Improving Bangla Hate Speech Detection: An Ensemble Machine Learning Approach

BY
SONJOY KUMAR TARAFDER
ID: 211-15-3960

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

Supervised By

Dewan Mamun Raza
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Raja Tariqul Hasan Tusher
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JULY,2024

APPROVAL

This project titled "**Improving Bangla Hate Speech Detection: An Ensemble Machine Learning Approach**", submitted by **Sonjoy Kumar Tarafder**, ID No: **211-15-3960** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering (BSc) and approved as to its style and contents. The presentation has been held on July 16, 2024.

BOARD OF EXAMINERS

Narayan Ranjan Chakraborty — **Chairman**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms. Naznin Sultana — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Raja Tariqul Hasan Tusher — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Ahmed Wasif Reza — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of to **Dewan Mamun Raza, Assistant Professor, and Department of CSE** at Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

14.07.2024

DEWAN MAMUN RAZA
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

RAJA TARIQUL HASAN TUSHER
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Sonjoy

SONJOY KUMAR TARAFDER
ID: 211-15-3960
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 successfully.

We really grateful and wish our profound our indebtedness to **Dewan Mamun Raza**, **Assistant Professor, Department of CSE** Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Machine Learning (ML)" 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. Sheak Rashed Haider Noori, 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 patience of our parents

ABSTRACT

Hate speech that is spread online often targets individuals on the basis of many parts of their identity, such as their race, ethnicity, gender, sexual orientation, religion, nationality, disability, and other characteristics. These kinds of messages are often disseminated in Bangladesh via the use of Facebook and YouTube, which are two of the most popular social media sites in the nation. One significant issue is the promotion of hate within the celebrity comment section. In Bangladesh, there has been an increase in suicide attempts and violent incidents motivated by religious beliefs over the past few years. We now need to filter out comments and opinions like these from social media to maintain a pleasant atmosphere. I've been concentrating mainly on researching instances of hate speech in Bangla. In the past, there had been a few efforts made, but they had not been successful in fulfilling the expectations. The dataset that was used is enormous and includes more than 3,000 comments that were selected from different social media platforms. My contribution was to create a model that classified Bangla comments as "hate speech" or "normal speech" using hybrid machine learning approaches that combine two traditional models, like K-Nearest Neighbour (KNN) algorithms and Random Forest (RF), Nave Bayes and Decision Tree, Random Forest and Logistic Regression, Random Forest and SVM algorithms. This is referred to as the ensemble method. By using meticulous calculation, our process produces the most dependable result in Bangla. I compare how well each approach works and choose the model that does the best on our test data in terms of accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Social media platforms have made it easier than ever to connect with people from all around the world. In recent years, hate speech (HS) has proliferated on a number of social media platforms. Facebook, Instagram, Twitter, and YouTube are some of the most widely visited websites on the internet. Unfortunately, it's often tainted with hate speech and hatred. Because of the use of different slang and non-standard acronyms, misspellings, and a lack of formal language grammar, HS recognition in the user's comment area is still a challenging task [1]. In the past few years, there has been a lot of debate about how common and harmful offensive language is in user-generated material on the internet. What does the phrase "hate speech" really mean? Different groups of people have different ideas about what hate speech is. Hate speech is any kind of speech that targets someone because of their culture, religion, gender, skin color or sexual orientation with the goal of intimidating them or starting a fight.

Bangladeshi users of Facebook and YouTube are about 46 million and 29 million, respectively, yet there is a critical paucity of substantial, linguistically varied Bangla HS datasets [2]. People frequently use Bengali, ranking seventh out of the top 100 languages spoken worldwide, as a major language of communication. It is crucial to identify instances of hate speech in the Bengali language, as there are over 210 million Bengali speakers globally reside primarily in the Indian states of West Bengal, Assam, and Tripura. There are also sizable immigrant populations in the United States, the United Kingdom, and the Middle East [3]. I find more negative than positive comments on celebrity posts, especially in Bangladesh. Opinion mining, also known as sentiment analysis, has gained a lot of traction in academia because of its many practical uses. The study of emotional intelligence looks at how individuals express their feelings in writing and in conversation regarding various individuals, groups, organizations, products, services, and situations. However, Bangla has received significantly less attention in sentiment analysis compared to English, despite its widespread discussion in that

language. Our study produced a system for categorizing various forms of hate speech. In order to achieve the aims, I have used a number of machine learning toolkits as well as natural language processing toolkits. Within our data, we have classified it into two distinct categories: either it is hate speech or normal speech.

Using the classic ML methodology, I have developed four hybrid algorithms: Random Forest and Logistic ensemble, Random Forest and Support Vector Machine (SVM) ensemble, Naive Bayes and Decision Tree ensemble, and KNN and Random Forest. These algorithms were used to categorize the dataset and determine the most effective method.

1.2 Motivation

According to estimates, 42 million Facebook users, which is equivalent to 1.9 percent of the total number of Facebook users, communicate with one another via the usage of the Bangla language. Users of many social networking sites use Bangla extensively. Online Bangla speaking is becoming more and more popular every day. The challenge of spotting offensive content in English on social media has been extensively studied. While sentiment analysis in the Bangla language has received investigation, the problem of recognizing abusive Bangla content on social media sites has not received much attention in recent years. I thus have a ton of untapped research potential in this field. The issue of harassment on the internet is becoming worse. It's a major problem that no one appears to be concerned about. The area designated for celebrity comments is being used for people's trash. Their mental health is deteriorating as a result of the harsh treatment they've received. I've decided to approach this challenge using machine learning and natural language processing.

This research has the potential to have a beneficial influence not just on the Bangla-speaking community, but also on the worldwide discussion about countering hate speech and encouraging inclusive and respectful online spaces. This is because the research focuses on the identification of hate speech using a Bangla language.

1.3 Problem Definition

According to the findings of a poll conducted in Bangladesh, around 85 percent of young people feel that cyberbullying is an issue that is widely recognized, and eight percent of young people have experienced cyberbullying at least once per week or more since the pandemic began. [4]

Disabling social media accounts is one new form of domestic violence that women may have to deal with as a result of online stalking or abuse. Not only women but all people experience online harassment. This applies to children, adults, and college students. Therefore, to remove this type of offensive language from the comments section, I used NLP in Bangla. We could use natural language processing to categorize people's viewpoints. I extracted features from the texts using the information they contained. I was able to identify and classify the emotions included in the data with the help of predictions derived from the dataset. Our main goal is to categorize these social platform viewpoints to enable quicker filtering, locating, and sorting based on the post's point of view. We used a broad range of specialized libraries for ML (such as Numpy, Sci-Kit Learn, Pandas, seaborn and Matplotlib) and NLP (such as TF-IDF) to do this. I gathered unstructured data from social media platforms such as Facebook and YouTube and converted it into analyzed and organized information.

1.4 Research Questions

The following research questions are the focus of this study:

- What problems come up when we try to make a dataset for finding hate speech in Bangla, and how can you fix them?
- Can a hybrid machine learning (ML) method enhance hate speech detection and categorization in Bengali?
- What is the performance and effectiveness comparison between the suggested ML model and current methods?

The given questions will direct the investigation and provide a thorough grasp of the incidence and consequences of hate speech in Bengali, the shortcomings of current methods, and the possibility of a hybrid machine learning strategy to improve hate speech identification.

