IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14405

Cyber Bullying Detection in Twitter Social Media Platforms

B Naga Raju[1], Vishnumolakala Mohith[2], V Venkata Naveen[3], S Abdul Azeez[4], Y Venkateshwaralu[5]

Assistant Professor, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, India[1]

Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, India[2-5]

Abstract: This project intends to develop a comprehensive categorization system based on machine learning to address the complex concerns of online harassment and discrimination on social media platforms. Inspired by recent research advocating for the use of machine learning in social media moderation, this project builds on existing methodologies to create a comprehensive framework capable of identifying various types of harmful content, such as cyber and non-cyber bullying, as well as discrimination based on ethnicity, gender, age, and religion. Machine learning models such as Naive Bayes, SVM, Random Forest, Decision Tree, and Sklearn classifiers are trained to detect patterns and subtle nuances indicative of online abuse and discrimination by utilizing diverse datasets representing instances of harmful behavior across multiple dimensions. The suggested categorization system's performance and flexibility are tested by comprehensive testing and assessment on real-world social media data. The technology provides timely and precise identification of hazardous information by combining different categorization tasks under a uniform framework, allowing social media platforms to handle its propagation proactively. Furthermore, the use of machine learning algorithms improves the scalability and effectiveness of content moderation activities, reducing the burden on human moderators and creating a safer and more inclusive online environment. This study contributes to a better understanding of the complex dynamics of online abuse and discrimination, enabling the creation of nuanced solutions for enhancing online safety and content control.

Keywords: child predators, cyber harassers, Twitter, machine learning.

I. INTRODUCTION

People of all ages choose to connect and socialize online using social media platforms such as Facebook, Instagram, Flickr, Twitter, and Facebook. Although these platforms allow people to interact and communicate in previously imagined ways, they have also given birth to bad activities such as cyberbullying. Psychological abuse, including cyberbullying, has a significant impact on society. The majority of young people who spend their time bouncing between social media platforms have seen an upsurge in cyberbullying incidences. Because of their extensive use and the anonymity provided by the Internet, social media sites such as Facebook and Twitter are particularly prone to cyberbullying (CB).

Cyberbullying, defined as a purposeful and repetitive act aimed at injuring or humiliating persons using information and communication technology, takes many forms, including online harassment, denigration, and flame. Its insidious nature, regardless of the bully's outward appearance, poses a Persistent danger in online environments. Recognizing the societal consequences of cyberbullying, it is critical to develop efficient detection tools. The automatic recognition of emojis, bully phrases, and audio characteristics across internet platforms, notably microblogging sites like Twitter and Facebook, as well as video-sharing platforms like YouTube, is critical to this effort.

In this case, the classification algorithm works inside a supervised framework to detect several types of harmful material, such as cyber and non-cyber bullying, as well as discrimination based on race, gender, age, or religion. Naive Bayes classifiers, based on Bayes' Theorem, are a set of classification algorithms used for this purpose. These classifiers function in a supervised framework, using tagged or labelled training datasets to detect the existence of bullying language inside a given communication. This technique allows the system to effectively identify instances of cyberbullying and prejudice, adding to continuous efforts to lessen their negative consequences on both individuals and communities.

DOI: 10.17148/IJARCCE.2025.14405

II. LITERATURE REVIEW

Online harassment has been a pervasive issue from the early days of social media, and it remains so now. The primary purpose of these research was to establish a computerized process that could recognize and report this sort of misbehavior [1]. Two methods—machine learning and deep learning—have been researched to prevent or identify incidents of sexual harassment and protect children from bullying in order to offer a safe environment. The authors of this work [2] used fuzzy logic and genetic algorithms to track the frequency of cyberbullying on social media sites. They identified and classified offensive, harassing, racist, and terroristic statements, as well as cyberbullying-related words and acts on social media [3]. The resulting F-measure was 0.91. To optimize parameters and achieve maximum performance, a genetic. The authors of this work [4] used fuzzy logic and genetic algorithms to track the frequency of cyberbullying on social media sites. They identified and classified offensive, harassing, racist, and terroristic statements, as well as cyberbullying-related words and acts on social media. The resulting F-measure was 0.91. A genetic algorithm is used to optimize parameters and performance.

The authors in ref [5] used three weighting techniques for Facebook message filtering: entropy, term frequency-inverse document frequency (TFIDF), and modified TF-IDF for feature selection. The support vector's A Support Vector Machine was used to measure recall, accuracy, and precision. According to the test results, the modified TF-IDF outperforms the other schemes, with a 96.50% accuracy. This work in [6] evaluated supervised machine learning (ML) algorithms for natural language learning with the evaluation of online harassment in Twitter tweets as part of the social media competitiveness and harassment (feature). Features were retrieved using Word2Vec embeddings and the TF-IDF algorithm. The results accurately covered more than 80% of the harassment subcategories considered in the data.

