CYBERBULLYING TEXT CLASSIFICATION USING MACHINE LEARNING APPROACHES

Submitted by
HAFIZ AL ASAD
(181 - 35 – 2484)
Department of Software Engineering

Supervised By
Dr.Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering
Daffodil International University

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Software Engineering.
Fall 2024

APPROVAL

This thesis titled on "**Cyberbullying Text Classification Using Machine Learning Approaches** ", submitted by **Hafiz Al Asad (ID: 181-35-2484)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Fazla Elahe
\-- **Chairman**

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

\-- **Internal Examiner 1**

Md. Khaled Sohel
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

\-- **Internal Examiner 2**

Khalid Been Md Badruzzaman
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

.9.24
\-- **External Examiner**

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

DECLARATION

We hereby declare that this research has been done by us under the supervision of **Dr.Md. Fazla Elahe , Assistant Professor & Associate Head, Department of Computer Software Engineering**, Faculty of Science and Information Technology, Daffodil International University. I also declare that neither this research nor any part of this research has been submitted elsewhere for the award of any degree.

Supervised by:

Fazla Elahe

Dr.Md. Fazla Elahe
Assistant Professor & Associate Head
Department of SWE
Daffodil International University

Hafiz Al Asad

Submitted by:

Hafiz Al Asad

ID: 181-35-2484
Department of SWE
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, I sincerely thank the Almighty for His blessings, which made it possible for me to successfully complete my final year thesis.

I am profoundly grateful to my supervisor, Dr. Md. Fazla Elahe, Senior Assistant Professor and Associate Head of the Department of Software Engineering, Faculty of Science and Information Technology, Daffodil International University, Dhaka. His vast knowledge, genuine interest, and invaluable guidance were crucial to the successful completion of my research titled "Cyberbullying Text Classification Using Machine Learning Approaches." His endless patience, expert advice, consistent encouragement, and constructive feedback throughout every stage of the project have been instrumental in achieving this outcome.

I would also like to express my heartfelt appreciation to my well-wishers, friends, family, and seniors for their unwavering support and motivation. This research is the result of relentless effort combined with the encouragement and inspiration I received from them.

Lastly, with the deepest respect, I acknowledge the constant support, patience, and encouragement of my parents, for which I will always be grateful.

ABSTRACT

Cyberbullying has become a significant issue in today's society, particularly among adolescents and teenagers. The rise of social media platforms and online communication tools has made it easier for individuals to harass others anonymously, often without accountability. In recent years, natural language processing (NLP) techniques have been employed to detect and classify instances of cyberbullying. These methods analyze the language used in online interactions to identify patterns and indicators of bullying behavior.

This study focuses on evaluating the effectiveness of NLP techniques in detecting and categorizing cyberbullying incidents. To achieve this, the research draws on various data sources, such as chat logs, social media posts, and other forms of online communication, to understand the diverse forms of cyberbullying. The ultimate goal is to enhance our understanding of cyberbullying dynamics and explore how NLP applications can help mitigate its adverse effects.

The research employs supervised learning techniques, which use labeled data to train algorithms for accurate predictions and classifications. As technology advances, it has impacted both the positive and negative aspects of life, with machine learning systems becoming increasingly effective in detecting aggressive language associated with cyberbullying.

This study categorizes cyberbullying into seven groups: "Not abusive," "gender," "ethnicity," "political," "insult," "age," and "religion." Among the machine learning classifiers tested, the Support Vector Machine (SVM) achieved the highest accuracy of 91.07% in identifying abusive or cyberbullying-related texts.

TABLE OF CONTENTS. page

FIGURE OF CONTENTS

FIGURE OF CONTENTS
Page

CHAPTER 1

Introduction

1.1 Introduction

Social media networks serve as valuable platforms for forging new connections. However, as the popularity of these platforms has surged, some individuals are using them in unethical and illegal ways. It's particularly alarming to see that teenagers and young adults are developing new tactics for online harassment. According to a Symantec study, over 25% of parents reported being aware of their child's involvement in a cyberbullying incident. In today's digital age, the issue of cyberbullying is becoming increasingly concerning, especially for young people. Cyberbullying refers to the use of digital communication methods—such as social media, text messaging, and email—to harass, threaten, or humiliate someone. The negative effects of cyberbullying can be severe, leading to anxiety, depression, low self-esteem, and, in extreme cases, suicidal thoughts. Unlike traditional bullying, which often occurs face-to-face and can be easier to identify and address, cyberbullying can happen discreetly and remotely, making it more challenging to combat. Cyberbullying can take various forms, including impersonating someone by creating a fake profile, sharing humiliating photos or videos, sending threatening or derogatory messages, and spreading rumors or falsehoods online. The emotional toll on victims can be significant, resulting in feelings of despair, anxiety, or even suicidal ideation. While cyberbullying is most commonly associated with teenagers and young adults, it can occur in any environment, including workplaces. Organizations must take proactive measures to prevent and address cyberbullying, as it can adversely affect employee morale, productivity, and retention. Companies need to implement strategies to combat and manage cyberbullying effectively.

1.2 Objectives

One of the most effective approaches to solving problems is through the application of machine learning. Specifically, supervised learning—a type of machine learning—can filter relevant content into categories such as "Not abusive," "gender," "ethnicity," "political," "insult," "age," and "religion." This process involves training a dataset, which must consist of labeled data, using powerful models to determine whether a given communication is harmful. The primary objective of supervised machine learning is to enable accurate classification and detection of harmful content. In the context of cyberbullying detection, machine learning techniques can be instrumental in identifying and addressing instances of online harassment.

1.3 Motivation

One of the key advantages of using supervised learning for identifying abusive behavior in cyberbullying is its potential for greater effectiveness compared to other methods, such as rule-based approaches or simple heuristics. Supervised algorithms can uncover complex patterns in the collected data, resulting in more accurate predictions. Furthermore, supervised learning models can be continuously improved and adapted over time as new data becomes available. This research employs supervised machine learning techniques, which represent a specific branch of machine learning. By using supervised learning algorithms, datasets can be utilized for both training and forecasting, allowing them to become informative and actionable. The process involves utilizing labeled data to select an appropriate model, followed by adjustments to enhance the model's fit. After evaluating the framework, the goal is to determine the most precise predictions possible. Try to develop a framework capable of identifying abusive text messages in incoming communications. Researchers have approached this problem from various angles to identify the strategy that yields the best results. Online bullying behavior often arises from several factors, predominantly linked to social media platforms. One major benefit of social media is the anonymity it offers users. This anonymity can embolden individuals to express things online that they might hesitate to say in person, as they think they can not caught or held accountable for their actions. Additionally, social media can create a sense of detachment, allowing individuals to dehumanize their targets and overlook the real-world consequences of their actions. Furthermore, platforms like Facebook and Twitter can facilitate the huge spread of misinformation, which can quickly raise into cyber harassment . The algorithms used by media platforms like socialmedia companies may also play a role in this phenomenon. In an effort to boost ad revenue and engagement, these algorithms tried to ensure inflammatory content, leading to the widespread dissemination of offensive or provocative information.

1.4 Rationale of the study

Cyber-bullying is a prevalent issue in today's society, particularly impacting youth and young adults, as the rise of social media and online communication technologies has made it easier for individuals to harass and threaten others. The negative effects can include anxiety and depression, with extreme cases leading to suicide. To effectively combat cyberbullying, it is crucial to understand the language used by bullies; however, manually classifying and analyzing large amounts of text data can be time-consuming and error-prone. This project aims to develop a natural language processing (NLP)-based classifier that can automatically categorize communications related to cyberbullying based on the language used in the messages. By doing so, this method can help assess the severity of cyberbullying incidents and facilitate targeted interventions to protect victims from further harm. The findings from this research could also contribute to the development of more effective guidelines and policies to prevent cyberbullying in online environments.

1.5 Research Questions

This study has been conducted with immense enthusiasm and dedication, despite the challenges encountered along the way. Crafting a plan that is fair, realistic, and accurate presents numerous difficulties. To assist researchers in grasping these concepts, the following questions have been formulated to provide clarity and guidance:

- Have I gathered raw data for my study?
- What ML models were used in this investigation?
- How is data pre-processing carried out?
- Which strategy is more effective in this circumstance?
- Where may someone start searching for a categorization of cyberbullying?
- Which algorithm is more effective at spotting inappropriate text?
- What is the best recommended model's accuracy?

1.6 Expected Outcome

The primary goal of this investigation is to determine whether terms related to bullying are present on social media platforms like Twitter and Instagram. The objective is to accurately classify user-submitted comments as either bullying statements or not. To achieve this, I have categorized the various types of remarks into seven groups: "Not abusive," "gender," "ethnicity," "political," "insult," "age," and "religion." The dataset used for this research is both current and substantial, which has allowed for results with a satisfactory degree of precision. Machine learning techniques have been employed in this study to derive some of the most accurate findings from the datasets. The accuracy of the model hinges on the quality of the training set and the effectiveness of the selected supervised learning methods. Our machine learning algorithms are designed to detect abusive text associated with cyberbullying, and we aim to achieve predictions with 100% accuracy for identifying offensive content. To ensure a comprehensive report, the precision, error rate, recall, and overall performance of various algorithms were thoroughly examined.

1.7 Report Structure

An overview of the project's justification, goals, research questions, and expected outcomes is pr ovided in Chapter One. It provides the study's background, highlighting the importance of comba ting cyberbullying and the demand for efficient detection techniques.This chapter also provides an overview of the report's structure, guiding the reader through the forthcoming sections and detailing how each part contributes to the overall research goals.

Chapter Two provides a comprehensive literature review, summarizing previous research and work conducted in the area of cyberbullying detection and the application of natural language processing (NLP) and machine learning techniques. It identifies existing gaps and limitations within the current body of knowledge, illustrating the scope of this study by highlighting areas that require further exploration. Additionally, this chapter discusses the main challenges and obstacles encountered during the research process, such as data collection issues, the complexity of language used in bullying, and the limitations of previous methodologies. By addressing these challenges, this chapter sets the stage for the need for innovative solutions, emphasizing the relevance of the current study.

Chapter Three delves into the theoretical framework of the study, providing an explanation of the statistical methods and analytical techniques used to assess the data. This chapter offers detailed insights into the machine learning algorithms employed in the research, including supervised learning approaches for classifying text as abusive or non-abusive. It outlines the process of data collection and preparation, emphasizing the importance of quality datasets in training effective models. The chapter also includes a thorough explanation of the confusion matrix analysis, which evaluates the model's performance and accuracy. By addressing the various algorithms techniques and methods used to solve problems , this chapter lays the groundwork for understanding how these methodologies contribute to identifying cyberbullying behavior.

