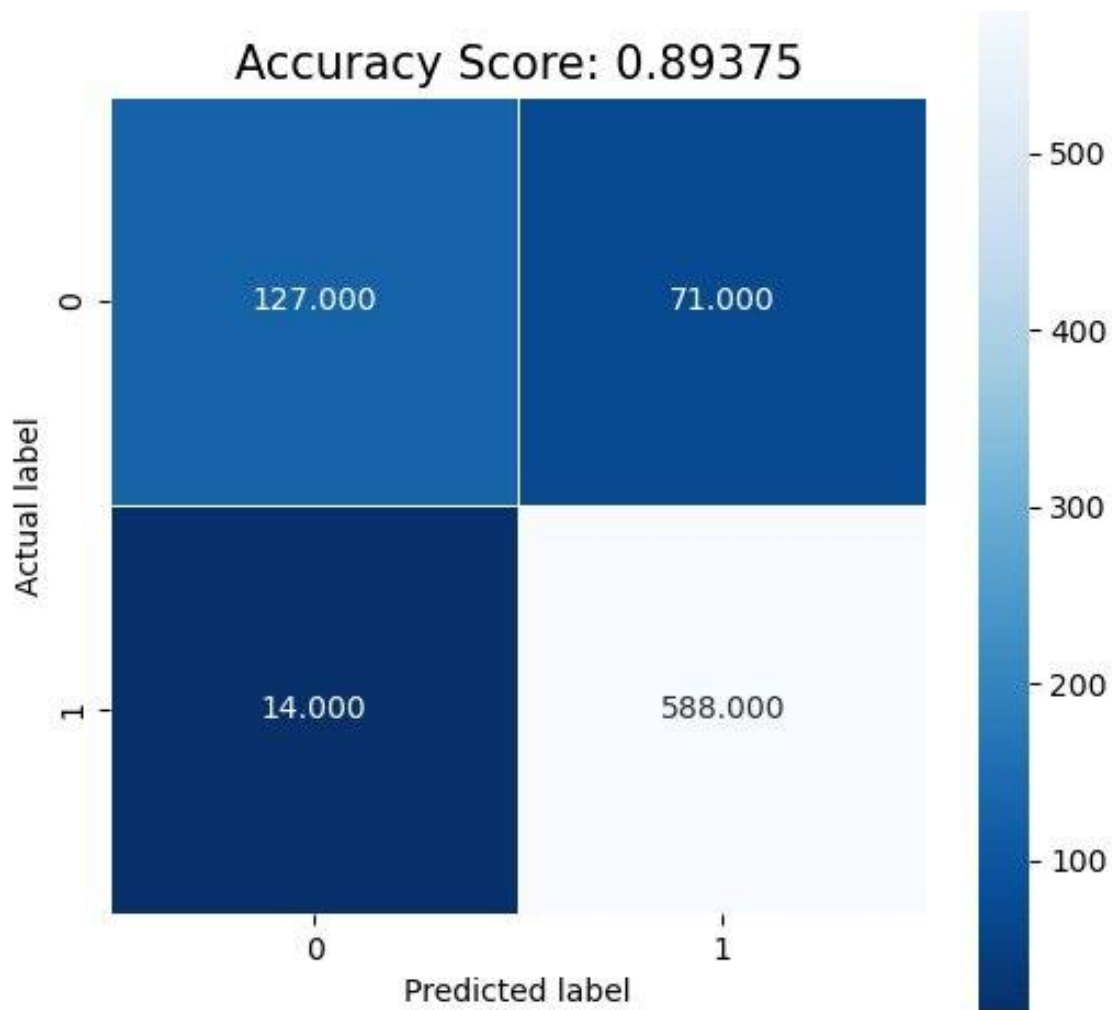Fig 3.4.8 SVM classification report

Fig 3.4.8 SVM confusion matrix

3.4.9 KNeighbors Classifier

RadiusNeighborsClassifier performs classification system is based on all neighborhood elements within a given radius, r, of the target point, t, whereas KNeighborsClassifier provides objective and accurate on voting by the closest k-neighbors of the target position, t.