This study [7] combines a modern sentencing vector approach with emotion analysis. As a novel way of predator sexual identification, word vectors are created using the Long-Short-Term Memory, Recurrent Neural Network (LSTM_RNN) language pattern. With a recall of 81.10%, the last stage of extracting the emotion value from the SoftMax layer outputs achieved a record-breaking accuracy rate. In reference [8], the authors create a Twitter post classification model for harmful intent by extracting characteristics from tags using convolution neural networks (CNN). A four-month Twitter dataset was examined to identify the story settings that expressed evil intent. They examined the significance of these occurrences in developing legislation against gender-based violence.

Sweta Karlekar describes the Safe City Web Community's efforts to categorize and study various types of sexual harassment in [9]. By sharing their tales, Safe City Web uses this information to help victims create web directories, provide more in-depth safety advice, and aid others in discovering relevant cases to prevent further sexual assault. The single-label CNN-RNN model has an accuracy of 86.5% for processing, linking, and annotating tags. Using Twitter, Espinoza [10] develops a new data set with four areas for identifying harassment. They classified the tweets using two distinct deep learning architectures: CNN and LSTM. F1's measurement during training was 55%; however, F1's results for the test set were just 46%. Arijit Josh Chowdhury [11] proposes a disclosure language paradigm. The ULMFiT fine-tuning architecture consists of a language model, a task-specific classifier, and a single mediator (Twitter). The overall comparison illustrates the benefits of some lightweight LSTM-based mean language models as well as an expanded vocabulary that captures linguistic intricacies seen in the deep text dealing with sexual harassment. Approximately 10,000 personal tales of sexual harassment were annotated to extract crucial elements and automatically classify the stories using neural network models, which yielded good results. With a 92.9% accuracy rate, additional advancements in categorization were made by evaluating the specificity of crucial features. In [12], the author presents an automated method for analyzing online discussion content to determine whether a participant or another user is a cyber-predator's target.

Each stage of the classification tasks employs a recurrent neural network (RNN), which has obtained an F0.5 score of 0.9058. The authors of [13] provide a novel method for categorizing sexual predators and sexual harassment in English by integrating word embedding, BiLSTM, and Gated Recurrent Unit (GRU) algorithms. In contrast to Extreme Gradient Boosting (XGBoost), which reached 90.10%, pre-trained Global Vectors (GloVe) words achieved 97.27%. The work in [14] describes a two-stage method for using machine learning classifiers to detect sexual predators in online interactions. Soft voting and Naive Bayes appear to be the most effective technique, with an F-0.5 score of 0.9348. [15] It is the most extensively used dataset for predatory identification tasks in online conversation, and it has been utilized in numerous recent related works.

In 2015, B. Sri Nandhini and J. I. Sheeba [16] published an approach for recognizing cyberbullying acts. is composed of the following stages: FuzGen learning technique, feature extraction, data preprocessing, and the Naive Classifier Technique. The genetic algorithm is part of the learning algorithm unit.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14405

Knowledge is expressed through an unclear set of rules. The primary goal is to change the way information is displayed for classification while retaining historical knowledge. The data is processed by a genetic algorithm and stored in a chromosomal population. Everyone's chromosome is attempting to predict how cyberbullying instances will be classified. The classifier technique employs the fitness value of the chromosome to categorize cyberbullying activities based on the learning unit results.

Batoul Haidar et al. (2018) presented an Arabic-language cyberbullying detection system. The Arabic dataset, which was obtained from Twitter, was manually annotated. The first step was to remove any emoticons, hyperlinks, and non-Arabic characters. The designations "0" and "1" represented non-bullying and bullying content, respectively. The dataset was tokenized into words, which deleted any unnecessary characters. Word embeddings were constructed using one-hot encoding. Using this Arabic dataset, the model was trained with a Feed-Forward Neural Network. The FFNN model was developed using four hidden layers. The dataset was split into 80% training and 20% testing, with the model programmed to shuffle it at each epoch. As a result, the performance measure was 91.17%. The accuracy of testing and validation was discovered.

III. SYSTEM ANALYSIS

To guarantee effective implementation, the proposed project's system analysis would include a detailed examination of its requirements, features, and limitations. This involves determining particular system needs such user registration, login, publishing, and content classification, as well as outlining the project's scope, objectives, and target audience. Functional analysis entails dividing the system's functionality into smaller components, producing use case diagrams to depict user interactions, and specifying input, output, and processing requirements. Data analysis includes identifying the types of data that the system will manage, selecting data sources and storage needs, and evaluating data. Processing requirements. This detailed study guarantees that the project fits the demands of users, follows best practices, and successfully handles the identification of hazardous information on social media platforms.

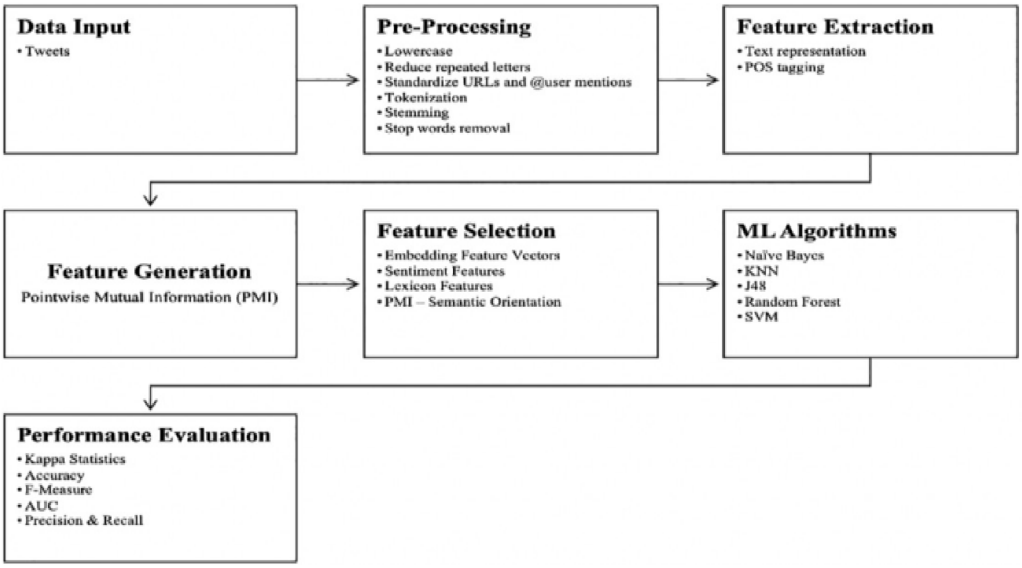

Figure 1: Model architecture

A. EXISTING SYSTEM

In the current system, strategies for locating child predators on the internet have been created, notably in gaming, voice chat, and other online entertainment platforms. These approaches enable parents to protect their children from potential sexual exploitation as they engage in internet activities. However, given the internet's ubiquitous effect in today's culture, many youngsters are increasingly using social networking sites as their major mode of communication. As a result of the lack of specific detection tools, minors on these platforms are susceptible to sexual predators.

The present technique employs five classification algorithms: the conversation-centered approach uses the Ridge or Naive Bayes Classifiers, both of which operate on the TF-IDF feature set, as well as the Neural Network Classifier, which also operates on the TF-IDF feature set. However, despite these efforts, present systems' accuracy in detecting harmful information, including cyber and non-cyber bullying, need improvement.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14405

B. PROPOSED SYSTEM

Our suggested method aims to create a comprehensive strategy for identifying harmful information on social media platforms, such as cyber and non-cyber bullying, as well as discrimination based on race, gender, age, and religion. To achieve this aim, we will use a number of machine learning approaches, such as K-Nearest Neighbors, Random Forest, Support Vector Machine (SVM), Naive Bayes, and Decision Tree algorithms.

The study will include training these models on datasets that include both regular and disturbing words and communications. These datasets will be meticulously chosen to include a diverse variety of hazardous behaviors observed on social media. Once trained, the algorithms will be used to evaluate user posts and automatically determine if they include harassing or ordinary material.

Our solution will be constructed with the Django framework, which provides a solid base for creating a web application specifically designed for identifying child predators and cyberbullies on social media sites. The application's backend logic will be created to detect and identify dangerous actions, including user registration, login, publishing, and text classification using machine learning techniques.

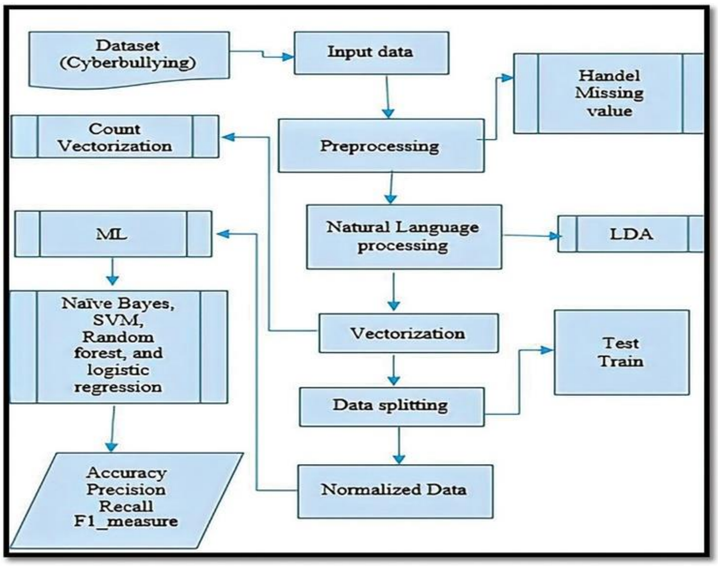

Figure 2: Proposed system architecture

Key elements and features of the proposed system include:

- The application will import Django modules and Python libraries for web development, data processing, machine learning, and database access.
- Data processing and machine learning activities will use global variables such as classifier, label_count, X, Y, and corpus.
- Django views will manage HTTP requests and provide HTML templates for application pages like Index, SendPost, Register, Admin, Login, AddCyberMessages, RunAlgorithms, MonitorPost, AddBullyingWords, Signup, UserLogin, AdminLogin, ViewUsers, ViewUserPost, word_count, prediction, cal_accuracy, and classifyPost. Each view will have a unique URL endpoint.
- The program will use the pymysql package to connect to a MySQL database and get and insert data, including user and post information.
- The dataset will be processed using machine learning methods, including SVM, Decision Tree, K-Nearest Neighbors, Random Forest, and Naive Bayes, to categorize text.
- Web Forms Handling: AddBullyingWords, Signup, UserLogin, and AdminLogin handle user-submitted forms for database changes and authentication.
- The File Upload component manages file uploads, including categorizing messages and user profile images.
- The user interface will be rendered using HTML templates, such as "index.html," "SendPost.html," "Register.html," and "Admin.html".
- Tokenizing words, deleting special characters, and converting to lowercase are part of the preprocessing stages for data.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14405

- Classification: Machine learning models will be used to classify text messages, and the findings will be displayed to users.
- Session Management: After successful login, a session.txt file is stored to manage user sessions.
- HTML templates will be used to convey classification results and related information to users. classification results and related information to users.

C. Data Preprocessing:

- Explain data preparation processes, including importing datasets, managing missing data, encoding categorical data, and dividing datasets.
- Understanding the significance of each preprocessing step for machine learning models.
- Python libraries used for data preparation include NumPy, Matplotlib, and Pandas.

D. TF-IDF Feature Extraction:

- Define and explain the TF-IDF (Term Frequency-Inverse Document Frequency) approach for natural language processing (NLP).
- Explained mathematically the components of TF (Term Frequency) and IDF (Inverse Document Frequency).
- TF-IDF plays a crucial role in assessing word importance in documents or corpora.
- Clarified how raw text data is transformed into numerical characteristics using TF-IDF.

The algorithms mentioned in the project analysis include a wide range of machine learning approaches that are essential for efficient data processing and classification jobs. Data preparation is highlighted as the first phase, highlighting its importance in preparing raw data for machine learning models by resolving difficulties such as missing values and encoding categorical data. TF-IDF feature extraction emerges as a crucial approach for detecting the importance of words in texts, which is required for natural language processing applications.

The Support Vector Machine (SVM) classifier is then investigated, notably in its use for detecting spam comments on social media sites using TF-IDF feature vectors. Decision tree classifiers are highly praised for their capacity to extract decision-making information from datasets, which improves interpretability and efficacy in classification tasks. Gradient boosting is also introduced as an ensemble learning strategy, which is recognized for combining weak learners such as decision trees to increase prediction performance. K-Nearest Neighbors (KNN) is notable for its simplicity and power in classification tasks, which rely on similarity measurements to categorize data points. The following section discusses logistic regression classifiers, highlighting their usefulness in assessing relationships between categorical dependent and independent variables. Finally, the Random Forest method is evaluated for its usefulness in ensemble learning, where it uses numerous decision trees to improve prediction accuracy, making it a versatile option for classification and regression tasks.

IV. DATASET

Our Dataset consists of different types of tweets, like

- **Ethnicity**: This class categorizes tweets that include allusions to certain ethnic groups or identities. Tweets may mention ethnicity, cultural heritage, or racial diversity and inclusiveness.
- **Gender**: Tweets with this classification are likely to contain debates, mentions, or references to gender or gender identity issues. This might involve conversations about gender equality, stereotypes, or specific gender identities.
- **Age:** This label indicates tweets that reference age-related topics or discussions about different age groups. Tweets may discuss age-related experiences, milestones, or issues affecting specific age demographics.
- **Religion:** Tweets categorized under this label likely contain references to religious beliefs, practices, or discussions about religious identity. This could include discussions about religious traditions, beliefs, or current events related to religion.
- **Bullying**: Tweets tagged "bullying" show instances of abusive conduct, harassment, or intimidation directed at specific persons or groups within the text. This category indicates tweets that use bullying language or conduct.
- **Non-Bullying**: Tweets categorized as "non-bullying" denote content that is free of bullying conduct. These tweets might contain neutral or positive comments, news updates, or general opinion unrelated to bullying.

