Bachelor of Science in Computer Science & Engineering

Religious Aggression Detection in Social Media from Bangla Textual Data Using Transformers

by

Suhana Binta Rashid

ID: 1804037

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

May, 2024

Religious Aggression Detection in Social Media from Bangla Textual Data Using Transformers

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Suhana Binta Rashid
ID: 1804037

Supervised by

Md. Shafiul Alam Forhad
Assistant Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

The thesis titled ' **Religious Aggression Detection in Social Media from Bangla Textual Data Using Transformers** ' submitted by ID: 1804037, Session 2021-2022 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

______________________________ Chairman

Md. Shafiul Alam Forhad
Assistant Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

______________________________ Member (Ex-Officio)

Dr. Abu Hasnat Mohammad Ashfak Habib
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

______________________________ Member (External)

Animesh Chandra Roy
Assistant Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Signature of the candidate

Date:

Acknowledgements

To begin, I am deeply grateful to the Almighty Allah for giving me the courage and determination to complete this thesis. I'd want to express my heartfelt gratitude to everyone whose supervision and assistance helped me complete this thesis work successfully. I am really grateful and respectful to my supervisor, Md. Shafiul Alam Forhad, Assistant Professor, Department of CSE, CUET, for his permanent leadership, continual support, and supervision throughout the thesis. I am also grateful to my external board member Animesh Chandra Roy, Assistant Professor, Department of CSE, CUET, for his guidance. My heartfelt gratitude goes to Prof. Dr. Abu Hasnat Mohammad Ashfak Habib, Head of the Department of CSE at CUET, for his constantly help and support. I am also very grateful to my parents for being very supportive. Lastly I want to thank my friends for their support throughout every stage of my life and for their encouragement during my periods of hard time.

Abstract

Religion is a broad cultural and social phenomenon that encompasses ideas, practices, rituals, and ethics governing how people conduct their lives. In Bangladesh, where the majority of the population holds strong religious views, these beliefs deeply influence daily life and cultural identity, makes religious discussions particularly sensitive. As social media platforms have become significant spaces for public debate and self-expression, they also operate the platform for sharing various types of opinion including religious. Unfortunately, some of these texts can be aggressive and potentially offensive to the religion followers. This has prompted our research, which seeks to identify and categorize instances of religious aggression in Bangla text. We divided into four classes so that we could carefully look over and look into the different features of texts that are religiously aggressively. This can help us to improve our knowledge and ability to manage possible conflicts arising by such content. We developed a dataset that includes 3,051 Bangla Text, carefully sorted into four specific categories: Hate Speech, Vandalism, Life Threatening, and Non-Aggressive. The data are collected from different social platforms including Facebook posts and comments, Youtube comments and some online news portal. We used different types of text classifier models to analyze this dataset. These included traditional machine learning algorithms like Multinomial Naive Bayes(MNB), Gradient Boosting (GB), and Random Forest (RF), as well as more advanced deep learning methods such as Long short-term memory(LSTM) and Recurrent Neural Network (RNN). We also tested transformer-based models such as BanglaBERT, m-BERT, and XLM-roBERTa. Among all models tested, transformer performed well on this dataset specially XLM-roBERTa achieved the highest accuracy of 72%. .

Keywords: BanglaBERT, m-BERT, XLM-RoBERTa, Sentiment Analysis, Aggression

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

In the modern era of technology, social media platforms such as Facebook, YouTube, Twitter, instagram , online news portals etc have become as key places for public conversation and expression. These platforms help people share ideas and also act as spaces where cultural identities and social rules are challenged and shaped. In Bangladesh, a country where most people have strong religious beliefs that are a big part of their everyday life and culture. Discussions about religion are sensitive.

1.1.1 Religious Composition of Bangladesh: A Demographic Overview

Religion has a major influence on Bangladesh and major of them follows Islam. Bangladesh is a multireligious country that is home to Muslims as well as Hindu, Buddhist, and Christian minorities. In Bangladesh, according to Wikipedia[1], the religious demographic is predominantly Muslim, with 91.04% of the population adhering to Islam. Hindus make up the second largest religious group, constituting 7.95% of the populace. Buddhists and Christians represent smaller fractions, at 0.65% and 0.30% respectively, while other religious groups account for a mere 0.22%. The information is visualized in the pie chart 1.1. This diverse religious composition reflects the multifaceted cultural and spiritual landscape of the nation.

[1] https://en.wikipedia.org/wiki/Religion_in_Bangladesh

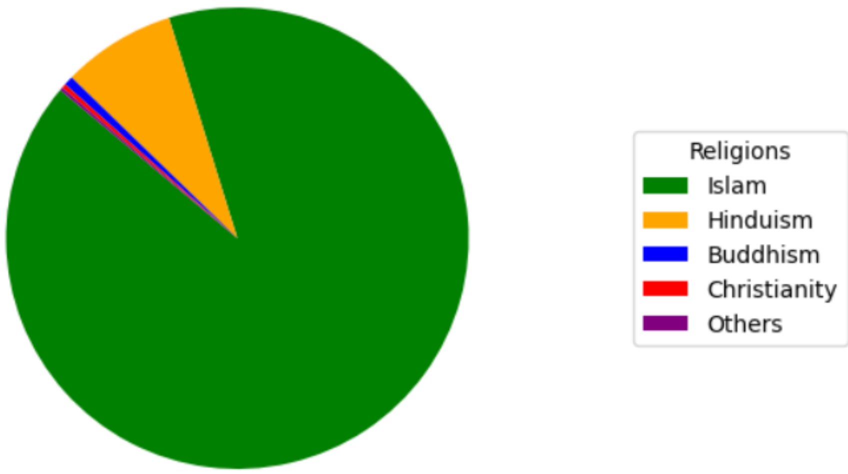

Figure 1.1: Religious Composition of Bangladesh [1]

Bangladeshis often exhibit a deep sense of commitment to and appreciation for religious authorities, houses of worship, and holy books. Mosques and other places of worship serve as significant hubs for religious communities, creating a sense of unity and cooperation among believers. Deep involvement with religion often leads to strong expressions that can become aggressive. These expressions can be harsh hate speech, threats based on religion, acts of vandalism connected to religious feelings, or normal texts. These interactions often show deeper social issues and can lead to real conflicts. This phenomenon can lead to significant disruptions in society, harming individuals and communities. A visual representation of cause of war is shown in Fig.1.2 and the information was taken from the a reliable source [2].

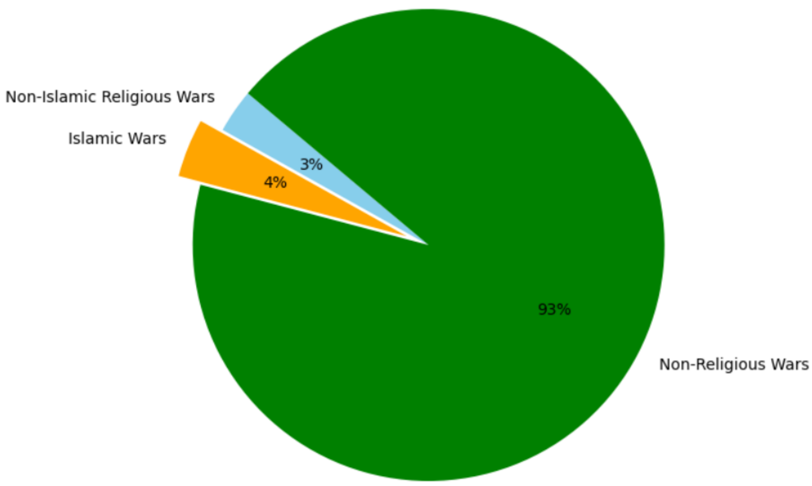

Figure 1.2: Cause of War [2]

[2]https://carm.org/atheism/the-myth-that-religion-is-the-1-cause-of-war/

By studying religious aggression, tools and methodologies are aimed to be developed that can help in identifying and mitigating such content before it escalates into real-world violence. This research is not only academically stimulating but also crucial for the betterment of society, offering a meaningful way for technology to be used to foster a safer and more respectful online environment.

Every day, massive amounts of data are generated on social media networks, making it extremely impossible to monitor manually. Therefore, this thesis introduces the use of advanced technology, including machine learning, deep learning, and specifically transformer-based models, to automatically identify and categorize types of religious aggression in Bangla text. This work makes use of a dataset of 3,051 items gathered from multiple sources, including Facebook posts and comments, YouTube comments, and online news portals. Each piece of data has been carefully divided into four categories: religious hate speech, religious vandalism, religious threats to life, and non-aggressive content. This method helps in understanding and managing the spread of religious aggression through these platforms.

The main goal of enhancing the methods for identifying and understanding religious aggression online is to make social media environments safer. By recognizing and addressing aggressive content swiftly and accurately, these social media platforms can become more secure. It also minimizes the possibility of conflicts that start online and then develop into real-world violence. This is vital for maintaining peace and order within communities, as online conflicts can quickly lead to real consequences, disrupting lives and disturbing social harmony.

1.2 Difficulties

Throughout the entire procedure, I experienced numerous challenges in achieving accurate results. These are explained below.

1. **Data Acquisition Challenge**

   Collecting religious-related data in Bangla language comes with some challenges. Gathering high-quality data that accurately reflects various religious perspectives has been difficult.

2. **Annotator bias:**

   Annotator bias is a significant issue in case of manually labeling data for models. Different annotator interpreted the data in their own ways, leading to inconsistent labels.

3. **Algorithm Selection:**

   I encountered difficulties in ensuring that the chosen algorithms not only performed well in specific contexts but also maintained its accuracy across various scenarios.

4. **RAM Crashed Problem:**

   When applying transformer models to my dataset, I encountered an issue where the system's RAM (Random Access Memory) crashed and it stopped working.

1.3 Applications

1. **Public Sentiment Tracking:**

   Utilizing sentiment analysis to continuously monitor and evaluate public sentiment in news articles and other social media platform about religious conflicts.

2. **Preventative Journalism:**

   By identifying aggressive texts before publication, news organizations can prevent the spread of harmful narratives that may create violence, promoting more responsible journalism.