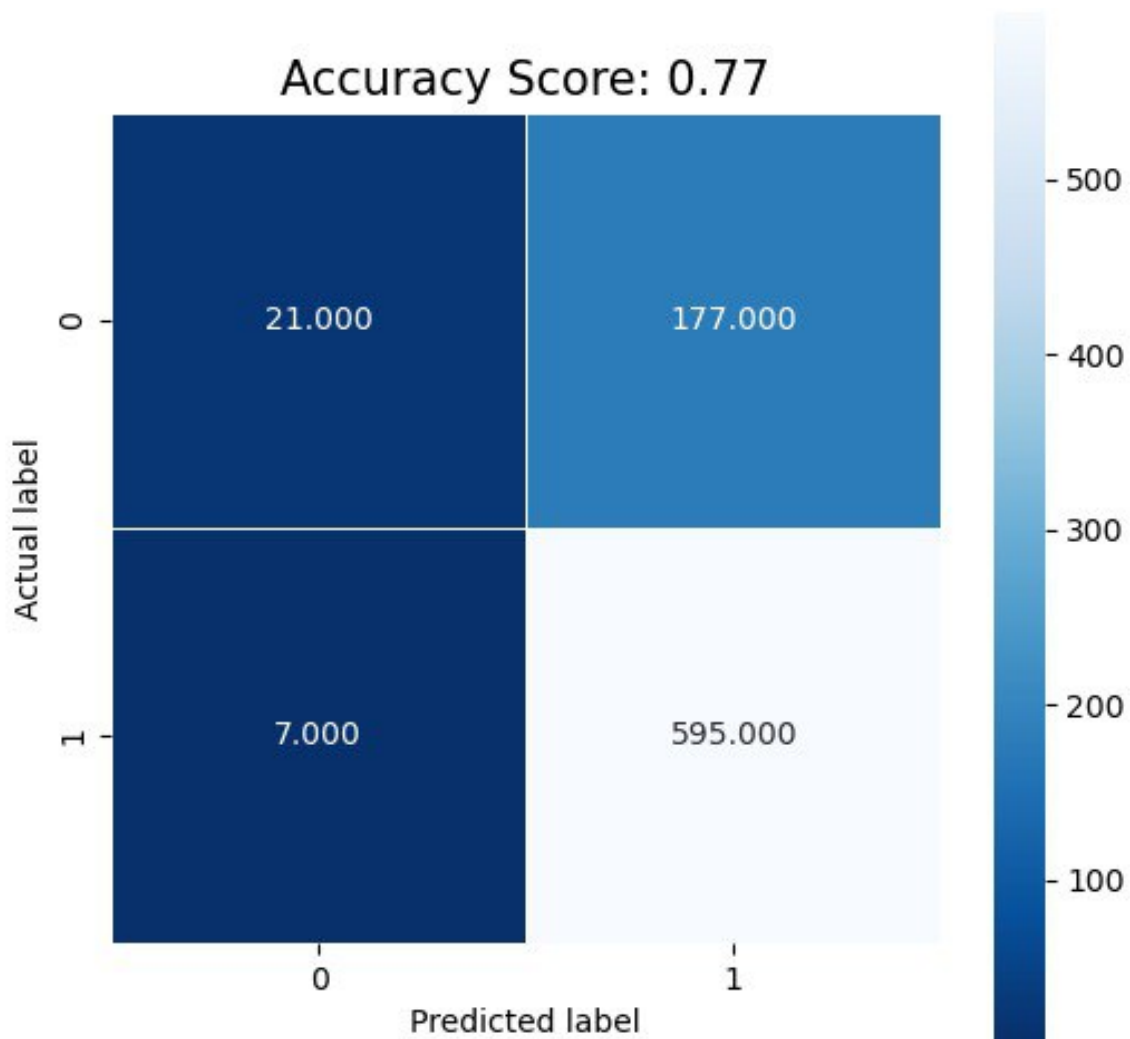|	precision	recall	f1-score	support
0.0	0.75	0.11	0.19	198
1.0	0.77	0.99	0.87	602
accuracy			0.77	800
macro avg	0.76	0.55	0.53	800
weighted avg	0.77	0.77	0.70	800

