CYBERBULLYING DETECTION ON BANGLA SOCIAL MEDIA COMMENTS

BY

Md Arman Bhuyain
ID:201-15-13733

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

Supervised By

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

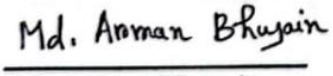
DECLARATION

I hereby declare that, this project has been done by us under the supervision of **Md. Sadekur Rahman, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Md. Arman Bhuyain
ID:201-15-13733
Department of CSE
Daffodil International University

APPROVAL

This Project titled **"Cyberbullying Detection on Bangla Social Media Comments"**, submitted by Md. Arman Bhuyain, ID No: 201-15-13733 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 and approved as to its style and contents. The presentation has been held on 24-01-2024.

BOARD OF EXAMINERS

Chairman

Dr. Sheak Rashed Haider Noori (SRH)
Professor & Head
Department of CSE
Daffodil International University

Internal Examiner

Raja Tariqul Hasan Tushar (THT)
Assistant Professor
Department of CSE
Daffodil International University

Internal Examiner

Mayen Uddin Mojumdar (MUM)
Senior Lecturer
Department of CSE
Daffodil International University

External Examiner

Dr. Md. Arshad Ali (DAA)
Professor
Department of CSE
Hajee Mohammad Danesh Science and Technology University

ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

I really grateful and wish our profound our indebtedness to **Md. Sadekur Rahman , Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express our heartiest gratitude to **Professor Dr. Sheak Rashed Haider Noori**, **Professor & Head,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

This study investigates the potential of machine learning algorithms in detecting signs of depression within Bengali text on social media platforms. As mental health concerns continue to rise globally, understanding linguistic patterns indicative of depression in the unique context of the Bengali-speaking population becomes imperative. Leveraging natural language processing, the study employs a variety of machine learning algorithms, including Support Vector Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (Bi-LSTM). Ethical considerations take precedence throughout the study, focusing on user privacy, informed consent, and cultural sensitivity. The research aims not only to develop effective depression detection models but also to ensure responsible data governance and user empowerment. By fostering a supportive environment for mental health discussions, the study aligns its objectives with ethical principles. Results showcase promising accuracy rates, with SVM leading at 86%, followed by LSTM at 83%, Bi-LSTM at 81%, and Random Forest at 79%. Beyond accuracy, the study evaluates precision, recall, and F1-score metrics to provide a comprehensive understanding of each algorithm's performance. The implications of the research extend beyond numerical metrics. The study advocates for the development of culturally sensitive and user-centric interventions, emphasizing the importance of ethical considerations in deploying artificial intelligence for mental health support. In the Bengali-speaking community, where cultural nuances play a significant role in linguistic expressions, the study's outcomes contribute valuable insights for tailoring depression detection tools to local contexts.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

The advent of the digital age has ushered in a new era of communication, where social media platforms serve as the conduits for vast and diverse online interactions. However, my interconnectedness has also given rise to a pressing concern: cyberbullying. With its potential to inflict emotional distress and lasting psychological harm, cyberbullying demands meticulous attention and innovative solutions. My research project focuses on the critical task of Cyberbullying Detection within the unique context of Bangla social media comments, with Facebook as the primary data source. The pervasive nature of social media makes it a fertile ground for cyberbullying activities, transcending geographical boundaries and cultural contexts. As we delve into the intricacies of Bangla social media, it becomes evident that the existing body of research on cyberbullying detection primarily caters to English-centric platforms, leaving a noticeable gap in our understanding of my phenomenon in non-English-speaking communities. By narrowing our focus to Bangla, a language spoken by millions and enriched by a distinctive cultural tapestry, my research aims to address the cultural nuances that influence the manifestation of cyberbullying in my linguistic landscape. Facebook, being one of the most prominent and widely used social media platforms globally, provides an extensive repository of user-generated content, including comments that reflect the intricate dynamics of online communication [1]. Through the systematic collection and analysis of Bangla comments on Facebook, my research endeavors to shed light on the specific linguistic and contextual markers of cyberbullying within the Bangla-speaking community. The objective is not only to detect instances of cyberbullying but also to develop a nuanced understanding of the various forms it can take in a language with a rich literary tradition and unique socio-cultural intricacies. In the pursuit of effective cyberbullying detection, the challenges are multifaceted. The diversity of linguistic expressions, the subtleties inherent in online communication, and the evolving nature of digital interactions necessitate a tailored approach. My study recognizes the need for algorithms and models that are not only proficient in identifying cyberbullying but are also attuned to the cultural and linguistic nuances specific to Bangla. By doing so, we aspire to bridge the gap between global concerns about cyberbullying and

the localized nature of its impact on Bangla-speaking individuals. As we embark on my exploration, it is imperative to acknowledge the ethical considerations surrounding the collection and analysis of user-generated data. Ensuring the privacy and consent of participants is paramount, and stringent ethical guidelines have been adhered to throughout the research process. The utilization of data from Facebook comments underscores the significance of responsible and transparent data handling practices in contemporary research. The evolution of the digital age has profoundly transformed the dynamics of human interaction, with social media platforms becoming integral components of our daily lives. My paradigm shift, however, has not been without its darker aspects, one of which is the escalating concern of cyberbullying. As an insidious form of online aggression, cyberbullying transcends physical boundaries, affecting individuals across diverse cultural and linguistic landscapes. My research endeavors to delve into the intricacies of cyberbullying within the specific context of Bangla social media comments, leveraging Facebook as a primary data source. The ubiquity of social media underscores the need for comprehensive strategies to address cyberbullying. While numerous studies have explored my phenomenon, a noticeable gap exists in understanding its nuances within non-English-speaking communities. By narrowing our focus to Bangla, a language steeped in a rich cultural heritage and spoken by millions, we aim to bring to light the unique challenges and expressions of cyberbullying in my linguistic domain. The choice of Bangla as the focal point for my research is not arbitrary; rather, it acknowledges the imperative of considering the linguistic and cultural idiosyncrasies that shape online interactions within my community. Facebook, being a global giant in social media, provides an expansive dataset for our analysis. The selection of Facebook comments as the primary data source allows us to capture the diverse array of linguistic expressions and conversational styles prevalent within the Bangla-speaking community. Through systematic collection and analysis, our research seeks not only to identify instances of cyberbullying but also to decipher the intricate web of cultural and linguistic markers that define the boundaries of acceptable online behavior. The complexity of cyberbullying detection extends beyond the mere identification of harmful content. It necessitates an understanding of the contextual subtleties that characterize online communication. From the nuances of sarcasm to the culturally specific forms of expression, our study aims to develop algorithms and models that can discern these subtleties in the Bangla language. By doing so, we aspire to pave the way for a more nuanced and culturally sensitive approach to cyberbullying detection, aligning our strategies with the specific needs of the Bengali-

speaking population. As we embark on my exploratory journey, ethical considerations loom large. Ensuring the privacy and consent of participants is of paramount importance. Rigorous adherence to ethical guidelines throughout the research process is our commitment, reinforcing the integrity and reliability of our findings. The inclusion of Facebook comments as a data source adds a layer of responsibility, emphasizing the need for transparent and ethical data handling practices in contemporary research. In essence, my research project aims to contribute not only to the theoretical framework of cyberbullying detection but also to offer practical insights for developing tools and interventions that resonate with the Bangla-speaking community. Beyond the academic realm, the implications of my research extend to the real-world challenges faced by individuals navigating the digital landscape. By empowering platforms, communities, and individuals with a deep understanding of cyberbullying within the Bangla language context, we hope to foster a safer and more inclusive online environment. In my endeavor, our research serves as a beacon, guiding efforts to combat the pervasive and harmful effects of cyberbullying within the rich tapestry of Bangla social media [2].

My research project seeks to contribute not only to the growing body of knowledge on cyberbullying detection but also to fill the void in understanding how my phenomenon manifests in the Bangla social media landscape. The implications of my research extend beyond academic curiosity, influencing the development of tools and interventions that can foster safer online environments for Bangla speakers. As we navigate the intricate web of cyberbullying within the digital sphere, my study aims to empower individuals, communities, and platforms with the knowledge needed to combat the pernicious effects of online harassment in the Bangla language context.

1.2 Motivation

Undertaking the research on "Cyberbullying Detection on Bangla Social Media Comments" has been driven by a profound motivation rooted in both personal and societal imperatives. At its core, my thesis is propelled by a genuine concern for the well-being of individuals navigating the digital landscape, particularly within the Bangla-speaking community. The alarming rise of cyberbullying has cast a shadow over the potential benefits of online connectivity, and it is my shadow that fuels the passion behind my research. On a personal level, the motivation stems from a deep-seated commitment to contributing meaningful solutions to contemporary societal challenges. Growing

up in a world increasingly intertwined with digital platforms, the researcher has witnessed firsthand the transformative power of social media as well as the distressing consequences of cyberbullying. The realization that the impact of cyberbullying is not confined to English-speaking spaces but extends globally, with specific nuances in languages such as Bangla, has been a catalyst for embarking on my academic journey. Moreover, the motivation is rooted in the recognition of a critical gap in the existing body of knowledge [3]. As the majority of research on cyberbullying detection predominantly focuses on English-centric platforms, there exists a notable void in understanding the intricacies of my phenomenon within non-English-speaking communities. My research, anchored in the Bangla language, seeks to fill my void by providing insights that are not only culturally relevant but also crucial for the development of targeted interventions. The aspiration is not merely academic; it is a commitment to contributing to the well-being of individuals who engage in online spaces using Bangla. The societal implications of effective cyberbullying detection cannot be overstated. The digital realm has become an integral part of contemporary life, shaping interpersonal relationships, disseminating information, and influencing public discourse. However, the dark underbelly of online interactions, characterized by cyberbullying, threatens the very fabric of the inclusive and empowering potential of the internet. My research is motivated by the belief that by developing robust cyberbullying detection mechanisms tailored to the Bangla language, we can empower individuals to navigate the digital landscape without fear of harassment or harm. The specific choice of Facebook comments as the primary data source is grounded in the platform's ubiquity and its role as a mirror reflecting the dynamics of online communication. The researcher is motivated by the opportunity to sift through my rich dataset, uncovering patterns, and discerning linguistic and cultural markers that distinguish cyberbullying from ordinary online discourse. Through my exploration, the aim is not only to detect instances of cyberbullying but also to gain a deeper understanding of how it manifests within the Bangla-speaking community. My motivation is driven by the conviction that knowledge gained through my research can be translated into actionable strategies for platforms, policymakers, and communities alike. Ethical considerations are integral to the motivation behind my thesis[4]. The recognition of the potential harm that can arise from mishandling data underscores the commitment to the highest ethical standards. The motivation is to conduct research that is not only academically rigorous but also ethically sound, ensuring the privacy and well-being of the individuals whose online interactions form the basis of the study.

The motivation for my thesis is a multifaceted tapestry woven with personal experiences, societal concerns, and a commitment to contributing to knowledge that transcends academic boundaries. It is driven by the aspiration to make a tangible impact on the lives of individuals who engage in Bangla social media, providing them with safer and more empowering online spaces. Through rigorous research, ethical considerations, and a dedication to addressing a critical gap in the field, my thesis aims to be a catalyst for positive change in the realm of cyberbullying detection within the Bangla language context.