3. **Policy Development and Adjustment:**

   This analysis can inform the development of new policies or the adjustment of existing ones to better handle the underlying causes of religious conflicts and promote societal harmony.

1.4 Motivation

Our project focuses on identifying signs of religious aggression on social media, a critical initiative to maintain peace and safety in Bangladesh. This is particularly important in a country like ours, where people of various religions live together in harmony. As Bangladesh works to meet its development goals, it is essential to manage and prevent the violence that can arise from religious disagreements. We are developing a tool to detect and categorize aggressive religious messages. This tool will aid police and policymakers in stopping conflicts before they explode.By doing so, we help to protect individuals from hate-motivated violence while also protecting public and private property. Ultimately, our main goal is to foster a safer and more inclusive online environment, which is crucial for allowing our diverse and harmonious society to thrive and remain stable.

1.5 Contribution of the thesis

Thesis and research studies are carried out to achieve a specific set of objectives, either by reducing the limitations of the existing works or by introducing new methodologies. In this thesis, the prime objective is to focus on precise authentication of a banknote. The primary contributions of the thesis are:

1. **Development of a Specialized Dataset:**
   We have compiled a dataset of 3051 Bangla texts categorized into Hate-Speech, Vandalism, Life Threatening, and Non-Aggression, to analyze sentiments related to religious matters.

2. **Analysis of Religious Sentiments:**
   Our primary contribution is the in-depth analysis of religious sentiments conveyed through social media.

3. **Development of a System:**
   We have developed a methodology that can identify the aggressiveness of Bangla text and categorize it into four distinct classes.

1.6 Thesis Organization

In the remaining sections of this paper, Chapter 2 offers a concise overview of previous research on this topic, highlighting their contributions and identifying some of their limitations. Chapter 3 details the proposed methodology, providing an overview of the framework and its implementation. Chapter 4 presents the findings and results, including a discussion on the dataset, performance analysis, and evaluation of the proposed method. Finally, Chapter 5 provides a summary of the entire thesis and offers future recommendations for this research.

1.7 Conclusion

This chapter presents a full explanation of our experiment, which focuses on detecting religious violence. It describes our reason for doing this research, highlighting the importance of better understanding and reducing religiously sensitive information on social media. In this chapter it is also explained how our research could be used to help law enforcement and community leaders solve problems before they occur. It also discusses the problems experienced during the project as well as the key contributions we made.

Chapter 2

Literature Review

2.1 Introduction:

Sentiment analysis has extensively explored fields such as mental health, political opinions, and emotional recognition. Various techniques have been created to assess sentiments in a broad range of subjects. Analyzing religious sentiment is a valuable method for understanding emotions and perspectives regarding religious topics. Nonetheless, there is a noticeable gap in research specifically targeting sentiment analysis of religious content, especially in Bangla language texts. However, a wealth of publications on sentiment analysis, covering both multiclasses and binary classes, is available. In this chapter, we will provide a summary of related works that have made significant contributions to the field of sentiment analysis.

2.2 Related Literature Review

In the research cited as [1], approximately 7,000 Bengali text documents are analyzed. This study categorizes the content into two groups: non-suspect and suspicious. Various machine learning classifiers with different features are applied to this corpus, achieving a maximum accuracy of 84.57%. However, a limitation of this study is its focus on only two classes.

Developing a sophisticated emotion analysis model in low-resource languages like Tamil poses significant challenges. According to the study [2], a method for categorizing emotions in Tamil text is introduced, which classifies a dataset of 22,200 entries into 11 categories: ambiguous, anger, anticipation, disgust, fear, joy, love, neutral, sorrow, surprise, and trust. This research evaluates several models, including transformer-based models (Multilingual-BERT, XLM-R), deep learning models (CNN, LSTM, BiLSTM), and machine learning models (LR, DT, MNB, SVM). The findings reveal that the XLM-R model surpasses the performance of all other models, achieving the highest macro F1-score of 0.33.

17,155 tweets has been used in the study [3] about e-learning that really became popular during the time of COVID-19. The performance of TextBlob, VADER, and SentiWordNet findings is compared to that of CNN (Convolutional Neural Network), LSTM (Long Short Term Memory), CNN-LSTM, and Bi-LSTM (Bidirectional-LSTM), as well as the classification results of deep learning and machine learning models. Additionally, subject modeling is used to identify the issues with e-learning. The results show that the top three issues are the unpredictability of campus opening dates, children's learning difficulties, and lacking effective networks for online education. With BoW (Bag of Words) characteristics, it gets the greatest accuracy of 0.95.

A proposed dataset includes 9312 Urdu reviews that have been carefully classified into three categories: positive, negative, and neutral by human experts in [4]. A manually annotated dataset is created for urdu sentiment analysis as well as applying rule-based, machine learning (SVM, NB, Adabbost, MLP, LR, and RF) and deep learning (CNN-1D, LSTM, Bi-LSTM, GRU, and Bi-GRU) to establish baseline outcomes. Word n-grams, char n-grams, and fast-trained pre-trained are four text representations.The classifers are trained using BERT word embeddings and text. The Multilingual BERT (mBERT) for sentiment analysis in Urdu has also been improved. Finally It achieved an 81.49% F1 rating.

The study [5] analyzed tweets from 179 individuals experiencing depression, with each person contributing between 200 to 3200 recent tweets. From this data, the 100 most common words were identified using Term Frequency-Inverse Document Frequency (TF-IDF) analysis. Additionally, 14 psychological attributes were evaluated using the Linguistic Inquiry and Word Count (LIWC) tool to categorize these words into emotional contexts. The goal was to classify the level of depression into three categories: High, Medium, and Low.

In [6] the author took bangla dataset for predicting if the text is suicidal or non-suicidal.The dataset includes 1700 posts that have been labeled of which 687 have been classified as suicidal (positive) and 1013 as non-suicidal (negative). With a 0.61 F1-Score, a deep learning-based model using CNN+BiLSTM beat Fasttext word embedding techniques.

The most recent research using deep learning to address issues with sentiment analysis, such sentiment polarity, is reviewed in this publication [7]. A number of datasets have been subjected to the application of models based on term frequency-inverse document frequency (TF-IDF) and word embedding. Finally, a comparison analysis of the experimental findings for various models and input attributes has been carried out.

It has been suggested in [8] a deep learning-based method (i.e., BiLSTM) for categorizing the restaurant evaluations submitted by customers into positive and negative polarity. The evaluation's findings on the test dataset demonstrate that the BiLSTM approach generated the best level of accuracy (91.35%).Also the suggested method is compared to existing machine learning algorithms in the analysis.

The author described the emotions that have dominated the popular debate on Twitter in Spain and in Spanish over the past 8 years con-cerning Islam and the Muslim population in the study [9]. This was accomplished by gathering and studying 190,320 messages that contained keywords associated with Muslim culture and religion.

Multiple deep learning and machine learning algorithms were fully evaluated in [10] to determine the most effective strategy for forecasting food quality. Among the approaches studied, logistic regression achieved the highest accuracy of 90.91%. The investigation used a large dataset of over 1,484 online evaluations from well-known food ordering services such as food Panda and HungryNaki. The possibility of developing mobile or web-based applications utilizing pre-trained models like BERT or Transformers was mentioned, which might improve performance even more. Initially, the study concentrated on binary class classifications, with reviews classified as good or negative.

In the study [11], a sentiment analysis was conducted on consumer comments collected from the Facebook pages of four prominent food delivery services: Food Panda, HungryNaki, Pathao Food, and Shohoz Food. A total of 843 positive and 557 negative reviews were analyzed using a range of machine learning (ML) and deep learning (DL) models, including extreme gradient boosting (XGB), random forest classifier (RFC), decision tree classifier (DTC), multinomial Naive Bayes (MNB), convolutional neural network (CNN), long short-term memory (LSTM), and recurrent neural network (RNN). It has been explained that among the machine learning algorithms, the XGB model achieved the highest accuracy with a rate of 89.64%. Notably, the LSTM model surpassed all other models, including the deep learning options, with an outstanding accuracy of 91.07%. This demonstrates the LSTM model's particular effectiveness in accurately predicting sentiment from the consumer reviews on these food delivery platforms.

A specialized dataset for Bangla e-commerce sentiment analysis was curated exclusively from the "Daraz" platform in the study [12]. Sentiment analysis in languages with limited resources like Bangla presents distinct challenges. The dataset compiled for this study consists of approximately 1000 genuine Bangla comments from "Daraz," categorized into binary (positive, negative) and multi-class (very positive, positive, negative, and very negative) categories, providing a valuable resource for developing and testing sentiment analysis models. Techniques such as transfer learning with LSTM and GRU, as well as Transformers like Bangla-BERT, were employed.

According to the survey, cyberbullying in the Bengali language is a big worry that is developing quickly. The study [13] shows aggressiveness exhibited in online bullying emphasizes the critical need for additional research on Bengali to avoid and detect such negative messages. A dataset of 10,512 entries was employed, with each entry classified by remark, class, and gender. The classes were classified as gibe, triggered, sexual, religious, and not bully. Exploratory Data Analysis was conducted using the Class and Gender categories, revealing insights on how bullying changes between genders. The study used a hybrid strategy, which demonstrated to produce superior results than standard methods. The deployment included three Deep Learning classifiers and three Machine Learning classifiers with an accuracy of 85%, .

The study [14], the issue of fake reviews proliferating across various online platforms was addressed, highlighting its significant impact on both consumers and businesses. It has been explained that a publicly available dataset named Bengali Fake Review Detection (BFRD) was developed for this purpose. This dataset comprises 7,710 non-fake and 1,339 fake food-related reviews, all collected from social media posts. To tackle the challenge of identifying fake reviews, multiple deep learning and pre-trained transformer language models were employed. The performance of these models was evaluated, achieving a weighted F1-score of 0.9843,

In that study [15], a Bangla annotated dataset was proposed for sentiment analysis concerning the ongoing Ukraine-Russia war. It was detailed that this dataset comprises 10,861 Bangla comments, categorized into three sentiment classes: Neutral, Pro-Ukraine (Positive), and Pro-Russia (Negative). Hyperparameter optimization was carried out on five different transformer language models, which included BanglaBERT, XLM-RoBERTa-base, XLM-RoBERTa-large, Distil mBERT, and mBERT. It has been explained that the highest accuracy achieved was 86%.