Chapter Four focuses on the experimental findings and performance assessment of the machine learning models used in this study. It presents the results of various tests, highlighting key metrics such as accuracy, precision, recall, and F1 score to evaluate the effectiveness of the models. This chapter also includes illustrative test images and examples to facilitate the reader's understanding of the project's implementation. The discussion section reflects on the results obtained, analyzing the strengths and weaknesses of the machine learning techniques applied. It also explores the implications of the findings for detecting abusive text in cyberbullying contexts and suggests potential improvements for future work.

Chapters Five and Six offer a comprehensive summary of the research, encapsulating the key findings and contributions of the study. These chapters outline recommended next steps for future research and practical applications of the study's results in real-world settings. They provide a conclusive overview of the study, demonstrating how the project aligns with established requirements and expectations in the field. Furthermore, the final sections address the broader impacts of the research on sustainability, society, and the environment, reflecting on the ethical considerations and limitations encountered during the study. The conclusion emphasizes the importance of continued efforts to combat cyberbullying and suggests pathways for future researchers to build upon the work presented in this project.

CHAPTER 2

Background

2.1 Introduction

Text detection for cyberbullying utilizes machine learning techniques, focusing on the development of datasets containing abusive language sourced from the internet and raw data. Our primary objective is to accurately identify abusive text, enabling the implementation of high-accuracy predictive models for improved performance. We propose a new formulation for the ideal model while reviewing previous research to support our approach. Cyberbullying text detection involves analyzing various categories within user-submitted data, and our aim is to establish a robust framework for classifying abusive content to combat unwanted messages effectively. Existing anti-spam strategies often struggle with high rates of false positives and negatives, highlighting the need for an intelligent and efficient cyberbullying text classification system. By leveraging current technologies, we can train our machine learning model to enhance its effectiveness and reliability in detecting cyberbullying. This comprehensive approach addresses the critical need for better detection methods and offers valuable insights for developing effective solutions against cyberbullying.

Fig 2.1: All Concept of Cyber Bullying.

The deliverable's word cloud depicting the entire cyberbullying notion is shown in Figure 2.1.

Many people today rely on the internet as their primary source of information about the world, and this trend is expected to continue. As mentioned, we will develop a cyberbullying text recognition algorithm designed to assess the accuracy of news and messages. Our model will provide users with a chart-marked map displaying the latest data on the primary keywords or sources disseminating the most inaccurate information. To address the widespread concern over misinformation, we will also include practical tips for preventing the spread of abusive text material online. This approach aims to empower users with the tools they need to identify and combat misinformation effectively, fostering a safer and more informed online environment.

2.2 Related Works

In [4], researchers systems to assess whether statements are hateful.. Their proposed model surpassed all individually employed machine learning techniques and ensemble methods, achieving a peak accuracy of 96% with a dataset extracted from Twitter.

In [5], the focus is on demonstrating how the program detects abusive posts, tweets, and other forms of online activity. Machine learning is advocated as a tool for identifying and mitigating bullying on Twitter. The true positive recognition rates for SVM and Naïve Bayes were reported at 71.25% and 52.70%, respectively. By utilizing the Twitter API, tweets are gathered, and a model assesses whether the content qualifies as bullying.

In [6], researchers constructed a global dataset consisting of 37,373 unique tweets for their analysis. They employed seven machine learning classifiers, including Logistic Regression (LR), which demonstrated superior performance with a median accuracy of around 90.57%. Among the classifiers, Stochastic Gradient Descent (SGD) achieved the highest accuracy of 0.968, SVM recorded the best recall at 1.00, and LR attained the best F1 score of 0.928.

The paper in [7] offers recommendations on applying the research findings to mitigate cyberbullying. A collection of 5,453 tweets labeled with the #Gamergate hashtag was meticulously annotated by human experts. The baseline algorithm utilized various Twitter-based features, including text, user, and network information. Cyberbullying detection significantly improved to a success rate of 91.88% when considering user dispositions and attitudes, alongside a weighted AUC of 0.97.

In [8], the goal is to develop an efficient approach for identifying online harassment and abusive comments by combining machine learning techniques with natural language processing. The research evaluates four distinct machine learning algorithms using two specific features: the Bag of Words (BoW) model and Term Frequency-Inverse Document Frequency (TFIDF).

In [9], the study investigates the potential for social network analysis features to enhance cyberbullying detection by examining the structure of social networks. A balanced training dataset was constructed using the 'SMOTE' technique, and the evaluation was conducted in realistically imbalanced scenarios, with 70% of all messages designated for the training set and the remaining 30% for testing.

According to a recent study from the Massachusetts Institute of Technology, efforts are underway to detect cyberbullying in YouTube video comments by analyzing textual context. The first step in the classification process involves identifying if the comment pertains to sensitive subjects such as sexuality, race/culture, IQ, or physical attributes, followed by theme selection. The overall success rate for identifying instances of online harassment in YouTube comments was 66.7%, with a support vector machine learner employed in this project [10].

Despite the limited number of research groups dedicated to cyberbullying detection, [11] notes that the CAW 2.0 organizers initiated a misbehavior detection project, with only a single contribution submitted. Yin and colleagues found that incorporating emotional and contextual information significantly improved the performance of the conventional text mining system using a bag-of-words approach; however, the support vector machine model achieved only 61.9% recall, even with the enhanced model.

In [12], researchers collected and prepared data from over 35,000 tweets on Twitter, which were then inputted into various intelligent machine learning algorithms. Five significant machine learning algorithms were applied to classify and predict the tweets into two primary categories: "offensive" and "non-offensive," with multiple performance metrics serving as the basis for the evaluation of the machine learning models.

2.3 Research summary

My first priority of the project is to explore various techniques available to the broader community. To achieve this, we utilized five different approaches and incorporated additional algorithms into our dataset. The dataset primarily consists of real-world information sourced from Instagram, which is publicly accessible online. As previously mentioned, our collection includes both newly acquired and previously used data, enabling us to assess the effectiveness of the five methods employed and examine the influence of the additional data obtained from the same platform. The presence of similar class structures and labeling types further facilitates this analysis. The machine learning techniques and text classification algorithms used for feature extraction and data preparation were primarily implemented in Python. For feature extraction, Python was my tool of choice, and I employed the SVM classifier methodology to categorize the data based on the predictor.

2.4 Scope of the problem

Essentially, our research involves evaluating the available data and constructing a model through the application of machine learning methods to facilitate the diagnosis of cyberbullying texts. This endeavor aims to make a significant impact on individuals within the community, as many people today encounter abusive comments without realizing the extent of their torment. They often become consumed by self-doubt and uncertainty about how to respond, leading to misguided initial assessments of their situations. To effectively identify the issue, there must be a way to discern whether the remarks they receive are indeed abusive. By utilizing natural language processing (NLP) tools, the project will gather data from various social media platforms and analyze it to identify trends and characteristics associated with cyberbullying behavior. Subsequently, this data will be used to train the system to accurately detect instances of cyberbullying. In this study, we examined tweets to uncover terms that could indicate potential cyberbullying.

2.5 Challenges

Considering the challenges involved in analyzing a single massive data file, the primary issue in this research lies in the integration and evaluation of multiple datasets. We employed various tools but the sheer size and complexity of the data, spanning different time periods and levels, required significant time and effort to achieve the desired results. The field also encompasses additional datasets, and since the research remains unfinished, I must put in considerable effort on related projects, which complicates my ability to quickly formulate the best responses. Cyberbullying is an increasingly pressing concern in today's digital age, capable of inflicting serious harm on those targeted. As a result, there has been a growing interest in developing (NLP) methods to automatically classify instances of cyberbullying, which could aid in detection and prevention. However, this project faces numerous challenges. A major obstacle is the lack of a universally accepted definition of cyberbullying. The varying definitions across different scholars and organizations complicate the development of a consistent approach. Additionally, the language used by cyberbullies can be nuanced and indirect, making it difficult for an NLP model to detect abusive comments. The challenge of classification intensifies when irony, sarcasm, and other figurative language are employed, further complicating the identification process.

- Gathering data,
- Importing data,
- Preprocessing and labeling data,
- Achieving accuracy levels above 90%,
- Making decisions based on test findings in machine learning models.

CHAPTER 3

Research Methodology

3.1 Introduction

This chapter outlines the methodology of the study, detailing the processes of dataset collection, conducting experiments, and utilizing models to enhance accuracy. It also introduces the proposed data workflow and modeling approach, aiming to improve clarity and simplify the presentation.

The objective of this section is to provide an overview of the techniques employed to identify cyberbullying text within data. The full methodology of the study is detailed here, covering various analysis methods and the selection of machine learning (ML) approaches. Since the study utilizes five distinct ML models, the process begins with building a dataset as a foundation for structuring and fitting algorithms. The data is then divided into training and testing sets, with feature selection applied at this stage. Training involves fitting data into multiple ML models, and subsequent evaluation focuses on model accuracy.

A process flow chart provides a high-level overview, complemented by equations and diagrams for further understanding. The methodology includes data collection, analysis, and the proposed model, with detailed explanations supported by equations, graphs, and visual representations. The study employs real-world data, including examples from platforms like Instagram and Twitter, and utilizes highly accurate.

3.2 Equipment Subject of the research

The research focuses on identifying and analyzing concepts for implementing, managing, and training models. Tools and techniques used include the Python programming language, Windows OS, and libraries such as NumPy, Sklearn, OpenCV, and others. The training and testing phases were conducted on the Google Colab platform, which allows Python programmers to write and execute ML and data science algorithms. These algorithms classify text into categories like “Not abusive,” “gender,” “ethnicity,” “political,” “insult,” “age,” and “religion.”

Libraries Used:

- **Matplotlib:** Provides tools for graphing and plotting data visualization.
- **NumPy:** Facilitates vector processing, matrix computations, and other numerical operations.
- **Sklearn:** A powerful tool for data analysis and prediction, integrating with libraries like NumPy and Matplotlib.
- **Seaborn:** Enhances data visualization with advanced plotting capabilities.
- **H5py:** Allows handling large datasets, particularly integers, using the HDF5 format.
- **TensorFlow:** A versatile AI framework for building and deploying ML models, especially neural networks.

- **Pandas:** A toolkit for systematic data analysis and manipulation, aiding in summary data management.
- **OS Module:** Enables interaction with operating system functionalities in Python.

3.3 Workflow
The workflow comprises various steps to achieve the desired outcomes:

Step 1: Data Collection
Data was gathered from online sources and real-world examples. Due to the difficulty of collecting data specific to abusive text detection and classification, the dataset was limited.