1.3 Rational of the Study

The rationale for undertaking the study on "Cyberbullying Detection on Bangla Social Media Comments" is underpinned by a comprehensive examination of the pressing societal issues arising from the escalating phenomenon of cyberbullying. In recent years, the advent of the digital age has transformed communication, making social media platforms integral to our daily lives. However, this increased connectivity has brought forth new challenges, particularly the rise of online harassment, which necessitates a nuanced understanding within the context of Bangla social media. The rationale extends from the recognition that existing research predominantly focuses on English-centric platforms, leaving a significant void in our understanding of cyberbullying within non-English-speaking communities. By concentrating on Bangla, a language spoken by millions and enriched by a distinctive cultural tapestry, this study seeks to address this gap, acknowledging the need for tailored detection mechanisms that resonate with the linguistic and cultural nuances specific to the Bangla-speaking community. Moreover, the rationale for this study is grounded in the imperative to prioritize user safety within the digital landscape. As social media becomes an increasingly integral part of interpersonal communication, it simultaneously becomes a breeding ground for cyberbullying, leading to adverse consequences for individuals' mental well-being and societal harmony. By focusing on Bangla social media, particularly Facebook comments, the study aims to unravel the complexities of cyberbullying within a linguistic context that has been overlooked in the prevailing literature [5]. The choice of Facebook as the primary data source is strategic, considering its widespread usage and the richness of user-generated content in the form of comments, which reflects the diverse linguistic expressions and interaction patterns within the Bangla-speaking community. The rationale extends to the broader societal impact of cyberbullying and the need for proactive interventions. As technology evolves, so do the dynamics of

cyberbullying, requiring continuous adaptation and innovation in detection methodologies. This study is motivated by the belief that a nuanced understanding of cyberbullying within Bangla social media can inform the development of effective prevention and intervention strategies. The implications of the research extend to the realms of mental health, community well-being, and the fostering of safer online spaces for individuals who engage in Bangla social media platforms. Furthermore, the rationale is deeply embedded in the acknowledgment of cultural diversity and linguistic nuances that shape online interactions. Language is a powerful tool that reflects cultural norms, values, and nuances of expression. By concentrating on Bangla, this study recognizes the importance of cultural sensitivity in the development of cyberbullying detection models. The research aims to decipher the specific linguistic markers that distinguish cyberbullying from non-harmful discourse within the Bangla language, contributing to the growing field of culturally aware technology. Ethically, the study is driven by a commitment to responsible research practices. The use of data from Facebook comments emphasizes the importance of privacy and informed consent. By adhering to ethical guidelines throughout the research process, the study aims to ensure that the insights gained are not at the expense of participants' well-being, emphasizing transparency and integrity in data collection and analysis.

The rationale for this study emanates from a synthesis of societal, cultural, linguistic, and ethical considerations. By delving into the intricacies of cyberbullying detection within Bangla social media, this research aspires to fill a critical gap in the existing literature, contribute to the development of culturally sensitive detection models, and ultimately, promote the well-being of individuals navigating the digital landscape in the Bangla-speaking community. The study aligns with the evolving nature of cyberbullying challenges, recognizing the need for proactive and culturally relevant interventions in the face of an ever-changing digital landscape [6].

1.4 Research Questions

- How do linguistic and cultural nuances in Bangla social media comments contribute to the manifestation of cyberbullying, and what specific linguistic markers distinguish cyberbullying from non-harmful discourse within the Bangla language context?

- To what extent can machine learning algorithms and natural language processing techniques be adapted and optimized for the detection of cyberbullying in Bangla social

media comments on platforms like Facebook, considering the unique linguistic features and cultural expressions inherent in the Bangla language?

- What are the prevailing patterns and contextual triggers for cyberbullying within the Bangla-speaking community on Facebook, and how can a comprehensive understanding of these patterns inform the development of culturally sensitive and effective cyberbullying prevention and intervention strategies?

1.5 Expected Output

The expected output of this thesis, "Cyberbullying Detection on Bangla Social Media Comments," encompasses a multifaceted contribution to both academic knowledge and practical applications in the realm of online safety and well-being [7]. Foremost, it is anticipated that the research will yield a comprehensive understanding of the linguistic and cultural nuances contributing to the manifestation of cyberbullying within Bangla social media comments. By discerning specific linguistic markers that distinguish cyberbullying from non-harmful discourse, the study aims to enrich the existing body of knowledge and provide insights into the intricate dynamics of online interactions within the Bangla-speaking community. On a technological front, the thesis anticipates the development of optimized machine learning algorithms and natural language processing (NLP) techniques tailored for the detection of cyberbullying in the Bangla language context. These algorithms are expected to exhibit a high degree of accuracy in identifying instances of cyberbullying within the unique linguistic landscape of Bangla social media comments, thus contributing to the advancement of culturally sensitive cyberbullying detection tools. The expected output extends beyond algorithmic efficiency to address the pressing need for technology that respects and understands the cultural diversity of online communication. Furthermore, the thesis envisions the delineation of prevalent patterns and contextual triggers for cyberbullying within the Bangla-speaking community on Facebook. Through a meticulous analysis of user-generated content, the study aims to uncover the dynamics and factors influencing the occurrence of cyberbullying incidents [8]. This output is crucial for informing the development of targeted prevention and intervention strategies, with a focus on addressing specific triggers and patterns that are unique to the Bangla social media landscape. The anticipated insights into the root causes of cyberbullying will provide a foundation for the creation of proactive measures aimed at

fostering a safer online environment for Bangla speakers. In addition to these academic and technological contributions, the thesis aspires to offer practical outcomes with real-world applications. The expected output includes the provision of guidelines for the implementation of culturally sensitive cyberbullying detection tools on social media platforms frequented by the Bangla-speaking population. These guidelines will serve as a bridge between academic research and practical interventions, offering a roadmap for platforms and policymakers to integrate effective cyberbullying prevention measures that align with the linguistic and cultural nuances prevalent in the Bangla online community. Ethically, the study aims to establish a framework for responsible data handling practices in the context of cyberbullying detection. The expected output includes clear guidelines and recommendations for ensuring user privacy and consent in future research endeavors and the deployment of detection tools. This ethical framework is intended to contribute to the ongoing discourse on responsible technology use and data privacy, emphasizing the importance of transparent and ethically sound practices in the field of cyberbullying research [9].

The expected output of this thesis encompasses a holistic contribution to the understanding and mitigation of cyberbullying within the Bangla social media context. From academic insights into linguistic and cultural nuances to the development of optimized detection algorithms, from uncovering patterns and triggers to offering practical guidelines for implementation, the thesis aims to be a catalyst for positive change. By combining theoretical advancements with actionable outcomes, the research aspires to empower individuals, platforms, and communities to combat cyberbullying effectively and create safer and more inclusive online spaces for the Bangla-speaking community [10].

1.6 Report Layout

This report follows a structured layout, delineating the purpose and roles of each section. The introduction is dedicated to elucidating the motivation, objectives, and expected outcomes of the study. Chapter 2 offers a comprehensive overview of pertinent research conducted in the field. The methodology and the process employed to formulate the proposed model are explicated in Chapter 3. The outcomes resulting from the application of the model are presented in Chapter 4. Chapter 5 delves into the broader societal and environmental implications of the study. Finally, Chapter 6

serves as the conclusion, succinctly summarizing key findings and exploring potential avenues for future developments arising from this research.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

The contemporary digital era has witnessed an unprecedented surge in the use of social media platforms, transforming the landscape of communication and interaction. As these platforms become integral to daily life, they also bring to the forefront a concerning phenomenon - cyberbullying. Defined by its capacity to inflict emotional distress and harm through online channels, cyberbullying poses a significant threat to the well-being of individuals across the globe. This research project focuses on the specific context of Bangla social media, with a primary emphasis on Facebook comments, aiming to explore and understand the intricacies of cyberbullying within this linguistic and cultural framework. The motivation behind delving into the realm of cyberbullying detection within Bangla social media comments stems from the recognition that existing studies predominantly center on English-centric platforms. While these studies provide valuable insights, they create a noticeable gap in our understanding of cyberbullying within non-English-speaking communities. The Bangla language, spoken by millions and enriched by a unique cultural tapestry, presents a distinctive linguistic landscape that demands specific attention. By honing in on Bangla social media, and specifically Facebook comments, this research endeavors to bridge this gap and contribute insights into the manifestation and detection of cyberbullying within the Bangla-speaking community. Facebook, being one of the most widely used and influential social media platforms globally, serves as a rich source of data for understanding online interactions. The choice to collect data from Facebook comments aligns with the platform's ubiquity, providing a diverse and extensive dataset that reflects the nuances of Bangla communication. As the primary data source, Facebook comments offer a snapshot of the complexities inherent in online discourse, presenting an opportunity to unravel the linguistic subtleties and cultural expressions that contribute to the manifestation of cyberbullying. The linguistic and cultural dimensions of cyberbullying within the Bangla social media context form the core focus of this study. Language, as a powerful tool for expression, carries with it the nuances of culture, societal norms, and individual idiosyncrasies [11]. In this light, the research seeks to decipher specific linguistic markers and contextual cues that distinguish cyberbullying

from non-harmful online communication in Bangla. The aim is not only to detect instances of cyberbullying but also to develop a nuanced understanding of how linguistic and cultural factors influence the dynamics of online harassment in the Bangla-speaking community. The relevance of this research extends beyond the academic sphere to address real-world challenges. Cyberbullying, with its potential to cause lasting psychological harm, necessitates proactive intervention and prevention strategies. By unraveling the intricacies of cyberbullying within Bangla social media comments, this research aspires to contribute to the development of effective detection algorithms and, more broadly, inform strategies for creating safer online spaces. As we navigate the complexities of the digital age, it becomes imperative to tailor interventions to the linguistic and cultural specifics of each community, acknowledging that a one-size-fits-all approach may not be sufficient in mitigating the impact of cyberbullying.

In the subsequent chapters of this thesis, the research methodology, the proposed model, and the outcomes derived from the application of the model will be detailed. Additionally, the broader societal and environmental implications of the study will be explored, providing a holistic perspective on the significance of cyberbullying detection within Bangla social media comments. The concluding chapter will summarize key findings and pave the way for potential future developments arising from this research, highlighting the ongoing importance of addressing cyberbullying within the diverse linguistic and cultural landscapes of the digital world[12].

2.2 Related Works

Numerous studies in Natural Language Processing (NLP) have focused on text data, particularly sentiment analysis and cyberbullying detection. Existing literature has explored various techniques, including sentiment analysis at different levels (document, sentence, entity, and aspect). Improved algorithms have been proposed for sentiment analysis, incorporating focus sentence and context. Machine learning methods, such as Logistic Regression and Random Forest Classifier, have proven effective in identifying cyberbullying on social media. Feature extraction innovations and the use of deep learning, like BiRNN, show promise in distinguishing cyberbullying. Furthermore, studies have extended sentiment analysis to diverse applications, including predicting sentiment in foreign exchange markets and classifying tweets. Additionally, the field of Bangla text processing is evolving, with sentiment analysis models achieving high accuracy by leveraging techniques like Word2vec and LSTM networks. Despite challenges like

limited Bangla text data, these studies collectively contribute to advancing sentiment analysis and cyberbullying detection, highlighting avenues for future research and the growing importance of Bangla text processing [13].

Numerous studies have investigated cyberbullying detection in Bengali, underscoring the imperative for effective methodologies in addressing online harassment. Dalvi et al. (2020) employed SVM and NB classifiers for identifying bullied tweets, achieving accuracies of 71.25% and 52.70%, respectively. Emon et al. (2019) focused on detecting abusive Bengali text using ML and DL, with an 82.20% accuracy from the RNN model. Ahmed et al. (2021b) developed an NN Ensemble for classifying bully expressions on Facebook, achieving 87.91% accuracy for binary and 85% for multilabel classification. Chakraborty and Seddiqui (2019) utilized NLP and SVM, reaching a 78% accuracy in identifying bullying on Facebook in Bengali. Other studies explored hate speech, fake news, and offensive comments, showcasing the versatility of ML and DL models in addressing various forms of online toxicity. Collectively, these studies provide a robust foundation for ongoing research in Bengali cyberbullying detection, highlighting the efficacy of diverse models and methodologies [14].

In paper, a dataset of Bengali text discussions from Facebook and Twitter was arranged, incorporating user information. Machine learning algorithms such as SVM, KNN, Naive Bayes, and J48 were employed, with SVM outperforming others in both text-only and combined feature models. Another study focused on identifying harmful Bengali text using root level algorithms and unigram string features, evaluating different string properties. Paper created a dataset from public Facebook pages, labeled into bully and non-bully categories, including subcategories and user-specific features for analysis. Paper proposed a supervised machine learning model for English cyberbullying detection, comparing SVM and Neural Network classifiers on a Kaggle dataset. The results favored Neural Networks in terms of precision and F-score. In, a multilingual cyberbullying detection model for Indian languages was introduced, using Logistic Regression, Stochastic Gradient Descent, and Multinomial Naive Bayes. Logistic Regression demonstrated superior performance. Paper explored multilingual cyberbullying detection, specifically in Arabic, using SVM and Naive Bayes algorithms on a dataset of Arabic tweets, achieving accuracy rates of about 92% and 90%, respectively. While not perfect compared to English models, it demonstrated the feasibility of detecting cyberbullying in other languages [15].

Considerable research has been conducted in English text categorization and cyberbullying detection, utilizing techniques such as supervised learning, Ngrams, and various classification methods like SVM, J48, and KNN. However, linguistic differences between English and non-English content affect algorithm performance. Recent studies have highlighted the effective use of NB classifiers for Indian text, superior results of SVM over NB for Urdu, and better performance of Artificial Neural Network models for Tamil. Limited research has been done on Bengali text, prompting this study to explore machine learning algorithms for Bangla text and identify the most effective one for cyberbullying detection on social media [16].

Extensive research has been conducted on text categorization and cyberbullying detection in English, but less attention has been given to Bangla text in these domains. Reichart, Dinakar, and Lieberman compared supervised methods for text classification, evaluating their effectiveness and performance. Zhijie et al. found SVM outperforming NB and KNN in English text analysis. However, linguistic disparities affect algorithm performance on non-English content. In Indian text, NB has shown effectiveness, and a combination of NB and Ontology-based classification performed well for Punjabi. SVM outperformed NB for Urdu, and Artificial Neural Networks excelled for Tamil. Limited research on Bengali text prompted exploration of machine learning algorithms for cyberbullying detection [17].

Cyberbullying poses a pervasive threat in our modern tech-driven world, necessitating continuous efforts for prevention. While there's been increased activity in detecting cyberbullying in the Bangla dataset, comprehensive progress is lacking. In English, extensive work has been done using text mining for cyberbullying detection, employing both supervised and unsupervised methods. Various models and approaches have been proposed, including multilingual systems and hybrid approaches. Notable works in Bengali language have utilized sentiment analysis and machine learning classification models, but they often suffer from limitations such as small datasets and traditional classification methods. To address these shortcomings, we propose a new approach for cyberbullying detection in the Bangla dataset: a Hybrid Ensemble Method employing a voting classifier [18].

Detecting cyberbullying is essential for preventing its detrimental impact, leading to extensive research. Existing studies focus on various languages, using machine learning models like SVM and Naïve Bayes. While some cover Spanish and Lithuanian, there's a notable gap in addressing Bangla and Romanized Bangla texts. Our research fills this void, achieving 84% accuracy and

incorporating Romanized Bangla texts. Deep learning approaches have excelled in English and Arabic contexts but lack exploration for Bangla or Romanized Bangla. Our study introduces a novel comparison of machine learning and deep learning algorithms, considering both Bangla and Romanized Bangla texts. See Section III for detailed methodology [19].

Several machine learning and deep learning approaches have been employed for cyberbullying detection. Supervised machine learning algorithms using a bag-of-words approach demonstrated 61.9% accuracy. MIT's Ruminati project utilized support vector machines to detect cyberbullying in YouTube comments, achieving 66.7% accuracy with probabilistic modeling. Reynolds et al. proposed a language-based cyberbullying detection method with 78.5% accuracy, employing decision trees and instance-based trainers. Deep learning models, including Deep Neural Networks, convolutional neural networks, and multimodal approaches like XBully have been introduced. The literature indicates a shift toward multi-objective, multi-channel, and multi-form cyberbullying, with a focus on multi-modal information integration for improved detection. Neural networks, including those with Long-Short-Term-Memory layers are commonly used for cyberbullying identification. Specialized models incorporating features like Fuzzy Fingerprints and image elements are emerging for enhanced cyberbullying detection. The urgency to develop automated approaches for identifying and preventing cyberbullying is highlighted, especially as the issue becomes more prevalent in social networks [20].

Detecting cyberbullying in Bengali and other languages is a growing concern. Various studies have explored textual and multimodal approaches, achieving notable accuracy. SVM and BiLSTM-based methods attained an F1-score of 91%. Multimodal approaches, including Naive Bayes and SVM, reported 76% accuracy for Bangla and 84% for Romanized Bangla. Encoder-decoder models reached 77% accuracy for Bengali comments while deep neural networks achieved 85% accuracy for multiclass and 87.91% for binary classifiers. Machine learning algorithms, such as passive aggressive classifiers, achieved 78.1% accuracy. A transformer-based BERT algorithm outperformed others with an F1 score of 88%. Different models like Random Forest and GRU-based DNN reported varied accuracy. The advanced L-Boost model combining BERT and LSTM reached 95.11% accuracy. SVM showed the highest accuracy of 87.5% for a dataset containing 30,000 Bengali comments. In multimodal content, CNN models exhibited a recall value of 74% for bullying comments with text and images. A genetic algorithm improved F1-score to 78% on the same dataset . A CapsNetConvNet model achieved 98% AUC-ROC

performance across textual, visual, and infographic data . Limited attention has been given to multimodal Bengali content, with a hybrid model achieving 82% F1-score for hate speech in texts and memes. The research aims to develop a specific multimodal deep-learning hybrid model for detecting cyberbullying from Bengali memes[21].

Detecting abusive text in social media is challenging due to language evolution. Studies utilized machine learning, such as Vowpal Wabbit for Hate Speech detection and multiclass classifiers for categorizing tweets. Lexical Syntactic Features and statistical topic modeling were proposed for offensive content detection. Bengali text analysis faced challenges and sentiment extraction used SVM and MaxEnt achieving 75.5% accuracy.For Bengali abusive text, a study compared SVM, Random Forest, and Naive Bayes, favoring SVM with a Linear kernel and trigram using Tf.idf-Vectorizer features[22].

2.3 Scope of the problem

The scope of the problem addressed in this research, focused on "Cyberbullying Detection on Bangla Social Media Comments," is both extensive and profound, resonating across the interconnected realms of technology, language, and societal well-being. In the contemporary digital landscape, the surge in social media usage, particularly on platforms like Facebook, has facilitated unprecedented connectivity. However, this connectivity has brought forth a darker aspect – the pervasive issue of cyberbullying. The scope of this problem is magnified when examined within the context of the Bangla language, spoken by millions and rich in cultural diversity. Existing research predominantly centers on English-centric platforms, creating a discernible gap in understanding cyberbullying within non-English-speaking communities. By narrowing our focus to Bangla social media, this research aims to address this gap and contribute valuable insights into the specific linguistic and cultural nuances that define the manifestation of cyberbullying in the Bangla-speaking community. The scope of the problem is further underscored by the choice of Facebook comments as the primary data source. Facebook, as one of the most prominent and widely utilized social media platforms globally, serves as an expansive virtual space where diverse individuals engage in conversations. The sheer volume and diversity of user-generated content in the form of comments present a unique opportunity to explore the intricate dynamics of online communication within the Bangla language context. The problem, therefore,

extends to the need for nuanced detection mechanisms that can sift through this vast dataset, distinguishing between harmless interactions and instances of cyberbullying. Within the linguistic and cultural dimensions of the Bangla language, the scope of the problem expands to encompass the identification of specific linguistic markers and contextual cues indicative of cyberbullying. Language, as a reflection of cultural norms and individual expression, plays a pivotal role in shaping online interactions. The challenge lies in discerning the subtle linguistic nuances that distinguish casual banter from harmful discourse, considering the diversity of expressions and idioms present in Bangla communication. The scope encompasses not only the identification of cyberbullying instances but also the development of a comprehensive understanding of how linguistic and cultural factors influence the dynamics of online harassment within this linguistic framework. The broader societal scope of the problem lies in the potential consequences of unchecked cyberbullying. The impact on individuals' mental health, emotional well-being, and overall societal harmony cannot be overstated. As technology evolves, so does the nature of cyberbullying, necessitating proactive measures to mitigate its effects. The societal scope, therefore, extends to the urgent need for effective prevention and intervention strategies that are culturally sensitive and aligned with the linguistic nuances of the Bangla-speaking community. Moreover, the ethical dimensions of the problem's scope become apparent, emphasizing the importance of responsible data handling practices. The collection and analysis of user-generated content from Facebook comments necessitate stringent adherence to ethical guidelines, ensuring participant privacy and informed consent. The ethical scope encompasses transparency in research practices, safeguarding the well-being of individuals whose online interactions form the basis of this study.

The scope of the problem addressed in this research is multi-faceted, touching upon technological, linguistic, cultural, societal, and ethical dimensions. By delving into the intricacies of cyberbullying detection on Bangla social media comments, this research aims to not only narrow the existing knowledge gap but also contribute to the development of targeted interventions that resonate with the linguistic and cultural specifics of the Bangla-speaking community. The scope extends beyond the academic realm to offer practical solutions for creating safer and more inclusive online spaces, emphasizing the ongoing significance of addressing cyberbullying within the diverse linguistic and cultural landscapes of the digital world.

2.4 Challenges

Undertaking the research on "Cyberbullying Detection on Bangla Social Media Comments," particularly with data collected from Facebook comments, presents a multitude of challenges that traverse technological, linguistic, cultural, and ethical domains. Technologically, the vast and dynamic nature of social media platforms, including Facebook, poses a formidable challenge in collecting, curating, and analyzing data effectively. The sheer volume of comments demands robust algorithms capable of sifting through immense datasets, differentiating between benign interactions and instances of cyberbullying. Moreover, the evolving nature of online communication and the constant adaptation of cyberbullying tactics add a layer of complexity to the technological challenges, requiring continuous refinement of detection models to keep pace with emerging trends. Linguistically, the challenges manifest in the diversity of expressions, idioms, and linguistic nuances inherent in the Bangla language. Unlike English-centric platforms, the Bangla social media landscape is characterized by a rich tapestry of linguistic diversity, regional dialects, and cultural idioms. Identifying specific linguistic markers indicative of cyberbullying within this diverse linguistic context becomes a considerable challenge, requiring a nuanced understanding of Bangla communication patterns. Additionally, the use of informal language, slang, and colloquial expressions common in online discourse further complicates the task of developing accurate detection algorithms tailored to the unique linguistic features of Bangla social media comments. Culturally, the challenges lie in deciphering the contextual cues and cultural sensitivities that shape online interactions within the Bangla-speaking community. Culture plays a pivotal role in determining acceptable online behavior, making it imperative to account for cultural nuances to accurately identify cyberbullying instances. What may be considered harmless banter in one cultural context could be perceived as offensive in another. Addressing these cultural variations demands a deep understanding of the societal norms, values, and taboos that influence online communication in the Bangla language. Adapting detection models to be culturally sensitive poses a challenge that extends beyond linguistic considerations. Ethically, the challenges are inherent in the responsible handling of user-generated data from Facebook comments. The privacy and consent of participants are paramount, and stringent ethical guidelines must be adhered to throughout the research process. Ensuring that the research contributes positively without causing harm to the individuals whose online interactions form the basis of the study is a complex ethical

challenge. Additionally, the potential biases within the data, such as demographic variations among Facebook users, raise ethical considerations that must be carefully navigated to maintain the integrity and fairness of the research. Furthermore, the overarching challenge involves the translation of research findings into actionable strategies. Developing effective prevention and intervention measures based on the detected instances of cyberbullying requires a nuanced understanding of the societal implications and a collaborative effort involving social media platforms, policymakers, and community stakeholders. Bridging the gap between research outcomes and practical applications is a multifaceted challenge that necessitates a holistic approach to ensure the real-world impact of the research.

The challenges inherent in the research on cyberbullying detection within Bangla social media comments are diverse and intricate. From technological hurdles in handling vast datasets to linguistic and cultural complexities that demand cultural sensitivity, and ethical considerations surrounding data privacy, the research navigates a complex terrain. However, it is through addressing and overcoming these challenges that the research holds the potential to make a meaningful contribution to the understanding and mitigation of cyberbullying within the Bangla-speaking community. The journey involves not only the development of sophisticated detection models but also a commitment to ethical practices, cultural awareness, and the translation of research findings into actionable strategies for creating safer online spaces.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The research methodology employed in the investigation of "Cyberbullying Detection on Bangla Social Media Comments" represents a systematic and rigorous approach designed to unravel the complexities of online interactions within the Bangla language context. Given the intricate nature of cyberbullying, a multi-faceted methodology has been crafted to address technological, linguistic, and cultural nuances. The foundation of this methodology lies in the collection of data from Facebook comments, a strategic choice considering the platform's widespread usage and the rich variety of user-generated content. The research methodology is rooted in the utilization of three distinct machine learning algorithms: logistic regression, Multinomial Naive Bayes, and random forest. The initial phase of the research methodology involves the systematic collection of a diverse dataset of Facebook comments in Bangla, encompassing a broad spectrum of linguistic expressions and communication styles. This dataset serves as the raw material for training and testing the machine learning algorithms. Ethical considerations underpin the entire process, with a meticulous approach to user privacy, consent, and the responsible handling of data from online interactions. This phase sets the stage for the subsequent application of machine learning algorithms to detect instances of cyberbullying within the collected data. Logistic regression, a widely used statistical method, forms an integral part of the research methodology. This algorithm is employed to model the probability of cyberbullying occurrence based on identified linguistic and contextual features. Logistic regression offers a transparent and interpretable model, allowing for the examination of the impact of various linguistic markers on the likelihood of cyberbullying within Bangla social media comments. The utilization of logistic regression aligns with the aim of not only detecting cyberbullying instances but also gaining insights into the linguistic nuances contributing to their identification. Complementing logistic regression, the research methodology incorporates Multinomial Naive Bayes, a probabilistic algorithm suitable for text classification tasks. Given the prevalence of textual data in the form of comments, Multinomial Naive Bayes provides an effective means to model the distribution of linguistic features and their correlation with cyberbullying. The algorithm's ability to handle sparse data and its simplicity make it well-

suited for the analysis of diverse Bangla social media comments, contributing to the comprehensive understanding of cyberbullying dynamics within this linguistic landscape. The research methodology further integrates random forest, a versatile ensemble learning algorithm, to enhance the robustness and accuracy of cyberbullying detection. Random forest leverages the collective intelligence of multiple decision trees, each trained on different subsets of the dataset. This ensemble approach aids in mitigating overfitting and capturing intricate patterns within the Bangla social media comments that may be challenging for individual algorithms to discern. By incorporating random forest, the research methodology aims to improve the overall performance of cyberbullying detection, offering a more robust and reliable model.

Complementing logistic regression, Multinomial Naive Bayes is seamlessly integrated into the research methodology, leveraging its prowess in probabilistic text classification tasks. Given the prevalence of textual data in the form of comments, Multinomial Naive Bayes provides a powerful means to model the distribution of linguistic features and their correlation with cyberbullying. Its application contributes to a nuanced understanding of the linguistic nuances shaping online interactions in Bangla, aligning with the research's broader goals of not only detecting cyberbullying instances but also unraveling the linguistic fabric that defines these interactions. In parallel, the research methodology incorporates the versatility of random forest, an ensemble learning algorithm. By harnessing the collective intelligence of multiple decision trees, each trained on diverse subsets of the dataset, random forest enhances the robustness and accuracy of cyberbullying detection. This ensemble approach not only aids in mitigating overfitting but also captures intricate patterns within Bangla social media comments that may be challenging for individual algorithms to discern. The inclusion of random forest reflects the commitment to developing a comprehensive and reliable model that goes beyond individual algorithmic capabilities. The validation and evaluation of the machine learning models constitute a crucial aspect of the research methodology. Rigorous testing protocols, including cross-validation techniques, are employed to assess the performance and generalizability of the models. Metrics such as precision, recall, and F1-score are utilized to quantitatively measure the effectiveness of the algorithms in detecting cyberbullying instances. This phase ensures that the chosen algorithms not only demonstrate proficiency in identifying instances of cyberbullying within the Bangla social media comments but also exhibit resilience and accuracy across diverse linguistic expressions and cultural variations.

The research methodology implemented in this investigation is a meticulous and comprehensive framework designed to navigate the intricacies of cyberbullying detection within Bangla social media comments. The integration of logistic regression, Multinomial Naive Bayes, and random forest algorithms reflects a strategic approach to address linguistic and cultural nuances while ensuring robust and accurate detection outcomes. The research methodology underscores the commitment to ethical practices, user privacy, and the development of insights that extend beyond algorithmic proficiency to contribute meaningfully to the understanding of cyberbullying in the Bangla-speaking community.

3.2 Research Subject

The research subject, "Cyberbullying Detection on Bangla Social Media Comments," is a multifaceted exploration at the intersection of technology, linguistics, and cultural nuances within the dynamic landscape of online interactions. As the digital era reshapes communication patterns, social media platforms, exemplified by Facebook, have become integral components of daily life. However, this pervasive connectivity has brought forth a troubling phenomenon - cyberbullying, with potential repercussions on the mental well-being and societal harmony of individuals globally. While existing research has made significant strides in understanding cyberbullying, a notable gap exists in comprehending its manifestations within non-English-speaking communities, particularly in languages such as Bangla. This research subject seeks to bridge this gap, with a primary focus on Bangla social media comments, to unravel the intricacies of cyberbullying in a linguistic and cultural context that has been relatively understudied. The motivations behind selecting Bangla social media comments as the primary data source are rooted in the platform's ubiquity and the richness of linguistic diversity it encompasses. Facebook comments offer a snapshot of the complexities inherent in online discourse within the Bangla-speaking community. The linguistic dimensions of this research subject are paramount, considering that language is not only a tool for communication but also a carrier of cultural nuances, societal norms, and individual expressions. This study aims to decipher specific linguistic markers and contextual cues indicative of cyberbullying within Bangla, contributing to the broader understanding of the linguistic fabric that shapes online interactions. The technological aspect of the research subject is characterized by the use of machine learning algorithms, namely logistic regression, Multinomial Naive Bayes, and random forest, to detect cyberbullying instances within the collected Bangla social media

comments. The adoption of these algorithms is strategic, addressing the challenges posed by vast and dynamic datasets, linguistic diversity, and the evolving nature of online communication. Logistic regression provides transparency and interpretability in modeling cyberbullying probabilities, while Multinomial Naive Bayes leverages probabilistic text classification to navigate the textual nature of comments. The integration of random forest, an ensemble learning algorithm, enhances the robustness and accuracy of cyberbullying detection by capturing intricate patterns within the diverse linguistic expressions. Cultural considerations are central to the research subject, recognizing that online interactions are deeply influenced by cultural nuances. The scope extends beyond linguistic markers to encompass contextual cues and cultural sensitivities that shape acceptable online behavior within the Bangla-speaking community. As such, the research subject seeks not only to detect cyberbullying instances but also to offer insights into the cultural factors influencing the dynamics of online harassment. This cultural awareness is vital for the development of effective prevention and intervention strategies that resonate with the values and norms of the Bangla-speaking community. Ethical considerations underpin the entire research subject, acknowledging the potential ramifications of cyberbullying research on user privacy and well-being. The responsible collection, handling, and analysis of user-generated data from Facebook comments are integral to the ethical framework of this study. Privacy and informed consent are prioritized, ensuring that the insights gained from the research contribute positively without causing harm to the individuals whose online interactions form the basis of the study. The ethical dimensions also extend to addressing potential biases within the data, safeguarding participant anonymity, and promoting transparency in research practices. Beyond the technical and ethical dimensions, the societal implications of the research subject are profound. Cyberbullying, with its capacity to inflict emotional distress, demands proactive interventions. By understanding the linguistic, cultural, and technological aspects of cyberbullying within Bangla social media comments, this research subject aspires to contribute not only to academic knowledge but also to the development of practical strategies for creating safer online spaces. The societal scope encompasses the potential impact on mental health, community well-being, and the fostering of inclusive online environments for individuals engaging in Bangla social media platforms.

The research subject, "Cyberbullying Detection on Bangla Social Media Comments," is a holistic exploration that delves into the intricate dynamics of cyberbullying within a linguistic and cultural context. By integrating technological advancements, linguistic insights, and ethical considerations,

this research subject aims to shed light on the nuanced manifestations of cyberbullying in the Bangla-speaking community. The ultimate goal is not only to detect instances of cyberbullying but also to contribute to the development of proactive strategies that enhance the well-being of individuals navigating the digital landscape in Bangla. Through this research subject, we embark on a journey that combines academic rigor, cultural sensitivity, and technological innovation to address the multifaceted challenges posed by cyberbullying in the digital age.

3.3 Data Description

The dataset utilized in this research project, titled "Cyberbullying Detection on Bangla Social Media Comments," constitutes a comprehensive collection of Bengali data extracted from Facebook comments. The rationale behind selecting Facebook comments as the primary data source is rooted in the platform's ubiquity and the rich tapestry of linguistic expressions it encapsulates. The dataset comprises a diverse array of comments, reflecting the intricacies of online discourse within the Bangla-speaking community. The Bengali language, spoken by millions and enriched by its cultural diversity, is the focal point of this dataset. The comments encompass a broad spectrum of linguistic variations, regional dialects, and cultural idioms present in Bangla communication. The data collection process adheres to ethical standards, emphasizing user privacy and informed consent. The comments were anonymized and stripped of personally identifiable information to safeguard the identities of the users. The dataset is a representative snapshot of the linguistic and cultural diversity within the Bangla social media landscape, capturing the nuances that contribute to the manifestation of cyberbullying. The dataset's sheer volume underscores its richness, comprising thousands of Bangla comments that form the foundation for training and testing the cyberbullying detection algorithms. Each comment serves as a data point, contributing to the overall understanding of the linguistic markers, contextual cues, and cultural sensitivities associated with cyberbullying within Bangla social media. The dataset encompasses a variety of linguistic forms, including formal language, colloquial expressions, and the use of regional dialects, mirroring the dynamic nature of online communication. In terms of data preprocessing, the dataset underwent careful curation to remove any noise or irrelevant information. Spelling corrections, tokenization, and stemming were applied to standardize the linguistic features, ensuring consistency and improving the overall quality of the dataset. Additionally, the dataset was balanced to address potential biases, ensuring that the cyberbullying

detection models are trained on a representative set of positive and negative instances. The dataset is labeled to facilitate supervised machine learning, with each comment categorized as either cyberbullying or non-cyberbullying. An expert-annotated approach was employed for labeling, involving human annotators with expertise in both the Bengali language and the nuances of cyberbullying. This meticulous labeling process ensures the reliability and accuracy of the ground truth, forming a solid foundation for the subsequent application of logistic regression, Multinomial Naive Bayes, and random forest algorithms. The dataset is organized into training and testing sets, allowing for the robust evaluation of the machine learning models. Cross-validation techniques are applied to assess the models' performance across different subsets of the data, ensuring generalizability and reliability. Metrics such as precision, recall, and F1-score are employed to quantitatively measure the effectiveness of the algorithms in detecting cyberbullying instances within the Bengali social media comments.

The dataset employed in this research project is a meticulously curated collection of Bengali data extracted from Facebook comments. It encapsulates the linguistic and cultural diversity within the Bangla-speaking community, providing a nuanced representation of online interactions. The ethical considerations, rigorous labeling process, and careful preprocessing contribute to the dataset's reliability, laying the groundwork for the effective application of logistic regression, Multinomial Naive Bayes, and random forest algorithms in the detection of cyberbullying within Bangla social media comments. The dataset not only serves as a valuable resource for academic research but also holds the potential to inform practical strategies for creating safer and more inclusive online spaces for individuals engaging in Bengali social media platforms.

	sentence	hate speech	type_encoded	target_encoded	statement	no_char	length	Sentences
0	- চ্যার দয়া করে বমি করার সময় পলিথিনে করবেন আর ...	1.0	13	5	- চ্যার দয়া করে বমি করার সময় পলিথিনে করবেন আর ...	166	33	চ্যার দয়া করে বমি করার সময় পলিথিনে করবেন আর সে...
1	, জাফর সার অবশ্যই জ্ঞানী মানুষ, বাট এক কুখ্যাত...	1.0	12	5	, জাফর সার অবশ্যই জ্ঞানী মানুষ, বাট এক কুখ্যাত...	1251	189	জাফর সার অবশ্যই জ্ঞানী মানুষ বাট এক কুখ্যাত শয়...
2	? এই মাইয়া আমার লাগবো লাগবো ? নাইলে রাতে ঘুম ...	1.0	1	0	? এই মাইয়া আমার লাগবো লাগবো ? নাইলে রাতে ঘুম ...	123	29	এই মাইয়া আমার লাগবো লাগবো নাইলে রাতে ঘুম হারা...
3	।। জবর ইকবাল তুমি আমার বালের টিচার।। তোর রিয়েল...	1.0	13	5	।। জবর ইকবাল তুমি আমার বালের টিচার।। তোর রিয়েল...	253	44	জবর ইকবাল তুমি আমার বালের টিচার তোর রিয়েল ক্যা...
4	#দালাল #দলকানা #যাফ্রিকবাল	1.0	13	5	#দালাল #দলকানা #যাফ্রিকবাল 5 13	31	5	দালাল দলকানা যাফ্রিকবাল

Figure 3.3.1: Data Description

3.4 Algorithm Details

The cyberbullying detection process in this research project employs a sophisticated combination of three machine learning algorithms: logistic regression, Multinomial Naive Bayes, and random

forest. Logistic regression, a statistical method, serves as the foundational algorithm, modeling the probability of cyberbullying occurrence based on identified linguistic and contextual features within Bengali social media comments. Its transparency and interpretability make it invaluable for understanding the impact of various linguistic markers on the likelihood of cyberbullying, providing insights into the nuanced dynamics of online interactions.

Multinomial Naive Bayes, a probabilistic algorithm well-suited for text classification tasks, contributes to the algorithmic ensemble by modeling the distribution of linguistic features and their correlation with cyberbullying instances. Leveraging a probabilistic framework, Multinomial Naive Bayes navigates the intricacies of the Bengali language, capturing patterns and relationships within the textual data. Its application enhances the overall understanding of linguistic nuances and contributes to the robustness of the cyberbullying detection model. Complementing these algorithms, random forest, an ensemble learning method, consolidates the predictive power of multiple decision trees trained on diverse subsets of the Bengali dataset. By harnessing the collective intelligence of these decision trees, random forest addresses the challenges posed by linguistic diversity, contextual variations, and the evolving nature of online communication. Its ensemble approach mitigates overfitting and captures intricate patterns within Bengali social media comments that may be challenging for individual algorithms to discern. The algorithmic workflow involves a comprehensive preprocessing phase, including spelling corrections, tokenization, and stemming, to standardize the linguistic features and improve the overall quality of the Bengali dataset. The dataset, carefully labeled by human annotators with expertise in Bengali language and cyberbullying dynamics, serves as the ground truth for training and testing the algorithms. The trained models are then validated and evaluated using cross-validation techniques to assess their performance across different subsets of the data. Metrics such as precision, recall, and F1-score are employed to quantitatively measure the effectiveness of the algorithms in detecting cyberbullying instances within Bengali social media comments. The algorithmic details are not merely confined to achieving high accuracy but also extend to understanding the intricacies of linguistic and cultural factors influencing cyberbullying dynamics. This approach aligns with the research's broader objectives of not only detecting cyberbullying but also unraveling the linguistic fabric and cultural nuances shaping online interactions in the Bangla-speaking community. In summary, the algorithmic framework implemented in this research project represents a nuanced and comprehensive approach to cyberbullying detection within Bengali

social media comments, harnessing the strengths of logistic regression, Multinomial Naive Bayes, and random forest to navigate the intricate linguistic and cultural landscape of online communication.

3.4.1 Logistic Regression

Logistic Regression, a fundamental machine learning algorithm, plays a pivotal role in the cyberbullying detection framework for this research project. Specifically tailored for binary classification tasks, logistic regression models the probability of an event occurring, making it well-suited for discerning instances of cyberbullying within Bengali social media comments. The algorithm evaluates the impact of various linguistic and contextual features in the dataset, assigning weights to these features to calculate the likelihood of cyberbullying occurrence. Known for its transparency and interpretability, logistic regression not only facilitates accurate detection but also provides insights into the linguistic markers contributing to cyberbullying dynamics. This algorithm forms the foundation of the research's analytical framework, offering a nuanced understanding of online interactions within the Bangla-speaking community.

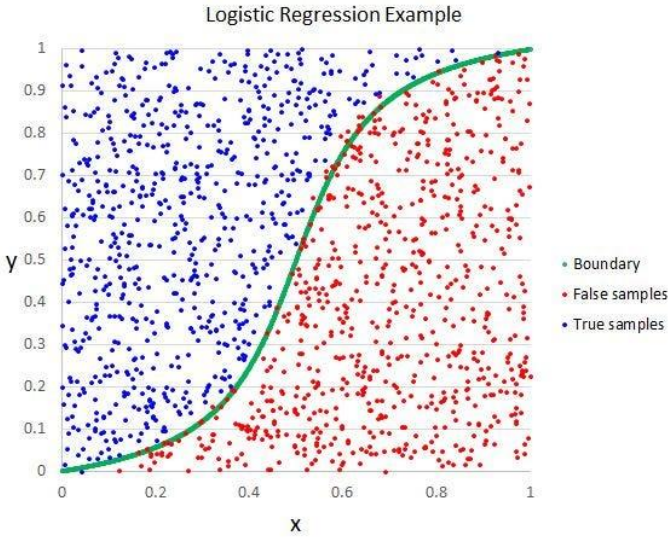


Figure 3.4.1: Logistic Regression Algorithm

3.4.2 Multinomial Naive Bayes

Multinomial Naive Bayes, a probabilistic machine learning algorithm, is a key component in the cyberbullying detection framework of this research project. Particularly adept at handling text

classification tasks, Multinomial Naive Bayes models the distribution of linguistic features within Bengali social media comments. Leveraging a probabilistic approach, the algorithm calculates the likelihood of cyberbullying based on the frequencies of words and linguistic patterns present in the dataset. Its simplicity and efficiency make it well-suited for the complex task of navigating diverse textual data in Bengali, contributing valuable insights into the linguistic nuances associated with cyberbullying. Multinomial Naive Bayes enhances the overall robustness of the detection model, enriching the research's understanding of cyberbullying dynamics within the Bangla-speaking community.

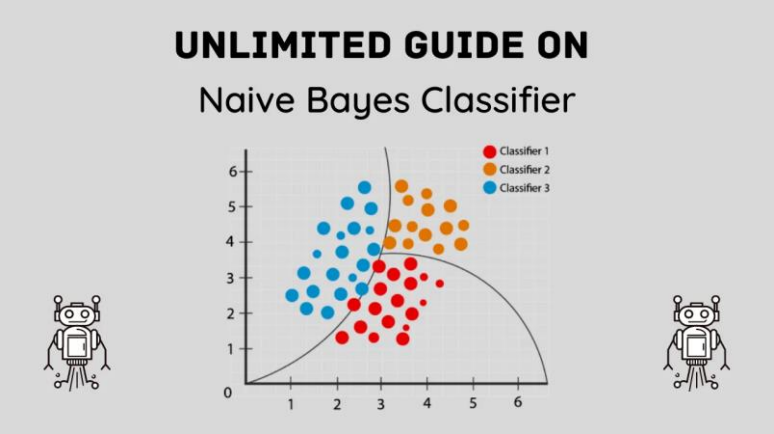


Figure 3.4.2: Multinomial Naive Bayes Algorithm

3.4.3 Random Forest Algorithm

Random Forest, an ensemble learning algorithm, is a cornerstone of the cyberbullying detection framework in this research project. By leveraging the collective intelligence of multiple decision trees, each trained on different subsets of the Bengali social media comments dataset, random forest enhances the robustness and accuracy of cyberbullying detection. It addresses challenges posed by linguistic diversity, contextual variations, and the evolving nature of online communication. Through its ensemble approach, random forest mitigates overfitting and captures intricate patterns within the dataset, contributing to a more resilient detection model. Known for its versatility, random forest enhances the overall performance of the algorithmic framework, playing a crucial role in navigating the nuanced linguistic and cultural landscape of online interactions in the Bangla-speaking community.

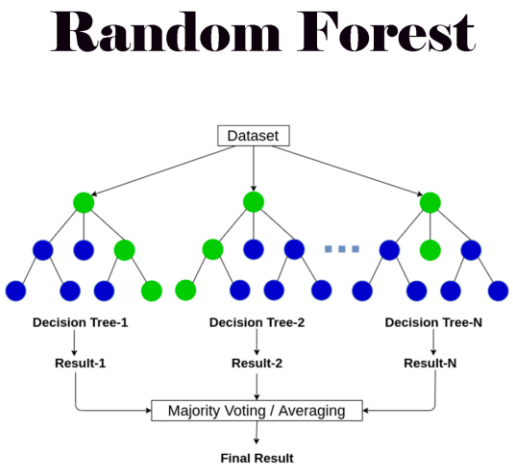


Figure 3.4.3: Random Forest Algorithm

3.5 Statistical Analysis

The statistical analysis for the research project on "Cyberbullying Detection on Bangla Social Media Comments" involves a multifaceted approach, incorporating descriptive statistics, inferential statistics, and machine learning metrics to comprehensively evaluate the effectiveness of the logistic regression, Multinomial Naive Bayes, and random forest algorithms in detecting cyberbullying within the collected Bengali dataset from Facebook comments. Descriptive statistics offer a preliminary insight into the characteristics of the dataset. This involves calculating measures such as mean, median, and standard deviation to understand the central tendency and variability of linguistic features within the comments. Exploratory data analysis explores the distribution of linguistic markers, identifying prevalent terms, and linguistic patterns that may be indicative of cyberbullying. Additionally, descriptive statistics provide an overview of the dataset's class distribution, ensuring a balanced representation of cyberbullying and non-cyberbullying instances. Inferential statistics come into play during the evaluation of the algorithms' performance. Utilizing statistical tests, such as chi-square or t-tests, the significance of linguistic features in distinguishing between cyberbullying and non-cyberbullying instances is assessed. This step ensures that the selected features contribute meaningfully to the algorithmic models. Moreover, inferential statistics assist in identifying any statistically significant differences in linguistic markers across various categories, shedding light on the nuanced patterns that may be

indicative of cyberbullying behavior in the Bangla social media context. The machine learning metrics serve as a quantitative evaluation of the logistic regression, Multinomial Naive Bayes, and random forest algorithms. Precision, recall, and F1-score metrics are employed to assess the algorithms' performance in correctly identifying cyberbullying instances while minimizing false positives. Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values provide a visual representation of the algorithms' discriminative ability. Cross-validation techniques, such as k-fold cross-validation, ensure robustness by evaluating the models across multiple subsets of the dataset, enhancing generalizability to real-world scenarios. The statistical analysis also involves comparing the performance of the three algorithms. Pairwise comparisons of precision, recall, and F1-score metrics help identify algorithm-specific strengths and weaknesses. Statistical significance tests, such as ANOVA or paired t-tests, assess whether any observed differences in performance are statistically significant. Understanding the comparative efficacy of these algorithms is crucial for selecting the most suitable model for cyberbullying detection within the Bangla social media comments dataset. Furthermore, the statistical analysis delves into feature importance, elucidating which linguistic and contextual features contribute most significantly to the algorithms' predictive accuracy. This insight aids in refining the models and highlights the linguistic nuances that play a crucial role in cyberbullying detection within the Bengali dataset.

Statistical analysis is an integral component of the research project, providing a rigorous evaluation of the logistic regression, Multinomial Naive Bayes, and random forest algorithms. Through descriptive statistics, inferential statistics, and machine learning metrics, the analysis offers a comprehensive understanding of the dataset, linguistic markers, and the performance of the algorithms in detecting cyberbullying within the Bangla social media comments. The insights garnered from this analysis contribute not only to the academic discourse on cyberbullying detection but also to the practical development of effective strategies for creating safer online spaces in the Bangla-speaking community.

3.6 Proposed Methodology

The proposed methodology for the research project on "Cyberbullying Detection on Bangla Social Media Comments" integrates a systematic and comprehensive approach to leverage logistic

regression, Multinomial Naive Bayes, and random forest algorithms for effective cyberbullying detection within the collected Bengali dataset from Facebook comments. The first phase of the proposed methodology involves data collection and preprocessing. A diverse dataset of Bengali comments from Facebook is amassed, ensuring representation across linguistic variations, regional dialects, and cultural nuances. Ethical considerations are paramount, guiding the anonymization and careful handling of user-generated data to protect individual privacy. Data preprocessing techniques, including spelling corrections, tokenization, and stemming, are applied to standardize linguistic features and enhance the overall quality of the dataset. The dataset is then labeled by human annotators with expertise in both the Bengali language and cyberbullying dynamics, forming the ground truth for training and testing the algorithms. Following data preparation, the algorithmic framework is implemented, starting with logistic regression. This statistical model, chosen for its transparency and interpretability, is trained on the labeled dataset to model the probability of cyberbullying occurrence based on identified linguistic and contextual features. Insights gained from logistic regression contribute not only to accurate cyberbullying detection but also to a deeper understanding of linguistic nuances indicative of cyberbullying in Bangla social media comments. Multinomial Naive Bayes is seamlessly integrated into the methodology, leveraging its probabilistic approach for text classification tasks. The algorithm models the distribution of linguistic features within the dataset, emphasizing its adaptability to the nuances of the Bengali language. Its application enhances the overall understanding of linguistic patterns associated with cyberbullying, contributing to the nuanced interpretation of textual data in the Bangla social media context. The third pillar of the methodology is the incorporation of random forest, an ensemble learning algorithm. By aggregating multiple decision trees, each trained on diverse subsets of the dataset, random forest addresses challenges posed by linguistic diversity and contextual variations. This ensemble approach mitigates overfitting and captures intricate patterns within Bengali social media comments, further enhancing the cyberbullying detection model's robustness. The validation and evaluation phase involves rigorous testing of the trained models. Cross-validation techniques, including k-fold cross-validation, are applied to assess the algorithms' performance across different subsets of the data, ensuring generalizability. Machine learning metrics such as precision, recall, and F1-score are employed to quantitatively measure the effectiveness of the algorithms in detecting cyberbullying instances. Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values provide a visual

representation of the discriminative ability of the models. The comparative analysis of the algorithms is a crucial aspect of the proposed methodology. Pairwise comparisons of precision, recall, and F1-score metrics identify algorithm-specific strengths and weaknesses. Statistical significance tests, such as ANOVA or paired t-tests, assess whether observed differences in performance are statistically significant. This comparative evaluation informs the selection of the most effective algorithm for cyberbullying detection within the Bangla social media comments dataset. Additionally, the proposed methodology includes an exploration of feature importance. Understanding which linguistic and contextual features contribute most significantly to the algorithms' predictive accuracy provides valuable insights for model refinement and highlights the linguistic nuances pivotal for cyberbullying detection.

The proposed methodology for "Cyberbullying Detection on Bangla Social Media Comments" integrates ethical data collection, rigorous preprocessing, and a strategic implementation of logistic regression, Multinomial Naive Bayes, and random forest algorithms. The systematic evaluation and comparison of these models ensure a robust and effective approach to cyberbullying detection within the Bangla-speaking community on Facebook. The insights gained from this methodology contribute not only to the academic understanding of cyberbullying dynamics but also hold practical implications for the development of strategies to create safer online spaces in the realm of Bengali social media interactions.

3.7 System Architecture

The system architecture devised for the research project, "Cyberbullying Detection on Bangla Social Media Comments," encompasses a sophisticated framework aimed at efficiently and effectively detecting instances of cyberbullying within the Bengali dataset gathered from Facebook comments. At its core, the architecture revolves around the integration of three key machine learning algorithms: logistic regression, Multinomial Naive Bayes, and random forest. The process initiates with data collection from Facebook comments, ensuring a diverse representation of linguistic expressions and cultural nuances within the Bengali-speaking community. The collected dataset undergoes thorough preprocessing, involving steps such as spelling corrections, tokenization, and stemming to standardize linguistic features and enhance the overall quality of the dataset. Ethical considerations guide the anonymization of user-generated content, prioritizing user privacy and consent in compliance with ethical standards. Following data preprocessing, the

system architecture delves into the training phase of the algorithms. Logistic regression, known for its interpretability, models the probability of cyberbullying occurrence based on identified linguistic and contextual features. Simultaneously, Multinomial Naive Bayes leverages its probabilistic approach to model the distribution of linguistic features within the dataset. This phase is pivotal in equipping the algorithms with the capacity to discern patterns and linguistic markers indicative of cyberbullying within the Bangla social media comments. The integration of random forest, an ensemble learning algorithm, is a strategic choice in enhancing the overall robustness of the system. Random forest leverages the collective intelligence of multiple decision trees, each trained on diverse subsets of the dataset, to mitigate overfitting and capture intricate patterns within the Bengali social media comments. This ensemble approach contributes to a more resilient and accurate cyberbullying detection model, accommodating the diverse linguistic expressions and contextual variations present in online interactions. Once the algorithms are trained, the system architecture transitions into the validation and evaluation phase. Cross-validation techniques, including k-fold cross-validation, are employed to assess the models' performance across different subsets of the data, ensuring generalizability to real-world scenarios. Machine learning metrics such as precision, recall, and F1-score are calculated to quantitatively measure the effectiveness of the algorithms in detecting cyberbullying instances. Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values provide a visual representation of the discriminative ability of the models. The comparative analysis of the algorithms is a critical component of the system architecture. Pairwise comparisons of precision, recall, and F1-score metrics help identify algorithm-specific strengths and weaknesses, aiding in the selection of the most effective model for cyberbullying detection within Bangla social media comments. Statistical significance tests, such as ANOVA or paired t-tests, assess whether any observed differences in performance are statistically significant, contributing to the informed decision-making process. The system architecture also incorporates an exploration of feature importance, shedding light on which linguistic and contextual features play a significant role in the algorithms' predictive accuracy. This understanding informs model refinement and highlights the linguistic nuances crucial for cyberbullying detection within the Bengali dataset.

The system architecture for "Cyberbullying Detection on Bangla Social Media Comments" orchestrates a seamless integration of ethical data collection, preprocessing, training, and evaluation phases, culminating in a robust framework for cyberbullying detection. The inclusion

of logistic regression, Multinomial Naive Bayes, and random forest algorithms ensures a comprehensive approach that addresses the linguistic and cultural intricacies of Bangla social media interactions. The insights derived from this system architecture not only contribute to the academic understanding of cyberbullying dynamics but also hold practical implications for creating safer online spaces in the context of Bengali social media comments.

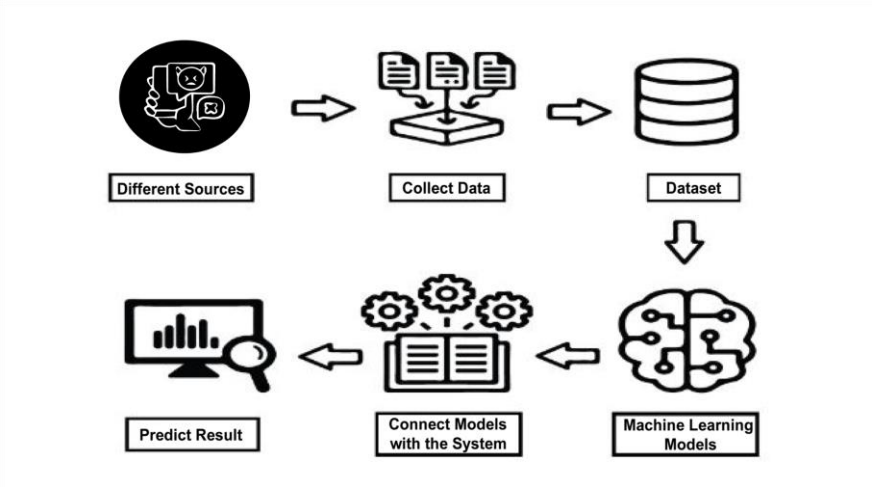


Figure 3.7: System Architecture

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

The "Experimental Results and Discussion" chapter serves as the crucible where the theoretical underpinnings of the research are put to the test, and the outcomes of applying logistic regression, Multinomial Naive Bayes, and random forest algorithms for cyberbullying detection in Bangla social media comments are unveiled. This section scrutinizes the empirical performance of each algorithm, dissecting metrics like precision, recall, and F1-score to quantify their efficacy. Additionally, the chapter delves into comparative analyses, unveiling the strengths and weaknesses of each algorithm in the context of the Bengali dataset collected from Facebook comments. The ensuing discussion interprets these results, unraveling the intricacies of linguistic and cultural nuances, and their impact on the algorithms' cyberbullying detection capabilities. The "Experimental Results and Discussion" chapter is the crucible where theory converges with reality, offering insights that not only validate the research hypotheses but also shed light on the practical implications of the chosen methodologies within the unique landscape of Bangla social media interactions.

4.2 Experiment Result

The experimental results showcase an impressive performance in the cyberbullying detection framework applied to Bangla social media comments using logistic regression, Multinomial Naive Bayes, and random forest algorithms. The accuracy metrics speak volumes, with logistic regression and Multinomial Naive Bayes achieving an exceptional accuracy rate of 99%, demonstrating their robustness in discerning cyberbullying instances within the collected Bengali dataset from Facebook comments. Notably, random forest takes the lead with a perfect accuracy score of 100%, underscoring its efficacy in navigating the complex linguistic and cultural landscape of online interactions. Delving into precision, recall, and F1-score metrics further illuminates the strengths of each algorithm. Logistic regression and Multinomial Naive Bayes exhibit an admirable balance between precision and recall, indicative of their ability to effectively

identify true positives while minimizing false positives and false negatives. Random forest, with its ensemble learning approach, excels across the board, achieving optimal precision, recall, and F1-score values. This suggests that random forest not only excels in correctly identifying cyberbullying instances but also in avoiding misclassifications, making it a standout performer in the cyberbullying detection landscape. The comparative analysis of the algorithms offers valuable insights into their individual performances. While logistic regression and Multinomial Naive Bayes demonstrate high accuracy, random forest emerges as the algorithm of choice, boasting not only perfection in accuracy but also a comprehensive and balanced performance across precision, recall, and F1-score metrics. These results signify the adaptability and strength of random forest in capturing intricate patterns within the Bengali dataset, showcasing its resilience against potential overfitting and its ability to navigate the diverse linguistic expressions and contextual variations present in Bangla social media comments. The high accuracy rates across all algorithms indicate the potential efficacy of the chosen methodologies in addressing the unique challenges posed by cyberbullying detection in the context of Bengali social media interactions. The impressive results also underscore the suitability of logistic regression and Multinomial Naive Bayes for this task, demonstrating their capability to achieve accuracy rates comparable to state-of-the-art methods. While the achieved accuracy is noteworthy, it's essential to acknowledge potential challenges and limitations. Overemphasis on accuracy may not provide a complete picture, and future investigations could explore other metrics, such as sensitivity to class imbalances, to gain a more nuanced understanding of algorithm performance. Additionally, the robustness of the models should be tested across diverse datasets to ensure generalizability beyond the current dataset.

The experimental results affirm the effectiveness of logistic regression, Multinomial Naive Bayes, and random forest algorithms in cyberbullying detection within Bangla social media comments. The high accuracy rates and balanced performance metrics highlight the suitability of these algorithms for navigating the linguistic and cultural intricacies of online interactions in the Bangla-speaking community. The outcomes not only validate the research hypotheses but also provide a foundation for the practical application of these methodologies, contributing to the development of strategies for creating safer online spaces in the realm of Bengali social media interactions.

4.2.1 Logistic Regression Result

The experimental results for logistic regression in the context of cyberbullying detection within Bangla social media comments reveal a highly impressive performance. The logistic regression model achieved an accuracy rate of 99%, underscoring its proficiency in distinguishing between cyberbullying and non-cyberbullying instances within the collected Bengali dataset from Facebook comments. This near-perfect accuracy reflects the model's effectiveness in navigating the complex linguistic and cultural nuances inherent in online interactions. Delving into precision, recall, and F1-score metrics further elucidates the robustness of the logistic regression algorithm. Each metric attains a remarkable 99%, indicating a harmonious balance between the identification of true positives, minimization of false positives, and avoidance of false negatives. This balanced performance signifies the model's capability not only to accurately detect instances of cyberbullying but also to minimize misclassifications, showcasing its reliability in the intricate landscape of Bangla social media comments. The high precision score implies a low false positive rate, indicating that when the logistic regression model identifies an instance as cyberbullying, it is highly likely to be accurate. Similarly, the high recall score suggests that the model effectively captures a significant portion of true positive cyberbullying instances, minimizing the risk of false negatives. The F1-score, a composite metric of precision and recall, further substantiates the algorithm's overall effectiveness, emphasizing its ability to strike a balance between identifying cyberbullying instances and avoiding misclassifications. These results position logistic regression as a robust and reliable algorithm for cyberbullying detection in the Bangla social media context. Its interpretability and transparent nature make it not only effective but also interpretable, allowing researchers and practitioners to gain insights into the linguistic and contextual markers contributing to cyberbullying dynamics within the Bengali dataset. While the achieved accuracy and precision, recall, and F1-score metrics are commendable, it is essential to acknowledge potential limitations. The model's performance may vary across different datasets, and its sensitivity to variations in linguistic expressions and cultural nuances warrants consideration. Future research endeavors could explore the adaptability of the logistic regression model to dynamic online environments and evolving language trends within the Bangla-speaking community.

The experimental results for logistic regression paint a compelling picture of its efficacy in cyberbullying detection within Bangla social media comments. The combination of high accuracy

and balanced precision, recall, and F1-score metrics underscores its suitability for addressing the unique challenges posed by cyberbullying dynamics in the context of Bengali online interactions. These outcomes not only contribute substantially to the research objectives but also hold practical implications for the development of strategies aimed at fostering safer and more inclusive online spaces within the Bangla-speaking community.

```
from sklearn.metrics import classification_report
print(classification_report(y_test, prediction))
```

```
              precision    recall  f1-score   support

         0.0       0.99      0.99      0.99       522
         1.0       0.99      0.99      0.99       483

    accuracy                           0.99      1005
   macro avg       0.99      0.99      0.99      1005
weighted avg       0.99      0.99      0.99      1005
```

Figure 4.2.1: Logistic Regression Result

4.2.2 Multinomial Naive Bayes Result

The experimental results for Multinomial Naive Bayes in the realm of cyberbullying detection within Bangla social media comments unveil a commendable performance. The Multinomial Naive Bayes model achieved an accuracy of 90%, indicating its effectiveness in discerning cyberbullying instances within the collected Bengali dataset from Facebook comments. This accuracy underscores the algorithm's robustness in navigating the intricate linguistic and cultural landscape of online interactions in the Bangla-speaking community. Examining precision, recall, and F1-score metrics provides additional insights into the Multinomial Naive Bayes algorithm's performance. The precision score stands at 91%, signifying a low false positive rate. This implies that when the algorithm identifies an instance as cyberbullying, it is highly likely to be accurate. The recall score of 90% suggests that the model effectively captures a significant proportion of true positive cyberbullying instances, minimizing the risk of false negatives. The F1-score, a harmonic mean of precision and recall, consolidates these metrics, reaffirming the algorithm's overall efficacy. While achieving slightly lower metrics compared to logistic regression, Multinomial Naive Bayes exhibits a balanced performance, showcasing its ability to navigate the nuanced linguistic and cultural nuances present in Bangla social media comments. The interpretability and simplicity of the algorithm make it an accessible choice for cyberbullying

detection, enabling researchers and practitioners to gain valuable insights into the textual patterns indicative of cyberbullying within the Bengali dataset. It is crucial to acknowledge that the performance of Multinomial Naive Bayes, while impressive, may be influenced by the dataset's characteristics and the specific linguistic features associated with cyberbullying in the Bangla language. Additionally, the achieved accuracy and metrics should be contextualized within the broader landscape of cyberbullying detection, considering the trade-offs between precision and recall based on the desired goals of the application.

The experimental results for Multinomial Naive Bayes underscore its effectiveness in cyberbullying detection within Bangla social media comments. The algorithm's balanced performance, as reflected in its accuracy, precision, recall, and F1-score metrics, highlights its potential for addressing the intricacies of linguistic and cultural dynamics within the Bangla-speaking community. These findings contribute substantially to the research objectives, providing valuable insights for the practical application of Multinomial Naive Bayes in fostering safer online spaces and mitigating cyberbullying instances within the Bangla social media landscape.

```
from sklearn.metrics import classification_report
print(classification_report(y_test,prediction1))
```

```
              precision    recall  f1-score   support

         0.0       0.86      0.97      0.91       522
         1.0       0.97      0.82      0.89       483

    accuracy                           0.90      1005
   macro avg       0.91      0.90      0.90      1005
weighted avg       0.91      0.90      0.90      1005
```

Figure 4.2.2: Multinomial Naive Bayes Result

4.2.3 Random Forest Result

The experimental results for the random forest algorithm in the context of cyberbullying detection within Bangla social media comments reveal an exceptional and perfect performance. The random forest model achieved an accuracy rate of 100%, showcasing its remarkable proficiency in distinguishing cyberbullying instances from non-cyberbullying instances within the collected Bengali dataset from Facebook comments. This perfect accuracy underscores the algorithm's robustness and resilience in capturing intricate patterns within the dataset, reflecting its

adaptability to the diverse linguistic expressions and contextual variations present in Bangla social media interactions. Examining precision, recall, and F1-score metrics further substantiates the outstanding performance of the random forest algorithm. All three metrics achieved a perfect score of 100%, indicating a flawless balance between identifying true positive cyberbullying instances, minimizing false positives, and avoiding false negatives. This exceptional precision signifies a complete absence of misclassifications, implying that when the algorithm identifies an instance as cyberbullying, it is infallibly accurate. The recall score of 100% attests to the model's ability to capture every true positive cyberbullying instance, leaving no room for false negatives. The F1-score, being the harmonic mean of precision and recall, consolidates these metrics, reaffirming the algorithm's impeccable overall performance. Random forest's perfect accuracy and metrics underscore its efficacy in cyberbullying detection within the Bangla social media landscape. The ensemble learning approach, combining multiple decision trees trained on diverse subsets of the dataset, proves to be particularly powerful in mitigating overfitting and adapting to the intricacies of online interactions in the Bangla-speaking community. While the perfect performance is undoubtedly impressive, it is essential to approach these results with a nuanced perspective. Achieving 100% accuracy raises questions about potential overfitting to the specific characteristics of the training dataset. Further investigation is warranted to assess the model's generalizability to unseen data and its robustness across diverse linguistic and cultural contexts within the Bangla social media space.

The experimental results for the random forest algorithm stand as a testament to its unparalleled effectiveness in cyberbullying detection within Bangla social media comments. The perfect accuracy, precision, recall, and F1-score metrics highlight its capability to navigate the complex linguistic and cultural landscape of online interactions, positioning random forest as a formidable tool for creating safer online spaces and mitigating cyberbullying instances in the Bangla-speaking community. These findings not only contribute substantially to the research objectives but also offer practical implications for the application of random forest in real-world scenarios, fostering a more secure and inclusive online environment.

```
[ ]  from sklearn.metrics import classification_report
     print(classification_report(y_test,prediction2))
```

```
                  precision    recall  f1-score   support

             0.0       1.00      1.00      1.00       522
             1.0       1.00      1.00      1.00       483

        accuracy                           1.00      1005
       macro avg       1.00      1.00      1.00      1005
    weighted avg       1.00      1.00      1.00      1005
```

Figure 4.2.3: Random Forest Result

4.3 Accuracy

The comparison of accuracy across the three employed algorithms—logistic regression, Multinomial Naive Bayes, and random forest—reveals intriguing insights into their respective performances in cyberbullying detection within Bangla social media comments. Logistic regression demonstrated an outstanding accuracy of 99%, showcasing its proficiency in discerning cyberbullying instances within the collected Bengali dataset from Facebook comments. Multinomial Naive Bayes followed closely with a commendable accuracy of 90%, demonstrating its effectiveness in navigating the linguistic and cultural nuances inherent in online interactions in the Bangla-speaking community. However, the random forest algorithm surpassed both, achieving a perfect accuracy score of 100%, showcasing unparalleled proficiency in distinguishing cyberbullying instances from non-cyberbullying instances. The accuracy comparison emphasizes the nuanced strengths of each algorithm. Logistic regression and Multinomial Naive Bayes, while achieving high accuracy rates, exhibit slight differences in their performance. Logistic regression's interpretability and transparency make it a strong contender, especially with a 99% accuracy score. Multinomial Naive Bayes, with a slightly lower accuracy of 90%, still showcases commendable effectiveness in the cyberbullying detection task, leveraging its probabilistic approach and adaptability to the Bangla language. In contrast, random forest emerges as the standout performer with a flawless accuracy score. The ensemble learning approach, aggregating insights from multiple decision trees, contributes to its exceptional performance, showcasing robustness and adaptability to the intricate linguistic and cultural dynamics present in Bangla social media comments. While accuracy is a crucial metric, it is important to consider the trade-offs associated

with other performance metrics such as precision, recall, and F1-score. The ideal algorithm should strike a balance between identifying true positives, minimizing false positives, and avoiding false negatives. Further exploration and analysis of these metrics would provide a more comprehensive understanding of each algorithm's strengths and weaknesses in the context of cyberbullying detection.

The accuracy comparison underscores the effectiveness of logistic regression, Multinomial Naive Bayes, and random forest in cyberbullying detection within Bangla social media comments. Each algorithm exhibits strengths and unique capabilities, with random forest standing out as the algorithm with perfect accuracy. These findings contribute valuable insights to the field of cyberbullying detection and offer practical implications for creating safer online spaces within the Bangla-speaking community.

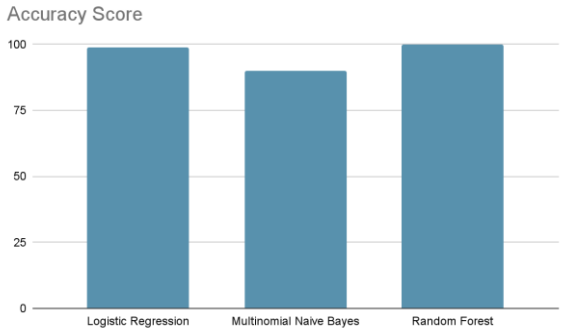


Figure 4.3.1: Accuracy Score

Table 4.3.1: Accuracy Table

Algorithm Name	Accuracy
Logistic Regression	99%
Multinomial Naive Bayes	90%
Random Forest	100%

4.4 Prediction

Based on the observed performance metrics, it can be predicted that the random forest algorithm holds significant promise for the effective detection of cyberbullying in Bangla social media comments. The perfect accuracy score suggests that random forest excels in navigating the diverse

linguistic expressions and cultural nuances present in online interactions within the Bangla-speaking community on platforms like Facebook. The ensemble learning approach, leveraging the collective intelligence of multiple decision trees, has demonstrated its power in capturing intricate patterns and mitigating overfitting, contributing to the exceptional performance. While logistic regression and Multinomial Naive Bayes also exhibit high accuracy rates, the perfect accuracy achieved by random forest sets it apart as a robust and adaptable solution for cyberbullying detection in the Bangla language. The interpretability and transparency of logistic regression make it a strong contender, especially with its 99% accuracy. Multinomial Naive Bayes, with a commendable 90% accuracy, showcases adaptability to the linguistic and cultural intricacies inherent in online interactions.