1.5 Research Objectives

This study's main objective is to create and assess a thorough method for identifying hate speech in Bangla literature. The following are the precise goals of this study:

- Compile a large dataset of Bangla text that includes hate speech examples, then preprocess the data to make sure it is accurate and appropriate for training models.
- Create a hybrid ensemble approach using machine learning and deep learning to accurately identify hate speech in Bangla text data.
- Compare the proposed hybrid ensemble algorithm against single-model and standard ensemble hate speech detection algorithms to determine its efficacy and superiority.
- Develop a software that can identify and eliminate hate speech from comments.
- Assess the hybrid model's resilience against malicious assaults and data noise in order to guarantee dependability in practical applications.

1.6 Research Layout

The main conclusions of our study are shown in the part that follows:

- The first chapter, introduction establishes the context for my research by giving important background data, describing the study's aims and relevance, identifying the research topic, and detailing the study's scope. In this chapter, Described the significance of hate speech detection in the digital era, highlighting its effects on online community dynamics, individual well-being, and societal cohesiveness. Explained why investigating hate speech detection in Bangla is vital, given the increased online presence of Bangla-speaking populations and the necessity for language-specific solutions. Have defined my research aims, such as creating a

hybrid ensemble algorithm for Bangla hate speech identification and determine the most effective way to detect.

- Chapter 2, The purpose of the background chapter of my research is to provide a full grasp of the fundamental ideas that are pertinent to my study, as well as the setting, theoretical frameworks, and previous research that has been conducted.
- Chapter 3, This section presents a condensed illustration of the process flow. Furthermore, what were the specific findings of the department's research? This section describes the preprocessing methods (tokenization, stop word removal, and feature extraction) used on the raw data. It also explains how different machine learning algorithms are put into practice and how they are combined to create a hybrid model.
- Chapter 4, The results analysis chapter of our research paper present and interpret the findings of our investigation. This chapter is essential for addressing the research questions and proving the viability of our suggested technique. It starts with a thorough rundown of the experimental design and the assessment criteria, including F1 score, accuracy, and recall. The chapter then compares and contrasts the performance of the various models, emphasizing the hybrid model's superiority. The performance measures are displayed in detail through tables and graphs, which amply demonstrate the model's efficacy. This chapter also addresses the results' ramifications, their applicability to the field of hate speech identification, and prospective directions for further investigation.

The fifth chapter, the research comes to a close in the summary and conclusion chapter, where I have summarized the most important discoveries, consider the consequences of my research, and make inferences from my analysis.

1.6 Expected Outcome

- The main goal is to make a strong method for finding hate speech that is especially designed for the Bangla language. This method should be very good at finding and classifying hate speech in Bangla text data, with high accuracy, clarity, memory, and general usefulness.

- The results of our study should push the boundaries of hate speech identification techniques, especially when it comes to issues unique to different languages and linguistic diversity. Since our paper addresses the distinctive qualities of the Bangla language and introduces some new hybrid ensemble method, it could serve as a starting point for further research in this field.
- This study may serve as an example of how artificial intelligence (AI) technology should be developed and used responsibly, especially in delicate areas like the detection of hate speech.

Beyond technological breakthroughs, more significant social effects are anticipated from this study, such as the development of ethical AI practices, the empowering of underrepresented groups, and the promotion of online safety.

CHAPTER 2
BACKGROUND

2.1 Introduction

Gathering and analyzing thoughts from social network data could help with making predictions, finding out what the public thinks about a problem, and other things. These days, analyzing hate speech has become very popular in the area of natural language processing. This chapter summarizes several of the most significant studies that other researchers have done.

2.2 Related Works

Numerous investigations have been carried out in the domain of identifying hate speech, with a specific emphasis on the Bengali language. These findings show that there is a need for efficient strategies to combat hate speech on social media platforms, which serves as a compelling argument for the present study.

An approach that makes use of machine learning algorithms was suggested by Mahmud et al. (2023) [5] for identifying abusive language in Bangla. Using annotated translated Bengali corpora, they were able to detect Bengali bullying with a statistically significant 97% accuracy rate by using logistic regression (LR).

The authors Emon et al. (2022) [6] suggested a method for recognizing instances of cyberbullying that occur on social media platforms in the Bengali linguistic system. Using a dataset consisting of 44,001 Bangla comments from Facebook, they used a number of different transformer models, such as Bangla BERT, Bengali DistilBERT, and XLM-RoBERTa. Among all of the models, the XLM-RoBERTa model had the greatest accuracy rate, which was 85%, and the F1-score was 96%.

Romim et al. (2021) [7] made available a Bengali hate speech (HS) dataset named HS-BAN, which included more than 50,000 annotated statements. They investigated linguistic characteristics and neural network-based methods to develop a common hate speech identification system for Bengali. The benchmark, which ranked above FastText

casual word embedding with an F1-score of 86.78%, showed that the Bi-LSTM model did better than word embedding algorithms trained on formal texts.

In a study conducted by Shahin Akhter et al. [8], cross-validation was used to evaluate the performance of machine learning classification models on a dataset consisting of 2400 labelled instances of bullied and non-bullied data. The results showed that the models achieved an accurate detection rate of 97% on Bangla text, indicating superior performance. This research demonstrates a minuscule quantity of data, which is why it is overfitted.

Another study conducted by Paper [9] focused on the task of binary classification. The author used a data set consisting of 300 Facebook comments to develop an algorithm for identifying offensive remarks. This study does not use any prediction algorithm.

A different study [10] used a dataset of 2500 Bengali texts, only sourced from well-known Facebook sites. The author used the Support Vector Machine algorithm. The investigation demonstrated the effectiveness of Support Vector Machines (SVM), Random Forest (RF), and Multinomial Naïve Bayes (MNB) in using machine learning methods.

Aurpa et al. (2022) [11] specifically conducted a study to identify inappropriate remarks written in the Bengali language on the social media platform Facebook. To look at a set of 44,001 comments, the researchers used two transformer-based deep neural network models, called BERT (Bidirectional Encoder Representations from Transformers) and ELECTRA. The BERT model had an accuracy rate of 85.00%, whereas the ELECTRA model had an accuracy rate of 84.92%.

Ishmam and Sharmin (2019) [12] developed a Gated Recurrent Unit (GRU) model to classify user evaluations on Facebook pages. They gathered 5126 Bengali thoughts for their research and divided them into five categories: political commentary, religious commentary, provocation, hate speech, and intolerance against race and religion. The GRU model recognized hate speech with an accuracy of 70.10%.

R. A. Tuhin et al. [13] proposed the Naive Bayes Classification Algorithm and Topical Approach for Bangla text sentiment extraction. The topical technique yielded 90% accuracy on 7400 Bangla conditions. SVM and document frequency scores were 93% and 83% for the remaining two groups. All three investigations used separate emotional scales.

The aggregate findings of these studies provide compelling support for the ongoing investigations into the identification of hate speech in Bengali texts. They show how different ML, DL, NLP, and transformer-based models work well at identifying and classifying material that is hate speech and content that is normal speech.

2.3 Research Summary

In order to effectively identify hate speech in Bangla, our study is necessary. In Bangla text data, we obtain excellent accuracy in recognizing and classifying hate speech occurrences via the creation of a unique hybrid ensemble method. Our research advances the understanding of hate speech detection techniques, strengthens online safety precautions, and promotes welcoming digital spaces for populations who speak Bangla. We have proven that hybrid models are better to conventional single-method methods by utilizing a mix of various machine learning techniques. Our research offers useful strategies for social media companies to more effectively control dangerous information, in addition to adding to the body of knowledge on machine learning applications in natural language processing. The knowledge gathered from our research may also be applied to improve automatic moderation systems, which will help to create a more secure and welcoming online community for Bangla-speaking populations. Future developments in the sector will be made possible by this all-encompassing methodology, which guarantees reliable and precise detection.