Fig 3.4.9 KNeighbors Classifier report

Fig 3.4.9 KNeighbors Classifier confusion matrices

3.5 Implementation Requirements

In our research we mainly used

KNN which accuracy is 77.8%, Logistic regression 88%, Random Forest 87.87%, DT is 82.8%, SVM which accuracy is 88% and NB 86.6%

The project initiation process in the SDLC is the most significant and important. The leading employees on the team carry it out using data from the customer, the representative, consumer research, and industry experts. That use this information, the comprehensive project methodology is then created, and the economic, organizational, and intellectual aspects of the initiative are then evaluated for preliminary viability.

It is also designed to detect project hazards and make new plans for the requirements of quality control during the strategic planning. The technology techniques that can be employed to complete the assignment efficiently and with the minimum risks are defined in the project feasibility original study findings.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

Performance appraisal is essential to understanding how well any competent data driven decisions classifier performs. For performance evaluation, we took into account the scatterplot, precision, recall, f1-score, accuracy, and AUC-ROC curve [28, [29]. We also demonstrated how many recommendations the classifiers made properly or wrong.

A critical performance evaluation factor is the confusion matrix. It incorporates many real and expected values. In order to calculate accuracy, precision, recall, the f1-score, and the AUC-ROC curve, the confusion matrix is very essential. The average accuracy [28] is determined by dividing the number of correct projections by the entire amount of input samples (6).

$$Accuracy = (TP+TN)/(TP+FP+FN+TN)$$

The precision value [28] is obtained by dividing the total number of correct correctly predicted by the total number of samples predictions generated by a classifier (7)

$$Precision = TP/(TP+FP)$$

The recall value is calculated by dividing the total quantity of correctly predicted samples were positive by the total amount of actually positive samples [28]. It is computed utilizing (8).

$$Recall = TP/(TP+FN)$$

F-Score

F1 Score takes into account recall and accuracy. It is the accuracy and recall's harmonics means (average).

Precision (p) and recall (r) should be balanced in the system for the optimum F1 Score. On the other hand, if one metric is enhanced at the price of the other, the F1 Score isn't as strong.

For instance, the F1 score is 0 if P is 1 and R is 0.

F1 Score is calculated as 2*(Recall * Precision) / (Recall + Precision).

