Design and Development of Improved Cyberbullying Detection Model

Mehul Parmar Shubham Kotak

Long Island University Long Island University

[MEHUL.PARMAR@my.liu.edu](mailto:MEHUL.PARMAR@my.liu.edu) [shubhamajaykumar.kotak@my.liu.edu](mailto:shubhamajaykumar.kotak@my.liu.edu)

Dhyey Parekh Pooja Modi

Long Island University Long Island University

[dhyeyvishal.parekh@my.liu.edu](mailto:dhyeyvishal.parekh@my.liu.edu) [poojaatulkumarmodi.modi@my.liu.edu](mailto:poojaatulkumarmodi.modi@my.liu.edu)

***Abstract*—The number of internet users is increasing in social media as the internet grows in popularity day by day. Depending on the utilization of people these social media sites can have both positive as well as negative consequences. People are increasingly sharing their photos, videos, ideas, status, experiences, and views on social media sites such as Facebook, YouTube, and Twitter, as their use of the internet has grown in recent years. Cyberbullying is one of the adverse repercussions of social media today. it includes acts of bullying and exploitation of digital technologies such as posting or sharing inappropriate or corrupt content about another person. It is a significant issue, and it has a serious negative impact on victims, which may be quite devastating. To keep these online platforms healthy and secure, systems for detecting these events have become vital. The discovery demonstrates how a machine learning classifier can be used to detect and predict the act of cyberbullying Due to the large amount of users and data on social media sites it is difficult to detect cyberbullying. As a result, detecting cyberbullying in efficient manner becomes critical. The best accuracy for detection can be achieved by identifying and integrating key features into a single model. Extensive tests have been carried out, and the results show that the proposed scheme detects cyberbullying with good efficiency and accuracy.**

***Keywords-Cyberbullying Detection, Social Media, Machine Learning, Sentiment Analysis***

**CHAPTER 1-INTRODUCTION**

1. **What is Cyberbullying?**

Bullying is not a new phenomenon, but the increasing adoption of new communication technology has allowed bullying behavior to migrate to cyberspace, a phenomenon generally referred to as ”cyberbullying.” Because of its expanding frequency and the fact that it has been linked as a factor in a number of teen suicides, cyberbullying is a major source of concern for parents, police, educators, and the general public.[4] People’s lives have become more complicated since the birth of the internet, with

Their lives becoming totally or largely dependent on it. Children and teenagers are accessing the internet in greater numbers, at younger ages, and in more diversified ways than ever before (e.g. smartphones, laptops and tablets) As a result, cyberbullying has become a significant cause of worry.[5] Cyberbullying is a conscious and repeated act of causing harm or humiliation to another person through the use of information and communication technology such as mobile phones, e-mails, and social media. Because the social lifestyle eliminates the physically human interaction and allows for inappropriate interaction with others, it is necessary to analyses and research the domain of cyberbullying. Further- more, because the majority of countries have not established a well-defined legal framework for cyberbullying, the knowledge to defend the issue is uncertain. Social media has changed the way we live and conduct business by allowing us to communicate with others in real-time. Along with the positive uses of social media, certain harmful effects are observed. One of the most extreme bad uses of digital media is cyberbullying, in which these platforms are used to irritate, intimidate, or humiliate another online user. Some notorious users are using these platforms to upload distorted data and images, to write insulting or harsh comments about others, and to post clips designed to hurt or shame others. Cyberbullying has been shown to have long-term impacts on victims, producing stress and placing them in a permanent state of distress or anxiety, resulting in sleep disorders and hunger problems. [6]

The automatic surveillance of cyberbullying has attracted a lot of attention in the field of computer science. The subjective nature of bully gestures, on the other hand, complicates computer recognition of abuse messages and its classification. As a result, numerous solutions to this problem have been presented. They can be divided into one of three types: supervised learning, lexicon-based learning, and rule-based learning approaches. This critical data, as well as textual information, is employed in this work to improve the detection efficiency of cyberbullying.[2]

1. **Its Impact**
   * Cyberbullying does not involve personal contact between the bully and the victim, but it is harmful for the victim psychologically.
   * 8
   * Physical impacts:- Sleep, Headache, Abdominal pain, Anorexia, Nausea are caused.
   * Emotional Impacts:- Anxiety, Loneliness and Depression.
   * Their self-esteem is reduced. Low self-esteem may also become the beginning of cyberbullying because indi- viduals with low self-esteem are an easy target for the aggressors. [8]

**Types of Cyberbullying**

Just like traditional bullying exists in many different types such as verbal abuse and physical violence, there are many different types of cyberbullying. Here is a list of few common types of cyberbullying. [9]

Types of Cyberbullying

1. **Harassment**

• It involves the bully sending offensive and malicious messages to an individual or a group and is often repeated multiple times.

• Cyberstalking is one form of harassment that involves continual threatening and rude messages, and can lead to physical harassment in the real, offline world.