Step 2: Data Processing
Each data component was individually evaluated. Processing involved cleaning inappropriate and irrelevant text and preparing the dataset for use.

Step 3: Dataset Preparation
Data preparation included organizing the content, removing stop words, and structuring it for training purposes. Minimal preprocessing was conducted before data splitting.

Step 4: Model Selection
To enhance reliability, a prediction model was selected, trained, and tested. Although multiple models were employed to improve efficiency and facilitate classification.
Step 5: Performance Evaluation
The final step involved evaluating model performance using metrics like F1 score, accuracy, and confusion matrices. These assessments determined whether the machine learning algorithms effectively identified cyberbullying through abusive language.

Step-6: Final Thoughts and Upcoming Tasks: Provide development roadmap.

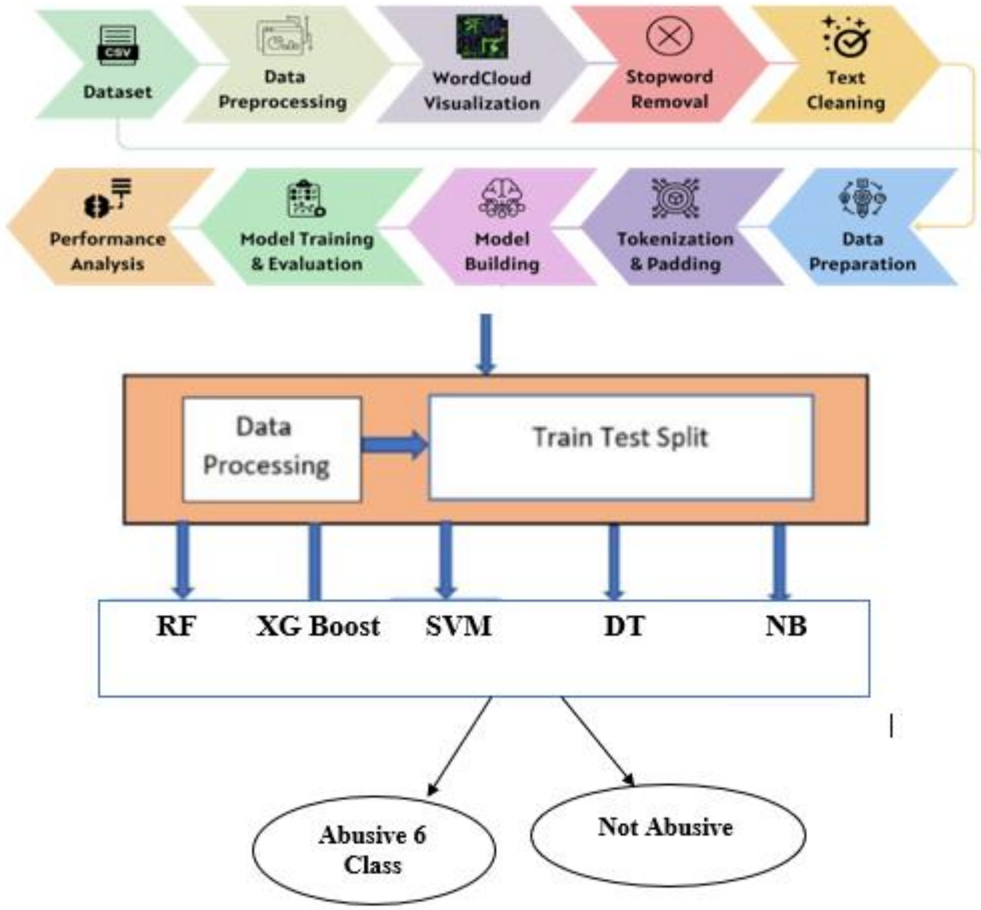


Fig 3.1: Model Workflow for the Entire Research Proposed.

The steps of our research methodology, designed to detect abusive cyberbullying texts, are outlined in Fig. 3.1. The data for our study was collected from various online sources, including publicly available Instagram accounts and tweets, ensuring that it reflects real-world scenarios. To guarantee the accuracy and relevance of the data, we thoroughly reviewed the datasets, removed unnecessary phrases, and refined the language. Word clouds were utilized as visual tools to highlight commonly used phrases, making them easier to interpret.

Our study explores machine learning techniques by building and enhancing models with the preprocessed data. To address the issue of class imbalance, we also applied permutation techniques to ensure fair representation and improve the model's overall performance. Additionally, our approach aims to reduce the occurrence of false positives and false negatives. By combining language analysis with machine learning, we present a comprehensive strategy for identifying and mitigating deception in abusive texts associated with cyberbullying.

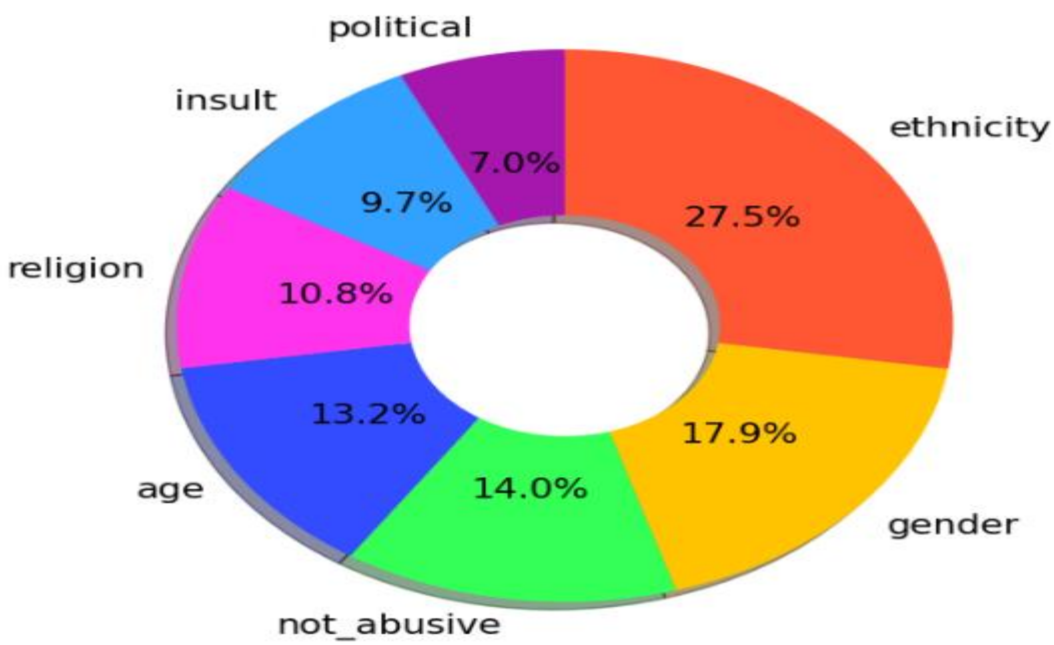
3.4 Data Collection

The final dataset used for this study was created by combining data from twitter called X and instagram post .A generated dataset containing 80873 total data points has two columns named "full_text" and "label." There two kind of data : test and train . Test has 16175 after duplicate data was eliminated 7 types, such as "Not abusive," "gender," "ethnicity," "political," "insult," "age," and "religion." Train has 64698 and categorizing data.

Table 3.1, data field is under the point

Table 3.1: . Column Description.

Name	Description of the Column
full_text	Description of cyberbullying text
label	"not_abusive," "gender," "ethnicity," "political," "insult," "age," and "religion."



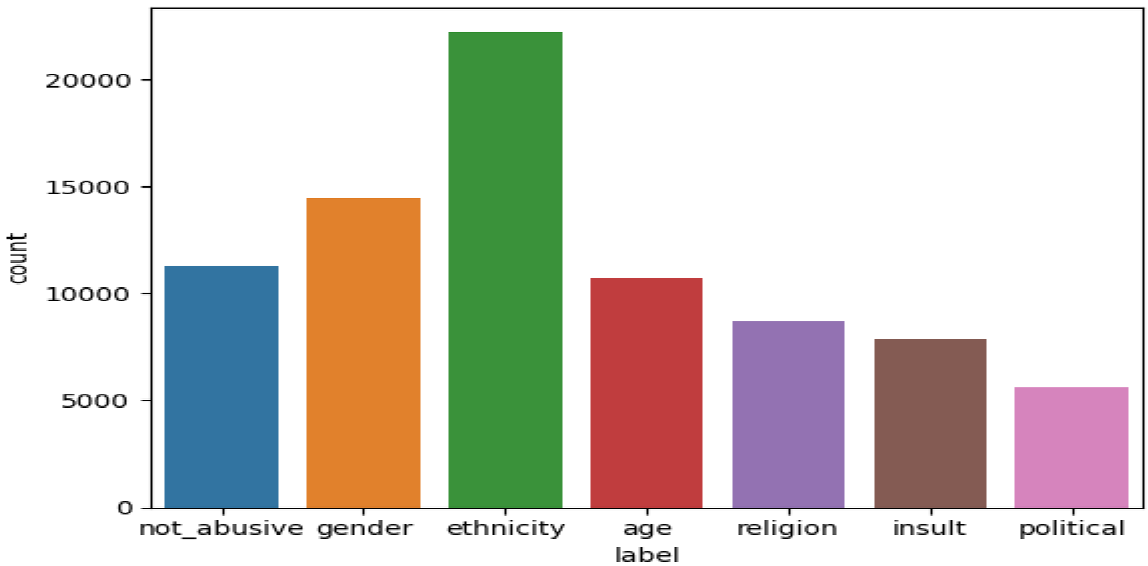


Fig 3.3: All 7 Class's Distribution

Label Encoding

Machine learning experts frequently work with datasets containing numerous features, regardless of their size. These features often include unique identifiers that may consist of both letters and numbers. In training, words are often used to classify data, enhancing user comprehension and convenience.

Label encoding is a technique for converting identifiers into a machine-readable numeric format. This step is essential for building a structure that represents numerical values, enabling effective processing. The decision on how to utilize these labels rests with the programmers developing the machine learning models. Before encoding, it is crucial to perform an initial analysis of the data, considering any changes or variations being monitored.

3.5 Statistical Analysis

3.5.1 Data Analysis

The goal of the data pre-processing phase is to transform raw textual input into a structure suitable for evaluation and model training. This begins with extracting the essential "full_text" and "label" features from datasets focused on detecting abusive text related to cyberbullying. Various datasets to create a single database for test and training.

Once the data collection is prepared, errors are corrected by removing unnecessary punctuation and irrelevant characters. These steps aim to improve word choice relevance and overall data quality. Word cloud visualization is used to display common word repetitions graphically, providing insights into language structure. Additionally, tokenization is applied, converting words into numeric vectors, making them easier for machine learning models to process. This comprehensive pre-processing approach enhances the model's ability to distinguish between harmful and non-harmful data and supports deeper analysis.

