BANGLA HATE SPECCH DETECTION USING MACHINE LEARNING

BY

Asfi Hossain Choudury

ID: 171-15-9571

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

Supervised By

Mr. Raja Tariqul Hasan Tusher

Senior Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Mr. Aniruddha Rakshit

Senior Lecturer

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

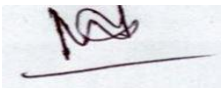
DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

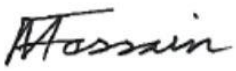
This project titled "**BANGLA HATE SPEECH DETECTION BY MACHINE LEARNING**", submitted by **Md. Asfi Hossain Choudury** , ID No: 171-15-9571 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements 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 2 January, 2022.

BOARD OF EXAMINERS



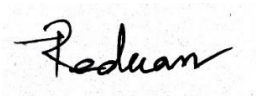
Chairman

Dr. Md. Ismail Jabiullah
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Dr. Md. Fokhray Hossain
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Md. Reduanul Haque
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Mohammad Shorif Uddin
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that this thesis has been done by us under the supervision of **Raja Tariqul Hasan Tusher, Senior Lecturer, 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:

Raja Tariqul Hasan Tusher
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Asfi Hossain Choudury
ID: 171-15-9571
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 Raja Tariqul Hasan Tusher, Senior Lecturer, 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 type of speech that takes place online intends to attack an individual or a group based on religion, ethnicity, gender, disability and even based on the color of their skin. Some popular social media in Bangladesh notably Facebook and Youtube filled with these types of speeches. The comment section of celebrities can be a perfect example of people spreading hatred. In recent times we can also see religious clashes and cases of suicide in Bangladesh because of the spread of hate speech. Filtering these types of comments and opinions has become a need to make social media free of negativity. So detecting hate speech in the Bangla language has been our primary goal. There had been a few previous works, but they were not up to the mark. A significantly large dataset is being used consisting of more than eight thousand comments collected from different social media platforms. We Introduced a model that classifies Bangla comments into normal speech and hates speech by implementing Support Vector Machine (SVM), Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbor(KNN) algorithms. Our model with the help of specific calculation provides the most dependable result in Bangla Language. After analysis results of all algorithm, we choose the best model that is produced best accuracy for test data.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE NO.**

LIST OF TABLES

TABLE NO. **PAGE NO.**

CHAPTER 1

INTRODUCTION

1.1 Introduction

Social media is a powerful tool for bringing people together and allowing them to network, share information, and express their personal views.

In 2020, almost 3.6 billion individuals will be accessing social media, with a growth rate of 49% as of January 2020.[1] One of the most popular digital hobbies nowadays is spending time on social media platforms such as Facebook, Instagram, Twitter, and YouTube.

Unfortunately, it is frequently contaminated with negativity, including hate speech. The use of offensive language in user-generated content on internet platforms, as well as its consequences, has gotten a lot of attention in recent years. What precisely is hate speech? There are no hard and fast guidelines for identifying hate speech for what it is. Hate speech, at its most basic, harasses, intimidates, or incites violence against someone because of their race, skin color, ethnic group, religion, gender, or sexual orientation, effectively harassing, intimidating, or inciting violence against them for being who they are. [2]

In 2020, 265 million native speakers spoke Bangla, making it the world's seventh most spoken native language by population. This total population represents 3.05 percent of the world's total population. [3]. In January 2021, Bangladesh had 47.61 million internet users [4] just people had easy access to the internet. But when a lot of people started coming online they started to write whatever they wanted on social media. We observe more hate comments than positive comments in celebrity posts, particularly in Bangladesh. Sentiment analysis, also known as opinion mining, has sparked a lot of interest and growth in the research community because it has a lot of practical and realistic applications. It analyzes people's feelings, opinions, behavior, attitudes, and emotions toward individuals, organizations, products, services, and situations, as well as their attributes, using written or spoken language. Sentiment analysis, on the other hand, is a well-studied area in English, but not so much in Bangla. In our research, we developed a model for determining whether a statement is hate speech or not. To classify the data set, we utilized various supervised learning algorithms such as KNN, Decision Tree, Support Vector

Machine (SVM), Random Forest, and Logistic Regression to see which approach performs best with the maximum accuracy.

We also employed NLP libraries such as NLTK (Natural Language Toolkit), TF-IDF, and TF-IDF, as well as separate ML toolkits such as learning Sci-kit, Numpy, Matplotlib, and Pandas to achieve these objectives.

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?

1.4 Research Methodology

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

1.5 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 This first section is a crucial part of the preliminary investigation.In addition, this chapter discusses why we decided to conduct such research. The most important component of this chapter is the problem definition. The study issue, as well as the challenge, are included in this segment.

Chapter 2 This is composed of an input analysis that gives a brief overview of the work done in this field. Here is a description of some of the related notable machine learning work.

Chapter 3 is a simple workflow summary. What was the outcome of the analysis in this segment?

Chapter 4 It's in the evaluation of the outcomes. It comprises the graphic analysis' outcomes.

Chapter 5 It's the final section of the research. This section explains the model's output. This portion also demonstrates the accuracy of the relationship. This part also includes the concept and performance's online implementation. The chapter ends with a discussion of the work's shortcomings. The study's potential was 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

Mining and evaluating sentiments from social network data can be useful in a variety of domains, including prediction, analyzing the general public's mood on a particular social issue, and so forth. Sentiment analysis has emerged as a significant focus area in natural language processing. This Chapter summarizes significant research conducted by other researchers.