These categories categorize tweet content, allowing for analysis and classification based on themes and features.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14405

V. IMPLEMENTATION & RESULTS

Figure 3: Illustration of the index URL page for Cyberbullying Detection on the Twitter Social Media Platform, including the home, user, registration, and administrator modules.

Figure 4: Visit the Register URL page to establish an account as a new user.

Figure 5: displays a confirmation screen when a user has successfully registered for an account. It provides a message verifying that the registration procedure is complete. If a previously existing user attempts to register again, it shows a notice indicating that the user already exists and does not register.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14405

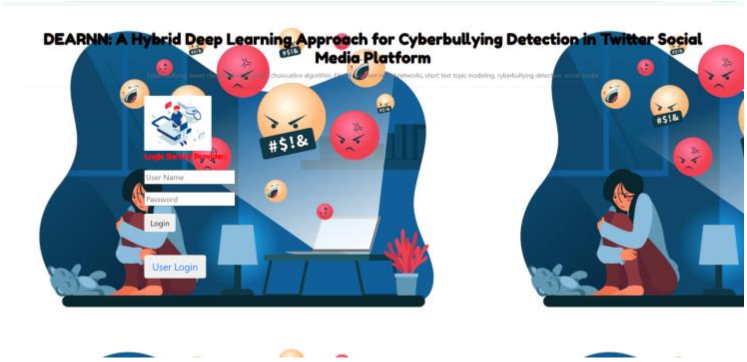

Figure 6: The admin URL page allows you to login as an admin using the username admin and password admin.

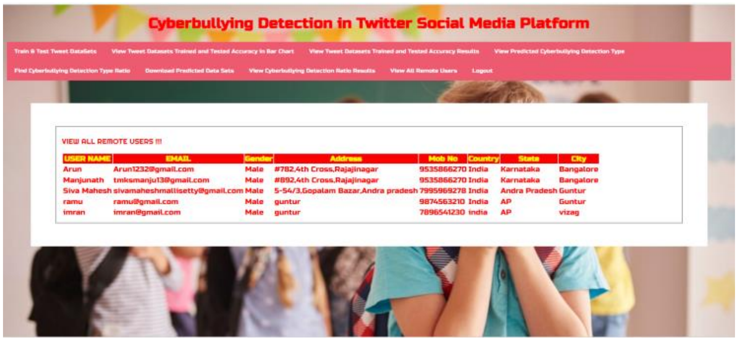

Figure 7: Admin login page displays modules such as Train & Test Datasets, see all remote users, monitor posts, and check cyberbullying detection ratio results. Find the ratio of cyberbullying detection types, then log out.

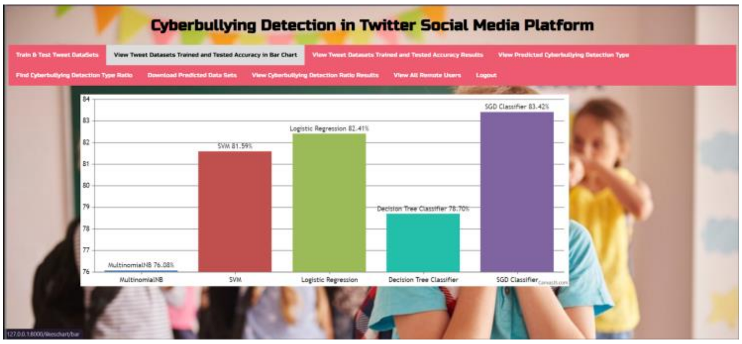

Figure 8: A bar chart of the Tweets dataset, which was developed and evaluated using several machine learning models.

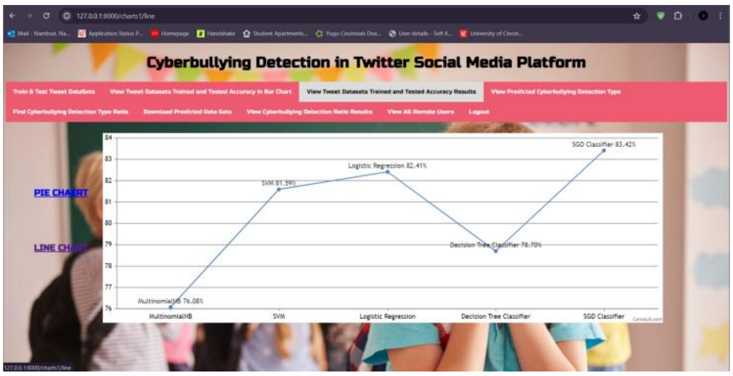

Figure 9: The train and test twitter datasets website displays many machine learning models, including SVM, naïve bayes, random forest, decision tree, and KNN classification.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪬ Peer-reviewed & Refereed journal ⪬ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14405

Figure 10: Type ratio for detecting cyberbullying through tweets.

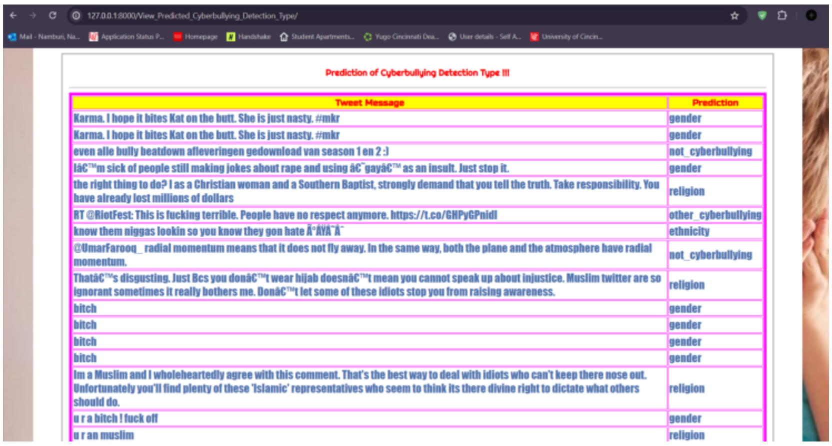

Figure 11: Monitor the post page, which displays registered users' messages, uploaded files, and the prediction status (bullying, not_cyberbullying), as well as religion, age, gender, and ethnicity.

Figure 12: Home page where users may submit their login information.

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14405

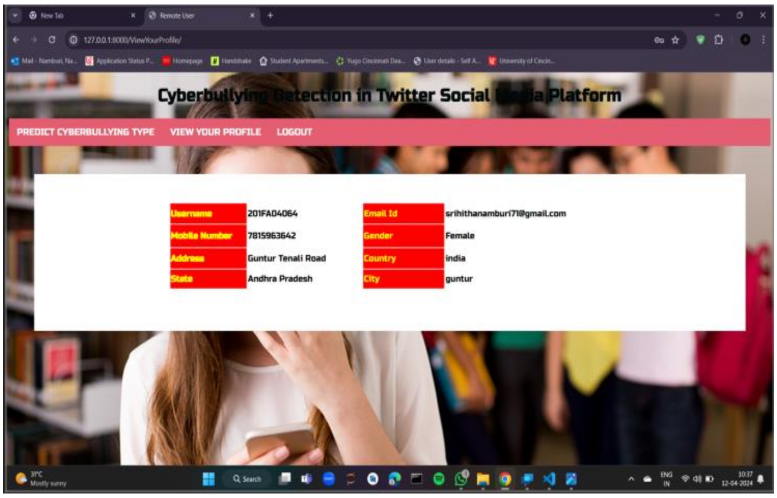

Figure 13: The user login URL page allows you to log in to the system and submit or see posts.

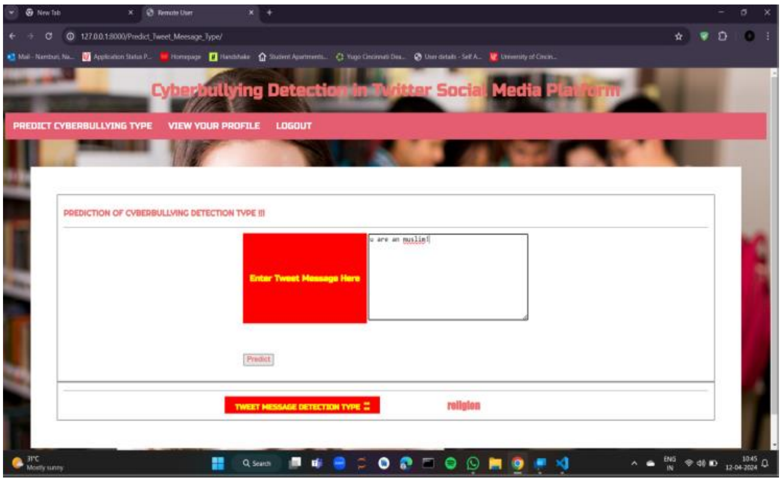

Figure 14: Represents the user interface's "Send post" page or module. This page allows users to compose and send messages. This page contains key aspects such as:

Text entry field: Users can enter their message into this box.
Predict Button: A button that enables people to submit their posts.

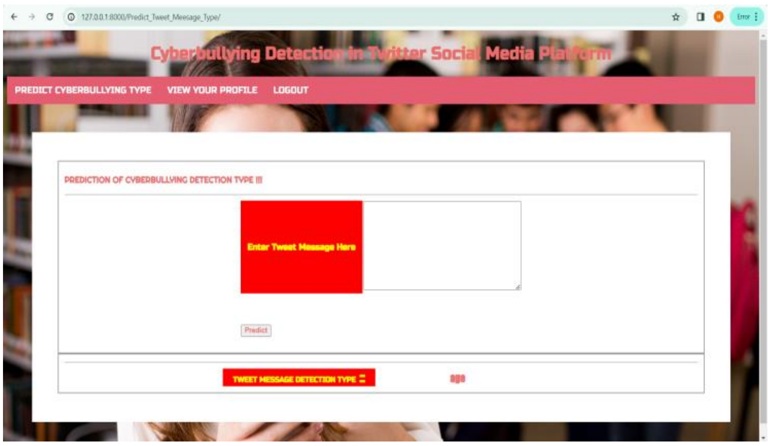

Figure 15: The proposed method may anticipate the message as cyber, non-cyber harassers, ethnicity, age, and gender using machine learning. Machine learning algorithms are used to predict if a phrase is harasser or non-harasser based on dataset records.

DOI: 10.17148/IJARCCE.2025.14405

VI. CONCLUSION

In response to the serious issue of cyberbullying and online safety, our work proposes a unique technique to improve cyberbullying detection on social media platforms. Using machine learning approaches, we created a strong tweet classification model that designed to increase effectiveness. Using topic models to detect incidents of cyberbullying. Our technique combines a wide range of machine learning algorithms, methodically fine-tuning parameters to obtain best classification performance.

We ran lengthy tests on a newly curated Twitter dataset that was rigorously collected using cyberbullying-related keywords. Our model's performance was thoroughly evaluated against recognized approaches such as Support Vector Machines (SVM), Random Forests (RF), Multinomial Naive Bayes (MNB), and others often used in cyberbullying detection tasks. The experimental results unambiguously illustrate the superiority of our machine learning-based technique in detecting cyberbullying episodes across a wide range of assessment measures, including accuracy, recall, F-measure, precision, and specificity.

Furthermore, our research goes beyond the identification of cyberbullying on Twitter, recognizing the importance of include additional social media sites such as Instagram, Flickr, YouTube, and Facebook. Such a broad approach is required to acquire a thorough grasp of cyberbullying patterns across various online groups. Furthermore, our future study will focus on combining numerous data sources to improve cyberbullying detection skills, widening the analysis to include user behavior patterns.

In addition to identifying cyberbullying through textual content, our study underlines the relevance of demographic parameters such as age, ethnicity, gender, and religion. By combining these characteristics, we want to give a more comprehensive understanding of cyberbullying dynamics and its effects on various demographic groups. Furthermore, we intend to investigate real-time cyberbullying detection with our machine learning-based technique, therefore contributing to the construction of safer online environments for all users.

Finally, our findings highlight the effectiveness of machine learning in fighting cyberbullying and call for more research in this area to enhance online safety and well-being. By tackling the multiple facets of cyberbullying and employing innovative analytical approaches, we want to create a more inclusive and courteous online environment for people of all ages and backgrounds.

VII. FUTURE WORK

Although the program is now outstanding, there is plenty of space for improvement and advancement. Enhanced user authentication and continual machine learning model optimization for cyberbullying detection through the examination of various methodologies, feature engineering techniques, and hyperparameter tuning are just a few.

REFERENCES

[1]. Wachs S, Wolf KD, Pan C. Cyber grooming: Risk factors, coping strategies and associations with cyberbullying. Psicothema 2012;24(4):628

[2]. F. Mishna, M. Khoury-Kassabri, T. Gadalla, and J. Daciuk, ‘‘Risk factors for involvement in cyber bullying: Victims, bullies and bully–victims,’’ Children Youth Services Rev., vol. 34, no. 1, pp. 63–70, Jan. 2012, doi: 10.1016/j.childyouth.2011.08.032. [2] K. Miller, ‘‘Cyberbullying and its consequences: How cyberbullying is contorting the minds of victims and bullies alike, and the law's limited available redress,’’ Southern California Interdiscipl. Law J., vol. 26, no. 2, p. 379, 2016.

[3]. A. M. Vivolo-Kantor, B. N. Martell, K. M. Holland, and R. Westby, ‘‘A systematic review and content analysis of bullying and cyber-bullying measurement strategies,’’ Aggression Violent Behav., vol. 19, no. 4, pp. 423–434, Jul. 2014, doi: 10.1016/j.avb.2014.06.008.

[4]. H. Sampasa-Kanyinga, P. Roumeliotis, and H. Xu, ‘‘Associations between cyberbullying and school bullying victimization and suicidal ideation, plans and attempts among Canadian schoolchildren,’’ PLoS ONE, vol. 9, no. 7, Jul. 2014, Art. no. e102145, doi: 10.1371/journal.pone.0102145.

[5]. M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, ‘‘Improving cyberbullying detection with user context,’’ in Proc. Eur. Conf. Inf. Retr., in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 7814, 2013, pp. 693–696.

[6]. A. S. Srinath, H. Johnson, G. G. Dagher, and M. Long, ‘‘BullyNet: Unmasking cyberbullies on social networks,’’ IEEE Trans. Computat. Social Syst., vol. 8, no. 2, pp. 332–344, Apr. 2021, doi: 10.1109/TCSS.2021.3049232.

DOI: 10.17148/IJARCCE.2025.14405

[7]. A. Agarwal, A. S. Chivukula, M. H. Bhuyan, T. Jan, B. Narayan, and M. Prasad, ''Identification and classification of cyberbullying posts: A recurrent neural network approach using under-sampling and class weighting,'' in Neural Information Processing (Communications in Computer and Information Science), vol. 1333, H. Yang, K. Pasupa, A. C.-S. Leung, J. T. Kwok, J. H. Chan, and I. King, Eds. Cham, Switzerland: Springer, 2020, pp. 113–120.

[8]. Z. L. Chia, M. Ptaszynski, F. Masui, G. Leliwa, and M. Wroczynski, ''Machine learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection,'' Inf. Process. Manage., vol. 58, no. 4, Jul. 2021, Art. no. 102600, doi: 10.1016/j.ipm.2021.102600.

[9]. N. Yuvaraj, K. Srihari, G. Dhiman, K. Somasundaram, A. Sharma, S. Rajeskannan, M. Soni, G. S. Gaba, M. A. AlZain, and M. Masud, ''Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking,'' Math. Problems Eng., vol. 2021, pp. 1–12, Feb. 2021, doi: 10.1155/2021/6644652.

[10]. B. A. Talpur and D. O'Sullivan, ''Multi-class imbalance in text classification: A feature engineering approach to detect cyberbullying in Twitter,'' Informatics, vol. 7, no. 4, p. 52, Nov. 2020, doi: 10.3390/informatics7040052.

[11]. A. Muneer and S. M. Fati, ''A comparative analysis of machine learning techniques for cyberbullying detection on Twitter,'' Futur. Internet, vol. 12, no. 11, pp. 1–21, 2020, doi: 10.3390/fi12110187.

[12]. R. R. Dalvi, S. B. Chavan, and A. Halbe, ''Detecting a Twitter cyberbullying using machine learning,'' Ann. Romanian Soc. Cell Biol., vol. 25, no. 4, pp. 16307–16315, 2021.

[13]. R. Zhao, A. Zhou, and K. Mao, ''Automatic detection of cyberbullying on social networks based on bullying features,'' in Proc. 17th Int. Conf. Distrib. Comput. Netw., Jan. 2016, pp. 1–6, doi: 10.1145/2833312.2849567.

[14]. L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, ''XBully: Cyberbullying detection within a multi-modal context,'' in Proc. 12th ACM Int. Conf. Web Search Data Mining, Jan. 2019, pp. 339–347, doi: 10.1145/3289600.3291037.

[15]. K. Reynolds, A. Kontostathis, and L. Edwards, ''Using machine learning to detect cyberbullying,'' in Proc. 10th Int. Conf. Mach. Learn. Appl. Workshops (ICMLA), vol. 2, Dec. 2011, pp. 241–244, doi: 10.1109/ICMLA.2011.152.

[16]. S. Agrawal and A. Awekar, ''Deep learning for detecting cyberbullying across multiple social media platforms,'' in Advances in Information Retrieval (Lecture Notes in Computer Science), vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds. Cham, Switzerland: Springer, 2018, pp. 141–153. VOLUME 10, 2022 25869 IEEE Access and Machine Learning,Volume:10,Issue Date:March.11,2022.

[17]. R. I. Rafiq, H. Hosseinmardi, R. Han, Q. Lv, S. Mishra, and S. A. Mattson, ''Careful what you share in six seconds: Detecting cyberbullying instances in vine,'' in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2015, pp. 617–622, doi: 10.1145/2808797.2809381.

[18]. N. Yuvaraj, V. Chang, B. Gobinathan, A. Pinagapani, S. Kannan, G. Dhiman, and A. R. Rajan, ''Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification,'' Comput. Electr. Eng., vol. 92, Jun. 2021, Art. no. 107186, doi: 10.1016/j.compeleceng.2021.107186.

[19]. A. Al-Hassan and H. Al-Dossari, ''Detection of hate speech in Arabic tweets using deep learning,'' Multimedia Syst., Jan. 2021, doi: 10.1007/s00530-020-00742-w.

[20]. Y. Fang, S. Yang, B. Zhao, and C. Huang, ''Cyberbullying detection in social networks using bi-GRU with self-attention mechanism,'' Information, vol. 12, no. 4, p. 171, Apr. 2021, doi: 10.3390/info12040171.