A dataset consisting of three sentiment classes related to Bangladesh Cricket was prepared, reflecting real people's sentiments in the study [16]. The data were preprocessed and features extracted using TF-IDF before applying various machine learning models for classification. It has been explained that the first preference for classification was the Support Vector Machine (SVM), chosen for its robust performance with smaller datasets. The dataset comprised 1,601 entries divided into three classes: praise (513 entries), criticism (604 entries), and sadness (484 entries). The study achieved an accuracy of around 64% in classifying these sentiments, demonstrating the effectiveness of the chosen methods under the constraints of the dataset's size and complexity.

The summary of above publications are shown below with Table 2.3

Table 2.1: Summary of Literature Review

Refe.	Goal	Dataset Type	Used Models	Limitations
[1]	Emotion of conversation about islam and Muslim population in Spain since 2015	Studying 1,90,320 messages	Gensim (used for topic modeling),	all the messages could not be accessed because of geolocation filter
[2]	Classifying whether Bengali Text suspicious or non-suspicious	7,000 Bangli Text	stochastic gradient descent (SGD), logistic regression (LR), decision tree (DT), random forest (RF), and multinomial naïve Bayes (MNB)	Small dataset with only two classess of output
[3]	Represents the textual emotion in Tamil classified into 11 classess	22,200 dataset	SVM, LR, DT, RF, MNB, CNN, LSTM, m-BERT	Small dataset with only two classes of output
[4]	By examining people's attitudes toward e-learning, this study aims to determine if it is successful	About 17,155 tweets about e-learning	Support vector machine (SVM), logistic regression (LR), decision tree (DT), random forest (RF), K-Nearest Neighbors (KNN),Gaussian Naive Bayes (GNB), AdaBoost	
[5]	Manually annotated dataset for Urdu sentiment analysis	9,312 Urdu reviews	Multilingual BERT (mBERT)	

Table 2.2: Summary of Literature Review (continue)

[6]	Analyzing tweets of depressive individuals catagorized into three classess of depression (High, Medium, Low)	1,56,511 tweets	1-DCNN, NN, SVM, RF to train the models	There is not more detailed information about depressive tweets like emojis, pictures, gifs etc
[9]	Whether restaurant reviews are positive or negative	A corpus of 8,435 reviews	BiLSTM approach	There is only two genenrated output
[10]	Whether restaurant reviews are positive or negative	A corpus of 8,435 reviews	BiLSTM approach	There is only two genenrated output
[11]	sentiment polarity analysis of Bangla food reviews	A corpus of 1484 reviews	LR, DTC, RFC, SVC, RNN, STM, GRU	Limited dataset
[12]	E-commerce sentiment analysis in Bangla for binary class classification	A corpus of 1400 reviews	LSTM, GRU and Transformers (Bangla-BERT)	outcome is not satisfied
[13]	Cyberbullying Detection into five class classsification	Total 10512 data	LSTM+BiLSTM	application of Transformer-based model
[14]	Bengali fake reviews detection system	Total 9049 data	BanglaBERT, XLM-RoBERTa-base, XLM-RoBERTa-large, Distil mBERT and mBERT	imbalance dataset

Table 2.3: Summary of Literature Review (continue)

[15]	Sentiment Analysis on Bangla Social Media Comments on Russia–Ukraine War Using Transformers	Total 10861 data	BanglaBERT, XLM-RoBERTa-base, XLM-RoBERTa-large, Distil mBERT and mBERT	imbalance dataset
[16]	Sentiment Analysis on public opinions of Bangladesh Cricket	Total 1601 data	SVM	limited dataset

2.2.1 Conclusion

In this chapter, we reviewed several research works related to sentiment analysis. By going through these related works we came to understand how different machine learning models, deep learning models and transfer based model are working and how the results are different from each other. We made an approach for detecting religious aggression, taking into account the strengths and limitations of previous tools.

Chapter 3

Methodology

3.1 Introduction

To accurately identify and classify religious aggression into four categories—Hate Speech, Vandalism, Life Threatening, and Non-Aggressive—we used a detailed approach that includes machine learning, deep learning, and transformer-based models to ensure high accuracy. We carefully followed a set of clear steps to perform multiclass sentiment analysis. This chapter explains our methods, showing how we used and improved each model to achieve our goals. We discuss the specific methods and strategies that allowed us to effectively use these advanced models. This ensures strong and dependable outcomes in detecting religious aggression.

3.2 Definition

Below, we define the task and provide full explanations of the four categories of religious aggression.

3.2.1 Definition of Task

We have annotated a dataset of 3,051 entries, which are to be categorized into four distinct classes: Hate Speech, Vandalism, Life Threatening, and Non-Aggressive. It has been visualized in the fig 3.1. Our task involves accurately classifying each entry into the appropriate category.

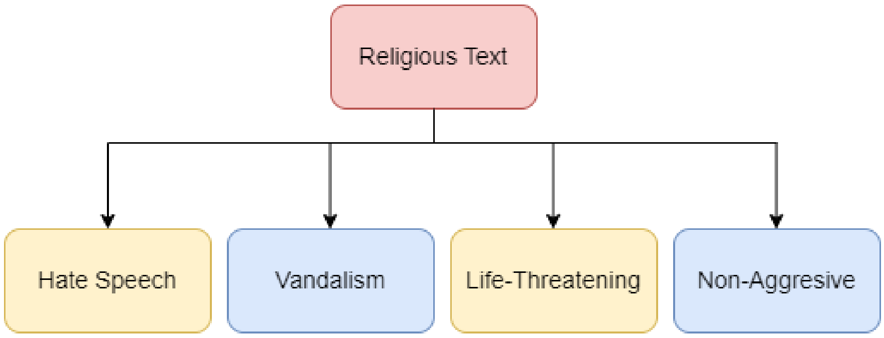

Figure 3.1: Classification of Data

3.2.2 Definition of Classes

The categorised class are explained here:

1. **Hate Speech** Text that sneers or mocks shows a lack of respect for religious views, and this often leads to religiously motivated hate speech. Words like opposing and condemning are more than just disagreeing; they actively put down and dismiss specific religious beliefs or practices or to the person who practices those beliefs.

2. **Vandalism** Texts that describe the demolition of mosques, temples, religious statues, and residences by others are often rooted in the opposition of religious opinion. It also contains religious conflicts and the fire attack.

3. **Life Threatening** Texts that discuss death, suicide, or massacres are categorized as life-threatening, especially when they stem from or are motivated by religious opinions or conflicts. These descriptions highlight the serious risks of religiously influenced conflicts.

4. **Non-aggression** Texts that engage in regular religious discussions without expressing any aggression provide insightful commentary on spiritual beliefs and practices, fostering understanding and dialogue rather than conflict.

3.3 Overview of Overall System Architecture

The diagram 3.2 provides a visual overview of the methodology used for detecting and classifying religious aggression in Bangla text data collected from social media platforms such as Facebook and YouTube. The process begins with data collection, where Bangla text data is gathered and stored in a CSV file. The data then undergoes a series of preprocessing steps including data cleaning, removal of duplicate words, stopword removal, and stemming. This cleaned data is split into training and testing datasets. Then training data is fed into various models for training.

