RECOGNISING CYBERBULLYING THROUGH MACHINE LEARNING TECHNIQUES: A COMPARATIVE ANALYSIS

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted by:

Mayank Singh Sajwan

2K21/ISY/14

Under the supervision of

Dr Jasraj Meena

Assistant Professor

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi

May 2023

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi

CANDIDATE'S DECLARATION

I, Mayank Singh Sajwan, Roll No. 2K21/ISY/14 of M.Tech. (Information Systems), hereby declare that the dissertation report titled "RECOGNISING CYBERBULLYING THROUGH MACHINE LEARNING TECHNIQUES: A COMPARATIVE ANALYSIS" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship, or other similar title or recognition.

Place : Delhi Mayank Singh Sajwan

Date :

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi

CERTIFICATE

I, hereby certify that the dissertation which is submitted by Mayank Singh Sajwan, Roll No. 2K21/ISY/14 (Information Systems), Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place : Delhi

Date :

Dr. Jasraj Meena

SUPERVISOR

Assistant Professor

Department of InformationTechnology

Delhi Technological University

ABSTRACT

Millions of people use social media to communicate and share information with one another. Facebook, Twitter, and other social media platforms provide the ability to interact and converse with anybody, at any point in time, and with a huge group of individuals. On a global scale, social media is used by over three billion people. It is a 100% web-based platform that is constantly growing and evolving. According to the National Crime Prevention Council, cyberbullying occurs when individuals use their phones, online game programmes, or other electronic devices to email or send text, photographs, or videos with the intent of intentionally injuring or humiliating others (NCPC).

Cyberbullying can occur at any time of day or week and can affect anyone who is online. Cyberbullying text messages, photographs, and videos can be anonymously posted and instantly distributed to a large audience. Maintaining track of those posts' reassessments may be difficult, if not impossible. Additionally, it is no longer possible to delete such communications after a specified time period has passed. Certain social networking sites offer a guide to avoiding cyberbullying. To protect your data, Facebook provides a section dedicated to detailing how to report cyberbullying and block the attacker. Users on Instagram can unfollow or block anyone who posts photographs or videos that make them feel uneasy. Violations of the Community Guidelines can also be reported directly from the app. According to Twitter, the consumer should be penalized for inappropriate, abusive, rude, or threatening behavior. Cyberbullying has been linked to social, emotional, and educational difficulties in adolescents, including despair and estrangement, as well as an increased risk of self-harm, including suicide. Numerous organizations are attempting to raise awareness about cyberbullying. In this thesis, we tested multiple machine learning models on text-based datasets.

Keywords: cyberbullying; hate speech; offensive language; machine learning;

ACKNOWLEDGEMENT

I am grateful to Prof. Dinesh Kumar Vishwakarma, HOD (Department of Information and Technology), Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, and all other faculty members of our department for their astute guidance, constant encouragement, and sincere support for this project work.

I would like to take this opportunity to express our profound gratitude and deep regard to our project mentor Dr Jasraj Meena, for his exemplary guidance, valuable feedback, and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout our project work. His perspective and criticism kept us working to make this project in a much better way. Working under him was an extremely knowledgeable experience for us.

We would also like to give our sincere gratitude to all our friends for their help and support.

Mayank Singh Sajwan

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Cyberbullying detection have been existed to identify negative comments, harmful messages, and mean or false content about someone. If someone shares or spreads private information and makes them embarrassed or humiliated, then these types of Cyberbullying can be detected using machine learning. Cyberbullying sometimes crosses the limit and means illegal or criminal behavior. However, this illegal activity can be detected by machine learning. The section presents the research's motivation in light of the primary objectives, research questions, and possible explanations. In addition, the motive for this is supplied in this section to demonstrate the problem spots suitable to the investigation, observed by the significance of the investigation. The general framework of the study is also supplied in this chapter. This analysis shall deliver the different aspects of machine learning to detect the different Cyberbullying activities.

Figure 1.1: Different types of Cyberbullying

(Source: https://theaseanpost.com/article/cyberbullying-rise)

1.2 BACKGROUND OF STUDY

Cyberbullying Detection looks for negative comments in user messages using a variety of codes and a Machine learning algorithm. Before deciding whether or not the comment should be changed, the algorithm gives the message a value using trained data. In the event that it is absolutely difficult, the framework will look through the confusing group of users to determine how each user interacts with other people in general and with the end user as a whole. Using this information, the system will determine whether the message needs to be altered. A number of models are used to change the sentence's negative parts into positive ones if this is the case. The initial algorithms then examine the transformed sentence. It is assigned a value, and the framework will continue to send the changed positive sentence to the end customer in the event that the value results in a positive sentence. Otherwise, the models will be used to insert the sentence once more. Through a developed web front end, users communicate with a central server. The users are referred to as clients. If any messages are altered, the user who is receiving them will also be informed.

1.3 OBJECTIVE

The objectives which can be attained are:

- To choose an appropriate machine learning method for detecting cyberbullying.
- To develop and improve the effective machine learning algorithm used for cyberbullying detection.
- To improve the machine learning algorithm's effectiveness and efficiency in detecting cyberbullying.

1.4 QUESTIONS

The following questions must be framed in light of the study's main objective and in order to comprehend how machine learning is used to identify cyberbullying:

1. Which is the suitable algorithm for detecting cyberbullying?
2. How can the algorithm be made better to identify cyberbullying?
3. How can the effectiveness and algorithm for the detection of cyberbullying be improved?

1.5 SIGNIFICANCE

Cyberbullying is the act of sending, receiving, and uploading offensive, defamatory, or malicious content about another person while using the internet or other online gaming or entertainment services. Social media bullying can take many forms, all of which operate in the same way such as threatening, slandering, and criticizing the individual. As a direct result of Cyberbullying the incidence of mental health issues has been exposed, particularly among the younger generation (Desai *et al.* [1]). It has developed in lower self-esteem and an increase in suicidal ideation. Thus, this problem must be mitigated as illegal activities can be detected or involving participants like perpetrators and sufferers can be identified. For this purpose, some automatic systems powered by machine learning can be implemented. Because the social condition exceeds the physical obstacle of human interplay and includes accessible contact with outsiders, it is necessary to examine the context of the topic. Cyberbullying makes the reverse understand that someone is being pursued anywhere as the web is just a pop away. As a result, the victim might experience physical, mental, and emotional effects. The majority of Cyberbullying takes the form of images or text posted on social media. If bullying text and non-bullying text can be distinguished, a machine can respond appropriately. An effective Cyberbullying detection system can help websites and other messaging apps combat such attacks and decrease the number of cases of Cyberbullying. The Cyberbullying location framework is used to distinguish and consider the significance of the Cyberbullying text. Before applying the previous data or visuals, one examines the text's various aspects to determine its context. A customized framework that can access such text really and productively should be created.