2.2 Related Works

Hate speech is a type of writing that defames and is likely to cause harm or danger to the victim on social media. A subjective sentence expresses feelings, points of view, or beliefs. When necessary, the subjective sentence can be divided into positive and negative classes. Sentiment Analysis is a sentimental analysis method that follows Aspect-based and selects a single subject and analyzes the feelings associated with that subject. With this approach,

Masum Billah et al.[5] proposed a model that uses machine learning to predict whether or not a person is depressed based on their Bangla Facebook status. The work is based on the data of 50 Facebook users, 17 of whom committed suicide. The SGDC with unigram + emoticons as a feature had the highest accuracy of 77.96 percent. Due to a lack of data, only traditional machine learning procedures were used in their paper.

Cyberbullying against people has also increased on social media.. M. T. Ahmed et al.[6] created a model to detect cyberbullying in Bangla and Romanized Bangla texts by utilizing Machine Learning and Deep Learning algorithms. In addition, They presented a comparison of the algorithms in terms of accuracy, precision, recall, f1-score, and roc area. From social media, They created three datasets: Bangla, Romanized Bangla, and a blend of the two.There were 5000 Bangla, 7000 Romanized Bangla, and a combination of 12000 Bangla and Romanized Bangla texts in the three datasets. The Bangla Dataset obtained the best results, with CNN obtaining an accuracy of 84 percent.

C. Chauhan et al.[7] used machine learning algorithms to distinguish between negative and positive feedback on a product for potential users to authenticate their reviews. They reviewed

various papers and concluded that Naïve Bayes produced promising results, but the results varied depending on the setup and methodology used, as well as the objectives.

Taimur Islam et al.[8] created a dictionary of unique words collected from political and nonpolitical postings and comments, and then trained it using the probability theory-based Naïve Bayes method. They used 200 Facebook posts to test their algorithm, and the results demonstrate that the approach can accurately classify posts or comments.

R. A. Tuhin et.al.[9] suggested two machine learning techniques to extract emotion from any Bangla text by using Naïve Bayes Classification Algorithm and the Topical Approach.. A data collection of 7400 Bangla phrases was used, with a topical approach providing 90% accuracy.They then compared their paper to two others, both of which scored 93 percent for SVM and 83 percent for document frequency.The emotional parameter in each of the three articles was different.

Namita, et al.[10] offered a method for increasing the coverage of the Hindi SentiWordNet in order to improve classification results. The paper described an educational program that investigates the Roman Urdu people's emotions through the genres of sports, software, food and recipes, drama, and politics. It contains 10,021 sentences culled from 566 internet discussions. Their proposed method yields an overall accuracy of 80.21 percent, with 82.89 percent for positive reviews and 76.59 percent for negative reviews.

Detecting abusive text in Bangla can help to avoid cybercrime such as online harassment, blackmail, and cyber bullying, which are becoming a major worry in Bangladesh these days. M. G. Hussain et al.[11] discovered abusive Bangla comments that were collected from numerous social media platforms where people share their sentiments, thoughts, and views. They trained our system with 250 comments and tested it with 50 comments.They proposed a root-level approach for detecting abusive content, as well as unigram string properties to improve the outcome.

Sigurbergsson et al.[12] created a Danish dataset including user-generated Reddit and Facebook comments. They construct four automatic classification systems, each of which is meant to function with both English and Danish. A macro averaged F1-score of 0.70 is achieved by the best performing system for Danish. In the detection of whether or not an

offensive post is targeted, the best performing system for English achieves a macro averaged F1-score of 0.62

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

The methodology is broken down into six stages, the last of which is seen in Figure 3.1. The following are the measures to take:

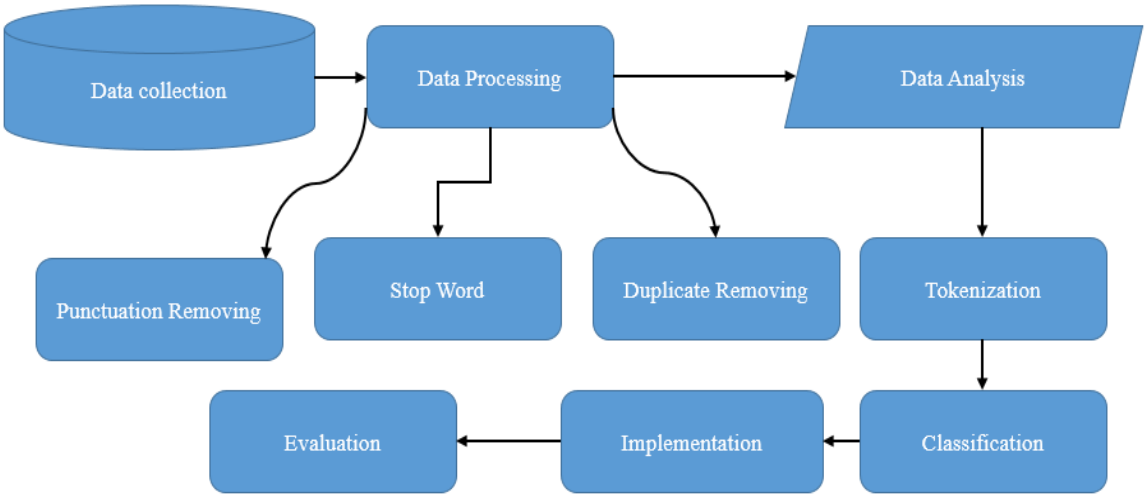


Figure 3.1: Methodology diagram

3.2 Data gathering

Every day, hundreds of billions of posts are posted on social media. Each point of view creates a specific emotion. We mainly focused on Bangla posts on social media because we mostly dealt with Bangla. We then gathered both credible and filthy Bangla texts via social media. For the conclusions of our research, we gathered about 6000 comments.