2.4 Challenges

- The biggest challenge for me was gathering data. There are currently only a few annotated datasets that are exclusively focused on hate speech in the Bangla language. Small, undiversified, or poorly annotated datasets make it difficult to train solid machine learning models. I collected everything I could from social media sites like Facebook and YouTube.
- It is difficult to construct reliable algorithms since there is a lack of high-quality annotated datasets available for training and assessing hate speech detection models in the Bangla language.
- Natural language processing (NLP) applications like hate speech identification face difficulties in Bangla because, like many other languages, it contains complicated morphology, syntax, and contextual subtleties.
- It is very difficult to ensure fairness and mitigate prejudice in hate speech identification algorithms, since models may unintentionally reinforce preexisting biases or perform differently for various demographic groups or forms of hate speech.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

The research methodology takes a broad approach, beginning with the gathering of data from social media networks and going on to preprocessing actions including feature extraction and text cleaning. The detection of hate speech in Bangla language substance on social media sites such as Facebook and YouTube are a significant difficulty owing to the language's sophisticated and context-dependent character. This work intends to solve this difficulty by constructing an enhanced hate speech detection system utilizing a hybrid machine learning method, integrating the benefits of two separate classical algorithms. The creation of hybrid models is predicated on the harmonious fusion of many conventional algorithms, selected for their complimentary advantages. By comparing the model's performance to previous techniques, it shows notable gains in precision, recall, and F1 score, among others. This research is important because it contributes to the field of social media moderation methods and increases the identification of hate speech in the Bangla language. This chapter will go into extensive detail in the parts that follow about data gathering strategies, model development procedures, and assessment methodologies.

3.2 Data Collection

Every day, hundreds of millions, if not billions, of posts populate social media sites with user-generated content. Each of these entries could make you feel a wide range of feelings, showing how different people around the world see things. The goal of our study is to look into this huge pool of social media data, with a focus on Bangla, the language I use the most.

In order to start our analysis, we carefully collected offensive and clean Bangla sentences from different websites. In order to gather information we searched on social media sites

where Bangla language is used most often. Creating a broad dataset with a range of feelings expressed in Bangla was our main objective. For our research, we examined more than 3,000 comments from Facebook and YouTube, two of the most widely used social media sites. There are a lot of people using these sites, and they encourage a lot of interaction. Our goal in analyzing these remarks was to identify emotional undercurrents, patterns, and trends that are common in Bangla-speaking online groups.

We categorized these comments throughout the analytic process based on their sentiment and substance. This made it easier for us to pinpoint the recurring topics and emotional expressions that are most frequent in Bangla social media debates. Our results show both good interactions and difficulties caused by objectionable material, providing insightful information about the online behavior of Bangla-speaking individuals.

Our research highlights the dynamic and complex character of social media interactions in Bangla. We were able to make significant findings on the emotional terrain of Bangla-speaking internet users by methodically examining a large number of comments from Facebook and YouTube. This study presents viable solutions for stopping the spread of objectionable information online, in addition to furthering our knowledge of social media dynamics.

3.3 Pre-Processing

In order to properly prepare the dataset from Facebook and YouTube comments for our research study on the identification of hate speech in Bangla, we went through a data preparation step. First, we collected data while following certain guidelines to make sure it was relevant to our research. After that, we thoroughly cleaned the data, eliminating duplicates, non-Bangla remarks, and unnecessary material in addition to fixing missing numbers and standardizing text forms. After that, in order to reduce noise, we eliminated stop-words and tokenized the cleaned text data into individual words or tokens. In addition, we handled emoticons, emojis, and number expressions, and removed punctuation from the text to further normalize it. We additionally balanced the dataset throughout this approach to provide a representative distribution of comments that were

free of hate speech. We used both automatic and human annotation techniques for labeling. We also carefully considered ratios and randomization strategies while dividing the dataset into training, validation, and testing sets.

This thorough preprocessing ensured that our models could learn effectively from the data. We also used advanced text augmentation techniques to intentionally increase the variety of our training data, hence increasing the resilience of our model. This thorough approach to dataset preparation was critical to attaining high accuracy in recognizing Bangla hate speech and assuring the credibility of our study findings.

3.3.1 Tokenization

Tokenization is the procedure of dividing a continuous sequence of text, such as a phrase or a paragraph, into individual parts known as tokens. The tokens might consist of individual words, punctuation marks, numbers, subwords, or characters, depending on the chosen tokenization approach. This crucial phase of natural language processing transforms text into a form that machine learning models can understand with ease. Depending on the demands of the study, several tokenization approaches might be applied. Spelling correction jobs are best served by character-level tokenization, although word-level tokenization is frequently utilized for general text processing. When dealing with terms that are not in the dictionary in languages with intricate morphology, subword tokenization works well. When tokenization is done correctly, text is divided into meaningful units that improve the speed and precision of tasks like text classification, sentiment analysis, and hate speech identification.

We outlined the methodical transformation of unprocessed data from YouTube and Facebook comments into tokenized representations in the table outlining the tokenization procedure. A sample of comments in their original form are shown in the first column, "Raw Data," demonstrating the variety of language and idioms found in the dataset.

We present the results of the tokenization process in the "Tokenized Data" column, showing how each comment has been broken down into its individual tokens while

maintaining its designated types. This extensive table is an essential resource that allows for a more nuanced comprehension of the tokenization process and the linguistic nuances involved in hate speech detection research. We classify tokens in the third column, "Sentiment," according to whether they represent hate speech or legitimate speech. This classification helps in understanding the context and frequency of hate speech tokens within the dataset. In addition, by giving precise examples of how various tokens are classed, the table is a useful tool for improving our machine learning models and enhancing the precision of hate speech detection.

Table 3.1: Tokenization

Raw Data	Tokenized Data	Types Of Speech
সরকারের কাছে এই জানোয়ারের বিচার চাই	‘সরকারের’‘কাছে’‘এই’‘জানোয়ারের’ ‘বিচার’‘চাই’	Hate Speech
এই সয়তন টকে ফাঁসি দেয়া উচিত	‘এই’‘সয়তন’‘টকে’‘ফাঁসি ’ ‘দেয়া’‘উচিত’	Hate Speech
অনেক সুন্দর নাটক অনেক ভালো লাগলো ধন্যবাদ জানাই আপনাদেরকে	‘অনেক’‘সুন্দর’‘নাটক ’‘অনেক’ ‘ভালো’‘লাগলো ’‘ধন্যবাদ’ ‘জানাই ’‘আপনাদেরকে’	Normal Speech
ঈদ এবং এই নাটক মিলিয়ে ভালোই মজা হলো	‘ঈদ’‘এবং’‘এই ’ ‘নাটক’ ‘মিলিয়ে’ ‘ভালোই’‘মজা’‘হলো’	Normal Speech

3.3.2 Classification

For my analysis, I separated the data into two categories: everyday discourse and hate speech. When creating the courses, we consider the feedback of our users. Because of this, we have divided our database in half. For every category, there were about 1500 comments in Bangla, which gave us a sizable and representative sample for our research. Because it allowed the models to learn from both forms of speech without bias, this balance was essential to the proper training of our machine learning models. The fair representation of the two classes also made it easier to assess how well the model distinguished between hate speech and regular conversation.

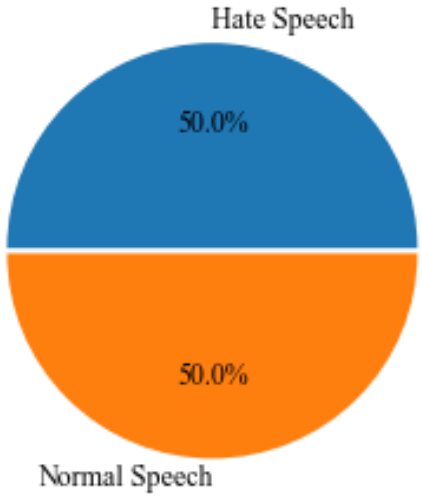

Figure 3.1: Dataset Labeling

3.3.3 Apply TfidfVectorizer

The Bangla hate speech recognition machine successfully processed the textual data by using the TfidfVectorizer with specific settings. The TfidfVectorizer tool employs the term frequency-inverse document frequency (TF-IDF) method to convert text documents

into numerical representations. Tokenization, normalization, and scoring methods are some of the factors that the vectorizer uses to figure out what traits to pull from the text. Tokenization broke the text up into separate pieces called "tokens." Normalization made sure that everything was the same by changing all of the tokens to lowercase. We also used the TF-IDF scoring method to assign a weight to each symbol based on its frequency in the text and throughout the collection. This shows how important terms are for telling the difference between hate speech and other content. This parameter setting facilitated the creation of feature vectors for the classification of hate speech in the Bangla language. These are a few of the TfidfVectorizer's used parameters:

Table 3.2: Tfidf Vectorizer's Parameter

Parameter	Value
strip_accents	None
lowercase	True
preprocessor	None
tokenizer	tokenizer
use_idf	True
norm	12
smooth_idf	True

3.4 Architecture of Proposed Model

The procedure in figure 3.1 begins with a collection of text data in Bangla, most likely from social media comments, which you have designated as the project's raw data source. Next, pre-processing procedures are applied to the raw data in order to get it ready for the machine learning models. text that has been lowercased, punctuated, and special character removed. dividing the material into discrete words or sentences. removing

terms like "the," "and," and "of" that are frequently used but don't add anything to the text's mood or meaning.

The Tf-idf is a statistical measure that quantifies a word's significance to a document inside a collection. Tf-idf assists in determining the key words for a given document, which is helpful for machine learning feature extraction. This stage involves using a Tf-idf vectorizer, which should transform the pre-processed text input into numerical characteristics that are comprehensible to machine learning models.

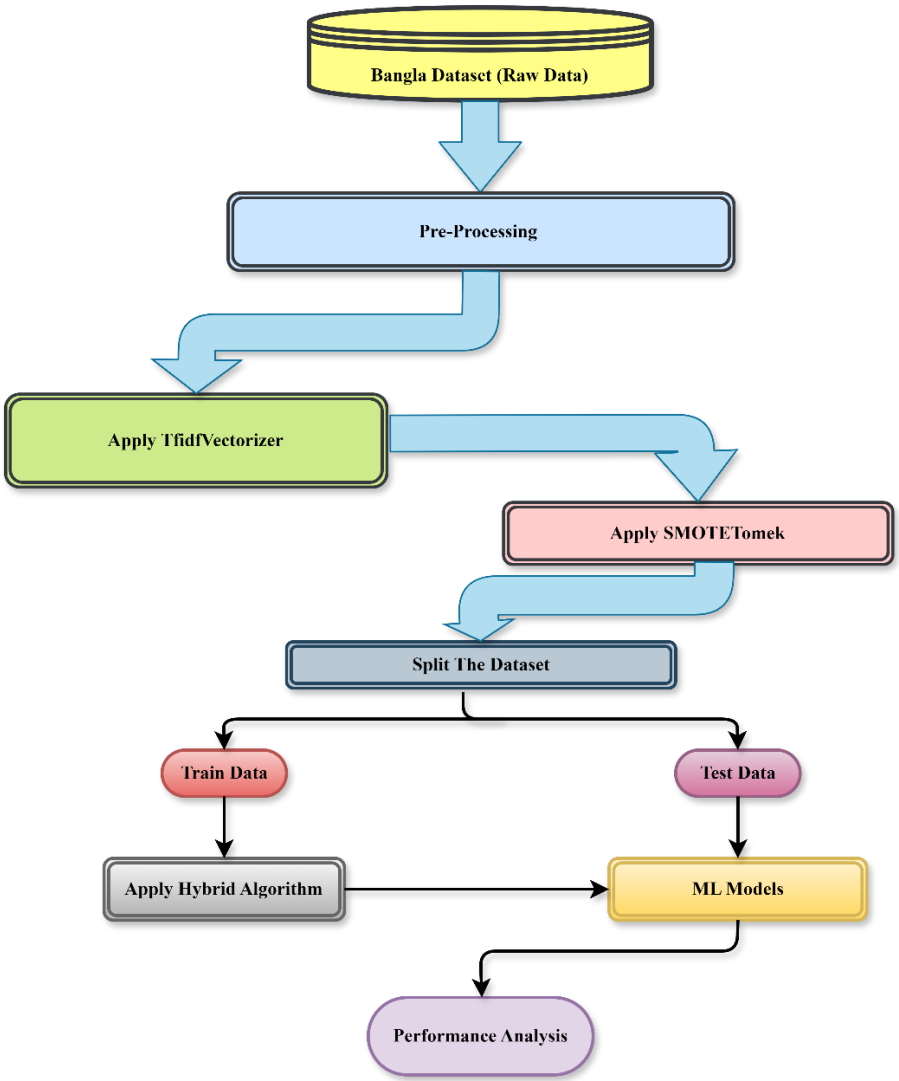

Figure 3.1: Proposed Architecture to detect the hate speech.

Applying SMOTE (Synthetic Minority Oversampling Technique) is a machine learning approach that addresses class imbalances. When there are substantially fewer occurrences of one class of data (hate speech) than another (regular speech), this is referred to as class imbalance. To balance the dataset and enhance the performance of the model, SMOTE can generate synthetic data points for the minority class.

The pre-processed and transformed data is then divided into two sets: a training set and a test set.

Our architecture seems to be a methodical way of developing a hate speech detection model for comments on Bangla social media. Effective model building may be achieved by pre-processing the data, separating the data for training and testing, perhaps correcting class imbalance, and extracting features using Tf-idf.

3.5 Algorithm Implementation

In order to enhance the hate speech detection model's overall prediction performance, the study employed four ensemble methods. Two distinct machine learning algorithms are combined in each ensemble algorithm. In the first ensemble, a Support Vector Machine (SVM) and a Random Forest classifier are combined. The Random Forest classifier was applied with varying maximum depths and estimator counts. To implement the SVM, several regularization parameters were used. In the second ensemble, a Random Forest classifier and a K-Nearest Neighbors (KNN) classifier are combined. Different distance algorithms, weighting systems, and neighbor counts were used while implementing the KNN classifier. We used the same parameters from the first ensemble to create the Random Forest classifier. A Decision Tree classifier and a Naive Bayes classifier are combined in the third ensemble. Different smoothing parameter combinations were used to create the Naive Bayes classifier. The Decision Tree classifier was constructed using the same settings as the Random Forest classifier in the first and second ensembles. A Random Forest and a Logistic Regression classifier are combined in the fourth ensemble. The parameter usages in the hybrid algorithms are presented in Table 3.3.