4.2 Experimental Results and Analysis

KNN which accuracy is 77.8%, Logistic regression 88%, Random Forest 87.87%, DT is 82.8%, SVM which accuracy is 88% and NB 86.6%
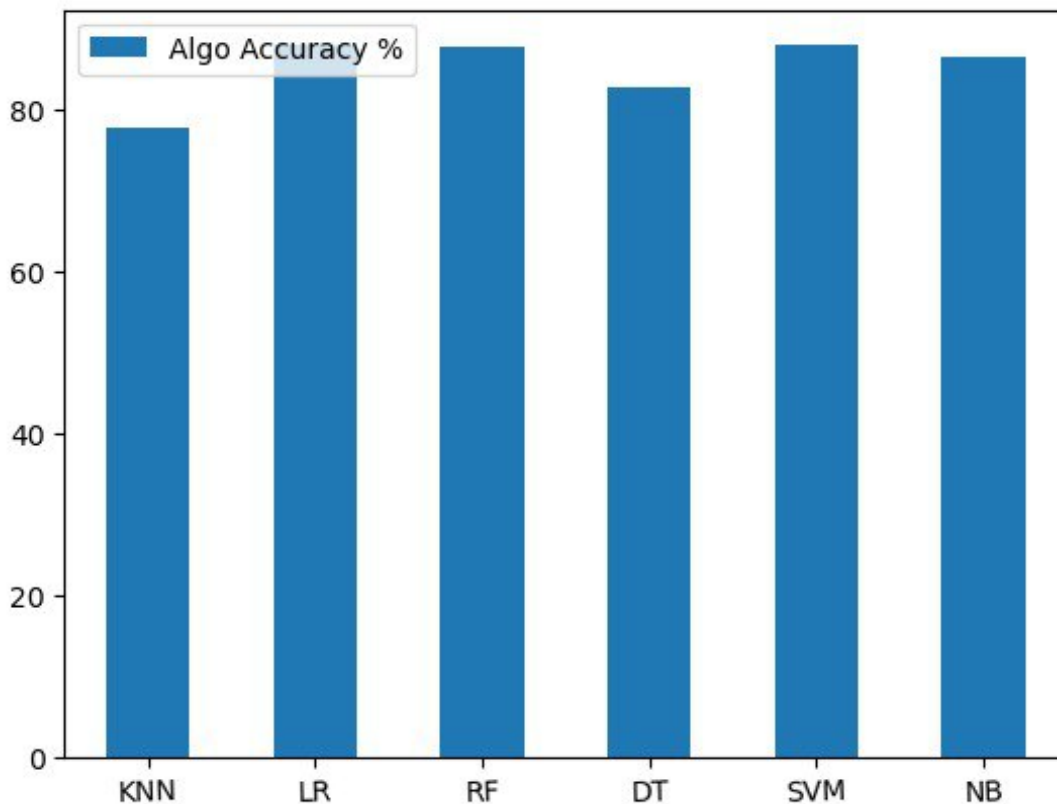
Graph:

Fig 4.2 compares analysis

Table 4.2 Algorithm

Algorithm	Accuracy
KNN	77.8%
LR	88.0%
RF	87.9%
DT	82.8%
SVM	88.0%
NB	86.8%

In our research, we applied that algorithm and get our expected accuracy. Our system can now detect online bowling. our best accuracy is provided by SVM algorithm which is 88.0%. For thus analysis and regression activities are carried out using a controlled machine learning method known as the Svm Based Machine (SVM). The most relevant term is taxonomy, even though we frequently use the phrase "regression worries." The objective of the SVM approach is to find a higher spatial dimensions in an N-dimensional environment that difference is considered the bits of input.

