VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Project Report on

Social Media Investigations against Cybercrime: CyberGuardian

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2023-24

Submitted by
Bhavesh Bhatia (D17A 07)
Chaitanya Limaye(D17B 37)
Chitra Atlani(D17C 04)
Siyona Singh(D17C 54)

Project Mentor
Mrs. Rupali Soni

(2023-24)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Certificate

This is to certify that ***Bhavesh Bhatia (D17A 07), Chaitanya Limaye(D17B 37), Chitra Atlani(D17C 04), Siyona Singh(D17C 54)*** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "***Social Media Investigations against Cybercrime: SocialSentinel***" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Mrs. Rupali Soni*** in the year 2023-24 .

This project report entitled ***Social Media Investigations against Cybercrime: SocialSentinel*** by ***Bhavesh Bhatia (D17A 07), Chaitanya Limaye(D17B 37), Chitra Atlani(D17C 04), Siyona Singh(D17C 54)*** is approved for the degree of ***B.E in Computer Engineering***.

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This thesis/dissertation/project report entitled **Social Media Investigations against Cybercrime: SocialSentinel** by **Bhavesh Bhatia (D17A 07), Chaitanya Limaye(D17B 37), Chitra Atlani(D17C 04), Siyona Singh(D17C 54)** is approved for the degree of ***B.E in Computer Engineering***.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:
Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Bhavesh Bhatia (D17A 07)

(Signature)

Chaitanya Limaye(D17B 37)

(Signature)

Chitra Atlani(D17C 04)

(Signature)

Siyona Singh(D17C 54)

Date: 12/04/2024

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

INDEX

Abstract

The main objective of our project is to develop a comprehensive framework that leverages social media intelligence and analysis techniques to proactively identify and mitigate cyber threats using the various phases of product design and development. By integrating social media investigations into the traditional cybersecurity protocols, this approach aims to enhance the overall security posture of digital products and services. In the past decade, cyberbullying has been an emerging phenomenon that has a socio-psychological impact on people of all age groups, especially adolescents. With the advancement of digital technology, youth is more attached to social media, resulting in elevated chances of cyberbullying. With the increasing usage of techno-savvy gadgets, social media applications are highly prevalent among the youth, which can be advantageous as well as disadvantageous. Social media allows sharing posts, photos, and messages publicly as well as privately among friends, while on the other hand, it involves an increase in cyberbullying by creating fake accounts on the apps.

Chapter 1

Chapter 1 : Introduction

1.1. Introduction to the project

Information and communication technologies (ICTs) have become an interactive context where leisure, entertainment, and learning take place, also within family life. ICTs bring many advantages, such as immediate access to a large amount of information or the ability to communicate instantly with anyone. Adolescents born in the digital age have grown up enjoying these advantages and communicate indistinctly in physical or virtual space. However, these technologies also carry risks and pose new challenges for both school education and family involvement and guidance. Cyberspace is a medium where adolescents interact, often away from family supervision and sometimes in an aggressive and immoral manner. In the last decade, both families and teachers have reported difficulties in teaching about the prevention of online risks, such as grooming, internet addiction, or cyberbullying [1] — the latter being one of the online risks that families are most concerned about due to its prevalence rates and harmful consequences cyberbullying has been defined as a particular form of aggression that occurs when an individual or group uses digital devices to harm a person intentionally and repeatedly, who finds it difficult to prevent this harassment from continuing.[2]

In the past, the cyberattacks were not as complicated as today's attacks and there were fewer computer-based systems to protect. However, the rapid advancement of technology has made cybercrimes easier and more sophisticated.[3] The increased usage of smartphones, IoT devices, social media platforms, cloud platforms, and crypto mining also escalated the effects of cybercrimes. As stated in several scientific research reports, cybercrimes cost a few trillion dollars to the world economy each year, and this cost is expected to increase every year.[4]

The National Center for Education Statistics (NCES) 2019 report indicated that 20.2% of students aged 12 to 18 experience bullying. Disturbingly, 15% of these incidents occurred online or via text, highlighting the prevalence of cyberbullying. Moreover, 41% of students who were bullied at school believed that the bullying would persist. These statistics shed light on the alarming reality of cyberbullying and its potential long term effects on victims. Research conducted by Armitage et al. [5] reveals that bullying has profound and lasting impacts on victims, beginning with mental health issues during childhood. Childhood bullying represents a significant public health concern as it heightens the risk of poor health, social challenges, and educational setbacks during formative years. These consequences extend into adulthood, manifesting as psychopathology, suicidal tendencies, and even involvement in criminal activities. Cyberbullying, the online counterpart of traditional bullying that has now expanded its reach into the digital landscape, emerged as a pervasive issue with its own set of challenges and consequences. The use of technology is becoming increasingly intertwined with our lives, and addressing the harmful impact of cyberbullying is

imperative in ensuring the well-being and safety of individuals in the digital realm. Cyberbullying has emerged as a growing concern, and it is defined as the intentional and repetitive use of a computer, mobile phone, or other electronic device to cause harm to someone. The term includes sending threats, posting or distributing libelous or harassing messages, and uploading or distributing hateful images or videos that harm others. Depending on the age of the group and the definition of cyberbullying, a range of 5 to 72 percent of youth experience cyberbullying. [6]Cyberbullying is a serious problem that affects teens and schools because of its psychological, emotional, behavioral, and physical consequences. Cyberbullying has legal implications that law enforcement officers must be aware of and understand. The growth of cell phones and Internet usage among teens has altered conduct and social norms. Cyberbullying is one of the most critical issues facing law enforcement today. According to reports and research, police across the country lack direction, resources, training, and support. Cyberbullying cases can be complex and time consuming to investigate, especially on social media, and many law enforcement agencies lack the necessary resources and expertise to handle them effectively [7]

The anticipated outcome of our project is twofold: First, it will significantly enhance the cybersecurity resilience of digital products and services by proactively identifying and mitigating potential cyber threats originating from social media channels. Second, it will contribute to the broader field of cybersecurity by promoting a multidisciplinary approach that bridges social media intelligence with product design and development practices.

Ultimately, this integrated approach is envisioned to empower organizations and developers to stay ahead of cybercriminals, safeguarding their products, and enhancing the overall security of digital ecosystems.

1.2. Motivation for the project

Although the computer today is a great convenience for many of us and brings many advantages, it also has disadvantages that people need to be aware of. Many computer users are utilizing the computer for the erroneous purposes either for their personal benefits or for others benefit. This gave birth to “Cyber Crime”.[8]

Cybercrimes are occurring more frequently than ever before. With the widespread availability of internet access and the growing number of connected devices, cybercriminals have a larger attack surface to target individuals, businesses, and governments. Cybercriminals engage in a wide variety of activities, including hacking, data breaches, identity theft, phishing, ransomware attacks, online harassment, cyberbullying, and more. These crimes can lead to financial losses, privacy violations, and emotional distress for victims. Also, cybercriminals are continuously developing more sophisticated and innovative attack methods. They often leverage advanced tools and techniques,

making it challenging for individuals and organizations to defend against these threats. Victims of cyberbullying, harassment, or online threats may experience emotional distress and psychological harm. The emotional toll of cybercrimes should not be underestimated, as it can have long-lasting effects on mental health and well-being.

Cybercrime has become a lucrative industry for criminal organizations. Activities such as ransomware attacks generate substantial profits, motivating cybercriminals to continue their illicit operations. Cybercrime tools, including malware kits and hacking services, are readily available on the dark web. As a result, even individuals with limited technical expertise can engage in cybercriminal activities.

Due to above mentioned factors, we observe that there are widespread effects of cyberbullying on the social, economic and even psychological aspects of the society. Hence, it is the need of the hour to develop a system which can detect such online activities and help in their investigations.

1.3. Problem Definition

We aim to develop a tool:

1. To collect and analyze data from various social media platforms language independent to aid in investigations related to cyberbullying & online harassment.
2. For detecting, investigating, and mitigating above cybercrime issues require efficient and comprehensive tools that can collect and analyze data from various social media platforms.
3. Which will efficiently collect, process, and analyze relevant data from diverse social media platforms, aiding investigators in uncovering instances of harmful behavior.

1.4. Lacuna of Existing Systems

Existing systems have following shortcomings or improvements needed:

1. The BERT model [7] used in existing systems can achieve more accurate results if provided with a large dataset.
2. The system can achieve better results in the cyberbullying detection process by considering all the features.
3. A combination of other models on top of the BERT model could have been used to create a state-of-the-art model for the specific NLP tasks in detecting cyberbullying.
4. The quality of datasets and the clarity in reporting classes are lower.
5. The capacity for fine-grained detection needs to be improved.
6. Increasing the reliability and reproducibility of models is also needed.

1.5. Relevance of the Project

With the increasing reliance on digital technologies and the internet, cybercrime has become more prevalent and sophisticated. Criminals often use social media platforms to carry out various forms of cybercrime, such as identity theft, phishing, cyberbullying, and fraud.

Privacy concerns related to social media are a significant issue. Cybercriminals can exploit vulnerabilities in social media security settings to access personal information, leading to privacy breaches and potential identity theft. Law enforcement agencies and cybersecurity experts need to stay ahead of cybercriminals by monitoring and investigating their activities on social media. This involves tracking their online presence, analyzing digital footprints, and identifying potential threats.

Social media investigations can provide valuable threat intelligence to cybersecurity professionals and organizations. Analyzing the tactics, techniques, and procedures used by cybercriminals on social media can help in developing proactive security measures. Cyberbullying and online harassment are growing problems on social media. Investigating and addressing such incidents is crucial for ensuring the safety and mental well-being of users, especially young people.

Hence, social media investigations against cybercrime are relevant in today's era due to the increasing prevalence of cyber threats on social platforms, the need to protect user privacy, the importance of gathering digital evidence, and the ongoing efforts to combat cybercrime at both the individual and international levels.

