AN ARTIFICIAL INTELLIGENCE APPROACH TO DETECT HATE SPEECH FROM BANGLA SENTENCE

BY

Name: Md. Shimul Islam

ID: 191-15-12531

Name: Rejoyana Kabir

ID: 191-15-12592

Name: Fahima Islam

ID: 191-15-12141

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

Supervised By

Ms. Samia Nawshin

Assistant Professor

Department of CSE

Daffodil International University

Co-Supervised By

Asma Mariam

Lecturer

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This project titled "**An Artificial Intelligence Approach to Detect Hate Speech from Bangla Sentence**", submitted by **Md. Shimul Islam, Rejoyana Kabir, Fhima Islam** 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 in 24 January, 2022.

<u>BOARD OF EXAMINERS</u>

Chairman

____

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

Internal Examiner

____

Abdus Sattar
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

____

Fatema Tuj Johra
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

24-01-23
____

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that this thesis has been done by us under the supervision of **Ms. Samia Nawshin, Assistant Professor, Department of CSE,** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

23.7.23

Ms. Samia Nawshin
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Asma Mariam
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Md. Shimul Islam
ID: 191-15-12531
Department of CSE
Daffodil International University

Rejoyana Kabir
ID: 191-15-12592
Department of CSE
Daffodil International University

Fahima Islam

Fahima Islam
ID: 191-15-12141
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

As a matter of first importance, we offer our ardent appreciation to Almighty God for His heavenly elegance, which has empowered us to effectively finish the last proposition.

We are truly appreciative and wish our significant obligation to Ms. Samia Nawshin, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Bangladesh. Profound Knowledge distinct fascination of our boss in the field of "AI and NLP" to complete this proposition. His endless resistance, scholarly course, tenacious consolation, steady and enthusiastic management, productive analysis, important counsel, perusing numerous helpless drafts, and amending us at all levels permitted us to complete this postulation.

We might want to offer our ardent thanks to Prof Dr. Touhid Bhuiyan, Professor and Head, Department of CSE, for his priceless help with finishing our proposal, just as to the other employees and staff of the Daffodil International University CSE office.

We might want to offer our thanks to the entirety of our Daffodil International University schoolmates who partook in this discussion as a feature of their course work.

Finally, we must recognize and acknowledge our parents; unwavering support and enthusiasm.

ABSTRACT

The kind of speech that occurs on the internet is intended to target a person or group on the basis of their religion, ethnicity, gender, disability, or even the color of their skin. Facebook and YouTube are two of the most prominent social media platforms in Bangladesh, and both of them frequently feature talks of this kind. People who preach hatred sometimes do it in the comment sections of celebrities' websites and social media accounts. Because of the proliferation of hate speech in recent years in Bangladesh, there have also been incidences of religious violence and suicides in that country. The removal of negative content from social media platforms has made it necessary to filter out certain types of comments and viewpoints. Consequently, the identification of hate speech expressed in the Bangla language has been our primary objective. There were a few works that came before this one, but they were not satisfactory in any way. A dataset that has more than eight thousand comments gathered from various social media platforms is being used, making it one of the most extensive datasets ever utilized. We implemented the Support Vector Machine (SVM), Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbor (KNN) algorithms in order to create a model that separates Bangla comments into normal speech and hate speech. The most reliable result in Bangla Language can be obtained by using our model, which is based on very specific calculations. Following an examination of the outcomes of each method, we selected the model that provided the highest degree of precision for the test data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

As a strong medium for communication and self-expression, social media facilitates connections between individuals and facilitates the dissemination of information. Almost 3.6 billion people will use social media in 2020, a 49% increase from January 2020.[1] Time spent on social networking sites like Facebook, Instagram, Twitter, and YouTube has become a mainstream digital pastime.

It's often tainted with negativity and hate speech, unfortunately. In recent years, there has been a lot of discussion about the prevalence of foul language in user-generated material on online platforms, as well as the repercussions of this trend. Just what does "hate speech" entail? There are no universally accepted criteria for determining what constitutes hate speech. At its core, hate speech is any form of expression that targets an individual on the basis of their race, ethnicity, religion, gender, or sexual orientation with the intent of frightening them into silence or instigating violence. [2]

Among the world's native languages, Bangla ranked #7 in 2020 with 265 million speakers. This number of people accounts for 3.05 percent of the global populace. [3]. By January 2021, 47.61 million Bangladeshis had access to the internet, proving that widespread use was a reality in the country. But as more and more individuals gained access to the internet, they felt free to express themselves anyway they liked on social media. In particular, in Bangladesh, we find that hateful comments outnumber pleasant ones in celebrity posts. Research into sentiment analysis, often called opinion mining, has flourished in recent years due to the numerous real-world applications it has shown. Emotional intelligence examines how people talk about and write about how they feel about specific people, groups, organizations, goods, services, and circumstances. On the other hand, while sentiment analysis has seen much research in English, this field has seen far less attention in Bangla. Our study offered a framework for identifying instances of hate speech. Several supervised learning algorithms were used to determine the most effective method for classifying the dataset. These included KNN, Decision Tree, Support Vector Machine (SVM), Random Forest, and Logistic Regression.

To accomplish these goals, we used ML toolkits such as learning Sci-kit, Numpy, Matplotlib, and Pandas, and NLP libraries including NLTK (Natural Language Toolkit), TF-IDF, and TF-IDF.

1.2 Motivation

42 million Facebook users, about 1.9 percent of all Facebook users, use the Bangla language to connect with one another. Users on other social media platforms also use the Bangla language. And the use of Bangla on all social media sites is growing by the day. In the topic of offensive text identification in the English language using social networks, a lot of study has been done.

Though there have been some studies on sentiment analysis in the Bangla language, there has been relatively little contemporary study on detecting abusive Bangla text on social media platforms. As a result, there is a lot of study potential in this sector for us. Online harassment has become common in social media. It is a great problem which is left unnoticed. The celebrity comment section has become a garbage disposal for common people. People are being very harsh towards them and that is affecting their mental health. Finally, we've chosen to use NLP and machine learning to solve this challenge. Algorithms, as we all know, do not comprehend strings directly. We must first convert the string to numeric format.

We utilized the TFIDF algorithm in this situation. We utilized a Machine Learning system to categorize each comment. We employed different parameters for each algorithm. And we chose these parameters since they generated the best results.

1.3 Problem Definition

Online harassment affects over 57 percent of women who use Facebook, the largest rate of any social networking site.

[4] Because to cyberstalking or abuse on social media, women may be forced to confront a new type of domestic violence: deactivating their social media accounts. Not just for women, but for practically everyone, whether an adult, a university student, or a youngster, internet harassment is a common occurrence. So to filter out this type of toxic speech form comment section we used Natural language processing(NLP) in Banglali language. We may use NLP to analyze people's opinion and categorize them into different categories. To extract features, text data was used. By running prediction based on the dataset, we were able to isolate the sentiment

of the data and divide it into multiple groups. Our first goal is to categorize these social platform opinions in order to make filtering, searching, and arranging easier based on the post's perspective. We used multiple Natural Language Processing libraries including, TF-IDF, and numerous Machine Learning toolkits like Numpy, sci-kit learn, Pandas, and Matplotlib to fulfill this goal We acquired raw data from social media platforms like Twitter, Facebook, and YouTube, then processed and changed it into labeled data.

1.4 Research Questions

- What strategies will be used to collect and prepare the dataset?
- Is it possible to appropriately define positive and negative groups?
- What criteria will be used to classify positive and negative?
- Is it possible for a machine learning algorithm to correctly estimate Positive and Negative classes?
- How will this work benefit the majority?

1.5 Research Methodology

This section will go over our workflow, which includes data processing, information processing, data classification, and algorithm implementation. Algorithm evaluation, model training.

1.6 Research Objectives

- To study customer data by employing or classifying classification algorithms.
- To create a model capable of reliably detecting positive and negative comments.
- To develop a specific scientific feeling, undertake research.
- Create a software application to filter out hate speech from the comment section..

1.7 Research Layout

The substance of our study is as follows:

Chapter 1 The first part of the preliminary research is just as important as the rest. In this chapter, we also explain why we chose to undertake this study. The problem definition is the key part of this chapter. Here we discuss the study's problem and the difficulty that it presents.

Chapter 2 An input analysis is included, providing a high-level summary of the relevant research. Some of the major machine learning work in this area is described here..

Chapter 3 a brief description of the process flow. What did the analysis reveal about this section?

Chapter 4 It's all about how the results are analyzed. It includes the findings from the graphical analysis.

Chapter 5 The study has reached its last chapter. The results of the model are described in this section. This part further substantiates the validity of the connection. This section also covers the incorporation of the idea and performance into an online environment. At the end of this section, the book's flaws are examined. Future possibilities for the research were also encoded.

1.8 Expected Outcome

- Hate speech in the comment section will be filtered out.
- We will stop online harassment.
- We want to strengthen the ICT Digital Security Act on online hate speech.

CHAPTER 2
BACKGROUND

2.1 Introduction

There are several applications for extracting and analyzing sentiments from social network data, including prediction, gauging public opinion on a certain social topic, and so on. Natural language processing now places a large emphasis on sentiment analysis. The major study undertaken by other researchers is summarized in this chapter.

2.2 Related Works

Hate speech on social media is any content that demeans the target and could put them in danger. Any remark that expresses the author's or speaker's personal attitude, belief, or opinion is considered subjective. It is possible to separate the subjective claim into positive and negative components. Sentiment analysis is a method of emotional analysis that takes a perspective-based approach to identifying and exploring the feelings surrounding a target topic of interest. The idea behind this tactic,

Masum Billah and colleagues [5] suggested a machine learning algorithm to determine whether a Bangla-speaking Facebook user is depressed based on their profile information. Data for the study was gathered from 50 Facebook users, 17 of whom later took their own lives. Unigram + Emoji resulted in a 77.96% accuracy for SGDC. The project could only use traditional machine learning techniques due to a lack of data.

The prevalence of online harassment and abuse is alarming. His coauthor M.T. Ahmed [6] used machine learning and deep learning techniques to create a method for recognizing Bengali cyberbullying and romanizing Bengali text. In addition, they compared algorithms with regards to correctness, accuracy, precision, recall, f1 value, and roc variety. They created three social media datasets: one in Bengali, one in Romanized Bengali, and one that combines the two. CNN achieved 84% accuracy on the Bangla dataset, making it the best-performing dataset overall.

Nave Bayes classifier and the thematic approach are two of the machine learning techniques that Tuhin et al. [7] offer for extracting sentiment from Bengali text. We employed a thematic approach with 90% accuracy on a dataset of 7400 Bengali sentences. The authors then compared their work to two others that received identical scores of 93% on the SVM and 83% on the document frequency. There is a wide range of emotions represented in these three songs.

When it comes to improving categorization results, Namita et al.[8] came up with a strategy to broaden the scope of Hindi SentiWordNet. Emotions in Roman Urdu are analyzed via the lenses of sports, software, food and recipes, theater, and politics in this article. The document includes 10,021 unique expressions taken from 566 online discussions. Overall, their method is accurate 80.21 percent of the time, with a favorable review yielding 82.89 percent and a bad review yielding 76.59 percent.

Bangladesh is facing a growing problem with cybercrime such online harassment, blackmail, and cyberbullying; detecting abusive language in Bengali can assist. Researchers Hussain et al. [8] analyzed Abusive comments were noticed in Bengali across many social networking platforms where people shared their thoughts and beliefs. For training, they used 250 replies, and for testing, they used 50. To find the problematic documents, they suggested a top-down approach, and to boost the quality of the results, they suggested using unigram string properties.

Danish data was compiled by Sigurbergsson et al. [9] using user-generated content on social media platforms like Reddit and Facebook. They developed four automatic classification algorithms that are expected to work in English and Danish. The best performing Danish system has an average F1 macro score of 0.70. The highest-performing method for identifying whether or not a communication is offensive in English provides an F1 macro score of 0.62 on average.

Models for detecting hate speech in Indonesian from text and speech input were developed by T. L. Sutejo et al. [10]. They compared the reliability of the oral and written forms. In this study, we compare different methods for identifying hate speech and find that textual features alone are superior to acoustic characteristics and to models that use both sets of information. Highest scoring model used textual features, outperforming lexical feature and auditory feature models by 82.5% and 86.98%, respectively.

Al-Hassan et al. [11] categorized Arabic tweets using five categories: general hostility, sexism, racism, and religion. The SVM model was compared to the LTSM, CNN + LTSM, GRU, and CNN + GRU deep learning models to set a standard. Eleven thousand tweets were collected and organized into a dataset. The results demonstrate that each of the four deep learning algorithms outperformed the SVM model in identifying hostile tweets. When compared to the average recall of the deep learning models (75%), the SVM achieves 74%. When CNN is combined with LTSM, however, the detection performance improves across the board. The new combined system achieves an F1 score of 73%, recall of 75%, and accuracy of 72%.

2.3 Research Summary

The above study shows different researches were conducted in the area of sentiment analysis. But the results were not as good as it hoped to be. Fewer resources can be one of the reasons. In the context of Bangla language, not much work has been done. We hope that research in his field will increase gradually with better output.

2.4 Challenges

Data collection was our biggest challenge in this research. We collected as much data as possible from different social media platforms. There were not much resources for us to work with as few works had been done in this field The data we gathered were not well optimized. We use some advanced machine learning algorithms to preporess our dataset for further processing.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

we followed these 7 steps to complete our work. Data collection, Preprocessing, dataset, labeling tokenization, algorithm implementation, evaluation is our main part. So main part like preprocessing, labeling, and dataset are divided into subcategories. Figure 3.1 represents methodology graph.

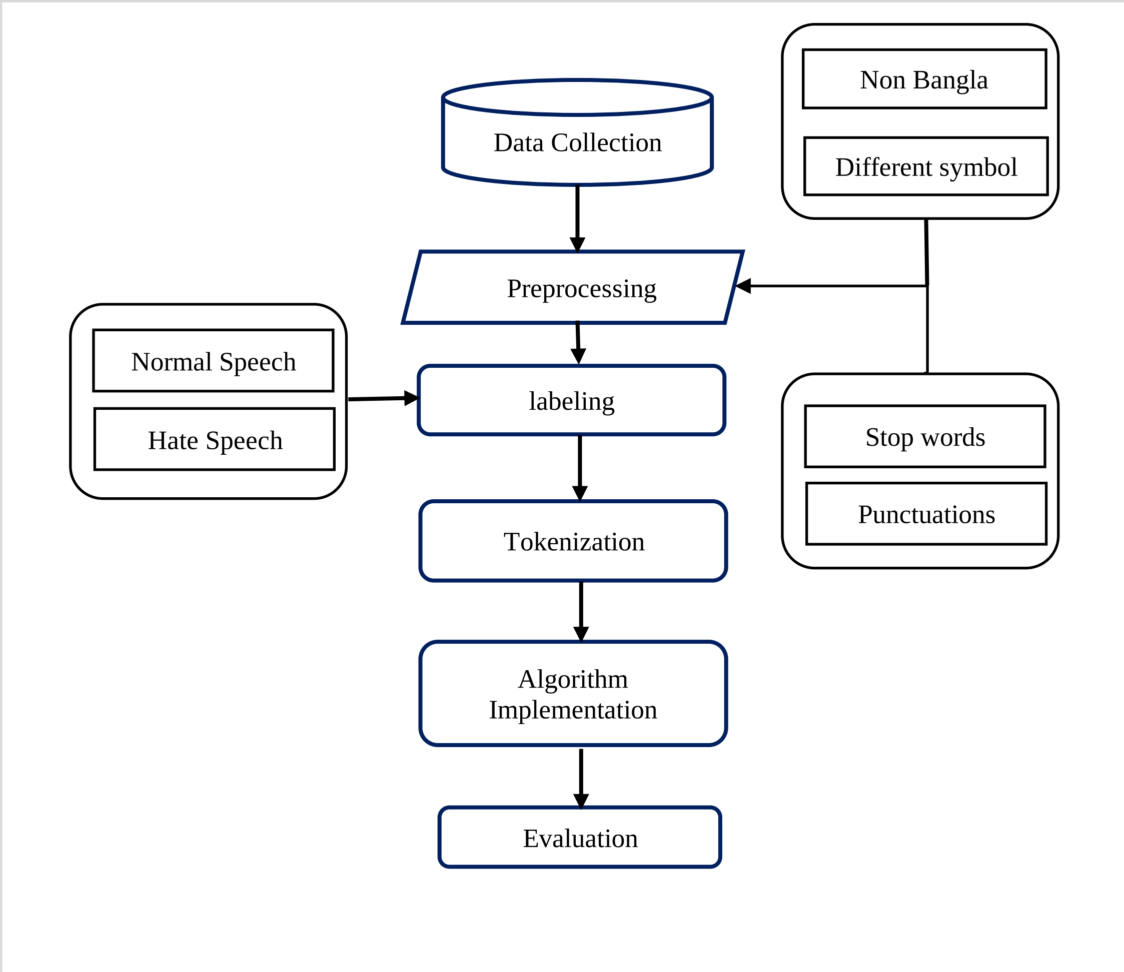

Figure 3.1: Methodology diagram

3.2 Data Collection

Hundreds of billions of postings are shared on social media every day. We gathered the necessary information from several social media platforms such as Facebook, YouTube, and many Bangla blog sites. Each point of view elicits a distinct emotion. Because we usually worked with Bangla, we concentrated on Bangla posts on social media. We then used social media to collect both respectable and filthy Bangla writings. We collected over 10,000 comments for our research results.

3.3 Dataset

TABLE 3.1: SAMPLE OF DATASET

Comments	**Sentiment**	**Labeling**
বেহাইয়া জামাই	Hate Speech	0
নতুন অভিনয়শিল্পী অভারএ্যাকটিং	Hate Speech	0
অসাধারণ নিশো বস্ আর অমি ভাইকেও	Normal Speech	1
আমার দেখা বেস্ট নাটক	Normal Speech	1

Table 3.1 represents the sample of our dataset. There are three columns are existing in our dataset. Comments contains sentences that we collected. Sentiment column represents the sentiments of each comments. And we set hate speech as 0 and Normal speech as 1, because machine learning algorithm never understand string.