4.3 Discussion

Our information was split into training and testing groups of 80/20. We utilized the 20% testing sets to assess performance after using 80% of the training information to train the classifications. We primarily get 4125 different sorts of textual information from social media platforms. We applied 7 different types of algorithms. KNN which accuracy is 77.8%, Logistic regression 88%, Random Forest 87.87%, DT is 82.8%, SVM which accuracy is 88% and NB 86.6%. our expected and best accuracy is provided by SVM and LR algorithm. We saw those two algorithms' classifications report. In SVM pression, recall and f1-score is 89%,89%, 88%.
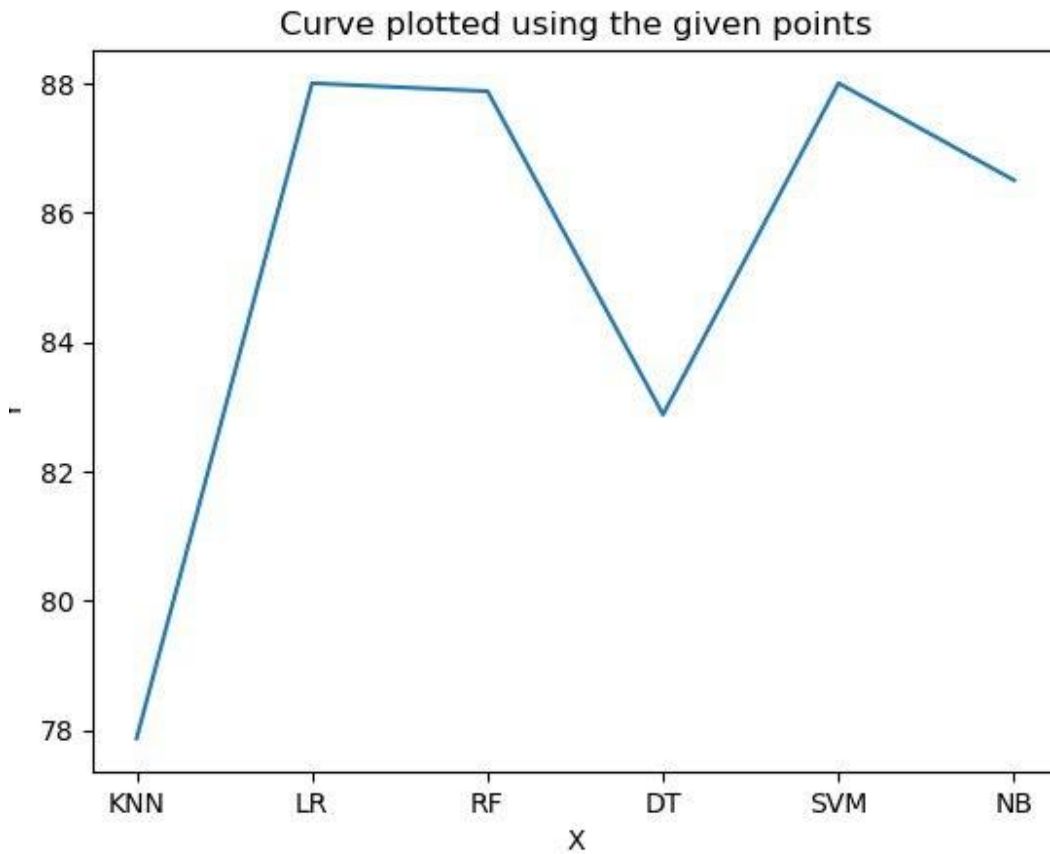
Curve plotted using the given points"

Fig 4.3 Curve plotted

As illustrated in Figure 4, it is evident that the greatest performing algorithm is multinomial naive Bayes. It performs likewise does pretty well, although SVM outscored it.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Social media bullying nowadays very trending problem in our country. Every age people fall a victim of it. Nowadays social site are uses in different reason like education, gaming, business or other purpose. woman's activities are huge now. In social media site girls are faces so many problems. Different sick mentality people bullying those girls through comment which is bulling. Cyberbullying is a form of verbally aggressive mistreatment that exclusively takes place online or through devices like smartphones, desktops, and video game systems. It is frequently designed to embarrass, disgrace, or deter someone by revealing private or sensitive data or by degrading another person. Bullying is a serious problem, despite the fact that it is prohibited in some settings and circumstances. For this reason, Teenagers' reactions to cyberbullying can take a variety of forms.

- poorer productivity and attendance to school
- elevated levels of stress and worry, loneliness and dread, poor attention, and sadness
- In severe circumstances, cyberbullying can result in suicide due to a decline in self - acceptance and self.

5.2 Impact on Environment

An atmospheric consequence is any alteration to the surrounding positive or negative—caused by a facility's activities or services. The consequences of people's decisions and actions on the environment may also be discussed.

5.3 Ethical Aspects

Bullying on digital networking is an Social media . social media and an ethical issue have taken over younger people's livelihoods. Workplace bullying is the term used to describe the use of cyberspace to harass an individual or a team with the goal of physically, psychologically, or socially harmful them. This article will go deeper into this subject and

touch on how sin, destiny, and a conscious morality affect those who engage in harassment. The Bible and Catholic socioeconomic principles will be used to illustrate this, and judgments on the relevance of Catholic teachings on cyberbullying will be made after that.