3.3 Pre-Processing

Data preprocessing is a data mining approach that transforms unprocessed data into a more useable and efficient format. It is impossible to overestimate the value of information preparation in acquiring knowledge. Information cannot be used directly in the execution of calculations. As a result, it must be preprocessed before the computation is run. As a result,

information must be transferred in a systematic manner, which necessitates preparation. We started by eliminating Bangla accentuation and Bangla punctuation marks from the data. Then we deleted it from the inside of the post emoticon. We inserted the emoticon at the conclusion so that we could precisely deconstruct the sentence's sentiment. We couldn't eliminate emoticons from postings as a component of the feeling indefinitely.

3.4 Data Analysis

We divided the dataset into two types based on the post-preprocessing viewpoint (normal and hate speech). Because the string isn't explicitly understood by computations. As a result, we needed to convert our digested data into numbers. We used the TF-IDF technique (Term Frequency–Inverse Text Frequency) to do this:

$$W_{i,j} = tf_{i,j} \times log\left(\frac{N}{df_i}\right) \quad (1)$$

W i, j is a TF-IDF score in the TF-IDF (1). tf I j calculates the number of instances of I in j. The number of papers is N. And df i denotes the number of documents that include the letter i.

TABLE: 3.2 TOKENIZATION

Raw Data	Type	Tokenized data
এই শহর দিন দিন ফালতু হয়ে যাচ্ছে	Hate	‘ এই’ ,‘ শহর ’ ,‘ দিন ‘,‘ দিন’, ‘ ফালতু’, ‘ ‘হয়ে’, ‘ যাচ্ছে’
ক্রিকেটে সাকিব দিন দিন উন্নতি করছে	Normal	‘ ক্রিকেটে ’ ,‘ সাকিব ’, ‘ দিন ’ ,‘ দিন ’, ‘ উন্নতি ’, ‘ করছে ’
তার মত বিরক্তিকর মানুষ আগে কখনো দেখিনি	Hate	‘ তার ’,’ মত ’ বিরক্তিকর ’,’ ‘ মানুষ ’ , ‘ আগে ’, ‘ কখনো ’, ‘ দেখিনি ’

3.5 Tokenization

Tokenization is a crucial aspect of NLP Tokenization separated particular terms from a statement, which is essential for sentence analysis. Table 3.1 attempts to show the tokenization technique. Before using the TF-IDF approach, any string must be tokenized. As displayed in Table 3.1, the social media form of raw data was used as an input. Punctuation and a stop word were then eliminated. Then, to tokenize our raw data, we used a porter stemmer, which is often used in tokenization techniques.

3.6 Classification

Our data was separated into two categories: normal speech and hate speech. The courses are created with the user's feelings in mind. As a result, we've divided our whole database into two groups. There were around 3000 Bangla comments in each class.

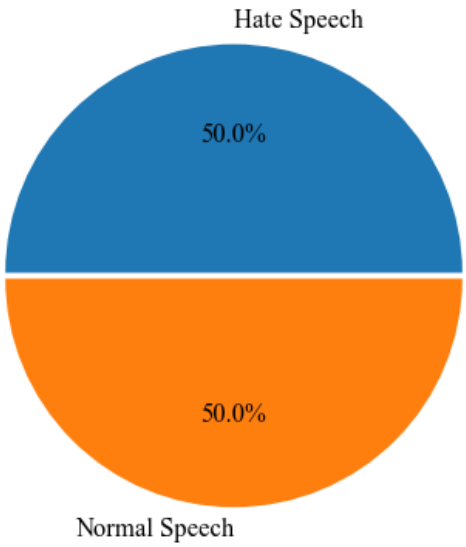


Figure 3.2: Classification of dataset.

3.7 Algorithm Implementation

By employing 50% of the test data and applying five computations listed in Table 3.2, we demonstrated that Logistic Regression provided the most precise results. Different computations also performed beautifully. We choose to use Logistic Regression to estimate the emotion of the Bangla post since it produces the greatest result. The values and other items that we used to execute the selected algorithms are shown in Table 3.2.

TABLE 3.2 PARAMETER USAGE

Algorithms	Details
Logistic Regression	max_iter=200 , cv = 5
SVM	Kernel = linear
Decision Tree	random_state = 42
Random Forest	Number of estimators = 150
KNN	random_state=0 , p=3, K=4

3.1Evaluation

We evaluated our selected logistic regression technique utilizing real-time data estimation and an uncertainty matrix. Initially, we collected 220 actual data points from which our model failed to learn. For each of the classes chosen, different social media platforms for Bangla comments were used. Figure 3.3 shows a comparison of the actual and predicted results. There are 90 positive reviews and 130 negative reviews in our dataset, which are represented by blue bars. The expected value is shown by the orange color. Three less favorable ratings are

predicted by our algorithm. The negative review model predicts same review. This is a minor flaw in our model. As a result, we may presume that our model worked well with real-world data. Confusion matrix can also be used to test this forecast.

