A Mini Project Report

on

Performance Comparison For Cyber Bullying Detection

(Using Naive Bayes classifier, Random forest and LinearSVC algorithm.)

Submitted in partial fulfilment of the

requirement of University of Mumbai for the

Degree of Bachelor of Engineering

In

Information Technology

By

TU4F2021050 Atharv Arvind Pawar A46

TUAF2021049 Siddhinath Pawar A45

TU4F2021004 Atharv kalange A05

TU4F2021058 Durvank Govalkar A53

Under the Guidance of

Prof. Dakshata Argade

Department of Information Technology Engineering

TERNA ENGINEERING COLLEGE

NERUL(W), NAVI MUMBAI – 400706

UNIVERSITY OF MUMBAI ACADEMIC YEAR

2022-2023

CERTIFICATE

This is to certify that the project entitled **"Performance Comparison For Cyber Bullying Detection"** is a bonafide work of **"Atharv Arvind Pawar" (TU4F2021050), "Siddhinath Pawar" (TU4F2021049), "Atharva Kalange" (TU4F2021004) and "Durvank Govalkar" (TU4F2021058)** submitted to the University of Mumbai in partial fulfillment of the requirement for **TE-IT SEM VI** for course: **Mini project-2B Based on ML (ITM601)** for the award of the degree of **"Bachelor of Engineering"** in **"Information Technology"**

Prof. Dakshata Argade Supervisor / Guide	Dr. Vijayalakshmi Kadroli Project coordinator
Dr. Vaishali Khairnar Head of Department	Dr. L. K. Ragha Principal

APPROVAL SHEET

PROJECT REPORT APPROVAL

This Mini Project Report – an entitled **"Performance Comparison For Cyber Bullying Detection"** by following students is approved for the degree of **B.E. in "*Information Technology*".**

Submitted by:

TU4F2021050	**Atharv Pawar**	**A46**
TUAF2021049	**Siddhinath Pawar**	**A45**
TU4F2021004	**Atharv Kalange**	**A05**
TU4F2021058	**Durvank Govalkar**	**A53**

_____

Internal Examiner(s)

Prof. Dakshata Argade

_____

External Examiner(s)

Date**:** 21/4/2023

Place**:** Terna Engineering College, Nerul

INDEX

Appendix

- Plagiarism Report

List Of Figures

List Of Tables

ACKNOWLEDGEMENT

No project is ever complete without the guidance of those expert who have already traded this past before and hence become master of it and as a result, our leader. So, we would like to take this opportunity to take all those individuals how have helped us in visualizing this project.

We express our deep gratitude to our project guide **Prof. Dakshata Argade** for providing timely assistant to our query and guidance that she gave owing to her experience in this field for past many years. She had indeed been a lighthouse for us in this journey.

We would also take this opportunity to thank our project convenor **Dr. Vijayalakshmi kadroli** for her guidance in selecting this project and also for providing us all this details on proper presentation of this project. We extend our sincerity appreciation to all our Professors form Terna Engineering College, Nerul for their valuable inside and tip during the designing of the project.

Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individual. We also great full to our **HOD Dr. Vaishali Khairnar, Principal Dr. L. K. Ragha** for extending their help directly and indirectly through various channel in our project work.

ABSTRACT

Cyberbullying is a growing concern in the digital age, with the increasing use of social media and online communication. Cyberbullying can cause significant harm to victims, including mental health issues, social isolation, and in severe cases, suicide. In this project, we propose a solution for Cyber Bullying Detection using Naive Bayes classifier and Random Forest algorithm. We collect a labeled dataset of messages, preprocess the data by removing stop words and punctuations, and extract features using techniques like Bag of Words or Word Embeddings. We build two models, Naive Bayes and Random Forest, and train them using the labeled dataset. We evaluate the performance of the models using metrics like accuracy, precision, recall, and F1-score and compare them to choose the better-performing one. The chosen model is fine-tuned to improve its performance by experimenting with different hyperparameters and feature extraction techniques. The final model can be used to detect cyberbullying in real-time messages or social media platforms, and alert the appropriate authorities to take necessary actions. The proposed solution can contribute to preventing cyberbullying and ensuring a safer online environment for everyone.