5.4 Sustainability

The experience and understanding digital era is driven by continual rivalry. A smartphone and tablet function were developed today, but soon long, another one with greater functionality will arise, devaluing the initial request. For this competition, we plan to establish our application work very actively. Customers will have the opportunity to get this Android program without cost. By surfing, you may locate any information more rapidly. Utilization will be far more pleasurable without ads. Customers of our software will profit from the agreements we reach with other countries' doctor scheduling, help, and features. For every problem, they have access to rapid help.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

The fundamental principle of harassment and its diagnostic procedure are covered in this chapter in greater detail. Support vector machines (SVM), J48 Decision Tree, Naive Bayes (NB), and k-nearest neighbor (1-Nearest and 3-Nearest) are among the well-known text-based classifiers that are discussed in the literature.

Additionally, traditional criteria and assessment techniques have been established for the automatic identification of cyberbullying on social media in order to evaluate the effectiveness of classification techniques.

6.2 Conclusions

Teenagers are increasingly concerned about the significant problem of cyberbullying, which frequently has a profoundly detrimental effect on the person. With the onset of the internet era, a new type of bullying has emerged that frequently results in stigmatization. Teenagers are using social media platforms more often, which has led to an upsurge in cyberbullying, which in some severe situations has led to the victims' deaths. Social science specialists have concentrated on personality, social ties, and psychological aspects affecting both of the aggressor and the victim to better understand the phenomena of cyberbullying.

Although somewhat study has been conducted to identify cyberbullying in English writing, nothing has been undertaken in Bangla. In this effort, the efficiency of four cutting-edge unsupervised machines learning algorithms, including NB, SVM, and KNN, on Bangla text was objectively compared. According to an exploratory finding, the SVM-based approach performs better and obtains the greatest accuracy when user-specific data is added. SVM outperforms alternative classification methods for text summarization

because of the high-dimensional input space, lack of unnecessary characteristics, and binary classification character of the text collection.

6.3 Implication for Further Study

The followings for additional research have been suggested because of the proposed system provides constrictions:

1. The suggested solution might be enhanced with several crucial capabilities, such as maintaining historical data and producing reports based on the data.

2. Notifications of a life-or-death situation through SMS or email to the relevant parties.

3. The recommended solution's Android version may introduce a new problem for this project.

4. This system employed four algorithms for machine learning to determine which was the best for this task: NB, J48, SVM, and KNN. To improve accuracy, more pertinent machine learning algorithms might be examined in the coming.

REFERENCES

[1] Akhter, S., 2018, December. Social media bullying detection using machine learning on Bangla text. In *2018 10th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 385-388). IEEE.

[2] Sarker, M., Hossain, M.F., Liza, F.R., Sakib, S.N. and Al Farooq, A., 2022, February. A Machine Learning Approach to Classify Anti-social Bengali Comments on Social Media. In 2022 International *Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)* (pp. 1-6). IEEE.

[3] Islam, M.M., Uddin, M.A., Islam, L., Akter, A., Sharmin, S. and Acharjee, U.K., 2020, December. Cyberbullying detection on social networks using machine learning approaches. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* (pp. 1-6). IEEE.

[4] Ahmed, M.F., Mahmud, Z., Biash, Z.T., Ryen, A.A.N., Hossain, A. and Ashraf, F.B., 2021. Cyberbullying detection using deep neural network from social media comments in bangla language. *arXiv preprint arXiv:2106.04506*.

[5] Kajla, H., Hooda, J. and Saini, G., 2020, May. Classification of online toxic comments using machine learning algorithms. In *2020 4th international conference on intelligent computing and control systems (ICICCS)* (pp. 1119-1123). IEEE.

[6] Emon, M.I.H., Iqbal, K.N., Mehedi, M.H.K., Mahbub, M.J.A. and Rasel, A.A., 2022, July. Detection of Bangla Hate Comments and Cyberbullying in Social Media Using NLP and Transformer Models. In *Advances in Computing and Data Sciences: 6th International Conference, ICACDS 2022, Kurnool, India, April 22–23, 2022, Revised Selected Papers, Part I* (pp. 86-96). Cham: Springer International Publishing.