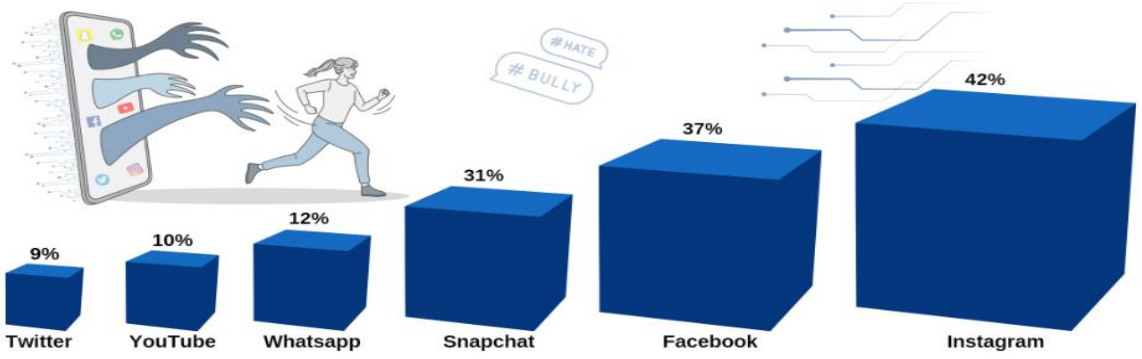

Figure 1.2: Where are People Cyberbullied

(Source: https://www.broadbandsearch.net/blog/cyber-bullying-statistics)

1.6 FRAMEWORK

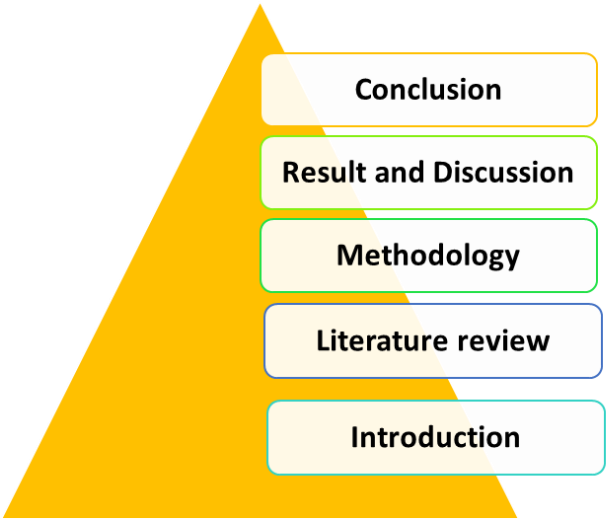

Figure 1.3: Framework

(Source: Self-created)

1.7 CONCLUSION

This section provides detailed information about the individual areas of the research topic. The thesis's aims and objectives have been constructed in order to explain the study's earlier objectives. This additionally covered the background, which highlighted the utilization of machine learning algorithms in Cyberbullying recognition. Likewise, the study questions are examined corresponding to the thesis objectives. The potential areas of "Machine learning algorithm" and Cyberbullying activities, as well as the issues associated with such activities, have been discussed. Furthermore, this chapter also successfully highlights the significance of doing a study on this high-burden topic. The comprehensive framework for the research is covered in the chapter's final portion of the thesis introduction.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Cyberbullying is extremely common and fits with the current commercial trend. Thus, several authors and investigators who have already explored this research topic previously will be evaluated and discussed from research papers, peer-reviewed journals, and research articles in this literature review chapter. In addition to that, this chapter also effectively highlights the research theories and models associated with the Machine learning algorithm used in Cyberbullying detection. However, this study also highlights the gap that remains in the previous research that has yet not been discussed. Finally, this chapter will be closed with a proper conceptual framework indicating the dependent and independent variables associated with this study.

2.2 EMPIRICAL STUDY

According to Rosa, *et al.* [2], perhaps the main issue during the time spent on programmed Cyberbullying recognition is the legitimate operationalization of Cyberbullying considering the essential standards illustrated in the important writing to accomplish the objective of programmed location frameworks, which is to unequivocally recognize Cyberbullying occasions. However, distinct standards should be established in order to develop appropriate advanced devices that combine programmed recognition features and capture the complexity of the peculiarity. This study provides a methodical consideration of current research based on the aforementioned definitions of Cyberbullying and focuses on four main criteria: a) language that is hostile or aggressive; b) the goal to hurt another person; c) Repeatedly engaging in the same behavior; and d) the amount that peers tend to have. This gives a predominant understanding of the factors considered in the customized disclosure of Cyberbullying. This investigation also focuses on methodological issues like inter-rater reliability, coder expertise, peer interaction, user consent, and peer interaction, all of which contribute to the validity of detection model classifiers. Many social factors, specifically related to AI, such as disasters, pandemics, war, psychological oppression, and wrongdoing, can be avoided using models based on techniques

from complexity science. The application of ML and Regular Language Handling techniques to the identification of Cyberbullying has been fraught with difficulties, making it a perplexing problem despite their success in a variety of text-based tasks. Cyberbullying is a fairly recent classification task. According to the author Cyberbullying mainly occurs via different digital devices such as computers, laptops, and mobile phones. And different illegal activities in terms of Cyberbullying mainly happened through email, and online chat, and it also happened via harmful messages or comments in regard to someone's post on social media platforms.

According to Kim *et al.* [3], a second, supporting set of studies has examined the crucial AI strategy used to combat Cyberbullying. The use of regulated learning, dictionary-based, rule-based, and blended drive methods for mechanized identification was discovered in a well-known review of specific scientists. Then again, various analysts proposed working on the strategic parts of AI in view of their particular audits. Yet, they didn't ponder the way that these upgrades need to come from real circumstances where the calculations could either help or hurt individuals they were intended for. However, a few Cyberbullying that additional research was required to take user and contextual aspects of cyber bullying into account. In reality, the researchers emphasized that while considering context, the majority of these studies have largely ignored the crucial problem. Some researchers recommended carefully considering user demographics when operationalizing the term Cyberbullying, while others in order to understand the context of incidents, suggested improved feature engineering rather than placing too much importance on feature selection and improvements in machine learning methodology. However, none of these studies suggested including bullies, victims, or bystanders in Cyberbullying incidents to capture this important context. Researchers have reported the importance of assuming other machine learning approaches, such as unsupervised and semi-supervised ones, in addition to the majority of the family of methods used to detect Cyberbullying. However, a few researchers also stated that methodological innovation and appropriate evaluation must coexist. As a result, experts have emphasized the importance of selecting an assessment metric that is unaffected by information skewness in order to avoid undesirable outcomes and uncertain outcomes. As evaluation metrics, the F-1 score or the area under the receiver-operating characteristic curve were recommended. Despite the fact that this is the case, the recent surveys did not consider the significance or value of efforts made by people to eliminate misclassifications.

According to Hani *et al.* [4], as more people engage in entertainment via the Internet, another kind of harassment is becoming more common. The latter is described as an ongoing, purposeful, or violent assault against a victim who is helpless to protect themselves due to infrequent communication messages. Life has always included harassment. Since the internet's inception, bullies have exploited this novel and opportunistic medium. Bullies were able to carry out their vile acts in complete secrecy and from a great distance away from the people they were targeting by utilizing services like email and instant messaging (Ptaszynski *et al [5]*). The main difference between Cyberbullying and traditional harassment is how the victim is affected. In contrast to Cyberbullying, which only affects the victim's emotional and mental state, traditional bullying can cause physical, emotional, and psychological harm. Cyberbullying must be identified and stopped as soon as possible because it hurts the victims. One of the useful methods that makes use of information to learn and create a model that naturally orders the right activities is artificial intelligence. PC-based knowledge, which can assume a huge part in perceiving risks' language plans, can be utilized to make a model that can be utilized to distinguish Cyberbullying ways to deal with acting. This paper's suggestion for a controlled AI strategy for identifying and preventing Cyberbullying is hence its key contribution.

Cyberbullying is a pressing online risk for adolescents, according to Cheng *et al.* [6], and it is highly problematic in society. With the aid of software, the researcher attempted to define the cyberbullying system. The most common internet threats relate to cyberbullying. This risk has been created for adolescents. The fast development in the use of “social media platforms” has dramatically improved the possibility of cyberbullying to happen. This desire may lead to unfavourable outcomes like low self-esteem, sadness, suicidal thoughts, and other forms of actions. This detection system mainly depends on the software system that can be controlled by the computer language. The researcher finds various kinds of problems and the researcher will try to state the issues in the proper way. The researcher tried to create a personal model for the detection system of cyberbullying.