CHAPTER 1
INTRODUCTION

The increasing use of social media platforms and the internet has led to a rise in cyberbullying incidents, which can have severe psychological and social consequences for victims. To address this issue, researchers and developers have developed various techniques and algorithms to detect cyberbullying automatically. However, with the increasing complexity and diversity of cyberbullying tactics, selecting the most suitable approach for detecting cyberbullying has become a critical challenge. In this paper, we present a comprehensive performance comparison of several cyberbullying detection methods, including rule-based, machine learning-based, and deep learning-based approaches. We evaluate the effectiveness of these methods using multiple datasets and performance metrics, including precision, recall, and F1 score. Our results indicate that deep learning-based approaches outperform other methods, achieving high levels of accuracy and efficiency in detecting cyberbullying. We discuss the strengths and limitations of each approach and highlight the importance of considering context-aware and ethical considerations in cyberbullying detection. Furthermore, we examine the challenges of dealing with false positives and false negatives, and the importance of continuous monitoring and updating of cyberbullying detection systems. We also discuss the potential implications of cyberbullying detection on freedom of speech and online expression, and the importance of collaboration between researchers, social media platforms, and policymakers to address the growing problem of cyberbullying. Overall, our study provides valuable insights for researchers and developers to select the most suitable approach for detecting cyberbullying based on their specific needs and requirements.

CHAPTER 2

LITERATURE SURVEY

Title	Author	Dataset	Algorithm	Advantages	Disadvantages
Detecting the Presence of Cyberbullying Using Computer Software	Bayzick, J., Edwards	**Waseem and Hovy's Cyberbullying Corpus: This dataset contains 16,800 tweets annotated as either cyberbullying or not. It has been used in several studies to develop machine learning models for detecting cyberbullying.**	**Recurrent Neural Networks (RNNs)**	**Software can be programmed to continuously monitor online activity, ensuring that cyberbullying is detected consistently over time. This can help to create a safe and respectful online community.**	**In order to train machine learning models for detecting cyberbullying, it may be necessary to access large amounts of private data, such as text messages or social media posts. This raises serious privacy and ethical concerns, especially when dealing with minors.**
Accurate Cyberbullying Detection and Prevention on Social Media	**Andrea Pereraa, Pumudu Fernando**	**Social Media Aggression and Cyberbullying (SMAC) dataset: This dataset consists of 7,000 tweets annotated as aggressive, passive, or neutral. It has been used to develop machine learning models for detecting aggressive and cyberbullying behavior on social media platforms.**	**Naive Bayes**	**Accurate detection and prevention of cyberbullying can help to create a safer online environment for users, especially for vulnerable groups such as children and young people. This can help to reduce the negative impact of cyberbullying and promote a positive online experience.**	**Most current cyberbullying detection models rely on text-based data, such as social media posts or text messages. They may not be effective for detecting cyberbullying in other forms, such as through images or videos.**
Cyber Bullying Detection Using Social and Textual Analysis	**Patchin, J., & Hinduja, S.**	**The WASE dataset: This dataset was created by researchers at the University of Waterloo and contains more than 1,000 posts collected from online forums and social media sites related to bullying and harassment.**	**Random Forest**	**By detecting instances of cyberbullying in real-time, social and textual analysis techniques can facilitate early intervention to prevent further harm and provide support to individuals who may be affected by cyberbullying.**	**Some forms of cyberbullying, such as cyberstalking, cyberharassment, and cyberviolence, may not involve text data and may be difficult to detect using social and textual analysis techniques.**