3.5.2 Word Cloud Visualization

Word cloud visualization is a key tool for analyzing patterns and trends in language across training and testing datasets. This includes separate visualizations for "not_abusive" and other classification categories, as depicted in Figures 3.4–3.7. By illustrating the most prominent terms in the text, this method helps identify implicit and distinctive linguistic features.

Word clouds provide essential insights into the evolution of the English language within the dataset, serving as a guide for categorizing data into classes. This visual representation bridges the gap between raw data and meaningful analysis, enabling well-informed decisions during the development of systems to reduce false positives.

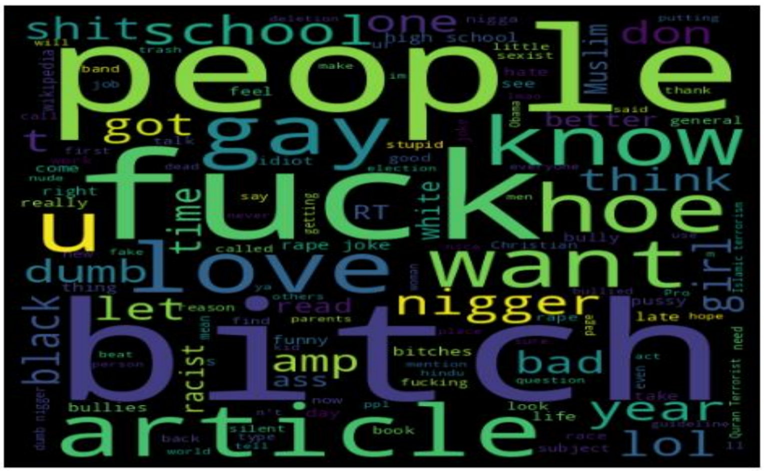


Fig 3.4: Word visualization of Train data.

Fig 3.5: Word visualization of Text Data.

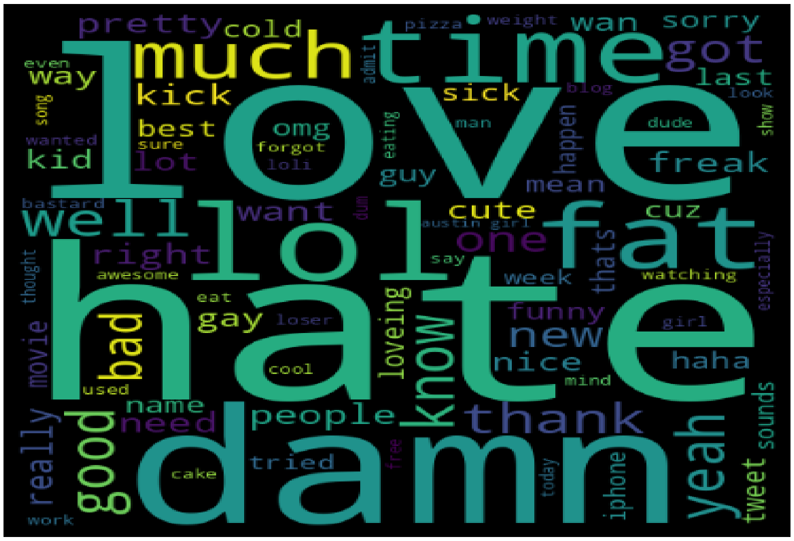


Fig 3.6: “not_abusive” class Word visualization

Fig 3.7: “gender” class Word visualization

3.5.3 Stop Word Removal

While some words enhance content, many commonly used terms—known as stop words—add little value to text classification or the overall language context. To address this, a custom module called “Stop words,” sourced from GitHub, is employed. This library includes a meticulously curated collection of English stop words and numbers, aligned with the syntactic structure of the language.

By incorporating these libraries, data processing is streamlined, reducing irrelevant characters. Punctuation is also removed to enhance text consistency and improve the accuracy of text classification models. These measures ensure cleaner, more efficient datasets, as demonstrated in prior examples.

.

```
# Define a function to remove specified words (stop words) from a sentence
def remove_stop_words_and_single_alphabets(sentence):
    words = sentence.split()
    stop_words = ["it", "ve", "!","re", "me", "the", "oh", "we", "you","so", "he", "is", "." ,"she", "lik
    "world", "back", "this", "if", "let", "mkr", "been", "thing", "should", "anything",
    "did", "its", "day", "still", "first", "too", "cant", "And", "had", "going", "make",
    "these", "only", "see", "has", "go", "why", "were", "there", "will", "because", "how",
    "the", "then", "an", "he", "if", "its", "Im", "no", "by", "at", "what", "u", "do",
    "amp", "i", "or", "so", "have", "be", "my", "who", "was", "are", "I", "to", "a", "the"]
    words = [word for word in words if len(word) > 1 and word.lower() not in stop_words]  # Remove single
    return ' '.join(words)
```

Fig 3.8: English Stop Words.

3.5.4 Clean the Text

This is a step in preparing the data collection . This process involves two main strategies to enhance the quality and relevance of the textual content. First, the text is filtered to retain only essential articles, with a predefined word limit of 100 words. This filtering ensures the content remains high and qualityful, legally compliant, and informative. Text correction process is implemented to systematically remove unnecessary elements such as tabs, stop words, emojis, special characters, and markers. Standard symbols, line breaks, and select English characters are also eliminated. A thorough preservation method is applied to ensure the remaining content is well-prepared for detailed analysis.

	full_text	label
0	yeah got backups hate happen strugglin week ha...	not_abusive
1	hate using love iphone tried new provided get ...	not_abusive
2	wow lol sounds lot piss hehehe	not_abusive
3	damn thang typical rap beef one person worryin...	not_abusive
4	well damn needed mother time	not_abusive
...		

Fig 3.9: After cleaned text.

3.5.5 Tokenization and Padding

Tokenization and padding are essential steps in modern data processing. Tokenization involves converting words into sequences of numerical values, enabling the model to interpret and process language. Each word is assigned a unique number, establishing a link between written language and symbolic representation. Padding ensures all sequences are of uniform length during training, increasing consistency and improving pattern recognition. By transforming sentences into numerical forms and managing variations in sequence lengths, our algorithms can effectively analyze the linguistic complexities of the text. This step is critical for enabling the system to accurately interpret the data and differentiate between abusive content or the six defined classes.

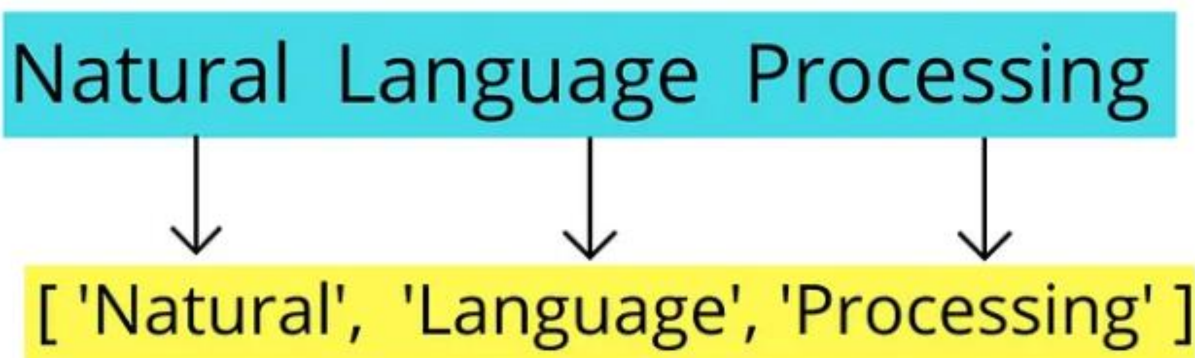


Fig 3.10: Tokenization .

3.5.6 Data Preparation

Even after removing duplicate entries, the data was not randomly split for model training and testing during the data preparation phase. The cyberbullying text dataset comprises a total of

80,873 data
points.from which we extract the most important "full_text" and "label" portions. Train and Test are the two sections into which the dataset is separated. The data was divided into the following c ategories after duplicates were eliminated: "not_abusive," "gender," "ethnicity," "political," "insu lt," "age," and "religion." Test has 1,6175 data, whereas Training contains 6,4698 data that are cl assified.

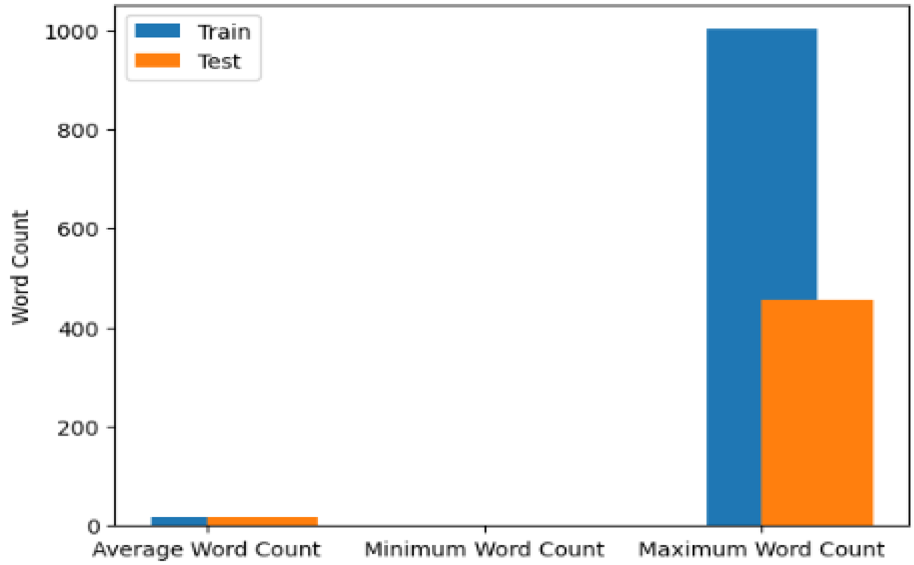


Fig 3.11: Comparison of Train and Test Data Word Count.

3.5.5 Models

Our analysis employs various credibility of cyberbullying texts. The models used include XGBoost, Naive Bayes, Random Forest Classifier, SVM, and Decision Tree. By leveraging the diverse capabilities of these models to represent patterns, context, and linguistic variations, we aim to develop a reliable method for detecting abusive content in cyberbullying.

1. **Decision Tree**:

The decision tree is a supervised machine learning technique applicable to both classification and regression problems. It is a tree-structured classifier where branching represents classification rules, internal nodes represent dataset attributes, and leaf nodes represent outcomes. A decision tree consists of two types of nodes:

- **Decision Node**: Contains multiple branches and represents decisions or tests based on dataset attributes.