[7] Eshan, S.C. and Hasan, M.S., 2017, December. An application of machine learning to detect abusive bengali text. In *2017 20th International conference of computer and information technology (ICCIT)* (pp. 1-6). IEEE.

[8] Chakraborty, P. and Seddiqui, M.H., 2019, May. Threat and abusive language detection on social media in bengali language. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1-6). IEEE.

[9] Aporna, A.A., Azad, I., Amlan, N.S., Mehedi, M.H.K., Mahbub, M.J.A. and Rasel, A.A., 2022, July. Classifying offensive speech of bangla text and analysis using explainable AI. In *Advances in Computing*

and Data Sciences: 6th International Conference, ICACDS 2022, Kurnool, India, April 22–23, 2022, Revised Selected Papers, Part I (pp. 133-144). Cham: Springer International Publishing.

[10] Tuhin, S.H., Islam, M.D. and Islam, M.D., 2022. *Cyberbullying Detection using Machine Learning from Social Media Comments in Bangla Language* (Doctoral dissertation, Brac University).

[11] Awal, M.A., Rahman, M.S. and Rabbi, J., 2018, October. Detecting abusive comments in discussion threads using naïve bayes. In *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)* (pp. 163-167). IEEE.

[12] Andročec, D., 2020. Machine learning methods for toxic comment classification: a systematic review. *Acta Universitatis Sapientiae, Informatica*, *12*(2), pp.205-216.

[13] Talpur, K.R., Yuhaniz, S.S. and Amir, N.N.B., 2020. Cyberbullying detection: Current trends and future directions. *Journal of Theoretical and Applied Information Technology*, *98*(16), pp.3197-3208.

[14] Alanazi, I. and Alves-Foss, J., 2020. Cyber Bullying and Machine Learning: A Survey. *International Journal of Computer Science and Information Security (IJCSIS)*, *18*(10).

[15] Apoorva, K.G. and Uma, D., 2022, August. Detection of Cyberbullying Using Machine Learning and Deep Learning Algorithms. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-7). IEEE.

[16] Defersha, N.B. and Tune, K.K., 2021. Detection of hate speech text in afan oromo social media using machine learning approach. *Indian J Sci Technol*, *14*(31), pp.2567-78.

[17] Ritu, S.S., Mondal, J., Mia, M.M. and Al Marouf, A., 2021, July. Bangla Abusive Language Detection using Machine Learning on Radio Message Gateway. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1725-1729). IEEE.

[18] Tarwani, S., Jethanandani, M. and Kant, V., 2019, April. Cyberbullying detection in hindi-english code-mixed language using sentiment classification. In *International conference on advances in computing and data sciences* (pp. 543-551). Springer, Singapore.

[19] Kotak, N. and Patil, B., 2021. STUDY ON CYBERBULLYING AND THE CHALLENGES IN USING HINDI IN SOCIAL MEDIA. *Journal on Future Engineering & Technology*, *16*(3).

[20] Mahmud, T., Das, S., Ptaszynski, M., Hossain, M.S., Andersson, K. and Barua, K., 2022, October. Reason Based Machine Learning Approach to Detect Bangla Abusive Social Media Comments. In *Intelligent Computing & Optimization: Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022)* (pp. 489-498). Cham: Springer International Publishing.

PLAGIARISM REPORT

CYBER BULLYING BENGALI WORDS FROM SOCIAL MEDIA
COMMENTS USING MACHINE LEARNING ALGORTIHMS

ORIGINALITY REPORT

25%	21%	10%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to BRAC University Student Paper	8%
2	Submitted to Daffodil International University Student Paper	4%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
4	Submitted to Columbia High School Student Paper	2%
5	lib.buet.ac.bd:8080 Internet Source	2%
6	fugumt.com Internet Source	1%
7	www.coursehero.com Internet Source	1%
8	Submitted to Aston University Student Paper	<1%
9	Abdhullah-Al-Mamun, Shahin Akhter. "Social media bullying detection using machine	<1%