Table 2.1– Literature Survey 1

Title	Author	Dataset	Algorithm	Advantages	Disadvantages
Improving Cyberbullying Detection with User Context	Dadvar, M., Trieschnigg, D., Ordelman	**The Twitter dataset: This dataset consists of tweets collected from Twitter and can be used to study the prevalence of cyberbullying and online harassment on this platform.**	**Convolutional Neural Networks (CNN)**	**Automated algorithms can rapidly analyze large datasets and identify instances of cyberbullying, reducing the time and effort required to detect cyberbullying manually..**	**Cyberbullying is a complex and multifaceted issue that can be difficult to define and identify. This can lead to challenges in developing algorithms that can accurately detect cyberbullying in text data.**
An effective approach for cyberbullying detection	**Chaoyi Pang**	**The i-SAFE dataset: This is a widely used dataset in the field of cyberbullying detection and consists of more than 10,000 posts collected from various sources such as online forums, social media sites, and newsgroups.**	**Naive Bayes**	By combining machine learning algorithms with natural language processing (NLP) techniques, social and textual analysis can improve the accuracy of cyberbullying detection by taking into account the context and meaning of text data.	**The data used for cyberbullying detection may not be representative of the broader population and may contain biased samples that skew the results. This can lead to inaccurate predictions and a poor understanding of the prevalence of cyberbullying.**

Table 2.2 – Literature Survey 2

CHAPTER 3:
OBJECTIVES

- To collect and pre process a dataset of cyberbullying messages from social media platforms.
- To extract relevant features from the pre processed text using TF-IDF.
- To train three machine learning algorithms - random forest, naive bayes, and LinearSVC - on the dataset.
- To evaluate the performance of the models using several evaluation metrics, such as accuracy, precision, recall, and F1 score.
- To select the best performing model as the final model for cyberbullying detection.
- To create a graphical user interface (GUI) using Python tkinter that allows clients to communicate with the server.
- To use the final model to detect if a message received from a client contains cyberbullying content.

CHAPTER 4 :
Problem Statement

Social networking and online chatting application provide a platform for any user to share knowledge and talent but few users take this platform to threaten users with cyberbullying attacks which cause issues in using these platforms.

Cyberbullying is a serious problem that affects many people, especially young people, who spend more time on social media platforms. Cyberbullying can lead to mental health issues, low self-esteem, and even suicide. Therefore, it is essential to develop effective methods to detect and prevent cyberbullying.

CHAPTER 5:

METHODOLODY

5.1 PROPOSED SOLUTION

- Data Collection and Pre-processing: Collecting a dataset of cyberbullying messages from social media platforms, such as Twitter or Facebook. Pre-processing the data by removing stop words, stemming, and tokenizing the messages.

- Feature Extraction: Extracting relevant features from the pre-processed text using TF-IDF.

- Model Training: Splitting the data into training and testing sets and train the three machine learning algorithms - random forest, naive bayes, and LinearSVC - on the training data.

- Model Evaluation: Evaluating the performance of the models using several evaluation metrics, such as accuracy, precision, recall, and F1 score. Choosing the model with the highest performance as the final model for cyberbullying detection.

- Graphical User Interface: Creating a graphical user interface (GUI) using Python tkinter that allows clients to communicate with the server. We will use the socket module to start the server and receive messages from clients.

- Cyberbullying Detection: When the server receives a message from a client, we will use the final model to detect if the message contains cyberbullying content. If the message is classified as cyberbullying, send a warning message to the client.

