iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56676

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Cyber Bullying

Ms. Swapna S Banasode[1], Nidhi K V[2], Sabhareesh Balaji P[3], Tanvi Kamath[4], Tushar S[5]
Department of Computer Science and Engineering, K S Institute of Technology, Bengaluru, India

Abstract: ***Cyberbullying involves the use of cell phones or social networking platforms like Facebook and Twitter to harass, threaten, or intimidate individuals, adversely impacting their mental health. Our project focuses on the detection of such harmful comments using a machine learning algorithm, categorizing them into two severity levels: "low" and "high." Given the ubiquitous nature of social media, especially in the post-pandemic era, it has become an integral part of our lives for communication across all age groups. While it facilitates easy communication, social media also presents challenges, such as the escalation of hateful comments beyond the realm of healthy discourse. Recognizing the potential harm caused by these comments, our solution aims to identify and classify them based on severity using machine learning, enabling appropriate intervention.***

I. INTRODUCTION

Bullying of children and teenagers is primarily conducted on social media. In their daily lives, people pay close attention to everything. Social networking is being used by many people to boost their careers. Prefer putting their skills to use and sharing those things on various social media networks. Such as social networking sites like Instagram. Simply leaving abusive remarks on someone else's posts qualifies as cyberbullying. their mental health is so disturbed. Many young people are being impacted by bullying. Cyberbullying is on the rise as social networking usage grows. Our goal is to pinpoint and mitigate hostile remarks by utilizing specific machine learning models. We categorize the severity of these remarks into three levels: High, Low, and Medium. The classification of remarks as hateful or not is contingent on their level of seriousness. Though there are various approaches to tackle media bullying, a considerable number of them have been primarily focused on text. This paper aims to showcase a software solution designed to identify hostile remarks made by individuals engaging in bullying behavior. Creating a false identity and sharing an embarrassing image or video are just two examples of cyberbullying; it also involves spreading unfavorable rumors and making threats. The victim's conduct is altered by the poster of such cruel comments. This has an impact on their emotions, as well as their self-confidence and sensation of terror. A comprehensive solution is therefore needed for this situation. Cyberbullying must be stopped. Using machine learning models, the issue can be resolved by identifying and preventing it.

II. AIM OF THE PROJECT

Project is to Detect such hateful comments and block them. So, here we use some Machine Learning Models to detect such comments based on the Severity. To find the solution for cyberbullying and detecting such hateful comments and blocking them. The ultimate goal is to put an end to cyberbullying.

Therefore, a comprehensive solution is needed for this issue. Cyberbullying must end. The issue can be resolved by employing a machine learning approach to detect and prevent it, but this needs to be done from a different angle. Our paper's major goal is to create an ML model that can identify and stop social media abuse so that no one has to experience it.

III. EXISTING SYSTEM AIM OF THE PROJECT

1) Identify Cyberbullying in social media through machine learning methods.
2) Detects harmful comments with the use of the Deep learning model.
3) All the hateful comments are automatically reported upon their detection.

A. Disadvantages

1) No actions are taken to reduce such hateful comments.
2) All the comments are detected as hateful comments irrespective of their level of severity.
3) Not all hateful comments are required to be reported automatically.

B. Objectives

The objective of proposed system is given by:

1) To detect such hateful comments based on the severity levels.
2) By using machine learning models, it will be classifying.
3) To reduce cyberbullying on the social media platform.

IV. METHODOLOGY

Each system component is described in the system architecture. The following list of modules makes up this project:

1) *Dataset:* The dataset includes over 40 thousand comments that have been categorized as hateful and was collected from the Kaggle website.
2) *Pre-processing:* Before the input is fed to the algorithm, several modifications are made to it. Data rescaling, binarization, and standardization are all included in this pre-processing.
3) *Feature extraction:* We are using TF-IDF algorithm for feature extraction from natural language comments. TF-IDF algorithm short for term frequency– inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. This converts each word in the comments into a vector of integers which is forwarded into the next module.
4) *Classification:* The dataset is classified into two types using SVM (Support Vector Machine): hateful and non-hateful remarks. The goal of the SVM algorithm is to determine the optimal line or decision boundary for classifying a space with n dimensions. A hyperplane is the optimal choice boundary.

