A Comparative Analysis of Identifying Multi Class Toxic Comment Classification Utilizing Deep Learning and Machine Learning Technologies

A project report submitted as a partial fulfillment of the criteria for the degree of Masters of Science (Engineering) in Information and Communication Technology

Submitted By

Sajib Sarker

ID:22209032

Depertment of Information and Communication Technology

Comilla University, Kotbari, Cumilla.

Submission: January 2025

I want to devote my project
To all my respectable teachers

Abstract

Despite extensive efforts to prevent and manage malicious human behavior, there are still many compelling issues in cyber platforms. When such individuals live in a world where technology is rapidly developing and where various online social media platforms like Twitter, Facebook, Instagram, etc. are widely accessible, those nefarious human acts rise. Because of this, extremely dangerous crimes like toxic comments, which pose a threat to users, groups, and even governments, have increased in frequency. This study aims to identify such harmful comments in Bangla text that are posted on social media sites, classify them using machine learning and deep learning algorithms, and take preventative measures. By combining TF-IDF with various machine learning algorithms, such as Naive Bayes, Decision Tree, Random Forest, AdaBoost Classifier, Stochastic Gradient Descent Classifier, Logistic Regression, KNN, and Support Vector Machine, we were able to detect toxic comments with an accuracy of 76.5%, 78.2%, 75.0%, 75.1%, 76.2%, 77.7%, 70.2%, and 78.5%, respectively. In comparison to previous machine learning techniques, we achieved an accuracy of 89% by employing deep learning algorithms.

Table of Contents

Figures Index

Figure **Page**

List of Tables

Table **Page**

CHAPTER 1

INTRODUCTION

1.1 Introduction

A toxic comment is a type of aggressive conduct or behavior that irritates or angers another person. This kind of premeditated behavior puts another individual in a terrible uncomfortable situation. Cyberbullying or online toxic comments are those types of bullying that are committed utilizing electronic media or another form of communication.

Malicious online behavior that aims to humiliate, intimidate, bully, or make toxic comments is referred to as cyberbullying. The widespread use of smartphones, the availability of the Internet, and the expectation of anonymity and immunity on social media platforms have all contributed to a recent dramatic rise in the prevalence of harassment. The use of many social media sites, including Facebook, Instagram, LinkedIn, YouTube, and others, has made them popular venues for cyberbullying. The toxic is defined as ":Comments in online forums that are unpleasant, disrespectful, or unjust are known as toxic comments, and they typically drive many members to leave the discussion.." [1].

Technology has advanced and spread everywhere in the modern era. In addition to those, it became overly accessible to the populace. This had a significant impact on social media platforms and online infrastructures. That offered the poisonous individuals a sizable window of opportunity to disseminate poison. As a result, accessing those internet channels exposes common users to bullying.

The growth of the social networking has a great significant affect in our lives. But along with the good side this is also putting people at risk and they are becoming victims in online platforms. In the month of July in 2022, almost 59% population all of the words use social media, which is half in number of the total population in the world. So, about 4.70 billion people are using social media platform around the world. Most alarming this is that they spend an average of 2 hours 29 minutes in average behind this [2]. Even in Bangladesh, in the beginning of 2022, in January, 2022 about 52.58 million people use internet in there. And over there 49.55 million were using social media. From kepios analysis, we found that in Bangladesh it is increased by 4.6 million between 2021 and 2022 at the rate of 10.1% and it's growing more [3]. So, it is becoming quite easy to harass someone using those social media platforms.

Around the world, some 300 million people speak Bangla as their first language, and another 37 million speak it as a second language. [4]. Bangla is the fifth most spoken mother tongue and seventh most spoken language considering by the total number of people whom are talking using this language all over the world [5]. Bangla is also increasingly being used on social media platforms because it is the primary language of communication for many people in India and Bangladesh. Therefore, it is necessary to filter the toxic and insulting comments that are widely available on those sites. We choose to work with the Bangla text because of this. Filtering out harmful remarks or harassment, however, is a difficult process in a language like Bangla. Additionally, the text that is available on social networking platforms is largely disorganized and

unstructured. In order to obtain the information, we must therefore use Natural Language Processing (NLP) to extract the key features from the textual input.

There are many different social media platforms, and among them, Facebook has over 2.94 billion active members worldwide, and Instagram has 1.45 billion. In Bangladesh, the platforms are extremely popular. Furthermore, 29.7% of the population of Bangladesh, or around 49.55 million individuals, uses various social media platforms. Among all different platforms Facebook and Instagram is most popular over there. Almost 44.7 million people in Bangladesh use Facebook and 4.45 million users use Instagram by using Bangla as their communication medium [3]. And occasionally, those exchanges turn toxic, hurting others in the process.

Recently, authorities have arrested a number of people for posting harmful and toxic content online. A well-known YouTuber named Shubham Mishra was detained by the Vadodara police last year after he posted a video to his millions of YouTube viewers in which he threatened stand-up comic Agrima Joshua. Furthermore, Donald J. Trump was banned from almost all social media platforms in January 2021 after he was accused of encouraging the riots in Washington. There is a problematic issue as a result, and it is crucial to identify certain content before it is submitted. Additionally, consumers are finding the internet to be a dangerous place as a result of these destructive contents.[6]

1.2 Research Motivation

We conducted this study in response to the demand for a comprehensive, practical, accurate, and reliable cyberbullying detection classifier. Cyberbullying is the term for a forceful and intentional act committed once or more against a helpless victim online by a person, group, or both, using a variety of communication channels (such as email, smartphones, and social media). Contrary to traditional bullying, cyberbullies use quick-changing techniques and structures that are more harsh and difficult to spot. For instance, spreading rumors about somebody on the internet and social media is easier and less likely to result in legal action. And whether the problem is linked to other problems. The result could be a serious issue. In modern culture, we frequently witness conflict between various groups as a result of offensive remarks on Facebook posts. Sometimes it gets out of hand and becomes a major issue. Numerous people are hurt, and some of them also pass away. Because of this, it is imperative to spot poisonous commentary in many national concerns.

The most frequent method utilized by the offenders is text-based cyberbullying. Additionally, various other types are frequently blended. As a result, Bangla is being used more and more on social media sites. Our population is becoming more politically interested. They frequently express their opinions and perceptions of a situation on social media. Spend a considerable amount of time online, and the majority of them will use foul language to comment on a piece if they don't like it or if the author is a member of a political bully group professional investigation to identify political harassment in Bangla Text has not been properly completed, which is an urgent necessity in the present. Because of this, we selected this as our research topic.

1.3 Research Methodology

Bullying essentially occurs everywhere online. Therefore, it is crucial to develop a suitable detection model that can be transported and modified. By re-implementing the models on an entirely new dataset, we can enhance our research. Using straightforward building blocks and a graphic, we will briefly outline our complete methodology in this part before moving on to the specifics. Here, we offer a useful technique for identifying cyberbullying on multi-label-toxic comments in Bangla, which are gathered from Instagram and Facebook postings. In this study, we offer a learning framework for supervised learning to identify cyberbullying in Bangla writings that contain multiple harmful comments.

This is the first attempt, to the best of our knowledge, to use machine learning models to identify political cyberbullying in Bangla. To sum up, we take the succeeding action described below:

- Explain the importance of social media monitoring for cyberbullying in formal terms.
- Manually gather information from Facebook and Instagram posts, as well as using a custom Python tool to get some information.
- We next saved the text data we had gathered into a csv file to create our own, entirely original dataset.
- We suggested using a model we had created to carry out the entire procedure.
- After that, divide the data into two halves. Both are classified as test data and training data, respectively.
- After training the machine learning classifiers with training data, test them with test data.
- Utilize the test data to assess the classifier once it has been trained using the training data.
- Calculate performance in terms of recall, accuracy, and precision using the F1-Score. These are computed in the Spyder Notebook.
- The output of different machine learning classifiers was compared.
- Finally, we selected the most effective machine learning algorithm by contrasting its performance with that of deep learning.

1.4 Project Descriptions

Chapter 1 explains cyberbullying and outlines the thesis's goals and objectives.

Chapter 2 mentions social media, text mining, when discussing the literature study, the identification of cyberbullying, the classification of offensive texts, and offensive text images on social media. Similar studies in several areas of cyberbullying detection are also covered in this chapter.

Chapter 3 offers image and training and test datasets for our suggested methodology's feature extraction and data preprocessing processes.

Chapter 4 represents the implementation of our proposed model. It thoroughly defines each and every machine learning model that we employ in our study. It also describes about the software we use. Code segment of the supervised machine learning algorithm.

Chapter 5 Contains data on machine learning classifiers and evaluations of the algorithms' performance. It includes all of the outcomes from the built-in models.

Chapter6 shows both the success of our project and its continued development.

CHAPTER 2

LITERATURE REVIEW

2.1 Related Works

There has been a sizable amount of research performed in text categorization or cyberbullying detection for English and a small amount for Bangla also by the classification of posts or conversations using text mining. But no research has been performed on detecting or preventing bullying on toxic comment issues on social media. A comparison of text classification techniques has been conducted by using various supervised approaches by Reichart, Dinakar, and Lieberman. When applying supervised learning for text classification, Yin, Xue, and Hong [4] labeled texts using the Ngrams methodology and weighted them using the term frequency (TF)-inverse document frequency method (IDF).

Kelly Reynolds used the Decision tree (J48) and k-nearest neighbor (KNN)[6]. Comments of YouTube videos are used in labeled manually & multiclass classification is used. In [7], SVM is applied for text classification but they did not get good performance. A comparison of NB & KNN with SVM is shown in [8] by Zhijie et al, SVM performed is better. All of the works are performed, about English text.

The performance & accuracy of the algorithm may vary due to the linguistic difference between English & non-English content. For Indian text NB show, a better result, and a combination of Ontology-based classification & NB provide a better result for Panjabi text. In [9], For Urdu text, SVM outperforms NB. The artificial neural network model-based approach performs better for Tamil text. [10]. NB classifier was used efficiently for Indian text[9] but they didn't specify the amount of data. Nandhini et al. [11] proposed Naïve Bayes machine learning approach and achieve 91% accuracy and got their dataset from MySpace.com, after that they proposed another model which is the combination of NB & FuzGen, which achieve 87% accuracy. Bunchanan et al. [12] collect the data from the chat of War of Tanks games and classified them manually.

Then compare the result with the Naïve classification and the result is too poor. Romsaiyud et al. [13] proposed NB and achieve 95.79% accuracy for the dataset get from Slashdot, Kongregate, and MySpace but the problem is the cluster processes don't operate concurrently. Isa et al [14] collect data from kaggle and used SVM & NB. Get the accuracy of 97.11% for SVM & 92.81% for NB but they did not specify the amount of the dataset used for training or testing, so the results might not be reliable.

Haidar et al[15] proposed a model to identify cyberbullying for Arabic text. They used Naïve Bayes and got a precision of 90.85%, then they used SVM and got a precision of 94.1%, but they also had a high proportion of false positives. Although they achieved excellent accuracy— 56% precision, 78% recall, and 96% accuracy—their dataset was imbalanced, which led to misleading findings and had an impact on their precision score. Using WEKA with a dataset obtained from Myspace.Maqsudur Rahman et al[29] proposed a model to predict ratings, based on customer text reviews. They collect the dataset from Kaggle and get that Random Forest Classifier provides the best performance for their model. Their dataset has 25 columns but they use only 2 of those.

Zhang et al[16] suggested using a unique pronunciation-based convolution neural network PCNN) to address the issue of noise and bullying data sparsely and address the issue of class imbalance. Twitter sent out 13,313 messages, and formspring.me sent out 13,000 messages. The uneven nature of the Twitter dataset prevented accuracy from being calculated. Md. Tofael Ahmed, Maqsudur Rahman el al.[30] proposed a model to detect the bullying in Bangle & Romanized Bangla text. They use 3 datasets. One for Bangla, one for Romanized Bangla and the last one is the combination of previous 2 datasets. They collect the data from the Youtube comments and preprocessed the data using machine learning and deep learning algorithms. CNN perform best for Bangla text and MNB for the others two Franciska

De Jong et al.[17][18][19][20]. In their first and second articles, they constructed a Support Vector Machine classifier. The primary difference between the two studies is that the second study used gender data in its classification, yielding 43% precision, 16% recall, and no accuracy information. Additionally, for their second piece, 4626 comments were gathered from 3858 distinct users.

Manual bullying and non-bullying are the labels assigned to those comments (9.7% and 93.3%, respectively). Their results, using the SVM classifier, showed up to 55% recall and 78% accuracy. Using three datasets, Md. Tofael Ahmed, Maqsudur Rahman, et al. [21] found that SVM performed best (76% accuracy) on the first dataset, whereas Multinomial Naive Bayes performed better on the second (84%) and third (80%) datasets.In order to identify instances of cyberbullying in Bangla texts, Md. Tofael Ahmed, Maqsudur Rahman, and colleagues [31] developed the PMI-SO model. They employ two datasets. The PMI dataset, which has 10277 Bangla texts, comes last. The first one contains 5000 Bangla texts that were gathered from social media. SVM gives them the lowest accuracy of 85%, while XGBoost gives them the highest accuracy of 93%.

Maqsudur Rahman el al [32] proposed a model for detecting the spam from Bangla text. For this they collect 2759 Facebook comments. Use KNN, SVC,Gradient Boosting & MNB for classification. But their dataset doesn't contain a handsome amount of data. It could have been a bit bigger. Parime et al [33] proposed another approach for detecting the bullying texts. They collect their dataset from MySpcae. They label the data manually and use SVM classifier for classification. The limitation what we think about this approach is, the use of a single classifier. They can use more classifier and it could make their approach more accurate and acceptable.

A feature extraction method is proposed by Chen et al. [34] called Lexical Syntactic Feature. They also use SVM classifier and got 77.9% accuracy. Furthermore, a technique is proposed by Ting et al. [35] which is based on the SVM. They collect their data from the Social media and after that SNA measurements and sentiments are used for feature extraction. Number of experiment were made and finally they got 97% precision and 71% recall. But what the problem in this proposed technique we think is. The result is not balanced and did not specify from which social media they collect the data to make their dataset.

Random Forest, N-grams, Decision Trees, and Neural Networks are further suitable text classification methods for non-English content. There is no work has been done for detecting political bullying in Bangla text. Thus, the purpose of this study was to investigate different

machine learning methods for detecting bullying in political issues in Bangla text & identifying the best one which provides the best performance.

Facebook data was gathered by Ahmed, Md. Faisal, et al. [22] and binary and multiclass classification models were employed. To extract features, Word2Vec was utilized. In this manner, they achieved 87.91% accuracy in binary classification and 85% accuracy in multiclass classification. However, for somewhat long sentences, this model displays false positive results.

Using the data gathered from Facebook, Hussain, Md Gulzar, Tamim Al Mahmud, and Waheda Akthar et al. [23] suggested a model. N-gram is used for feature extraction in this model, which is suggested using a root level technique. This study demonstrates that root level operations are ineffective.

Talpur et al. [24] combined PMI-semantic analysis with language presentation and embedding mostly for feature extraction and data interpretation. They used the KNN, Random Forest, Naive Bayes, Decision Tree, and SVM algorithms in conjunction with a variety of machine learning models. Test data results indicate that the binary and multi-class models provide much superior results for the suggested model. This also provides a better way to distinguish between cyberbullying and its effects on social media networks.

Haque et al. [25] implemented review analysis system which is used both in Bangla text and Phonetic Bangla text, they used reviews based on restaurant that detect reviews by using several supervised ML. tecliniques. An Identification which is of vectorizers based implemented on different classifiers, but with Support Vector Machine (SVM) they achieve the highest accuracy of 75.58%

Zhang et al. [41] proposed a method which is based on the Deep Learning and Neural Networks approach. In that research paper Pronunciation convolution neural network (PCNN) is used, thereby mitigating bullied data for overcoming the imbalance in the class. 1313 messages and 13000 messages were collected from twitter and formspring.me respectively. Twitter data was imbalanced and for this reason twitter dataset was ignored in case of accuracy calculation. Through the research the gained 56% precision, 78% recall and 96% accuracy, that clearly indicate that data set is imbalance as precision is only 56% and we are getting fake accuracy.

An NLP and machine learning-based system was proposed by Ahmed, Md. Tofael, et al. [26]. The YouTube API provided the dataset's data. Three datasets and the machine learning classifiers SVM, Multinomial NB, XGBoost, and Logistic Regression were employed. Multinomial NB obtained height accuracy of 84% and 80% for the remaining datasets, whereas SVM obtained 76% accuracy for one dataset.

2.2 Multi-Label-Toxic-Comment and its Impact

Making comments in online debates has grown to be a significant means of exercising one's right to free expression. However, this fundamental freedom is being violated since harmful users stifle respectful conversations with their offensive remarks. A nasty, disrespectful, or irrational comment is referred to as toxic if it is likely to cause other users to quit a debate. Classifying offensive comments is a part of sentiment analysis. Here, we'll offer a fine-grained classification system for toxic remarks and describe why it's important to find them in online discussions.

2.3 Classes of Toxicity

1. Vulgar:

Language that is rude or obscene can be referred to as vulgarity in the sense of vulgar discourse. "Cursing" is the word most frequently connected to the verbal manifestation of vulgarity. There are other subcategories of filthy terms, though. Example: "পরী তকমা লাগানো, হট সেক্স ভিডিও কবে আপলোড করেছে, আমি একবার লাগাতে চাই.". Swear words are discussed in the first class. In the above example, the toxicity of this remark can be summed up in the phrase "bullshit." As is customary for this class, if at least one offensive word has been discovered, the entire statement need not be considered. For detection, straightforward blacklists of offensive words can be utilized. Malicious users frequently use these words in different ways or with different spellings to avoid blacklists.

2. Insults:

Example: "এই রমজানে ও তুই লোচ্ছামি করতেছিস হারামি শুয়ার..তুই নিজে নিজের গু'য়ের গন্ধ শুকে মরে জা না?". Statements concerning specific people or groups are not included in the preceding class of comments, but they are in the "insults" class. "Insults" are harsh or disrespectful comments about a specific person or group of people. The comment in the example is addressed to another user directly, which is typical but not required.

3. Threats:

Example: "প্রধানমন্ত্রী হক সাহেবের ক্ষতি হলে জাতির স্বার্থে কেনো কোনো বাম পক্ষ কে ছাড় দেয়ার উচিত না.". A frequent threat made in online debates is to close another user's account. Comments that are extremely harmful make threats against the lives of other users or the users' families. This category includes declarations or arguments that call for punishing, torturing, harming, or damaging oneself or others.

4. Troll:

Online posts or remarks that are intended to 'bait' someone into an argument or an emotional response Example: আগে কুত্তাটাকে নির্বাচক থেকে অপসারণ করতে হবে।একটা ছাগলের বাচ্চা । এইসবকে কি করে নির্বাচক করে বুঝি না

5. Religious:
Example: এই মালাউন পবিত্র কুরআনে আগুন দিয়েছে ওর ফাঁসি চাই

2.4 Machine Learning:

An area of artificial intelligence called "machine learning" that studies how computers can recognize patterns in databases and solve problems on their own. In other words, machine learning helps computers to identify patterns within the confines of already created algorithms and data sets and to deliver pertinent solutions. Both new technology and machine learning have advanced. As a result of using experience, machine learning creates artificial knowledge. People must take action before the computer can generate solutions on its own. For instance, the systems must be prepared with exact methods and data, which calls for the development of analytical standards for confirming data patterns. Following the completion of these two steps, the system may carry out the following tasks using machine learning:

- Machine learning models can generate an accurate prediction based on the analytical data (training data).
- They can calculate the probability of specific outcomes using data.
- They are capable of conducting pertinent data searches, extractions, and summaries.
- Machine learning-powered smartphones may recognize faces while taking photos or unlocking themselves.
- Machine Learning has the ability to adapt the machine to its specific changes on its own.
- On the basis of recognized tendencies, they can optimize processes.
- The machine, which used Machine Learning as its operating system can recover them if any of those functions fail during the operation.
- Machine Learning can perform a great rule in the medical field by analyzing previous data which was used to train the machine.
- It can also reduce the human work and suffer by doing automatic language translation, help to predict the probability in the stock market, detecting cyberbullying, fraud detection and many more operations.

Machine learning functions identically to human learning because of human learning. For example, if a baby is shown pictures with certain things on them, he or she will be able to tell them apart. Machine learning works in a similar way. Takes the data as an input and specific commands and then the machine trained how to identify and distinguish between different things. It learns from the input pattern. For this reason, the algorithm or program is trained and saved with data. Iterative machine learning is significant because it permits models to change on their own when they are presented with fresh facts. In order to generate accurate and dependable outcomes, they rely their conclusions on prior computations. In short what the main procedure occurs in the

machine learning is, at first human train the machine by some data called training data and based on this training machine will work or perform its functions with the new data.

Because of new computing technologies, modern machine learning is different from machine learning from the past. The topic of whether computers could learn from data was prompted by pattern recognition and the idea that they might learn without being programmed to carry out certain tasks, which interested researchers in artificial intelligence sought to answer. Machine learning's iterative aspect is essential because models may independently modify as they are exposed to new data. They gain knowledge from past calculations in order to deliver reliable, reproducible judgments and results. Although it is an ancient science, it has recently gained new momentum.

The same forces that have increased the use of Bayesian analysis and data mining have also led to a renaissance of interest in machine learning. Examples include growing volumes and varieties of data, more powerful and cost-effective processing, and more reasonably priced data storage. Even on a very large scale, these components enable the rapid and automatic development of models that can evaluate more complex and vast data and provide faster, more accurate results. By creating precise models, a company can increase its chances of seeing promising possibilities or avoiding unidentified risks.

2.4.1 Types of Machine Learning

Machine learning algorithms can be trained using a variety of methods, each of which has benefits and drawbacks. We must first examine the types of data that each type of machine learning employs in order to comprehend the benefits and drawbacks of each type. Unlabeled data and labeled data are the two forms of data that machine learning employs.
Although labeled data includes input and output characteristics in a completely machine-readable form, labeling the data initially necessitates a significant amount of human effort. Unlabeled data has one or no parameters in a format that can be read by machines. Although this does away with the requirement for human work, it calls for more difficult answers. Some machine learning algorithms also have highly specific applications. There are numerous machine learning algorithms, including:

- **Supervised Learning:**

One of the most fundamental varieties of machine learning is supervised learning. In this case, labeled data are used to train the machine learning system. Despite the fact that precise data labeling is required for this method to function, supervised learning is quite effective when applied in the appropriate situations. A tiny training dataset is provided to the ML algorithm in supervised learning. This training dataset, which is a smaller fraction of the total dataset, provides the algorithm with a general understanding of the issue, the solution, and the data points that need to

be handled. The training dataset gives the algorithm with the labeled parameters it needs to finish the task and shares many properties with the final dataset. To create a cause-and-effect relationship between the variables in the dataset, the algorithm essentially looks for connections between the supplied parameters.

Following training, the algorithm has a general understanding of how the data work and how input and output relate to one another. The final dataset is used to test this solution, and it gains the same knowledge from it as it did from the training dataset. This means that supervised machine learning algorithms will continue to advance even after being put to use, identifying fresh connections and patterns as they learn from fresh data.

The steps of supervised learning are as follows.

- Determine the kind of training dataset first.
- Use labels to collect the training data.
- From the training dataset, create test, validation, and training datasets.
- Determine the training dataset's input features, which must be sufficiently specific to provide reliable output prediction.
- Select the strategy that the model responds to the best, such as a decision tree or a support vector machine.
- Use the practice data to apply the algorithm. It is occasionally necessary to use validation sets, a subset of training datasets, as control parameters.
- Utilizing the test set, evaluate the model's accuracy. If the model predicts the outcome correctly, it is deemed to be accurate.

Types of Supervised Learning:

Problems with supervised learning can be further broken down into two categories:

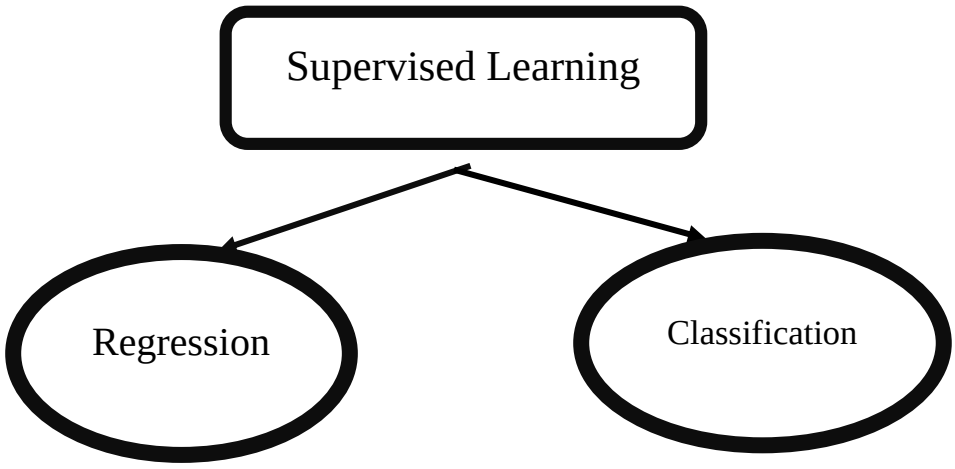

Figure 1. Supervised Learning Types

Regression:

If there is a correlation between the input and output variables, regression techniques are employed. It is used to predict continuous variables such as the weather, market patterns, and other things. a selection of popular supervised learning regression methods. Following are listed:

- Bayesian Linear Regression
- Non-Linear Regression
- Polynomial Regression
- Regression Trees
- Linear Regression

Classification:

Classification methods are used when the output variable is categorical, which means there are two classes, such as Yes-No, Male-Female, True-False, etc.o Logistic Regression

- Decision Trees
- Support vector Machines
- Random Forest

- **Unsupervised Learning:**

Unsupervised machine learning can be advantageous when used with untagged data. The software can now manage far larger datasets because the dataset can be converted into machine-readable data without a person's help. The supervised learning labels enable the system to identify the specific type of link between any two data elements. Unsupervised learning, on the other hand, creates hidden structures since it lacks labels to base its activity upon. Without requiring human input, the algorithm abstractly interprets relationships between data points. Unsupervised learning systems are flexible because of these hidden structures. Without a preset and detailed issue description, unsupervised learning algorithms can adapt to the situation.

Example: Consider supplying the unsupervised learning system with an input dataset that includes images of various cat and dog breeds. The algorithm never receives training on the supplied dataset, therefore it is unaware of its properties. Allowing the visual features to speak for themselves is the unsupervised learning algorithm's task. This task will be carried out by an unsupervised learning algorithm, which will group the image collection into categories based on visual similarity.

Types of Unsupervised Learning:

The unsupervised learning method can be divided into two different categories of issues.

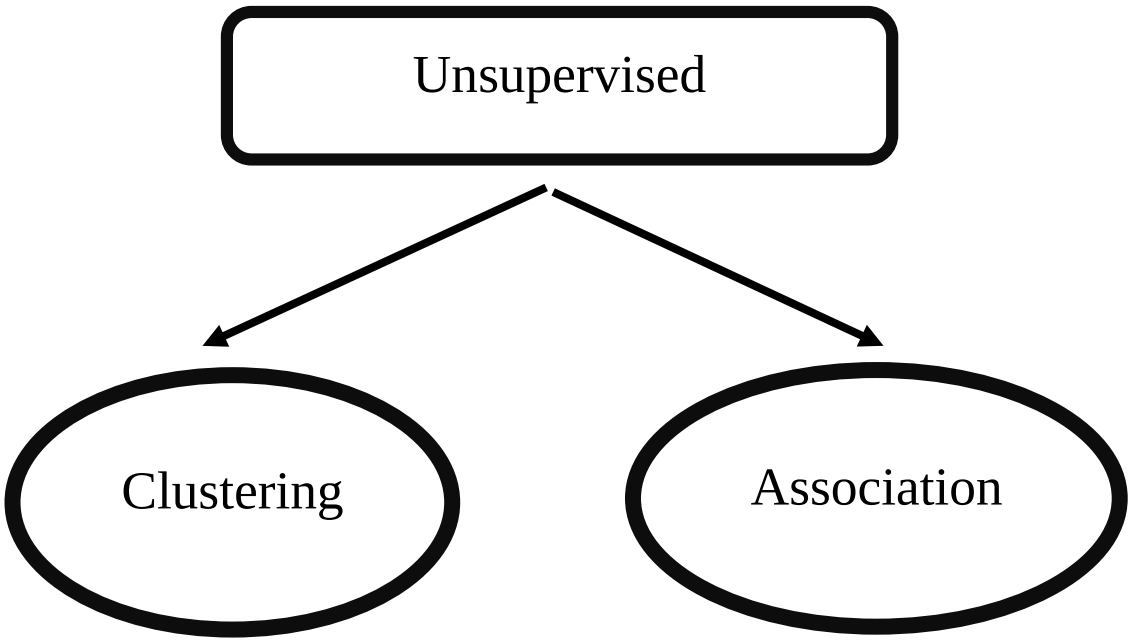

Figure 2.Unsupervised Learning Types

Clustering:

Items are classified into clusters using the clustering technique so that those with the greatest similarity remain in one group and have little to no similarity with those in another. Based on the presence or absence of cluster analysis-discovered commonalities, the data objects are categorized.

Association:

The associations between the variables in a huge database are found using an association rule, an unsupervised learning technique. By doing so, the collection's co-occurring pieces are established. Due to the association rule, marketing strategy is more effective. If X is bread, then Y is butter or jam, and people who buy X also buy Y regularly. An association rule in action is demonstrated by market basket analysis.

- **Reinforcement Learning**

The way people learn from facts in their daily lives is quite similar to how reinforcement learning works. It has a self-improving algorithm that adapts to changing circumstances and learns from errors. Positive results are "reinforced" or encouraged while negative results are "punished" or discouraged. Reinforcement learning, which is based on the psychological idea of conditioning,

functions by placing the algorithm in a context with an interpreter and a reward system. The interpreter determines whether or not the outcome is favorable after receiving the output result from each algorithm iteration.

If the program chooses the right answer, the interpreter improves the response by commending the algorithm. If the result is unfavorable, the algorithm must continue the process until a better conclusion is obtained. The effectiveness of the outcome is typically directly correlated with the reward system. The answer is not an absolute value in common reinforcement learning use cases, such as determining the fastest path between two points on a map. An efficiency rating with an a% value is used as a substitute. The algorithm gets rewarded lavishly the higher this percentage amount. The computer has been programmed to provide the greatest solution for the best reward as a result.

Example Consider an AI agent who is in a maze and whose job it is to find the diamond. The agent interacts with the environment by taking certain activities, and based on those actions, the agent state is alerted and it also receives feedback in the form of rewards or penalties. The agent keeps engaging in those behaviors in order to engage with the world, signal his presence in it or continue to do so, and gather feedback. In doing so, he develops the ability to explore his surroundings.

Principal Elements of Reinforcement Learning:

- The agent in Reinforcement Learning is not given instructions regarding the environment or the appropriate course of action.
- It is founded on the hit-and-miss method.
- In reaction to input from the previous action, the agent executes the succeeding action and changes its states.
- The agent might then be given a reward.
- Agent must investigate the stochastic environment in order to maximize positive rewards.

Reinforcement learning components:

The four primary elements of reinforcement learning are as follows:

1. Policy

2 .Reward signal

3. Value function

4. Model for environment

Policy:

A particular agent's behavior at that particular time is defined by their policy. It establishes a link between the perception of environmental circumstances and the responses to them. A policy is the

core element of the RL because it is the only thing that can determine how an agent will behave. In some circumstances, it might just be a simple function or lookup table, but in others, it might require general computation like a search algorithm. It could be a stochastic or deterministic policy: Deterministic Policy: $a = \pi(s)$; Stochastic Policy: $\pi(a \mid s) = P[At = a \mid St = s]$

Reward Signal:

The aim of reinforcement learning is determined by the reward signal. At every state, the environment immediately delivers a signal known as a reward signal to the learning agent. Based on the agent's successful and unsuccessful acts, these prizes are awarded. The agent's principal goal is to increase the overall supply of rewards for moral behavior. The policy may be altered by the reward signal. In the event that an agent's actions result in a poor reward, for instance, the policy might be altered to choose different behaviors going forward.

Value Function:

The value function informs an agent of the advantages of the circumstance, the action, and the prospective reward. A reward provides the immediate signal for both good and bad behavior, whereas a value function specifies the desired condition and action for the future. Since value cannot exist without it, the reward is an integral part of the value function. Value estimation can be used to gain more benefits.

Model:

The final element of reinforcement learning is the model, which duplicates the activities of the environment. The model allows for the drawing of conclusions regarding the behavior of the environment. For instance, if a state and an action are specified, a model can predict the state and reward that will happen. Because it is used for planning, the model enables the choice of a course of action while taking into account all conceivable consequences before those outcomes actually happen. Model-based techniques are methods for using models to solve RL issues. A model-free method, however, doesn't.

2.5 Text Analysis

The use of offensive language is required in any discussion of cyberbullying by natural logic. The computer used in the suggested text mining method won't look up hazardous terms in a glossary. A dataset of labeled Bangla text conversations will be used to detect cyberbullying [26]. The following measures must be completed in order to put the suggested approach into practice:

a) Collecting data from social media (Facebook)

b) Labeling the data manually

c) Perform preprocessing approaches

d) Training and testing the classifier by using training and testing data.

2.6 Deep Learning:

Deep learning methods used in machine learning (ML) and artificial intelligence (AI) construct algorithms in numerous layers to produce Artificial Neural Networks (ANNs) that can learn on their own and make intelligent decisions on their own. Actually, machine learning is a subfield of deep learning. It is a crucial element of data science that incorporates statistical modeling and forecasting. The deep learning method is highly useful for data scientists that work with a lot of data. This method makes it simpler and quicker.

2.6.1 How deep learning Works

Artificial neural network (ANN) architecture, which has a layered structure, is used by the majority of deep learning approaches. Because of this, deep learning models are also known as deep neural networks. The construction of a neural network uses the same principle as the biological network of neurons in the human brain. Deep learning requires extensive training in order for a machine to be capable of making the right decisions, but once this functional deep learning begins to function as intended, it is regarded as the foundation of AI.

In this context, the hidden layer in neural networks is referred to as "deep" in deep learning. On the other hand, a deep neural network may have as many as 150 hidden layers.

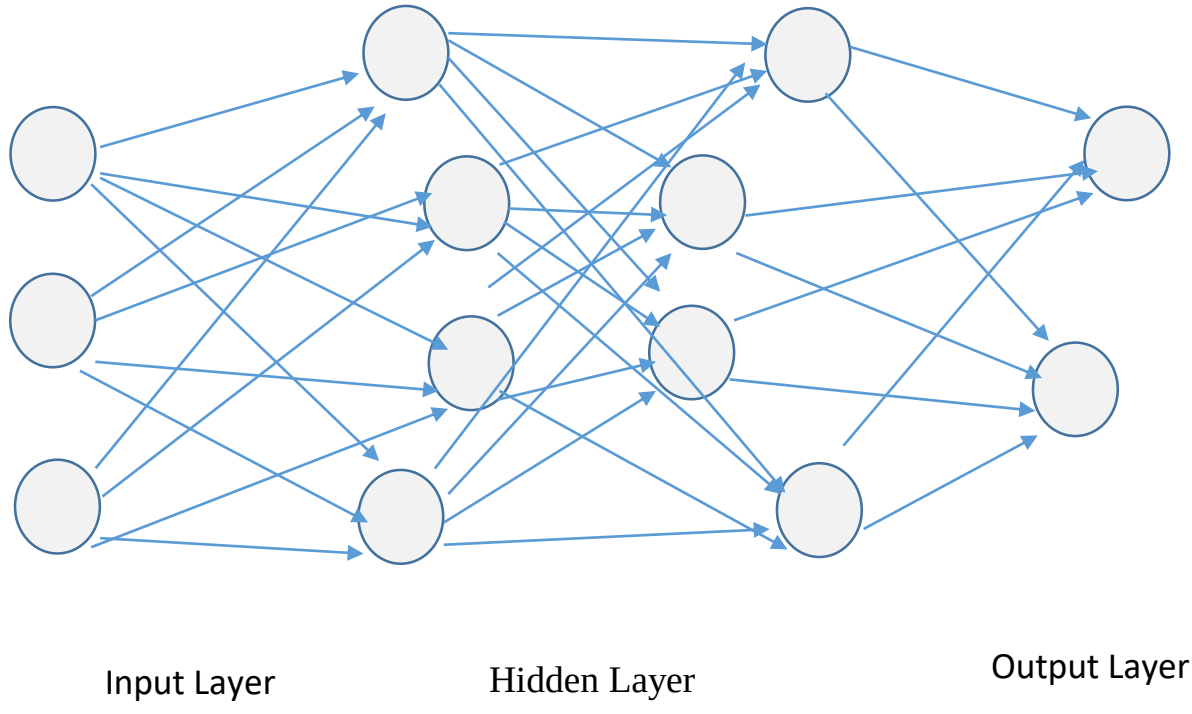

Figure 3. Neural networks organized in layers with interconnected nodes

In deep learning method models are trained by neural network and by making use of large set of data labeled data learns feature directly from data without need to any kind of neural feature extraction.

2.6.2 Types of Deep Learning

Numerous sorts of algorithms are employed in deep learning methods. Among them, we will discuss about the most popular methods. Those are

1. **Conventional Neural Network:**

CNNs, a sort of deep learning technique built specifically for object detection and image processing, are conventional neural networks .CNNs processes and makes future data extractions using a number of layers. They are as follows:

- Convolution Layer
- Rectified Layer Unit (ReLU)
- Pooling Layer
- Fully Connected Layer

The technique of filtering and analyzing each element of an image is known as "convolution." Convolutional neural networks, a subfield of artificial intelligence that enables machines to understand how to interpret the visualization of the outside environment, are primarily used in computer versions. Take face recognition technology as an illustration.

2. **Long short-Term Memory Networks:**

A long short-term memory network, or LSTM, is a sort of recurrent neural network (RNN) that can learn new information and remember long-term dependencies. The LSTM's default tendency is to recall the prior data over extended periods of time. Over time, LSTM retain knowledge. They are helpful in predicting time series because they remember past input. The LSTM's operational stage is listed below:

- At first, they forget the unnecessary portion of the previous step
- Then, cell state is update selectively
- Finally, the cell state contains the output of certain parts

LSTMs are basically used for task like speech recognition, music composition etc.

3. **Recurrent Neural Network:**

Recurrent neural network (RNNs) allow the algorithms to recall the past data points by using own built-in-feedback loops. Recurrent neural networks use their past memory to understand current event and also try to predict the future. A deep learning based neural network can sometimes can

think an optimum way when it has good context. For example let consider a maps app which is powered by a recurrent neural network and if it has previous experience then this can remember when traffic usually got worse. Then it can use its previous experience to recommend an alternative route.

2.7 Machine Learning vs Deep Learning:

Machine learning algorithms include deep learning algorithms. Instead than focusing on specific differences, it would be preferable if we could consider what makes deep learning unique after being a branch of machine learning. It is the design of an artificial neural network (ANN) algorithm, which requires little human interaction and a lot of data.

Traditional Machine learning algorithm have simple structure as compared with deep learning algorithm. As opposed to that, deep learning follows the structure of artificial neural network (ANN).when those are multi layered then deep learning acts like human brain.

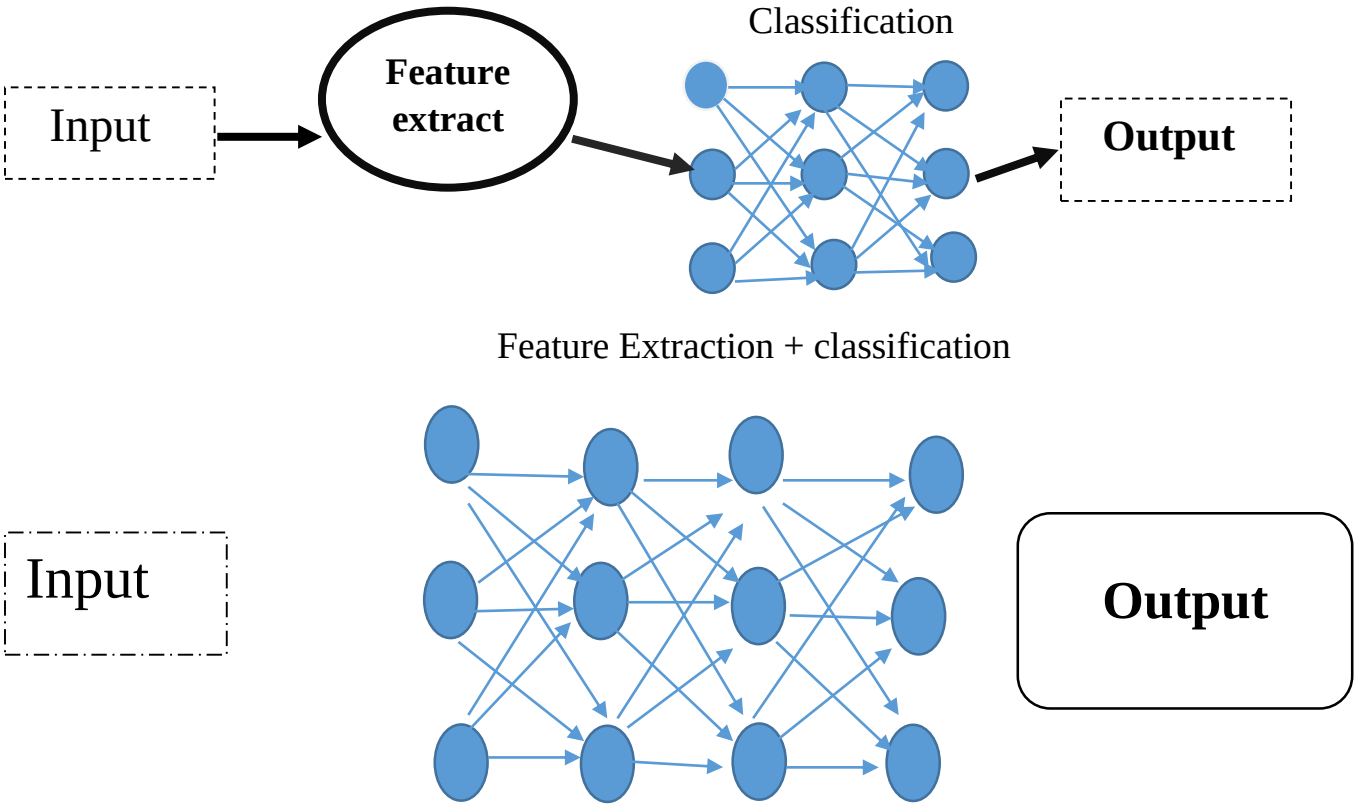

Figure 4.Machine learning Vs. Deep learning based on feature Extraction

Deep learning algorithm required less human interaction.Like traditional machine learning, deep learning doesn't need feature extraction separately.Feature are extracted autometically and deep learning algorithm learns from own error.

Again Deep learning requires huge data than traditionmal machine learning algorithm for proper functionality.For example,if machine learning with thousand of data machine learning works with millions of data.Deep learning need huge data for multi layered structure[27].

2.8 Natural language Processing (NLP):

Computer science, the humanities, and artificial intelligence are all subfields of natural language processing, or NLP for short. Modern machines can analyze, alter,

And interpret human languages thanks to technological developments. It performs tasks including translation, automatic summarization, named entity recognition (NER), audio recognition, relationship extraction, and topic segmentation to help developers organize knowledge.

Computers can now understand human language thanks to natural language processing, or NLP. Behind the scenes, NLP analyzes phrase structure and word meaning separately, then uses algorithms to extract meaning and offer results. In other words, it decodes human language to do a variety of tasks automatically. Virtual assistants like Google Assist, Siri, and Alexa are probably the most well-known NLP applications. NLP turns spoken and written words into numbers to make it easier for machines to understand.

An application of NLP that is widely used is chatboots. They help support staff solve issues by automatically deciphering requests in common language and responding. You probably use a lot more everyday apps that have included NLP without even realizing it. Offer to translate a Facebook post that is written in a different language or send text recommendations in order to filter out superfluous marketing emails and place them in the spam folder. In a nutshell, natural language processing's goal is to make it simple for machines to understand human language, which is intricate, confusing, and enormously varied.

NLP components:

NLP consists of the following two elements:

1. Natural Language Understanding (NLU):
   Natural language understanding (NLU) enables computers to grasp and evaluate human language by extracting text's metadata, such as concepts, entities, keywords, emotions, relations, and semantic roles. NLU is mostly used in business applications to understand written and spoken client issues.
   Following are the tasks involved in NLU:
   - The given input is mapped into a usable form using it.
   - It is employed to examine many facets of language.
2. Natural Language Generation (NLG):

The process of transforming digital data into natural language representation, or NLG, is known as natural language generation. The three essential elements are text planning, sentence planning, and text realization.

2.9 Challenges:

It's difficult to understand Bangla text writing. In the case of Bangla data, vectorizing category data is a very challenging task. Furthermore, NLTK's built-in stopword library does not exist for the Bangla language. Bangla is also more sophisticated than English in terms of word phrasing and structure. It could be difficult at times to tell them apart.

CHAPTER 3

RESEARCH MEHODOLOGY

3.1 Flow Chart of Proposed Methodology:

The research's objective is to identify political bullying in Bangla writing. We suggested a model for this. Following are the elements of the proposed model.

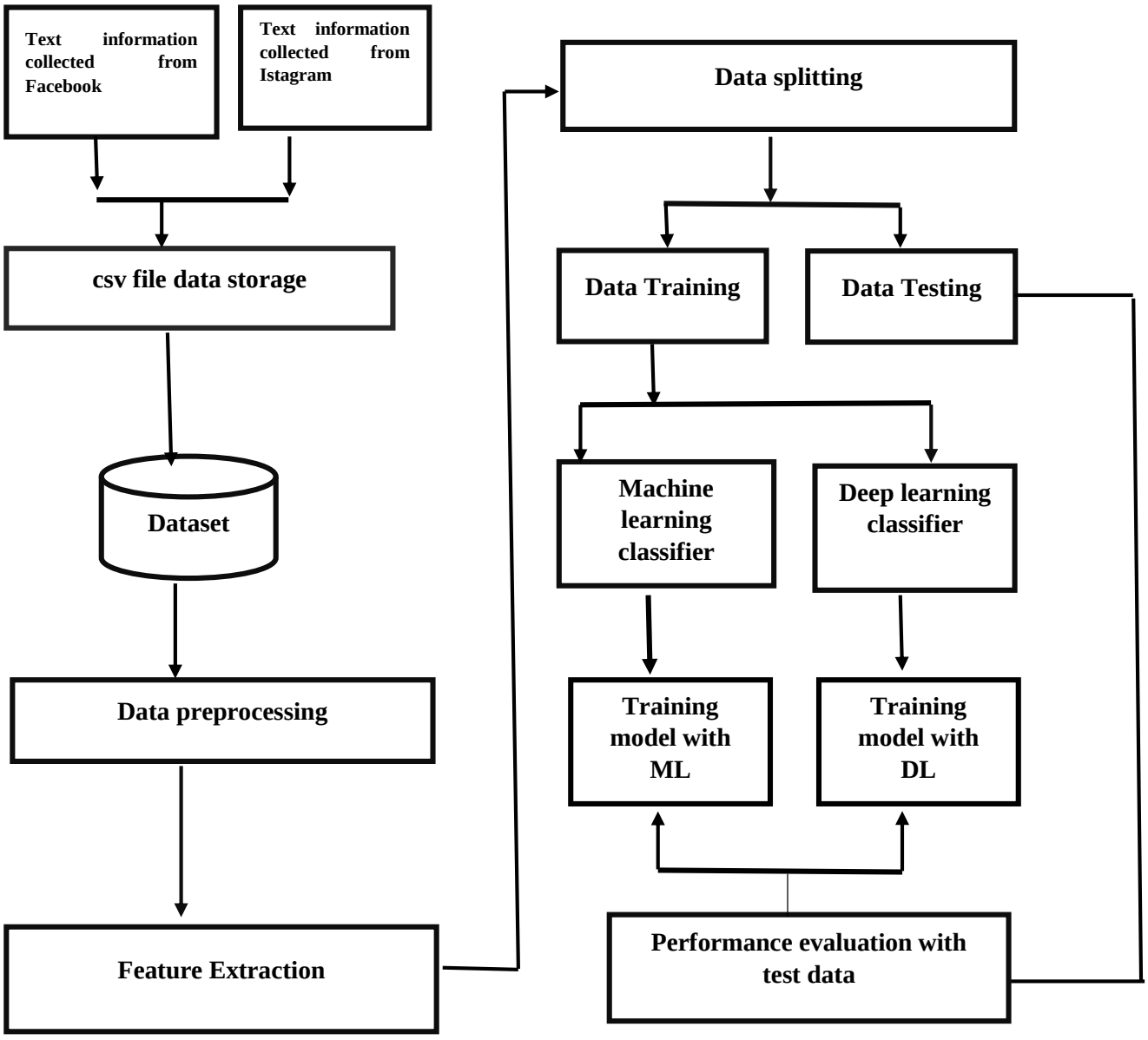

Figure 5. Flow chart of proposed methodology

3.2 Data Collection:

To use machine learning and deep learning techniques, we need a sizable amount of data to train our model. The data is the most important part of that. There are times when a significant amount of data that is more relevant to our enquiry can have an impact on a model's performance.

Social media has a vast amount of data. People can usually find poisonous comments here, and there are a lot of publicly accessible statistics. Therefore, we get information from publicly accessible comments on Facebook and Instagram. We use apify, a web application, as well as manual data collection .For this, we checked public posts where the majority of the comments were in Bangla. In doing so, we gather random texts. We checked public posts where the majority of comments were in Bangla for this. We gathered 16074 random texts in Bangla from Facebook and Instagram, with 85% of the comments coming from Facebook and 15% from Instagram. Then, for our research, we keep those remarks in a csv file.

We used supervised machine learning techniques in our research. Because supervised techniques work with labeled data, we can train and test our models and demonstrate how well they perform as classifiers. However, because the data from Facebook and Instagram were not categorized, we had to do it ourselves. We utilize a data dictionary to classify our data, which includes phrases that are frequently used in harmful online comments. We then personally review each label to ensure accuracy. That much massive data set's labeling required a lot of time and work. In our label dataset, 1 denotes data related to bullying, while 0 denotes data unrelated to bullying.

Table 1. Dataset Statistics

Category	Number
Total number of comments	16074
Total Comments collected from Facebook	14050
Total Comments collected from Instagram	2024
Total number of bullying comments	8488
Total number of bullying comments	7586

3.3 Data Preprocessing

A crucial step in machine learning is data processing. Data organization and cleaning are the main steps in data processing, which prepares raw data for use in a machine learning model. In conclusion, it can be said that data processing in machine learning is a type of data mining strategy where data is transformed into a form that is preferred by the machine learning model and is relevant and comprehensible.

As part of the processing procedure, data cleaning entails removing extraneous text and reducing noise from the data. The preprocessing steps involves some steps. The steps are the following:

I. **Tokenization:** The technique of tokenizing involves separating a sentence into its component words. Only words can offer the distinct meaning that each thing has, which is necessary in order to develop the model. It cannot be an elaborate sentence or a letter because each word has a unique significance. As a result, we treat the comments or text as paragraphs or sentences in this step and break them up into individual words.

II. **Stop words removal:** In the sentence, there are some words that doesn't convey any important or meaningful informatio9n.These are usually called stop words. In natural language stop words are commonly used preposition, quantifiers, linkers etc. Words like 'the', 'of'. 'is', 'in', 'am' etc are the most common word in English accent .Like this way in Bangla language there are similar kind of stop words as well for example: 'ও',' এবং', 'অথবা', 'কিন্তু' etc are such kind of stop words. From a github repositories we found a set of Bangla stop words [28] and remove those stop word from our raw sentence.

III. **Remove URLs:**The dataset contains a large number of urls, but none of them express any form of feeling or sentiment. Therefore, certain urls must be eliminated in order to improve the dataset's efficiency and speed up the system's performance.

IV. **Emoji Removal:** Emoji are essentially digital symbols that are frequently used in comments or communications to convey various emotions or feelings. It has a few facial icons with a range of expressions, such as joyful, sad, and emotional ones. These are some examples of those emotions: etc. Since we are only working with text data, we eliminate those emoji.

V. **Remove punctuation marks**: Bangla text uses a few punctuation marks. These include the comma (,), semicolon (;), question mark (?) , and others. It is preferable to omit the punctuation marks because they lack any form of emotion.

VI. **Word Correction**: Wrong spelling of words doesn't express the expression. That's why spelling correction is very important for the sentence. We used Contextual Spell checker for Bangla [36].

3.4 Feature Extraction

After preprocessing the data set, feature extraction step is involved in the proposed model. Feature extraction is the process which makes the raw data much more feasible form for machine because the machine can't process with the raw text directly in machine learning. So, by using some technique we convert the raw text into a matrix or vector. The transformed matrix or vector that we processed shows like the original form like before and in the same time it doesn't lose any information [37].

Feature extraction is also important because it reduces the unnecessary and common data from dataset. The reduction of data from dataset. The reduction of data helps the model to work with less attempt. That boost the speed of machine learning process.

3.4.1 Feature Extraction techniques

There are some feature extraction techniques. Some of most popular feature extraction methods are:

I. **BoW :** BoW stands for Bag of words. This is a common method of transforming token into feature. This is basically applied in document classification and here each individual word is considered as a feature and train the classifier.
II. **TF-IDF:** The full form of TF-IDF is term frequency inverse document frequency. In our project we applied TF-IDF for feature extraction .In the later section this briefly described.
III. **CountVectorizer:**Converting all words to lowercase and deleting any special characters before breaking a sentence or any text down into its individual words.

3.4.2 TF-IDF

TF-IDF is a quantitative standard that evaluate the importance of a word in a sentence form a set of sentences. In TF-IDF, the input data feature is extracted [38] and then avail them in the feature list. In TF-IDF is that when that works on any text, it gets the weights of the word with respects to the importance of that word in the sentence or documents.

In feature extraction, text data is transformed into a format that can be used in machine learning techniques by implementing feature extraction. To bring out that features, we implemented Term Frequency and Inverse Document Frequency. It deals with the text by determining the weights of words in a document. By this, they got numerical data and that data is put into the machine learning models.

Two parameter are used in that case:

1. How many times a term actually appeared in a document

2. The word inverse document frequency over a collection of documents.

Term Frequency (TF):

Term frequency (TF) is the number of times a word or term t appears in a document d.TF is basically the ratio of those two terms which is calculated by the following equation:

$$TF=\frac{t}{d} \quad \text{..... (1)}$$

Here t=frequency of word in a document

d=summation of all word in the document.

Inverse Document Frequency (IDF):

The relevance of a term is essentially determined by Inverse Document Frequency (IDF). IDF is calculated by dividing the total documents by the total papers that contain a specific phrase. The following equations give a description of it:

$$IDF=\log\left(\frac{N}{DFt}\right) \quad \text{(2)}$$

Here N=number of document

DFt= number of document that contains t term

The term frequency (TF) and inverse document frequency (IDF) are multiplied, and this is shown by the following equations:

$$TF\text{-}IDF=TF(t,d)*IDF(t) \quad \text{(3)}$$

Here t is the term of a document d in a document set.

Suppose, in a document, if we get a high term frequency and a low document frequency of a term in the entire document set, then at the time of tf -idf computation we get a high weight value [39]. That's how tf-idf plays an important role.

3.4 Training and Testing:

We must separate our dataset into a training and testing data set after the feature extraction procedure. This is necessary because, in the supervised machine learning technique, the train data is used to train the machine learning algorithm. After training, we test the model's effectiveness by using test data as input and comparing the model's output label to the label of the test data. As a result, we split the dataset into an 80:20 training and testing set.

Figure 6 represent data splitting into test and train set:

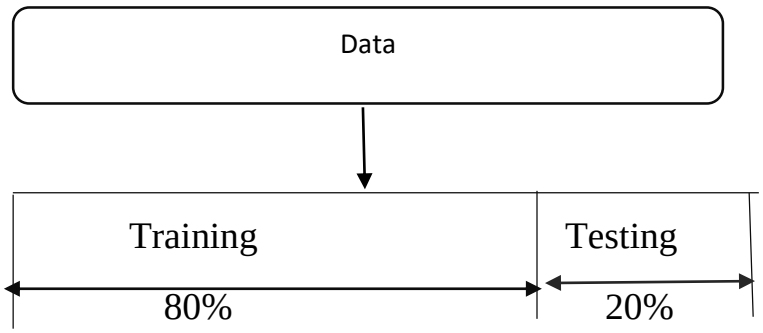

Figure 6. Dividing data into train and test set

A training dataset is a collection of data used to hone a model's ability to forecast future outcomes.

A testing data set is a set of data used to check if the model was trained correctly or not.

Table 2 below shows how many comments are included in various data sets.

Table 2. Number of data in train and test data set

Category	Number
Number of comments overall before splitting	16074
Total comments in the train set	12860
Total comments in test set	3214

We use training sets to develop our machine learning models, and testing sets to evaluate them.

CHAPTER 4

CLASSIFICATION & IMPEMENTATION

4.1 Machine Learning Classifiers:

Machine Learning classifier used the classify the data labels whether it is bullying data or not by ng machine learning algorithm. There are basically 4 different types of machine lasifier [40]. Those are

1.Binary Classification-Two class labels are used in the binary classification approach. whether a certain category is present or not, that is.

2.Multi Class Classification- Multi Class classifiers assign the item to a variety of classes or categories.

3.Multi Label Classification- An object is divided into several labels using this kind of classifier. This classifier initially categorized an object based on one category before reclassifying it based on a different category.

4.Imbalanced Classification- The distribution of the object inside the class is not equal in the imbalanced classifier.

Depending on the needs of the research, many classifier types are used. Based on the requirements, we constructed a Binary Classifier in our thesis to classify our data as either bullying or non-bullying.

Several machine learning methods can be used to categorize our data. These employ a total of eight machine learning algorithms. In those

- Naive Bayes
- Random Forest
- Decision Tree
- AdaBoost
- SGD Classifier
- Logistic Regression
- KNN
- SVM

Short Description about those algorithms given below:

(a) Naive Bayes:

A probabilistic machine learning technique called Naive Bayes or NB is used for classification tasks and is based on the Bayes Theorem. Bayes Theorem formed by a simple mathematical equation or formula that is used for computing conditional probabilities.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Here, P(A|B) is the probability of A occurring where B is already occurred.

P(A) is the probability of A occurring.

P(B) is the probability of B occurring.

P(B|A) is the probability of B occurring where A is already occurred

Narve Bayes calculate each label probability for a given text and select the label of height probability. By moving to different label this algorithm calculate whether this data belongs to that particularly category or not.

(b) Decision Tree:

Decision Tree or DT is a type of binary tree that split the dataset recursively till they reached the pure leaf node. That means, each data clearly falls on a particular category For example, there are two different types of nodes in a decision tree classifier. Those are

I. **Decision node:** Decision node contains the information or condition that split the data.
II. **Leaf node:** Leaf node doesn't contain any condition based on which data will be split. We can say, leaf node is a decision makings node. As, the particular data label is selected so don't need to split this node anymore.

The following figure represents the decision tree:

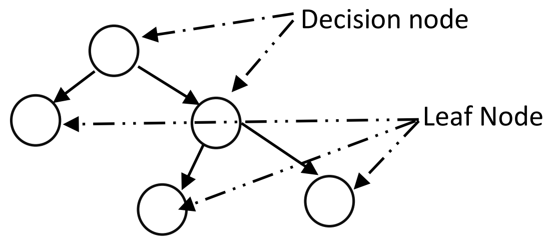

Figure 7.Decision Tree

(c) Random Forest

A supervised machine learning strategy called RF, commonly referred to as Random Forest, handles classification and regression problems. We do Bootstrapping in the random forest first, followed by Aggregation, also known as Bagging. This is how random forest is created. Because we employed two distinct random processes, such as bootstrapping and random feature selection, we may refer to this algorithm as random. We execute bootstrapping and feature selection because bootstrapping reduces the sensitivity of our model to the training data by not utilizing the same data point for every tree. Because all features would result in identical nodes in the decision tree, feature selection reduces correlations between the trees. Researchers found that values close to the log and sqrt of the total number of features works well.

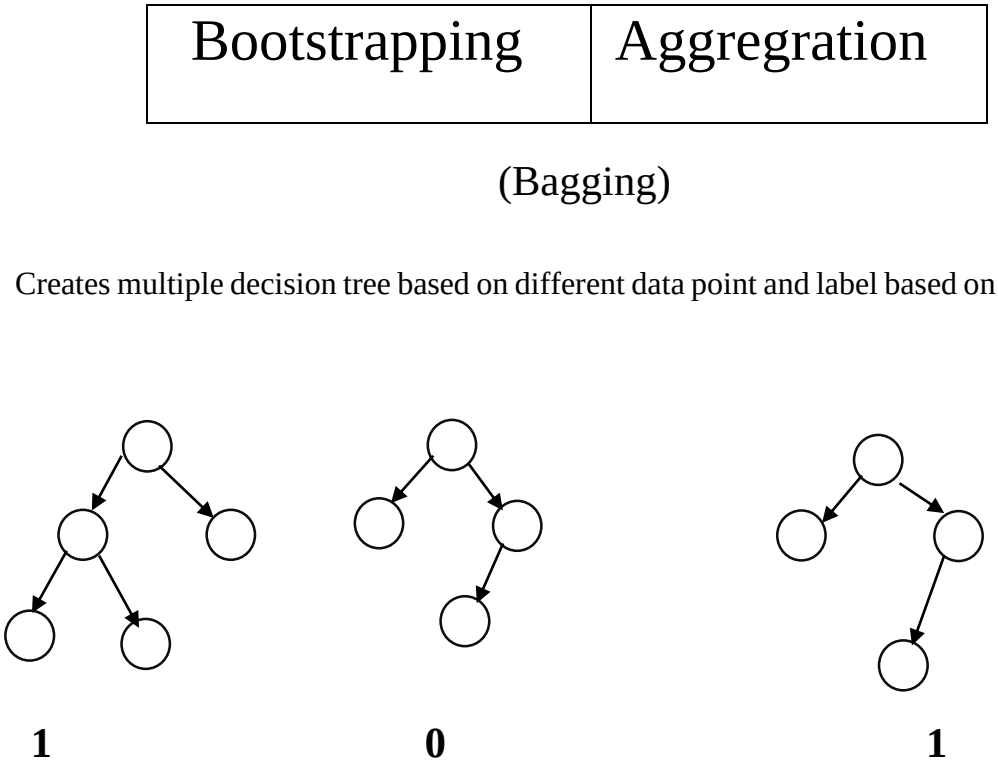

Figure 8.Bootstrapping and Aggregation in Random forest

(d) AdaBoost Classifier

AdaBoost classifier algorithm merge multiple classifiers to enhance the accuracy of the model. AdaBoost classifier is an iterative group of method. Basically, AdaBoost combines some classifier whose are bad in performance and by merging them it enhances the performance. AdaBoost classifier set weights to the classifier and in the continuous iteration

process it trains the data. When AdaBoost classifier tries to predict the label, there are errors, which means some data is incorrectly classified. Those errors are forwarded to the Base Learner. Base Learner are kind of decision tree which is not like random forest decision tree.Instead of that .it is created with one depth.

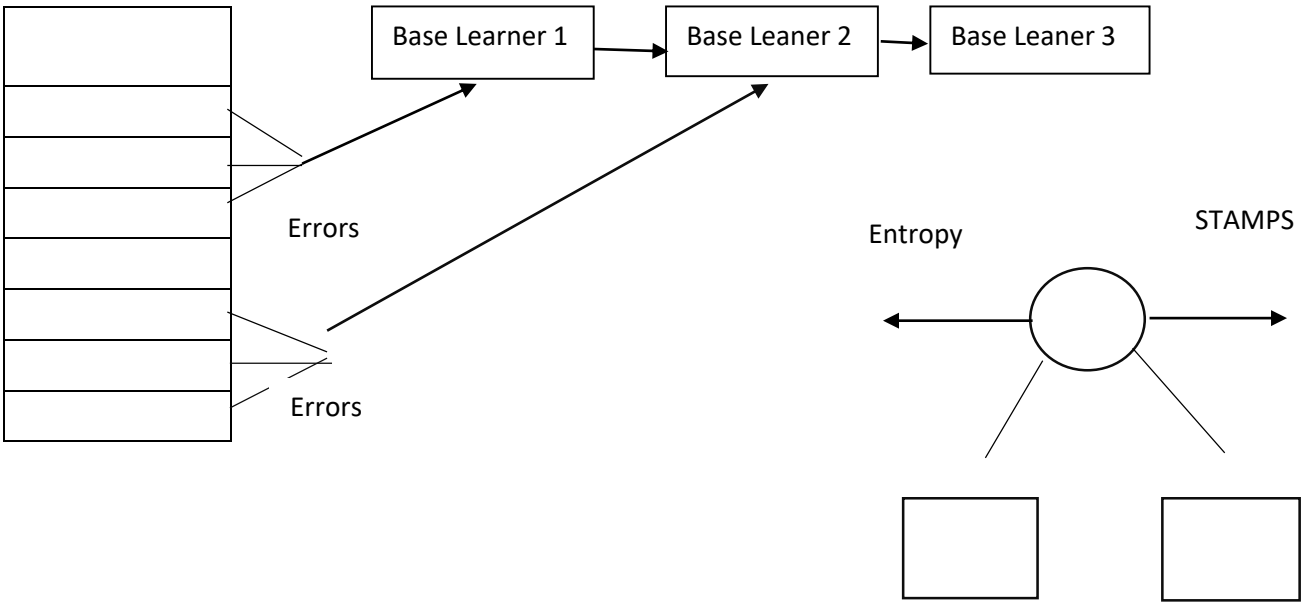

Figure 9. AdaBoost Classifier

(e) SGD Classifire:

SGD classifier for stochastic gradient.This is an optimizer algorithm that is used to finalize values of parameter to reduce loss function.

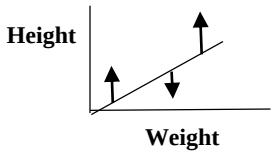

Figure 10. SGD Classifier

SGD classifire only focus on the loss function. Create the new weight by taking the derivative of the lost function in relation to the weight that has to be optimized.

(f) SVM

Support Vector Machine is the name of it. among the most prevalent machine learning methods is SVM. This approach is also a supervised learning approach. This is used to solve classification and regression-related issues. It is, nevertheless, frequently applied to categorization problems.

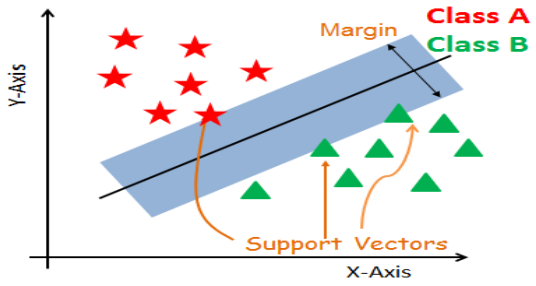

Figure 11.SVM Classifier

In support vector machine we have to select hyper plane in such a way that maximize the marginal plane. We want more generalized model to reduce the errors.

(g) Logistic Regression

A supervised machine learning algorithm is logistic regression. Instead of having the term regression in the name, it lies in the classification category Logistic regression deals with both continuous and discrete value but only discrete value gives the output.

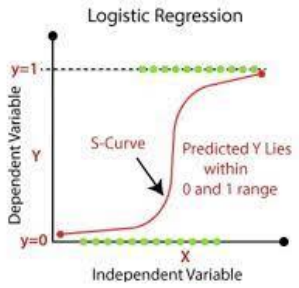

Figure 12. Logistic Regression Classifier

There is a function called sigmoid function. The function follows the following equation:

$F(x)= \frac{1}{1+e^{x}}$ ..... (5)

This function is called sigmoid function. By using that function, logistic regression draws the curve.

(h) KNN

K-Nearest Neighbor is referred to as KNN. KNN is a straightforward machine learning method that uses the Euclidean distance as its foundation. It calculates the distance between data points and close the point of nearest one.

Figure 13. KNN algorithm.

Category A and Category B are the two categories, as seen in the above image. To find the Kth neighbor, we also get a new piece of data. Determine the category for a new data item, and then observe which is the closet.

4.2 Deep Learning Classification Model:

Our data can be classified using a variety of deep learning algorithms. We employ two deep learning algorithms among them. They are:

- CNN Classifier
- LSTM Classifier
- CNN and LSTM combined classifier.

Short Description about those algorithms given below:

(a)CNN

Convolution Neural Network, or CNN Artificial neural networks (ANN) work well for machine learning. ANN used for text, image, and other classification. CNN and ANN both have the same parameters. We use CNN to create a model in our research study.

(b) LSTM

Long Short-Term Memory is LSTM. A challenging area of deep learning is LSTM. The LSTM is a type of recurrent neural network (RNN) that has the ability to learn on its own and remember long-term dependencies. Because LSTM retains knowledge across time, it can recollect the prior information for extended periods of time. They aid in time series prediction because they can recall past input.

(c) CNN-LSTM combined classifier

In our research we proposed a model that is combined of CNN and LSTM deep learning algorithms. The architecture of CNN and LSTM involves using for feature extraction for feature extraction by using Convolutional Neural Network (CNN) layers then combined with LSTM to support order prediction

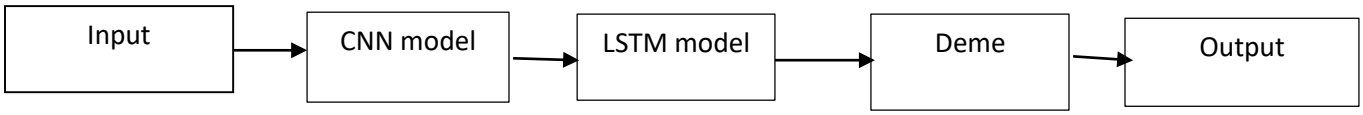

Figure 14. CNN-LSTM architecture

Here two different model is used. CNN for feature extraction and LSTM is for explaining the feature with respect to time.

4.3 Jupyter Notebook

Jupyter notebook is a web application that is used for generating and spreading computing documents [40]. Everyone can use this web-based platform for free. The creation of open-source software, standards, and services is accomplished through this effort. We may now share documents that contain text, code, equations, and visualizations. A fork of the IPython project, Jupyter Notebooks once had its own IPython Notebook project. The three programming languages that Jupyter supports. They are

• Julia,

•Python

•R

Jupyter is named from them. Jupyter supports IPython kernel, that allow users to develop Python programs.

4.4 Python Software

The Python community is very large and diverse. It is much welcomed that this community wants to expand to include a broad population. Many websites employ Python APIs (application programming interfaces) as their end point because Python is used in web development. Python is crucial to machine learning and data science, additional to web development.

Different Python modules allow to access the data and analyze them as well as transfer them when necessary. Python may also be used for Natural Language Processing. Machine learning had to deal with data continuously. Although it is a well-known Python toolkit for NLP implementation using Python, the Natural Language Toolkit (NLTK) is somewhat difficult to comprehend. On the

other hand, Text Bolb is also a python library which process text data. Text Bolb also provide API which is used for diving the NLP or natural language processing tasks such as sentiment analysis, translation, classification and many more.

4.5 Data Analysis Using Jupyter Notebook

We preprocessed and labeled the data after gathering text data from Facebook and Instagram. We can retrieve our data in csv format and import it into the Jupyter notebook. The screenshot of a little amount of data is provided below.

	text	vulgar	hate	religious	threat	troll	Insult
0	প্রধানমন্ত্রী হক সাহেবের ক্ষতি হলে জাতির স্বার...	0	0	0	1	0	0
1	আমি বললাম, 'দেন'	0	0	0	0	0	0
2	অসাধারণ তানজিন তিশা আমার বালো লাগার একজনকাতার ...	0	0	0	0	0	0
3	তার উপর ২ জন মেয়র	0	0	0	0	0	0
4	পলাশের কাজ এতো ভালো হবে কল্পনাও করি নাই তৌহিদে...	0	0	0	0	0	0

Figure 15. Sample of the labeled dataset

We can see that multi label distribution in dataset through pie chart for more clear visualization

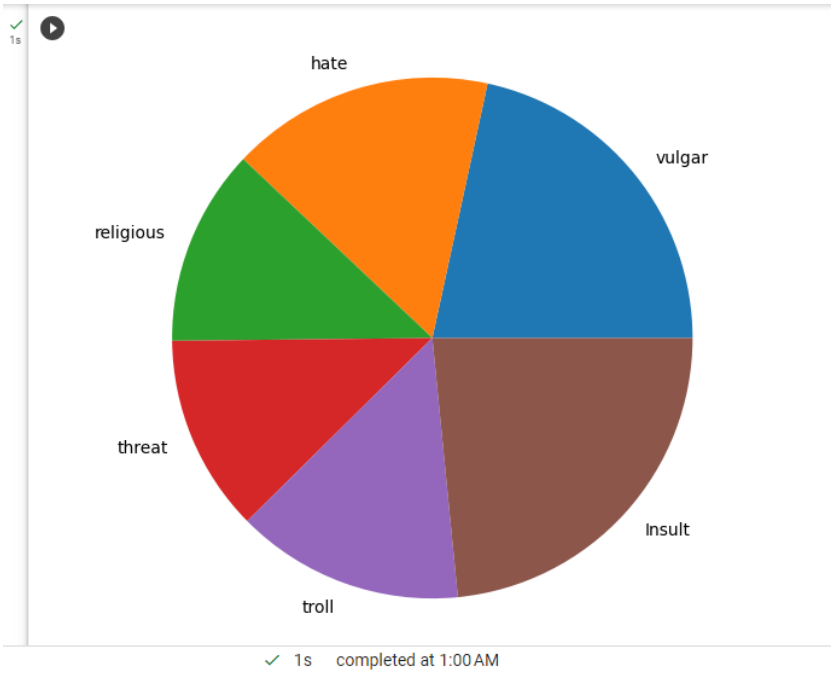

Figure 16. A pie chart of multi-label data distributions

CHAPTER 5

RESULT & DISCUSION

5.1 Performance Analysis

Our deep learning and machine learning models were both trained using the train data set. Now, using the test dataset, we must test those models. We'll use a few performance metrics to try and evaluate the performance. The performance indicators are:

- Accuracy
- Precision
- Recall
- . Fl-score

Accuracy

The percentage of predictions made by our model that were accurate is called accuracy. The most widely used criteria for performance analysis are. This truly gauges the accuracy of our model. This represents the ratio of accurate positive and negative forecasts to all predictions. The equation is as follows

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad \text{(6)}$$

Precision

Precision describes the proportional data points that were both relevant and, according to our model, existed in the pertinent class. It is the proportion between actual positive results and all positive projections. The formula is as follows:

$$\text{Precision}=\frac{TP}{TP+FP} \quad \text{(7)}$$

Recall

Recall is a measurement of a model's ability to find all the data points in the data set that belong to the class in which we are interested. In essence, it is a true positive ratio. It establishes the proportion of actual positives that were correctly predicted out of all anticipated positive events. It results from the ratio of true positives to false positives and negatives. The equation reads as follows:

$$\text{Recall}=\frac{TP}{TP+FN} \quad \text{(8)}$$

FI-Score

FI-Score basically calculates the harmonic mean of precision and recall

Here, we select B=1, when both false negative and false positive are important. It goes by following equation It goes by following equation:

F1-score$\frac{2*\text{precision}-\text{recall}}{\text{recall precision recall}}$..... (9)

Here, TP True Positive,

TN- True Negative,

FP- False Positive,

FN- False Negative.

5.2 Result and Discussion:

The outputs of our model will be displayed in this part, along with an explanation of the results. We have determined the accuracy, precision, recall, and f1-score of our model. The table below allows us to assess how well machine learning classifiers worked:

Table 3. Performance of machine learning classifier.

Algorithm	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	0.76	0.76	0.76	0.76
Random Forest	0.78	0.78	0.78	0.78
Decision Tree	0.75	0.75	0.75	0.75
AdaBoost	0.75	0.74	0.71	0.75
SGD	0.76	0.77	0.77	0.77
Logistic Regression	0.78	0.78	0.78	0.78
KNN	0.70	0.71	0.70	0.70
SVM	0.79	0.78	0.79	0.77

In the above Table 3 we can see the performance of our machine learning classifier. Among those different machine learning algorithm, Naïve Bayes algorithm achieves 76% accuracy, precision, recall and fl-score. From Random Forest classifier, we got 78% accuracy and recall, precision and fl-score. On the other hand, from Decision Tree algorithm, we got 75% accuracy and recall, precision and fl-score. Then, for AdaBoost classifier, we got 75% accuracy, 74% precision, 71`% recall and 75% fl-score. From SGD classifier, we got 76% accuracy, 77% precision, recall and fl-score. Logistic Regression algorithm achieves 78% accuracy, precision, recall and fl-score. For KNN algorithm, we got 70 % accuracy, 71 % precision, 70% recall and 70% fl-score. Finally, from SVM algorithm we achieve 79 % accuracy, 78% precision, 79% recall and 77% fl-score which performs comparatively better than all other machine learning algorithms.

We shall see the effectiveness of our deep learning classifiers in the table below:

Table 4. Performance of deep learning classifier.

Algorithm	Accuracy	Precision	Recall	F1-score
CNN	0.89	0.88	0.87	0.89
LSTM	0.89	0.88	0.88	0.88
CNN+LSTM	0.89	0.89	0.89	0.89

In the above Table 4 we can see the performance of our deep learning classifier. Among those different deep learning algorithms, CNN algorithm achieves got 89 % accuracy, 88 % precision, 87% recall and 89% fl-score. From LSTM classifier, we got 89 % accuracy and 88% recall, precision and fl-score. Finally, from CNN and LSTM combined classifiers, we achieve 89% accuracy, precision, recall and fl-score which is comparatively more balanced than all other deep learning algorithm.

After analyzing all different machine learning and deep learning classifier, we can see deep learning classifier performs comparatively batter in that case.

Below we can see the comparison of different classifiers based on individual performance parameters:

5.2.1 Accuracy

The most widely used performance metrics for performance evaluation are accuracy. Below is a graph that displays the accuracy of various classifiers.

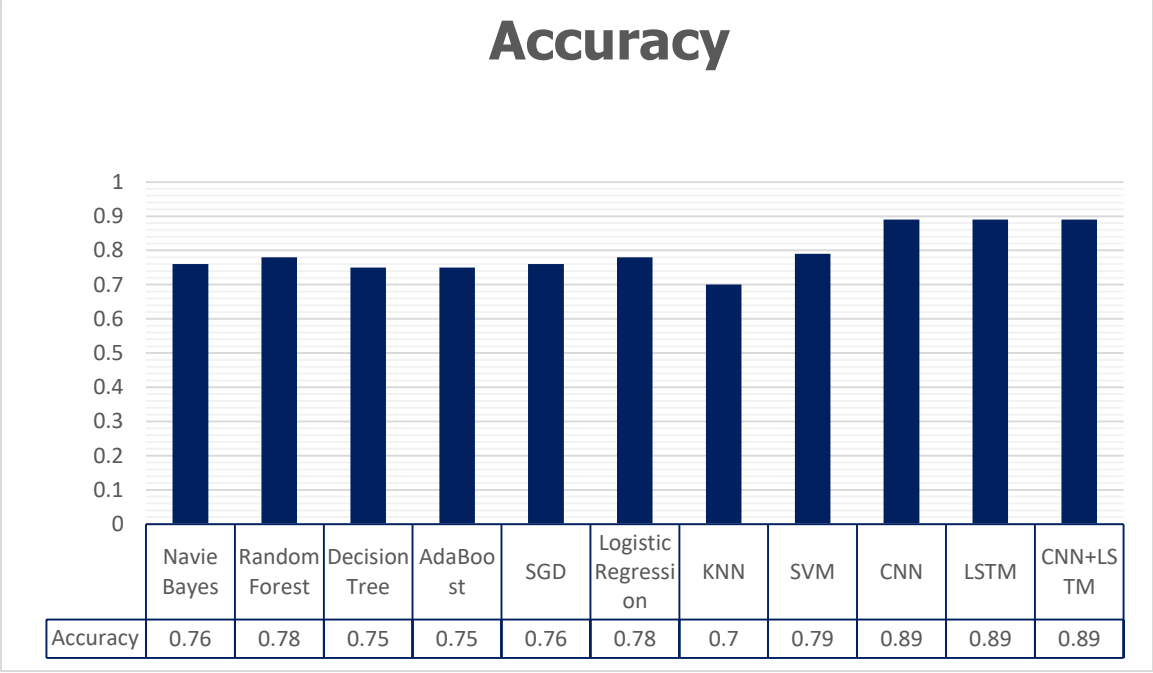

Figure 17. Accuracy of Different Classifier

In the above figure, we can see accuracy of different classifiers. Here, Naive bayes achieved 76% accuracy, Random Forest 78%, Decision Tree 75%, AdaBoost Classifier 75%, SGD Classifier 76%, Logistic Regression 78%, KNN 70%, SVM 79%, CNN 89%, LSTM 89% and CNN+LSTM 89%.

After the analysis of the result, we can see, in our research, Deep learning algorithm achieves better accuracy than machine learning algorithms. And CNN, LSTM and CNN+LSTM performs equally well.

5.2.2 Precision

Precision is an important parameter in performance analysis. In the below figure, we can see precision of different classifiers. Here, Naive bayes achieved 76% accuracy, Random Forest 78%, Decision Tree 75%, AdaBoost 74%, SGD 77%, LR 78%, KNN 71%, SVM 78%, CNN 88%, LSTM 88% and CNN+LSTM 89%.

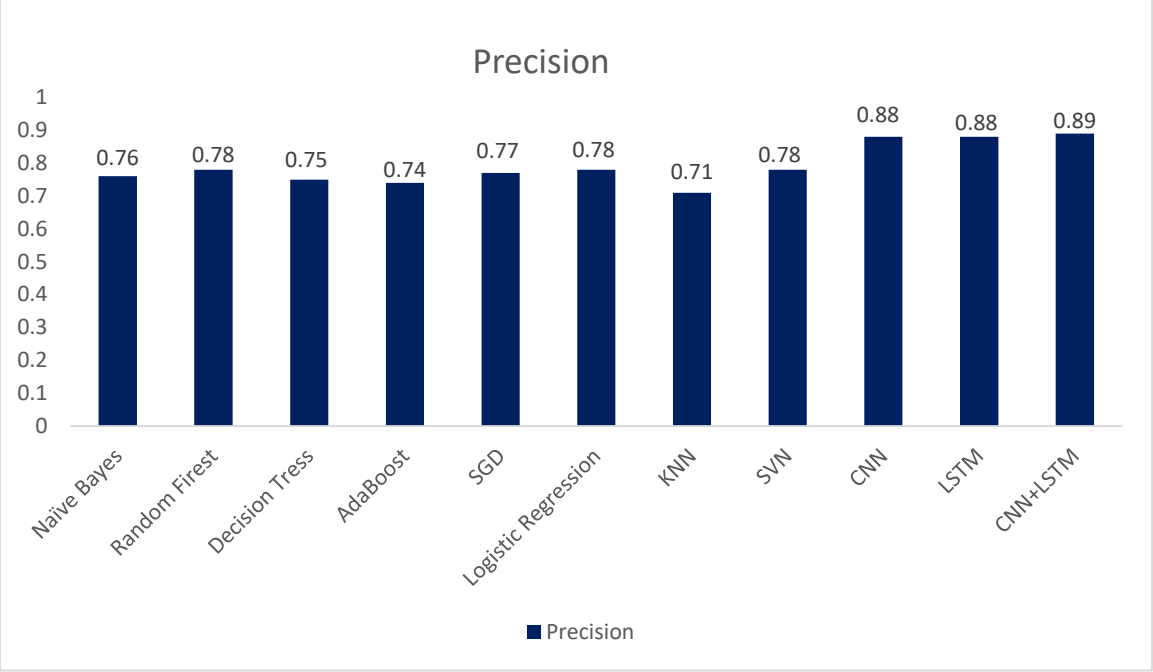

Figure 18. Precision of Different Classifier

After the analysis of the result, we can see, in our research, Deep learning algorithm achieves better precision than machine learning algorithms. And by using CNN+LSTM we achieve best precision.

5.2.3 Recall

Recall is an important parameter in performance analysis.

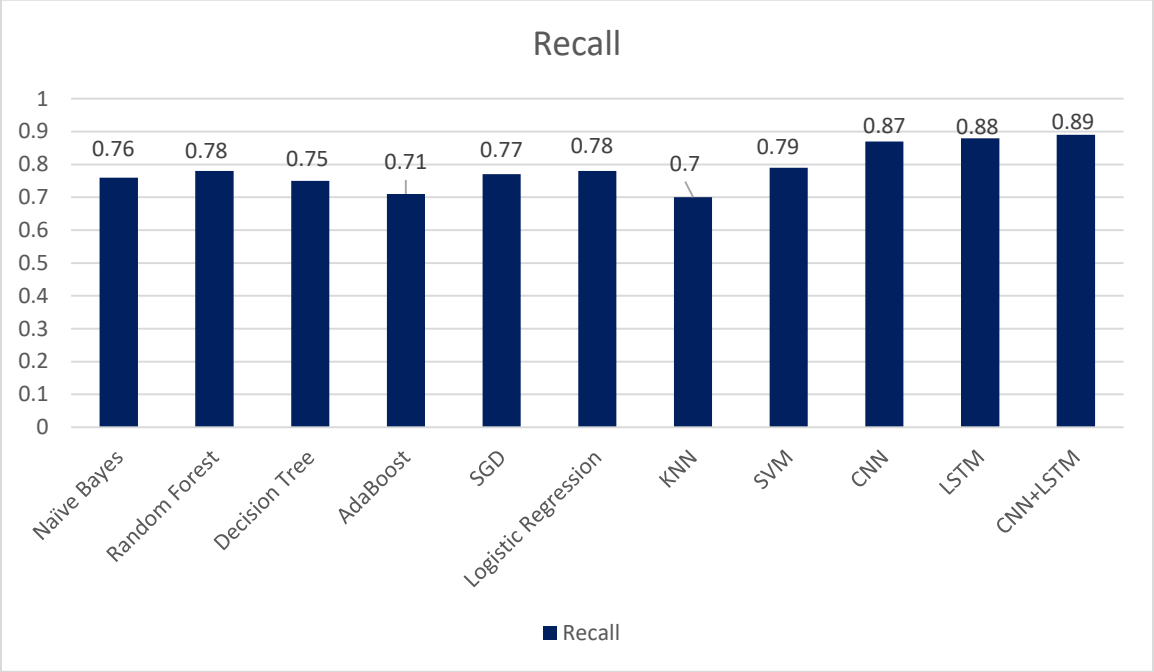

Figure 19. Recall of different Classifier

In the above figure, we can see accuracy of different classifiers. Here, Naive bayes achieved 76% accuracy, Random Forest 78% Decision Tree 75%, AdaBoost 71% SGD Classifier 77%, Logistic Regression 78% KNN 70% SVM 79% CNN 87%, LSTM 88% and CNN LSTM 89%

After the analysis of the result, we can see, in our research, Deep learning algorithm achieves better recall than machine learning algorithms. And by using CNN+LSTM we achieve best recall.

5.2.4 F1-Score

F1-Score is an important parameter in performance analysis. Fl-Score basically calculates the harmonic mean of precision and recall.

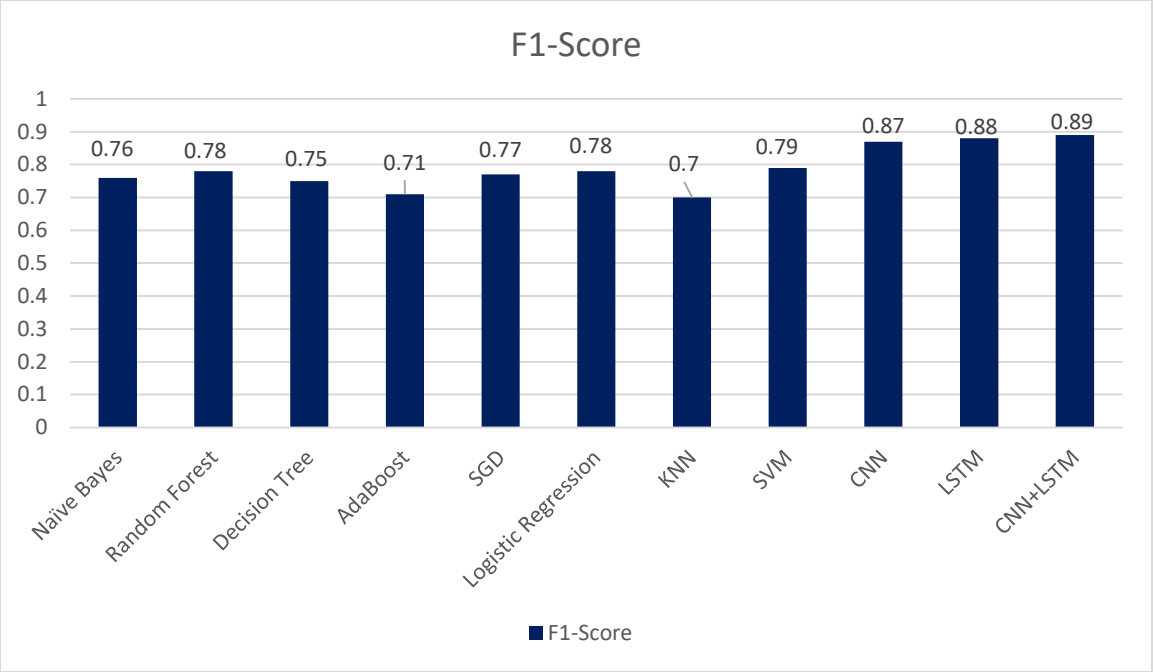

Figure 20. F1-Score of Different Classifier

In the above figure, we can see accuracy of different classifiers. Here, Naive bayes achieved 76% accuracy, Random Forest 78%, Decision Tree 75%, AdaBoost Classifier 71%, SGD Classifier 77%, Logistic Regression 78%, KNN 70% SVM 79%, CNN 89%, LSTM 88% and CNN+LSTM 89%.

After the analysis of the result, we can see, in our research, Deep learning algorithm achieves better recall than machine learning algorithms. And by using CNN and CNN LSTM we achieve best Fl-Score.

CHAPTER 6

CONCLUSION & FUTURE WORK

6.1 Conclusion

Almost everyone is active on numerous different social media sites, and social media use has become a crucial aspect of daily life. As a result, users, especially teens, frequently experience online abuse in the form of harmful comments. We evaluate and analyze various research projects that have been carried out by researchers as part of the research aim. And it is obvious from their efforts that they are attempting to effectively address the bullying issue on various social media platforms.

To our knowledge, many attempts have been done to date to discover offensive statements in Bangla. We look for the most effective means of quickly identifying poisonous comments. Therefore, our objective was to create a model that could effectively identify harmful remarks with multiple labels in Bangla language.

We gathered Bangla text from comments on Facebook and Instagram to create the dataset. In total, there were 16530 of them. Out of this, 46% were bullying-related statistics and 54% were not. Before training our data, we employed the TF-IDF approach for feature extraction.

It can be difficult to find harmful comments on several social media networks. Additionally, it is more difficult for Bangla text because various people from different regions use it differently and because there are more spelling errors in comments.

In spite of these difficulties, the CNN+LSTM combined method in deep learning yields the highest accuracy as well as other performance metrics including precision, recall, and fl-score. With the combined CNN+LSTM approach, we obtained 89% accuracy, precision, recall, and fl-score. SVM method obtains 79% accuracy, 78% precision, 79% recall, and 77% fl- score in machine learning classifier.

6.2 Future Work

Improvement and Development in research is a continuous process. With the time being we tries to improve ourselves. After completing that research, we figure out, there are some scopes in where we can make improvement but because of the shortage of our time we couldn't make it. So, we will implement those features in our future work.

Below we can see some scope where we can develop this research near future:

- We'll work to increase the dataset's overall data count.
- Now a days, a lot of Romanized text are being used. Will add Romanized text along with Bangla text in the dataset.
- . We will reorganize our model in order to make it more efficient and to make the detection system automated.

- To compare the performance of several machine learning and deep learning algorithms, other ones will be implemented.
- We will add new features to improve the accuracy of the findings.
- We will try to implement an extension for browser and app for mobile users so that it could identifies the person who harass other in different social media platforms.
- To test our data, we can utilize a reinforcement machine learning model.

We believe, the suggestions that we presented in this section can assist in making the social media world a lot more secure and safe environment for the general users. We also believe that the research community will find our thesis findings beneficial in future investigations.

References

[1] A. T. Abhishek Aggarwal, "Multi Label Toxic Comment Classification using," *International Journal of Recent Technology and Engineering (IJRTE),* Vols. Volume-10 , no. Issue-1, p. 158, May 2021.

[2] "Global social media statistics research summary 2022", dataportal [online]. Available:https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research.

[3] "Digital 2022: Bangladesh", dataportal [online]. Available: https://datareportal.com/reports/digital-2022-bangladesh.

[4] "Bangali Language at Ethnologue" https://www.ethnologue.com/language/ben.

[5] "Summary by language size" Ethnologue. 2019. Archived from the original on 24 April 2019.

[6] A. T. Abhishek Aggarwal, "Multi Label Toxic Comment Classification using Machine Learning Algorithms," *International Journal of Recent Technology and Engineering (IJRTE),* pp. 158-161, May 2021.

[7] S. Ding, H. Zhu, X.-L. Liu, and L. Zhang," Appropriateness in applying SVMs to text classification ". Comput Eng Sci, vol. 32, pp. 106-108, 2010.

[8] S. Shuicai, L. Kun, L. Xueqiang, and L. Zhijie," Study on SVM Compared with the other Text Classification Methods" - in Education Technology and Computer Science (ETCS), 2010 Second International Workshop on, 2010, pp. 219-222.

[9] Shikhar Kumar Sarma, Gogoi, and Moromi," Document classification of Assamese text using Naïve Bayes approach.". International Journalof Computer Trends and Technology 30.4 (2015):1-5.

[10] Palanivel, S., Palanivel, S., Ramalingam, V., Ganesan, M., and Rajan, K. Automatic Classification of Tamil documents using Vector Space Model and Artificial Neural network" Expert System with Applications, Elsevier, Volume 36 Issue 8, DOI= 10.1016/j.eswa.2009.02.010(2009).

[11] JI Sheeba and B Nandhini."Cyberbullying detection and classification using information retrieval algorithm" In Proceedings of the 2015 International

Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015), page 20. ACM, 2015.

[12] William J. Buchanan, Adrian Smales, Gordon Russell, and Shane Murnion," Machine learning and semantic analysis of in-game chat for cyberbullying." Computers & Security, 76:197– 213, 2018.

[13] Kodchakorn na Nakornphanom, Pimpaka Prasertsilp, Piyaporn Nurarak, and Pirom Konglerd, Walisa Romsaiyud." Automated cyberbullying detection using clustering appearance patterns" In Knowledge and Smart Technology (KST), 2017 9th International Conference on, pages 242– 247. IEEE, 2017.

[14] Livia Ashianti, Sani Muhamad Isa," Cyberbullying classification using text mining." In Informatics and Computational Sciences (ICICoS), 2017 1st International Conference on, pages 241–246. IEEE, 2017.

[15] Maroun Chamoun, Ahmed Serhrouchni, and Batoul Haidar, "A multilingual system for cyberbullying detection: Arabic content detection using machine learning." Advances in Science, Technology and Engineering Systems Journal, 2(6):275–284, 2017.

[16] Edward Dillon, Jamie Macbeth, Hongxin Hu, Jonathan Tong, Nishant Vishwamitra, Elizabeth Whittaker, Joseph P. Mazer, Robin Kowalski, and Zhang." Cyberbullying detection with a pronunciation based convolutional neural network." In 2016 15thIEEE International Conference on Machine Learning and Applications(ICMLA), pages 740–745. IEEE, 2016.

[17] Franciska De Jong and Maral Dadvar."Cyberbullying detection: a step toward a safer internet yard." In Proceedings of the 21st International Conference on World Wide Web, pages 121–126. ACM, 2012. [18] Franciska De Jong and Maral Dadvar," Improved cyberbullying detection using gender information", Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012). University of Ghent, 2012.

[18] Franciska De Jong and Maral Dadvar," Improved cyberbullying detection using gender information", Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012). University of Ghent, 2012.

[19] Franciska De Jong and Maral Dadvar,” Improving cyberbullying detection with user context”, In European Conference on Information Retrieval, pages 693–696. Springer,2013.

[20] Franciska De Jong and Maral Dadvar,” Experts and machines against bullies: A hybrid approach to detect cyberbullies.”, In Canadian Conference on Artificial Intelligence, pages 275– 281. Springer, 2014.

[21] Md. Tofael Ahmed, Maqsudur Rahman, Shafayet Nur, Abu Zafor Muhammad Touhidul Islam, Dipankar Das,” Natural language processing and machine learning based cyberbullying detection for Bangla and Romanized Bangla texts”, TELKOMNIKA Telecommunication Computing Electronics and Control, Vol. 20, No. 1, February 2022, pp. 89~97

[22] Ahmed, Md Faisal, et al. "Cyberbullying detection using deep neural network from social media comments in bangla language." arXiv preprint arXiv:2106.04506 (2021).

[23] Hussain, Md Gulzar, Tamim Al Mahmud, and Waheda Akthar. "An approach to detect abusive bangla text" 2018 International Conference on Innovation in Engineering and Technology (ICIET). IEEE, 2018.

[24] Talpur, Bandeh Ali, and Declan O'Sullivan. "Cyberbullying severity detection: A machine learning approach." PloS one 15.10 (2020): e0240924.

[25] Haque, Fabliha, Md Motaleb Hossen Manik, and M. M. A. Hashem. "Opinion mining from bangla and phonetic bangla reviews using vectorization methods. 2019 4th International Conference on Electrical Information and Communication Technology (EICT), IEEE, 2019.

[26] Ahmed, Md Tofael, et al. "Natural language processing and machine learning based cyberbullying detection for Bangla and Romanized Bangla texts." TELKOMNIKA (Telecommunication Computing Electronics and Control) 20.1 (2021): 89-97,

[27] Deep Learning VS. Learning Machine What's The Difference?

https://levity.ai/blog/difference-machine-learning-deep-learning

[28] D. Gene, Stopwords Bengali (BN), (2016), GitHub repository link: github.com/stopwords-iso/stopwords-bn/

[29] Maqsudur Rahman, Tofael Ahmed, Md. Iqbal Hossain, A. Z. M. Touhidul Islam, ”Forecast the Rating of Online Products from Customer Text Review based on Machine Learning Algorithms”, 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 27-28 February, Dhaka

[30] Md. Tofael Ahmed, Maqsudur Rahman, Shafayet Nur, Azm Islam ,Muhammad Touhidul Islam, Dipankar Das, “Deployment of Machine Learning and Deep Learning Algorithms in Detecting Cyberbullying in Bangla and Romanized Bangla text: A Comparative Study”, 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-7281-5791-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAECT49130.2021.9392608

[31] Md. Tofael Ahmed, Maqsudur Rahman, Shafayet Nur, Azm Islam ,Muhammad Touhidul Islam, Dipankar Das, “Introduction of PMI-So Integrated with Predictive and Lexicon Based Features to Detect Cyberbullying in Bangle Text Using Machine Learning”, Proceedings of 2nd International Conference on Artificial Intelligence: Advances and Applications,ICAIAA 2021

[32] Maqsudur Rahman, Pushpita Shil, Umma Saima Rahman, Md Shariful Islam, “An Approach for Detecting Bangla Spam Comments on Facebook”, 2021 International Conference on Electronics, Communications and Information Technology (ICECIT)

[33] Vaibhav Suri and Sourabh Parime, “Cyberbullying detection and prevention: Data mining and psychological perspective”. In Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on,pages 1541–1547. IEEE, 2014.

[34] Sencun Zhu, Yilu Zhou, Ying Chen, and Heng Xu. “Detecting offensive language in social media to protect adolescent online safety”. In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), pages 71–80. IEEE, 2012.

[35] I-Hsien Ting, Giovanny Mauricio Tarazona Bermudez, Wun Sheng Liou, Dario Liberona, Shyue-Liang Wang, “Towards the detection of cyberbullying based on social network mining techniques.” In Behavioral, Economic, Socio-cultural Computing (BESC), 2017 International Conference on, pages 1–2. IEEE, 2017.

[36]Lifeparticle,https://GitHub.com/lifeparticle/BengaliAlphabet#puncutation(Git Hubrepository.

[37] Partha C. Ahmed S, Yousuf MA, Azad A, Alyami SA, Moni MA (2021) A human-robot interaction system calculating visual focus of human's attention level.IEEE Access 9:93409- 93421.

[38] Akiko Aizawa. An information-theoretic perspective of tf-idf measures. Information Processing & Management, 39(1):45-65, 2003

[39] Ahmed, Md Tofael, et al. "Deployment of machine learning and deep learning algorithms in detecting cyberbullying in bangla and romanized bangla text: A comparative study." 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE, 2021.

[40] Open Standards for Interactive Computing, link: https://jupyter.org/

[41] Zhang, Xiang, et al. "Cyberbullying detection with a pronunciation based convolutional neural network." 2016 15th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2016.