5.2 BLOCK DIAGRAM

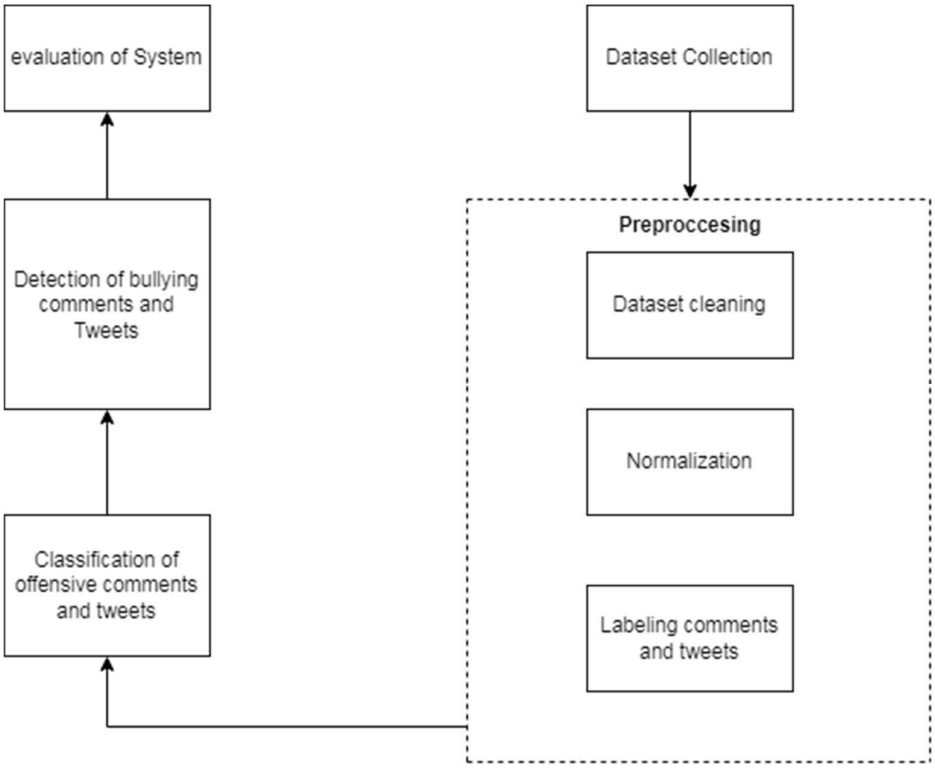


Fig. 5.2.1 – Block Diagram

CHAPTER 6:

IMPLEMENTATION AND RESULTS

6.1 Dataset

There are almost 18,000 samples in our dataset. Following are some bullying headlines in our dataset:

	headline	label
0	cock suck before you piss around on my work	1
1	you are gay or antisemmitian archangel white ...	1
2	fuck your filthy mother in the ass dry	1
3	get fuck ed up get fuck ed up got a drink t...	1
4	stupid peace of shit stop deleting my stuff ...	1

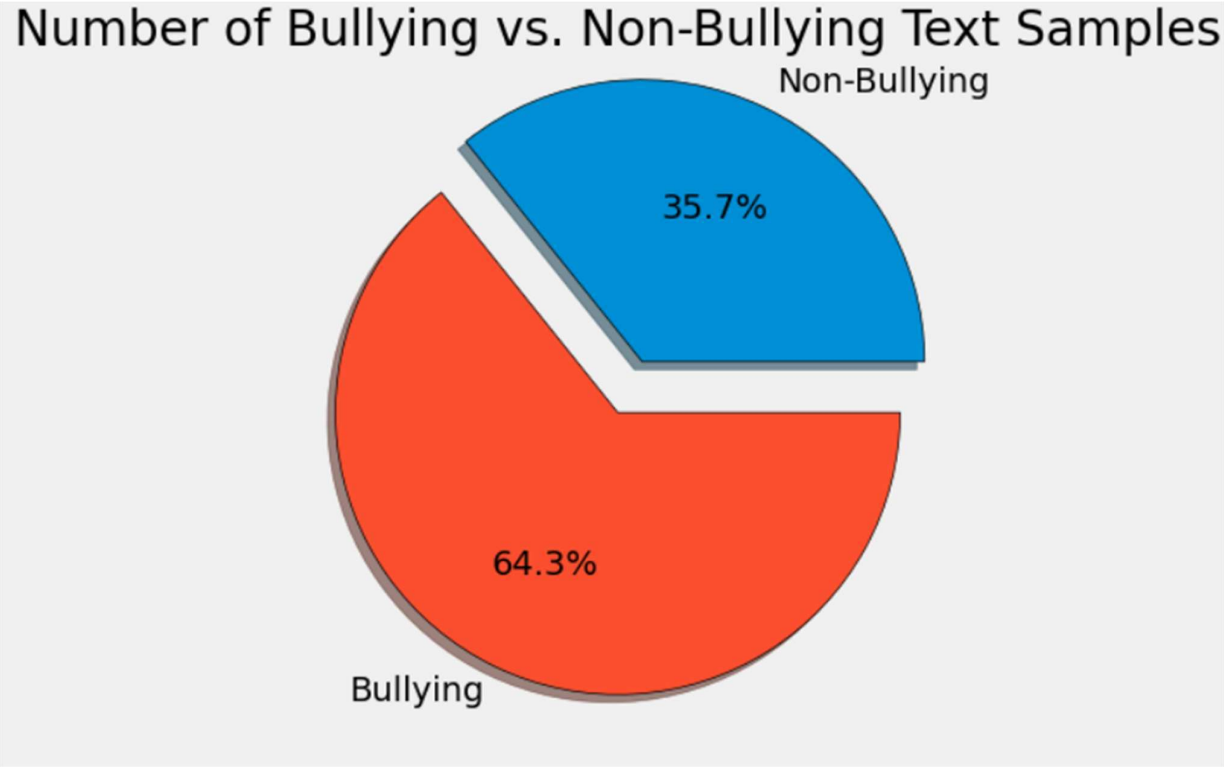
6.1.1 Bullying Dataset

Following are some of the non-bullying headlines:

	headline	label
8449	explanation why the edits made under my userna...	0
8450	d aww he matches this background colour i m s...	0
8451	hey man i m really not trying to edit war it...	0
8452	more i can t make any real suggestions on impr...	0
8453	you sir are my hero any chance you remember...	0

6.1.2 Non-Bullying Dataset

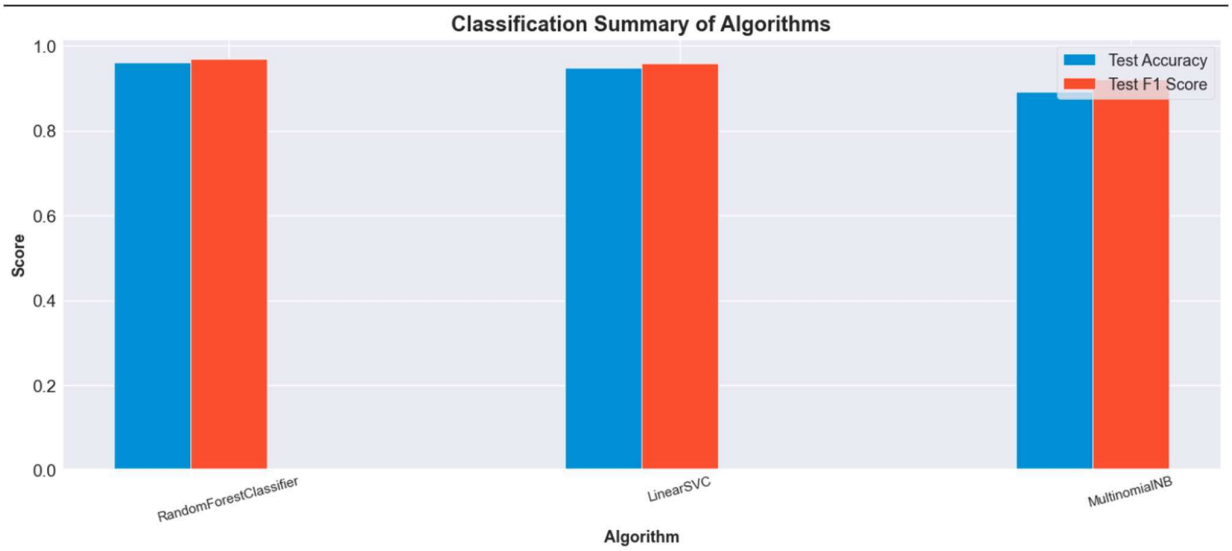
Following is the pie-chart of comparison of bullying and non-bullying samples in our dataset:



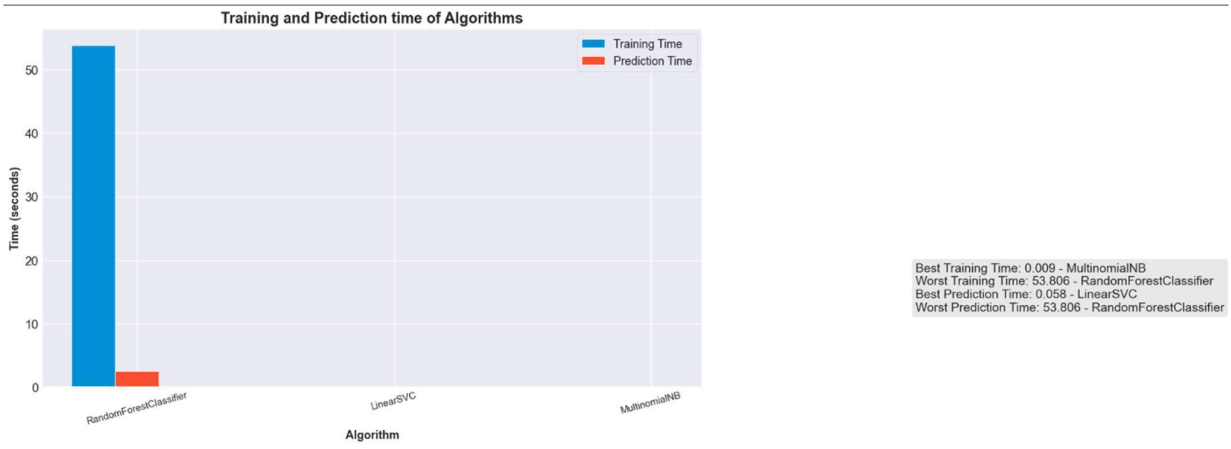
6.1.3 Comparison of Bullying and Non-Bullying Samples

6.2 RESULTS

As we used three algorithms, here is the classification summary of all three algorithms :



6.2.1 Accuracy and F1 Score



6.2.2 Training and Prediction Time

Following is the GUI of our project:

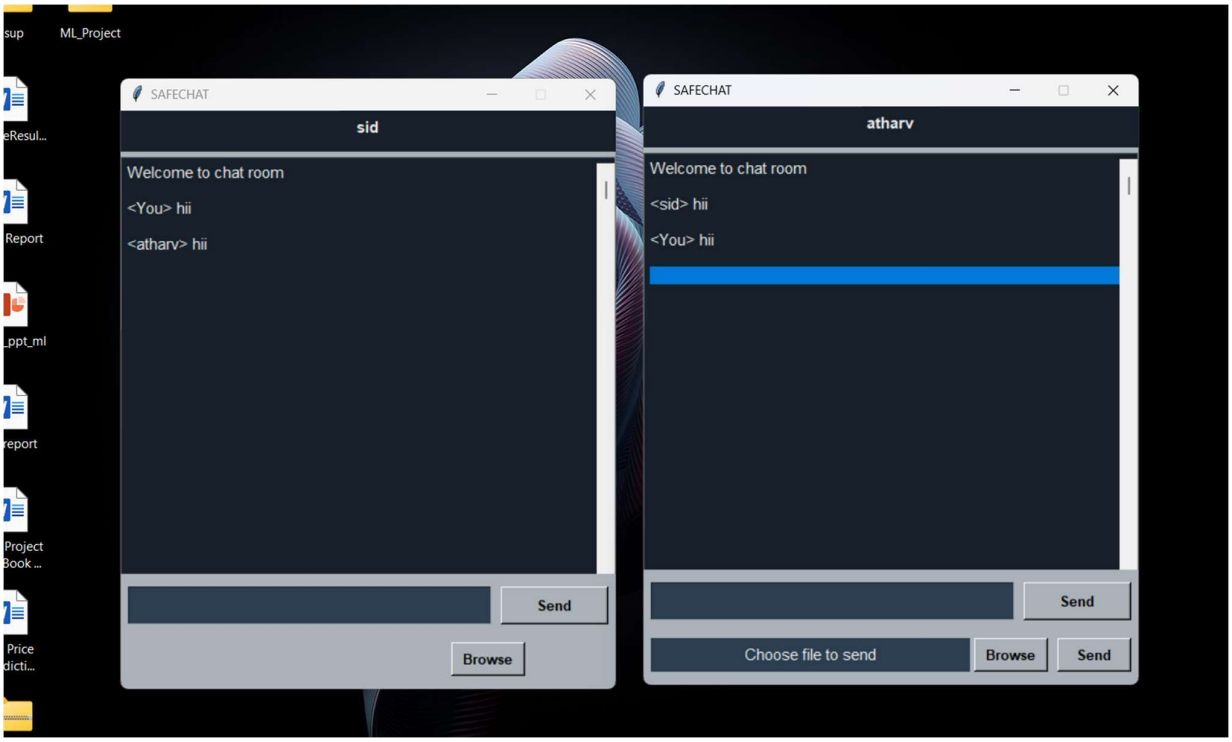


Fig. 6.2.3 – GUI

But if we enter any bullying message such as following then system will detect it using our trained model and will give bullying warning.