1. **Flaming:** Flaming is similar to harassment, but it refers to an online fight ex- changed via emails, instant messaging or chat rooms. It is a type of public bullying that often directs harsh languages, or images to a specific person.
2. **Exclusion:** Exclusion is the act of purposely singling out and excluding someone from an online group, such as a chat room or a website. The gang then makes derogatory comments and harasses the individual they singled out.
3. **Outing:** When a bully publicly discloses personal and private information, images, or videos or any other sensitive information about someone, this is known as outing. When a person’s information is widely distributed on the internet, he is “outed.”
4. **Masquerading:** Masquerading is a situation where a bully creates a fake identity to harass someone anonymously. In addition to creating a fake identity, the bully can impersonate someone else to send malicious messages to the victim.

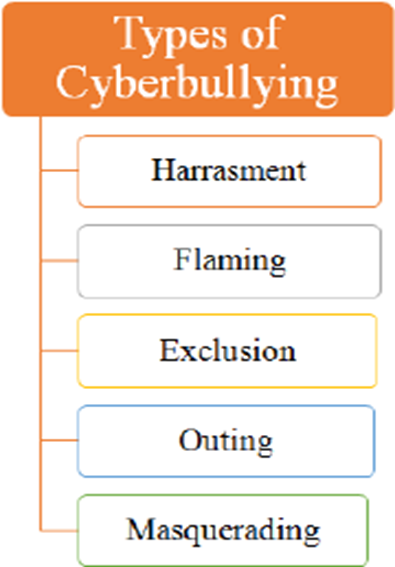


Figure 1.1: Types of Cyberbullying

1. **Countermeasures by Social Media**

Social networks can help you have a more secure browsing experience. The following are some tools that can be used to preserve one’s privacy:

* Twitter© provides users with the following tools (”Learn How”, 2017).

1. Unwanted followers can be blocked, muted, or unfol- lowed by users.
2. Filters on alerts allow users to filter out any unsolicited answers or mentions from accounts they do not follow.
3. Twitter has been notified of the inappropriate behavior
4. Before displaying sensitive content, warn the user. It only works with photographs and videos.
5. Tagging privacy for photos enables the user to control who can and cannot tag him or her in photographs.
6. Twitter only permits 280 characters.
7. Any inappropriate content and account can be blocked and reported to the Twitter which may lead into Termination of the account.

• Facebook ® has the following tools (”How to Report”, 2017) (Basic Privacy, 2017):

– Block a story from appearing in their News Feed. However, there are no built-in methods on social networking platforms to detect cyberbullying. They only take action in response to reported incidents. [7]

1. **Problem Definition and Motivation**

Cyberbullying is terrifying and devastating enough to make victims contemplate suicide, and in the most severe or unpleasant instances, it can result in suicidal attempts and long-term mental harm.

Countless deadly cyberbullying episodes have been observed around the world, indicating the devastating impact. As a result, relevant actions must be taken against bullies, and social media users must be prevented from be- coming victims. We attempted to develop an automatic system that could detect bullying activity in online social networking platforms so that the government could take the necessary steps to prevent bullies from causing harm to a large number of users.

Students, in particular, and members of society in general, have come to believe that cyberbullying is ”no big issue,” or that harassing others repeatedly is accept- able if a large number of incidents go unpunished. Furthermore, the entire schoolyard will realize the gravity of the matter and the reality that these actions will have consequences.

This, however, cannot be accomplished simply relying on the victim or bystander to report the incidents. This is why the vast majority of cyberbullying instances go unreported and consequently penalized. Due to the vast volume of data created on social media, manual identification of this work would be too impossible to perform. Although there are over 100 social media websites, Facebook, Ask.FM, and Twitter have been identified as the most likely causes of cyberbullying. As a result, the focus of this study is on identifying cyberbullying activity in a publicly available Twitter dataset. Twitter is a popular microblogging platform. Registered users can read or publish messages of no more than 280 character, known as “Tweets.” Registered users can also share photos and videos on Twitter. Unregistered individuals can read tweets that have been made public. Although cyberbullying can take many forms, such as publishing embarrassing messages, images, and videos, for the sake of this thesis, cyberbullying detection is limited to textual cyberbullying detection in a Twitter dataset.

1. **Objectives**

The goal of this study is to develop a prototype that can automatically detect cyberbullying and abusive behavior on social media and online communities by:

* Extracting, collecting, and labelling the data set.
* Improve accuracy by pre-processing, cleaning, and exper- imenting with various features.
* Text, comment, or post classification into one of the many classes.
* The best model’s evaluation and analysis.

1. **LITERATURE REVIEW**

***Improving Cyberbullying Detection using Twitter Users’ Psychological Features and Machine Learning***

**Publication details:** Computers and Security, Elsevier

**Authors:** Vimala Balakrishnan, Shahzaib Khan ,Hamid R. Arabnia [1]

**Year of Published:** 2020

**Publication details:** Computers and Security, Elsevier

**Authors:** Vimala Balakrishnan, Shahzaib Khan ,Hamid R. Arabnia [1]

# Year of Published: 2020

1. ***Objective:****:* This research focuses on improving the cyberbullying detection system by tapping into Twitter users’ psychological characteristics, such as personalities, sentiments, emotions and Twitter features.
2. ***Summary:****:* This model for detecting cyberbullying takes into account user personalities, sentiments, emotions, and Twitter-based features.

Personality is determined based on the Big Five model (Costa Jr. and Mc-Crae,1992; John and Srivastava, 1999) which is as follows:

* Extraversion:- The person is socially active pays more attention to events and is full of excitement.
* Agreeableness:- The person is kind, gentle, and gets along with other people.
* Conscientiousness:- it presents how much a person pays attention to others when making decisions.
* Openness:- The person is creative, broad-minded, adjustment making. Dark Triad is used to see the darker side of the user’s personality.
* Machiavellianism:- Person having Empathy lacking and impulsive behavior.
* Agreeableness:- The person is kind, gentle, and gets along with other people.
* Psychopathy:- Having a tendency to manipulate others.
* Sentiments:- Concerning Cyberbullying, sentiments are thoughts and opinions that can be categorized as bullies, victims, and non-bullies.
* Emotion Analysis:- Generally the common feelings that are expressed through texts while cyberbullying behaviors are, anger, embarrassment, empathy, fear, pride, relief, and sadness which are called Emotions.
* Twitter features:- Features like the time when the account was created and text/content features like the number of hashtags, number of symbols, number of user mentions, etc. are considered. The overall model has three main stages, namely, Twitter data collection, feature extraction, and cyberbullying detection and classification. Here Twitter Dataset with 9484 annotated tweet IDs was used. The three classes of the labeling dataset were as follows: Bully, Aggressor, and Spammer.

While Pre-processing non-English tweets were removed and tweets containing only special characters such as numbers, punctuations, and stop words were removed. Twitter user profiles were examined so as to include only active users. The final.

Twitter dataset contained 5453 tweets in which Abusive users make up about 8% of the dataset. The classifiers used were trained which classify the tweets in cyber bullying detection model into bully, aggressor, spammer, or normal. Random Forest and J48 were performed for the cyberbullying classifications.

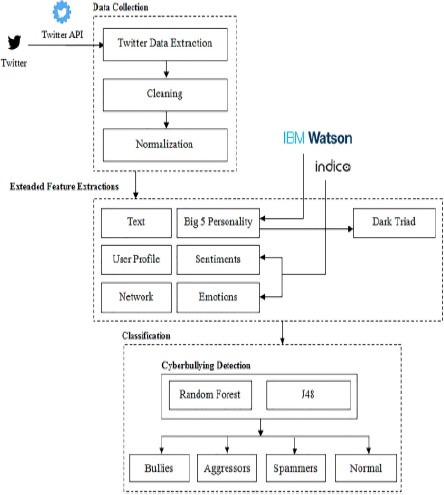


Figure 2.1: Multiple Feature Cyberbullying Detection Model [[1](#_bookmark85)]

***Early detection of cyberbullying on social media networks***

**Publication details:** Future Generation Computer Systems, Elsevier

**Authors:** Manuel F. Lopez-Vizcaino, Francisco J. Novoa, Victor Carneiro [2]

# Year of Published: 2021

1. ***Objective:****:* To explore different approaches that take into account the time in the detection of cyberbullying in social networks and to investigate the early detection of cyberbullying.
2. ***Summary:****:* A supervised learning method with two different specific early detection models, named threshold and dual is performed. Two groups of features and two early detection methods are specifically designed for this problem. The problem of early detection of Cyberbullying is defined and characterized. Two machine learning models (i.e. threshold and dual) for the cyberbullying early detection problem and two sets of features contribute to improving the performance in the cyberbullying early detection problem. Extensive experiments using a real-world dataset and following a time- aware evaluation were performed. In this paper, for social media sessions, the function posts from 1 to k act as input, and three possible values are returned: 0, 1, and 2. Now,

0 corresponds to a session that is considered normal 1 is considered cyberbullying, and 2 if no definitive decision can be made.

The threshold model is trained using characteristics such as the number of followers and followings of the user, number of likes and comments, and percentage of negative commands, which incorporates a decision function based on the class probabilities to assess whether there is enough evidence to proceed with a firm judgment. The dual model is made up of two separate learning models.

One model is trained to recognize positive cases, while the other is taught to detect negative cases. Based on the use of all features for the identification of positive cases along with low thresholds to produce early detections, and simpler features for profile owners for the negative model, the dual model consistently provides the best performance for the early detection of cyberbullying. Features of the model can be extended by considering comments, as these concentrate most of the information for early detection. Early detection can be achieved by investigating an evaluation based on time, instead of a number of posts. Experimentation with other datasets from some other social media platforms can be done to validate the approach and generalize the results.

***Exploring the hidden patterns of cyberbullying on social media***

**Publication details:** Procedia Computer Science, Elsevier **Authors:**Swaranjit Singh, Vivek Thapar, Sachin Bagga [12]

**Year of Published:** 2020

1. ***Objective****:* Social Network Analysis is done to analyze the Twitter network of Momo-challenge
2. ***Summary****:* In this research paper, Node XL, an open- source tool managed by a social media research foundation, is used to collect and analyze data. Three analysis tech- niques are developed in the proposed system for discovering cyberbullying. They are network analysis, content analysis, and graph-based network visualization analysis. Network and content analysis is used to scrutinize the ways by which the victims are targeted, the behavior of the users, and how the information spreads and sentiment analysis of the user tweets to detect opinions of the users about cyberbullying.

Graph- based visualizations are used to find the network patterns of MomoChallenge. For the Momo challenge, the current study has new methods for detecting cyberbullying on Social media networks’ important parts and network patterns. The authors also proposed machine learning classification techniques, such as the Random Forest classifier, can be used to conduct behavioral analysis of users and to investigate hidden patterns of cyberbullying. The research can be extended by using the same detection method on other social networking platforms and in other languages than English.

***Automatic classification of particular roles in cyberbullying: can we detect victim, bullies and bystanders in social media text?***

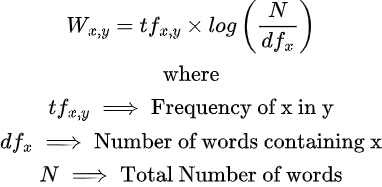
**Publication details:** Cambridge University Press

**Authors:** Gilles Jacobs, Cynthia Van Hee and Veronique Hoste [13]

**Published:** 2020

1. ***Objective:****:* To automatically detect different participant roles involved in textual cyberbullying, including bullies, victims, and bystanders.
2. ***Summary:****:* Two cyberbullying corpora (a Dutch and English corpus) were taken, both manually annotated with bullying types and participant roles. Fine-grained annotation guidelines were developed to enable the annotation of participant roles in cyberbullying. Series of multi-class classification experiments are performed to determine the feasibility of text-based cyberbullying participant role detection. Two different experimental setups are carried out, one where they optimized and compared linear task-specific classification algorithms, another one where they explored the performance of fine-tuning pre-trained transformer models for this task.

In both experimental setups, the participant roles can be classified with satisfactory results. The transformer-based models RoBERTa and RobBERT achieved the highest scores for English and Dutch. While Roberta outperformed the best Cascading classifier by 5%, the difference between RobBERT and the Voting classifier is less (1.6%). This research can be carried ahead by investigating other conversational data genres that allow for the incorporation of user and context information and by investigating various methodologies for using DNNs in problems with a high level of class imbalance.

**CHAPTER 3 PROPOSED WORK**

***3.1 Proposed Model***

We present our cyberbullying detection model in this section. We examined Twitter profiles and the tweets associated with them to check for indicators of bullying. This section includes information on data collecting, necessary data pre- processing, feature engineering, implementation, and results.

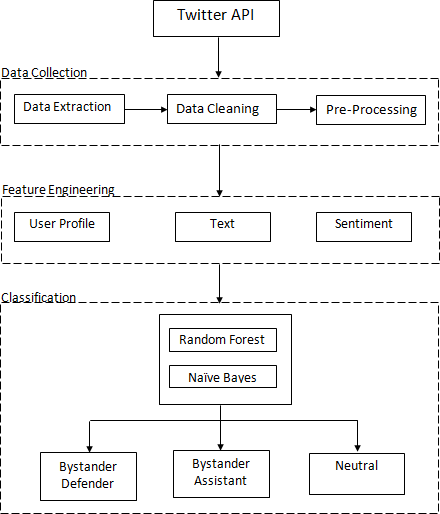
***3.0.1 Data Collection:***By Using the Streaming API data can be downloaded on real time from Twitter. A Python Library called Tweepy is used here to connect to the Twitter API and download the tweets from there.

Figure 3.1: Cyberbullying Detection Model

The tweets are downloaded with specific keywords or hashtags. After extracting the tweets are saved in .CSV file.

· Import libraries such as tweepy, Sentiment Intensity Analyser, pandas, numpy, Counter, re, pyplot, json, nltk, train test split, stopwords

* First of all we need scrape data from twitter api
* Assign tweepy.cursor to tweets variable
* For tweets in list tweets do:

Username is equal to tweet.user.screen name, Description is equal to tweet.user.description

Location is equal to tweet.user.location

Following is equal to tweet.user.friends count Followers is equal to tweet.user.followers count

Total tweets is equal to tweet.user.statuses count Retweet count is equal to tweet. Retweet count

Hashtags is equal to tweet. Entities[’hashtags’]

* Create a list of hashtext

***3.0.2 Data Pre-processing:***Data Pre-processing is done for conversion of aren’t to are not. Regular expression (re) is used for preprocessing the tweets. The tweets are pre-processed by converting the string into Lower-cases, Removing usernames, Removing URLs, Removing all digits, Removing &, Removing all single characters, Replacing con- tractions, Removing all punctuation and Replacing double spaces with single spacing. Lemmatize is also used for conversion of words into root words like Changing: change, Played: Play

* Now load Contractions from contractions.json file using pandas read the scraped data.csv file.
* After that replace emojis with its emoji sentiment then remove all single character

Then after replace all contraction present in the tweet using Contractions () function defined before also, remove all punctuation, double spacing present in it

Return processed tweet

* Using the vectorize function of numpy, iterate

through all the ‘text’ present in the dataset and process each with process text function defined before, while adding this process text to subsequent data rows in the new ‘processed text’ column.

***3.0.3 Feature Engineering:***

**Profile features:** While scraping the data from Twitter following are the features which are considered. Username, id description, user location, user friends count, user followers count, total tweets, retweet count and hashtags.

**Feature Extraction:** TfidfVectorizer is used which employs an in-memory vocabulary to map the most common words to feature indices, resulting in a sparse word oc- currence frequency matrix. The term tfidf stands for term frequency–inverse document frequency.

It is a mathematical statistic intended to reflect the importance of a word to a record in a collection or corpus. The tfidf esteem develops in direct proportion to the number of times a word appears in the document.

***3.0.4 Sentiment Feature::***Each tweet’s positive and negative, neutral and compound score is calculated using Vader Sentiment Analyzer. It is an abbreviation for Valence Aware

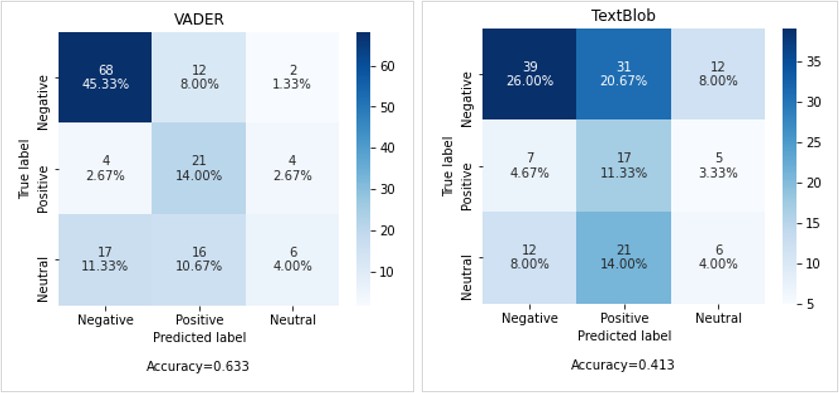
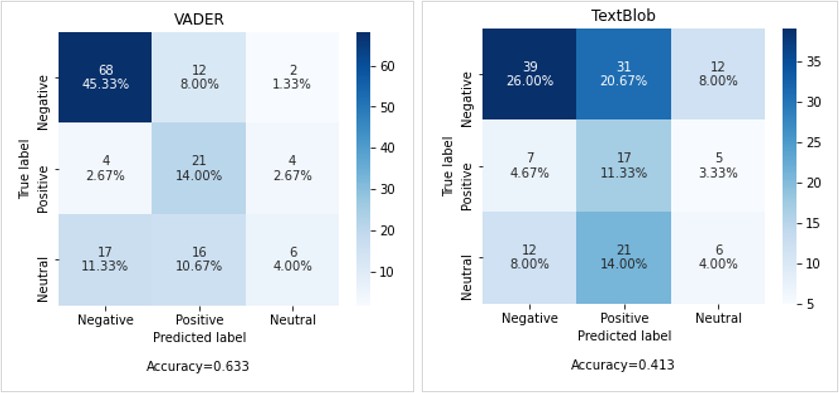
Dictionary and Sentiment Reasoner. It is a sentiment analysis lexicon and simple rule-based model. It can handle vocabularies, abbreviations, capitalizations, repeated punctuation, and emoticons with ease and is commonly used on social media platforms to express one’s sentiment.

Figure 3.2: Difference between Vader and Text Blob

***3.1 Steps for bullying detection***

Step 1: Extract the data by Twitter API.

Step 2: Clean tweet text by removing links, special characters and other impurities.

Step 3: Removal of words that don’t contribute to the sentiment.

Step 4: Use “vadersentiment” to identify the sentiment compound score for classifying tweets.

Step 5: Applying Naıve Bayes, Random Forest and Pro- posed Algorithm.

**CHAPTER 4- IMPLEMENTATION**

## Extract Twitter Data

For extracting Twitter data, a python program was developed. The user’s User- name, id description, user location, user friends count, user followers count, total tweets, retweet count and hashtag data were collected using the Twitter API and the information was saved as a text/csv file. In the program, it asks for keyword to be searched as a input from user and also the date since which we need tweets to be fetched. For getting data through Twitter API, Authentication was done.

## Data Pre-processing

Data pre-processing is a type of processing that converts data into a readable format. Here the tweets are pre-processed by converting the string into Lower cases, Removing user- names, Removing URLs, Removing all digits, Removing (&quot;), Removing all single character, Replacing contractions, Removing all punctuation’s and Replacing double spaces with single spacing. Lemmatizer is also used for conversion of words into root words like Changing: change, Played: Play. Stop words are also removed from the tweets as it takes up space and the processing time.

## Feature Engineering

For transforming text into a vector/matrix, Text vectorization algorithm namely TF-IDF vectorizer is used. Tf-idf is a measure of origanility of work by comparing the number of times the word appeared.

## Classify Tweets

Vader Sentiment Analyzer is used to calculate the positive and negative, neutral, and compound scores for each tweet. The compound score is the sum of the positive, negative, and neutral scores, normalized between -1 and 1. In this case, 1 represents the most severe negative and +1 represents the most extreme positive. The closer the Compound score is to +1, the more optimistic the text. Also Naive Bayes and Random Forest Classifier are used for the training and testing of data.