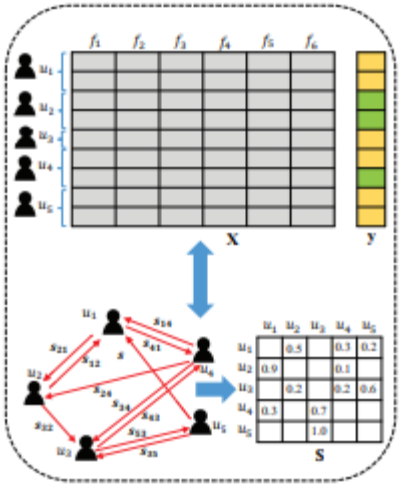

Figure 2.1: Detection system

(Source: https://par.nsf.gov/servlets/purl/10110257)

The figure shows the detection system that detects all users who are online on the server. All users are directly attached to each other and the detection system's results show in a table. Mathematical values are needed to get perfect detection values.

Social media sites like Facebook, Twitter, Flickr, and Instagram have evolved into web-based venues for posting and interpersonal interaction, according to Murshed *et al*. [7]. While these stages make it feasible for individuals to convey and cooperate in manners that were earlier impossible, they have additionally added to destructive ways of behaving like cyberbullying. A type of mental maltreatment known as cyberbullying essentially affects society. Occasions including cyberbullying have been on the rise, especially among youngsters who invest the majority of their energy bouncing between web-based entertainment stages. Virtual entertainment networks are particularly defenseless to CB due to their renown and the confidentiality given to victimizers by the Web. The author claims that 14% of incidents are said to occur online, and 37% of these occurrences include kids. Moreover, cyberbullying may bring about emotional wellness problems and damage. Nervousness, sorrow, stress, and social and personal troubles welcomed on by cyberbullying occasions represent most of the suicides. Therefore, a strategy for spotting cyberbullying in virtual entertainment posts, tweets, and remarks is fundamental. The cyberbullying area inside the Twitter stage has by and large been pursued through tweet gathering and somewhat with subjects showing draws near. Regularly

used text plans that take into account directed artificial intelligence (ML) models describe tweets into bothersome and non-torturing tweets.

According to Bozyiğit *et al.* [8], ML-based cyberbullying watchwords are one more heading of cyberbullying discovery, which has been utilized generally by a few specialists. Besides, AI (ML) is a part of man-made brainpower innovation that gives frameworks the capacity to gain and grow naturally for a fact without being uniquely modified, frequently sorted as regulated, semi-directed or solo calculations. A few preparation occasions in managed calculations are used to construct a model that produces the ideal forecast, for example, in light of clarified or marked information. Conversely, unaided calculations are not in light of information and are fundamentally used for bunching issues. The developers show how a web-based tool would be used by Japanese school employees and parents, with a requirement to identify poor materials on unofficial optional sites. The goal is to notify government experts about occurrences of cyberbullying; by using SVMs in this effort, they were able to achieve 79.9% accuracy. In order to safeguard underage users from sex provocations and cyberbullying, they have proposed a Facebook structure. In an effort to identify changes in behavior, the framework analyzes the content of images and videos as well as client activities. Another study suggests using provocation, flaming, psychological oppression, and prejudice to distinguish and categorize cyberbullying. The findings are average in terms of accuracy (about 40%) because the creator utilizes a flimsy characterization rule, however implementing more rules enhanced the classifier competency by up to 90%. In view of the Opinion examination in code-blended Hindi-English language, the developers have promoted a cyberbullying location model. With an eye towards the Instagram and YouTube stages, the creators finished their testing. In light of top, the authors employ an eight gauge classifier hybrid model that outperforms top with a precision and f1-score of 80.26% and 82.96%, respectively.

According to Iwendi et al. [9], the advancement of networking and information technology has made it possible for transparent online communication channels. Trolls have, however, abused this contemporary technology to conduct assaults and cyberattacks. According to data, about 18% of kids in Europe have experienced bullying or harassment from others via the web or mobile devices. Online misconduct that is upsetting and worrisome is termed as cyberbullying.

In numerous social media platforms, it shows up in a textual style and takes on numerous forms. Automation of this incident monitoring requires intelligent systems. This problem has been addressed in a few recent trials using conventional machine learning algorithms. The majority of the subjects have only ever been used on a single social network. Different models constructed using "deep learning algorithms" have been shown to have an effect upon identifying instances of cyberbullying in the most recent research. While some of these types of detection systems have constraints of "standard identification versions", others have effectively identified instances. This study employs empirical research to assess the viability and the efficiency of deep learning algorithms in identifying social commentary attacks. For the purposes of the research, the Gated Recurrent Units (GRU), Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BLSTM) and Long and Recurrent Neural Network (RNN) deep learning models were employed. As part of the data pre-processing stages, the researchers used text cleaning, tokenization, stemming, lemmatization, and the elimination of stop words. After pre-processing, clean text data is given to deep learning algorithms for predictions. According to the findings, the BLSTM model outperformed LSTM, GRU and RNN in terms of accuracy and F1-measure values. The thorough analysis of the research's results showed that deep learning models, when directly compared to others, proved to be the most effective against cyberbullying, opening the door for potential future blended methods that might be utilized to treat this serious online issue.

Ates *et al*. [10] claim that the degree of social media use has essentially increased as a result of the social networks' current rapid growth. People currently live in a time when they create virtual profiles through internet leisure for a number of reasons, such as to express their emotions and provide items to other people. Text, video, picture, and the voice data formats can be shifted after the virtual benefit upgrade, which often begins with enlistment to electronic diversion objections. Despite data transfers, internet entertainment is a correspondence environment where people can communicate by sending each other messages like status updates or comments. Electronic entertainment has expanded quickly over the past few years, which has led to both good and unpleasant changes. Cyberbullying is one of the most commonly noted harmful effects. It is the act of purposefully mistreating another person online using tailored, innovative tools. Cyberbullying is the incitement or fomenting of conflict through messages, cellphones, video

games, and other digital distractions. The evaluations based on the Turkish language are incredibly inadequate because the majority of examinations concentrating on differentiating cyberbullying are English-based. It is challenging to study "Turkish language" because it is a head-last language. It is spoken in Turkey in particular, as well as many other places on the earth. When estimates of online entertainment usage in Turkey are explicitly analyzed, it becomes clear how widely used virtual entertainment is there. As a result, it is now required to differentiate cyberbullying in Turkish literature. Cyberbullying is the act of employing mechanized devices against another individual or group of people with disproportionately large amounts of power. Cyberbullying causes more serious harm than actual torture. On the internet and in a virtual amusement setting, things that are prone to harassment might disappear swiftly. Along with the rapid growth of technology, hoodlums are motivated to perform these performances by their capacity to conceal the identities of the offenders through various methods in virtual settings and the difficulties in doing so.

Cyberbullying assumes the essential standards described in the important writing to complete the objective of programmed area frameworks, which is to unequivocally identify Cyberbullying occasions. However, distinct measures should be established in order to design suitable refined devices that combine programmed recognition features and capture the complexity of the peculiarity (Arif *et al.* [11]). This study delivers a methodical consideration of current research based on the aforementioned definitions of Cyberbullying and focuses on four main criteria: a language that is hostile or aggressive with the goal to hurt another person; repeatedly engaging in the same behavior; the amount that peers manage to maintain. This offers a prevailing understanding of the factors considered in the customized exposure of Cyberbullying. This investigation also concentrates on methodological issues like inter-rater reliability, coder expertise, peer interaction, user consent, and peer interaction, all of which contribute to the validity of detection model classifiers. Many social factors, specifically related to AI, such as disasters, pandemics, war, psychological oppression, and wrongdoing, can be detoured by operating models based on techniques from complicated science. The application of ML and Regular Language Handling methods to the identification of Cyberbullying has been fraught with complications, making it a perplexing problem despite their conquests in a variety of text-based tasks. Cyberbullying is a fairly recent classification task (Islam *et al* .[12]).