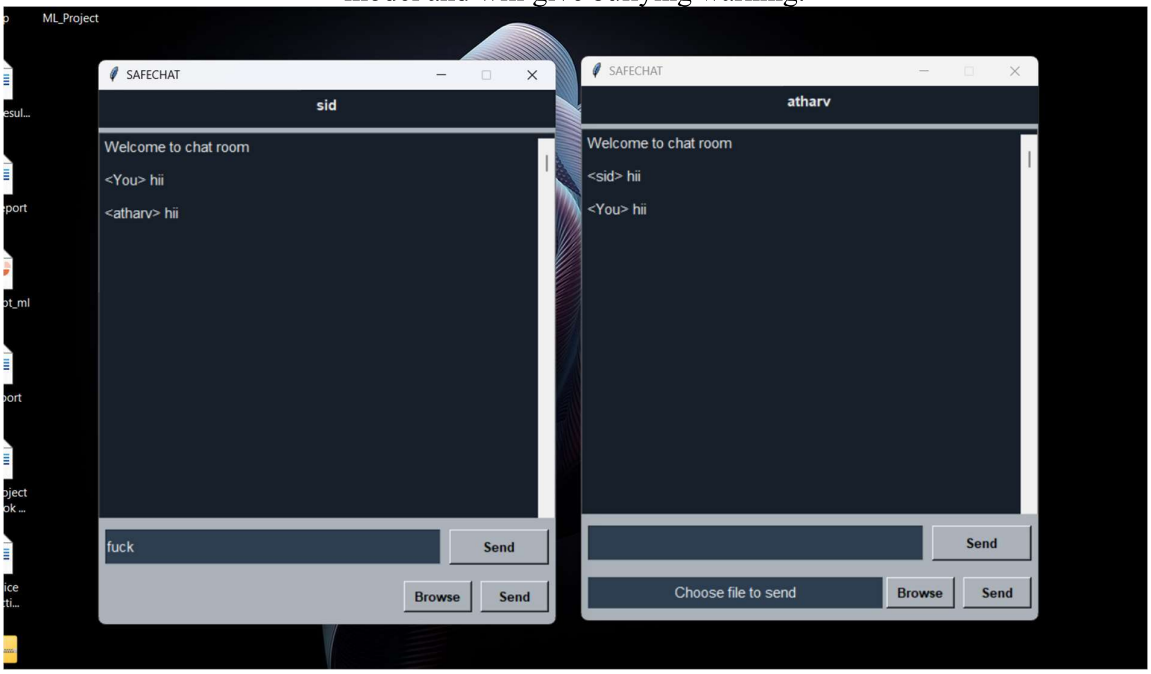


Fig. 6.2.4 – Harassing message

As you can see in following screenshot the system has detected the bullying message and has given warning to the user.