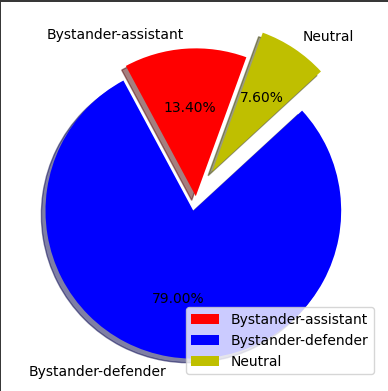


Figure 4.1: Sentiment Analysis

## E. Performance measures

Precision, recall, F1-score and accuracy were the four evaluation metrics which are used.

*Precision:* It is defined as the ratio of correctly classified data elements (for both the minority and majority classes) to the total number of classified instances.

P =TP/ (TP + FP)

***Recall:*** The proportion of correctly classified minority class instances to the total number of minority class instances.

R =TP/ (TP + FN)

***F-Measure Calculation:*** Precision and Recall are used to calculate F-measure. Precision & Recall’s harmonic mean is used to calculate it. It’s possible to say that it’s essentially the average of the two percentages. It significantly simplifies the comparison of classifiers.

F-measure = 2 / (1/R + 1/P)

# Confusion Matrix

* 1. **True Positive (TP):** The True positive value is where the actual value and predicted value are the same.
  2. **False Positive (FP):** The False-positive value for a class will be the sum of values of the corresponding column except for the TP value.
  3. **True Negative (TN):** True Negative value for a class will be the sum of values of all columns and rows except the values of that class that we are calculating the values for.
  4. **False Negative (FN):** The False-negative value for a class will be the sum of values of corresponding rows except for the TP value.