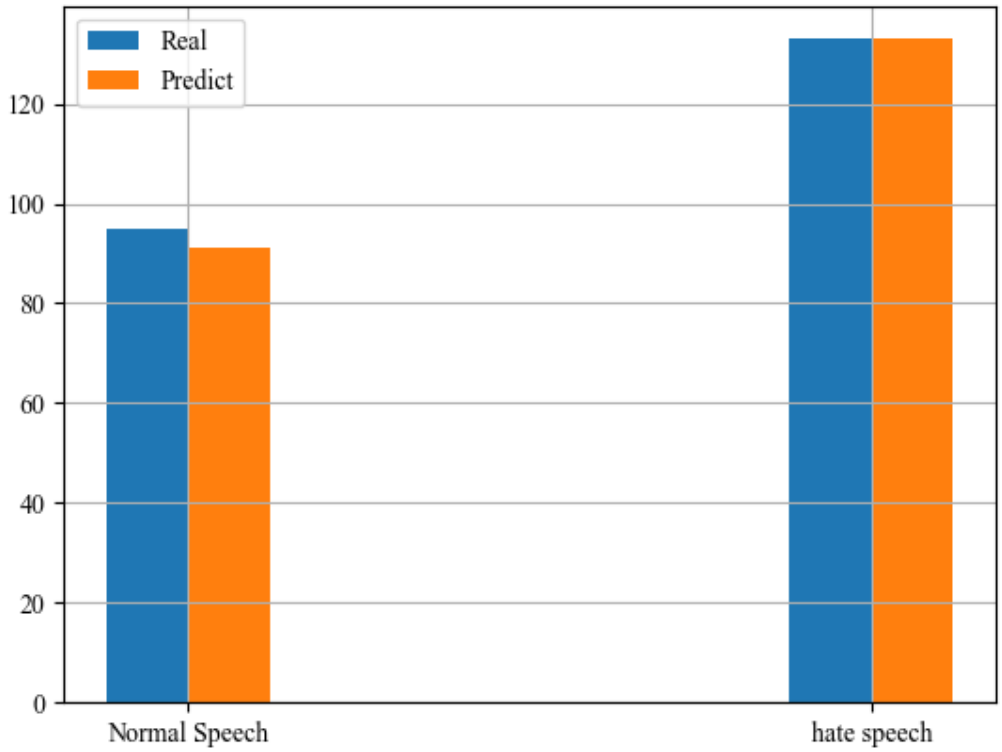


Figure 3.3 Real and expected classifications are compared.

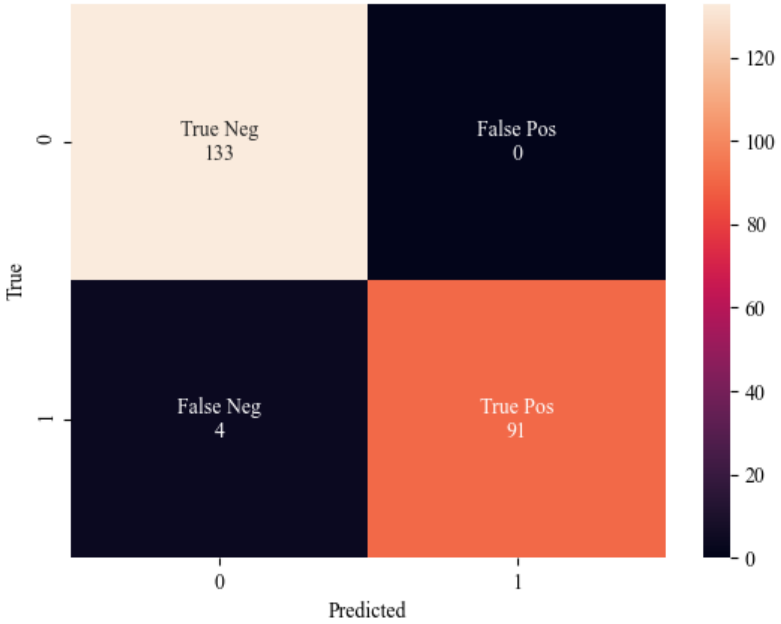


Figure 3.4: Confusion Matrix

$$\text{Accuracy} = \frac{133+91}{133+0+4+91} = 0.9824 * 100$$

$$= 98.24\%$$

$$\text{Error} = 1 - 0.9824 = 0.0176 * 100 = 1.76\%$$

Recall rate for positive:

$$\frac{91}{91+4} = .958 * 100 = 95.79\%$$

Recall rate for Negative: $\frac{133}{133+0} = 1 * 100 = 100\%$

To detect overall outcomes, we employed the Confusion Matrix. The validation dataset uncertainty matrix is shown in Figure 3.4. In the assessment procedure, we have a precision of 98.24%. This also means that our model works with both visible and hidden data. The percentage of positive memory is 95.79 percent, whereas the rate of negative recall is 100 percent. Rather than being a positive for bad reviews, it's an excellent illustration for our model.

CHAPTER 4
RESULT ANALYSIS

4.1 Introduction

The primary subjects of Chapter 4 are the descriptive analysis of the data utilized in the research, as well as the experimental results of our investigation.

4.2 Experimental Result

We used five distinct approaches using pre-processed data to calculate the accuracy of our work. Because we can readily comprehend and distinguish between these algorithms based on their findings, an accuracy score was calculated using the accuracy value provided by these algorithms. Table 4.1. To evaluate consistency, We utilized five different complete data training sets: 30 percent, 40 percent, 50 percent, 60 percent, and 70 percent. A terrific result was obtained by comparing the five methods. The logistic regression approach, which used 30% training data and had an accuracy rate of 97.09 percent, gave the highest accuracy of the five algorithms with an accuracy rate of 97.09 percent, as shown in Table 4.1 by the red rectangular border box. A yellow point in each column represents the maximum accuracy of such algorithms employing varied percent of data use-values.