3.4 Pre-Processing

Preprocessing data is a data mining technique that turns unstructured data into a format that is easier to use and more effective. The importance of information preparation in learning cannot be overstated. Calculations cannot be performed directly using information. Therefore, the computation must be preprocessed before being performed. As a result, it is necessary to provide information methodically, which calls for planning. During the preprocessing phase, we adhered to the following four steps: Taking away Non-Bangla There are no longer any emoticons, stop words, or punctuation. The first phrase in this table denotes the removal of non-bangla words, the second one, the removal of symbols, and the last two, the removal of stop words and punctuation, respectively. Table 3.2 represents the data preprocessing steps

TABLE 3.2: PREPROCESSING STEPS

Operation Name	**Input Sample**	**Output sample**
Non Bangla removing	Just Waste Of Time.আমাদের দেশের নাটকের এ কি দুর অবস্থা।	আমাদের দেশের নাটকের এ কি দুর অবস্থা।
Symbol Removing	সময় টাই ফাউল গেলো	সময় টাই ফাউল গেলো
Stop word Removing	অনেক ভালো লেগেছে অসাধারণ	অনেক ভালো লেগেছে অসাধারণ
Punctuation Removing	অনেক অনেক সুন্দর।	অনেক অনেক সুন্দর

3.5 Tokenization

Based on the post-preprocessing perspective, we separated the dataset into two kinds (normal and hate speech). because calculations do not clearly understand the string. We therefore had to turn our digested data into numbers. To do this, we applied the Term Frequency-Inverse Text Frequency (TF-IDF) method. We base our work on word level. Because of this, we had to tokenize each word in a phrase. We utilized the TFIDF technique to transform words to numerical vectors after tokenization. This table provides a tokenization sample. Table 3.3 represents tokenization process.

TABLE 3.3: TOKENIZATION

Raw Data	Type	Tokenized data
ট্রেলার দেখে অপেক্ষায় ছিলাম অপেক্ষা সার্থক হলো অনেক ভালো লাগছে	Normal	‘ট্রেলার’ ‘দেখে’ ‘অপেক্ষায়’,‘ছিলাম’ ‘অপেক্ষা’ ‘সার্থক হলো’, ’অনেক’ ,’ ভালো’. ’ লাগছে’
এরকম বিশ্রী এবং অশ্লীল নাটক বন্ধ করুন	Hate	‘এরকম’, ‘বিশ্রী’, ‘এবং’, ‘অশ্লীল’ ‘নাটক’, ‘বন্ধ ‘করুন’

3.6 Labeling

Our data was split into two categories: speech that is hateful and speech that is commonplace. When designing the courses, user emotions are taken into account. As a result, we divided our whole database into two groups. In our sample of 10,000 data points, 33.1% were deemed to be hate speech, while 66.9% were deemed to be normal. The imbalance in our dataset is shown in this graph 3.2.

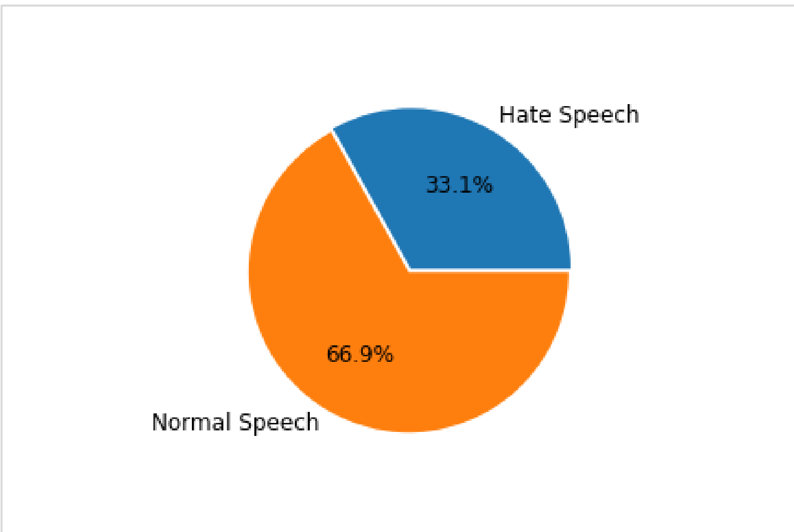

Figure 3.2: Classification of dataset.

3.7 Algorithm Implementation

We showed that Support vector machine produced the most accurate results using 30% of the test data and the five calculations mentioned in Table 3.4. Numerous calculations also worked flawlessly. Since it yields the best results, we decide to evaluate the emotional content of the Bangla post using logistic regression. Table 3.4 displays the values and other elements that we utilized to carry out the chosen algorithms.

TABLE 3.4 PARAMETER USAGE

Algorithms	Details
KNN	K=3,p=2,random_state=0
Decision Tree	random_state= 0
SVM	Kernel = linear
Random Forest	Number of estimators = 100
SGD	random_state= 0

3.8 Evaluation

With the use of an uncertainty matrix and real-time data estimate, we assessed the effectiveness of our chosen logistic regression approach. To begin with, we gathered a total of 195 hatred and everyday phrases that our algorithm never picks up on for the final evaluation of the model represented by figure 3.3. Hate speech is correctly identified in about 118 of 122 instances. For typical speech, our algorithm accurately detects 67 out of 73 instances, which are shown by blue bars. The orange hue indicates the anticipated value. Our model predicts three less favorable evaluations. The model of unfavorable reviews predicts similar reviews. Our model has a slight defect like this. We may assume that our model performed well with data from the actual world as a consequence. This prediction can also be tested using a confusion matrix.

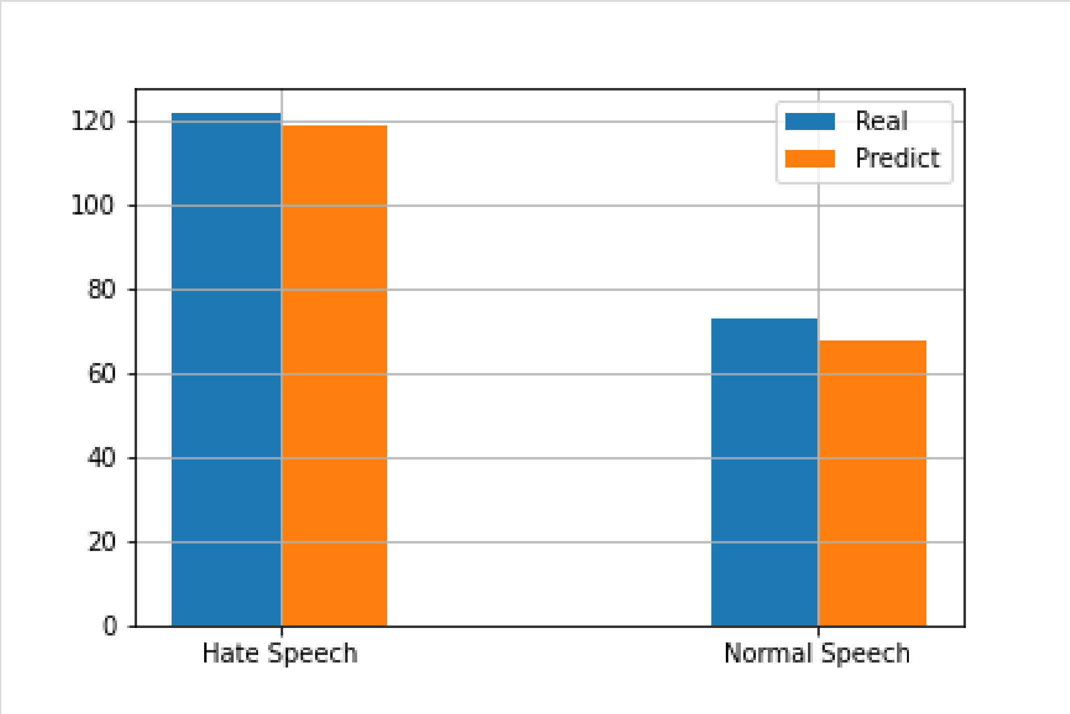

Figure 3.3 Real and expected classifications are compared.