- **Leaf Node**: Contains one or two branches and represents the outcome of decisions.

This model operates by testing attributes of the given dataset and making decisions based on the results. Decision trees are intuitive and easy to understand, as they mimic human decision-making processes and provide a clear, tree-like structure for reasoning.

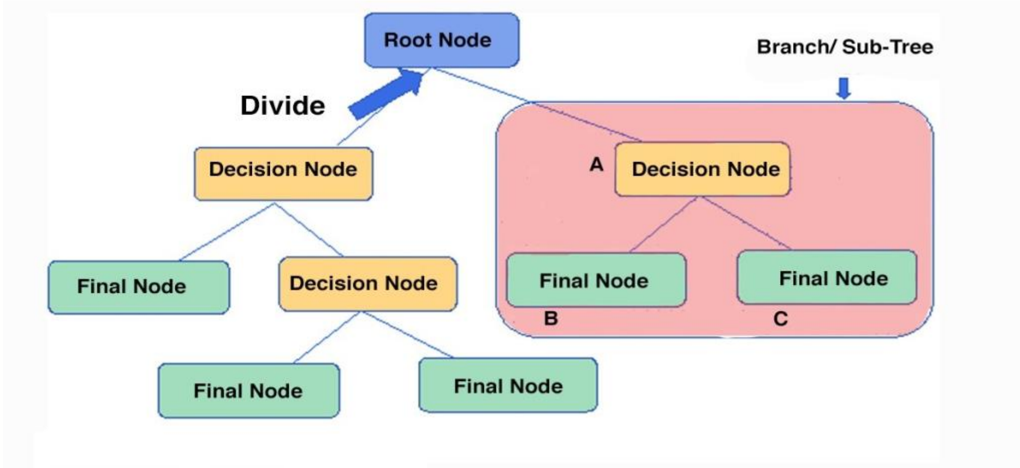


Fig 3.12: Working process of Decision Tree classifier.

1. Random Forest

The Random Forest (RF) Classifier is a tree-based on Machine Learning method that can be used for both regression and classification tasks. It constructs a collection of hierarchical decision trees using a technique. For classification, the model generates numerous decision trees and averages their outcomes, addressing issues like overfitting. This algorithm is widely applicable and effective when working with large datasets, making it highly popular in machine learning applications. Random Forest is particularly favored in business settings due to its versatility and advantages over other techniques. Initially introduced in years of Nineteen ninety seven, it was specifically made for handle extensive data collection efficiently.

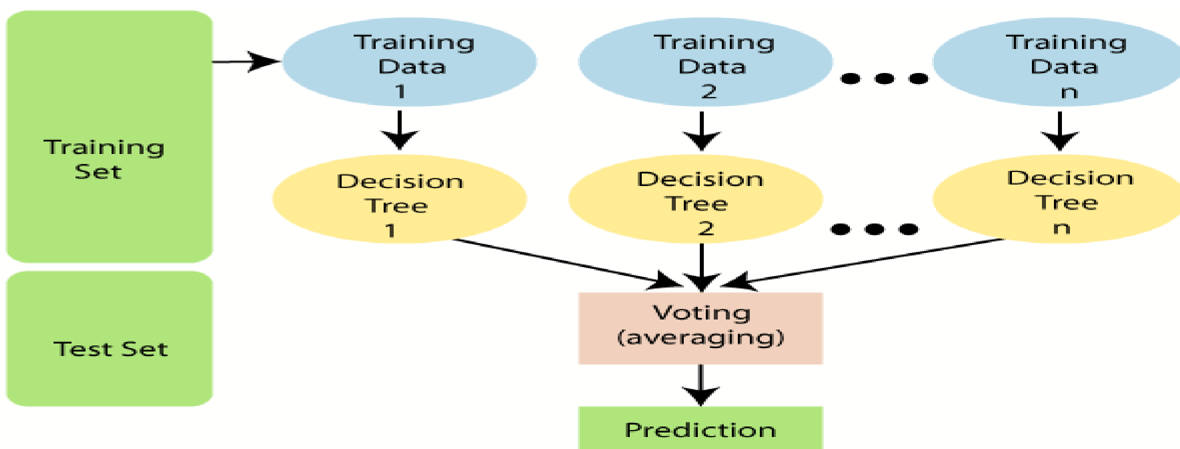


Fig 3.13: Working process of Random Forest Classifier.

1. **SVM:** SVM, or Support Vector Classifier, is the abbreviation for it. Since SVC does not make any assumptions about the underlying structure of the data, such as the number of groups and their relative sizes, it is by definition a probabilistic clustering technique.

Because of this, if your data is highly dimensional, you will usually need to do some preliminary processing, such principal component analysis. It has been demonstrated by earlier studies to function effectively with low-dimensional data. The original approach has been modified in many ways that provide specific ways to compute the clusters by looking at a subset of the regions included in the the relationship matrices. There are several updates available.

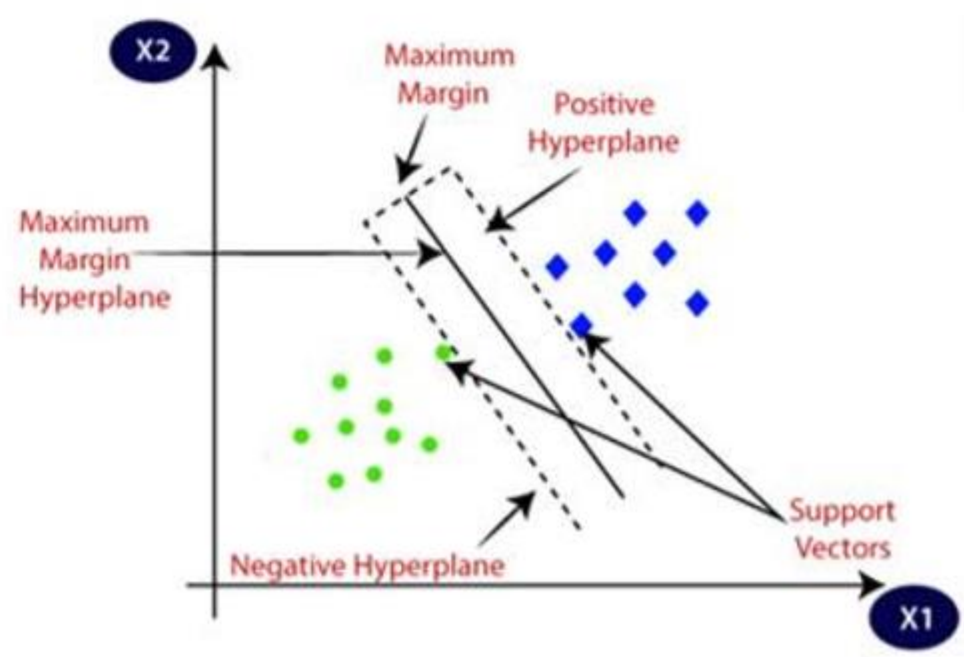


Fig 3.14: SVM.

1. Naive Bayes

This is highly recommended for handling large datasets, even those with millions of records, due to its ability to process vast amounts of data efficiently. It excels in natural language processing (NLP) tasks, such as abusive text analysis, and is known for its simplicity and speed. To understand the Naive Bayes classifier, it's important to first grasp Bayes' theorem, which is based on conditional probability. Conditional probability refers to the likelihood of an event occurring given the occurrence of another event. This theorem allows us to calculate the probability of an event using prior knowledge and conditional chances. Naïve Bayes algorithms are widely used in areas like recommendation systems, sentiment analysis, and spam filtering. However, their main limitation is the assumption that predictors are independent, which can negatively affect performance in real-world applications where prediction parameters are often interdependent.

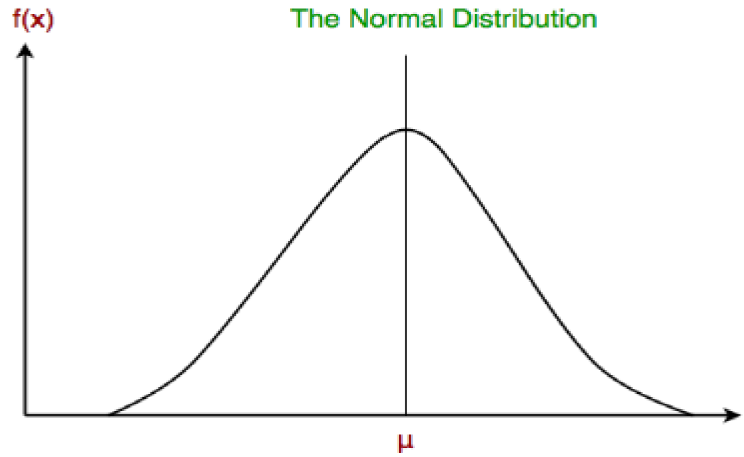


Fig 3.15: How Naive Bayes Functions works.

1. XGBoost

XGBoost, or Extreme Gradient Boosting, is a highly effective and widely used machine learning technique for solving both regression and classification problems. This method enables the rapid and efficient implementation of gradient-augmented decision trees. As an ensemble learning approach, XGBoost enhances model performance by combining the predictions of multiple weak learners, typically decision trees. Each weak learner corrects the errors of the previous one in a step-by-step manner, progressively improving the model. To prevent overfitting, XGBoost incorporates several regularization techniques with penalty terms, ensuring efficient training and optimal generalization. The method is capable of detecting and learning from non-linear data patterns and is fully functional and ready to use without the need for further setup.

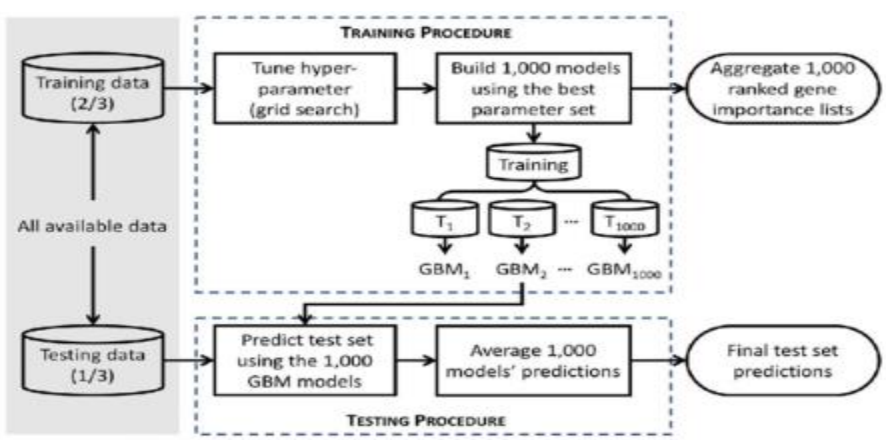


Fig 3.16: Structure of XGBoost model.