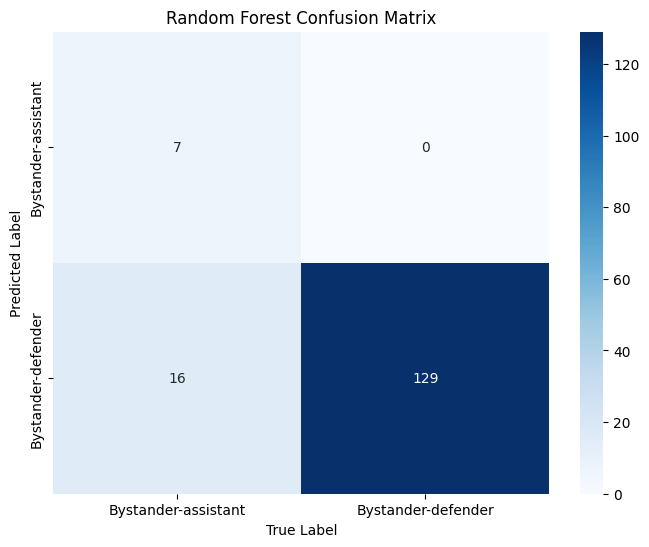


Figure 4.2: Confusion Matrix

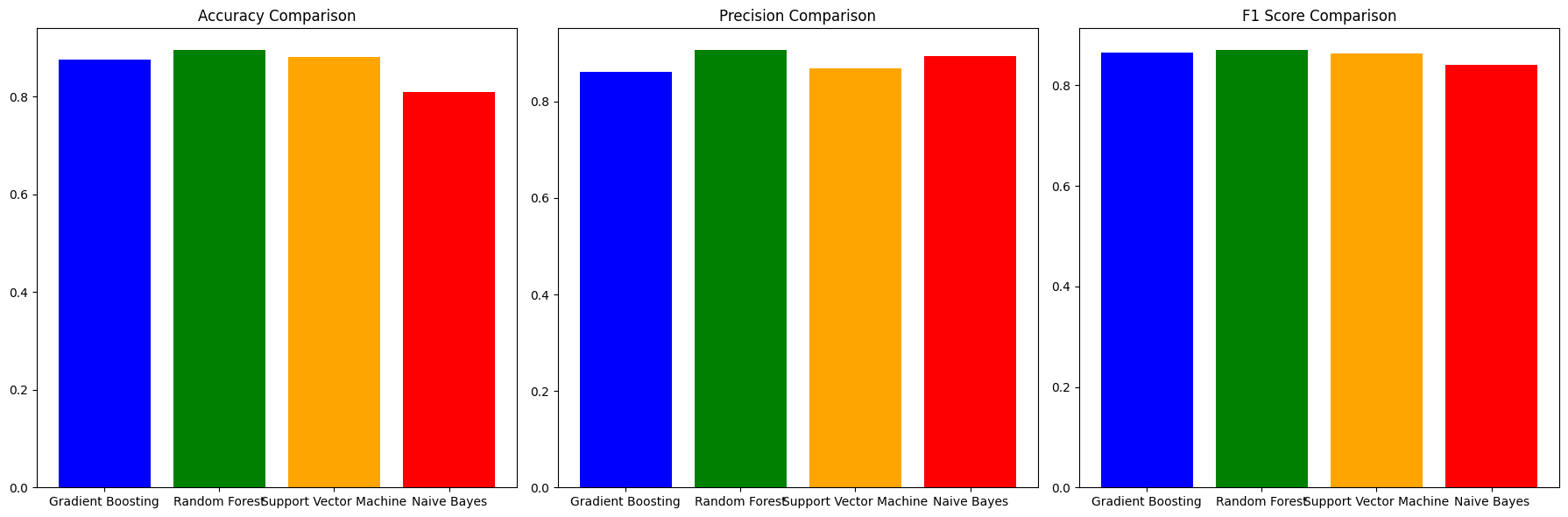
## F. Experimental Results

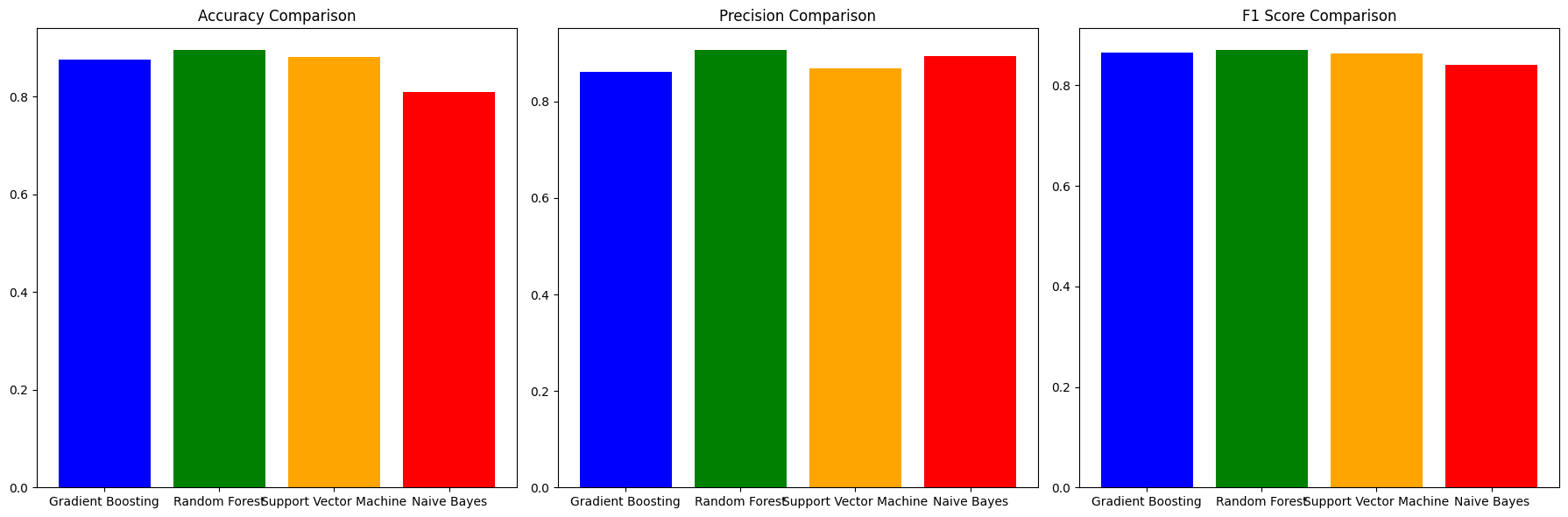
The Python programming language was used to conduct all of the experiments. Table shows the results of the analyses, including precision, recall, F1- score and accuracy. The performance of various machine learning algorithms is evaluated as per the live data collected on the basis of keyword search. The data is divided into two sets, one for training and the other for testing, both of which are mutually exclusive.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| Navie Bayes | 79.3% | 0.83 | 0.79 | 0.81 |
| Random Forest | 84.8% | 0.87 | 0.85 | 0.81 |
| SVM | 84.2% | 0.84 | 0.84 | 0.81 |
| Gradient Boosting | 85.4% | 0.86 | 0.85 | 0.83 |

Table 4.1: Performance of cyberbullying detection on twitter data

The training set are divided into the batches of 30%, 20% and 15% of the dataset, while the respective testing set makes up 70%, 80% and 75% of the dataset. For a single machine learning algorithm, different classifiers and methodologies exist. According to the results of the above analysis, the proposed methodology is the best machine learning methodology.