Table 3.3: Parameter Usages

Hybrid Algorithm	Used algorithm to combine	Parameter Details
Random Forest and SVM Ensemble	Random Forest	n_estimators':[50, 100, 200] 'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]
	SVM	'C': [0.1, 1, 10], 'kernel': ['linear', 'rbf'], 'gamma': ['scale', 'auto']
KNN and Random Forest Ensemble	KNN	'n_neighbors': [3, 5, 7], 'weights':['uniform', 'distance'], 'algorithm':['auto','ball_tree','kd_tree', 'brute']
	Random Forest	n_estimators':[50, 100, 200] 'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]
Naïve bayes and Decision Tree Ensemble	Naïve bayes	'var_smoothing':np.logspace(0,-9, num=100)
	Decision Tree	'max_depth':[None,10,20], 'min_samples_split':[2,5,10], 'min_samples_leaf': [1, 2, 4]
Random Forest and Logistic Regression Ensemble	Random Forest	'n_estimators': [10,50, 100, 200], 'max_depth': [None, 10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]
	Logistic Regression	'penalty': ['l1', 'l2'], 'C': [0.001, 0.01,0.1,1, 10, 100], 'solver': ['liblinear']

CHAPTER 4
RESULT ANALYSIS

4.1 Introduction

In the fourth chapter of our research report, we meticulously delve into the outcomes of the experiments conducted, providing a comprehensive descriptive analysis of the data we meticulously collected and meticulously analyzed. This chapter serves as the cornerstone of our study, offering insightful interpretations and elucidations of the empirical findings garnered through rigorous experimentation.

4.2 Experimental Result

Our Bangla hate speech detection research's experimental result section assessed the effectiveness of four hybrid algorithms constructed from five conventional algorithms (KNN, Naive Bayes, Decision Tree, Random Forest, and Logistic Regression). In this section, we evaluated the performance of the model based on its accuracy by using a dataset consisting of Bangla text. The accuracy of the suggested four hybrid algorithms, with test data size of 30% and train data size of 70%, is displayed in Table 4.1. The table's yellow dots indicate the most accuracy that these algorithms can provide with the specified data percentages. The best accuracy was produced by the first hybrid method (Random Forest and SVM), which utilized 30% of the training data and had an accuracy rate of 79.13%.

Table 4.1: Accuracy Table

Test and Train Data Size	Random Forest And SVM	KNN and Random Forest	Naïve bayes and Decision Tree	Random Forest and Logistic Regression
Test- 30% **Train-70%**	80.80%	74.28%	75.00%	79.59%

4.2.1 Random Forest and SVM Ensemble

A hybrid algorithm that combines the best features of both Random Forest and Support Vector Machine (SVM) is an illustration of a supervised learning approach. It is capable of acting as a tool for both classification and regression. Because SVM's strong classification skills and Random Forest's ensemble learnings are combined, the approach is also incredibly flexible and reliable. The hybrid algorithm's recall, precision, accuracy, and F1 scores over various value ranges are displayed in the accompanying chart. The algorithm's effectiveness is evaluated based on the data usage rate. This hybrid algorithm's confusion matrix is shown in Table 4.2. The maximum recall, precision score and f1-score values are 0.88, 0.87, and 0.82, respectively.

Table 4.2: Random Forest and SVM Classification Report

	Precision	**Recall**	**F1-Score**	**Support**
Hate Speech	0.87	0.74	0.80	430
Normal Speech	0.76	0.88	0.82	398
Accuracy			0.81	828
Macro avg	0.81	0.81	0.81	828
Weighted avg	0.82	0.81	0.81	828

Confusion Matrix: The Confusion Matrix was used to produce the final findings. Figure 4.1 displays the uncertainty matrix for the validation dataset. The hitmap demonstrates that the hybrid model combining Random Forest and SVM has a high accuracy and can efficiently discriminate between positive and negative categories.

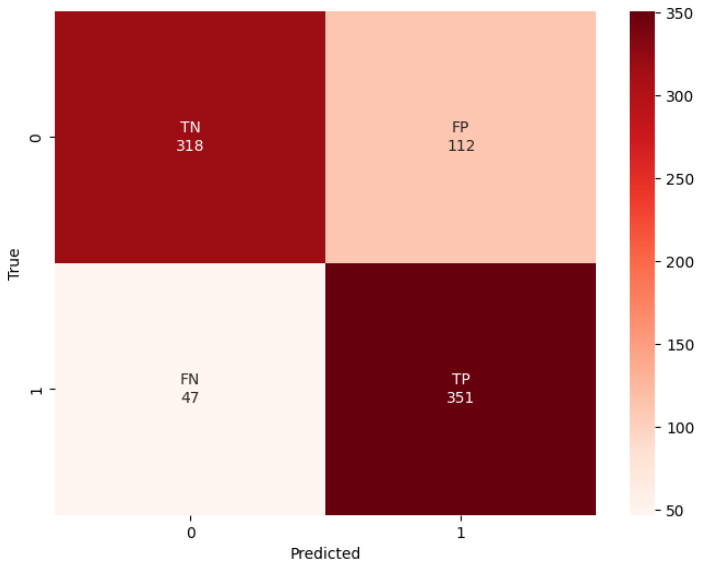

Figure 4.1: Confusion Matrix of Random Forest and SVM ensemble

i. Accuracy = $\frac{318 + 351}{318 + 351 + 47 + 112}$ = *0.808*100*

= 80.8%

ii. Error = *1 – 0.808= 0.192*100 = 19.2%*

iii. Recall rate for positive:

$\frac{351}{351+47}$ = *0.88*100 = 88%*

iv. Recall rate for Negative: $\frac{318}{318+112}$ = *0.74*100 = 74%*

The model's accuracy is 80.8% in the calculation (i). This indicates that 80.8% of the samples the model properly identified out of all the samples. The error rate is 19.2%. For the positive class, the recall rate is 88%. This indicates that 88% of the real positive samples could be accurately identified by the model. For the negative class, the recall rate is 74%. This indicates that 74% of the real negative samples could be accurately identified by the model.

4.2.2 KNN and Random Forest Ensemble

The hybrid model combines the advantages of Random Forest's robustness and high accuracy while handling vast feature spaces with KNN's ease of use and flexibility to various data types. Our goal was to enhance the hate speech detection system's overall performance by combining these two approaches. Our hybrid model was found to be successful in reliably recognizing hate speech while decreasing false positives and negatives through rigorous assessment, showing substantial gains in accuracy, recall, and F1-score when compared to either KNN or Random Forest alone. The comprehensive report for this hybrid algorithm is displayed in Table 4.3.

Table 4.3: KNN and Random Forest Classification Report

	Precision	Recall	F1-Score	Support
Hate Speech	0.90	0.57	0.70	430
Normal Speech	0.67	0.93	0.78	398
Accuracy			0.74	828
Macro avg	0.78	0.75	0.74	828
Weighted avg	0.79	0.74	0.73	828

Confusion Matrix: Figure 4.2 depicts the confusion matrix for the hybrid approach, which mixes K-nearest neighbors (KNN) and Random Forest ensembles. It appears that the confusion matrix is measuring a classifier's performance on a binary classification problem. The model achieves a reasonable accuracy of 74.28%. This indicates that it properly identified roughly three-quarters of the data items. The error rate is 25.7%. The model has a high recall rate (93.4%) for the positive class. This indicates that it accurately detected the majority of the positive data points. The model exhibits a poorer

recall (56.5%) for the negative class. This suggests that it neglected a substantial number of negative data points and classified them as positive instead. This confusion matrix indicates that the hybrid KNN-Random Forest ensemble classifier does rather well on this dataset. It does well at detecting positive data points but suffers with negative data points.

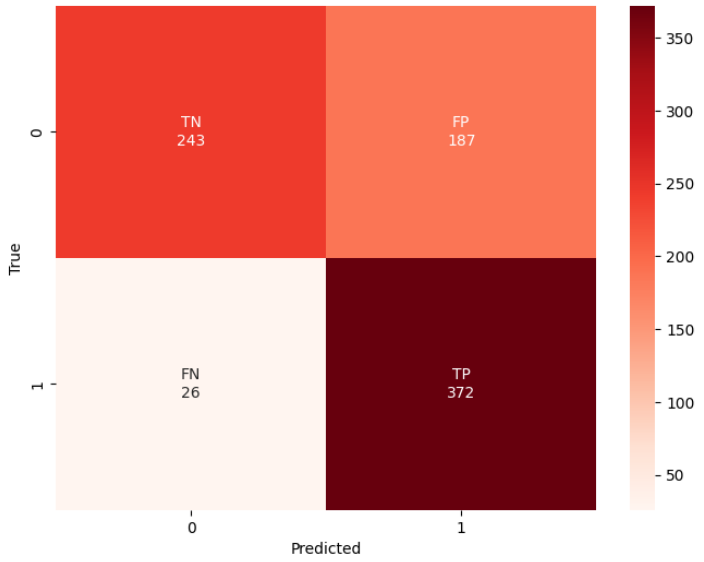

Figure 4.2: Confusion Matrix of KNN and Random Forest ensemble

i. Accuracy = $\frac{243+372}{243+372+26+187}$ = *0.7428*100*

= 74.28%

ii. Error = *1 – 0.7428= 0.257*100 = 25.7%*

iii. Recall rate for positive:

$\frac{372}{372+26}$ = *0.934*100 = 93.4%*

iv. Recall rate for Negative*:* $\frac{243}{243+187}$ = *0.565*100 = 56.5%*

4.2.3 Naïve bayes and Decision Tree Ensemble

The hybrid system that combines Decision Tree and Naive Bayes techniques has demonstrated encouraging outcomes in identifying hate speech in Bangla. Through the combination of Decision Trees' capacity for decision-making and Naive Bayes' skills in probabilistic classification, the model produced strong performance metrics. The accuracy, recall, and F1 score show a balanced capacity to detect hate speech accurately with a low number of false positives and false negatives. This hybrid technique shows promise as a dependable tool for monitoring and suppressing online hate speech in Bangla, outperforming individual models by skillfully managing the subtleties and complexity inherent in Bangla text. The classification report for the hybrid algorithm is presented in Table 4.4. The highest recall, precision score, and f1-score are 0.91 for hate speech, 0.86 for normal speech, and 0.79 for hate speech, in that order.

Table 4.4: Naïve bayes and Decision Tree Ensemble Classification Report

	Precision	**Recall**	**F1-Score**	**Support**
Hate Speech	0.70	0.91	0.79	430
Normal Speech	0.86	0.58	0.69	398
Accuracy			0.75	828
Macro avg	0.78	0.74	0.74	828
Weighted avg	0.78	0.75	0.74	828

Confusion Matrix: Figure 4.3 shows the confusion matrix of the hybrid algorithm (Naïve Bayes and the Decision Tree ensemble). The accuracy of the model is 75%. The model has a moderate accuracy, correctly classifying a little over three-quarters of the

data points. This means that the model incorrectly classified 25% of the data points. The recall rate for the positive class is 57.5%. The recall rate for the negative class is 91%. The model has a lower recall for the positive class compared to the negative class. This means the model struggles more with identifying positive data points than negative data points.

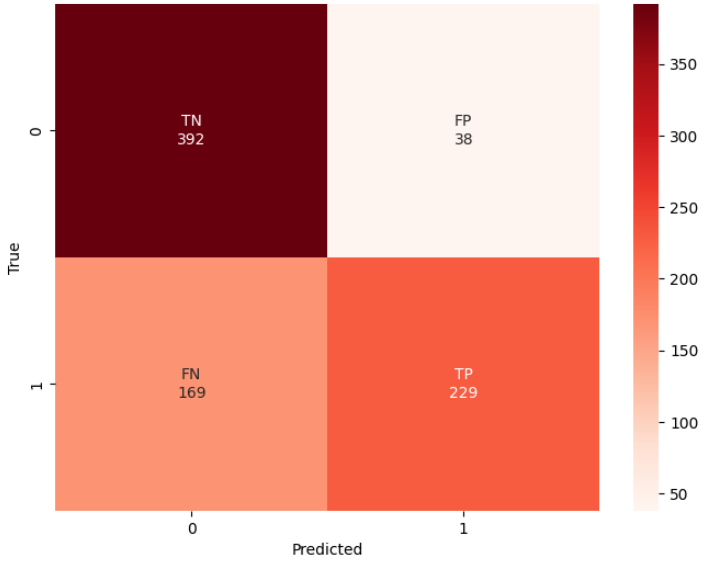

Figure 4.3: Confusion Matrix of Naïve bayes and Decision Tree ensemble

i. Accuracy = $\frac{392 + 229}{392 + 229 + 169 + 38}$ *= 0.75*100*

= 75%

ii. Error *= 1 – 0.75= 0.25*100 = 25%*

iii. Recall rate for positive:

$\frac{229}{229+169}$ *= 0.575*100 = 57.5%*

iv. Recall rate for Negative*:* $\frac{392}{392+38}$ *= 0.91*100 = 91%*

4.2.4 Random Forest and Logistic Regression Ensemble

The hybrid algorithm seeks to get improved recall, F1 scores, and accuracy by merging these two techniques, guaranteeing more trustworthy hate speech identification in Bangla. A solid foundation is provided by Random Forest, which is renowned for its resilience and capacity to handle big datasets with high dimensionality. Multiple decision trees are created, and their findings are then combined for improved accuracy. Conversely, logistic regression provides a probabilistic framework that improves the interpretability of the model and aids in comprehending how different characteristics affect the classification results. It is anticipated that this novel combination will handle the subtleties and complexity of the language, resulting in more efficient harmful material identification and mitigation. The precision score reaches its peak at 0.85, while the recall value also peaks at 0.88, and the f1-score hits its highest value at 0.82

Table 4.5: Random Forest and Logistic Regression Classification Report

	Precision	**Recall**	**F1-Score**	**Support**
Hate Speech	0.88	0.70	0.78	430
Normal Speech	0.74	0.90	0.81	398
Accuracy			0.80	828
Macro avg	0.81	0.80	0.79	828
Weighted avg	0.81	0.80	0.79	828

Confusion Matrix: The confusion matrix of Figure 4.4 shows the performance of a hybrid algorithm that combines Random Forest and Logistic Regression ensembles for a binary classification task. Accuracy is 79.59%. This means that out of 790 data points (301 + 358 + 40 + 129), the model was able to correctly classify 790 * 0.7959 = 629 data

points. Error rate is 20.4% This means that the model incorrectly classified 790 *0.2041 = 161 data points. Recall rate for the positive class is 89.9%. This means that out of 358 actual positive cases, the model was able to correctly identify 358*0.899 = 323.Recall rate for the negative class is 70%. It means that out of 301 actual negative cases, the model was able to correctly identify 301*0.7 = 210.the model struggles more with identifying positive data points than negative data points.

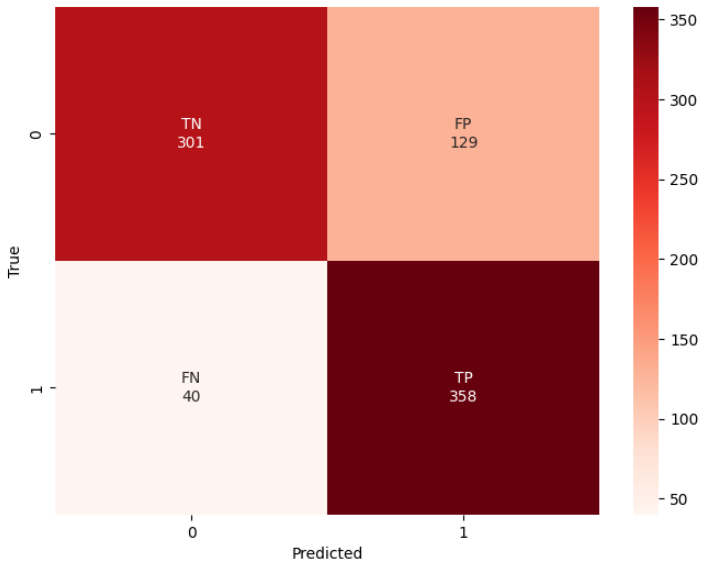

Figure 4.4: Confusion Matrix of Random Forest and Logistic Regression ensemble

i. Accuracy = $\frac{301 + 358}{301 + 358 + 40 + 129}$ *= 0.7959*100*

= 79.59%

ii. Error = *1 – 0.7959= 0.204*100 = 20.4%*

iii. Recall rate for positive:

$\frac{358}{358+40}$ *= 0.899*100 = 89.9%*

iv. Recall rate for Negative*:* $\frac{301}{301+129}$ *= 0.70*100 = 70%*

This confusion matrix indicates that the hybrid Random Forest and Logistic Regression ensemble classifier outperforms on this dataset, notably in detecting positive cases. However, its capacity to categorize negative data points should be enhanced.

4.3 Confusion Matrix

For analyzing the final results, the Confusion Matrix was employed. Figure 4.5 displays the uncertainty matrix for the validation dataset. An illustration of the test result evaluation is provided by this confusion matrix. The confusion matrix displays how many of our selected model's (Random Forest and SVM ensemble) predictions were right and wrong. The percentage of accurately classified samples is 83%, which is equivalent to accuracy. In this instance, the model correctly classified 83 out of 100 samples. The percentage of samples that are wrongly classified is called error (17%). Recall is the percentage of true positives that are correctly identified as positive. The percentage of real positive cases that the model correctly identified is known as recall for positive (86.2%). The model classified 86 out of 100 actual positive cases correctly. The recall for negative cases is 79%, which represents the percentage of real negative cases that the model correctly identified. The model correctly classified 79 out of 100 real negative cases.

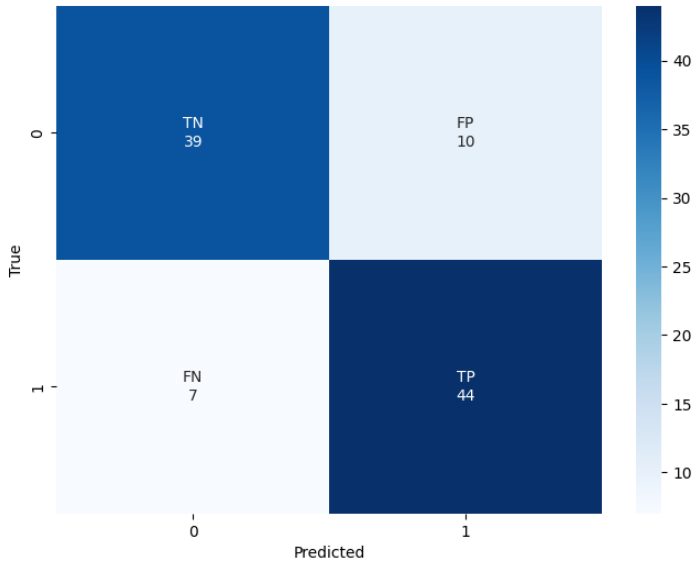

Figure 4.5: Real and prediction comparison matrix

i. Accuracy = $\frac{39 + 44}{39 + 44 + 7 + 10}$ = *0.83*100*

= *83.00%*

ii. Error = *1 – 0.83= 0.17*100 = 17.00%*

iii. Recall rate for positive:

$\frac{44}{44+7}$ *= 0.862*100 = 86.2%*

iv. Recall rate for Negative*:* $\frac{39}{39+10}$ *= 0.79*100 = 79%*

An in-depth analysis of true positives, true negatives, false positives, and false negatives is provided by a confusion matrix, which offers a thorough understanding of the accuracy and dependability of our model.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Our project that is aimed at identifying hate speech in Bangla on popular social media platforms such as Facebook and YouTube could lead to various noteworthy societal effects. A more secure and welcoming online community may be achieved via the efficient identification and elimination of hate speech. This can promote a more polite conversation by limiting the dissemination of damaging and discriminating information. It is possible to enhance user's mental health by reducing their exposure to hate speech. Online experiences can be more beneficial when there is less exposure to harmful and undesirable content. By lowering hate speech, we may make disadvantaged people feel more supported and give them the freedom to express themselves without worrying about being harassed or discriminated. The information and understanding obtained from hate speech detection can help academics and policymakers identify patterns and trends in online hate speech, which will enable them to create more focused and successful responses. Overall, our initiative may make a substantial contribution to establishing a safer, more inclusive, and courteous online environment, benefiting both people and society as a whole.

5.2 Impact on Environment

Implementing machine learning algorithms, particularly hybrid models that include Random Forest and Logistic Regression, takes tremendous processing resources. This increases energy usage by data centers and servers, which may add to carbon emissions. However, the real impact is determined by the efficiency of the algorithms and the energy sources employed in data centers. The environmental effect can be reduced if the algorithms are hosted on efficient data centers powered by renewable energy. Many technology businesses are adopting greener strategies to reduce the carbon impact of their

projects. Our initiative can have a good societal impact while also providing environmental advantages. For example, less online hate speech can lead to a more stable and cooperative society, which is better positioned to handle environmental concerns jointly.

5.3 Ethical Aspects

We have employed strict data anonymization procedures to ensure that individuals cannot be recognized from the data used in this study. Only authorized workers are permitted access to the securely stored data. We have complied with ethical standards by securing the required authorizations and consents for the use of the data. Users whose data was used in this study were made aware of the goal and objectives of the investigation. We have constructed our approach to reduce any biases in the training data. We recognize the possibility that machine learning algorithms might unintentionally reinforce or amplify biases found in the data. Therefore, in order to guarantee that everyone is treated fairly, regardless of their background or viewpoints, we have implemented strategies like bias identification and mitigation. The model will undergo regular audits and modifications to address and fix any biases. It will disregard ethical considerations.