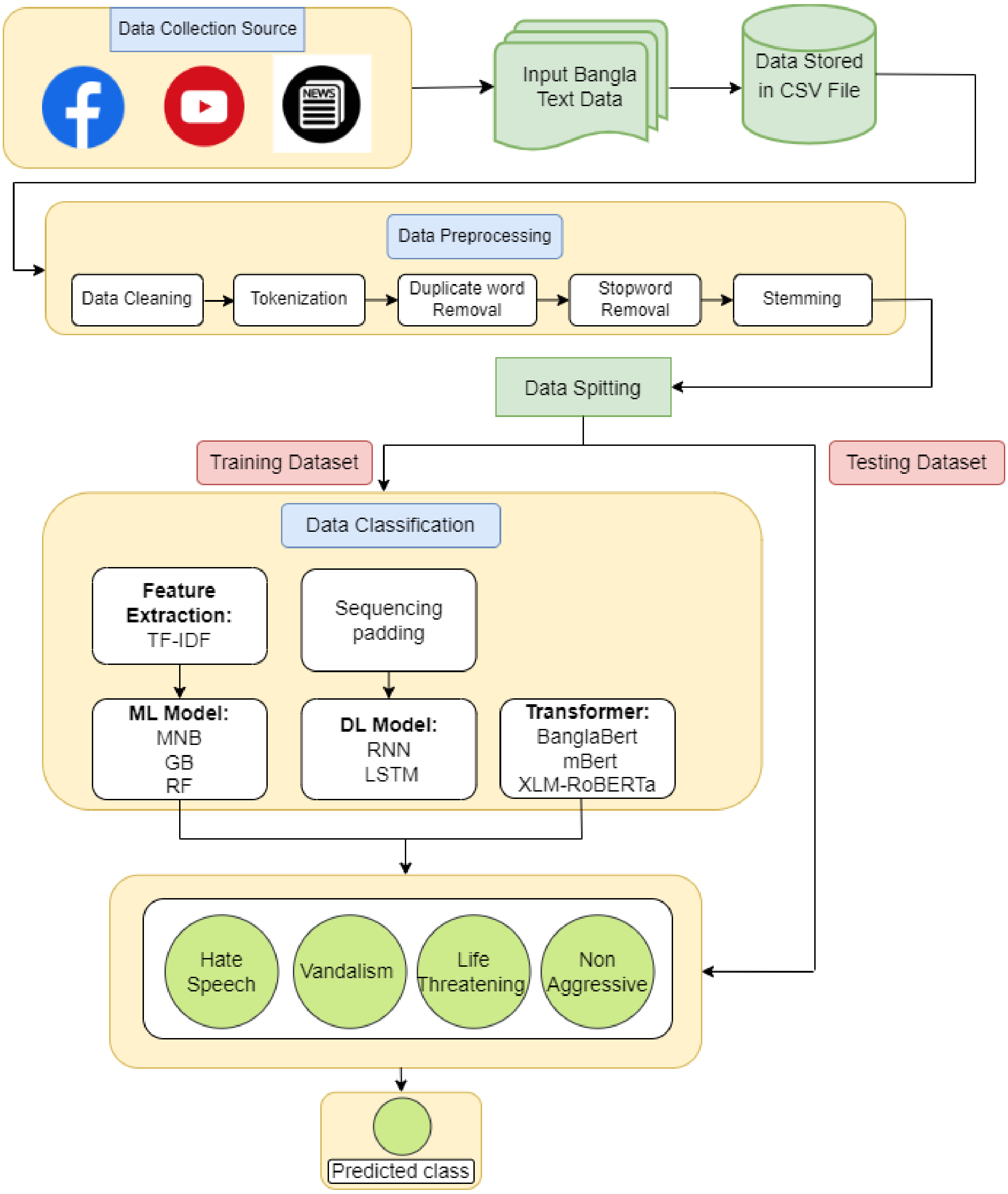

Figure 3.2: Overall System Architecture of Religious Aggression Detection

3.4 Dataset Development

The proposed approach utilizes Bangla text data collected from various social media platforms, including YouTube comments, Facebook posts, Facebook comments, and online newspapers. After gathering the raw data, it undergoes preprocessing to prepare it for further analysis and modeling. Once the annotation process is complete, the dataset is ready for use.

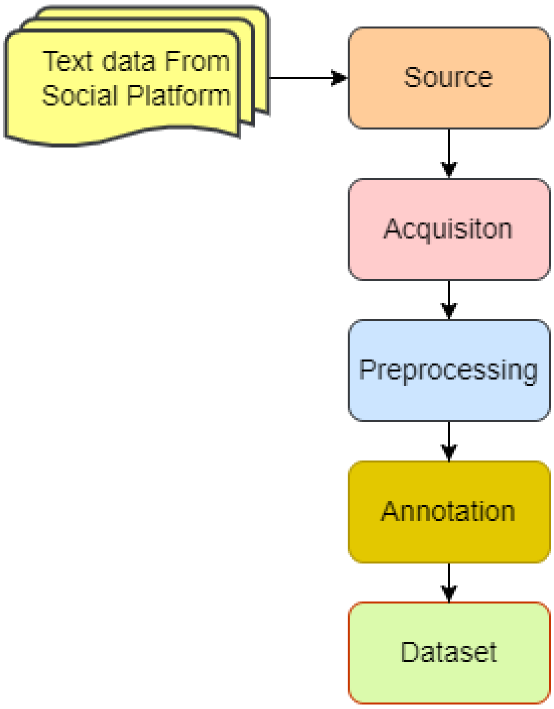

Figure 3.3: Dataset Development Step

3.4.1 Dataset Acquisition

The full dataset is collected from various digital platforms, including online news portals, Facebook, and YouTube . For news portals, major part of the data was collected from BBC Bangla, particularly the headlines. On Facebook, data was taken from religiously controversial posts and their associated comments. Additionally, comments from some religious videos on YouTube also contributed to the dataset acquisition. Data Acquisition amount has been shown in the Table 3.1. It is also visually represented in the Fig 3.4

Table 3.1: Data Acquisitin Summary

Index	Media	Type	Amount
1	Online News Portal	Headline and main News	1665
2	Facebook	Post and Comments	801
3	Youtube	Headline and Comments	585

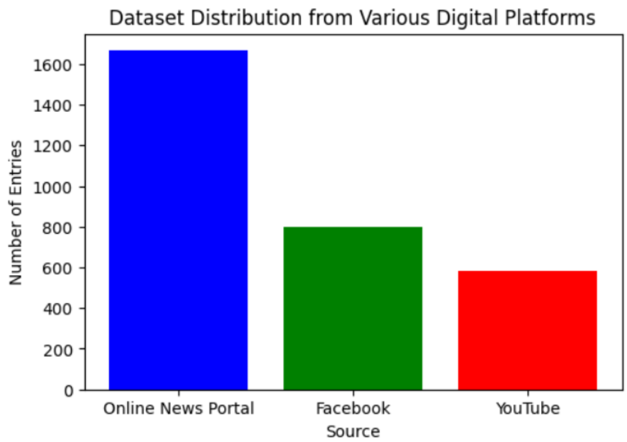

Figure 3.4: Distribution of Dataset Entries from Various Digital Platforms

3.4.2 Data Preprocessing

In the preprocessing phase, the raw data was cleaned by removing null values, followed by the removal of short-length text. Punctuation was then removed, duplicate words were eliminated, and stopwords were filtered out. Finally, stemming was applied to reduce words to their root forms.

3.4.3 Annotation

The importance of an experienced annotator lies in their ability to accurately and consistently label data. This ensures the quality and reliability of the dataset. To annotate our collected dataset we had three annotator. Two annotator labelled the data manually and the expert one solved disagreement between two annotators. The summary about the annotators are shown in Table 3.2

Table 3.2: Summary about the Annotators

Annotator Feature	**Delegation**	**Age**	**Experience**
Annotator 1	Undergraduate Student	22-25	1 year
Annotator 2	Undergraduate Student	22-25	1 year
Expert 1	Assistance Professor	26-30	4 year

3.4.3.1 Annotation Guidelines

Annotation guidelines provide a structured framework for annotators, ensuring consistency and accuracy in labeling data. The annotation guidelines are given in the Fig 3.5

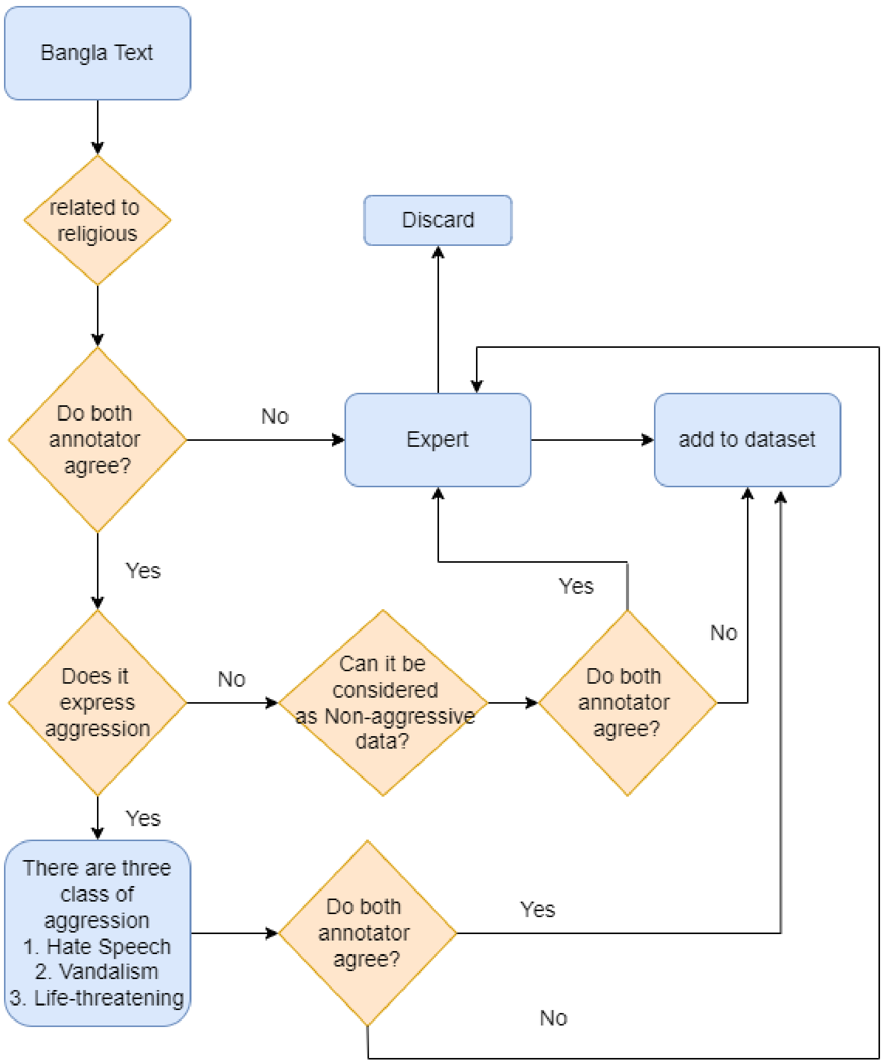

Figure 3.5: Annotation Process

3.4.3.2 Annotation Quality

The quality of the annotation is ensured by Cohen's Kappa value. The equation for Cohen's Kappa is given below:

$$k = \frac{P_a - P_r}{1 - P_r} \tag{3.1}$$

1. P_a is defined as probability of agreement
2. P_r is defined as probability of random agreement

In case of my dataset the kohen's cappa value is 87%

3.4.3.3 Overall Dataset

Our dataset comprises a total of 3,051 data entries. Among these, 1,052 texts are categorized as Non-Aggressive (NoAg), 454 texts are identified as Hate Speech (ReHate), 534 texts are related to Vandalism (ReVan), and 254 texts are about Life Threatening (ReLife) situations.

Text	Label	Classification
আমার মনে হয় আমি সবসময়ই মুসলিম ছিলাম।	NoAg	Non-Aggression
আমরা কখনো হিন্দু-মুসলমান নিয়ে আলোচনা করি না।	NoAg	Non-Aggression
ধর্ম নিয়ে কটাক্ষ করে উষ্কানি মুলক পোস্ট	ReHate	Hate Speech
এরা কি আদৌ মুসলমান??? এদের মধ্যে কি একজনও মুসলমান নাই!???	ReHate	Hate Speech
২০১২ সালের ২ জানুয়ারি নিউইয়র্ক শহরের একটি হিন্দু উপাসনালয়ে অগ্নিসংযোগ করা হয়।	ReVan	Vandalism
সেখানে অবশিষ্ট হিন্দুদের বিরুদ্ধে অকথ্য অত্যাচার করা হয়। মন্দিরগুলি ভেঙে ফেলা হয়। নারীদের অপহরণ ও ধর্ষণ করা হয়।	ReVan	Vandalism
ধর্ম অবমাননার অভিযোগে লালমনিরহাটে এক ব্যক্তিকে পিটিয়ে হত্যা, মৃতদেহে আগুন	ReLfth	Life Threatening
যেভাবে মৃত্যু বরণ করবে কিয়ামত পযন্ত তার সেই আযাব হতে থাকবে !	ReLfth	Life Threatening

Figure 3.6: Example of Dataset

The total dataset is divided into four categories as shown in the pie chart 3.7. Non-Aggressive texts make up 34.1% of the data with 1,041 entries. Hate Speech accounts for 26.5% with 807 entries. Vandalism comprises 21.9% with 667 entries, while Life Threatening texts represent 17.6% with 536 entries. This distribution highlights the varying prevalence of different types of religious aggression within the collected data.

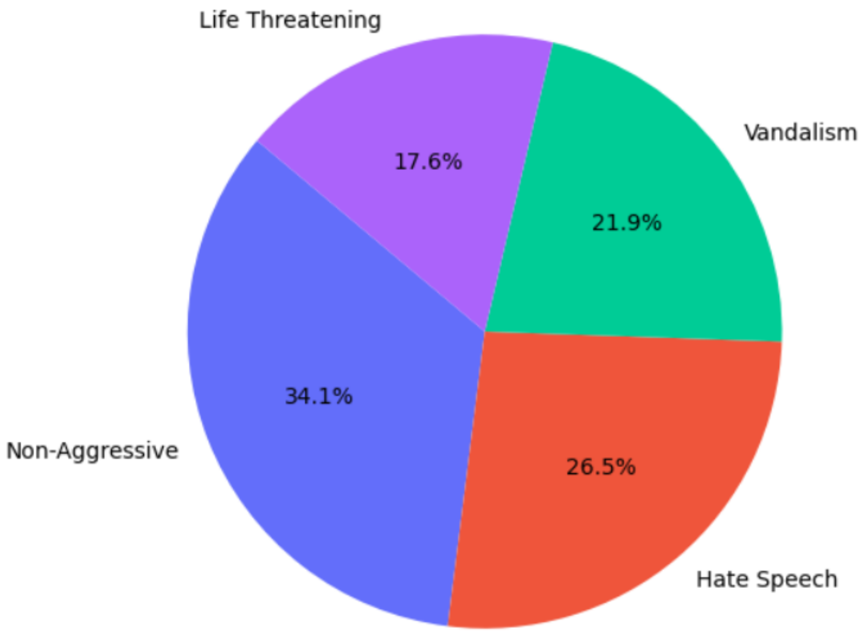

Figure 3.7: Distribution of Dataset by Four Class

3.5 Model Development

We used several models to achieve better accuracy after comparing the performance of these models.

1. Machine learning
   (a) Multinomial Naive Bayes (MNB)
   (b) Gradient Boosting (GB),
   (c) Random Forest (RF)
2. Deep learning
   (a) Recurrent Neural Network (RNN)
   (b) Long short-term memory (LSTM)
3. Transformer
   (a) Bangla-BERT

(b) Multilingual-BERT

(c) XLM-RoBERTa

3.5.1 BanglaBERT

BanglaBERT is a transformer model built on the BERT architecture customized only for the Bangla language. It has been pre-trained on a large dataset of 27.5GB of Bangla text, ensuring a thorough comprehension of the language's linguistic and contextual complexities. Notably, it has the fewest trainable parameters of the models we tested, at around 110 million. This small number of parameters helps to its computational efficiency, allowing it to properly analyze and comprehend Bangla text while keep high performance. Despite its small size, BanglaBERT is an effective tool for different natural language processing applications in the Bangla language.

3.5.2 Multilingual-BERT

The Multilingual BERT model was pre-trained on a massive corpus of 104 languages. This makes it one of the most comprehensive language models that was obtained from Wikipedia. It has a total of 177 million trainable parameters. Multilingual BERT stands out for its ability to handle a wide range of languages. This model is remarkable for its training data sampling strategy, known as exponential smoothing. This technique purposely under-samples high-resource languages like English. This allows for a larger emphasis on low-resource languages like Bangla. As a result, Multilingual BERT is especially well-suited to applications requiring less commonly represented languages. This increases its versatility and efficacy over a broad linguistic range.

3.5.3 XLM-RoBERTa

XLM-RoBERTa is a multilingual version of RoBERTa. The pre-trained dataset includes 100 languages and 2.5TB of filtered CommonCrawl data. It features a total of 278 million parameters. The model has a 12-layer design with 768 hidden states. The authors of XLM-RoBERTa reported that it outperforms mBERT by up to 23% in cross-lingual classification for low-resource languages.

3.6 Conclusion

This section explains the methodology of our work. It also describes the complete development of our dataset. Additionally, it provides information about the models that have been used. In the following chapter, we will discuss the performance of these models and compare them.

Chapter 4

Results and Discussions

4.1 Introduction

We addressed dataset development and model selection in the last chapters. In this chapter, we will discuss and analyze the outcomes of these models. We will also look at the performance metrics. In addition, we will evaluate the models and choose the best one.

4.2 Dataset Description

The dataset comprises 3,051 entries of religious text, meticulously collected from diverse sources such as YouTube, Facebook, and online news portals. This dataset is highly authentic, offering valuable insights into various forms of religious discourse. The entries are classified into four distinct categories, providing a structured approach to analyzing religious content. Notably, the highest frequency of a single word is 1516, with a total unique word count of 16784. Before preprocessing, the sentences contain a maximum of 106 words and a minimum of 4 words. Additionally, only a very small amount of text is interspersed with English words and numbers, ensuring the dataset remains primarily focused on its intended language. The text length of the dataset is shown in Fig 4.1

4.3 Impact Analysis

Impact identification is the process of identifying and expressing a study or initiative's possible benefits and contributions. Listed below are a few of them:

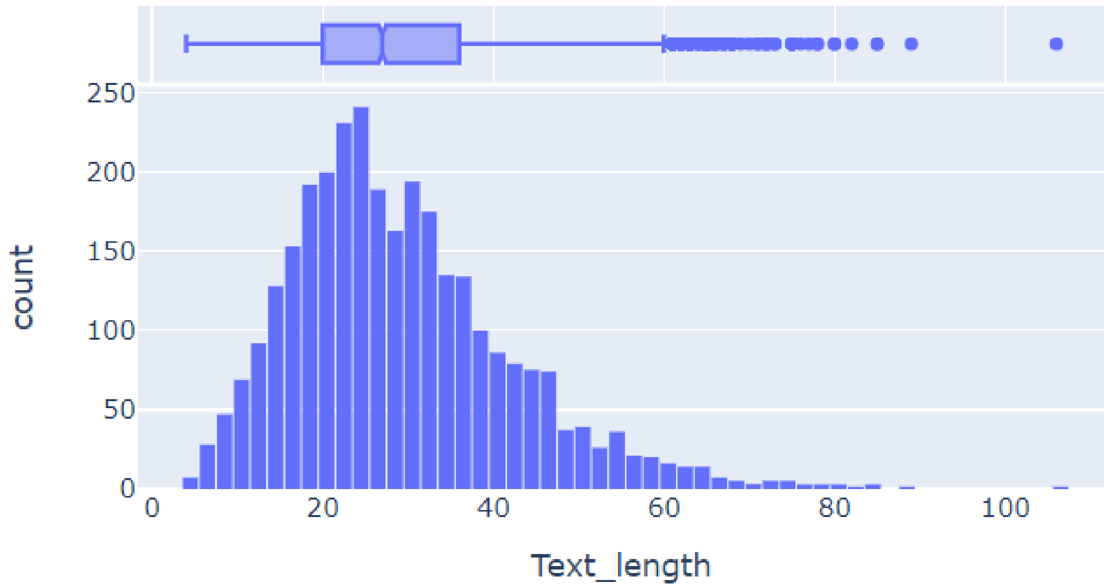

Figure 4.1: Text length of the dataset

4.3.1 Social and Cultural Impact

The presence and effects of religiously aggressive content on social media and other platforms can be brought to light by recognizing such content. Examining religious writings can help people learn the views and beliefs of other religious groups, fostering mutual understanding and appreciation of cultural diversity.

4.3.2 Environmental Impact

Recognizing aggressive religious content can help make online environments safer, encourage constructive discussion, and reduce harassment.

4.3.3 Sustainability Plan

By fostering a more polite and respectful online community, sentiment analysis for religious material can help social media companies remain successful. Providing a welcoming and pleasurable online experience to users can increase their conversion rates, which is good for the long-term sustainability of online platforms.

4.4 Evaluation of Performance

This section will show the overall performance analysis of our work. By utilizing different evaluation matrices, a detailed analysis and explanation regarding performance of our models will be discussed throughout the section.

4.4.1 Performance Evaluation Matrices

We evaluated our proposed system by employing multiple evaluation metrics, such as accuracy, precision, recall, and F1 score. Comprehensive explanations for each of these evaluation metrics are provided herewith.

Confusion Matrix: A confusion matrix is a crucial tool in statistics and machine learning classification for assessing the performance of classification model. It is a table that lets you see how well an algorithm performs by contrasting the actual target values with the values the model predicts. In the Fig 4.2 the visualization of the confusion matrix is given and it was taken from a trustworthy source [1].

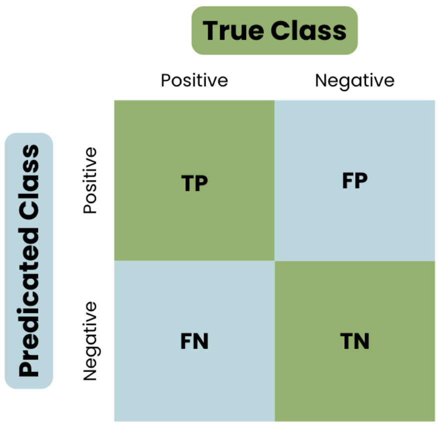

Figure 4.2: Confusion Matrix [1]

An explanation of the main elements of confusion matrix is given below:

- **True Positives (TP):** True positives are the cases where the prediction of model for a positive outcome is correct.
- **True Negatives (TN):** True negative are the cases where the model's prediction for a negative outcome is correct.