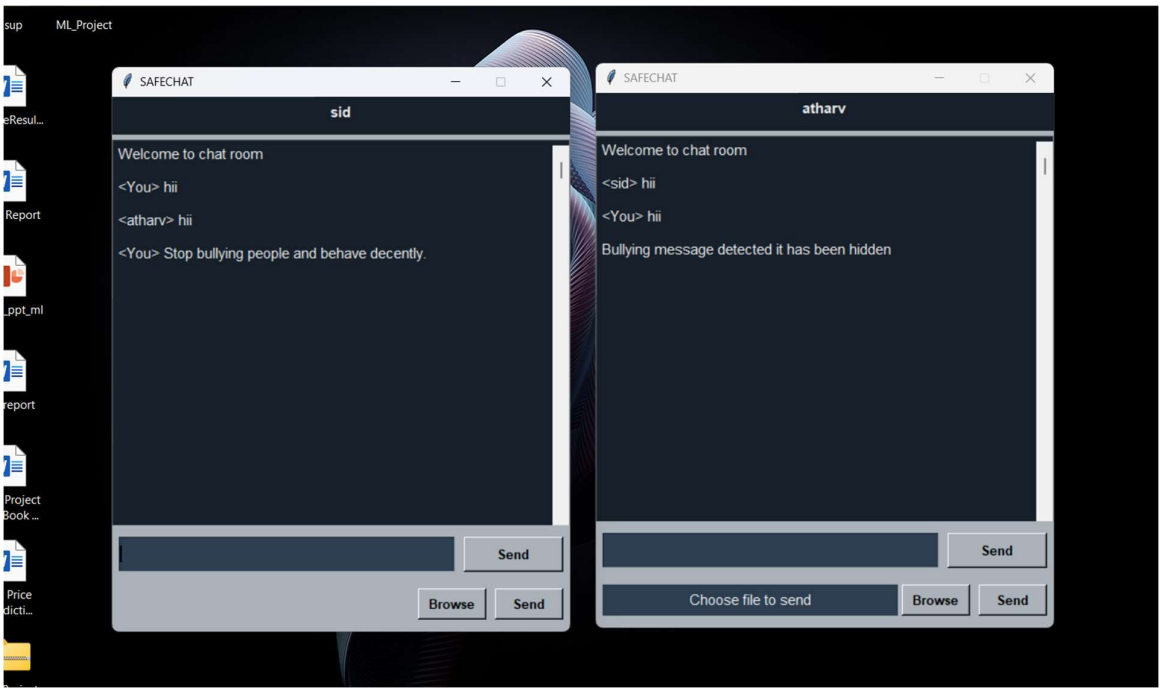


Fig. 6.2.5 – Cyber Bullying Detection

CHAPTER 7:

CONCLUSION

In conclusion, the cyberbullying detection project using machine learning and a graphical user interface (GUI) developed with Python tkinter and the socket module is an effective tool for detecting cyberbullying in social media messages. The project achieved its objectives by collecting and preprocessing a dataset of cyberbullying messages, extracting relevant features, training three machine learning algorithms, and evaluating their performance using several evaluation metrics.

The best-performing algorithm, determined through evaluation metrics, was selected as the final model for cyberbullying detection. The GUI developed using tkinter allowed clients to communicate with the server and receive warning messages if their messages were classified as cyberbullying.

Future work could involve improving the accuracy of the model, adding new features to the GUI, and expanding the system to work with other social media platforms.

CHAPTER 8 :
REFERENCES

[1] Bayzick, J., Edwards, A., & Edwards, L. (2023). Detecting the Presence of Cyberbullying Using Computer Software. ResearchGate. https://www.researchgate.net/publication/267964832_Detecting_the_Presence_of_Cyberbullying_Using_Computer_Software

[2] Perera, A., & Fernando, P. (2021). Accurate Cyberbullying Detection and Prevention on Social Media. Procedia Computer Science, 181, 605–611. https://doi.org/10.1016/j.procs.2021.01.207

[3] Huang, Q., Singh, V., & Atrey, P. K. (2014). Cyber Bullying Detection Using Social and Textual Analysis. In ACM Multimedia. https://doi.org/10.1145/2661126.2661133

[4] Dadvar, M., Trieschnigg, D., Ordelman, R., & De Jong, F. (2013). Improving Cyberbullying Detection with User Context. In Lecture Notes in Computer Science (pp. 693–696). Springer Science+Business Media. https://doi.org/10.1007/978-3-642-36973-5_62

[5] Pang, C. (2013). An Effective Approach for Cyberbullying Detection. ResearchGate. https://www.researchgate.net/publication/257029740_An_Effective_Approach_for_Cyberbullying_Detection

version 1

ORIGINALITY REPORT

20%	20%	7%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijsrd.com Internet Source	4%
2	pdfcoffee.com Internet Source	3%
3	www.coursehero.com Internet Source	3%
4	ir-library.mmarau.ac.ke:8080 Internet Source	2%
5	www.daernika.com Internet Source	2%
6	www.ijisae.org Internet Source	1%
7	mimoza.marmara.edu.tr Internet Source	1%
8	lrec2020.lrec-conf.org Internet Source	1%
9	oa.upm.es Internet Source	1%

10	polipapers.upv.es Internet Source	1 %
11	link.springer.com Internet Source	1 %

Exclude quotes	Off	Exclude matches	< 17 words
Exclude bibliography	Off		