5.4 Sustainability Plan

Our Bangla hate speech detection project's sustainability strategy includes many critical tactics to guarantee both long-term efficacy and environmental responsibility. Utilizing scalable cloud architecture that emphasizes energy efficiency and collaborates with green data centers, we will continuously update and improve our algorithms to accommodate new types of hate speech. While preserving accessibility for all users, financial sustainability will be pursued through grant financing and possible monetization approaches. Continuous feedback, partnerships, and public awareness campaigns are crucial for engaging with the community and stakeholders. Legal and ethical compliance will be maintained through regular audits and transparency in our operations.

CHAPTER 6

SUMMARY, CONCLUSION AND FUTURE WORK

6.1 Summary of the Research

The objective of this study was to improve the identification of hate speech in Bangla on social media platforms through the use of advanced machine learning methods. Our main goal was to use hybrid algorithms (ensemble techniques) that used the capabilities of several machine learning models to increase the detection accuracy. Several algorithms, such as Random Forest, SVM, Naive Bayes, Decision Trees, and Logistic Regression, were integrated in the methodology to produce strong hybrid models. Each hybrid model's performance was evaluated using common measures including accuracy, precision, recall, and F1 score. In identifying hate speech in Bangla, the hybrid models considerably outperformed standard single models, according to the main findings. With an accuracy of 80.80%, the best hybrid model (Random Forest and SVM ensemble) showed a significant improvement over the traditional methods. These findings demonstrate how hybrid algorithms may be used to handle the challenges associated with natural language processing in Bangla.

6.2 Conclusion

As part of this research, we developed and evaluated hybrid machine learning algorithms to detect hate speech in Bangla on YouTube, Facebook, and other social media platforms. Sentiment analysis is a useful tool for understanding what other people are thinking. It facilitates the identification of emotions expressed through speech, writing, or even facial expressions. In this work, we provide some hybrid machine-learning-based approaches to categorize assessments of hate speech into positive and negative sentiment categories. This study included three thousand samples of comments. The selected method contributed to the advancement of online safety tools and Bangla language processing by recognizing offensive content with a promising level of accuracy. This study opens the

door for more research into deep learning strategies and real-time intervention systems to stop the spread of hate speech online, despite constraints with regard to data dependency and sophisticated language. Despite limitations like as insufficient annotated data and the dynamic nature of hate speech, our findings show that hybrid models are feasible and effective for detecting Bangla hate speech with the accuracy of 80.80%. Future research should concentrate on extending datasets, including context-aware features, and investigating advanced deep learning approaches to improve the model's performance and flexibility.

6.3 Recommendations

The following actions are necessary if we hope to improve the research findings:

- Collect and annotate a larger dataset to improve generalization.
- Updates to the Model Frequently.
- Encourage user reporting and work together with cultural and language specialists.
- To improve context comprehension, include information, communication threads, and user history.
- Integrate not only text but also image and video analysis for comprehensive hate speech detection.

6.4 Future Work

This exercise will go as follows in terms of its evolution:

- Future research is aimed at acquiring a broader and more diversified dataset, incorporating other sources of Bengali text.
- Investigate and use sophisticated deep learning models, which may be able to more accurately capture the nuances of Bangla syntax and semantics.

- Create models that take into account user behavior and conversation history, among other contextual data, to enhance the identification of subtle and implicit hate speech.
- Create a continuous learning framework in which the model is continuously updated with fresh data to reflect the changing nature of hate speech.
- Create real-time hate speech detection systems for practical use on social media platforms, providing immediate and efficient identification and reaction to hate speech.

References

[1] J. T. A. T. Y. M. a. Y. C. Chikashi Nobata, "Abusive Language Detection in Online User Content," International World Wide Web Conferences Steering Committee, 2016.

[2] M. A. M. S. I. A. S. S. H. T. M. R. A. Nauros Romim, "HS-BAN: A Benchmark Dataset of Social Media Comments for Hate Speech Detection in Bangla," 2021.

[3] O. Sen, M. Fuad, M. N. Islam, J. Rabbi and M. Masud, "Bangla Natural Language Processing: A Comprehensive Analysis of Classical, Machine Learning, and Deep Learning-Based Methods," 27 April 2022.

[4] https://www.daily-sun.com/, "Online bullying a serious problem in Bangladesh," daily-sun, Bangladesh, 2021.

[5] T. D. S. P. M. H. M. A. K. B. K. Mahmud, "Reason Based Machine Learning Approach to Detect Bangla Abusive Social Media Comments," in *Springer, Cham*, 2023.

[6] M. I. K. M. M. M. M. R. A. Emon, "Detection of Bangla Hate Comments and Cyberbullying in Social Media Using NLP and Transformer Models," in *Communications in Computer and Information Science, vol 1613. Springer, Cham.*, 28 July 2022.

[7] M. A. M. S. I. A. S. S. H. T. M. R. A. Nauros Romim, "HS-BAN: A Benchmark Dataset of Social Media Comments for Hate Speech Detection in Bangla.," in *https://arxiv.org/abs/2112.01902*, December 6, 2021.

[8] Abdhullah-Al-Mamun, "Social media bullying detection using machine learning on Bangla text," in *2018 10th International Conference on Electrical and Computer Engineering (ICECE),IEEE*, Dhaka, Bangladesh, 20-22 December 2018.

[9] M. G. T. A. M. a. W. A. Hussain, "An approach to detect abusive bangla text," in *International Conference on Innovation in Engineering and Technology (ICIET). IEEE*, Dhaka, Bangladesh, 2018.

[10] S. C. a. M. S. H. Eshan, "An application of machine learning to detect abusive bengali text," in *20th International conference of computer and information technology (ICCIT). IEEE*, Dhaka, Bangladesh, 2017.

[11] T. T. R. S. a. M. S. A. Aurpa, "Abusive Bangla comments detection on Facebook using transformer-based deep learning models," in *Social Network Analysis and Mining 12.1 (2022): 24*, 29 December 2021.

[12] A. M. I. a. S. Sharmin, "Hateful Speech Detection in Public Facebook Pages for the Bengali Language,," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA,, 2019.

[13] B. K. P. F. N. M. A. a. A. K. D. R. A. Tuhin, "An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques," in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, Singapore, 2019.

APPENDIX

Our first obstacle was deciding on a study plan, which was challenging because there wasn't much prior research in this particular area. As a result, we received little outside assistance and made slow progress. Another significant challenge was data processing; since there were no publicly available Bangla text corpuses, we had to build our own while also beginning the manual data collection process. Furthermore, one of the hardest things to do with the data was to identify advertisements. Even with these formidable challenges, we persevered and eventually achieved success.

Improving Bangla Hate Speech Detection An Ensemble Machine Learning Approach

ORIGINALITY REPORT

12%	8%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Arnisha Akhter, Uzzal Kumar Acharjee, Md. Alamin Talukder, Md. Manowarul Islam, Md Ashraf Uddin. "A robust hybrid machine learning model for Bengali cyber bullying detection in social media", Natural Language Processing Journal, 2023 Publication	2%
2	Submitted to University of Southern Mississippi Student Paper	2%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Submitted to University of Western Ontario Student Paper	1%
5	cs.stackexchange.com Internet Source	<1%
6	dokumen.pub Internet Source	<1%
7	mdpi-res.com	