3.6 Implementation

After completing the previous steps, it was essential to collect the data to ensure accuracy. For our project, ten tasks were necessary to complete the basic setup. Each of these tasks had to be accomplished to meet our goal:

- Data Collection
- Data Preprocessing
- Word Cloud Visualization
- Removal of Stop Words
- Text Cleaning
- Data Preparation
- Padding and Tokenization
- Model Creation for All Five Algorithms
- Building a Text Classification Prototype
- Discussing Results and Accuracy

To bring the concept to life, we started by developing the code. We tested the accuracy of five different methods and analyzed the performance of the algorithms after completion.

Based on our evaluation, we concluded that this model would be the most suitable for our needs, particularly regarding the analysis of social media posts. After thoroughly examining the relevant theoretical and numerical approaches, we defined a set of requirements necessary for any attempt to classify abusive cyberbullying text. The essential results are as follows:

1. **Hardware and Software Requirements**
   - Operating System: Windows 7 or higher
   - Hard Disk: Minimum 1 TB
   - RAM: Minimum 4 GB
2. **Development Tools**
   - Python Environment
   - PyCharm
   - Google Colab
   - Visual Studio Code

CHAPTER 4

Experiment Results and Discussion

4.1 Introduction

This section outlines the process of classifying abusive text associated with cyberbullying. The model construction involved several key steps, including model selection, data collection and analysis, generating word clouds, removing stop words, cleaning the text, and evaluating performance based on abusive text detection outcomes. The results of the experiment are presented in the subsequent section.

4.2 Experimental Results

It is widely understood that no technology can deliver flawless results. Similarly, the parameters of our model can be fine-tuned during training to enhance accuracy. Nevertheless, by employing various techniques, we achieved relatively high accuracy. This section provides a visual summary of our research findings, including heatmaps, recall, precision, F1 scores, and support metrics. These visuals utilize our data to identify the categories most frequently associated with texts from the seven cyberbullying classifications.

4.3 Best ML Classifiers

The outcomes of our research varied based on the strategies employed. Using five machine learning algorithms, we successfully predicted the abusive text labels linked to cyberbullying with notable accuracy. These methods were applied to evaluate the effectiveness of each component within the framework, and multiple validation processes were conducted to finalize the predictions. After curating the complete dataset, each model utilized a single dataset comprising both original research data and publicly available online sources. MATLAB and its pre-configured libraries were used to assess algorithm performance after data preparation was completed.The second phase involved using a similar dataset To determine whether the content could be categorized as cyberbullying, we analyzed the data using various machine learning models and classification techniques.The categories included “Not abusive,” “Religion,” “Gender,” “Age,” “Insult,” “Political,” and “Ethnicity.” A detailed analysis of various models was conducted using key performance metrics such as overall F1 score, accuracy, precision, and recall, providing comprehensive insights into the algorithms’ effectiveness.

Table 4.1: Table OF Accuracy.

Classifier	**Accuracy Score (AUC)**
Decision Tree	83.04%
Random Forest	87.79%
SVM (classifier)	91.07%
Naïve Bayes	79.61%
XGBoost	87.48%

This section showcases the performance of the classifiers used in the research. Two free tools, PyCharm and Google Colab, were utilized throughout the process. Five classifiers were utilized in the implementation: Random Forest, Decision Tree, XGBoost, Support Vector Classifier (SVM), and Naïve Bayes.

```
              precision    recall  f1-score   support

         age       0.92      0.88      0.90      2121
   ethnicity       0.91      0.92      0.91      4441
      gender       0.89      0.85      0.87      2947
      insult       0.90      0.90      0.90      1605
 not_abusive       0.89      0.96      0.93      2211
   political       0.96      0.97      0.96      1150
    religion       0.96      0.94      0.95      1700

    accuracy                           0.91     16175
   macro avg       0.92      0.92      0.92     16175
weighted avg       0.91      0.91      0.91     16175
```

Fig 4.1: SVC classifier classification reports.

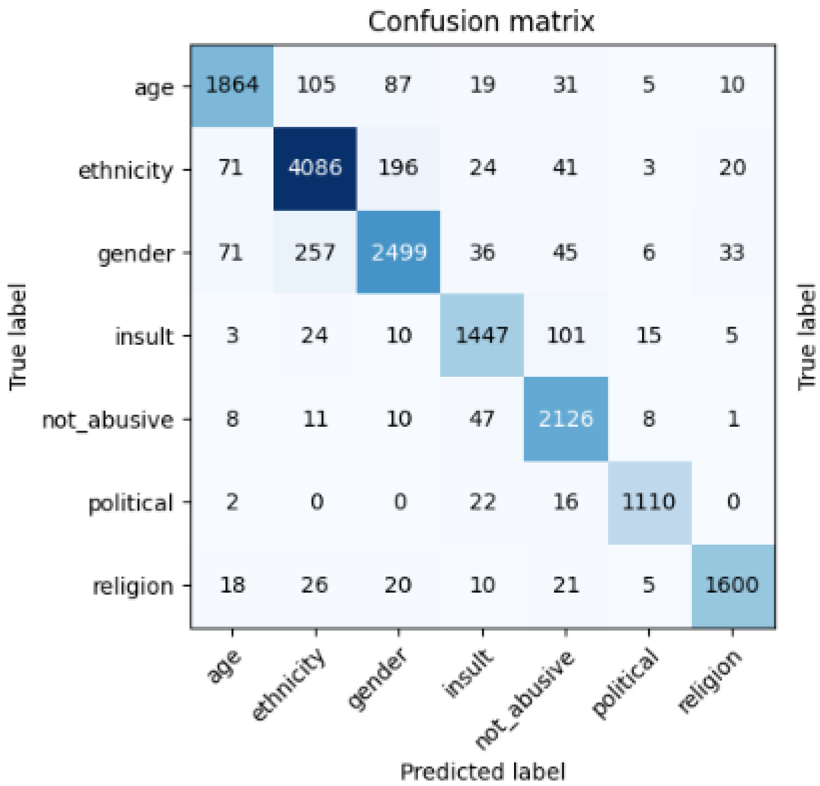


Fig 4.2: SVC classifier Confusion matrix.

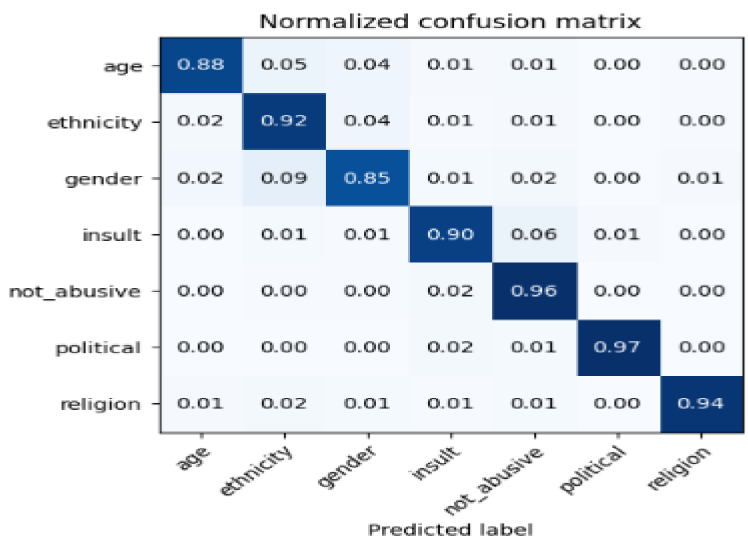


Fig 4.3: SVC classifier

.

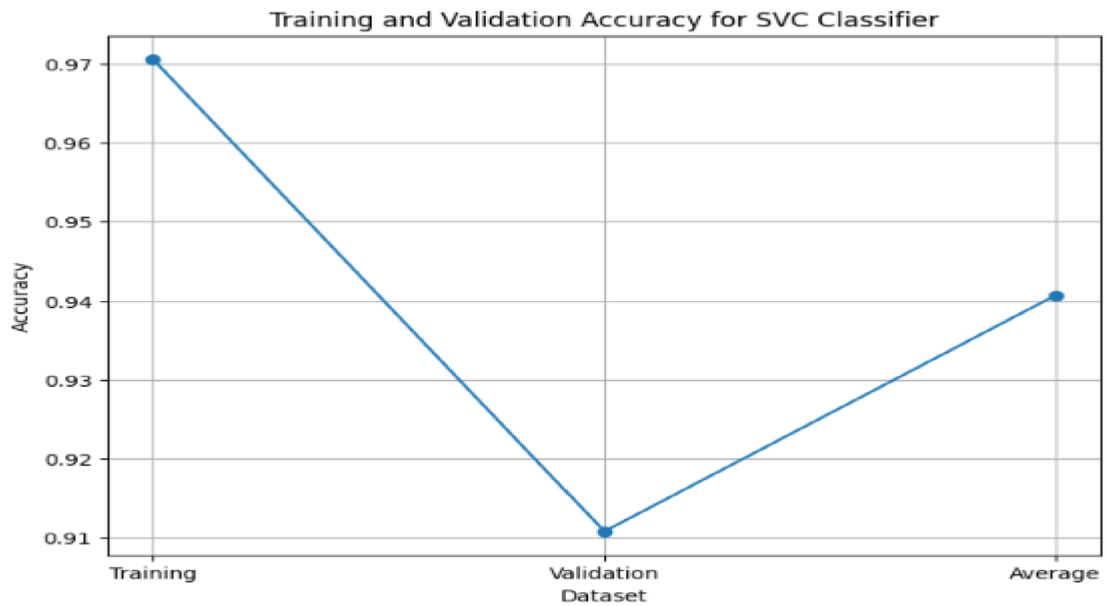
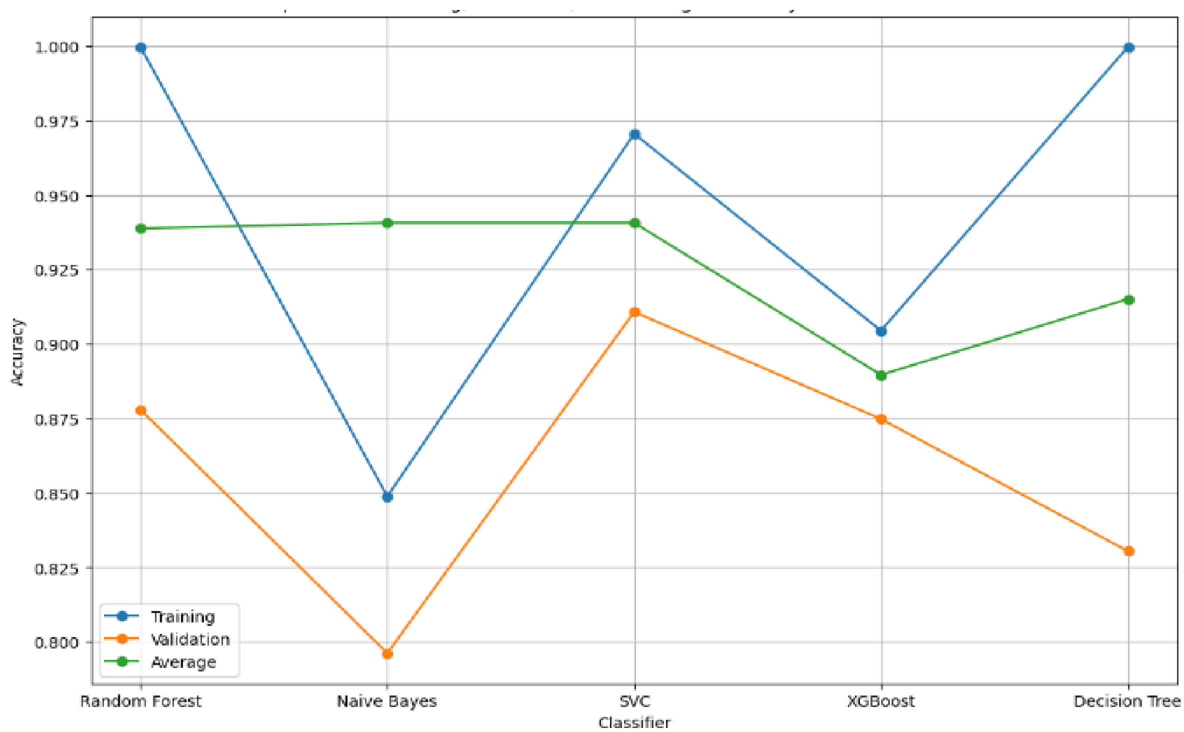