Chapter 2

Chapter 2: Literature Survey

A. Brief Overview of Literature Survey

Systems do exist in this domain, but they have some or the other setback, like they are either suitable for very specific conditions, or give results that are less accurate. Following is the description of some of the various resources and what they lack.

B. Related Works

In [7], a semi-supervised method for identifying cyberbullying by leveraging five distinct features(Sentimental Features, Sarcastic Features, Syntactic Features, Semantic Features and Social Features)that characterize cyberbullying content, employing the BERT model. Focusing solely on sentiment features, their BERT model achieved an accuracy of 91.90% after dual-cycle training, surpassing conventional machine learning models. The potential for the BERT model to deliver even higher accuracy hinges on access to extensive datasets. Integrating all proposed features outlined in their study could promise enhanced cyberbullying detection capabilities. An application harnessing these features could effectively identify and flag bullying content for further action. Additionally, combining complementary models with BERT can hold promise for developing a cutting-edge solution tailored to the nuances of cyberbullying detection in natural language processing tasks.

In [8], the goal is to develop an effective method for identifying and addressing online abusive content by combining NLP and ML to create a model that detects offensive language in English and Hinglish. They found that CV slightly outperforms TFIDF in accuracy and that Linear SVC and SGD offer better results in classifying bullying messages in Hinglish, with faster training times. However, deeper analysis in sentiment, semantics, and syntax could improve accuracy further. Integrating the model with various social media platforms can aid in reducing cyberbullying. The main challenge is obtaining large, accurately labeled datasets in Hinglish for ML training, as existing datasets are limited in size and reliability.

The work in [9] introduces a model designed to automatically detect cyberbullying content across multiple languages, addressing a crucial need in managing social media content and safeguarding users from the harmful effects of toxic remarks such as verbal attacks and offensive language. Their study evaluates the performance of various neural network models, with the CNN-BiLSTM network demonstrating the highest accuracy. Unlike the CNN model, which focuses solely on local word n-grams, the CNN-BiLSTM model can capture both local characteristics and global features, including long-term dependencies. The paper can explore the integration of image and video analysis to determine if cyberbullying detection can be automated across multimedia content. This is part of their future work.

In [10], current approaches to cyberbullying detection, specifically focusing on session-based detection in social media was explored. Social Media Session-based Cyberbullying Detection (SSCD) framework and review research progress within this framework, including model and dataset creation related to social media sessions was introduced. Their comparative experiments assess state-of-the-art models on SSCD datasets and suggest avenues for future research. They highlight the need to consider key cyberbullying characteristics like repetition and power imbalances in model design and dataset creation, emphasizing the potential of fine-grained detection methods. During the detection of cyberbullying in social media sessions, they encountered significant challenges like enhancing transparency and quality in dataset management, advancing fine-grained detection capabilities, and ensuring model reliability and reproducibility that need addressing.

In [11], a machine learning-based approach for cyberbullying detection and conducted evaluations using two classifiers, SVM and Neural Network, along with TFIDF and sentiment analysis algorithms for feature extraction was introduced. Their model achieved 92.8% accuracy with Neural Network using 3-grams and 90.3% accuracy with SVM using 4-grams when combining TFIDF and sentiment analysis. Notably, the Neural Network demonstrated superior performance with an average f-score of 91.9%, compared to SVM's average f-score of 89.8%. Additionally, their model surpassed related work in accuracy and f-score metrics. Despite these advancements, detecting cyberbullying patterns is constrained by training data size. Larger datasets can be used which are essential for improved performance, making deep learning techniques more suitable due to their proven superiority over machine learning methods with larger datasets.

[19] addresses the oversight of sentence semantics in existing academic methods by utilizing word2vec to train customized word embeddings. They develop an LSTM-CNN architecture tailored for cyberbullying detection, surpassing traditional approaches on Twitter data. Their model, with a 97% ROC AUC score, includes a web application for toxicity-based tweet classification and a Telegram Bot for cyberbullying prevention. They also create Chrome Extensions for NSFW content moderation on WhatsApp Web. While their solution is effective, future work includes transitioning to Attention-based Transformers, expanding platform compatibility, and incorporating multimedia analysis and language support.

[20] focuses on cyberbullying detection in Roman Urdu, facing challenges due to its linguistic limitations and diverse structures. They employ advanced preprocessing techniques and a deep learning architecture tailored for Roman Urdu cyberbullying detection. Experimentation led to RNN-LSTM and RNN-BiLSTM models outperforming CNN after 20 epochs. Future directions include developing ensemble models for improved detection of harassment and hate speech, incorporating context-specific features, and addressing morphological variations for enhanced outcomes.

2.1. Research Papers

a. Abstract of the research paper

In [7], a model based on various features that should be considered while detecting cyberbullying and implement a few features with the help of a bidirectional deep learning model called BERT.

The focus of [8] is to design and develop an effective technique to detect online abusive and bullying messages by merging Natural language processing and machine learning to develop a model that can detect offensive or hateful words in English and Hinglish language.

[9] proposed a deep learning framework that will evaluate real-time twitter tweets or social media posts as well as correctly identify any cyberbullying content in them. Recent studies has shown that deep neural network-based approaches are more effective than conventional techniques at detecting cyberbullying texts. Additionally, their application can recognise cyberbullying posts which were written in English, Hindi, and Hinglish (Multilingual data).

[6] proposes an open-source intelligence pipeline using data from Twitter to track keywords relevant to cyberbullying in social media to build dashboards for law enforcement agents. They discuss the prevalence of cyberbullying on social media, factors that compel individuals to indulge in cyberbullying, and the legal implications of cyberbullying in different countries also highlight the lack of direction, resources, training, and support that law enforcement officers face in investigating cyberbullying cases. The proposed interventions for cyberbullying involve collective efforts from various stakeholders, including parents, law enforcement, social media platforms, educational institutions, educators, and researchers. Their research provides a framework for cyberbullying and provides a comprehensive view of the digital landscape for investigators to track and identify cyberbullies, their tactics, and patterns.

[10] defines a framework that encapsulates four different steps session-based cyberbullying detection should go through, and discusses the multiple challenges that differ from single text-based cyberbullying detection. Based on this framework, theory provides a comprehensive overview of session-based cyberbullying detection in social media, delving into existing efforts from a data and methodological perspective. Our review leads us to proposing evidence-based criteria for a set of best practices to create session-based cyberbullying datasets. In addition, they perform benchmark experiments comparing the performance of state-of-the-art session-based cyberbullying detection models as well as large pre-trained language models across two different datasets. Through their review, we also put forth a set of open challenges as future research directions.

[11] proposes a supervised machine learning approach for detecting and preventing cyberbullying. Several classifiers are used to train and recognize bullying actions.The evaluation of the proposed approach on cyberbullying dataset shows that Neural Network performs better and achieve accuracy

of 92.8% and SVM achieves 90.3. Also, NN outperforms other classifiers of similar work on the same dataset.

b. Inference drawn from the paper

[7] proposed a semi-supervised approach in detecting cyberbullying based on the five features that can be used to define a cyberbullying post or message using the BERT model. While considering just one of the features, which was sentimental features, the BERT model achieved 91.90% accuracy when trained over dual cycles which outperformed the traditional machine learning models. The BERT model can achieve more accurate results if provided with a large dataset. To achieve even better results in the cyberbullying detection process, all the features can be considered.

With the increasing use of social media by teenagers, the need for automated cyberbullying detection is evident. In [8], Count Vectorizer (CV) outperformed TF-IDF in accuracy. Among various machine learning models, Linear SVC and SGD excelled in classifying bullying messages in Hinglish with faster processing. Further improvements can be made through sentiment analysis and integration with deep learning models. This integrated model could help combat cyberbullying on social media platforms and raise awareness.

The model for automatically detecting cyberbullying text on multilingual data is addressed and proposed in [9]. Solving this issue is critical for controlling social media material in multiple languages and protecting users from the negative impacts of toxic comments like verbal assaults and offensive language. The performance of their various models of neural networks is examined. The CNN-BiLSTM network has the best accuracy. While the CNN alone can only train local characteristics from word n-grams, with its LSTM layer, the CNN-BiLSTM can also learn global features and long-term dependencies. Future research will look at both picture and video elements to see if cyberbullying can be detected automatically.

[4] presents an open-source intelligence dashboard for tracking and analyzing cyberbullying incidents using Twitter data and Splunk software. It also highlights legal issues and the need for better resources and training for law enforcement.

The open-source intelligence pipeline empowers stakeholders to monitor and address cyberbullying on social media. This approach shows the potential of social media data for investigations. The dashboard provides a powerful tool for data collection and analysis, aiding organizations in addressing cyberbullying.Given the continued growth of social media and the internet, it's vital to explore new methods to combat cyberbullying and enhance online safety. This open-source intelligence dashboard encourages further research in this area.

In [10], they focus on cyberbullying detection, particularly session-based detection, introducing the Social Media Session-based Cyberbullying Detection (SSCD) framework. They review research on SSCD's four key components, model and dataset creation, and conduct benchmark experiments.

They emphasize the need to address repetition and power imbalances as inherent characteristics of cyberbullying during model and dataset design. Fine-grained detection research is still developing, offering insights into cyberbullying's nature and advancing the field. Their survey serves as a valuable reference for those exploring this emerging research trend.

[11] achieved 92.8% accuracy using NeuralNetwork with 3-grams and 90.3% accuracy using SVM with 4-grams while using both TF IDF and sentiment analysis together.They found that our Neural Network performed better than the SVM classifier as it also achieves average f-score 91.9% while the SVM achieves average f-score 89.8%. Furthermore, they compared our work with another related work that used the same dataset, finding that our Neural Network outperformed their classifiers in terms of accuracy and f-score. However, detecting cyberbullying patterns is limited by the sizeof training data. Thus, a larger amount of cyberbullying data is needed to improve the performance. Hence, deep learning techniques will be suitable in the larger data as they are proven to outperform machine learning approaches over larger size data.

2.2 Comparison with the existing system

Research Paper	Authors	D. O. P	Advantages	Disadvantages
An Application to Detect Cyberbullying Using Machine Learning and Deep Learning Techniques	Mitushi Raj, Samridhi Singh, Kanishka Solanki, and Ramani Selvanambi	26 July 2022	1. Recent studies has shown that deep neural network-based approaches are more effective than conventional techniques at detecting cyberbullying texts. 2. A deep learning framework will evaluate real-time twitter tweets or social media posts and correctly identify any cyberbullying content in them. 3. The application can recognise considers English, Hindi, and Hinglish (Multilingual data). 4. The CNN-BiLSTM network has the best accuracy.	1. With its LSTM layer, CNN-BiLSTM can also learn global features and long-term dependencies. 2. Future research will look at both picture and video elements to see if cyberbullying can be detected automatically.

An Application to Detect Cyberbullying Using Machine Learning and Deep Learning Techniques	Mitushi Raj, Samridhi Singh, Kanishka Solanki, and Ramani Selvanambi	26 July 2022	1. Recent studies has shown that deep neural network-based approaches are more effective than conventional techniques at detecting cyberbullying texts. 2. A deep learning framework will evaluate real-time twitter tweets or social media posts and correctly identify any cyberbullying content in them. 3. The application can recognise considers English, Hindi, and Hinglish (Multilingual data). 4. The CNN-BiLSTM network has the best accuracy.	1. With its LSTM layer, CNN-BiLSTM can also learn global features and long-term dependencies. 2. Future research will look at both picture and video elements to see if cyberbullying can be detected automatically.
Cyber Bullying Detection on Social Media using Machine Learning	Aditya Desai, Shashank Kalaskar, Omkar Kumbhar, and Rashmi Dhumal	9 August 2021	1. A semi-supervised approach in detecting cyberbullying based on five features that can be used to define a cyberbullying post or message using the BERT model. 2. Only considering sentimental features the model achieved 91.90% accuracy when trained over dual cycles which outperformed the traditional machine learning models.	1. The BERT model can achieve more accurate results if provided with a large dataset. 2. Achieve better results in the cyberbullying detection process by considering all the features. 3. A combination of other models on top of the BERT model can also be used in the future to create a state-of-the-art model for the specific NLP tasks in detecting cyberbullying.

Table 2.2.1 Comparison of various models in existing systems

Chapter 3

Chapter 3: Requirement Gathering for the Proposed System

3.1 Introduction to requirement gathering

The most important part in requirement gathering is data collection. There is a need to extract data from social media applications so that our dataset has a variety of the message content. The Dataset has various types of messages. The model has to be trained using many epochs using suitable DL & ML models. So GPU is preferred. The data collection, cyberbullying detection & countermeasures should be as per the standards of the law. Reporting tools gathering is required to pass the case to concerned authorities.

3.2 Functional Requirements

1. User Authentication and Authorization: The system should have secure user authentication and authorization mechanisms to ensure that only authorized personnel can access and use the system's features.
2. Data Collection: The ability to collect and store social media data from various platforms, including text, images, and videos.
3. Monitoring and Analysis: Real-time monitoring of social media platforms and the capability to analyze data for potential cybercrime indicators, such as threats, harassment, or fraud.
4. Alerting and Reporting: The system should generate alerts and reports when suspicious or criminal activity is detected, with customizable alert thresholds.
5. Data Search and Retrieval: The ability to search and retrieve historical data for investigative purposes, including data from different timeframes and sources.
6. Integration with External Tools: Integration with external tools and databases for cross-referencing and enhancing the investigative process.
7. User Collaboration: Support for multiple users to collaborate on investigations, share findings, and communicate securely within the system.
8. Data Visualization: Tools for visualizing data, trends, and relationships to aid investigators in understanding the context of cybercrimes.
9. Case Management: The system should provide a case management feature to organize and track investigations, including assigning tasks and monitoring progress.
10. Compliance: Ensure compliance with legal and ethical standards for data collection and privacy, adhering to relevant laws and regulations.

3.3 Non-Functional Requirements

1. Scalability: The system should be scalable to accommodate a growing volume of data and users.
2. Performance: Real-time processing and analysis of social media data with low latency to provide quick responses to potential threats.
3. Security: Strong encryption, secure storage, and access controls to protect sensitive data and maintain the system's integrity.
4. Reliability: The system should be available and reliable 24/7 to ensure continuous monitoring and investigation.
5. Usability: The user interface should be intuitive, and training should be minimal for users to become proficient in using the system.
6. Interoperability: The ability to interact with various social media platforms and other investigative tools.
7. Data Retention and Archiving: A system for long-term data retention and archiving to meet legal requirements and historical analysis.

3.4.Hardware, Software , Technology and tools utilized

Hardware:

Processor: Intel Core i5 or equivalent

Disk Space: 100 GB to 200 GB

RAM: 8 GB or more

Software:

Frontend: HTML, CSS, JavaScript, and a frontend framework like React or Vue.js.

Backend: Python with a web framework like Flask.

Database: MySQL or other suitable databases.

Data Analysis: Python libraries like NLTK, spaCy, scikit-learn for NLP and machine learning tasks.

Technology:

1. Keyword and Sentiment Analysis:

Sentiment analysis is performed using natural language processing (NLP) algorithms, and abusive content is detected based on specific keywords and phrases.

2. Machine Learning Models:

Machine learning models have been developed to classify and detect cyberbullying and harassment, leveraging historical data. Supervised learning is used to create models that can automatically recognize abusive language and behavior, with continuous training for adaptation.

3. User Behavior Analysis:

Patterns of behavior and interactions between users are analyzed to identify potential harassers or targets. Suspicious activities, such as excessive reporting or blocking, are flagged based on changes in user behavior.

Tools:

1. Python nltk libraries
2. ML & DL Algorithms like CNN & BERT

3.5 Constraints

1. Budget Constraints: The project may have budget limitations that impact the development and deployment of the system.
2. Legal and Ethical Constraints: Compliance with legal and ethical standards, including privacy laws and regulations, may impose constraints on data collection and usage.
3. Data Availability: The system's effectiveness depends on the availability of public and relevant social media data. Some platforms may restrict access to their data.
4. Technological Constraints: The choice of technology stack and infrastructure can impose constraints, such as compatibility with existing systems and hardware limitations.
5. Geographical Constraints: Data sovereignty and jurisdictional issues may affect data collection and storage across different regions.
6. User Skill Set: The skillset of the investigative team and users may influence the complexity and features of the system.

Chapter 4

Chapter 4: Proposed Design

4.1 Block diagram of the system

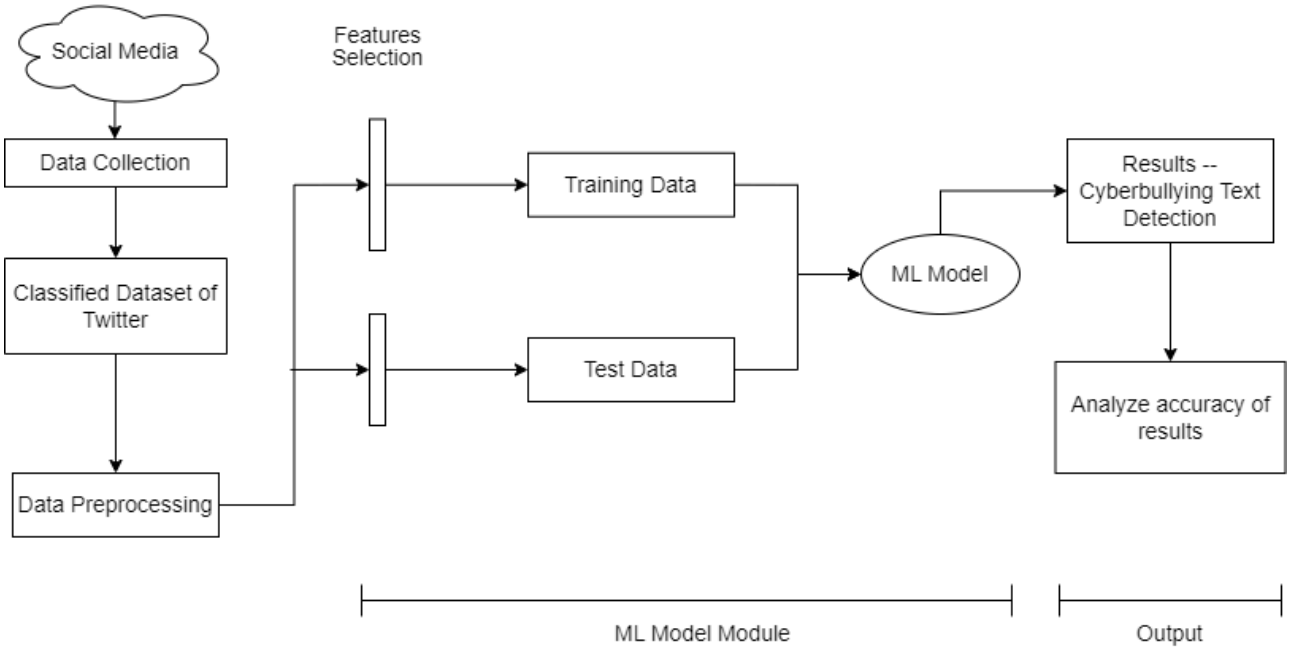


Fig. 4.1.1 Block Diagram

The above block diagram gives the flow of the project starting from data collection, training the model after feature engineering & analyzing the accuracy after achieving results.

4.2 Modular design of the system

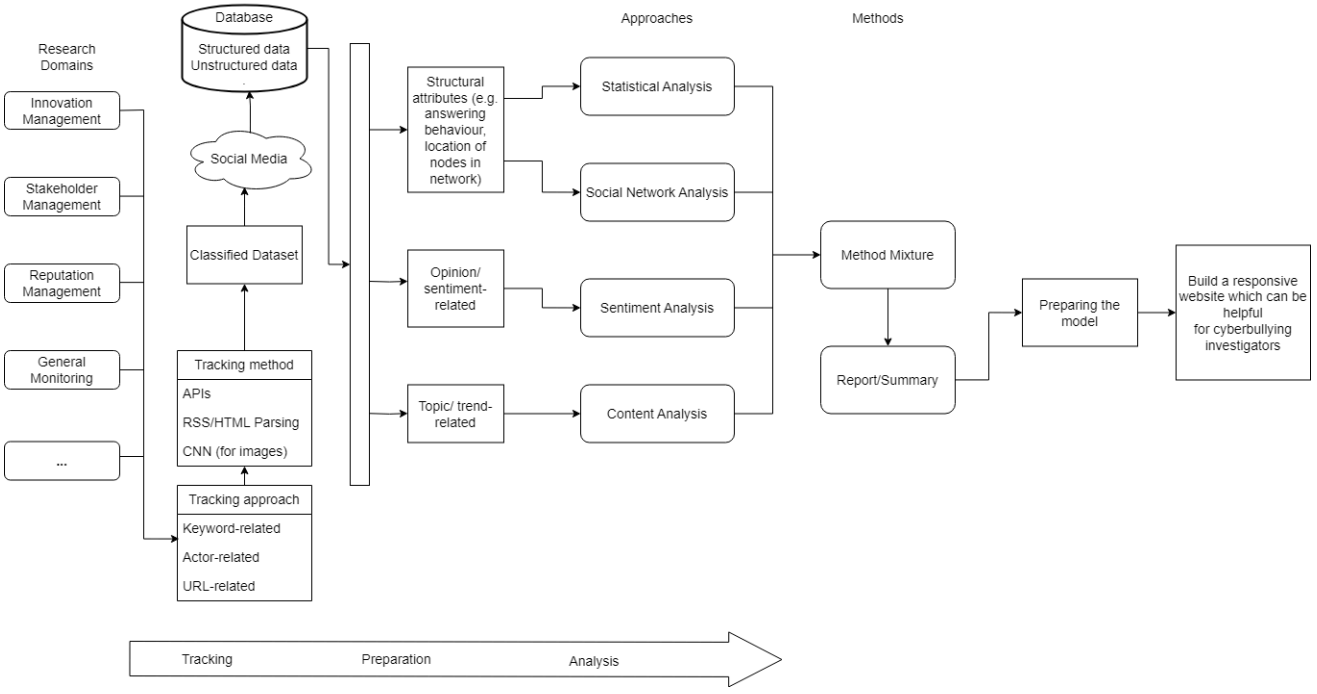


Fig. 4.2.1 Modular Diagram

The above modular diagram gives description about block diagram i.e data collection from various sources, making the data structured, applying the trained models & different types of analysis.

4.3 Detailed Design

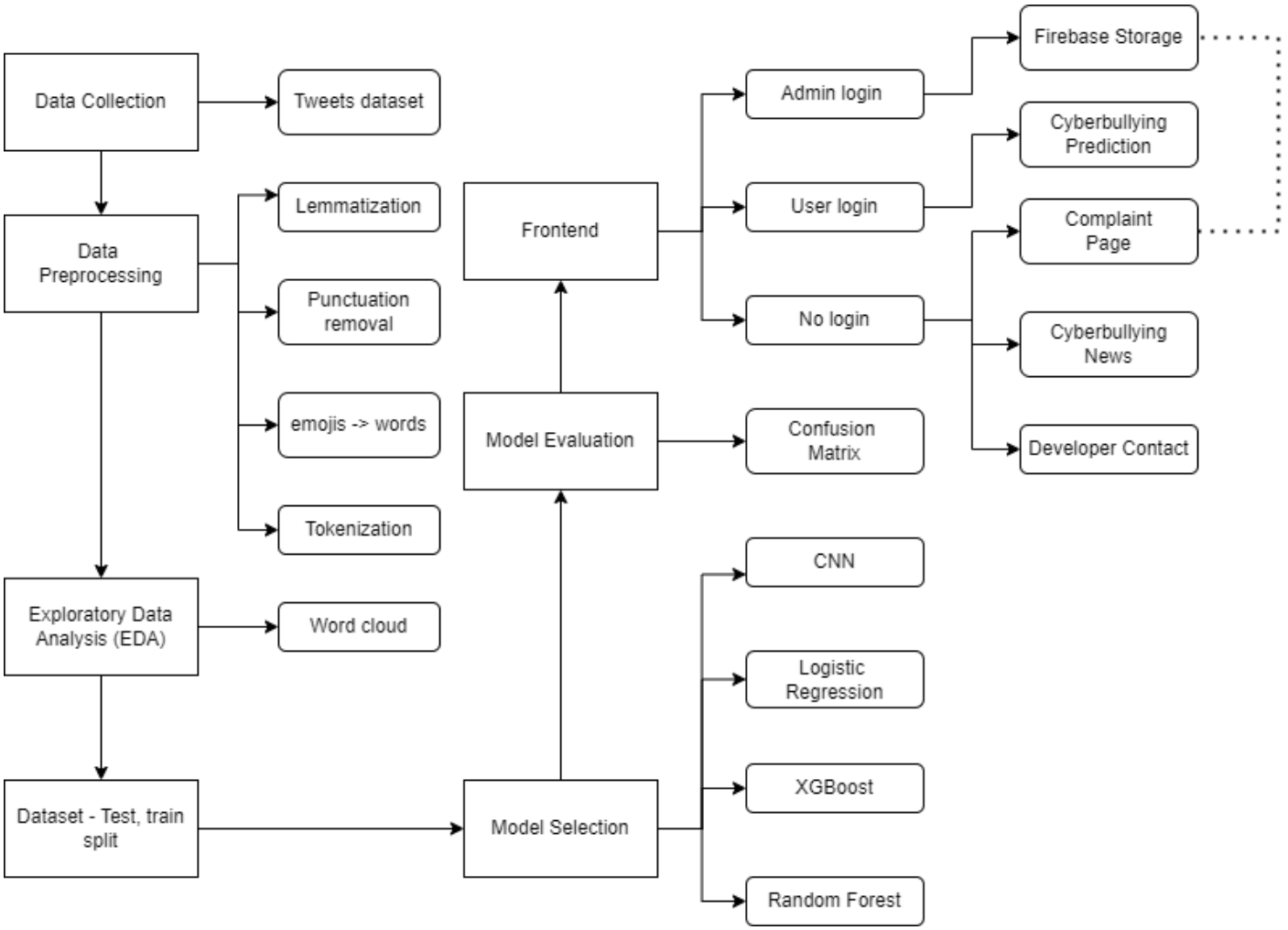


Fig. 4.3.1 Detailed Design

The above detailed design describes the whole architecture and structure of our project. Beginning from data collection, to data preprocessing, EDA, further to model selection and evaluation, we reach to the representation of our system using Frontend, which has three levels of access- one is without login where user can access latest cyberbullying news and contact to our developers, but most importantly, user can lodge a complaint, which is stored in Firebase. Once a user logs in, only then can they check for cyberbullying in the text they wish to enter. The admin has control of everything, including everything the user can access along with the Firebase backend which can only be accessed by the admin.

Chapter 5

Chapter 5: Implementation of the Proposed System

5.1. Methodology employed for development

1. Multi-platform Data Collection:
    - Data collection involves gathering text data from various sources where cyberbullying posts and tweets can be found, such as social media platforms, forums, or online communities.
    - You may need to use web scraping techniques, APIs, or crowdsourcing to collect a diverse set of data from different platforms.
2. Exploratory Data Analysis (EDA):
    - EDA is the process of understanding and summarizing the dataset to gain insights into its structure and characteristics.
    - Tasks in EDA include:

      Examining the dataset's size, distribution, and basic statistics.

      Visualizing the distribution of cyberbullying and non-cyberbullying posts.
    - Analyzing text length, word frequency, and other relevant attributes.
3. Data Pre-Processing & Enrichment:
    - Data preprocessing is essential to clean and prepare the text data for modeling.
    - Steps include:
        - Text Cleaning:
        - Removing HTML tags, URLs, and special characters.
        - Converting text to lowercase for consistency.
        - Removing punctuation marks.
        - Replacing emojis with their textual descriptions (e.g., 😃 to "smile").
        - Tokenization
        - Splitting text into individual words or tokens.
        - Stop Words Removal
        - Eliminating common stopwords (e.g., "the," "and," "is") to reduce noise.
4. Use of NLP Algorithms - Lemmatization:
    - Lemmatization is the process of reducing words to their base or root forms to standardize the text. For example, "running" becomes "run" after lemmatization.
    - Lemmatization helps reduce the dimensionality of the data and improve feature quality.

5. Splitting the Data into Test and Train:
    - Divide the pre-processed data into two subsets: a training set and a test set
    - Common splits are 70-30 or 80-20, where a majority of the data is used for training and the rest for testing the model.
6. Training the Model:
    - Convolutional Neural Network (CNN)
    - CNNs are typically used for image data, but they can also be adapted for text classification by using techniques like 1D convolutions
    - We have designed a neural network architecture, including convolutional layers, pooling layers, and dense layers, to learn patterns in text data
    - Logistic Regression
        - Logistic regression is not commonly used for text classification. It's typically used for regression problems where the target variable is continuous. You may want to consider using logistic regression for binary classification tasks like cyberbullying detection
    - XGBoost (Extreme Gradient Boosting)
        - XGBoost is a popular gradient boosting algorithm suitable for text classification.
7. Evaluation:
    - After training the models, we evaluate their performance using appropriate metrics like accuracy, precision, recall, F1-score, and ROC AUC.

8. Fine-Tuning and Model Selection:
    - Based on the evaluation results, you may need to fine-tune hyperparameters or try different model architectures to improve performance.
    - Ultimately, select the best-performing model for your cyberbullying detection task.

9. Deployment and Monitoring:
    - Once you have a model with satisfactory performance, you can deploy it to identify and respond to cyberbullying content in real-time on the platforms where you collected the data.
    - Monitor the model's performance over time and periodically retrain it with new data to maintain its accuracy.

5.2 Algorithms for the respective modules developed

1. Logistic Regression

1. Extract the suitable dataset
2. Clean the dataset by removing null values, duplicate rows
3. Preprocess the dataset by removing emojis, punctuation marks,stop words etc
4. Use tf idf(Term Frequency Inverse Document Frequency) for feature representation & handling sparse data.
5. Train the logistic regression model using the vector component & the prediction for that component.
6. Obtain all the confusion matrix parameters such as accuracy, precision, recall, support & f1-score.

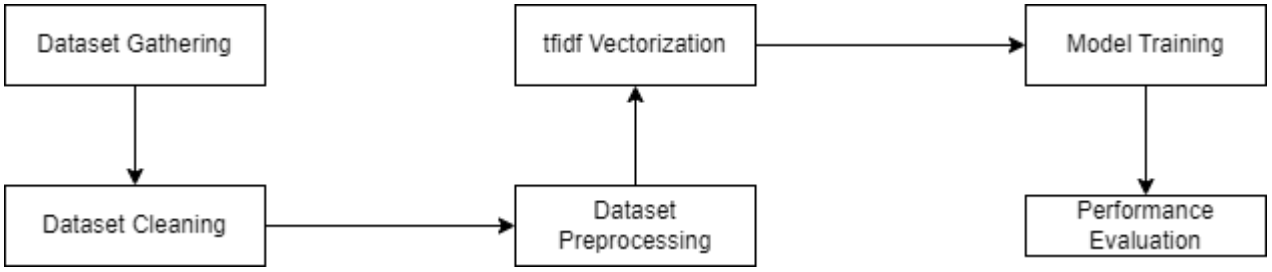


Fig. 5.2.1 Logistic regression

2. XGBoost:

Extract the suitable dataset

1. Clean the dataset by removing null values, duplicate rows
2. Dataset Preprocessing
3. Use tf idf(Term Frequency Inverse Document Frequency) for feature representation & handling sparse data.
4. Train the XGBoost model.
5. Obtain all the confusion matrix parameters such as accuracy, precision, recall, support & f1-score.

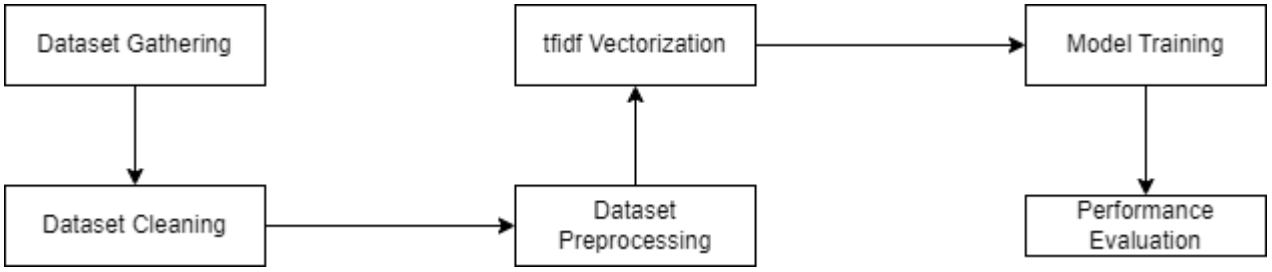


Fig. 5.2.2 XGBoost

3.Linear SVC:

1. Extract the suitable dataset
2. Clean the dataset by removing null values, duplicate rows
3. Get sentiment of every message content.
4. Preprocess the dataset by removing emojis, punctuation marks,stop words etc
5. Use tf idf (Term Frequency Inverse Document Frequency) for feature representation & handling sparse data.
6. Train the Linear SVM model using the vectorized component & the prediction for that component.
7. Obtain all the confusion matrix parameters such as accuracy, precision, recall, support & f1-score.

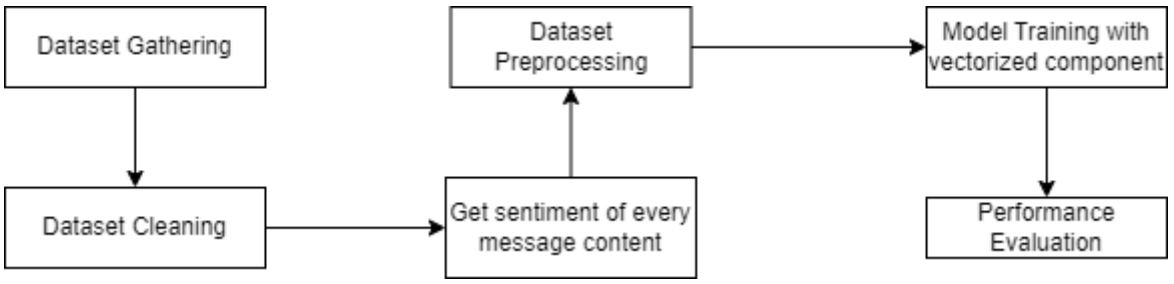


Fig. 5.2.3 Linear SVC

4.CNN:

1. All NLP steps performed
2. Counter: Using Counter for counting hashable objects and obtaining insights into word frequency.
3. Splitting Data into Test and Train: Performing a train-test split to estimate model performance.
4. Ordinal Encoder: Encoding categorical features as integer arrays.
5. Training the Model: Implementing CNNs to process textual data through multiple layers.

About CNN Deep Learning Layers

6. Input Layer: Here, we give input to our model. The number of neurons in this layer is equal to the total number of features in our data. Trained with the help of Keras.
7. Embedding: A technique used to represent categorical data, often used for converting words or tokens into dense vectors.
8. Convolutional Layer (Conv1D): A layer that applies convolutional operation on 1D inputs, often used for feature extraction in sequential data.
9. Max Pooling Layer (MaxPooling1D): A layer that performs down-sampling by taking the maximum value over a fixed-size window, reducing the dimensionality of the input.
10. Flatten Layer: A layer that flattens the input into a one-dimensional tensor, often used to connect convolutional layers to fully connected layers.
11. Dense Layer: A fully connected layer where each neuron is connected to every neuron in the previous layer, responsible for learning non-linear transformations in data.

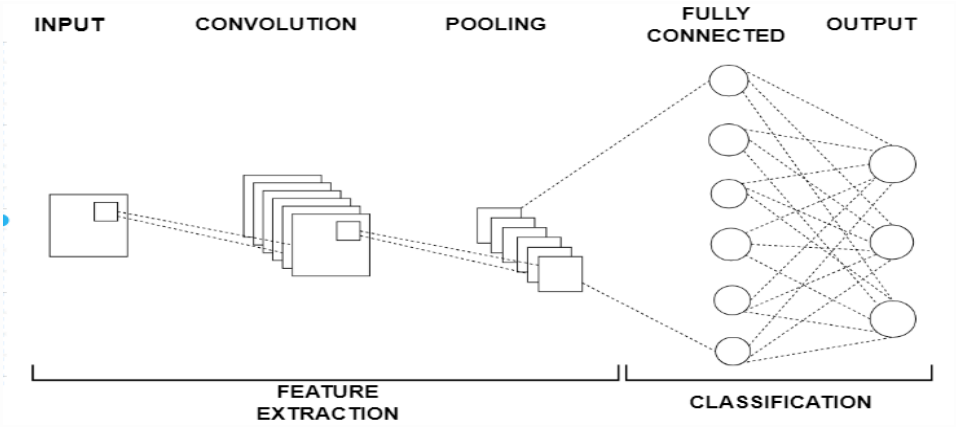


Fig. 5.2.4 CNN

5.3 Datasets source and utilization

In the era of widespread social media use heightened by the Covid-19 crisis, instances of cyberbullying have surged to unprecedented levels. To address this concerning trend, it is imperative to develop models that can automatically detect and flag potentially harmful tweets. By analyzing the patterns of hate speech and negative behavior, we can take proactive steps to mitigate the impact of cyberbullying.

This dataset is particularly significant as social media has become an integral part of daily communication across all demographics. Its omnipresence means that cyberbullying can affect individuals of all ages and backgrounds, with the anonymity afforded by the online environment making it challenging to combat compared to traditional forms of bullying.

UNICEF, on April 15th, 2020, issued a cautionary statement highlighting the heightened risk of cyberbullying during the Covid-19 pandemic. Factors such as widespread school closures, increased screen time, and reduced face-to-face social interactions have contributed to this risk. Disturbingly, statistics reveal that a significant percentage of middle and high school students have either experienced or witnessed cyberbullying, leading to adverse effects like academic decline, emotional distress, and even suicidal ideation.

This dataset comprises over 47,000 tweets categorized based on various forms of cyberbullying, including age-related, ethnicity-based, gender-specific, religious, and other forms of cyber harassment. It has been meticulously balanced to include approximately 8,000 instances of each cyberbullying category.

It's essential to note that the content of these tweets may contain descriptions of bullying incidents or offensive content. Therefore, users should exercise caution and engage with the dataset to a level where they feel comfortable, keeping in mind the potential triggering nature of the content.

Chapter 6

Chapter 6: Testing of the Proposed System

6.1 . Introduction to testing

In the development of our cyberbullying detection system, rigorous testing is pivotal to ensure its efficacy, reliability, and accuracy. Testing encompasses a series of processes designed to validate that the system accurately identifies instances of cyberbullying across various digital platforms, adhering to the predefined requirements and specifications. The overarching aim of this phase is to assess the system's performance through a multitude of scenarios, ensuring it functions optimally in real-world conditions.

6.2. Types of tests Considered

Our testing strategy for the Logistic Regression model was multi-faceted, encompassing:

- Unit Testing: We started with unit testing to validate the correctness of individual components within our Logistic Regression model. This included testing the data preprocessing, feature extraction, and logistic regression layers, ensuring each component functioned as expected independently.
- Integration Testing: Following unit testing, we conducted integration testing to examine the data flow and interaction between these components. This phase was crucial in ensuring seamless integration and data handling within the logistic regression model.
- System Testing: System testing allowed us to evaluate the logistic regression model's overall performance as a complete entity. This included assessing its user interface and API integrations, and ensuring compatibility with external environments.
- Performance Testing: We conducted detailed analysis of the logistic regression model's response time, accuracy, and resource utilization. This testing phase aimed to ensure that the model met operational requirements under various data volumes and system loads.
- Validation Testing: Finally, validation testing was carried out to confirm that our logistic regression model accurately achieved its objective, such as predicting outcomes or classifications, aligning with user expectations and requirements.

6.3 Various test case scenarios considered

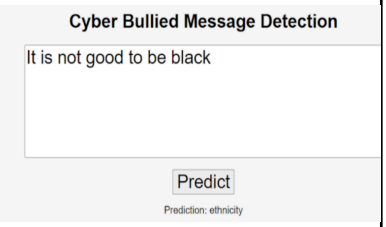
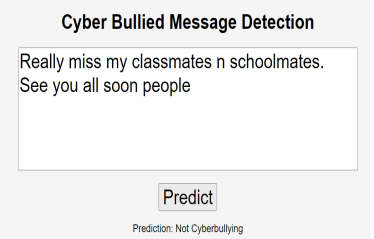
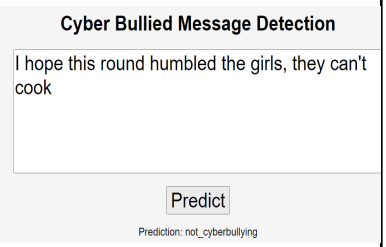
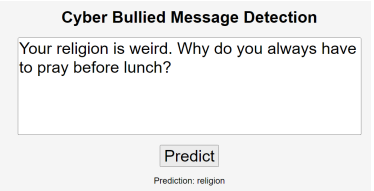
Test Case	Description	Example
Positive Test Case	We curated a dataset comprising explicit instances(types) of cyberbullying, including derogatory comments, threats, and harassment. The model was expected to correctly identify and flag these instances.	
Negative Test Cases	The same dataset with benign content, such as general discussions and positive interactions, was also tested. Here, the model's ability to discern and not falsely flag content was critical.	
Edge Cases	Ambiguous content that blurs the line between cyberbullying and non-cyberbullying was introduced to evaluate the model's precision and decision-making capabilities.	
Real-world Scenarios	Utilizing anonymized real-world data allowed us to observe the model's applicability in practical situations, ensuring its robustness and adaptability.	

Table 6.3.1 Explanation of test cases

6.4. Inference drawn from the test cases

The testing phase of our Logistic Regression model yielded insightful outcomes:

- The model demonstrated a high accuracy rate of 81%, in detecting cyberbullying content.
- Performance tests revealed that the model processes and classifies content efficiently, maintaining a steady accuracy rate even under increased data volumes, showcasing its scalability and robustness.
- Our analysis highlighted a low false positive rate of 4%, indicating a strong ability to distinguish between cyberbullying and non-cyberbullying content accurately.
- Despite its high efficacy, the model encountered challenges with very subtle or nuanced cyberbullying instances, suggesting an area for future enhancement, possibly through deeper contextual analysis or incorporating larger, more diverse data.

Chapter 7

Chapter 7: Results and Discussion

7.1. Screenshots of User Interface (UI) for the respective module

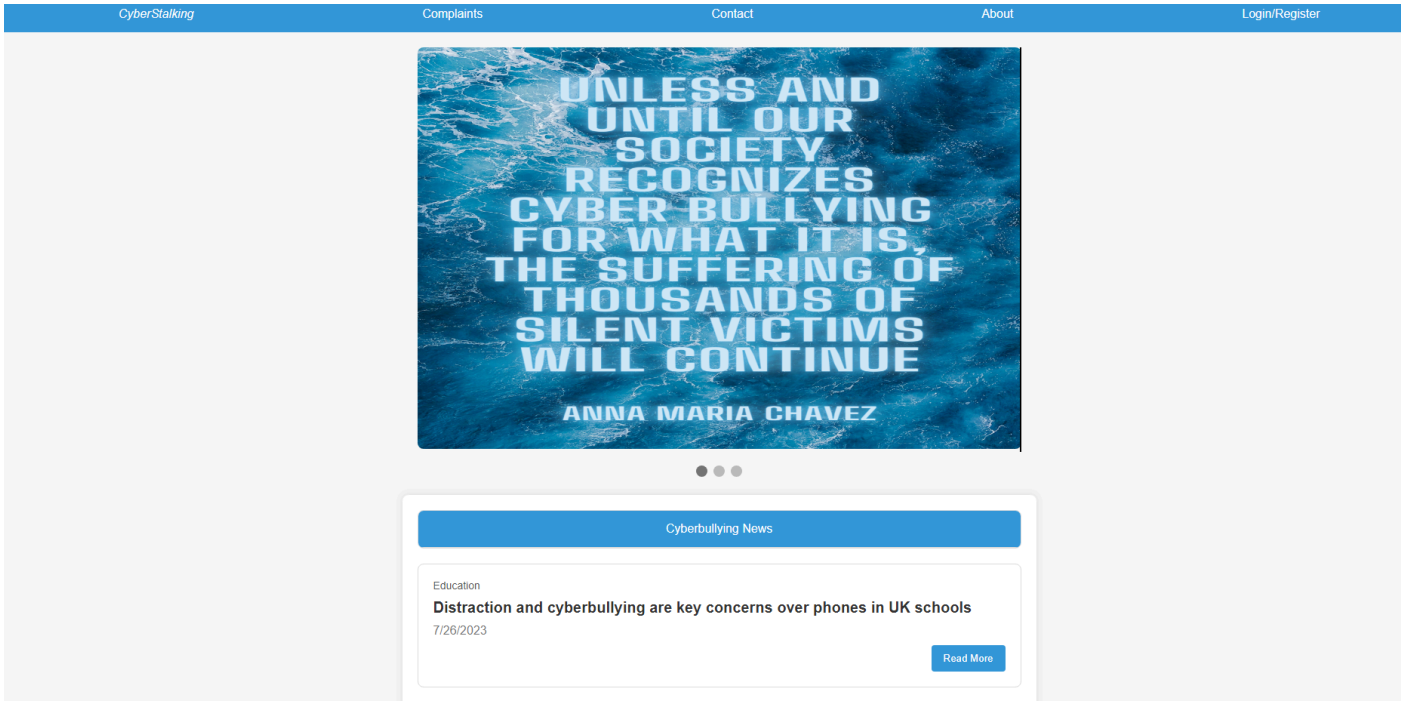


Fig. 7.1.1 Frontend

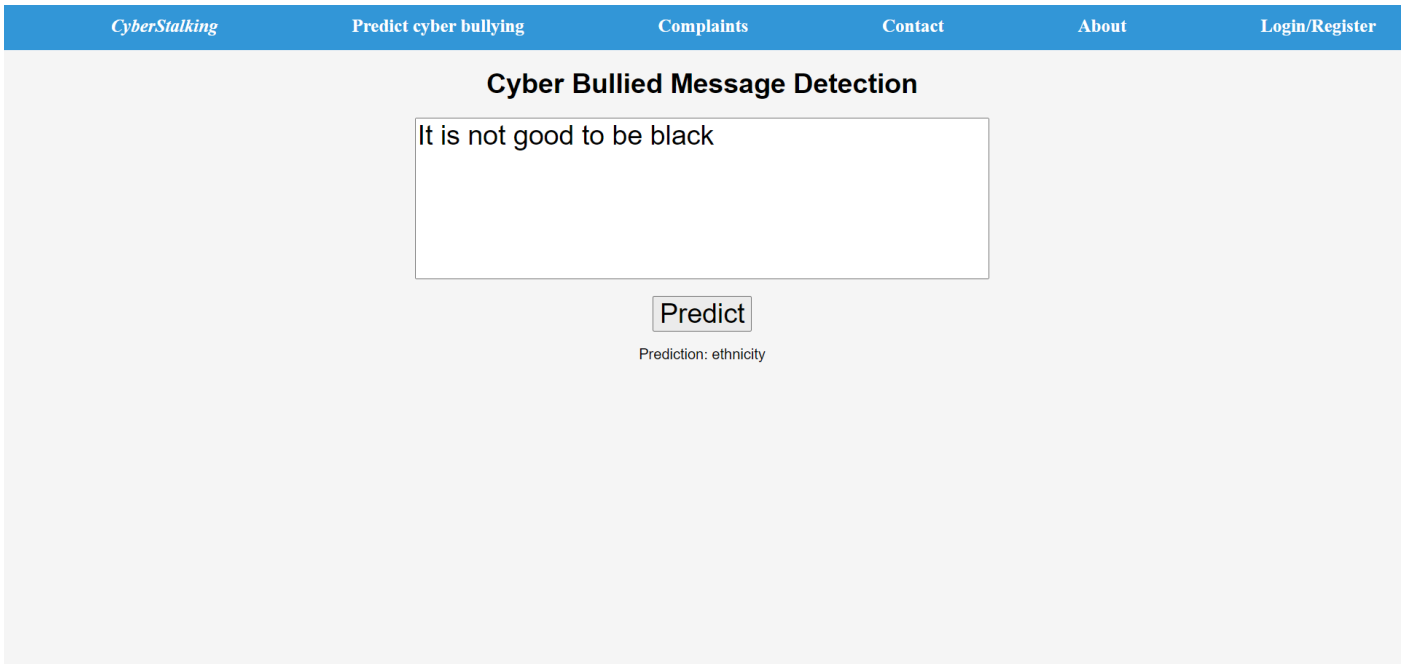


Fig. 7.1.2 Frontend

7.2. Performance Evaluation measures

Performance evaluation measures are used to assess the effectiveness and accuracy of machine learning models, particularly in classification tasks where the model predicts categorical outcomes. Here are some commonly used evaluation metrics:

1.Accuracy: Accuracy is one of the most straightforward metrics and represents the proportion of correctly classified instances among the total instances. It is calculated as the ratio of the number of correct predictions to the total number of predictions made by the model.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Where:

TP (True Positives): The number of instances correctly predicted as positive.

TN (True Negatives): The number of instances correctly predicted as negative.

FP (False Positives): The number of instances incorrectly predicted as positive.

FN (False Negatives): The number of instances incorrectly predicted as negative.

2.Precision: Precision measures the accuracy of positive predictions made by the model. It is the ratio of true positive predictions to the total number of instances predicted as positive, regardless of whether they were actually positive or negative.

Precision = TP / (TP + FP)

3.Recall (Sensitivity): Recall measures the ability of the model to identify all relevant instances, i.e., the proportion of actual positives that were correctly predicted by the model. It is the ratio of true positive predictions to the total number of actual positive instances.

Recall = TP / (TP + FN)

4.Support: Support represents the number of actual occurrences of the class in the specified dataset. In other words, it is the number of instances in each class.

5.F1 Score: F1 score is the harmonic mean of precision and recall. It provides a single score that balances both precision and recall. F1 score is especially useful when there is an uneven class distribution (class imbalance) in the dataset.

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

The F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0. It is a good way to compare models with different precision-recall trade-offs.

7.3. Input Parameters / Features considered

1)TF-IDF (Term Frequency-Inverse Document Frequency):

Term Frequency (TF): It measures the frequency of a term (word) in a document. It is calculated as the number of times a term appears in a document divided by the total number of terms in the document. The idea behind TF is that terms that appear more frequently in a document are more important.

Inverse Document Frequency (IDF): It measures how important a term is across a collection of documents. IDF is calculated as the logarithm of the total number of documents divided by the number of documents containing the term. The idea behind IDF is that terms that appear rarely in a collection of documents are more informative than terms that appear frequently across all documents.

TF-IDF: It combines TF and IDF to give a weight to each term in a document, reflecting both the importance of the term within the document and its rarity across the entire corpus. The higher the TF-IDF score of a term in a document, the more important or relevant the term is to that document.

2)Vocabulary (Vocab):

In the context of NLP, the vocabulary refers to the set of unique words or terms present in a corpus (collection of documents). Each word in the vocabulary represents a unique token that can be used to represent textual data. Building the vocabulary involves extracting and tokenizing words from the documents, removing stop words (common words like "the", "and", "is", etc.), and potentially applying stemming or lemmatization to reduce words to their base form.

3)CountVectorizer:

CountVectorizer is a technique used to convert a collection of text documents into a matrix of token counts. It essentially creates a document-term matrix (DTM) where each row represents a document and each column represents a term from the vocabulary. The value in each cell of the matrix represents the frequency of the corresponding term in the corresponding document. CountVectorizer works by tokenizing the text, building the vocabulary, and then counting the occurrences of each term in each document.

CountVectorizer ignores the structure of the text, such as word order or grammar, and only focuses on the occurrence of individual terms. This approach is useful for many text analysis tasks, including text classification, clustering, and information retrieval.

7.4. Graphical and statistical output

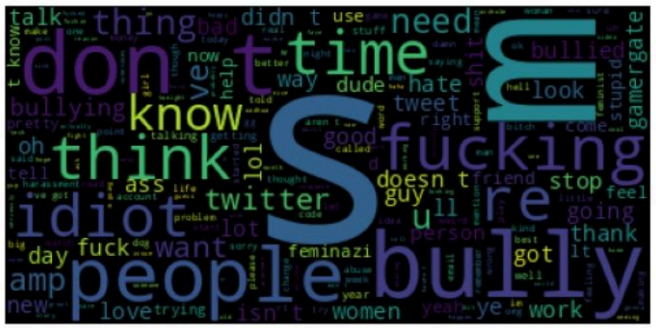
Age:

Gender:

Ethnicity:

Religion:

Other:



Not Cyberbullying:

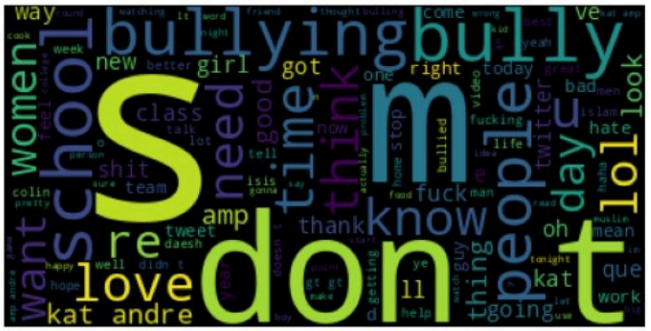


Fig. 7.4.1 Word clouds on Cyberbullying based on (a) Age, (b) Gender, (c) Ethnicity, (d) Religion. Other categories include (e) Other cyberbullying, (f) Not cyberbullying

Test Accuracy: 0.81

	precision	recall	f1-score	support
age	0.96	0.96	0.96	567
ethnicity	0.97	0.95	0.96	543
gender	0.89	0.80	0.84	550
not_cyberbullying	0.55	0.56	0.56	564
other_cyberbullying	0.58	0.63	0.61	567
religion	0.93	0.93	0.93	548
accuracy			0.81	3339
macro avg	0.81	0.81	0.81	3339
weighted avg	0.81	0.81	0.81	3339

Fig. 7.4.2 Different parameters for different types of cyberbullying Logistic Regression

Accuracy: 0.8025998532340917
Precision: 0.8040052930345933
Recall: 0.8025998532340917
F1-score: 0.8031305544976983

Fig. 7.4.3 Linear SVC

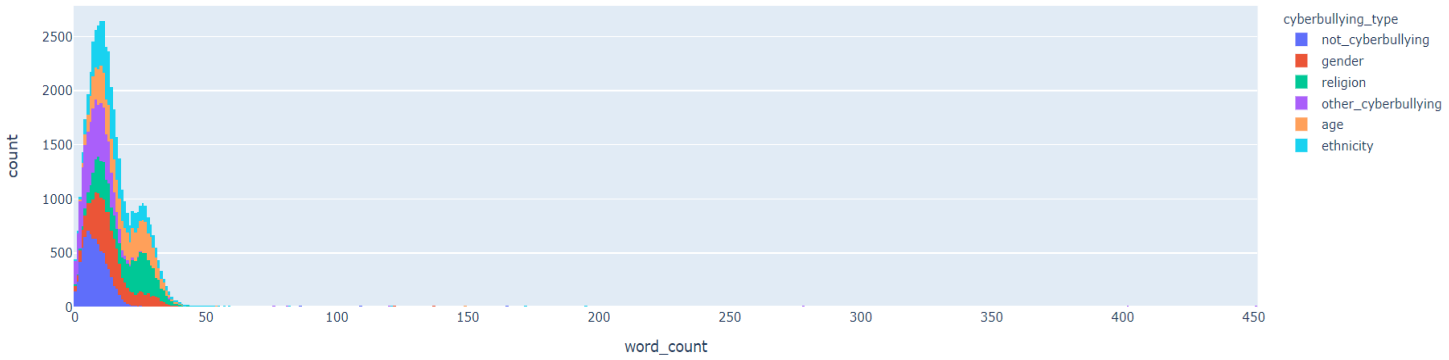


Fig.7.4.4 Graphical representation of cyberbullying words present in the dataset

7.5. Inference drawn (table of accuracies)

Model Name	Accuracy	Precision	Recall
CNN	0.80	0.78	0.78
Logistic Regression	0.81	0.84	0.835
XGBoost	0.81	0.53	0.55
Linear SVC	0.80	0.80	0.80

Table 7.5.1 Inference

Chapter 8

Chapter 8: Conclusion

8.1 Limitations

Our project can handle only text content. It can't handle images, audio or video directly. Users have to use some other model for extracting text from it. Development is done only for the English language which is used worldwide. However, it can't handle any regional language i.e model is not trained for multilingual content.

8.2 Conclusion

Through the incorporation of natural language processing (NLP) techniques, sentiment analysis, and machine learning algorithms, our tool can detect patterns indicative of cyberbullying and online harassment. By analyzing user profiles, activities, and connections, the tool aids in distinguishing genuine accounts from potential imposters, strengthening the foundations of trust in the digital realm.Initially we had chosen CNN model as it gives good accuracy on training data after running a number of epochs but it has low accuracy compared to Logistic Regression for test data. XGBoost has the same accuracy as Logistic Regression but other performance parameters are very low. We have chosen the Logistic Regression model as it gives good accuracy on test data & other performance parameters are also good.

The successful development of our tool will empower investigators, law enforcement agencies & organizations to proactively detect & respond to cyber bullying, online harassment & digital identity verification issues on social media platforms. It empowers investigators to swiftly identify and assess instances of harmful content, facilitating timely intervention and promoting a safer online environment. Our tool will streamline the investigative process, promote online safety & contribute to a healthier digital environment.

8.3 Future Scope

We can extend our project for cyberbullying detection from image dataset as well as audio & video datasets. Text has to be extracted from image, audio or video & then our present project can be used. We can incorporate other languages in our model. The successful development of our tool will empower investigators, law enforcement agencies & organizations to proactively detect & respond to cyber bullying, online harassment & digital identity verification issues on social media platforms.