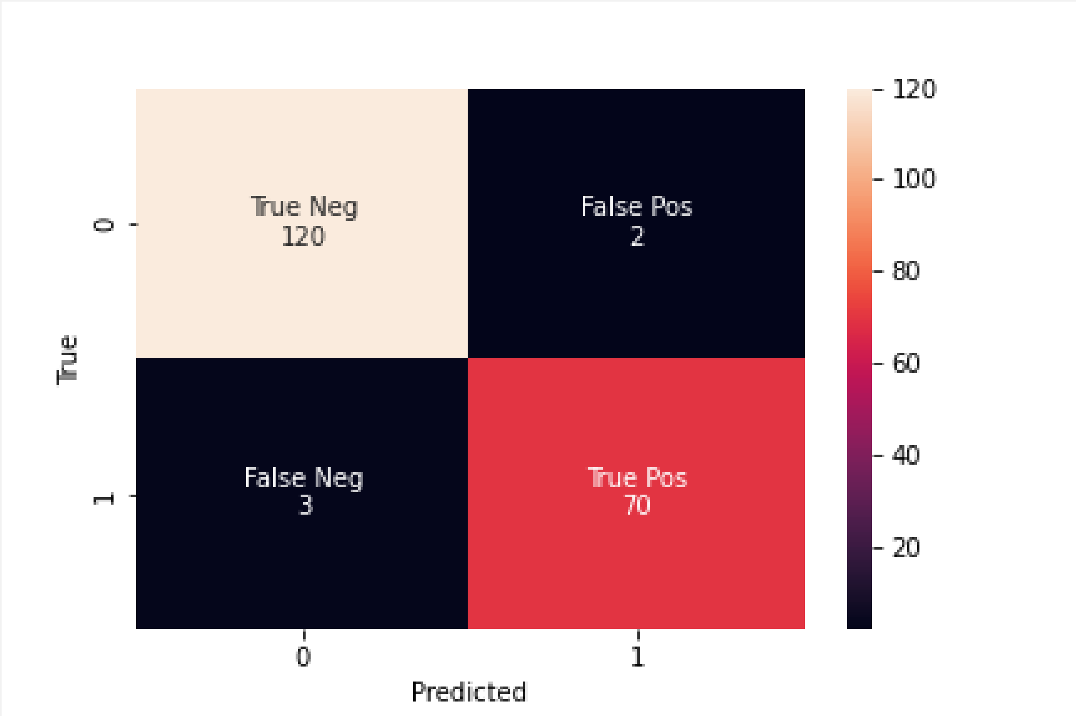

Figure 3.4: Confusion Matrix

$$\text{Accuracy} = \frac{120+70}{120+70+2+3} = 0.9743 * 100$$

$$= 97.43\%$$

$$\text{Error} = 1 - 0.9743 = 0.025 * 100 = 2.5\%$$

Recall rate for positive:

$$\frac{70}{70+3} = .958 * 100 = 95.89\%$$

$$\text{Recall rate for Negative: } \frac{120}{120+2} = 0.98 * 100 = 98\%$$

We used the Confusion Matrix to find overall results. Figure 3.4 displays the uncertainty matrix for the validation dataset. The assessment of test results is shown by this confusion matrix. 97.43% of test results are accurate. There is a 2.5 percent mistake rate. recall rates for positive and negative are respectively 98 percent and 85.89 percent. We may conclude from this calculation that our approach is excellent for communication that is critical or hateful. It serves as a wonderful example of our model rather than being a positive for negative feedback.

CHAPTER 4
RESULT ANALYSIS

4.1 Introduction

The field of applied machine learning relies heavily on controlled experimentation. Due to their complexity, machine learning approaches are resistant to traditional forms of examination. As a result, we need to have an empirical understanding of how algorithms perform on the problems we care about. For this, we employ carefully orchestrated experiments. The experimental outcomes of our examination and a descriptive analysis of the data used in the research are the focus of Chapter 4.

4.2 Experimental Result

To evaluate the precision of our results, we employed five unique methods and pre-processed data. And for better result we tried to create two different datasets. On is with stop word and another is without stop word. And applied both of dataset to our designed model. So at first we analysis the accuracy of two datasets and then apply another parameter for result experiment. An accuracy score was determined utilizing the accuracy value returned by these algorithms since we can easily interpret and differentiate between them based on their findings. We used 30-40-50-60-70-percent, 50-60-70-percent, 60-70-percent, and 70-80-percent complete data training sets to test for consistency. Incredibly, after comparing the five approaches, we got a fantastic outcome. Table 4.1 displays the results of these five methods of accuracy with stopword. Where the red rectangle border box indicating that the SVM produced the best accuracy about 96%. The highest precision achievable by such algorithms with varying data utilization percentages is shown by a yellow dot in each column.

TABLE 4.1 ACCURACY TABLE WITH STOPWORD

Test Data usage rate	Algorithms				
	KNN	*Decision Tree*	*SVM*	*Random Forest*	*SGD*
30%	82.90%	92.26%	96.28 %	94.83%	94.96%
40%	82.79%	92.19%	95.74 %	94.88%	94.69%
50%	83.02 %	91.73 %	96.01%	94.49%	94.80%
60%	82.11%	91.13%	95.64%	94.32%	94.93%
70%	81.32%	90.49%	95.29%	94.25%	94.08%

TABLE 4.2 ACCURACY TABLE WITHOUT STOPWORD

Test Data usage rate	Algorithms				
	KNN	*Decision Tree*	*Random Forest*	*SVC*	*SGD*
30%	84.64%	93.35%	94.58%	95.18%	93.23%
40%	84.40%	91.76%	94.37 %	95.18%	93.30 %
50%	83.19 %	91.19%	93.63%	94.86%	93.18%
60%	82.31%	94.51%	93.33%	94.51%	92.95%
70%	81.81%	94.09%	93.10%	94.09%	92.72%

Table 4.2 represents the accuracy table without stopwords. We can see that the highest accuracy is achieved by SVC algorithm also but if we compare previous data then we can see that model performed better with stopword dataset. so we have decided to use dataset with stopword as our final dataset of this work.

TABLE 4.3 F1-SCORE MATRIX TABLE

Score Matrix	Algorithms				
	KNN	*Decision Tree*	*SVM*	*Random Forest*	*SGD*
F1 Score	80.81	92.09	96.20	95.11	94.82
Recall	68.37	90.22	94.84	93.64	92.30
Precision	98.78	94.05	97.60	96.62	97.48
sensitivity	68.37	90.22	94.84	93.64	92.3
Specificity	75.6	90.5	94.92	93.77	92.61

As accuracy is not only the factor to choose machine learning model. Because accuracy only work with true positive and true negative. But if we see confusion matrix then we can see that false positive and false negative also there. And f1 score precision, sensitivity and specificity can measure the parameter also. When we used 30 comments, the SVM gave the greatest F1 score with 96 and RF gave 95 percent, as seen in table 4.3. so we have decided to use svm as our final model.

KNN

Below is a graphic that compares KNN's recall, precision, accuracy, and F1 scores across a range of values represented by figure 4.1. This demonstrates the algorithm's efficacy in terms of its data utilization rate. With a score of 82.9, KNN is the most accurate method. Differences in recall, accuracy, F1 Score, and precision may be seen in the graph.

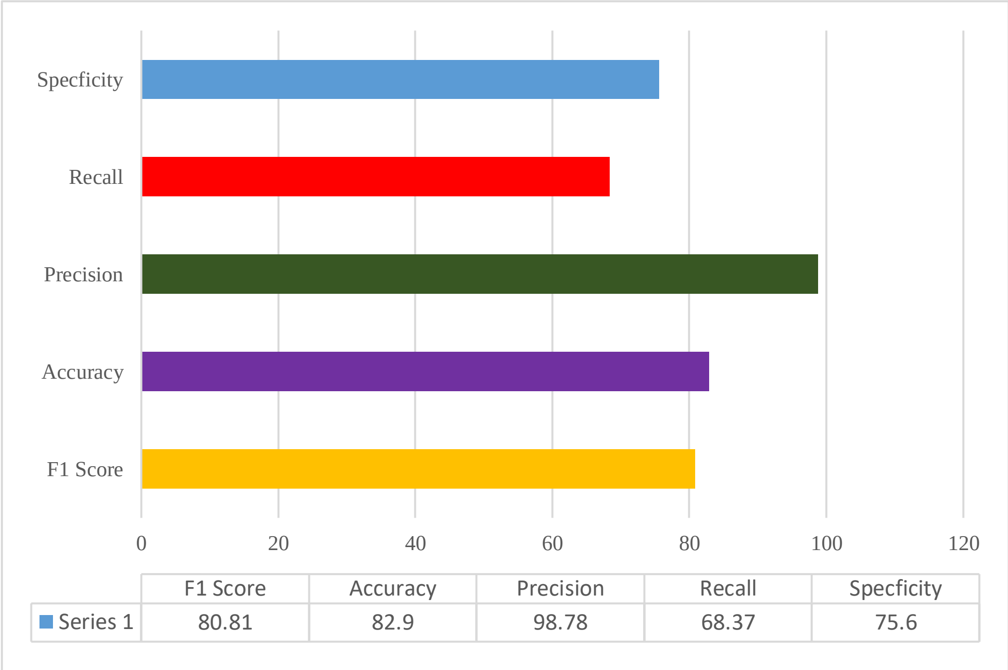

Figure 4.1 KNN's Different Score Comparing Graph

Decision Tree

Below is a graphic that compares Decision Tree's recall, precision, accuracy, and F1 scores across a range of values represented by figure 4.2. This demonstrates the algorithm's efficacy in terms of its data utilization rate. The highest precision score was 92.26, which was achieved by the Decision Tree algorithm. Variations in recall, accuracy, F1 Score, and precision are shown on the graph.

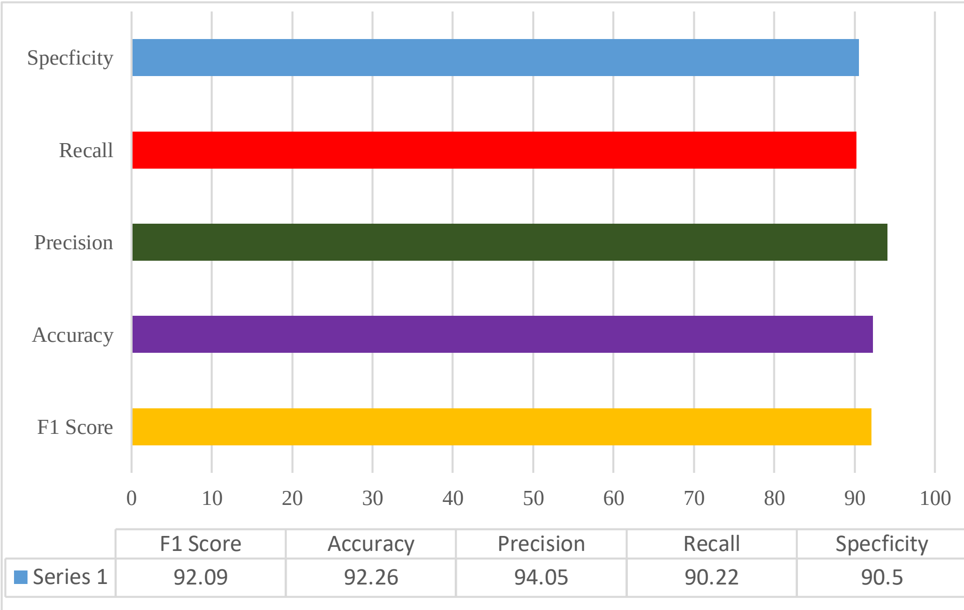

Figure 4.2: Decision Tree 's Different Score Comparing Graph

SVM

This chart compares and contrasts the SVM recall, precision, accuracy, and F1 scores represented by figure 4.3. The data utilization rate allows us to evaluate the algorithm's efficiency. SVM achieves a perfect 96.28 percent accuracy. Differences in recall, accuracy, F1 Score, and precision are displayed on the graph. When compared to other scoring matrices, Svm always performed better.

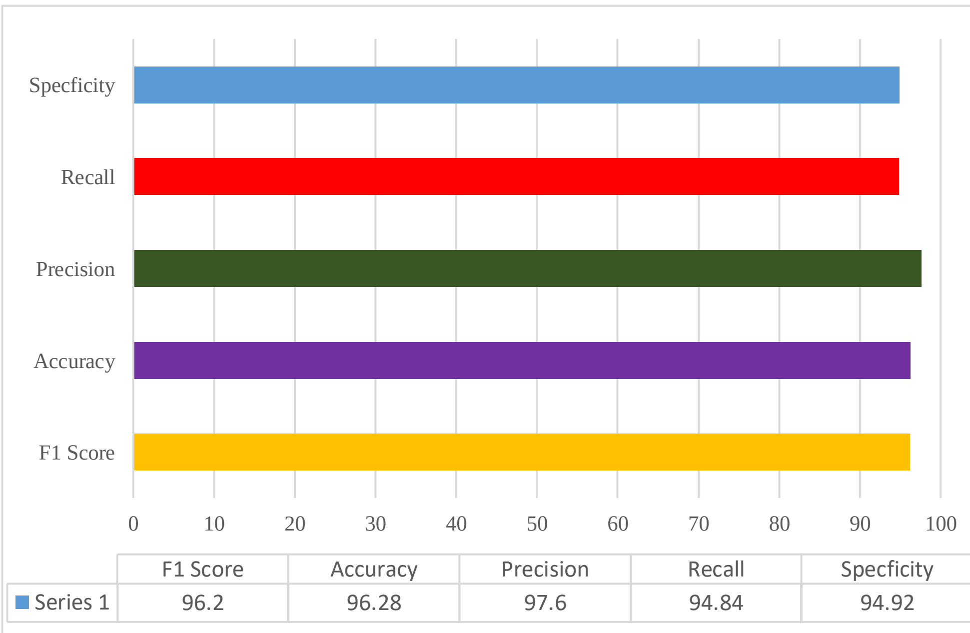

Figure 4.3: Different Score comparison graph of SVM

Random Forest

In other words, random forest is a supervised training technique. You may employ it for both regression and classification. Furthermore, it is the most flexible and intuitive algorithm available. Below is a comparative table for Random Forest's recall, precision, accuracy, and F1 scores represented by figure 4.4. This demonstrates the algorithm's efficacy in terms of its data utilization rate. With a 94.64 percent accuracy rate, Random Forest is the best method. Recall, accuracy, F1 Score, and precision performance gaps are shown graphically in Figure 6.

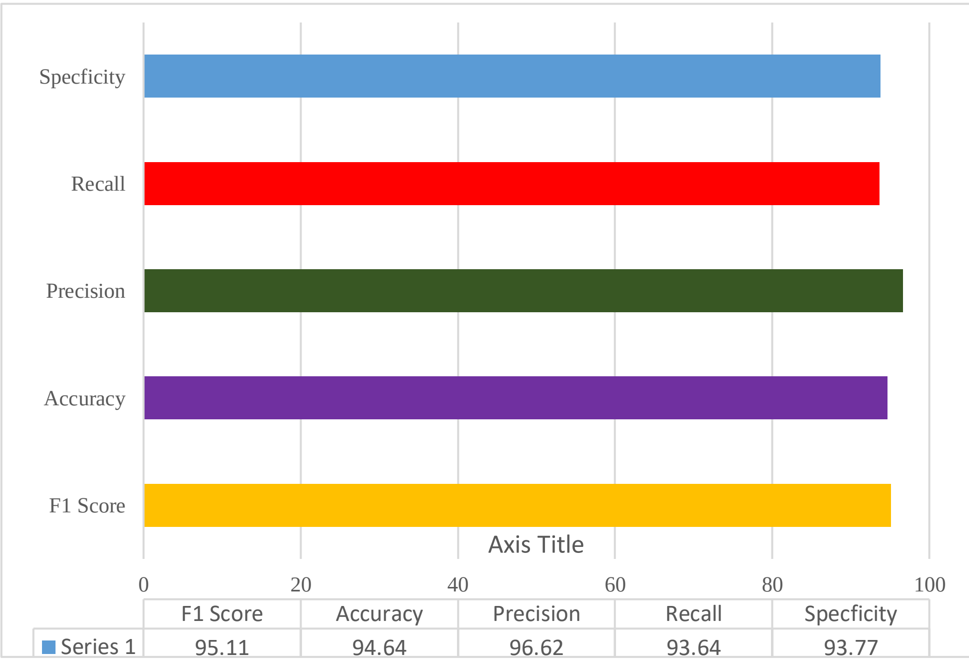

Figure 4.4: Different Score comparison graph of Random Forest

SGD algorithm

Check out this chart to see how different SGD algorithm metrics, such recall, precision, accuracy, and F1, are related to one another. SGD's numerous outputs are seen in Figure 4.5. Precision, recall, and F1 Score are three crucial metrics to look at when evaluating the efficacy of a machine learning method. All of them contributed to successful outcomes in our studies. Finally, we settled on SGD as the classification system for the advertisements. The data utilization rate allows us to evaluate the algorithm's efficiency. When it comes to accuracy, Logistic Regression is tops with 94.64%.