Fig 4.4: SVC classifier Training and Validation Accuracy.

To achieve the highest accuracy, we focus on the SVM classifier, which categorizes all reports and utilizes the normalization matrix. The process of detecting cyberbullying text is demonstrated using the SVM classification technique, along with the labeled English social media data, as illustrated in Fig. 4.10. The figure highlights two types of misclassified data that the classifier identifies as either harmful (six classes) or not. Employing machine learning algorithms, rather than traditional detection or prediction methods, proves to be one of the most effective ways to gain insights without supervision. This approach has demonstrated exceptional performance compared to earlier methods.

```
Text: ''m proud inability deal cold hate treadmill much cold' is 'not_abusive'.
Text: 'want baddest bitch world right lap' is 'gender'.
Text: 'pussy prolly taste heaven' is 'ethnicity'.
Text: 'eagles fans dont fact mcoy trash year everybody figured spread offense' is 'age'.
Text: 'big oleeeee middle finger bitch' is 'age'.
Text: 'this guy said slept girls let puke now know girl dudes slut whore hoe' is 'age'.
Text: 'via charlie angus sending made china canadian flag pins conservatives' is 'religion'.
Text: 'snapchat got hacked hoes explaining' is 'age'.
Text: 'let remove neutrality disputed template future once citations accumulated future safely neutrality disputed template off that assuming bi
Text: 'perhaps history country thoroughly disgraced excoriated james comey released inspector generals report ashamed himself' is 'political'.
Text: 'Okay all my sexy cougar aunties in the comments 😍👏👏👏' is 'gender'.
```

Fig 4.5: Abusive text detection.



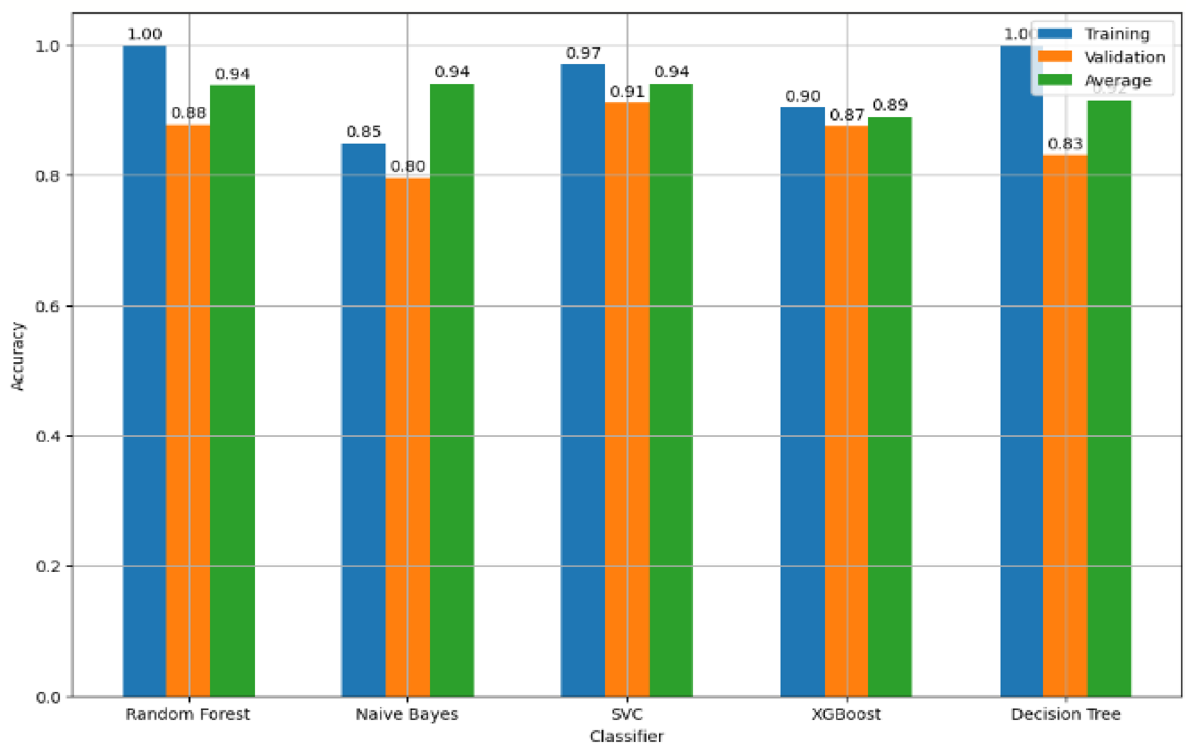


Fig 4.6: Comparison of five classifiers' average accuracy, validation, and training

4.4 Discussion

Our study employs predictive machine learning (ML) techniques to identify abusive text associated with cyberbullying. In any field of research, every word plays a significant role in the classification process. The primary objective of this study has been to detect harmful text effectively. Data serves as a critical element in any research, and the results of an experiment can vary significantly depending on the dataset used. Since we combined multiple datasets, we recognized that the results obtained by others using previously published public datasets might differ from ours when paired with our actual dataset.

We achieved our objective by leveraging a variety of machine learning methods to ensure reliability and averaged scores. A total of five algorithms were used in this project. Before starting, we identified several key requirements. Once the algorithm was selected, we proceeded with its implementation and evaluated the accuracy of each approach. Among the methods used, the SVM classifier demonstrated the highest accuracy at 91.07%, outperforming the other models. Additionally, a web-based tool was developed to detect abusive texts related to cyberbullying.

Precision: Precision measures the ability of a model to correctly identify true positives out of all predicted positives. It is calculated using the following formula:

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall evaluates the model's ability to identify all actual positive instances out of the total true positives. The equation is structured as follows:

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The F1 score provides a balanced evaluation of the model's performance by combining precision and recall. It is calculated using the formula below:

$$F1\ score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall}$$

Accuracy: Accuracy represents the proportion of correctly predicted instances to the total instances in the dataset. It is calculated as:

$$Accuracy = \frac{TP + TN}{Total\ Instances}$$

\

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Cyberbullying has profound effects on both its victims and society as a whole. While traditional bullying can be equally or more harmful, cyberbullying leads to significant emotional distress, including anxiety, depression, and in extreme cases, suicidal ideation or behavior. Beyond individual victims, cyberbullying negatively impacts society by fostering an environment of violence and intolerance, undermining empathy, and eroding trust within communities. It disrupts efforts to create inclusive, safe, and supportive spaces online and offline. Addressing cyberbullying requires proactive measures from individuals, schools, and local communities. These measures include educating young people about responsible online behavior, setting clear expectations and consequences for cyberbullying, and promoting positive digital citizenship. By collaborating on these initiatives, we can foster respect and safety in the online space, ultimately making it better for everyone.

5.2 Impact on the Environment

While cyberbullying itself does not directly affect the environment, the tools and strategies used to combat it can have unintended environmental consequences. The reliance on digital devices and online platforms, essential to addressing cyberbullying, contributes to resource consumption and environmental degradation through greenhouse gas emissions and hazardous waste from device production and disposal. Additionally, implementing interventions and awareness programs may involve activities that increase the carbon footprint. Although these measures are vital for creating safer online spaces, it is essential to consider their environmental implications and strive to minimize any negative impacts while addressing the issue of cyberbullying.

5.3 Ethical Aspects

Cyberbullying raises significant ethical concerns, as it involves intentional harm inflicted on others through digital means. Understanding the motivations and consequences of such behavior is critical for fostering a supportive learning environment where students can grow into responsible digital citizens. Further research is needed to explore the ethical dimensions of cyberbullying, including how moral awareness, personal accountability, and the consequences of one's actions can deter such behavior. This deeper understanding will help build a culture of ethical conduct online and reduce the prevalence of harmful behavior.

5.4 Sustainability Plan

This research aims to provide long-term strategies to reduce the occurrence of cyberbullying on social media. Based on the findings, effective security measures can be implemented, such as identifying false accounts and analyzing the emotional state of potential bullies. These techniques can enhance the ability to detect and prevent abusive behavior. Additionally, improving user experience—such as providing ad-free services—can encourage engagement with safer platforms. Agreements with international support networks can also ensure immediate assistance for users experiencing issues.

Key benefits of this research include:

- **Advancing knowledge:** Researchers will gain deeper insights into abusive text detection techniques.
- **Practical applications:** The findings can inform tools and methods to effectively address cyberbullying in digital spaces.

By integrating these approaches, we can build a more sustainable, supportive, and inclusive online environment.