According to the author Cyberbullying mainly appears via different digital devices such as computers, laptops, and mobile phones. The application of ML and Regular Language Handling strategies to the designation of Cyberbullying has been fraught with problems, completing it a mystifying issue despite their success in a variety of text-based tasks. Cyberbullying is a fairly recent classification assignment. Cyberbullying mainly emerges via different digital appliances such as computers, laptops, and mobile phones. And different illegal movements in terms of Cyberbullying mainly occurred through e-mail and online chat, and it also happened via harmful statements or analyses in regard to someone's post on social media platforms. For instance, Reynolds and associates designed a straightforward language-specific method that was able to identify Cyberbullying on a limited Form spring dataset by registering the portion of swear words and insults in a post. Created a program that, when used for Myspace posts, detected a window of time (Alam et al. [13]). In recent years, feature engineering has become the process for noticing Cyberbullying that is utilized the most. This approach uses the field knowledge of linguistic signals in Cyberbullying to add additional features and measurements to the standard text-bag-of-words representation in an effort to enhance each classical classifier's performance individually.

AI strategy utilized to battle Cyberbullying. The benefit of regulated learning, dictionary-based, rule-based, and blended movement strategies for mechanized markers was found in a well-known review of specific scientists. Then again, various critics suggested working on the strategic parts of AI in view of their particular audits. Yet, they didn't consider the way that these upgrades need to come from real possibilities where the calculations could either help or hurt the someone's they were planning for. However, a few Cyberbullying that additional research was demanded to take user and contextual elements of cyber bullying into account (Bayari et al. [14]). In reality, the researchers noted that the main problem has generally been covered up when discussing background in most of these investigations. The actual part that social media opposition plays in encouraging Cyberbullying. Some researchers recommended carefully considering user demographic characteristics when operationalizing the concept of Cyberbullying, while others argued for better feature engineering to take into account the rich context of the incidents rather than overemphasize feature selection and machine learning

methodological advancements. However, none of these studies included bullies, victims, or bystanders in Cyberbullying happenings to apprehend this significant context. Researchers have declared the significance of assuming other machine learning approaches, such as unsupervised and semi-supervised ones, in addition to the majority of the family of methods utilized to find Cyberbullying. However, a few researchers also stated that methodological innovation and proper evaluation must coexist.

Life has continually possessed harassment. Since the internet;s beginning, bullies have manipulated this novel and opportunistic medium. Bullies were able to take out their vile acts in full secrecy and from a great space away from the people they were targeting by operating services like email and instant messaging. The main difference between Cyberbullying and conventional harassment is how the victim is simulated. In contrast to Cyberbullying, which only affects the victim;s emotional and mental state, traditional bullying can drive physical, emotional, and psychological harm. Cyberbullying must be determined and ceased as soon as possible because it hurts the victims (Liu et al.[15]). One of the useful procedures that makes use of details to learn and create a model that naturally demands the right activities is artificial intelligence. PC-based knowledge, which can suppose a huge part in perceiving risks language plans, can be utilized to make a model that can be utilized to indicate Cyberbullying ways to deal with acting. As a result, the main commitment of this paper is its recommendation for a managed AI procedure for noticing and controlling Cyberbullying.

Teenagers' greatest "pressing online risk" is cyberbullying, which has significant negative social effects. Using the software, the investigator tried to identify the cyberbullying technique. There is a common understanding that cyberbullying refers to the intentional victimization of others through electronic forms of communication. Cyberbullying has the most standard online risks. This risk has been devised for adolescents (Larochelle *et al.* [16]). The fast development in the usefulness of “social media platforms” has dramatically enhanced the possibility of cyberbullying to happen. The negative impact of cyberbullying on targets serves as the primary motivation for research that increases the "cyberbullying detection" accuracy. This goal can stop undesirable outcomes like low self-esteem, stress, suicidal thoughts, and many types of behaviours. This detection system specifically relies on the software system that can be

possessed by the computer language. The researcher discovers various kinds of problems and the researcher will try to state the problems in the proper way. Social media like Facebook, Twitter, Flickr, and Instagram have converted into ready toward web-based steps for posting and socialization among individuals, things being what they are. While these steps construct it possible for individuals to convey and cooperate in forms that were earlier impossible, they have additionally added to damaging ways of acting like cyberbullying.

A type of cognitive maltreatment known as cyberbullying essentially simulates society. Circumstances including cyberbullying have existed on the upgrade, especially among youngsters who support the majority of their energy bouncing between web-based entertainment Stages (Tulkarm *et al.* [17]). Moreover, cyberbullying may bring about heavy dynamic wellness issues and impairment. Nervousness, sorrow, stress, and social and personal risks welcomed by cyberbullying occasions symbolize most of suicides. Therefore, a strategy for spotting cyberbullying in virtual entertainment posts, tweets, and remarks is fundamental. The cyberbullying area inside the Twitter stage has by and large been followed through tweet gathering and somewhat with subject show draws near. Text plan considering directed artificial intelligence (ML) models is routinely utilized for describing tweets into hassling and non-torturing tweets. In addition, machine learning-based classifiers have been employed to distinguish between tweets that include disturbing content and those that don't. However, required classifiers perform poorly if the class names are rigid and unrelated to the latest developments.

Yao et al.'s [18] recognition of the repetitive nature of cyberbullying on social media—defined as a series of hostile messages sent from a bully to a victim with the intent to harm—allows them to significantly reduce the amount of features used in classification while maintaining high accuracy. With this method, accuracy, scalability, and timeliness are given top attention. On an Instagram dataset obtained by snowball sampling and labelled manually (to a limited extent) by a group of subject matter experts, models are trained using semi-supervised machine learning techniques. Some of the limitations included the use of a single data set that was only appropriate for Instagram, the inability to verify the veracity of labels, and the time commitment.

Cyberbullying may be recognised by combining textual data with social network characteristics, claim Huang et al. [19]. This is done by analyzing the structure of users' social networks and figuring out measures like friend count, network embedding, and relationship centrality. The study claims that social media skills have not been properly utilized in earlier studies. This study recommends that, along with textual analysis, social environment should also be taken into account when detecting cyberbullying. Before using Naive Bayes, J48, SMO, Bagging, and dagging, balanced data is created using the Twitter corpus from December 2008 to January 2009 and the synthetic minority oversampling method (SMOTE) approach.

Rakib et al. [20] cleaned up the Reddit database corpus in order to identify cyberbullying and trained a word embedding model based on the word2vec skip-gram model. The random forest classifier was trained to categorize the cyberbully's statements using the parameters of this model. This novel word embedding model outperformed four previously trained word embedding models in addition to four brand-new feature extraction techniques.

During their research, Hitesh S. and associates[21] employed the Support Vector Machine (SVM), Random Forest Classifier (RF), Gradient Boosting Machine, and Regression (LR),. The most precise activity classifier was chosen by comparing the outcomes. Our data was compiled using databases from Twitter and YouTube, both of which are open-source tools. It was found that the research employed a feature stack that consists of features from users, text-based, network, and lexical grammar. The performance of the LR and RF Classifiers greatly surpassed that of the Support Vector Machine and the Gradient Boosting Machine.

Cyberbullying is a severe cybercrime that can cause a great deal of mental anguish for the person who is the victim. Individuals who are subjected to cyberbullying can suffer a variety of negative consequences, spanning from strain and anxiety to life-threatening problems such as self-destruction and suicide, among others. Pew Research Center research [22] found that 60 percent of web users in the United States have experienced cyberbullying, with adolescent females being particularly vulnerable to such acts.