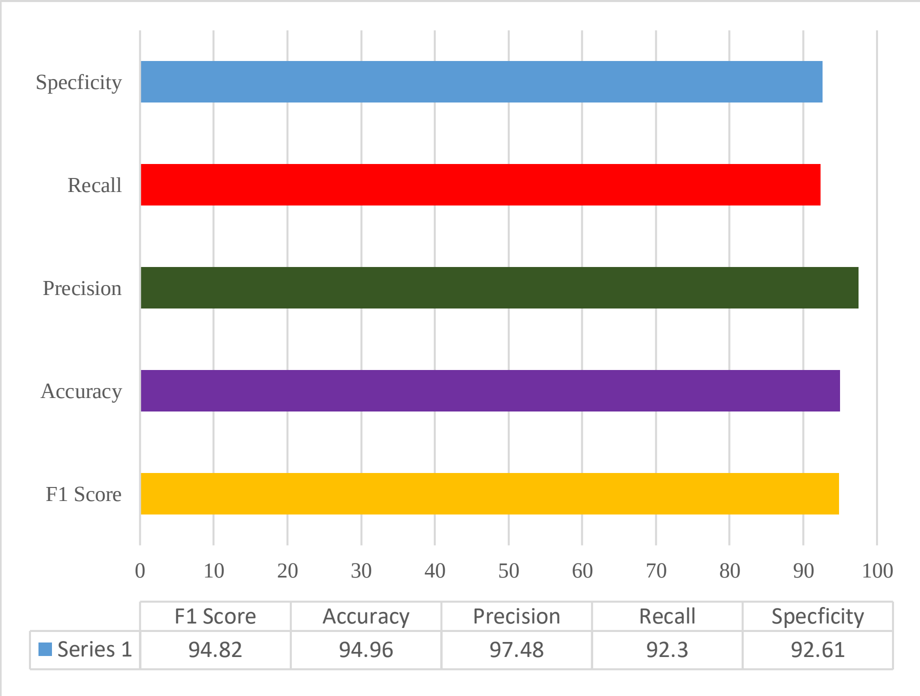

Figure 4.5: Different Score comparison graph of SGD algorithm.

CHAPTER 5

SUMMARY, CONCLUSION AND FUTURE WORK

5.1 Summary of the Research

While machine learning has been the subject of a lot of study, not much of it has been done in Bangladesh. Research in predictive styles may be a term in the field of computer science education, but the practical benefits that have resulted from this study have been quite astonishing. The lack of similar studies on the Bangla language, however, is grounds for serious concern. We expect many researchers from many nations to investigate this problem. Several factors are used to categorize our Bangla advertisements in our studies.

5.2 Conclusion

Due to the anonymity and portability of social media platforms, as well as the ever-changing political context in many areas of the world, hate speech has increased in frequency in recent years.

Sentiment analysis is a powerful method for gleaning insights from the minds of others. It helps us identify feelings communicated through speech, writing, or even physical manifestations. This study presents a machine-learning-based approach to sentiment classification, which is able to classify evaluations of hate speech into positive and negative categories. We conducted this study with data from a massive dataset consisting of 10,000 comments and 2,500 examples of hate speech. By combining the techniques of the Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Stochastic Gradient Descent (SGD), and K-Nearest Neighbor (KNN) algorithms, we were able to develop a model that can distinguish between neutral and offensive language in Bangla comments. The research found that a 96 percent accuracy rate could be achieved using the SVC method.

5.3 Recommendations

Here are a few excellent recommendations to consider in this regard:

- If we want more reliable results from this research, we need to improve data collection.
- In addition, expanding the dataset's size might help provide better outcomes.
- The use of Deep Learning is recommended.

5.4 Future Work

Here are some guidelines on how this project should go in the future:

- We are working on a web-based API to identify hate speech as a means to this end.
- In the future, we want to use deep learning methods to create an intelligent system.
- larger dataset may improve the accuracy of the algorithms in the future.
- Future expansion of this project's categorization structure may improve its efficiency.

REFERENCE

[1] Bangladesh Sees Rise In Disinformation, Hate Speech And Violence Against Religious Minorities During The Covid-19 Pandemic Available at <<https://www.media-diversity.org/bangladesh-sees-rise-in-disinformation-hate-speech-and-violence-against-religious-minorities-during-the-covid-19-pandemic>> last accessed on 07-10-2022.

[2] D. Chaffey. (17 april 2020). Global social media research summary 2020 Available: <<https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>> last accessed on 08-11-2022.

[3] S. KEMP. (17 FEBRUARY 2020). DIGITAL 2020: BANGLADESH. Available at << https://www.dhakatribune.com/world/2020/02/17/bengali-ranked-at-7th-among-100-most-spoken-languages-worldwide >> last accessed on 10-12-2022.

[4] Share of women worldwide who have ever experienced abuse or harassment on selected websites and social media platforms as of July 2017. Available at <https://www.statista.com/statistics/784821/harassment-women-websites-social-media-platforms> > last access on 08-30-2022

[5] Billah, Masum, and Enamul Hassan. "Depression Detection from Bangla Facebook Status using Machine Learning Approach." International Journal of Computer Applications 975: 8887.

[6] M. T. Ahmed, M. Rahman, S. Nur, A. Islam and D. Das, "Deployment of Machine Learning and Deep Learning Algorithms in Detecting Cyberbullying in Bangla and Romanized Bangla text: A Comparative Study," 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2021, pp. 1-10, doi: 10.1109/ICAECT49130.2021.9392608.

[7] C. Chauhan and S. Sehgal, "Sentiment analysis on product reviews," 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 26-31, doi: 10.1109/CCAA.2017.8229825.

[8] T. Islam, A. R. Bappy, T. Rahman and M. S. Uddin, "Filtering political sentiment in social media from textual information," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016, pp. 663-666, doi: 10.1109/ICIEV.2016.7760084.

[9] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter and A. K. Das, "An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019, pp. 360-364, doi: 10.1109/CCOMS.2019.8821658.

[10] Mittal, Namita, et al. "Sentiment analysis of hindi reviews based on negation and discourse relation." Proceedings of the 11th Workshop on Asian Language Resources. 2013.

[11] M. G. Hussain, T. A. Mahmud and W. Akthar, "An Approach to Detect Abusive Bangla Text," 2018 International Conference on Innovation in Engineering and Technology (ICIET), 2018, pp. 1-5, doi: 10.1109/CIET.2018.8660863

APPENDIX

Choosing a methodological approach for our inquiry was the first challenge we confronted during the research process. The work was unusual, as the area had seen relatively little action up to that point. So nobody was really able to assist us out.

Data processing also proven to be a difficult obstacle for us. Due to the lack of publicly available data for a Bangla text pre-processing system, we built our own corpus. Furthermore, we initiated a manual data collection process. In addition, it is difficult to properly categorize numerous advertisements. After much effort, we may be able to succeed.

PLAGIARISM REPORT

AI report

ORIGINALITY REPORT

19%	15%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	11%
2	Submitted to Daffodil International University Student Paper	2%
3	Md Gulzar Hussain, Tamim Al Mahmud, Waheda Akthar. "An Approach to Detect Abusive Bangla Text", 2018 International Conference on Innovation in Engineering and Technology (ICIET), 2018 Publication	1%
4	Submitted to University of the Free State Student Paper	1%
5	infoautoclassification.org Internet Source	1%
6	Rashedul Amin Tuhin, Bechitra Kumar Paul, Faria Nawrine, Mahbuba Akter, Amit Kumar Das. "An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques", 2019 IEEE 4th	1%