[1] https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning

- **False Negative (FN):** False negative are the cases where the model incorrectly predicts a negative outcome, missing a positive instance.
- **False Positive (FP):** False positive are the cases where the model inaccurately predicts a positive outcome, indicating a positive when it is not. These numbers are used to compute other evaluation metrics.

Precision: Precision is the ratio of actual positive predictions to all of the positive predictions of the model as given in equation 4.1

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.1}$$

High precision indicates that the model has a low false positive rate, meaning it is very accurate when it predicts a positive outcome.
Recall: Recall is a significant performance parameter used in machine learning for evaluation of classification models. It is sometimes referred to as sensitivity or true positive rate. It assesses how well a model can locate the relevant instances in a dataset, with a particular emphasis on how well it can find true positive cases. The equation is given in 4.2.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.2}$$

High recall indicates that the model is effective at capturing the positive instances, minimizing the number of false negatives.
F1 Score: The F1 Score is a performance metric used in classification to assess the balance of precision and recall. It is especially useful when dealing with unbalanced datasets in which one class outnumbers the others. The F1 Score is the harmonic mean of precision and recall, yielding a single metric that accounts for both false positives and false negatives. F1 score can be obtained from equation 4.3.

$$F_1Score = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{\text{2TP}}{\text{2TP} + \text{FP} + \text{FN}} \tag{4.3}$$

Accuracy: Accuracy is one of the most basic and widely used metrics for assessing the performance of a classification model. It calculates the proportion of correctly classified instances out of all instances in the dataset. The accuracy can be obtained from equation 4.4.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.4}$$

where TP = True positive; FP = False positive; TN = True negative; FN = False negative

4.4.2 Comparison of Performance

The performance of various models is summarized in the tables below.

Table 4.1 displays the accuracy of machine learning models, while Table 4.2 presents the accuracy of deep learning methods. For transformer-based models, the accuracies are shown in Table 4.3.

Table 4.1: Machine Learning Performance Comparison

Architecture	**Accuracy**	**F1 score**	**Precision**	**Recall**
Multinomial Naive Bayes	0.55	0.53	0.59	0.55
Gradient Boosting	.58	0.56	0.62	0.58
Random Forest	0.55	0.49	0.66	0.55

Table 4.2: Deep Learning Performance Comparison

Architecture	**Accuracy**	**F1 score**	**Precision**	**Recall**
RNN	0.59	0.56	0.59	0.56
LSTM	0.44	0.37	0.39	0.44

Table 4.3: Performance Comparison of Transformer Based Models

Architecture	**Accuracy**	**F1 score**	**Precision**	**Recall**
BanglaBert	0.71	0.71	0.71	0.71
MBert	0.68	0.68	0.68	0.68
XLM-RoBERTa	0.72	0.72	0.72	0.72

Based on the provided analysis, it seems that the performance of transformers, specifically XLM-RoBERTa, is superior to traditional machine learning and deep learning models. Among the transformers, XLM-RoBERTa has achieved the highest accuracy of 72% compared to BanglaBERT and Multilingual-BERT (M-BERT).

4.4.3 Confusion Matrix Diployment

Confusion Matrix of eight different model are presented here.

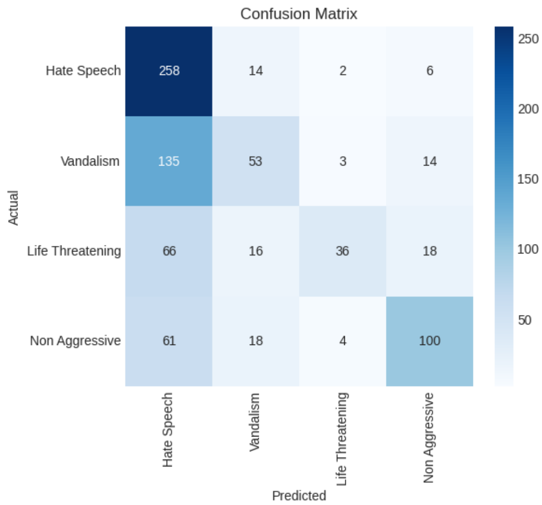

Figure 4.3: Confusion Matrix of Multinomial Naive Bayes

The confusion matrix in Fig 4.3 illustrates the performance of the classification model, showing that 'Hate Speech' is the most accurately predicted category, while 'Life Threatening' and 'Vandalism' have higher misclassification rates.

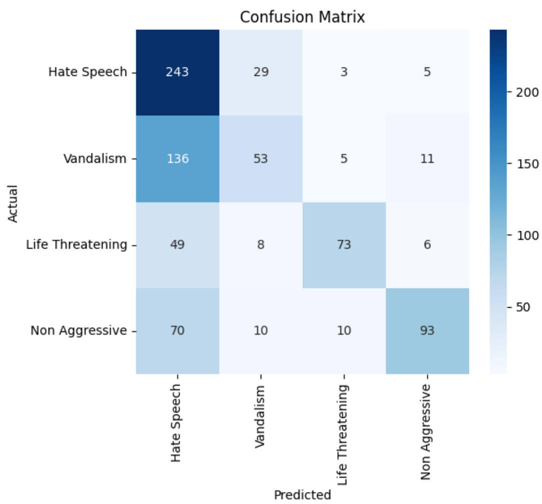

Figure 4.4: Confusion Matrix of Gradient Boostings

The confusion matrix in Fig 4.4 for the Gradient Boosting model shows that 'Hate Speech' is predicted most accurately, while 'Life Threatening' and 'Vandalism' categories have higher misclassification rates.

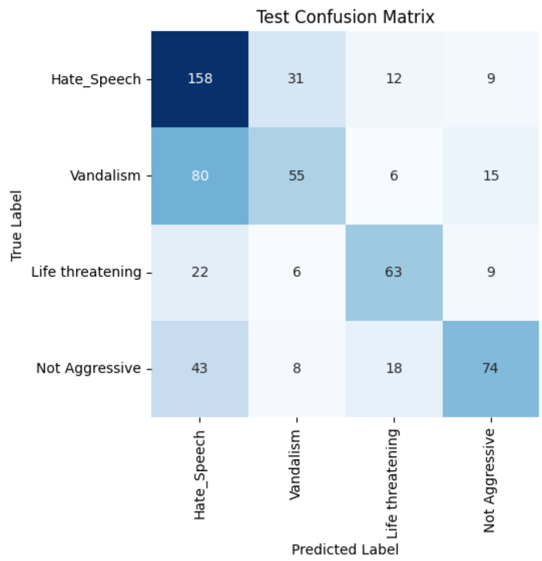

Figure 4.5: Confusion Matrix of Random Forest

The confusion matrix in Fig 4.5 for the Random Forest model shows that 'Hate Speech' is the most accurately predicted category, while 'Vandalism' and 'Life Threatening' have higher misclassification rates.

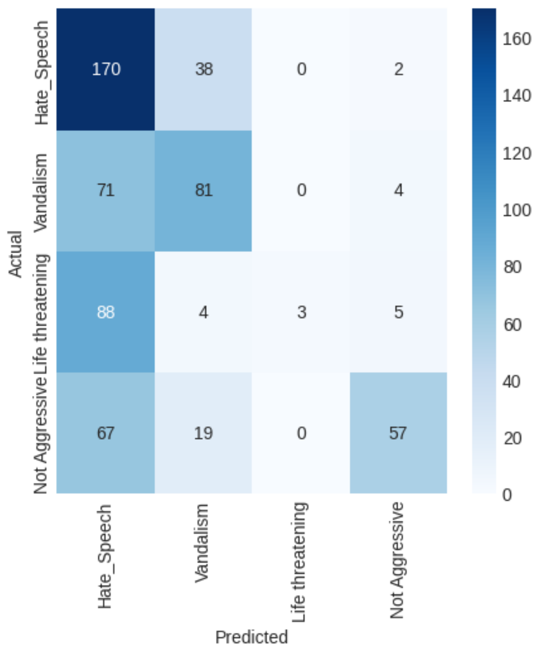

Figure 4.6: Confusion Matrix of Recurrent Neural Network

The confusion matrix in Fig 4.6 for the Recurrent Neural Network (RNN) model shows that 'Hate Speech' and 'Vandalism' are relatively well predicted, while 'Life Threatening' and 'Non-Aggressive' categories have significant misclassification rates.

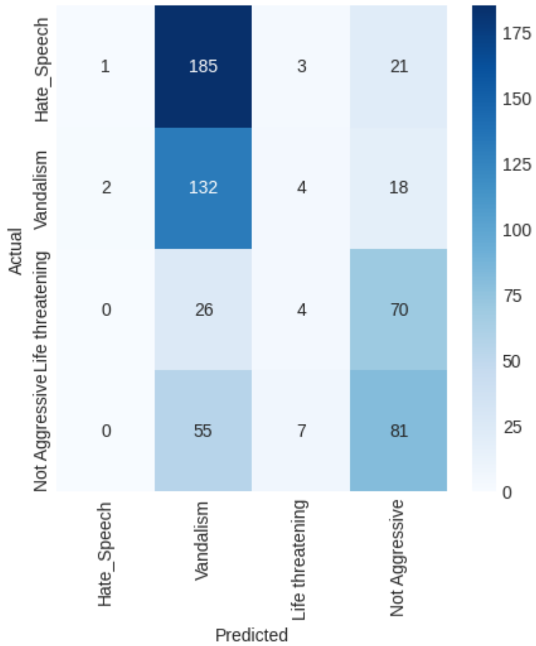

Figure 4.7: Confusion Matrix of Long short-term memory

The confusion matrix in Fig 4.7 for the Long Short-Term Memory (LSTM) model shows that 'Hate Speech' is predicted with high accuracy, while 'Vandalism' and 'Life Threatening' categories have higher misclassification rates.