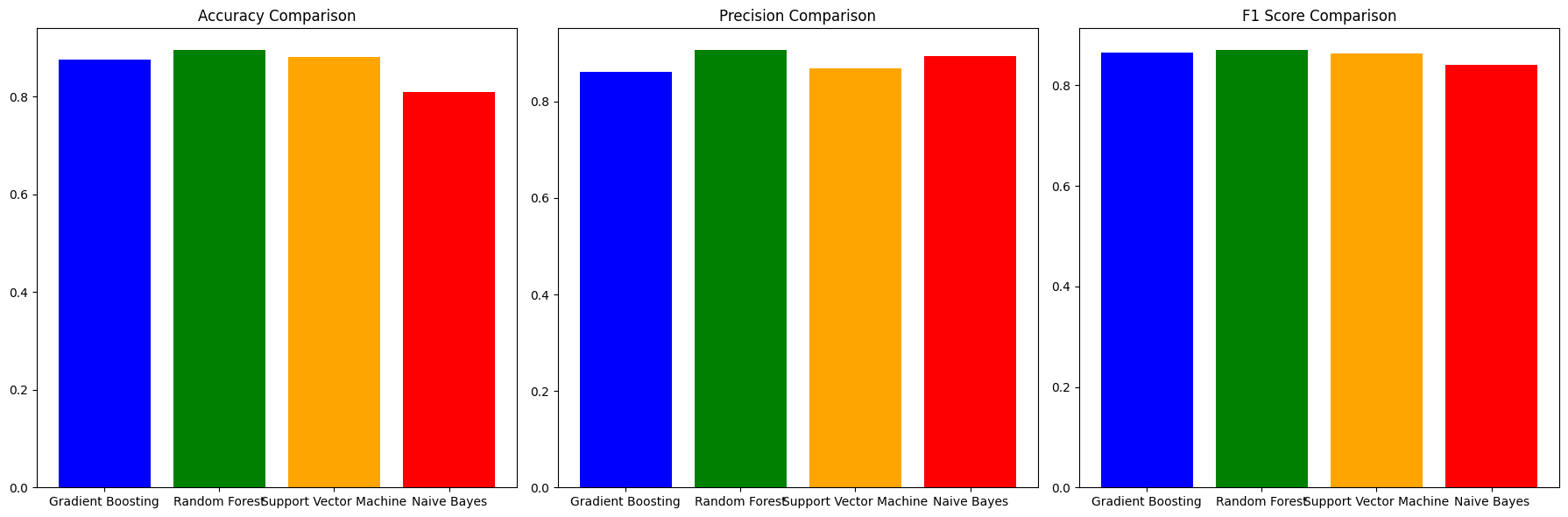


Figure4.3: Precision, Recall, F1-score of Proposed Algorithm

**CHAPTER 6- CONCLUSION AND FUTURE WORK**

This study examined available literature to detect cyberbullying on social media websites using several methodologies. Using machine learning approaches, we synthesized and determined the key parameters for detecting cyberbullying.

By reviewing existing literature we conclude four aspects of detecting bullying messages by using machine learning approaches, namely, data collection, Data Pre-processing, Sentiment Analysis and train & test data. The Proposed Naive Bayes achieves the best accuracy among the available machine learning methods, according to an experimental result.

In the future, the goal is to conduct in-depth literature reviews and social network analysis of various social media platforms. Various more machine learning classification approaches can be utilized to conduct behavioral analysis on users and investigate cyberbullying tendencies and Age and gender detection using machine learning classifiers.

**REFERENCES**

[1] V. Balakrishnan, S. Khan, and H. R. Arabnia, Improving cyberbullying de- tection using Twitter users’ psychological features and machine learning,” Comput. Secur., vol. 90, no. March, 2020, doi:10.1016/j.cose.2019.101710.

[2] M. F. Lopez-Vizcaino, F. J. Novoa, V. Carneiro, and F. Cacheda, ”Early detection of cyberbullying on social media networks,” Futur.Gener. Comput. Syst., vol. 118, pp. 219-229, 2021, doi: 10.1016/j.future.2021.01.

[3] Agrawal, S., & Awekar, A. (2018). Deep Learning for Detecting Cyberbullying Across Multiple Social Media Platforms. Advances in Information Retrieval, 141–153. doi:10.1007/978-3-319-76941-7-11

[4] Department of Justice, Canada.”Cyberbullying and the Non- consensual Distribution of Intimate Images” last accessed on 10 May 2022. tml

[5] C. Van Hee et al., Automatic detection of cyberbullying in social media text,” PLoS One, vol. 13, no. 10, pp. 1-22, 2018, doi: 10.1371/jour- nal.pone.0203794.

[6] P. K. Smith, J. Mahdavi, M. Carvalho, S. Fisher, S. Russell, and N.Tippett, Cyberbullying: Its nature and impact in secondary school pupils,” J. Child Psychol. Psychiatry Allied Discip., vol. 49, no. 4,pp. 376-385, 2008, doi: 10.1111/j.1469-7610.2007.01846.x.

[7] Amrita Mangaonkar, MSc. Thesis, “Collaborative Detection of Cyberbullying Behaviour in Twitter Data.” Faculty of Purdue University, August 2017.

[8] K. Kopecky and R. Szotkowski, ”Cyberbullying, cyber aggression and their impact on the victim - The teacher,” Telemat. Informatics, vol.34, no. 2, pp. 506-517, May 2017, doi: 10.1016/J.TELE.2016.08.014.

[9] Tae Hoon Kim ”5 Different Types of Cyberbullying” last accessed on 10 May 2022.

[10] Kargutkar, S. M., Chitre, V. (2020). A Study of Cyberbullying De- tection Using Machine Learning Techniques. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc48092.2020.iccmc-000137.

[11] M. Anzovino, E. Fersini, and P. Rosso, Automatic identification and classifi- cation of misogynistic language on twitter,” in Natural Language Processing and Information Systems, Cham: Springer International Pub- lishing, 2018, pp. 57-64.

[12] S. Singh, V. Thapar, and S. Bagga, ”Exploring the hidden patterns of cyber- bullying on social media,” Procedia Comput. Sci., vol. 167, pp. 1636-1647, 2020

[13] G. Jacobs, C. Van Hee, and V. Hoste,”Automatic classification of participant roles in cyberbullying: Can we detect victims, bullies, and bystanders in social media text?” Natural Language Engineering, pp. 1- 26, 202