TABLE 4.1 ACCURACY TABLE

Test Data usage rate	Algorithms				
	KNN	*Decision Tree*	*SVM*	*Random Forest*	*Logistic Regression*
30%	81.96%	93.35%	96.78%	95.64%	97.09%
40%	81.56%	92.61%	96.84%	95.49%	97.09%
50%	80.95%	91.59%	96.14%	94.86%	96.61%
60%	80.49%	91.34%	96.06%	94.61%	96.17%
70%	79.57%	90.00%	95.56%	93.86%	95.59%

TABLE 4.2 F1-SCORE MATRIX TABLE

Score Matrix	Algorithms				
	KNN	*Decision Tree*	*SVM*	*Random Forest*	*Logistic Regression*
F1 Score	0.7752	0.9206	0.9643	0.9479	0.9659
Recall	0.6351	0.8987	0.9458	0.9387	0.9474
Precision	0.9944	0.9435	0.9835	0.9572	0.9852
sensitivity	0.6351	0.8987	0.9458	0.9387	0.9474
Specificity	0.7351	0.9048	0.9486	0.9408	0.9501

The F1 score of our implemented methods is shown in Table 4.2. We also looked at F1 scores from different percentages of the test data. We also discovered an almost same outcome in terms of accuracy. Among all algorithms, Logistic Regression had the highest score.

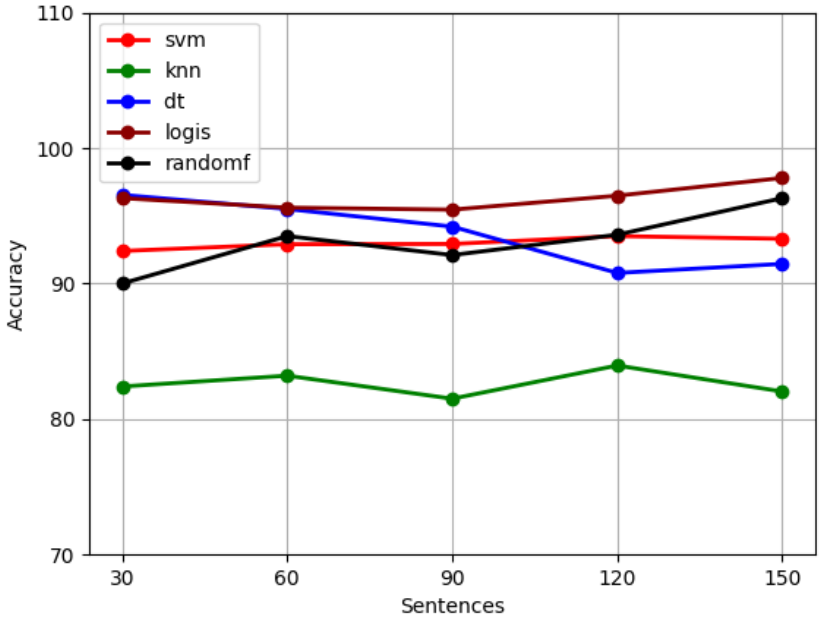


Figure 4.1 Accuracy plot

The DT and Logistic Regression give the best accuracy with 97 percent when we utilized 30 comments, as seen in this graph. Then, when we used 60 comments, we found that DT and Logistic Regression had the greatest accuracy (96%). When we utilized 90 comments, the greatest accuracy was 96 percent using Logistic Regression, while the DT was 95 percent. When we used 120 comments, the greatest accuracy was 97 percent for Logistic Regression and 94 percent for RF and SVM. When we utilized 150 comments, the greatest accuracy was 98 percent for Logistic Regression and 97 percent for RF. The overall accuracy of the Logistic Regression method is the best.

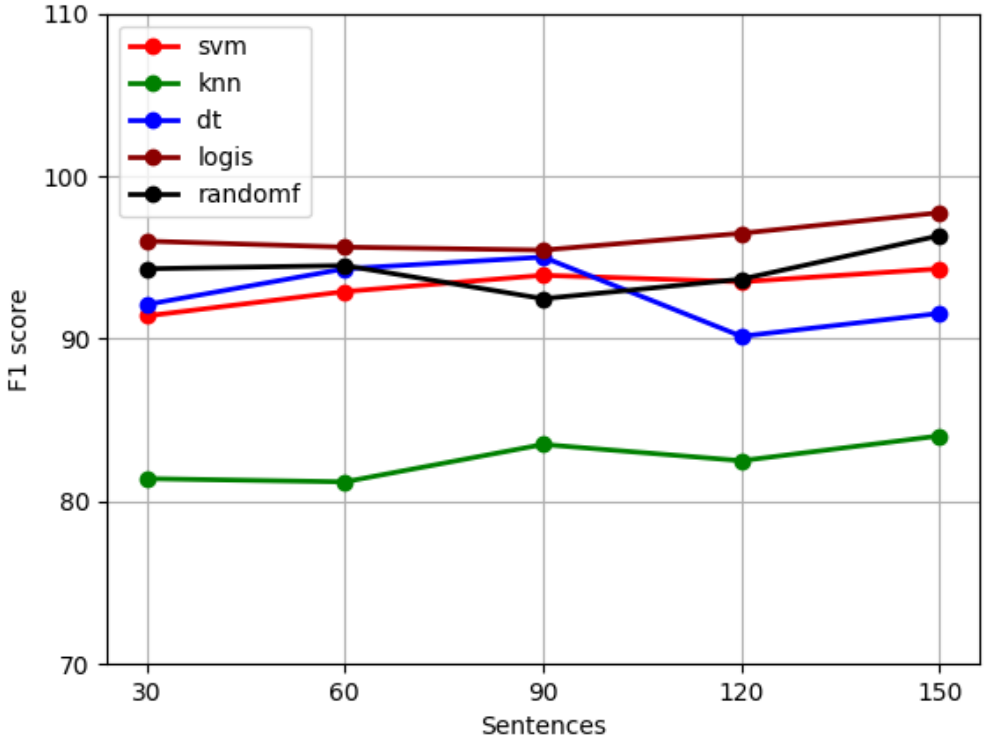


Figure 4.2 F1 score plot

When we used 30 comments, the Logistic Regression gave the greatest F1 score with 96 and RF gave 95 percent, as seen in this graph. Then, after using 60 comments, we discovered that Logistic Regression were the most F1 score with 96 and RF and DT gave 95 percent. when we used 90 comments, we found that DT and Logistic Regression had the greatest accuracy (96%). The highest F1 score was 97 percent for Logistic Regression and 94 percent for RF and SVM when we used 120 comments. The maximum F1 was 98 percent for Logistic Regression and 97 percent for RF when we used 150 comments. The Logistic Regression approach has the highest overall accuracy.

KNN

A chart containing varied recall, precision, accuracy, and F1 scores, as well as a comparison of each score for KNN, is shown below. The algorithm's performance in terms of data use rate is shown here. KNN has the greatest accuracy score of 92.79. The graph depicts performance deference in terms of recall, accuracy, F1 Score, and precision.

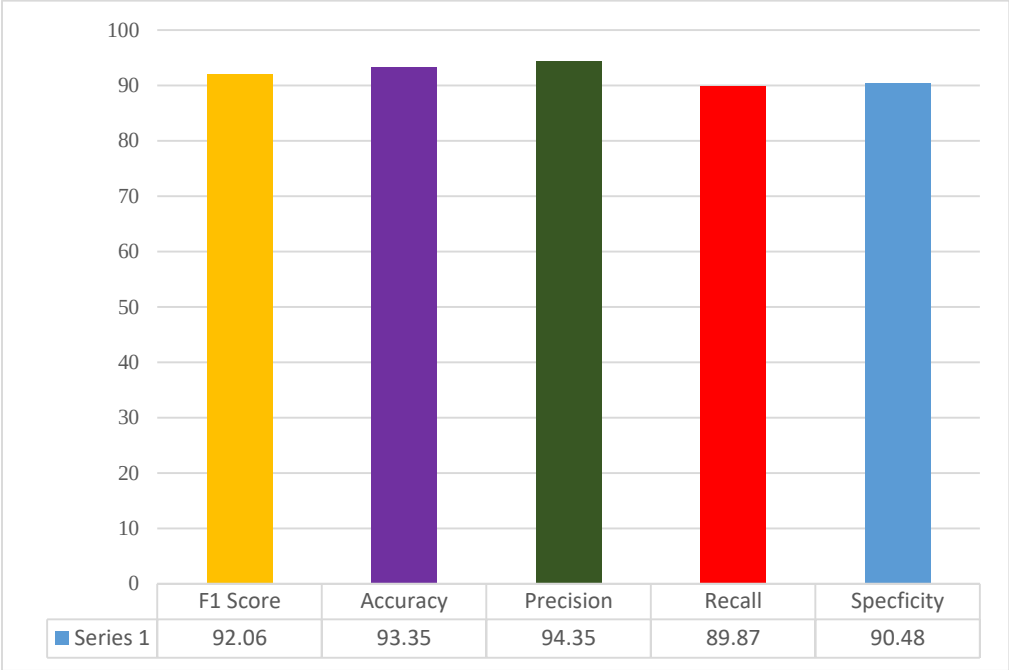


Figure 4.3 KNN's Different Score Comparing Graph

Decision Tree

A chart containing varied recall, precision, accuracy, and F1 scores, as well as a comparison of each score for Decision Tree, is shown below. The algorithm's performance in terms of data use rate is shown here. Decision Tree has the greatest accuracy score of 93.35. The graph depicts the performance differences in terms of recall, accuracy, F1 Score, and precision.

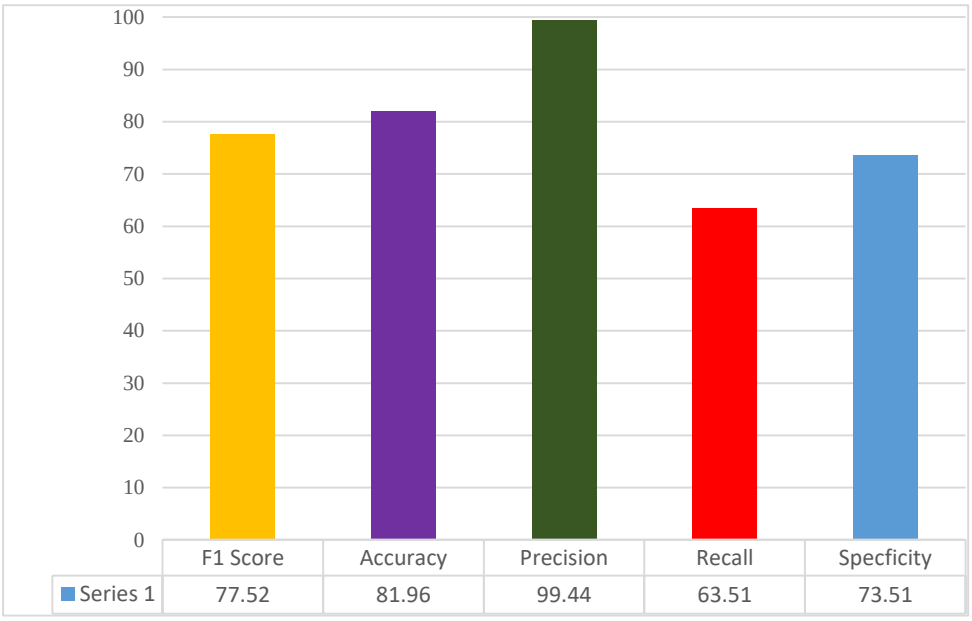


Figure 4.4: Decision Tree 's Different Score Comparing Graph

SVM

This graph shows the various recall, precision, accuracy, and F1 scores, as well as a comparison of each score for SVM. Here we can examine the algorithm's performance in terms of data use rate. SVM has the greatest accuracy score of 96.84. The graph depicts performance deference in terms of recall, accuracy, F1 Score, and precision.