2.3 THEORIES AND MODELS

Cyberbullying is a common issue in the computerized age, and it can genuinely affect people. With the multiplication of virtual entertainment stages and the web, Cyberbullying has become more common and more testing to recognize and address. Machine Learning has shown extraordinary commitment in distinguishing Cyberbullying and giving early mediation to forestall its heightening. There are a few kinds of ML models utilized in Cyberbullying discovery, including supervised, unsupervised, and deep learning. Supervised learning models are prepared on marked informational collections, and that implies they gain from instances of Cyberbullying occurrences that have proactively been recognized and named by people. These models utilize this data to foresee regardless of whether another occurrence is Cyberbullying. Then again, unsupervised learning models don't need marked information. All things considered, they distinguish designs in the information and gather comparable occurrences together. This approach is valuable in distinguishing new and arising types of Cyberbullying.

Profound learning models, explicitly RNNs(Recurrent Neural Networks) and CNNs(Convolutional Neural Networks), have shown extraordinary potential in identifying Cyberbullying in normal language text. CNNs are utilized to recognize explicit examples and highlights in the text, like explicit words or expressions, while RNNs can catch the unique situation and succession of the text. One of the difficulties in utilizing ML models for Cyberbullying discovery is the class irregularity issue. Cyberbullying occurrences are generally intriguing contrasted with non-Cyberbullying episodes, making it trying to prepare models to precisely recognize them. Procedures like information increase and oversampling can be utilized to resolve this issue. ML models for Cyberbullying recognition can be incorporated into virtual entertainment stages to give constant observation and mediation. At the point when an example of Cyberbullying is identified, fitting moves can be made, like obstructing the client or giving assets and backing to the person in question.

All in all, ML models have shown extraordinary commitment in distinguishing Cyberbullying and giving early mediation to forestall its heightening. These models can be prepared on marked informational collections or utilized unaided to figure out how to distinguish arising types of Cyberbullying. Profound learning models, explicitly CNNs and RNNs, are helpful in

determining Cyberbullying in normal language text. By coordinating ML models into online entertainment stages, constant observing and mediation can be given to address Cyberbullying occurrences.

2.4 LITERATURE GAP

Cyberbullying is a developing concern around the world, and it has turned into a huge test to distinguish and forestall such a way of behaving. Machine learning strategies have been utilized to identify Cyberbullying by investigating web text and virtual entertainment posts. Notwithstanding, there are still some examination holes in this space that should be tended to.

One research gap is the absence of an extensive dataset for Cyberbullying location. Most investigations have depended on little, high-quality datasets, which may not be a delegate of certifiable situations. A bigger and more different dataset could assist with working on the precision and generalizability of Cyberbullying discovery models. Another research gap is the absence of regard for non-literary types of Cyberbullying, like pictures, recordings, and sound accounts. Cyberbullying can take different structures, and the identification models ought to be prepared to precisely perceive these structures.

In addition, most examinations on Cyberbullying recognition utilizing AI have zeroed in on the location of unequivocal Cyberbullying. There is little examination on distinguishing secretive or inconspicuous types of Cyberbullying, for example, perceived hostilities or latent forceful way of behaving, which can likewise essentially affect the people in question. Finally, the ethical implications of Cyberbullying discovery models should be thought of. The potential for bogus up-sides and the utilization of these models for reconnaissance purposes could prompt unseen side-effects, for example, control and encroachment of security privileges. In this way, research on the moral ramifications of Cyberbullying discovery models is fundamental.

2.5 CONCEPTUAL FRAMEWORK

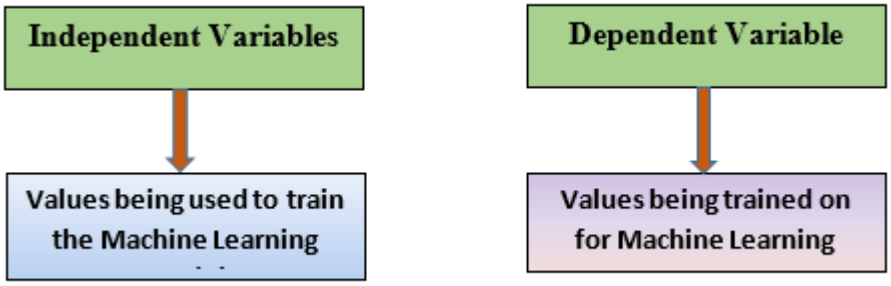

Figure 2.2: Conceptual Framework

(Source: Self-created)

2.6 CONCLUSION

Different facets of "Machine learning algorithm" and Cyberbullying activities and their impact on social life as researched by various authors and researchers are adequately described in the literature review chapter. The analysis of this study totally depends upon the understanding of the research topic and background of this study based on the particular research problems. This literature review chapter also highlighted the appropriate theory and models associated with this research topic. Finally, this chapter also presented the conceptual framework with different dependent and independent variables of this topic.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This methodology part is a very important section for each and every thesis report. The methodology contains various procedural methods that are assumed during the steps of performing the research. This chapter summarizes the steps affected by the strategic planning of the research method, and also it helps to develop various processes in relation to data collection, obtaining outcomes, and also performing the analysis. This methodology section covers the all important areas that are involved in the research performing method. The researcher will include a research outline, research design, research approaches, research strategies, and research methods. These points are very important and also directly connected to the thesis report. This methodology section mainly provides the proper data collection system of the research and also discusses the limitations of the thesis report.

3.2 METHOD OUTLINE

To obtain the desired study results, the research will constantly adhere to a system procedure from the design stage to the execution step. The planning of the study is largely influenced by its goals. The planning stage of research also involves setting up different strategies and approaches to reach the research goal and also to fulfill the objectives of the entire study (Rezvani *et al.* [23]). After selecting the appropriate research strategy, approaches, and method, the data collection process has been started. The data collection method is considered the most crucial step of the entire thesis from which all the required information can be gathered to obtain the most appropriate outcomes of the study. After collecting all the information, the most appropriate data has been identified from various resources which have been done by different authors. Then, using various methodologies, this data can be analyzed to produce the most insightful findings that can address all of the study's issues. As a result, the thesis procedure entails the full planning of the research project, including the approach, strategy, relevant methods, and finally the data gathering method and data analysis techniques.

3.3 PHILOSOPHY

A study is always conducted based on many research philosophies that the researcher always adheres to in order to achieve the study's genuine objectives in accordance with its primary goal. The researcher can advance the study by choosing an actual research philosophy before moving on to the research approach, research strategy, and lastly the research method. Research Onion presents a variety of research philosophies, including "interpretivism, objectivism, positivism, constructivism, realism, and pragmatism".

However, since the entire investigation will be based on the test hypothesis and the research questions that were established at an earlier stage of the study, the positivist research philosophy will be used for this study. Additionally, based on the background of the research issue, this study will also provide quantitative results in the context of well-defined knowledge. This research philosophy focuses on the quantifiable outcomes of the investigation to analyze various research topics to obtain the important research outcomes.

3.4 APPROACHES

The research onion presents many research methods to carry out the full research and to carry out the research analysis. The deductive research approach and inductive research approach are the two main research methodologies that are most widely used. The emphasis of the deductive research approach is on producing research findings that are based on accepted research theories. The formation of entirely new research theories based on the research background is the only emphasis of inductive research approaches, on the other hand. The background of the study topic and the research issues, however, have a significant impact on the choice of the research approach. Finally, in this particular research, the positivist research philosophy has been selected which focuses on the uptake of the “deductive research approach” to proceed with the particular research topic and to analyze the entire research. The deductive research approach focuses that the research will proceed with the existing information and the facts based to perform the potential results that will help to address and explain the actual research questions.

3.5 DESIGN

To successfully complete the entire study, research must be planned after choosing the suitable philosophy and approach. There are primarily two approaches to design research, i.e quantitative research and qualitative research. Qualitative research study is also known as the subjective study where the study is based only on information and theory with respect to the background of the entire research. The subjective examination is an exploratory investigation that aims to show the emotional experiences, perspectives, convictions, and behavior patterns of individuals from top to bottom. The method chosen depends on the research question and goals, and both qualitative and quantitative research methods have advantages and disadvantages. As opposed to quantitative examination, which can be utilized to test theories and draw speculations about a populace, subjective exploration can be utilized to research novel peculiarities and create novel ideas.

On the other hand, quantitative research is entirely dependent on numerical data, which helps to obtain quantifiable outcomes, statistical findings, and will help to achieve the study's true aim. This approach takes use of non-numerical data, such as words, images, and observations, to shed light on complicated events. Examples of qualitative research techniques include in-depth interviews, focus groups, ethnography, case studies, and content analysis. In qualitative research, data analysis techniques involve interpreting and classifying data into themes and patterns (Singh *et al.* [24]). Under this technique, accurate strategies are utilized in information assortment and examination to track down examples, connections, and circumstances and logical results connections. Surveys, experiments, and quasi-experiments are among the quantitative research methods. For quantitative research, statistical analysis methods like regression analysis, ANOVA, and correlation analysis are typically used in data analysis techniques.

However, this study is followed by the positivism research philosophy which emphasizes doing the study in a more scientific way thus, to meet this research purpose the deductive research approach is most appropriate which also has been upvoted for this particular study. In addition to that, quantitative research can only provide measurable outcomes either by gaining numerical results or by gaining the statistical result of the entire research based on the existing background knowledge and theory-based information.

Quantitative exploration is a planned approach that uses mathematical data to evaluate information and test theories.

3.6 STRATEGY

Research strategy is the stage of the entire study where the researcher identifies and describes all the research approaches to achieve the research goal and to achieve the study's true purpose. The research strategy is one of the steps of the entire methodology portion which describes the actual step needed to reach the resources of this study from where the information of this study can be gathered to obtain the actual result of this study. This step involved the involvement of the participants or the researcher if the study will proceed by primary analysis. Research can be conducted in mainly two ways either by primary analysis or by secondary analysis (Chen *et al.* [25]). The primary analysis can be carried out primarily by two processes, such as the interview method or the survey method, by adhering to the research philosophies. The interview or survey can be carried out using a questionnaire, in which a list of questions has been put together, the participants have been asked all the questions, statistical analysis has been done based on their responses, and finally the researcher can reach their research objectives and also obtain the answers to all the research questions by analyzing their responses. The secondary analysis, however, only relies on gathering data using the archival strategy mode. The secondary research analysis was only used to collect information from the resources where the research has been done on the basis of that particular topic and represent the proper and knowledge-based information that will help to get the proper results of study to reach the objectives and to meet actual purpose of the study.

3.7 METHOD

The research method is the step where all techniques of the research have been described. The research method describes all the techniques that must be applied to analyze the research by utilizing the gathered information collected from various research sources. Depending upon the utilizing the research resources the research method can be selected. There are mainly three research methods that can be described: the mono-method, the multi-method, and eventually the mixed-method of study. The mono-research method is the kind of approach that solely uses qualitative and quantitative research analysis techniques to carry out the complete investigation.

On the other hand, the mixed research method can be described as the research method where the research objectives can be met by using the mixed results of both the quantitative and the qualitative research. And finally, the multi-research method emphasizes the result just focusing on the result from both the qualitative and quantitative research.

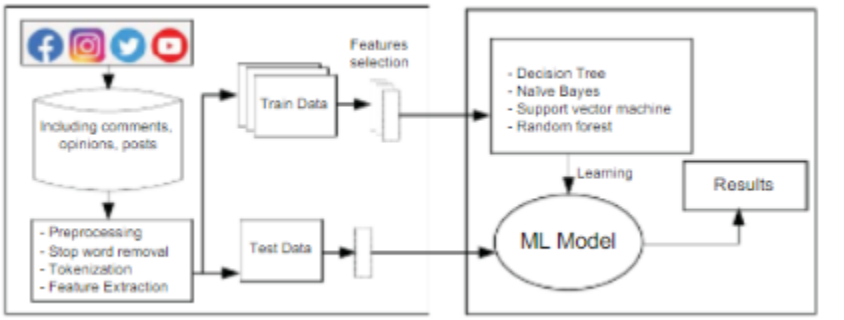

Figure 3.1: Proposed method

(Source: Self-created)

However, for this particular study, both quantitative and qualitative research can be followed to obtain the most successful result of this study. This study is software-based where the cyberbullying activities mainly on social media can be detected by a "machine learning algorithm". Thus this type of software-based research always provides measurable outcomes which means it follows quantitative research. However, this research also adheres to information and theory-based information that will assist in providing answers to theory-based research questions set up to fulfill the study's objectives and focus on the most important aspect of this subject.

3.8 DATA COLLECTION

To produce the final results, we utilized the Dataturks Tweet Dataset for Cybertroll Detection from Kaggle. Despite the fact that we also looked at many other datasets, many of them either lacked certain characteristics, had poor quality, or had data that, after manual inspection, was irrelevant. As a result, after examining a variety of open-source datasets, we decided on it because it appeared to meet all the necessary requirements.

Below is a thorough overview of the dataset:

1) Data in the dataset has been manually marked.

2) Total 20001 occurrences

Tweet and label are 2 attributes in the dataset. [1 is equivalent to yes, while 0 is equivalent to no]

3.9 DATA CLEANING

The dataset is in JSON format. Since the dataset's fields are straightforward to interpret, the original set of fields in the annotation attribute were deleted and replaced with the label values to simplify the next step. Table 3.1 displays the number of instances for each class present in the dataset.

Table 3.1: Instances

Total number of incidents	20001
Non- CyberBullying incidents	12179
CyberBullying incidents	7822

3.10 DATA PREPROCESSING

The following were the preprocessing steps:

1. Words Tokenization : A single item that acts as the main idea of a phrase or paragraph is referred to as a token. Our text is divided into individual words in a list using word tokenization.
2. Stopwords Stopwords are removed by using the nltk.corpus.stopwords.words('english') function, which returns the list of stopwords from the English lexicon. Stop words are unimportant terms like "the," "a," and "an" that have no influence on how the facts to be evaluated should be interpreted.
3. String.punctuation can be used to confirm that we only save non-punctuation characters when removing punctuation.
4. Stemming: Stemming is a linguistic normalization procedure that lowers words to their root word. To obtain the stemmed tokens, we use nltk.stem.porter.PorterStemmer.
5. Digit removal: Since numerical information doesn't promote cyberbullying, we also filtered out any such information.

6. The TF-IDF Transform in Python's sklearn module was used to extract features for use with ML algorithms in the following phase. Words that are common to several texts are given less weight (importance) in the TF-IDF approach, rendering them useless for differentiating across the papers. The result matrix includes the importance (weight) determined by tf * idf (matrix values) and each document's row and word's column. If a term has a high tf-idf, suggesting that it most likely appeared there, it must be absent from other texts. Therefore, the term must be a signature word.

3.11 LIMITATIONS

- The machine learning algorithm that will be used to identify cyberbullying actions on social media forms the basis of this thesis. However, even though there are numerous models employed in machine learning, none of them can be utilized to detect cyberbullying behavior.
- Machine learning activities have no such accuracy power to detect cyberbullying activities.
- Cyberbullying activities also can be detected by both the deep learning and hybrid methods which have not been used for this particular study.

3.12 CONCLUSION

The research process is completely explained in the chapter. The most crucial methods, approaches, frameworks, and models used by the researchers are fully presented in this chapter. The methodology section, which also covered the relationships between data collection, analysis, and planning in great detail, has also investigated the general order of procedures necessary to accomplish the research's goals. The chapter discussed the many processes in the research process, the considerations that were made when conducting the research, and the research constraints pertinent to the technique of the study.

CHAPTER 4

RESULT AND DISCUSSION

4.1 INTRODUCTION

This component is essential for completing the thesis procedure. Even making all of the adjustments that are absolutely essential to accomplish this event contributes to the best possible performance of the project's results, which also demonstrates its accuracy when used in research. That even creates the ideal possible opportunities for the research's understanding procedure. It creates the internal goals of the investigation, which are crucial to carry out in this study.

4.2 MACHINE LEARNING MODELS ADOPTED

Five machine learning (ML) models—Random Forest, Decision Tree, Logistic Regression, Naive Bayes, and Adaboost—are used in our study. The experiment revealed that Random Forest performed better than every other model in every parameter.

4.2.1 Logistic Regression

Any statistical set of guidelines that forecasts possibilities rather than classes is referred to as "logistic regression." A hyperplane is constructed using the logistic function to classify recorded factors within the given classes. Textual attributes are included in the set of rules that are used for generating predictions about a data point that belongs to a particular class.

4.2.2 Naive Bayes

Naive Bayes classifiers are classification algorithms that classify data using Bayes' Theorem. These methods are all guided by the same principle: each pair of features recognised is distinct from all previous pairs of features detected.

4.2.3 Decision Tree

A tree can be compared to a variety of real-world objects and has an impact on a number of machine learning applications, including classification and regression, among others. A decision tree can be used in decision analysis as a visual, explicit depiction of decisions and decision-making. Information is organized using a paradigm that resembles a decision tree, as implied by

the name. Making a strategy for achieving certain goals is frequently utilized in both machine learning and data mining.

4.2.4 Random Forest

The random forest classifier is the type of classifier that applies a number of decision tree classifiers at random to different sub-samples of the dataset in order to maximize projected accuracy and reduce overfitting.

4.2.5 AdaBoost

The stagewise addition principle, which employs a number of weak learners in order to build strong learners, is also the basis for the boosting method AdaBoost. In this case, the value of the alpha parameter will be inversely correlated with the learner's error.

4.3 PERFORMANCE METRICS

The evaluation of machine learning models involves the calculation of various metrics to assess their performance. These metrics include the model’s accuracy, precision, recall, and F-score. By applying different machine learning models, we can compare their accuracies and other performance evaluation metrics, enabling us to make informed decisions about the effectiveness of each model.

4.3.1 Accuracy

A predictive model's accuracy in the context of machine learning refers to its capacity to correctly categorize or forecast instances within a given dataset. It serves as a crucial evaluation parameter for evaluating the effectiveness and dependability of machine learning algorithms. By dividing the number of cases that are successfully predicted by total number of instances in the dataset, accuracy is often reported as a percentage. The resultant value is a measure of the model's accuracy in making predictions.

$$\text{Accuracy} = (\text{True Positive (TP)} + \text{True Negative (TN)}) / \text{Total Predictions}$$

4.3.2 Precision, Recall And F1-Score

Precision: can be described as the ratio of correctly classified attacks (TP) with total flows classified as the attack (TP+FP).

Precision = True Positive (TP) / (True Positive (TP) + False Positive (FP))

Table 4.1: Different machine learning models' performance metrics

Machine Learning Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Logistic Regression	0.80	0.81	0.80	0.81
Naive Bayes	0.62	0.79	0.62	0.59
Decision Tree	0.85	0.88	0.85	0.85
Random Forest	0.92	0.92	0.92	0.92
AdaBoost	0.71	0.74	0.71	0.72

Recall: Recall is the ratio of successfully detected attacks (TP) to all flows that are considered attacks (TP+FN).

Recall = True Positive (TP) / True Positive (TP) + False Negative (FN)

F1-Score: is the harmonic mean(HM) calculated by considering precision and recall.

F1-Score = 2*(Precision*Recall) / (Precision + Recall)

4.4 Conclusion

This chapter shows the result and discussion of the research. Various models were implemented and it was observed that random forest performed best among the machine learning models.

CHAPTER 5

CONCLUSION

5.1 INTRODUCTION

The fundamental analysis of the research served as the foundation for this chapter. That also makes all of the possible chances to complete the research. That even generates the proper aspects of the research. It maintains the condition of the research. That displays the efficiency of the research. With the help of the above discussion part, the conclusion part is a process. That leads to the interest to complete the research. However, a few researchers again displayed that methodological invention and proper evaluation must coexist. For instance, when the number of positive annotated models is extremely lower than the general content discovered on social media platforms, the majority of the Cyberbullying datasets frequently exhibit significant class inequality. As a result, experts have highlighted the significance of deciding on an assessment metric that is unaffected by information skewness in order to avoid unpleasant outcomes and uncertain developments. As a metric for evaluation, the area under the receiver-operating distinctive curve was proposed. Despite that, the current surveys did not scrutinize the effectiveness or value of human assistance in eliminating misclassifications. All of these aspects are necessary to perform the measurement or the purposes of the research. Machine learning algorithms, for instance, may be used to recognize language that is often used in cyberbullying and trends in how individuals interact online. Overall, there is some success in using machine learning to identify cyberbullying. There are still a few problems that need to be fixed in order to use artificial intelligence for cyberbullying detection. Additionally, it's vital to keep in mind that human judgment should never be replaced by machine learning algorithms. This entails keeping a distance from bullies online, banning them, and reporting any incidents to the authorities.

5.2 INKING WITH OBJECTIVES

- To choose a suitable Machine learning algorithm for the detection of Cyberbullying.

The fast development of social networks nowadays, the degree of social media use has basically developed. People are in a period in which people make a virtual profile through web-based entertainment for purposes, for instance, as conveying their sentiments and

presenting things to other people. After the virtual benefit modification, which, generally speaking, starts with enlistment to electronic diversion objections, data types, for instance, text, sound, picture, and video can be driven. Despite data transfers, internet entertainment is a correspondence environment where people can communicate by sending each other messages like status updates or comments. The quick extension in the use of the electronic amusement throughout the span of the last years has brought various positive and negative developments. All of these detection techniques are required to reduce the cyberbullying process. That also reduces the chances of cyber threats..

- To develop and improve the effective Machine learning algorithm used for Cyberbullying detection.

Transparent online communication channels' have been created by the evolution of "networking and information technology". However, this has taken advantage of this modern technology to launch attacks and cyber attacks. 18% of kids in Europe have experienced bullying or harassment from others via the web or mobile appliances. Online misconduct that is upsetting and worrisome is termed as cyberbullying. In numerous social media platforms, it displays up in a textual style and takes on numerous forms. Automation of this incident monitoring needs intelligent systems. This problem has been addressed in a few recent attempts utilizing conventional machine learning algorithms. The maturity of the subjects has only ever been employed on a single social network. Different models produced using "deep learning algorithms' have been shown to have an effect on identifying specimens of cyberbullying in the most recent research. To solve this kind of issue or problem the machine learning process is also used.

- To improve the accuracy and efficiency of the Machine learning algorithm that can be used for the detection of Cyberbullying.