CHAPTER 6

Conclusion

6.1 Summary

This study has provided significant insights into the topic of cyberbullying. Despite its importance, predicting cyberbullying remains a debated issue. Cyberbullying detection systems have seen declining use due to limitations in identifying abusive content. To address this, machine learning (ML) was used to classify seven key risk groups: non-abusive, religious, gender-related, age-related, insulting, political, and ethnic. ML models were also applied to predict whether a text is abusive based solely on its features.

The primary aim of this research was to explore the subject in depth using authentic data combined with information from online tools. These tools categorized content from platforms like Twitter and Instagram into seven classifications. Using this labeled data, five ML techniques were tested, with the SVM classifier achieving exceptional accuracy in identifying the precise “full_text” and its associated label. This approach facilitated accurate identification of abusive communication. While initial challenges arose, the intended goals were successfully met. It was observed that outcomes varied among students, and this was elaborated further in the study.

6.2 Conclusion

Cyberbullying poses a severe threat to individuals’ emotional and mental health, manifesting through various harmful behaviors such as offensive posts, false rumors, or online trolling. Regardless of one’s age, gender, or social status, cyberbullying can lead to anxiety, depression, and even suicide. It is crucial to raise awareness and take concrete actions to combat and prevent cyberbullying. These include promoting responsible online behavior, encouraging victims to report incidents, and offering support to those affected. Parents, schools, and governments have vital roles in fostering safe online environments.

One effective solution involves using supervised machine learning algorithms trained on labeled datasets of cyberbullying-related texts. Factors like the choice of algorithm, its configuration, and the quality of training data impact the success of detection systems. This study highlights the challenges and successes of using five classifiers for detecting abusive text, with SVM emerging as the most accurate, achieving a 91.07% accuracy rate. Other classifiers also performed well, reinforcing the effectiveness of ML techniques in this area.

6.3 Possible Impacts

The outcomes of this study suggest that results may vary depending on the context or observation site. Therefore, accuracy and legitimacy were prioritized by integrating both real-world and online datasets. ML models offer the potential to detect abusive language earlier than traditional methods, enabling timely intervention and improved outcomes. However, the widespread application of ML in platforms and mobile apps increases the likelihood of encountering cyberbullying content. By adopting a more user-focused approach, ML tools can reach and support individuals who might otherwise hesitate to seek help.

6.4 Future Work

This research has been an ongoing learning process, and further exploration is necessary. Cyberbullying, particularly among adolescents, continues to be a significant issue in today's digital age. Natural language processing (NLP) can play a pivotal role in detecting cyberbullying by analyzing the structure and content of online messages. Supervised ML algorithms can help classify these messages as cyberbullying or non-cyberbullying. By training classifiers on labeled datasets, it becomes possible to automatically detect instances of cyberbullying in new online content. Future work will focus on enhancing these models and expanding their application to create safer online spaces

References

1. D. Poeter, “Study: A Quarter of Parents Say Their Child Involved in Cyberbullying,” [Online], 2011.
2. Maurya, C., Muhammad, T., Dhillon, P., & Maurya, P. (2022). “The effects of cyberbullying victimization on depression and suicidal ideation among adolescents and young adults: A three-year cohort study from India.” *BMC Psychiatry, 22*(1).

3. Cucu, P. (2022, June 20). “Cyberbullying: Facts, statistics, and how to stop and prevent it.” *Heimdal Security Blog*. [Online]. Available: https://heimdalsecurity.com/blog/how-to-stop-and-prevent-cyberbullying/. [Accessed: April 15, 2023].

4. Alam, K. S., Bhowmik, S., & Prosun, P. R. (2021). “Cyberbullying detection: An ensemble-based machine learning approach.” *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*.

5. R. R. Dalvi, S. Baliram Chavan, and A. Halbe, “Detecting a Twitter cyberbullying using machine learning,” *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020.

6. A. Muneer and S. M. Fati, “A comparative analysis of Machine Learning Techniques for cyberbullying detection on Twitter,” *Future Internet*, vol. 12, no. 11, p. 187, 2020.
7. Balakrishnan, V., Khan, S., & Arabnia, H. R. (2020). “Improving cyberbullying detection using Twitter users’ psychological features and machine learning.” *Computers & Security, 90,* p. 101710.

8. Islam, M. M., Uddin, M. A., Islam, L., Akter, A., Sharmin, S., & Acharjee, U. K. (2020). “Cyberbullying detection on social networks using machine learning approaches.” *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE).*

9. Huang, Q., Singh, V. K., & Atrey, P. K. (2014). “Cyberbullying detection using social and textual analysis.” *Proceedings of the 3rd International Workshop on Socially-Aware Multimedia.*

10. K. Dinakar, R. Reichart, and H. Lieberman, “Modeling the Detection of Textual Cyberbullying,” *Proc. IEEE International Fifth International AAAI Conference on Weblogs and Social Media (SWM’11)*, Barcelona, Spain, 2011.

11. D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, “Detection of Harassment on Web 2.0,” *Proc. Content Analysis of Web 2.0 Workshop (CAW 2.0)*, Madrid, Spain, 2009.
12. Jain, V., Saxena, A. K., Senthil, A., Jain, A., & Jain, A. (2021). “Cyber-bullying detection in social media platforms using machine learning.” *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART).*

13. Maity, K., Kumar, A., & Saha, S. (2022). “A multitask multimodal framework for sentiment and emotion-aided cyberbullying detection.” *IEEE Internet Computing, 26*(4), pp. 68–78.

14. Lin, H., Siarry, P., Gururaj, H. L., Rodrigues, J., & Jain, D. K. (2022). “Special issue on deep learning methods for cyberbullying detection in multimodal social data.” *Multimedia Systems, 28*(6), pp. 1873–1875.

15. Maity, K., Saha, S., & Bhattacharyya, P. (2022). “Emoji, sentiment, and emotion-aided cyberbullying detection in Hinglish.” *IEEE Transactions on Computational Social Systems,* pp. 1–10.

16. K. Kao, “Social Media Addiction linked to cyberbullying,” 30-Mar-2021.

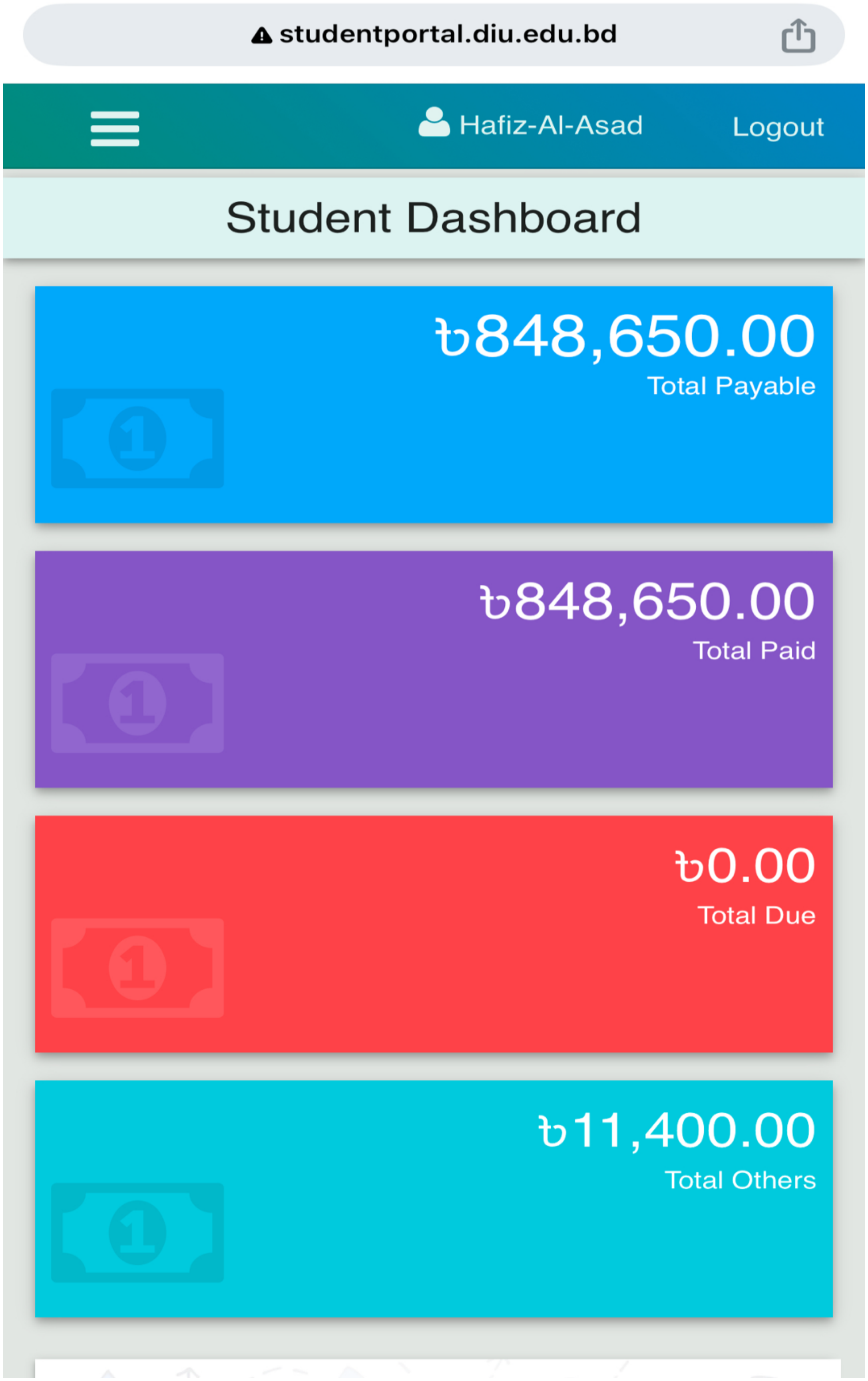
Cyber bully

ORIGINALITY REPORT

23%	19%	15%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Md. Shafiur Rahman Aronno, Md.Thoufiq Zumma, Rashed Prodhan, Fatema Tuz Zohora, Nazmus Sakib, K.B.M. Tahmiduzzaman. "A Study of Cyber Bullying Classification Using Social Media and Texual Analysis Based on Machine Learning Approches", 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023 Publication	4%
3	Submitted to Daffodil International University Student Paper	2%
4	Submitted to BRAC University Student Paper	2%
5	www.ijsrd.com Internet Source	1%
6	Mohamed Lahby, Al-Sakib Khan Pathan, Yassine Maleh. "Combatting Cyberbullying in	1%


studentportal.diu.edu.bd
Hafiz-Al-Asad
Logout
Student Dashboard
৳848,650.00
Total Payable
৳848,650.00
Total Paid
৳0.00
Total Due
৳11,400.00
Total Others