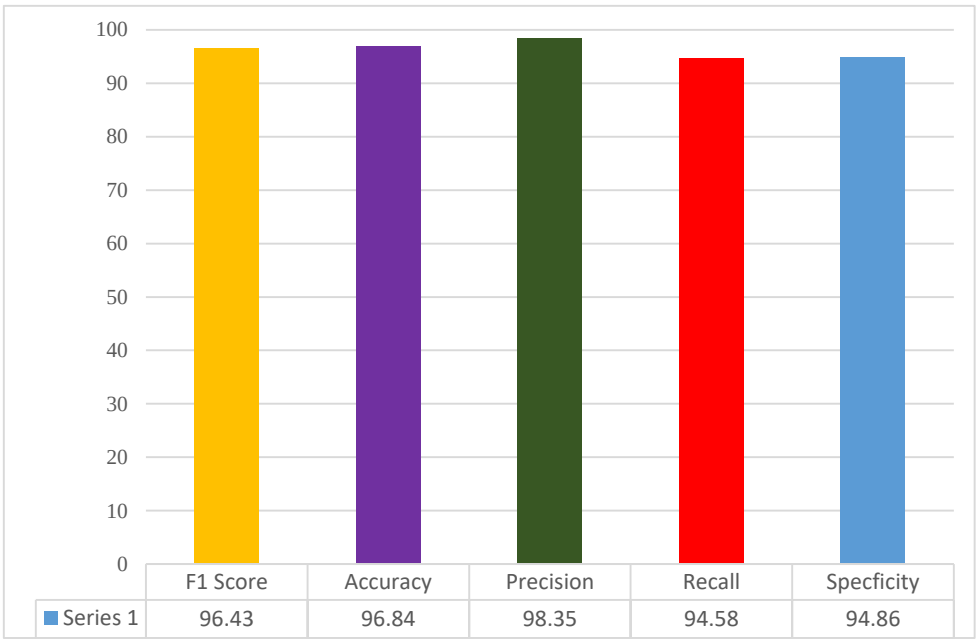


Figure 4.5: Different Score comparison graph of SVM

Random Forest

Random forest is a supervised training method. It may be used for both classification and regression purposes. It's also the algorithm that's the most adjustable and user-friendly. A chart containing varied recall, precision, accuracy, and F1 scores, as well as a comparison of each score for Random Forest, is shown below. The algorithm's performance in terms of data use rate is shown here. Random forest has the greatest accuracy score of 95.64. The graph in Figure 6 depicts performance deference in terms of recall, accuracy, F1 Score, and precision.

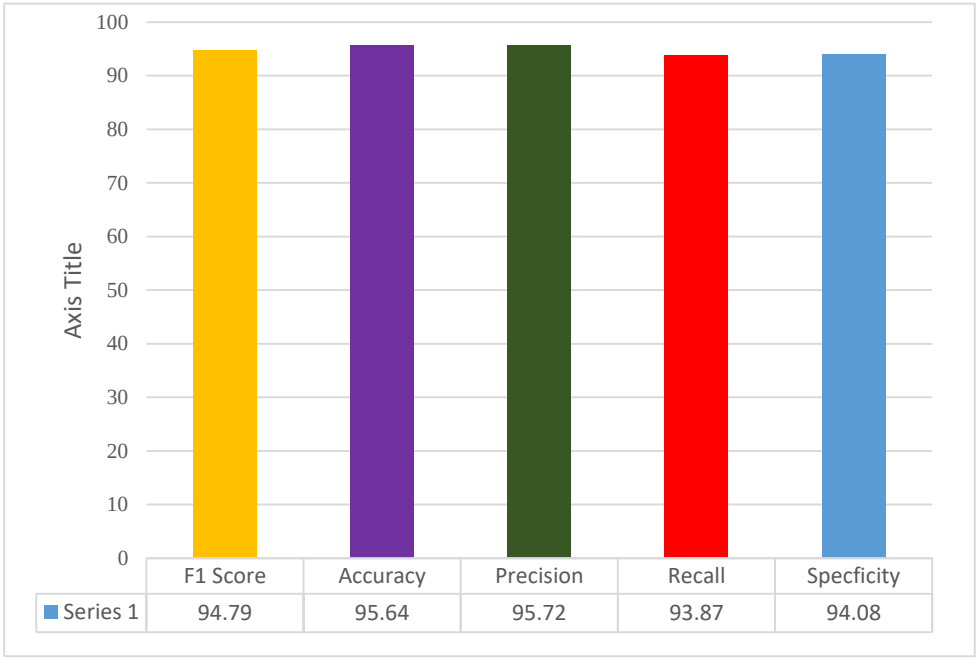


Figure 4.6: Different Score comparison graph of Random Forest

Logistic Regression

Here is a graph that shows the relationship between different scores for Logistic Regression, such as recall, precision, accuracy, and F1 score. Figure 4.5 shows the various results of the Logistic Regression technique. When analyzing the validity of any machine learning approach, precision, recall, and F1 Score are all significant factors to examine. In our research, all of those factors yielded good results. So, in the end, we decided to use Logistic Regression to categorize the postings. Here we can examine the algorithm's performance in terms of data use rate. Logistic Regression has the highest accuracy score of 97.09.

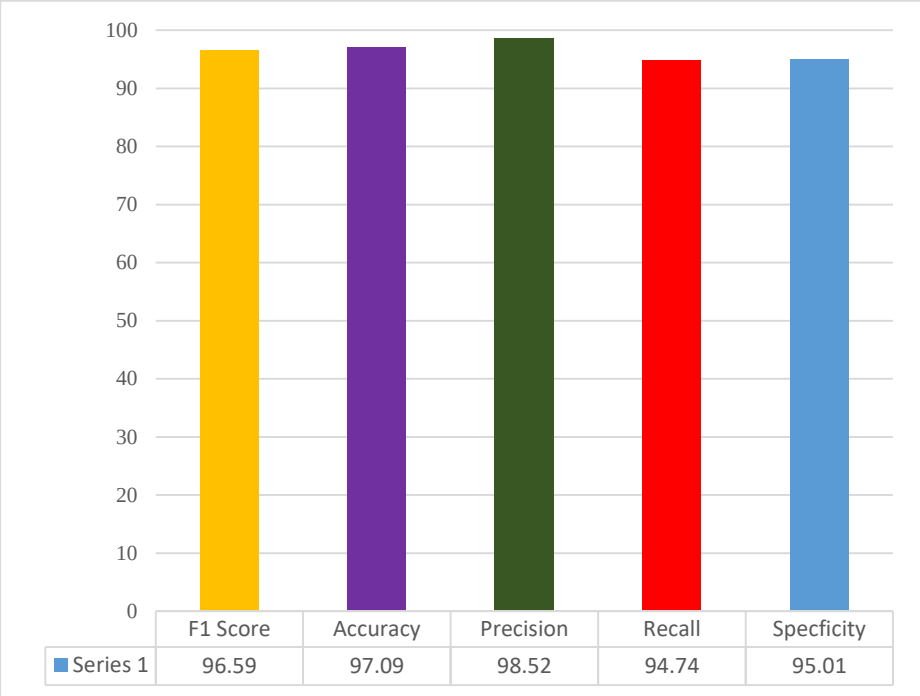


Figure 4.7: Different Score comparison graph of Logistic Regression

CHAPTER 5

SUMMARY, CONCLUSION AND FUTURE WORK

5.1 Summary of the Research

Machine learning has been the subject of a lot of study, however in Bangladesh, the quantity of research done is quite low. Despite the fact that work in predictive styles is a common word for computer education, we obtain some exceptional real-life applications as a result of such study operations. The paucity of similar research on the Bangla language, on the other hand, is a major subject of worry. We believe, however, that many experts from other nations will begin to investigate this area. We categorize our Bangla posts in a few distinct ways in our study.

5.2 Conclusion

Hate speech has been more common on social media in recent years, owing to the anonymity and mobility of such platforms, as well as the shifting political situation in many parts of the world.

Sentiment analysis is an important tool for extracting useful information from people's thoughts. It assists us with detecting human emotions from voice, text, and even facial expressions. This paper gives a sentiment classification technique based on machine learning that can categorize sentiment into positive and negative categories from hate speech reviews. We used a huge dataset of 6000 comments and 2500 hate speech in this research. We implemented the Support Vector Machine (SVM), Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbor (KNN) algorithms to create a model that classifies Bangla comments into normal speech and hate speech. The Logistic Regression technique generated the best results in the research, with a 97.09 percent accuracy rate.

5.3 Recommendations

The following are a few noteworthy suggestions in this regard:

- To increase the accuracy of data collecting in order to get better outcomes from this study.

- Increasing the data size of the dataset might also assist to improve the results.
- It would be preferable to utilize Deep Learning.

5.4 Future Work

The following is future direction on the development of this task:

- To reach this goal, we are presently developing a web-based API to detect hate speech.
- We will develop an intelligence system based on deep learning techniques in the future.
- In the future, a larger dataset might increase the algorithms' accuracy even further.
- In future, adding more categories to this project might help it run more smoothly.

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 access on 07-07-2021.

[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/>

[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 8-08-2021.

[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 08-30-2021

[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

[12] Sigurbergsson, Gudbjartur Ingi, and Leon Derczynski. "Offensive language and hate speech detection for Danish." arXiv preprint arXiv:1908.04531 (2019).

APPENDIX

The first problem we faced while conducting the research was deciding on a methodological technique for our investigation. It wasn't typical job, and little had been done in this region before. As a result, we were unable to receive much help from anyone.

Another roadblock was data processing, which proved to be a major challenge for us. Because there was no accessible source for a Bangla text pre-processing system, we created a corpus for data collection. We also started gathering data by hand. Furthermore, classifying multiple postings is a challenging process. We might be able to achieve it after a long time of hard labor.

PLAGIARISM REPORT

Turnitin Originality Report

Processed on: 31-Dec-2021 01:05 +06
ID: 1736518768
Word Count: 6148
Submitted: 1

BANGLA HATE SPECCH DETECTION BY MACHINE LEARNING By Asfi Hossain Choudury

Similarity Index	Similarity by Source	
15%	Internet Sources:	N/A
	Publications:	N/A
	Student Papers:	15%

2% match (student papers from 04-Apr-2018)
Submitted to Daffodil International University on 2018-04-04

1% match (student papers from 02-Apr-2018)
Submitted to Daffodil International University on 2018-04-02

1% match (student papers from 04-Apr-2018)
Submitted to Daffodil International University on 2018-04-04

1% match (student papers from 02-Apr-2019)
Submitted to Daffodil International University on 2019-04-02

1% match (student papers from 31-Mar-2019)
Submitted to Daffodil International University on 2019-03-31

1% match (student papers from 19-Oct-2021)
Submitted to Daffodil International University on 2021-10-19

1% match (student papers from 25-Jan-2021)
Submitted to Daffodil International University on 2021-01-25

1% match (student papers from 29-Oct-2019)
Submitted to Manchester Metropolitan University on 2019-10-29

1% match (student papers from 25-Oct-2021)
Submitted to University of the Free State on 2021-10-25

1% match (student papers from 23-Dec-2021)
Submitted to American Public University System on 2021-12-23