V. PROPOSED SYSTEM ARCHITECTURE

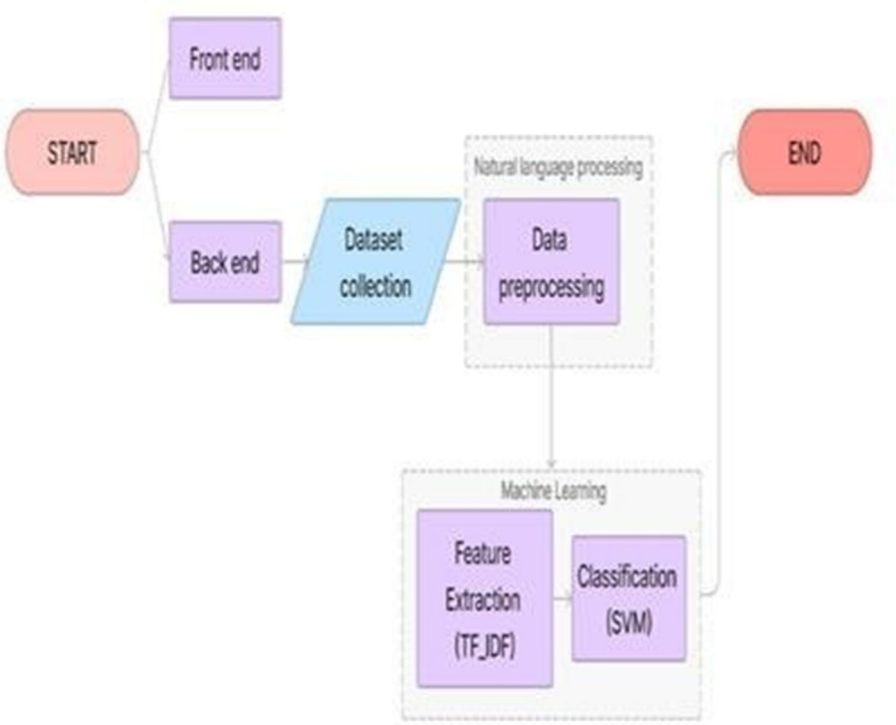


Fig.1 Block Diagram of Proposed System Architecture

A. Use Case Diagram

At its most basic level, a use case diagram isa representation of how a user interacts with the system that shows how the user is related to all of the many use cases in which they are involved. It is possible to identify the various system users and use cases using a use case diagram, which is typically complemented by other types of diagrams. The usage cases are shown as circles or ellipses, respectively.

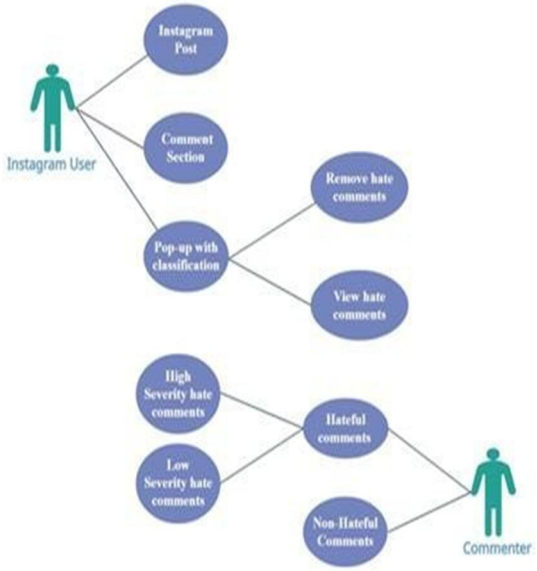


Fig.2 Use Case DiagramInitially we collect the data from the end user and save it in database.

We extract relevant data from the database and conduct thorough data cleaning and transformations to ensure its conformity to the required format. Following these transformations, the prepared data is then fed into the predictive model for making predictions.
The result is presented to the end user through a visual representation in the form of graphics.

1) Module 1

Data provider: Dataset of 20,000 comments is retrieved from Kaggle which includes the classification of hateful comments. It includes the distribution of 12,000 non- hateful comments and 8,000 hateful comments. To obtain the severity of the comments, we conducted a survey of those 8,000 hateful comments to classify them as low-severity hateful comment and high-severity hateful comment. Using which the classification model is trained.

2) Module 2

Data Preprocessing: Pre-processing refers to the modifications that are performed on the data before it is fed into the algorithm. Tokenization, deleting stop words, removing punctuation and numerals, and Porter Stemmer are the pre-processing techniques employed. Tokenization is a technique used in natural language processing to divide paragraphs and phrases into smaller parts that may be ascribed meaning more readily. The process of reducing a word to its stem that affixes to suffixes and prefixes or to the roots of words known as “lemmas" is known as stemming. After that, the data is ready for feature extraction.

3) Module 3

The TF-IDF technique is used to extract features from natural language comments. The TF-IDF method, short for term frequency-inverse document frequency, is a numerical statistic designed to indicate the importance of a word in a collection or corpus of documents. phrase frequency works by examining the frequency of a certain phrase in relation to the document. Inverse document frequency assesses the frequency (or rarity) of a term within the corpus.

4) Module 4

In the classification process, we utilize the Random Forest algorithm, a machine learning method designed to address both classification and regression challenges. Random Forest employs ensemble learning, a technique that resolves intricate problemsby amalgamating multiple classifiers.
The algorithm determines the outcome based on predictions from decision trees, making predictions through the averaging or combination of results from various trees. The accuracy of the outcome improves proportionally with the increase in the number of trees.

5) Module 5

Frontend: The post page of Instagram web application is duplicated to show the practical application of the project. The tech stack used is React JS and it is connected to the python backend using Flask as the middleware.

VI. FUTURE IMPLEMENTATION

1) The hateful comments are classified into 2 levels based on their severity:
2) Highly severe (E.g.- Go Die, Here the Word 'Die' is Highly Severe).
3) Constructive criticisms/Low severe (E.g.- Color is Not Good, Here the Word 'Not Good' is Low Severe).
4) Highly severe comments will be automatically reported, and the low-level comments will be tagged so that the commenter will know their comment is a hateful one.
5) Users will be given a choice of whether they would like to view the low-level hateful comments, or it will be removed from their view.