Conventional models have been used in the past to automatically display Cyberbullying on social media. Thus, this difficulty must be mitigated as illegal activities can be detected or affect participants like perpetrators and sufferers can be determined. For this purpose, some automatic procedures powered by machine learning can be executed. Because the social condition surpasses the physical obstruction of human interplay and includes convenient contact with

outsiders, it is necessary to examine the context of the topic. Cyberbullying makes the reversal understand that someone is being pursued anywhere as the web is just a pop away. As a result, the victim might participate in physical, mental, and emotional effects. The majority of Cyberbullying takes the form of images or text assigned on social media. If bullying text and non-bullying text can be distinguished, a device can respond properly. Thus it showed its capability to detect all of the cyberbullying processes.

5.3 RECOMMENDATION

There are irregular instances where using machine learning to identify cyberbullying is successful. To begin with, massive datasets must be utilized for training in order for algorithms that use machine learning to be effective. This might be difficult since it requires substantial access to information that might not be available. Additionally, algorithms for machine learning must be able to identify minute similarities in data that may go unnoticed by humans. This can be tough since algorithms are needed to be able to spot patterns that are hard for individuals to notice. The ability to discriminate between cyberbullying and other types of online contact is required by machine learning algorithms, which come in second. This can be challenging since the algorithms must be able to distinguish minute variations across various forms of online interactions. The ability to discern between cyberbullying and other online interactions that are not necessarily malevolent is another requirement for machine learning algorithms. All of this is also helpful to reduce the chances of the cyberbullying system.

5.4 CONCLUSION

The research issue, which is the application of a Machine learning algorithm to identify Cyberbullying actions, is clearly defined in the literature review chapter. This topic is very popular and concedes with the recent market movement. Thus, several authors and detectives who have already investigated this research topic earlier will be considered and consulted from research papers, peer-reviewed journals, and research articles in this literature review chapter. In addition to that, this chapter also actually highlights the analysis theories and models associated with the machine learning algorithm used in Cyberbullying detection. However, this study again emphasizes the gap that is discussed also. Finally, this chapter will be approached with a proper conceptual framework showing the dependent and autonomous proposes of the research.

Reference list

[1] Desai, A., Kalaskar, S., Kumbhar, O. and Dhumal, R., 2021. Cyber Bullying Detection on Social Media Using Machine Learning. In ITM Web of Conferences (Vol. 40, p. 03038). EDP Sciences.

[2] Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P.C., Carvalho, J.P., Oliveira, S., Coheur, L., Paulino, P., Simão, A.V. and Trancoso, I., 2019. Automatic "Cyberbullying" detection: A systematic review. *Computers in Human Behavior*, *93*, pp.333-345.

[3] Kim, S., Razi, A., Stringhini, G., Wisniewski, P.J. and De Choudhury, M., 2021. A human-centered systematic literature review of "Cyberbullying" detection algorithms. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), pp.1-34.

[4] Hani, J., Mohamed, N., Ahmed, M., Emad, Z., Amer, E. and Ammar, M., 2019. Social media "Cyberbullying" detection using machine learning. *International Journal of Advanced Computer Science and Applications*, *10*(5).

[5] Ptaszynski, M., Lempa, P., Masui, F., Kimura, Y., Rzepka, R., Araki, K., Wroczynski, M. and Leliwa, G., 2019. Brute-force sentence pattern extortion from harmful messages for cyberbullying detection. *Journal of the Association for Information Systems*, *20*(8), pp.1075-1127.

[6] Cheng, L., Li, J., Silva, Y., Hall, D. and Liu, H., 2019, August. PI-bully: Personalized cyberbullying detection with peer influence. In *The 28th International Joint Conference on Artificial Intelligence (IJCAI)*.

[7] Murshed, B.A.H., Abawajy, J., Mallappa, S., Saif, M.A.N. and Al-Ariki, H.D.E., 2022. DEA-RNN: A hybrid deep learning approach for cyberbullying detection in Twitter social media platform. IEEE Access, 10, pp.25857-25871.

[8] Bozyiğit, A., Utku, S. and Nasibov, E., 2021. “Cyberbullying” detection: Utilizing social media features. Expert Systems with Applications, 179, p.115001.

[9] Iwendi, C., Srivastava, G., Khan, S. and Maddikunta, P.K.R., 2020. “Cyberbullying” detection solutions based on deep learning architectures. Multimedia Systems, pp.1-14.

[10] Ates, E.C., Bostanci, E. and Guzel, M.S., 2021. Comparative performance of machine learning algorithms in cyberbullying detection: Using Turkish language preprocessing techniques. arXiv preprint arXiv:2101.12718.

[11] Arif, M., 2021. A systematic review of “Machine learning algorithms in “Cyberbullying”detection: future directions and challenges. Journal of Information Security and Cyber Crimes Research, 4(1), pp.01-26.

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

[13] Alam, K.S., Bhowmik, S. and Prosun, P.R.K., 2021, February. Cyberbullying detection: an ensemble based machine learning approach. In 2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV) (pp. 710-715). IEEE.

[14] Bayari, R. and Bensefia, A., 2021. Text mining techniques for cyberbullying detection: state of the art. *Adv. Sci. Technol. Eng. Syst. J*, *6*, pp.783-790.

[15] Liu, Y., Zavarsky, P. and Malik, Y., 2019. Non-linguistic features for cyberbullying detection on a social media platform using machine learning. In *Cyberspace Safety and Security: 11th International Symposium, CSS 2019, Guangzhou, China, December 1–3, 2019, Proceedings, Part I 11* (pp. 391-406). Springer International Publishing.

[16] Larochelle, M.A. and Khoury, R., 2020, December. Generalization of cyberbullying detection. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 296-300). IEEE.

[17] Tulkarm, P., 2021. Approaches to cyberbullying detection on social networks: A survey. *Journal of Theoretical and Applied Information Technology*, *99*(13).

[18] Yao, Mengfan, Charalampos Chelmis, and Daphney? Stavroula Zois. ”Cyberbullying ends here: Towards robust detection of cyberbullying in social media.” The World Wide Web Conference. 2019.

[19] Huang, Qianjia, Vivek Kumar Singh, and Pradeep Kumar Atrey. ”Cyberbullying detection using social and textual analysis.” Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. 2014.

[20] T. Bin Abdur Rakib, L. K. Soon, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag, 2018), vol. 10751 LNAI, pp. 180–189.

[21] Hitesh Kumar Sharma, K Kshitiz, Shailendra, “NLP and Machine Learning Techniques for Detecting Insulting Comments on Social Networking Platforms”, Proceedings of the International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France, June 2018.

[22] A. Geiger, “How and why we studied teens and cyberbullying,”Pew Research Center, 2018.

[23] Rezvani, N. and Beheshti, A., 2021. Towards Attention-Based Context-Boosted Cyberbullying Detection in social media. *Journal of Data Intelligence*, *2*(4), pp.418-433.

[24] Singh, V.K. and Hofenbitzer, C., 2019, August. Fairness across network positions in cyberbullying detection algorithms. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 557-559).

[25] Chen, H.Y. and Li, C.T., 2020. Henin: Learning heterogeneous neural interaction networks for explainable cyberbullying detection on social media. *arXiv preprint arXiv:2010.04576*.

Ptaszynski, M., Lempa, P., Masui, F., Kimura, Y., Rzepka, R., Araki, K., Wroczynski, M. and Leliwa, G., 2019. Brute-force sentence pattern extortion from harmful messages for cyberbullying detection. *Journal of the Association for Information Systems*, *20*(8), pp.1075-1127.

PAPER NAME

Thesis_Mayank.pdf

WORD COUNT

11421 Words

CHARACTER COUNT

65486 Characters

PAGE COUNT

44 Pages

FILE SIZE

2.2MB

SUBMISSION DATE

May 30, 2023 6:27 PM GMT+5:30

REPORT DATE

May 30, 2023 6:27 PM GMT+5:30

12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 8% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 10% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 8 words)