References

[1] Helsper, E.J.; Kalmus, V.; Hasebrink, U.; Ságvári, B.; de Haan, J. Country Classification: Opportunities, Risks, Harm and Parental Mediation; EU Kids Online, The London School of Economics and Political Science: London, UK, 2013, DOI: http://eprints.lse.ac.uk/id/eprint/52023

[2] Ibrahim Arpaci & Omer Aslan (2023) Development of a Scale to Measure Cybercrime-Awareness on Social Media, Journal of Computer Information Systems, 63:3, 695-705, DOI: 10.1080/08874417.2022.2101160

[3] R. Armitage, "Bullying in children: Impact on child health," BMJ paediatrics open, vol. 5, no. 1, 2021, DOI: 10.1136/bmjpo-2020-000939

[4] Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel J. G. van Eeten, Michael Levi, Tyler Moore & Stefan Savage, Measuring the Cost of Cybercrime, 2013,

DOI:https://doi.org/10.1007/978-3-642-39498-0_12

[5] EMERGING TRENDS OF CYBER CRIME IN INDIA: A CONTEMPORARY REVIEW, Tanya Gupta, 2023, DOI**:** http://dx.doi.org/10.37253/jlpt.v8i1.7839

[6] A Secure Open-Source Intelligence Framework For Cyberbullying Investigation, Sylvia Worlali Azumah, Victor Adewopo , Zag ElSayed , Nelly Elsayed, Murat Ozer. DOI: arXiv:2307.15225v2

[7] Cyber Bullying Detection on Social Media using Machine Learning, Aditya Desai, Shashank Kalaskar, Omkar Kumbhar, and Rashmi Dhumal, 2021, DOI:https://doi.org/10.1051/itmconf/20214003038

[8] Cyber Bullying Detection for Hindi-English Language Using Machine Learning, Ninad Mehendale, Karan Shah, Chaitanya Phadtare, and Keval Rajpara, 2022, DOI:http://dx.doi.org/10.2139/ssrn.4116143

[9] An Application to Detect Cyberbullying Using Machine Learning and Deep Learning Techniques, Mitushi Raj, Samridhi Singh, Kanishka Solanki, and Ramani Selvanambi, 2022, DOI: https://doi.org/10.1007/s42979-022-01308-5

[10] Session-based cyberbullying detection in social media: A survey, Peiling Yi, Arkaitz Zubiaga, 2023, DOI:https://doi.org/10.1016/j.osnem.2023.100250.

[11] Social Media Cyberbullying Detection using Machine Learning, John Hani Mounir, Mohamed Nashaat, Mostafaa Ahmed, and Zeyad Emad, 2019, DOI: Link

[12] Alon Jacovi, Oren Sar Shalom, and Yoav Goldberg. "Understanding Convolutional Neural Networks for Text Classification." Affiliations: 1 Computer Science Department, Bar Ilan University, Israel; 2 IBM Research, Haifa, Israel; 3 Intuit, Hod HaSharon, Israel; 4 Allen Institute for Artificial Intelligence,2020,DOI: 1809.08037.pdf (arxiv.org)

[13] "Convolutional Neural Network: Text Classification Model for Open Domain Question Answering" by S. M. Kamruzzaman and A. Mustafa 2020 DOI: Link

[14] "Understanding Convolutional Neural Networks for Text Classification" by Denny Britz, 2018, DOI: https://aclanthology.org/W18-5408/

[15] Convolutional Neural Networks for Sentence Classification by Yoon Kim,20 DOI:https://aclanthology.org/D14

[16] R. R. Dalvi, S. Baliram Chavan and A. Halbe, Detecting A Twitter Cyberbullying Using Machine Learning, ICICCS, pp. 297–301, doi:10.1109/ICICCS48265.2020.9120893. (2020)

[17] S. M. Kargutkar and V. Chitre, A Study of Cyberbullying Detection Using Machine Learning Techniques, ICCMC, pp. 734–739, doi:10.1109/ICCMC48092.2020.ICCMC-000137. (2020)

[18] Trana R.E., Gomez C.E., Adler R.F. (2021) Fighting Cyberbullying: An Analysis of Algorithms Used to Detect Harassing Text Found on YouTube. In: Ahram T. (eds) Advances in Artificial Intelligence,

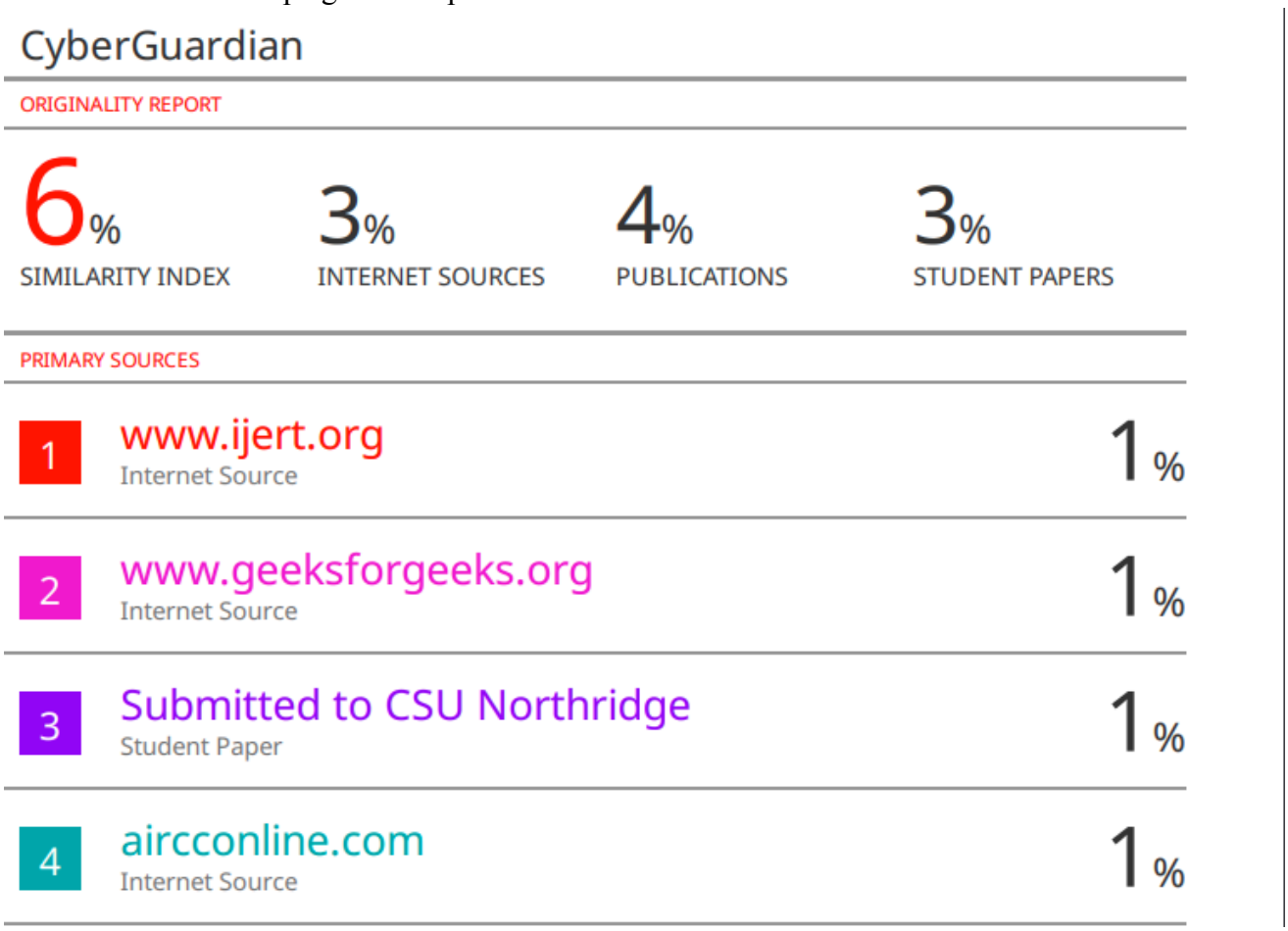
Software and Systems Engineering. AHFE 2020. Advances in Intelligent Systems and Computing, vol 1213. Springer, Cham. https://doi.org/10.1007/978-3-030-51328-3_2. (2020)

[19] Mihir Gada, Kaustubh Damania, Smita Sankhe, Cyberbullying Detection using LSTM-CNN architecture and its applications, doi:https://doi.org/10.1109/ICCCI50826.2021.9402412 (2021)

[20] Amirita Dewani, Mohsin Ali Memon, Sania Bhatti, Cyberbullying detection: advanced preprocessing techniques & deep learning architecture for Roman Urdu data, doi: https://doi.org/10.1186/s40537-021-00550-7(2021)

Appendix

i. Paper published
Link: Research Paper
ii. Plagiarism report
Recent plagiarism report:



CyberGuardian

ORIGINALITY REPORT

6%	3%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijert.org Internet Source	1%
2	www.geeksforgeeks.org Internet Source	1%
3	Submitted to CSU Northridge Student Paper	1%
4	aircconline.com Internet Source	1%

iii. Project review sheet

Inhouse/ Industry _Innovation/Research:

Class: D17 A/B/C

Group No.: 40

Sustainable Goal: Peace, Justice & Strong Institutions

Project Evaluation Sheet 2023 - 24

Title of Project: Social Media Investigations against Cybercrime : SocialSentinel

Group Members: Bhavesh Bhatia (D17A/07), Chaitanya Limaye (D17B/37), Chitra Attari (D17C/04), Siyona Singh (D17C/54)

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
3	3	3	2	3	2	2	2	2	2	2	2	2	1	4	35

Comments: understand the evaluation metrics clearly so that the performance of the System can be improved. Integration of UI & backend is remaining. Compare the evaluation of all models.

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
03	04	04	02	03	02	02	02	02	02	02	02	02	01	04	37

Comments:

Date: 10th february, 2024

Name & Signature Reviewer 2

Inhouse/ Industry _Innovation/Research: Peace, Justice & Strong Institution

Class: D17 A/B/C

Group No.: 40

Sustainable Goal:

Project Evaluation Sheet 2023 - 24

Title of Project: Social Media Investigations against Cybercrime : Social Sentinel

Group Members: Bhavesh Bhatia (D17A) 07, Chaitanya Limaye (D17B) 37, Chitra Attari (D17C) 04, Siyona Singh (D17C) 54

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
3	4	4	3	3	2	2	1	2	1	2	3	2	2	3	37

Comments: Need to implement Admin dashboard for complaint functionality. Predictions results can be improved using other NLP algorithms.

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)

Comments:

Date: 9th March, 2024

Name & Signature Reviewer 2