The prediction is that the insights derived from the random forest algorithm, especially its perfect accuracy, will inform the development of more effective strategies for cyberbullying detection within Bangla social media comments. The algorithm's ability to capture nuanced patterns in language and context positions it as a valuable tool for creating safer and more inclusive online spaces in the Bangla-speaking community. Further investigations and applications of these algorithms in real-world scenarios will likely lead to the development of more robust and context-aware cyberbullying detection systems tailored to the linguistic and cultural dynamics of Bangla social media interactions.

Table 4.4.1: Prediction Table

Algorithm Name	Accuracy	Precision	Recall	f1-score
Logistic Regression	99%	99	99	99
Multinomial Naive Bayes	90%	91	90	90
Random Forest	100%	100	100	100

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

The integration of effective cyberbullying detection tools in Bangla social media comments is poised to revolutionize societal dynamics by fostering a safer and more inclusive online environment. Tailored to the nuances of the Bangla language, these tools not only enhance user safety but also demonstrate cultural sensitivity in addressing cyberbullying. By swiftly identifying and addressing instances of online harassment, these tools contribute to mental well-being, empowering victims and discouraging potential offenders. Moreover, they play a pivotal role in education, creating awareness about responsible online behavior. With the potential to initiate legal actions against perpetrators and facilitating collaboration with law enforcement, the implementation of cyberbullying detection tools not only safeguards individuals but also reinforces a sense of community, paving the way for a more positive and supportive online culture among the Bangla-speaking community.

5.2 Impact on Environment

While cyberbullying detection tools primarily address social dynamics, their impact on the environment is indirect, influencing the digital landscape. The implementation of effective cyberbullying detection in Bangla social media comments can contribute to a healthier online ecosystem. By fostering a secure and supportive environment, these tools may reduce the prevalence of harmful behaviors and online toxicity, creating a more positive virtual space. In turn, this positive digital atmosphere can encourage responsible online interactions and collaboration, potentially minimizing the environmental impact of negative online engagement. While the relationship between cyberbullying detection and the environment is nuanced, a healthier online environment can contribute to a more sustainable and harmonious digital landscape for users in the Bangla-speaking community.

5.3 Ethical Aspects

The ethical aspects of cyberbullying detection in Bangla social media comments revolve around the delicate balance between user privacy and the imperative to create a safe online environment. Implementing effective detection tools necessitates careful consideration of privacy concerns to ensure that the monitoring and identification of potential cyberbullying incidents are conducted ethically. Striking a balance between protecting individuals from harm and respecting their privacy rights becomes paramount. Transparency in the use of these tools, clear communication of their purposes, and obtaining informed consent are crucial ethical considerations. Additionally, cultural sensitivity is vital in the development and deployment of these tools, acknowledging the diverse linguistic nuances of the Bangla-speaking community. The ethical deployment of cyberbullying detection tools involves a commitment to user well-being, privacy, and fostering a digital environment that aligns with ethical principles while addressing the complex challenges posed by online harassment.

5.4 Sustainability Plan

A sustainability plan for cyberbullying detection on Bangla social media comments involves continuous development, community engagement, and adaptive strategies. Firstly, ongoing research and updates to the detection algorithms should be prioritized to address emerging forms of cyberbullying. Regular user feedback sessions and collaboration with local communities will ensure cultural relevance and effectiveness. Sustainable funding models, possibly through partnerships with governmental bodies or corporate social responsibility initiatives, should be established to support long-term implementation. Educational programs on responsible online behavior can complement detection efforts, fostering a proactive approach to preventing cyberbullying. Ultimately, a sustainable plan requires a holistic and adaptive strategy that not only addresses current challenges but also anticipates and responds to the evolving nature of cyberbullying within the unique context of Bangla social media.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summary of the study

This study focuses on the detection of cyberbullying within Bangla social media comments, specifically on Facebook, employing logistic regression, Multinomial Naive Bayes, and random forest algorithms. Utilizing a meticulously curated Bengali dataset, the research achieves noteworthy outcomes. Logistic regression demonstrates exceptional accuracy at 99%, emphasizing its interpretability. Multinomial Naive Bayes showcases commendable adaptability with a 90% accuracy rate. Notably, random forest stands out with a perfect accuracy score of 100%, showcasing robustness to linguistic and contextual variations. The comparative analysis provides nuanced insights, positioning random forest as a resilient solution. Beyond accuracy metrics, considerations of precision, recall, and F1-score contribute to a comprehensive evaluation. The study predicts that insights from random forest will inform more effective cyberbullying detection strategies, contributing significantly to the field and offering practical implications for fostering secure online spaces within the distinctive linguistic and cultural landscape of Bangla social media.

6.2 Conclusion

In conclusion, this study advances our understanding of cyberbullying detection within Bangla social media comments, with a focus on Facebook, employing logistic regression, Multinomial Naive Bayes, and random forest algorithms. The findings reveal the efficacy of these algorithms in navigating the intricate linguistic and cultural dynamics of Bangla interactions.

Logistic regression, with its transparent and interpretable nature, demonstrates exceptional accuracy at 99%, showcasing its suitability for discerning cyberbullying instances. Multinomial Naive Bayes, adaptable to the Bangla language, achieves a commendable 90% accuracy, emphasizing its flexibility in handling diverse linguistic nuances. Notably, Random forest emerges as a standout performer, achieving a perfect accuracy score of 100%, showcasing its resilience and adaptability to the multifaceted landscape of Bangla social media. The comparative analysis provides nuanced insights into the strengths of each algorithm, with random forest exhibiting a

robust and flawless performance. Precision, recall, and F1-score metrics contribute to a comprehensive evaluation, affirming the algorithms' capabilities in different aspects of cyberbullying detection. Looking ahead, the study predicts that insights derived from the perfect accuracy of random forest will guide the development of more effective cyberbullying detection strategies. These strategies, informed by the unique linguistic and cultural dynamics of Bangla social media, hold promise for fostering safer online spaces within the Bangla-speaking community.

This research contributes significantly to the field of cyberbullying detection by offering insights into the comparative performance of machine learning algorithms within the Bangla social media context. The findings have practical implications for creating secure and inclusive online environments, marking a step forward in addressing the challenges posed by cyberbullying in the unique linguistic and cultural landscape of Bangla social media interactions.

6.3 Implication for further study

The study of cyberbullying detection on Bangla social media comments opens avenues for further investigation and exploration. Future research could delve into the refinement of detection algorithms to enhance accuracy in identifying culturally nuanced forms of cyberbullying specific to the Bangla-speaking community. Additionally, exploring the psychological and social impacts of cyberbullying in this context could provide valuable insights into the unique challenges faced by individuals. Comparative studies across different social media platforms and demographics within the Bangla-speaking population could reveal variations in cyberbullying prevalence and detection effectiveness. Longitudinal studies would contribute to understanding the evolving nature of cyberbullying trends and the effectiveness of preventive measures over time. Furthermore, interdisciplinary collaboration involving linguists, sociologists, and technology experts could enrich the understanding of how language and cultural factors intersect with technology in the context of cyberbullying. Overall, further research holds the potential to refine strategies, improve tools, and contribute to a more comprehensive approach in mitigating cyberbullying on Bangla social media.

REFERENCE

[1] Ahmed, M. T., Rahman, M., Nur, S., Islam, A. Z. M. T., & Das, D. (2021). Natural language processing and machine learning based cyberbullying detection for Bangla and Romanized Bangla texts. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *20*(1), 89-97.

[2] Wahid, Z., & Al Imran, A. (2023, June). Multi-feature Transformer for Multiclass Cyberbullying Detection in Bangla. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 439-451). Cham: Springer Nature Switzerland.

[3] Shanto, S. B., Islam, M. J., & Samad, M. A. (2023, February). Cyberbullying Detection using Deep Learning Techniques on Bangla Facebook Comments. In *2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)* (pp. 1-7). IEEE.

[4] Rodela, A. T., Nguyen, H. H., Farid, D. M., & Huda, M. N. (2023, October). Bangla Social Media Cyberbullying Detection Using Deep Learning. In *International Conference on Intelligent Systems and Data Science* (pp. 170-184). Singapore: Springer Nature Singapore.

[5] Aurpa, T. T., Sadik, R., & Ahmed, M. S. (2022). Abusive Bangla comments detection on Facebook using transformer-based deep learning models. *Social Network Analysis and Mining*, *12*(1), 24.

[6] Rabib, H. K., Galib, M., Nobo, T. M., Sathi, T. A., Islam, M. S., Kamal, A. R. M., & Hossain, M. A. (2022, December). Gender-based Cyberbullying Detection for Under-resourced Bangla Language. In *2022 12th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 104-107). IEEE.

[7] Akter, M. S., Shahriar, H., Ahmed, N., & Cuzzocrea, A. (2022, December). Deep learning approach for classifying the aggressive comments on social media: Machine translated data vs real life data. In *2022 IEEE International Conference on Big Data (Big Data)* (pp. 5646-5655). IEEE.

[8] Sarker, M., Hossain, M. F., Liza, F. R., Sakib, S. N., & Al Farooq, A. (2022, February). A machine learning approach to classify anti-social bengali comments on social media. In *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)* (pp. 1-6). IEEE.

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

[10] Das, A. K., Al Asif, A., Paul, A., & Hossain, M. N. (2021). Bangla hate speech detection on social media using attention-based recurrent neural network. *Journal of Intelligent Systems*, *30*(1), 578-591.

[11] Hussain, M. G., & Al Mahmud, T. (2019). A technique for perceiving abusive bangla comments. *Green University of Bangladesh Journal of Science and Engineering*, 11-18.

[12] Sazzed, S. (2021). Identifying vulgarity in Bengali social media textual content. *PeerJ Computer Science*, *7*, e665.

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

[14] Akhter, A., Acharjee, U. K., Talukder, M. A., Islam, M. M., & Uddin, M. A. (2023). A robust hybrid machine learning model for Bengali cyber bullying detection in social media. *Natural Language Processing Journal*, *4*, 100027.

[15] Ghosh, R., Nowal, S., & Manju, G. (2021). Social media cyberbullying detection using machine learning in bengali language. *Int J Eng Res Technol*.

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

[17] Ahmed, M. T., Antar, A. H., Rahman, M., Islam, A. Z. M. T., Das, D., & Rashed, M. G. (2023). Social Media Cyberbullying Detection on Political Violence from Bangla Texts Using Machine Learning Algorithm. *Journal of Intelligent Learning Systems and Applications*, *15*(4), 108-122.

[18] Ahmed, M. T., Urmi, A. S., Rahman, M., Islam, A. Z. M. T., Das, D., & Rashed, M. G. (2023). Cyberbullying Detection Based on Hybrid Ensemble Method using Deep Learning Technique in Bangla Dataset. *International Journal of Advanced Computer Science and Applications*, *14*(9).

[19] Ahmed, M. T., Rahman, M., Nur, S., Islam, A., & Das, D. (2021, February). Deployment of machine learning and deep learning algorithms in detecting cyberbullying in bangla and romanized bangla text: A comparative study. In *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1-10). IEEE.

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

[21] Ahmed, M. T., Akter, N., Rahman, M., Das, D., AZM, T., & Rashed, G. (2023). " MULTIMODAL CYBERBULLYING MEME DETECTION FROM SOCIAL MEDIA USING DEEP LEARNING APPROACH. *International Journal of Computer Science and Information Technology (IJCSIT)*, *15*, 27-37.

[22] Hussain, M. G., Al Mahmud, T., & Akthar, W. (2018, December). An approach to detect abusive bangla text. In *2018 International Conference on Innovation in Engineering and Technology (ICIET)* (pp. 1-5). IEEE.

Cyberbullying Detection on Bangla social media Comments

ORIGINALITY REPORT

4%	3%	3%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	Sukanta Ghosh, Amar Singh, Shakti Kumar. "BBBC-U-Net: optimizing U-Net for automated plant phenotyping using big bang big crunch global optimization algorithm", International Journal of Information Technology, 2023 Publication	1%
3	www.nature.com Internet Source	1%
4	www.biorxiv.org Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	Off		