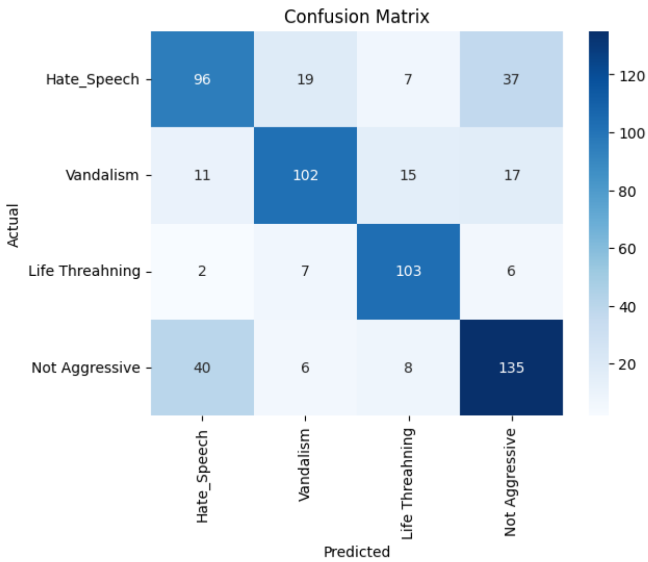

Figure 4.8: Confusion Matrix of BanglaBERT

The confusion matrix in Fig 4.8 for the BanglaBERT model indicates that 'Vandalism' and 'Non-Aggressive' categories are predicted with relatively high accuracy, while 'Hate Speech' and 'Life Threatening' have higher misclassification rates.

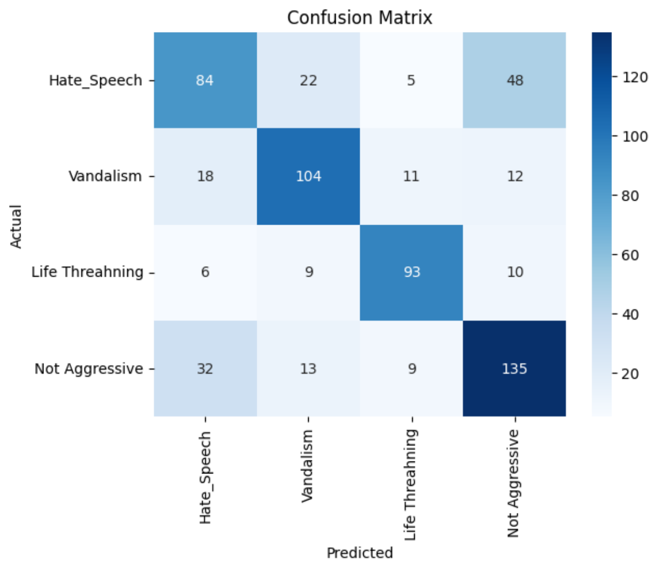

Figure 4.9: Confusion Matrix of Multilingual-BERT

The confusion matrix in Fig 4.10 for the Multilingual-BERT model shows high accuracy in predicting 'Vandalism' and 'Non-Aggressive' categories, while 'Hate Speech' and 'Life Threatening' have higher misclassification rates.

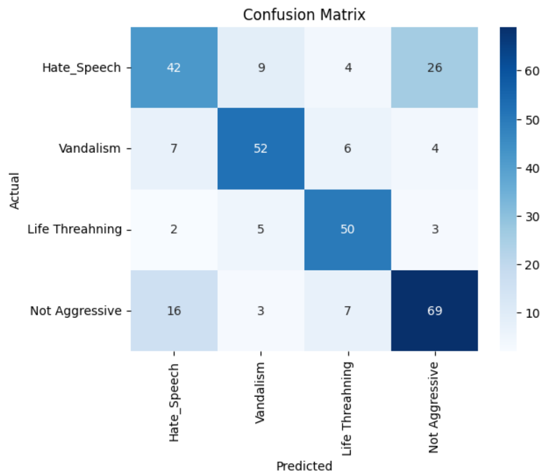

Figure 4.10: Confusion Matrix of XLM-RoBERTa

The confusion matrix in Fig 4.9 for the Multilingual-BERT model shows high accuracy in predicting 'Vandalism' and 'Non-Aggressive' categories, while 'Hate Speech' and 'Life Threatening' have higher misclassification rates.

4.4.4 ROC curve Diployment

Below, we will showcase the ROC (Receiver Operating Characteristic) curve curves to visually compare the performance of the models where class 0, class 1, class 2, class 3 represent 'Hate speech', 'Vandalism', 'Life Threatening' and 'Non-Aggressive' respectively

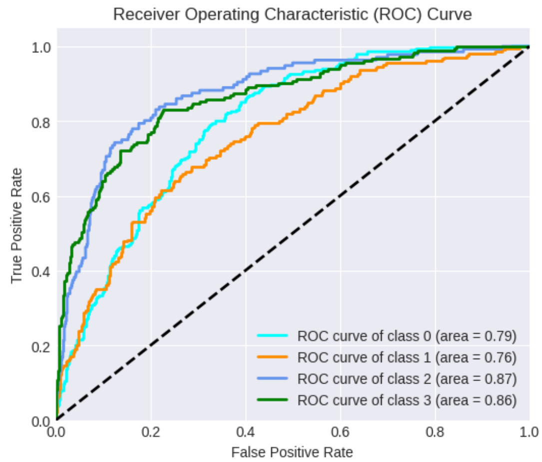

Figure 4.11: ROC Curve of Multinomial Naive Bayes

The ROC curve 4.11 for the Multinomial Naive Bayes model shows varying performance across the four classes, with the highest AUC for class 2 at 0.87 and the lowest for class 1 at 0.76.

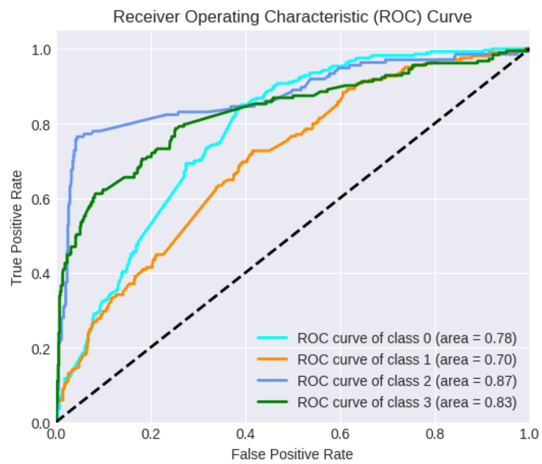

Figure 4.12: ROC Curve of Gradient Boostings

The ROC curve 4.12 for the Gradient Boosting model shows the highest AUC for class 2 at 0.87 and the lowest for class 1 at 0.70, indicating strong performance in predicting class 2 but weaker performance for class 1.

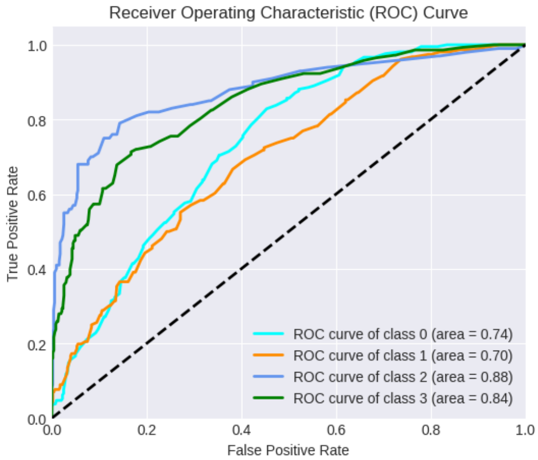

Figure 4.13: ROC Curve of Random Forest

The ROC curve 4.13 for the Random Forest model shows the highest AUC for class 2 at 0.88 and the lowest for class 1 at 0.70, indicating strong performance in predicting class 2 but weaker performance for class 1.

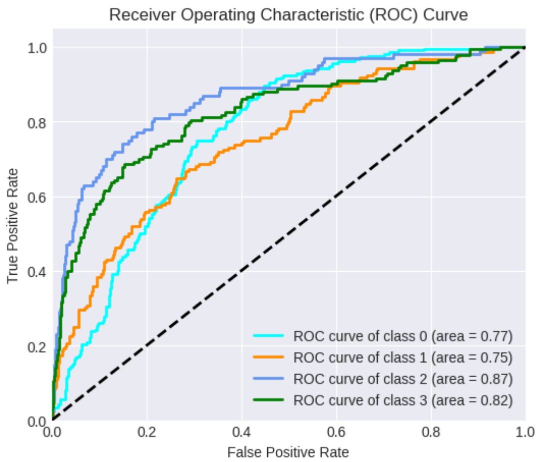

Figure 4.14: ROC Curve of Recurrent Neural Network

The ROC curve 4.14 for the Recurrent Neural Network (RNN) model shows the highest AUC for class 2 at 0.87 and the lowest for class 1 at 0.75, indicating strong performance in predicting class 2 but weaker performance for class 1.

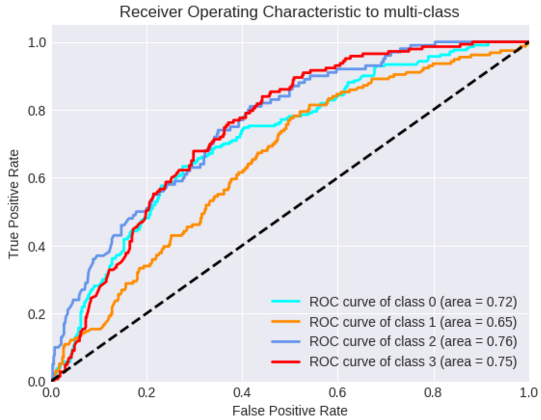

Figure 4.15: ROC Curve of Long short-term memory

The ROC curve 4.15 for the Long Short-Term Memory (LSTM) model shows the highest AUC for class 3 at 0.75 and the lowest for class 1 at 0.66, indicating moderate performance overall. This suggests that while the LSTM model is reasonably effective, it needs improvement in accurately identifying class 1.

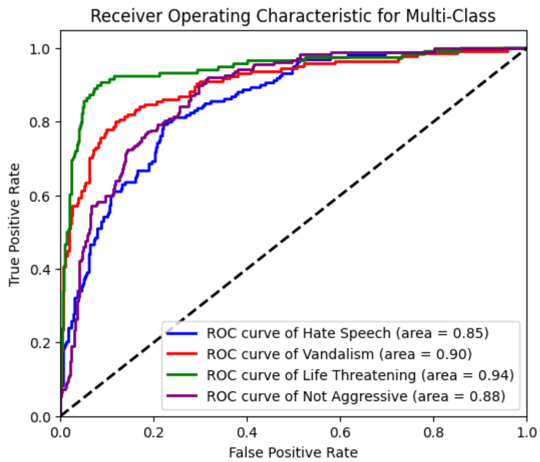

Figure 4.16: ROC Curve of BanglaBERT

The ROC curve 4.16 for the BanglaBERT model shows the highest AUC for 'Life Threatening' at 0.94 and the lowest for 'Not Aggressive' at 0.80, indicating excellent performance in predicting 'Life Threatening' and strong performance across other classes, though with room for improvement in identifying 'Not Aggressive' sentiments.

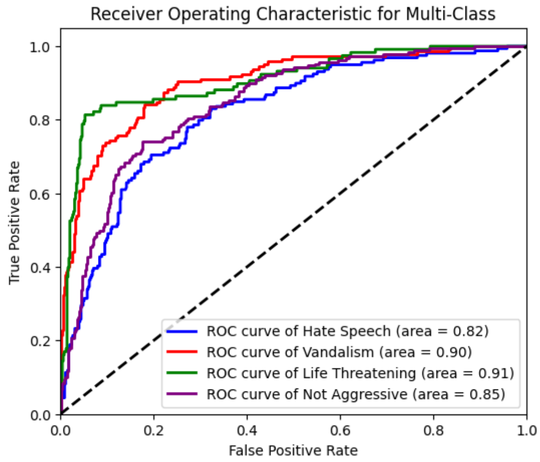

Figure 4.17: ROC Curve of Multilingual-BERT

The ROC curve 4.17 for the Multilingual-BERT model shows the highest AUC for 'Life Threatening' at 0.91 and the lowest for 'Hate Speech' at 0.82, indicating strong performance in predicting 'Life Threatening' and 'Vandalism' categories, while slightly less effective for 'Hate Speech'.

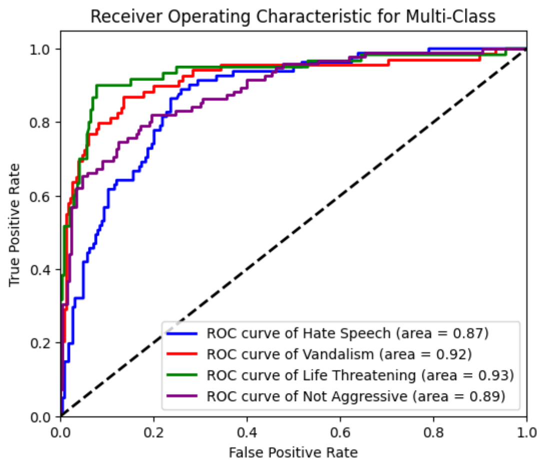

Figure 4.18: ROC Curve of XLM-RoBERTa

The ROC curve 4.18 for the XLM-RoBERTa model shows the highest AUC for 'Life Threatening' at 0.93 and the lowest for 'Hate Speech' at 0.87, indicating strong performance across all categories, with particularly high accuracy in predicting 'Life Threatening' and 'Vandalism' sentiments.

4.5 Conclusion

In this chapter, we have presented the results of various models, including Multinomial Naive Bayes, Gradient Boosting, Random Forest, RNN, LSTM, BanglaBERT, M-BERT, and XLM-RoBERTa. We have thoroughly evaluated the performance of these models, providing a comprehensive comparison to highlight their strengths and weaknesses.

Chapter 5
Conclusion

5.1 Conclusion

Social media is a platform where anything can be expressed, including aggressive religious views, which can potentially lead to significant conflicts. In this research, we initiated the task of building a unique dataset specifically designed for sentiment analysis for religious aggression detection, with the goal of classifying data into four distinct categories: Hate Speech, Vandalism, Life Threatening, and Non-Aggressive. The dataset, which is a comprehensive collection of 3,051 data points, serves as a foundation for our analysis.

To determine the most effective approach for sentiment classification, we experimented with a range of techniques, encompassing traditional machine learning algorithms, advanced deep learning architectures, and state-of-the-art transformer models. Through rigorous experimentation and evaluation, it became evident that transformer models significantly outperformed both machine learning and deep learning methods.

Among the transformer models we tested, XLM-RoBERTa emerged as the most accurate, achieving a noteworthy accuracy rate of 70%. This finding underscores the superiority of transformer-based models in handling complex sentiment analysis tasks, highlighting their ability to capture nuanced language patterns and context. Our study demonstrates the potential of advanced transformer models, particularly XLM-RoBERTa, in achieving high performance in sentiment analysis, setting a new benchmark for future research in this domain.

5.2 Future Works

The fundamental goal of this thesis was to find a strong and dependable model for correctly detecting religiously classified data. However, there are possibilities for future developments and breakthroughs in this area of research.

- In future, we will try to train the model with a more expanded dataset to achieve better performance.
- To improve accuracy, future work should focus on utilizing more fine-tuned models

References

[1] O. Sharif, M. M. Hoque, A. Kayes, R. Nowrozy and I. H. Sarker, 'Detecting suspicious texts using machine learning techniques,' *Applied Sciences*, vol. 10, no. 18, p. 6527, 2020 (cit. on p. 7).

[2] N. Mustakim, R. Rabu, G. M. Mursalin, E. Hossain, O. Sharif and M. M. Hoque, 'Cuet-nlp@ tamilnlp-acl2022: Multi-class textual emotion detection from social media using transformer,' in *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, 2022, pp. 199–206 (cit. on p. 7).

[3] S. Vernikou, A. Lyras and A. Kanavos, 'Multiclass sentiment analysis on covid-19-related tweets using deep learning models,' *Neural Computing and Applications*, vol. 34, no. 22, pp. 19 615–19 627, 2022 (cit. on p. 8).

[4] L. Khan, A. Amjad, N. Ashraf and H.-T. Chang, 'Multi-class sentiment analysis of urdu text using multilingual bert,' *Scientific Reports*, vol. 12, no. 1, p. 5436, 2022 (cit. on p. 8).

[5] R. U. Mustafa, N. Ashraf, F. S. Ahmed, J. Ferzund, B. Shahzad and A. Gelbukh, 'A multiclass depression detection in social media based on sentiment analysis,' in *17th International Conference on Information Technology–New Generations (ITNG 2020)*, Springer, 2020, pp. 659–662 (cit. on p. 8).

[6] S. Islam, M. S. A. Forhad and H. Murad, 'Banglasapm: A deep learning model for suicidal attempt prediction using social media content in bangla,' in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, IEEE, 2022, pp. 1122–1126 (cit. on p. 9).

[7] N. C. Dang, M. N. Moreno-García and F. De la Prieta, 'Sentiment analysis based on deep learning: A comparative study,' *Electronics*, vol. 9, no. 3, p. 483, 2020 (cit. on p. 9).

[8] O. Sharif, M. M. Hoque and E. Hossain, 'Sentiment analysis of bengali texts on online restaurant reviews using multinomial naïve bayes,' in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1–6. DOI: `10.1109/ICASERT.2019.8934655` (cit. on p. 9).

[9] K. L. Tan, C. P. Lee and K. M. Lim, 'A survey of sentiment analysis: Approaches, datasets, and future research,' *Applied Sciences*, vol. 13, no. 7, p. 4550, 2023 (cit. on p. 9).

[10] A. Amin, A. Sarkar, M. M. Islam, A. A. Miazee, M. R. Islam and M. M. Hoque, 'Sentiment polarity analysis of bangla food reviews using machine and deep learning algorithms,' *arXiv preprint arXiv:2405.06667*, 2024 (cit. on p. 9).

[11] A. K. Bitto, M. H. I. Bijoy, M. S. Arman, I. Mahmud, A. Das and J. Majumder, 'Sentiment analysis from bangladeshi food delivery startup based on user reviews using machine learning and deep learning,' *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 2282–2291, 2023 (cit. on p. 10).

[12] Z. Ahmed, S. S. Shanto and A. I. Jony, 'Advancement in bangla sentiment analysis: A comparative study of transformer-based and transfer learning models for e-commerce sentiment classification.,' *Journal of Information Systems Engineering & Business Intelligence*, vol. 9, no. 2, 2023 (cit. on p. 10).

[13] M. T. Ahmed, A. S. Urmi, M. Rahman, A. Z. M. T. Islam, D. Das and M. G. Rashed, 'Cyberbullying detection based on hybrid ensemble method using deep learning technique in bangla dataset,' *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, 2023 (cit. on p. 10).

[14] G. Shahariar, M. T. R. Shawon, F. M. Shah, M. S. Alam and M. S. Mahbub, 'Bengali fake reviews: A benchmark dataset and detection system,' *Neurocomputing*, p. 127 732, 2024 (cit. on p. 11).

[15] M. Hasan, L. Islam, I. Jahan, S. M. Meem and R. M. Rahman, 'Natural language processing and sentiment analysis on bangla social media comments on russia–ukraine war using transformers,' *Vietnam Journal of Computer Science*, vol. 10, no. 03, pp. 329–356, 2023 (cit. on p. 11).

[16] S. A. Mahtab, N. Islam and M. M. Rahaman, 'Sentiment analysis on bangladesh cricket with support vector machine,' in *2018 international conference on Bangla speech and language processing (ICBSLP)*, IEEE, 2018, pp. 1–4 (cit. on p. 11).