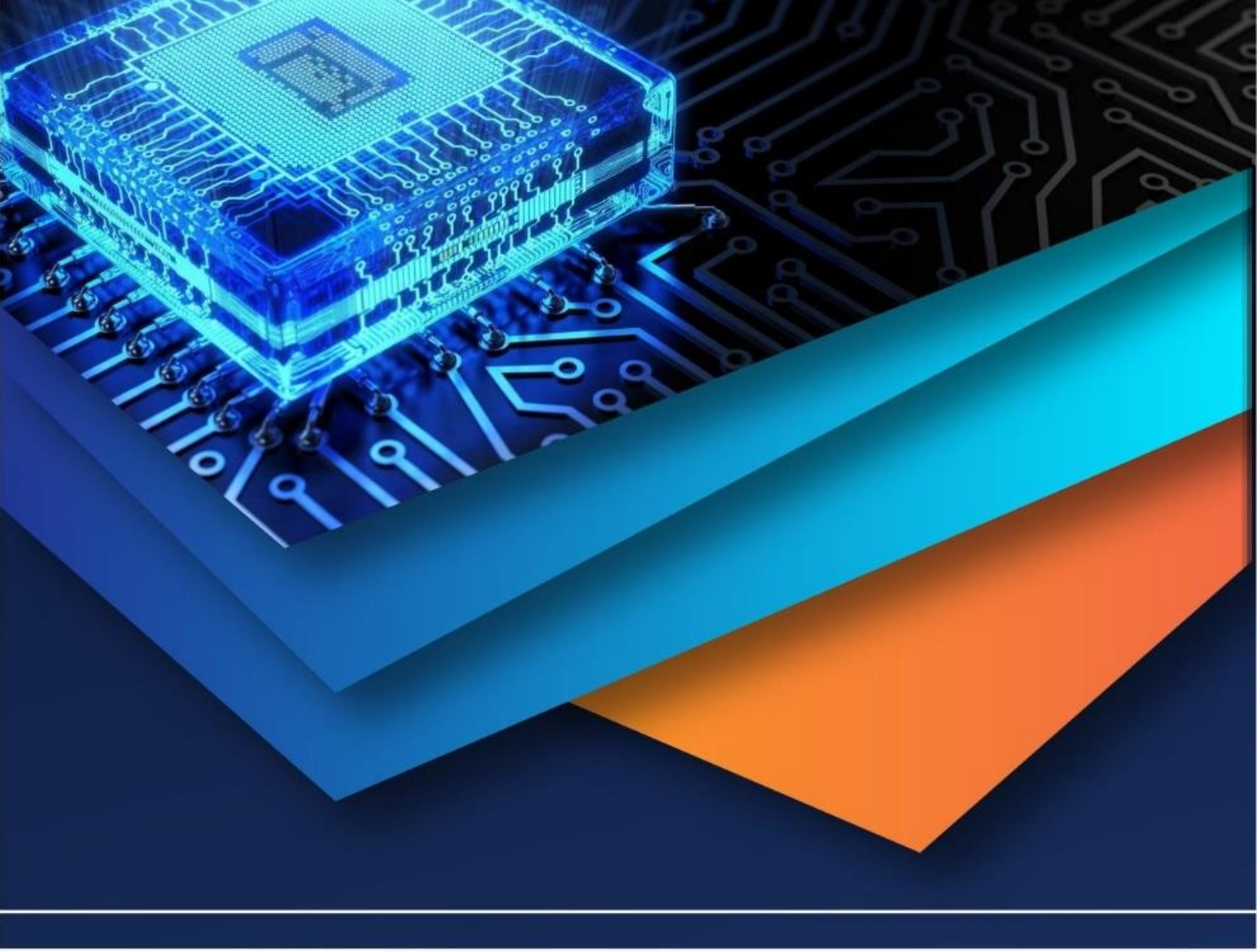
OBJECTIVES

The objective of Implementation is given by:

a) To detect such hateful comments based on the severity levels.
b) By using machine learning models, it will be classifying.
c) To reduce cyberbullying on the social media platform.

REFERENCES

[1] A Study of Cyberbullying Detection Using Machine Learning Techniques– Saloni Mahesh Kargutkar -Prof.Vidya Chitre IEEE 2020
[2] USE OF MACHINE LEARNING TO IDENTIFY CYBERBULLYING - Gandhali Jadhav l, Mrs. J.C. Pasalkar International Journal of Research Publication and Reviews, Vol 3, no II. pp 2744-2747 November 2022
[3] Detection of Cyberbullying on social media Using Machine learning -Varun Jain -Vishant Kumar IEEE2021
[4] CYBERBULLYING DETECTION USING MACHINE LEARNING - Shree Nidhi B S1, Mohammed Zaid Hulikatti2, Nafey A H3, Neha M R4 , Shradha S- International Journal of Creative Research Thoughts(IJCRT) -© 2022 IJCRT |Volume 10, Issue 4 April 2022



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY