NLP-INTEGRATED MACHINE LEARNING APPROACH TO PROTECT CHILDREN AND COMBAT CYBER HARASSMENT ON SOCIAL MEDIA

K. Mahesh[1*], Ayesha Nida[1], Seema Fatima[1], Pruthvika Reddy[1]

[1]Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, Telangana, India

ABSTRACT

Social media has become an integral part of our daily lives, facilitating connections between people across the globe. These platforms have brought about numerous advantages, but they have also exposed vulnerable individuals, particularly children, to online risks. Individuals who engage in harmful behavior, such as child predators and cyber harassers, exploit the anonymity and wide reach of social media platforms to cause harm to others. In the past, these risks were addressed through manual reporting and the involvement of human moderators. Users reported suspicious activity, and human moderators reviewed the content to ensure it complied with platform guidelines. Unfortunately, this reactive approach frequently resulted in delayed action, which allowed harmful content to spread. Researchers have embraced machine learning, a powerful form of artificial intelligence that enables computers to learn from data and make accurate predictions, in order to develop more proactive and streamlined solutions. Our objective is to develop a sophisticated automated system capable of efficiently detecting online child predators and cyber harassers through advanced machine learning techniques. The proposed machine learning-based approach offers several benefits compared to current methods. By significantly reducing response time, platforms can swiftly eliminate harmful information and individuals. Machine learning algorithms have the ability to uncover patterns and relationships in large data sets that human moderators might overlook, resulting in enhanced detection accuracy. By incorporating machine learning into social media moderation, human moderators can devote their attention to more complex tasks that require judgment and action. This enhances the efficiency of content moderation and alleviates the workload of moderators. Utilizing machine learning to detect and prevent online child predators and cyber harassers is a significant advancement in enhancing online safety. Machine learning technologies play a crucial role in combating online abuse by being proactive and accurate in mitigating social media threats posed by harmful users.

Keywords: Online safety, Cyber harassment, Child predator detection, Social media moderation, Automated content monitoring, Machine learning.

1. INTRODUCTION

In today's digital age, social media platforms have transformed the way we communicate and connect with others. These platforms have undoubtedly brought about numerous opportunities for global interaction, but they have also given rise to significant challenges [1]. Among these challenges are the presence of online child predators and cyber harassers, individuals who exploit the anonymity and reach of the internet for harmful purposes. Safeguarding individuals, particularly children, from these online threats has become a pressing concern [2]. The history of addressing online threats such as child predators and cyber harasser's dates back to the early days of the internet when these issues first emerged. Over the years, various efforts have been made to combat these threats. These efforts have included legal measures, educational campaigns aimed at users, and the development of technology-based solutions [3]. As technology advanced, particularly in the fields of machine learning and data analysis, new possibilities emerged for more effective identification and mitigation of these online dangers. The need for an automated system for identifying online child predators and cyber harassers arises from practical considerations:

- Scale: The sheer volume of online content makes manual monitoring unfeasible. Automated solutions are essential for processing and analyzing vast datasets effectively.
- Speed: Threats in the online world can escalate rapidly. Rapid detection and response are critical to preventing harm.
- Complexity: Identifying predatory or harassing behavior often involves analyzing text, images, and user behavior patterns. Machine learning and data analysis techniques can significantly enhance this process.
- Accuracy: Reducing false positives and false negatives is paramount. Striking the right balance ensures that innocent users are not wrongly targeted, and potential threats are not overlooked.

2. LITERATURE SURVEY

Online harassment has been a pervasive issue since the early days of social media, and it still exists today. These studies began with an attempt to develop an automated system to detect and report this type of misconduct. Studies have been started to reduce or detect cases of sexual harassment and protect children from bullying to create a safe environment using two approaches: machine learning and deep learning. In this study [4], the authors tracked the occurrence of cyberbullying on social media, using fuzzy logic and genetic algorithms. They identified and classified cyberbullying words and activities on social media, including inflammatory, harassing, racist, and terroristic comments. The resulting F-measure was 0.91. A genetic algorithm is used to optimize parameters and achieve correct performance.

In Facebook message filtering, the authors in ref [5] used three weighting schemes, namely entropy, term frequency-inverse document frequency (TFIDF), and modified TF-IDF was used as feature selection. A Support Vector Machine (SVM) was used to measure the accuracy, precision, and recall of a support vector. The test results indicated that the modified TF-IDF, with an accuracy of 96.50%, outperforms the other schemes. This work in [6] was carried out to compare supervised machine learning (ML) algorithms for natural language learning and assessment of online harassment in Twitter messages as part of social media competition and harassment (a feature). The feature extraction was done by TF-IDF and Word2Vec embeddings. The results accurately included over (80%) of all types of harassment considered within the data. This study [7] combines feeling analysis with a modern approach to sentencing vectors. The language pattern of Long-Short-Term-Memory, Recurrent Neural Network (LSTM_RNN) is used as a new means of predators Sexual Identification to produce word vectors. A record-breaking accuracy rate has been achieved in the last phase of determining the feeling value from the SoftMax layer outputs, with a recall of 81.10%.

The authors in ref. [8] use convolution neural networks (CNN) for extracting features from tags and creating a Twitter post classification model for malicious intent. They analyzed a four-month Twitter dataset to examine the narrative contexts that expressed malicious intent. They talked about the significance of such cases in developing gender-based violence regulations. Sweta Karlekar in [9] presents the work of the SafeCity Web Community to categorize and analyze different types of sexual harassment. SafeCity Web uses this information to help victims compile web directories, provide more detailed safety advice services by sharing their accounts, and help individuals uncover relevant cases to prevent further sexual violence. The single-label CNN-RNN model achieves 86.5% accuracy in processing, linking, and annotating tags. Espinoza [10] uses Twitter for a new data set with four classes of harassment detection. They applied two various deep learning architectures (CNN and LSTM) to classify the tweets. When training the data, the measurement for F1 was equal to 55 percent, while the results for the test set alone hit 46 percent for F1.

Arijit Josh Chowdhury [11] proposes a language model for disclosure. The ULMFiT fine-tuning architecture consists of a language model, a specific mediator (Twitter), and a task-specific classifier. The complete comparison shows the benefits of using particular and lightweight LSTM-based mean language models and an improved vocabulary reflecting linguistic nuances in the deep text that challenges sexual harassment. The neural network models with high results demonstrated about 10,000 personal sexual harassment stories annotated to extract core elements and automatically categorize the stories. Additional categorization improvements were accomplished through the details of key features with an accuracy of 92.9%. This author in [12] provides an automatic method for analyzing the text of online communication and determining if a cyber-predator's prey is another user or one of the participants. Classification tasks provide a RNN (recurrent neural network) in each stage, which achieved an F0.5 score of 0.9058. In [13], the authors provide a novel way to classify sexual harassment and sexual predators in English using BiLSTM with Gated Recurrent Unit (GRU) algorithms and word embedding. The pre-trained Global Vectors (GloVe) words achieved 97.27%, compared with Extreme Gradient Boosting (XGBoost), which reached 90.10%.

3. PROPOSED METHODOLOGY

This project implements a web application using the Django framework. The application appears to be related to the identification of online child predators and cyber harassers in social media environments. In addition, this project represents the backend logic of a web application designed for identifying and monitoring online child predators and cyber harassers in social media. Users can register, log in, send posts, and run machine learning algorithms to classify text messages as potentially harmful or not. The results are presented on web pages for users and administrators to monitor.

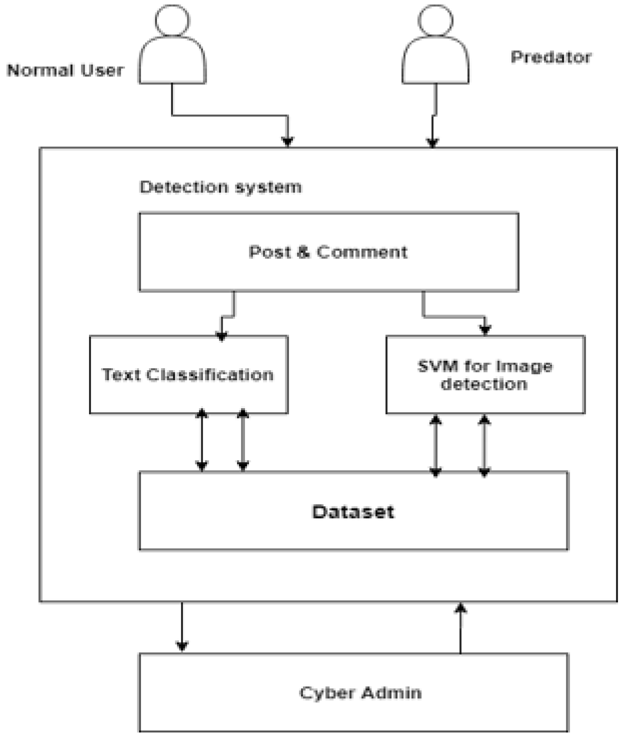


Figure 1: Proposed system architecture.

TF-IDF Feature Extraction

TF-IDF, short for Term Frequency-Inverse Document Frequency, is a commonly used technique in NLP to determine the significance of words in a document or corpus. To give some background context, a survey conducted in 2015 showed that 83% of text-based recommender systems in digital libraries use TF-IDF for extracting textual features. That's how popular the technique is. Essentially, it measures the importance of a word by comparing its frequency within a specific document with the frequency to its frequency in the entire corpus. The underlying assumption is that a word that occurs more frequently within a document but rarely in the corpus is particularly important in that document.

TF (Term Frequency) is determined by calculating the frequency of a word in a document and dividing it by the total number of words in the document.

- TF = (Number of times the word appears in the document) / (Total number of words in the document)
- IDF (Inverse Document Frequency), on the other hand, measures the importance of a word within the corpus as a whole. It is calculated as:
- IDF = log((Total number of documents in the corpus) / (Number of documents containing the word))

The tf-idf value increases in proportion to the number of times a word appears in the document but is often offset by the frequency of the word in the corpus, which helps to adjust with respect to the fact that some words appear more frequently in general. TF-IDF use two statistical methods, first is Term Frequency and the other is Inverse Document Frequency. Term frequency refers to the total number of times a given term t appears in the document doc against (per) the total number of all words in the document and The inverse document frequency measure of how much information the word provides. It measures the weight of a given word in the entire document. IDF show how common or rare a given word is across all documents. TF-IDF can be computed as tf * idf

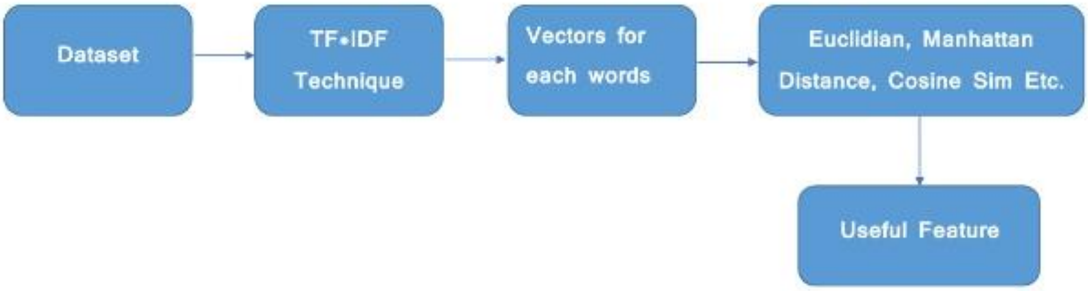


Fig. 2: TF-IDF block diagram.

TF-IDF do not convert directly raw data into useful features. Firstly, it converts raw strings or dataset into vectors and each word has its own vector. Then we'll use a particular technique for retrieving the feature like Cosine Similarity which works on vectors, etc.

Terminology

t — term (word)

d — document (set of words)

N — count of corpus

corpus — the total document set

Step 1: Term Frequency (TF): Suppose we have a set of English text documents and wish to rank which document is most relevant to the query, "Data Science is awesome!" A simple way to start out is by eliminating documents that do not contain all three words "Data" is", "Science", and "awesome", but this still leaves many documents. To further distinguish them, we might count the number of times each term occurs in each document; the number of times a term occurs in a document is called its term frequency. The weight of a term that occurs in a document is simply proportional to the term frequency.

$$tf(t,d) \ = \ count\ of\ t\ in\ d\ /\ number\ of\ words\ in\ d$$

Step 2: Document Frequency: This measures the importance of document in whole set of corpora, this is very similar to TF. The only difference is that TF is frequency counter for a term t in document d, whereas DF is the count of occurrences of term t in the document set N. In other words, DF is the number of documents in which the word is present. We consider one occurrence if the term consists in the document at least once, we do not need to know the number of times the term is present.

$$df(t) \ = \ occurrence\ of\ t\ in\ documents$$

Step 3: Inverse Document Frequency (IDF): While computing TF, all terms are considered equally important. However, it is known that certain terms, such as “is”, “of”, and “that”, may appear a lot of times but have little importance. Thus, we need to weigh down the frequent terms while scale up the rare ones, by computing IDF, an inverse document frequency factor is incorporated which diminishes the weight of terms that occur very frequently in the document set and increases the weight of terms that occur rarely. The IDF is the inverse of the document frequency which measures the informativeness of term t. When we calculate IDF, it will be very low for the most occurring words such as stop words (because stop words such as “is” is present in almost all of the documents, and N/df will give a very low value to that word). This finally gives what we want, a relative weightage.

$$idf(t) \ = \ N/df$$

Now there are few other problems with the IDF, in case of a large corpus, say 100,000,000, the IDF value explodes, to avoid the effect we take the log of idf . During the query time, when a word which is not in vocab occurs, the df will be 0. As we cannot divide by 0, we smoothen the value by adding 1 to the denominator.

$$idf(t) \ = \ log(N/(df \ + \ 1))$$

The TF-IDF now is at the right measure to evaluate how important a word is to a document in a collection or corpus. Here are many different variations of TF-IDF but for now let us concentrate on this basic version.

$$tf - idf(t,d) \ = \ tf(t,d) \ * \ log(N/(df \ + \ 1))$$

Step 4: Implementing TF-IDF: To make TF-IDF from scratch in python, let’s imagine those two sentences from different document:

first sentence: “Data Science is the sexiest job of the 21st century”.

second sentence: “machine learning is the key for data science”.

4. RESULTS AND DISCUSSION

This project implements a web application that combines Django web development with machine learning techniques to detect and monitor cyberbullying messages. It involves user registration, login, data collection, model training, and post classification. Below are the interactions between various components:

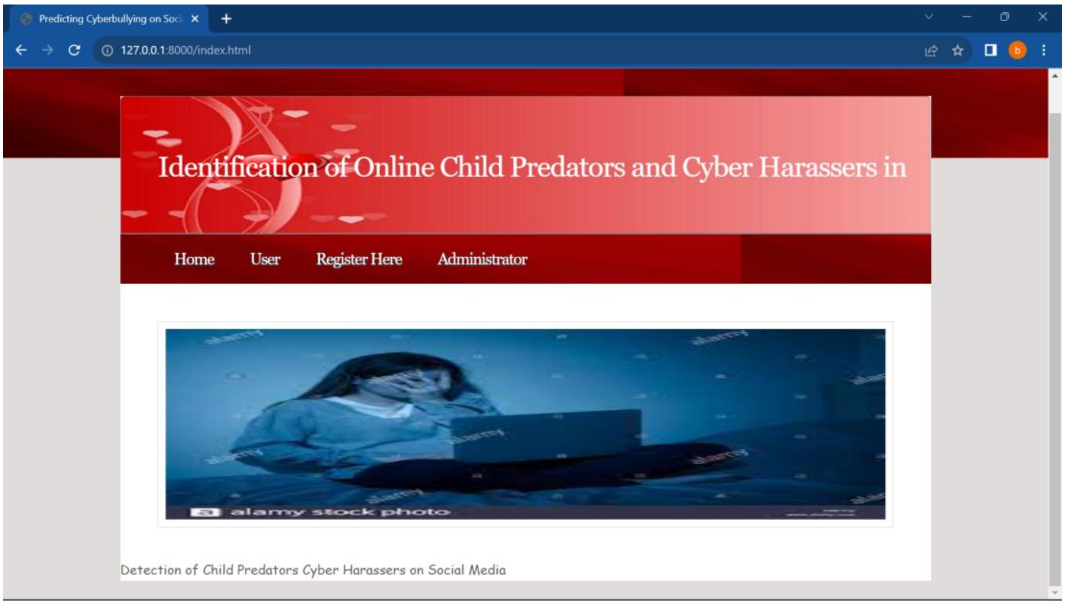
- User Interface: This section represents the user interface where users interact with the application. Users can access various pages such as the home page (Index), Send Post, Register, and Login.

- Database: The database section represents the MySQL database and file system used to store data and files. Users' requests, including database queries and file storage, are shown here.
- Admin Panel: Admins can access the Admin page, Add Bullying Words functionality, and View Users functionality. These actions also involve requests to the database.
- Machine Learning: This section depicts machine learning algorithms used for classification. The algorithms are trained and predict based on data supplied by the database. Model outputs and predictions are returned to the database.
- User Interaction: This part illustrates user interactions, such as sending posts, logging in (creating sessions), and user registration. These interactions result in data being saved in the database.
- Monitoring and Reporting: Admins can access the Monitor Posts and Run Algorithms functionalities, which involve database requests. The database supplies data to these functionalities, and they may generate reports or monitor posts.

Results description

Figure 3 represents the main landing page of the web application. It's called the "index" page, and it contains modules for different functionalities related to the identification of online child predators and cyber harassers in a social media environment. The modules include:

- Home: A link to the homepage or main dashboard.
- User: A link for users to log in or access their user-specific functionalities.
- Register: A link for new users to create an account.
- Administrator: A link for administrators to login and access admin-specific functionalities.

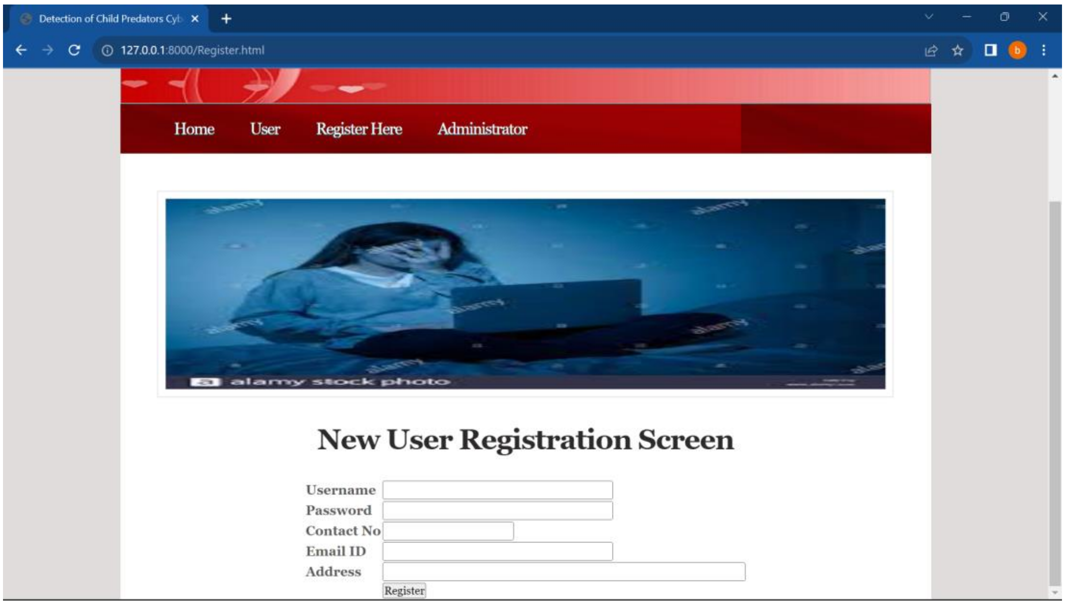


Identification of Online Child Predators and Cyber Harassers in

Detection of Child Predators Cyber Harassers on Social Media

Figure 3: Illustration of index URL page with home, user, register, and administrator modules of identification of online child predators and cyber harassers in social media environment.

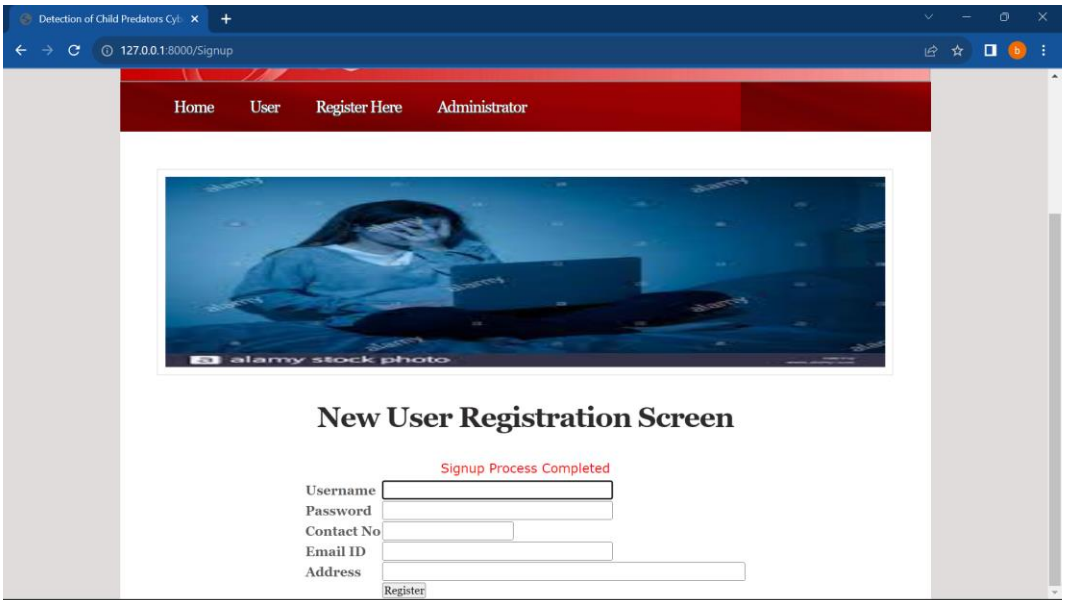
Figure 4 represents the registration page of the application. Users can access this page by clicking on the "Register" link from the main page. On this page, new users can sign up for an account by providing their registration details, such as username, password, contact information, email, and address. Figure 5 shows a confirmation page that appears after a user successfully registers for an account. It displays a message confirming that the registration process has been completed. Figure 6 represents the login page for administrators. Administrators can access this page by clicking on the "Administrator" link from the main page. They need to provide their login credentials, typically with a default username and password (username as 'admin' and password as 'admin') to gain access to admin-specific features and functionalities.


Detection of Child Predators Cyb
127.0.0.1:8000/Register.html
Home
User
Register Here
Administrator
alamy stock photo

New User Registration Screen

Username
Password
Contact No
Email ID
Address
Register

Figure 4: Register URL page for new user registration to create an account.



New User Registration Screen

Signup Process Completed

Username
Password
Contact No
Email ID
Address
Register

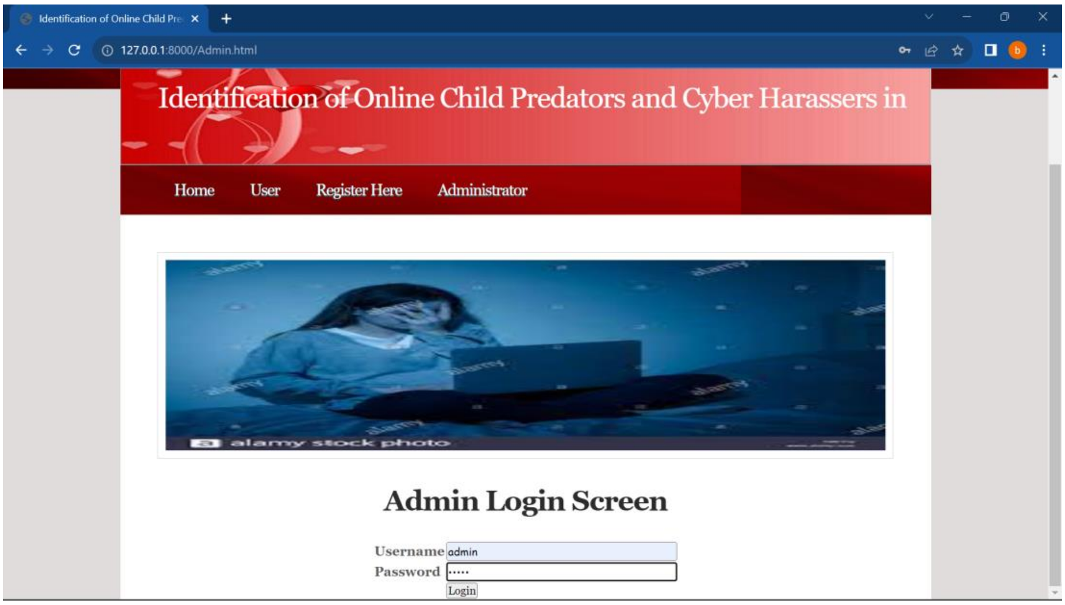
Figure 5: Signup URL after registering the new user account.

After logging in as an administrator, Figure 7 displays various modules and functionalities available to administrators. These modules include:

- View Users: This module allows administrators to view and manage registered users and their details.
- Monitor Post: Administrators can monitor posts made by users, including details like the messages, uploaded files, posting time, and the status of the bullying (e.g., bullying, or cyber harassment).

- Add Bullying Messages/Words: Administrators can add words or messages to the dataset that are considered bullying or related to cyber harassment.
- Run Machine Learning Algorithms: This module enables administrators to run machine learning algorithms, likely for the purpose of analyzing and detecting cyberbullying content.

Figure 8 represents the "View Users" page accessible to administrators. It displays a list of registered users along with their details, which include usernames, passwords, contact information, email addresses, addresses, and user statuses. Figure 9 shows the monitor post page that provides administrators with a view of registered users' posts. It includes information such as the sender's name, uploaded files (possibly images), the messages posted, the posting time, and the status of each post (e.g., categorized as bullying or cyber harassment). Figure 10 represents a page or interface where administrators can add specific words or messages to a dataset. These words are typically considered bullying or related to cyber harassment. This dataset is used for training machine learning models to detect such content.

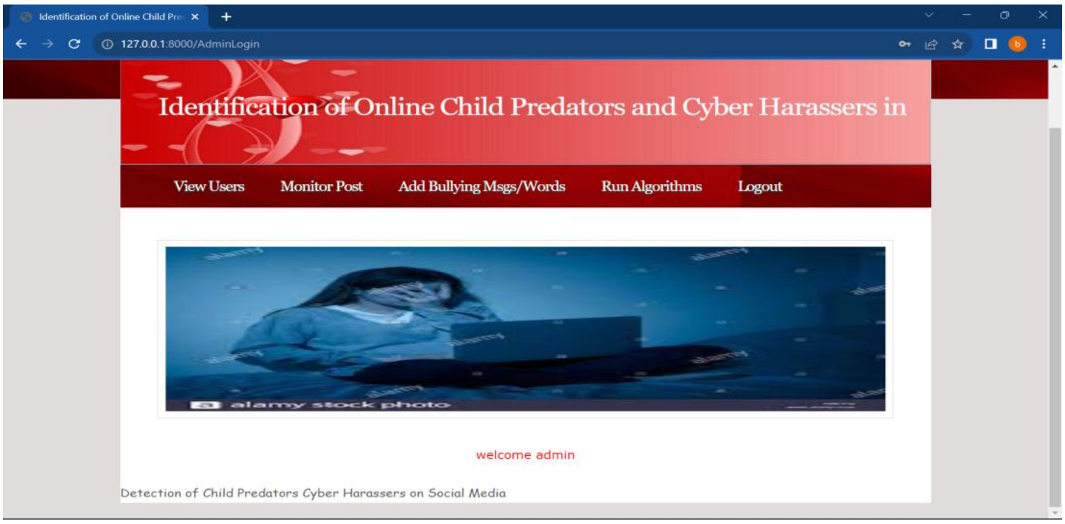


Identification of Online Child Predators and Cyber Harassers in

Admin Login Screen

Username admin
Password ·····

Figure 6: Admin URL page for login as an admin with a username as admin and password as admin.

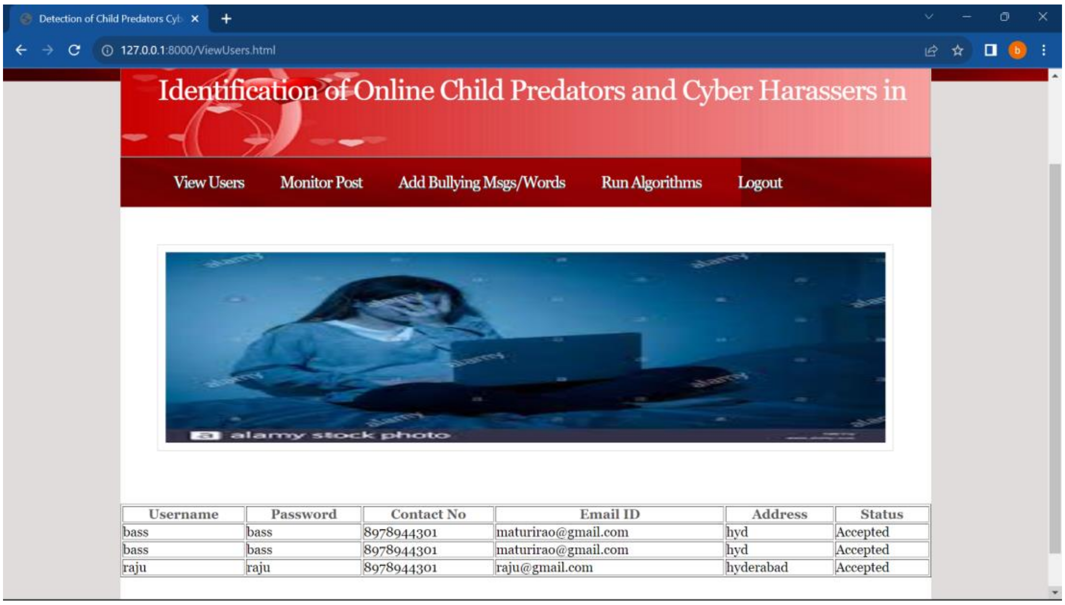


Identification of Online Child Predators and Cyber Harassers in

welcome admin

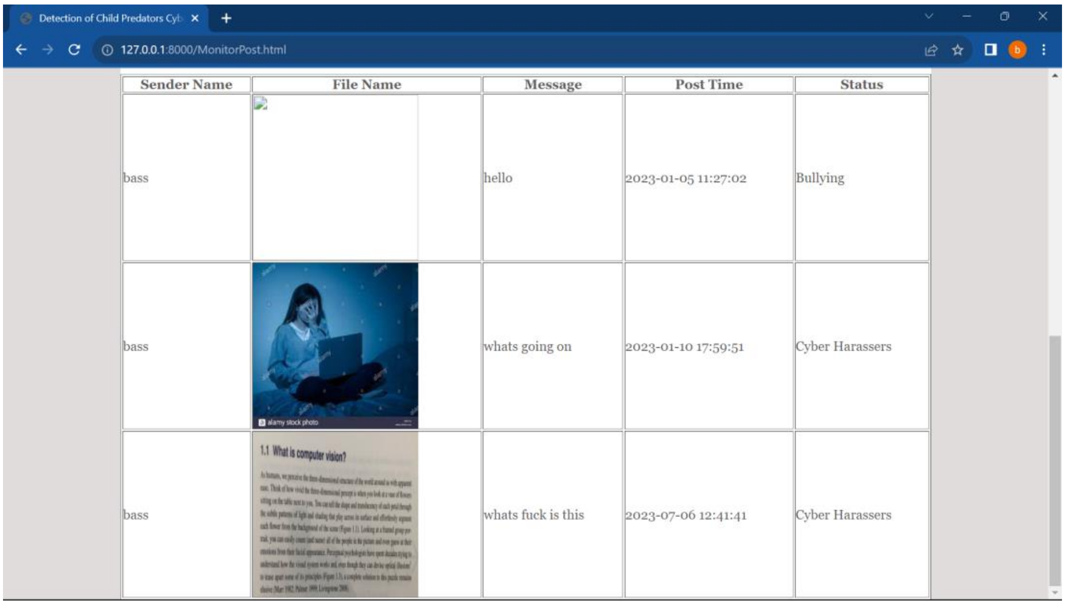
Detection of Child Predators Cyber Harassers on Social Media

Figure 7: Admin login page showing the modules like view users, monitor post, add bullying messages/words, and run machine learning algorithms.


Detection of Child Predators Cyb
127.0.0.1:8000/ViewUsers.html
Identification of Online Child Predators and Cyber Harassers in
View Users
Monitor Post
Add Bullying Msgs/Words
Run Algorithms
Logout
alamy stock photo

Username	Password	Contact No	Email ID	Address	Status
bass	bass	8978944301	maturirao@gmail.com	hyd	Accepted
bass	bass	8978944301	maturirao@gmail.com	hyd	Accepted
raju	raju	8978944301	raju@gmail.com	hyderabad	Accepted

Figure 8: View users page showing the registered users and their details.


Detection of Child Predators Cyb
127.0.0.1:8000/MonitorPost.html
alamy stock photo
1.1 What is computer vision?

Sender Name	File Name	Message	Post Time	Status
bass		hello	2023-01-05 11:27:02	Bullying
bass		whats going on	2023-01-10 17:59:51	Cyber Harassers
bass		whats fuck is this	2023-07-06 12:41:41	Cyber Harassers

Figure 9: Monitor post page showing the registered users with the messages, uploaded files with the posting time and the status of the bullying i.e., bullying, or cyber harassment.

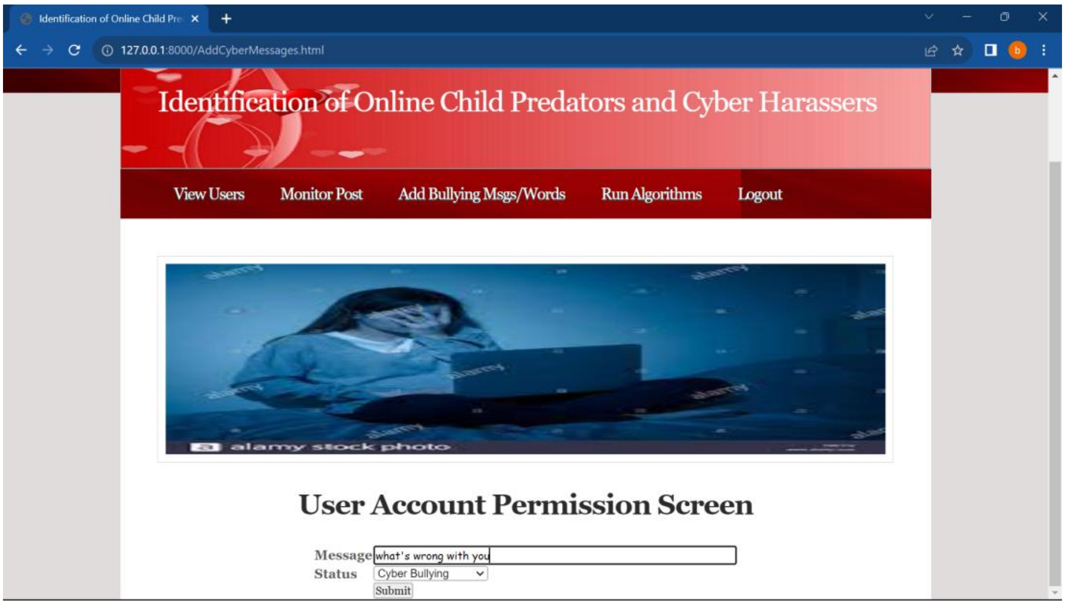
In Figure 11, admin must select each algorithm and click on ‘Submit’ button to train model and the accuracy will be shown for each algorithm. Admin must repeat this step whenever first time he starts the server or upon adding new bulling messages. He must run at least one algorithm to perform automatic detection of harasser’s or non-harassers.

Figure 12 represents the login page for regular users who want to access the system. Users need to provide their login credentials (e.g., username and password) to login and gain access to their account. The main functionalities accessible to users from this page are:

- Sending a post: Users can compose and send a message along with a photo.
- Viewing posts: Users can access a page to view posts made by themselves or others.
- Logout: There may be an option to log out of the system, terminating the current session.

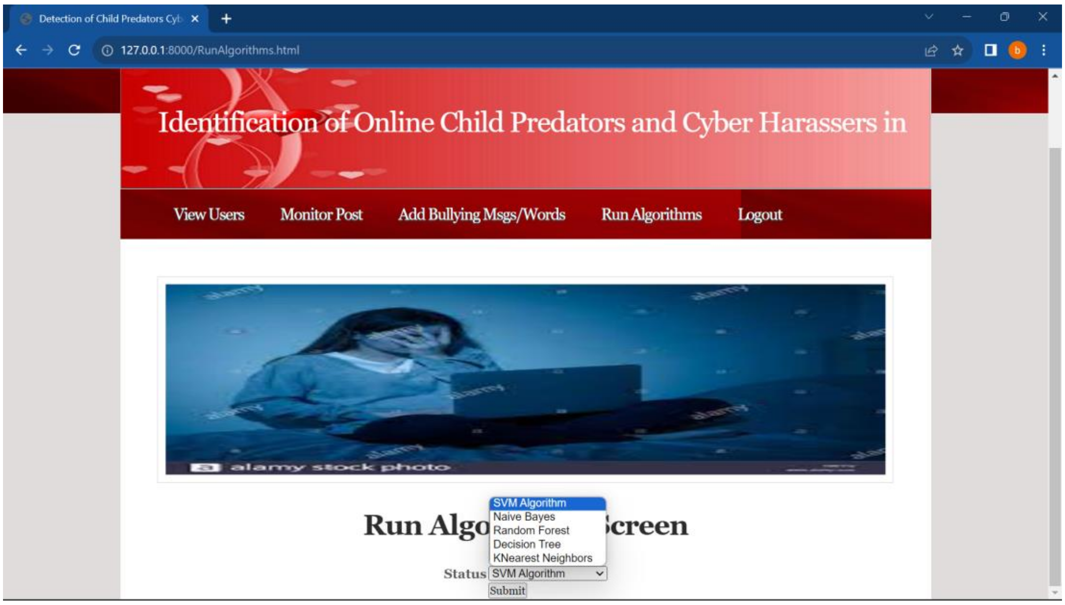
Figure 13 is an illustration of the user login page after successfully logging in. The page presents several modules or options to users:

- Send post: Users can click on this module to create and send a new post. This likely includes composing a message and attaching a photo.
- View post: Users can access a section where they can view posts, which could be their own posts or posts from other users.
- Logout: There may be a logout option to allow users to log out of their account when they're finished with their session.


Identification of Online Child Pre
127.0.0.1:8000/AddCyberMessages.html
Identification of Online Child Predators and Cyber Harassers
View Users
Monitor Post
Add Bullying Msgs/Words
Run Algorithms
Logout
alamy stock photo
User Account Permission Screen
Submit

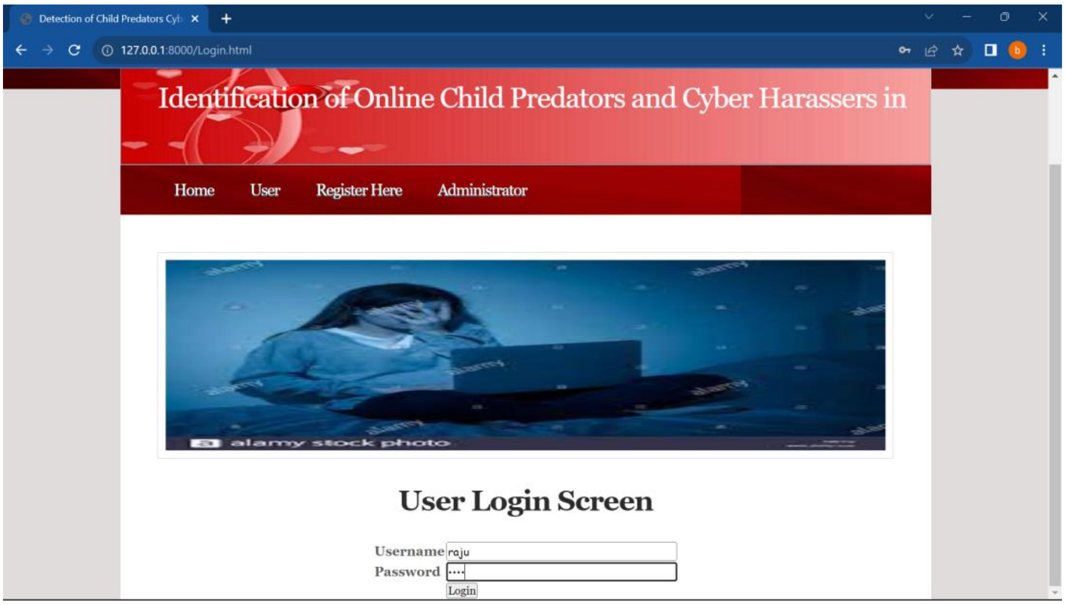
Message: what's wrong with you
Status: Cyber Bullying

Figure 10: Adding the bullying words to the dataset.



Status: SVM Algorithm

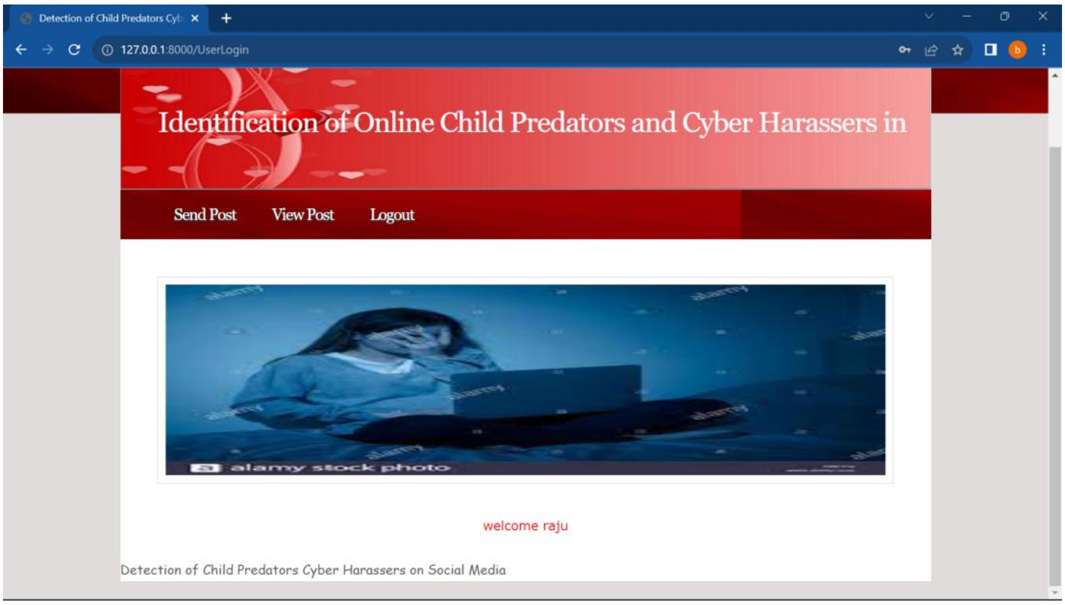
Figure 11: Run machine learning algorithms page showing different ML models i.e., SVM, naïve bayes, random forest, decision tree and KNN classification.


Detection of Child Predators Cyb
127.0.0.1:8000/Login.html
Identification of Online Child Predators and Cyber Harassers in
Home
User
Register Here
Administrator
alamy stock photo
Login

User Login Screen

Username raju
Password ····

Figure 12: User login URL page for logging into system to send a post and view a post.


Detection of Child Predators Cyb
127.0.0.1:8000/UserLogin
Identification of Online Child Predators and Cyber Harassers in
Send Post
View Post
Logout
alamy stock photo

welcome raju

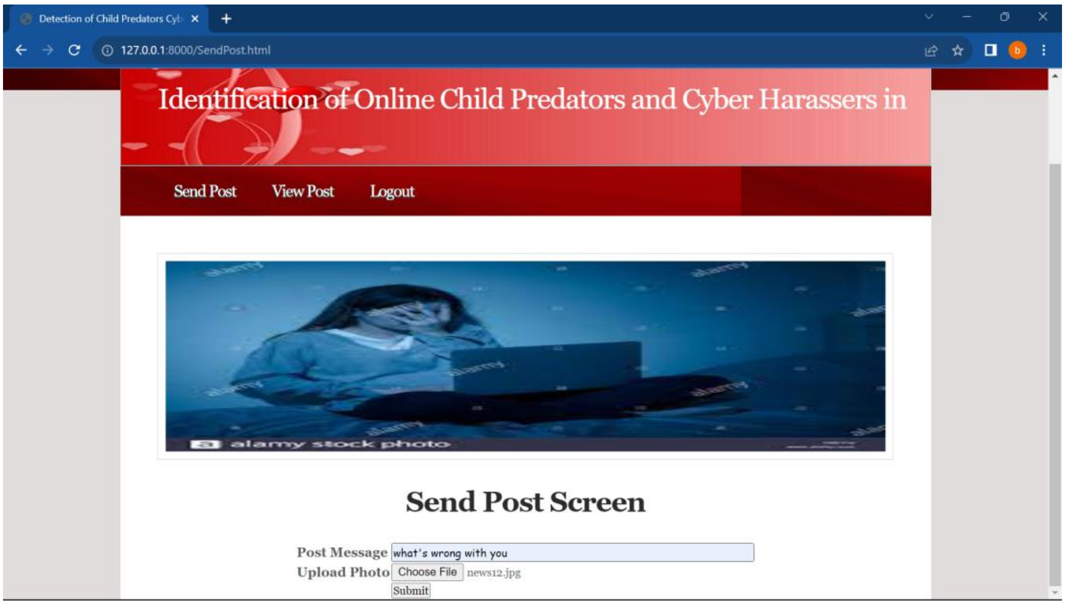
Detection of Child Predators Cyber Harassers on Social Media

Figure 13: User login page showing send post, view post and logout modules.

Figure 14 represents the "Send post" page or module within the user interface. On this page, users can compose and send a message along with a photo. Key elements on this page include:

- Text input field: Users can type their message in this field.
- File upload: Users can upload a photo or image to include in their post.
- Send button: A button that allows users to submit their post.

Figure 15 shows the posts from all users, and it is observed that with the help of machine learning, the proposed system can predict the message as cyber or non-cyber harassers. Here, machine learning models are utilized to predict the harasser or non-harasser word based on the dataset records. So, admin can add all the possible harasser and non-harasser words to dataset by using ‘add words’ module from admin as discussed in earlier steps. After adding words then run algorithms link to train model and then proposed application predicts the harasser or non-harasser automatically.



Detection of Child Predators Cyb

127.0.0.1:8000/SendPost.html

Identification of Online Child Predators and Cyber Harassers in

Send Post View Post Logout

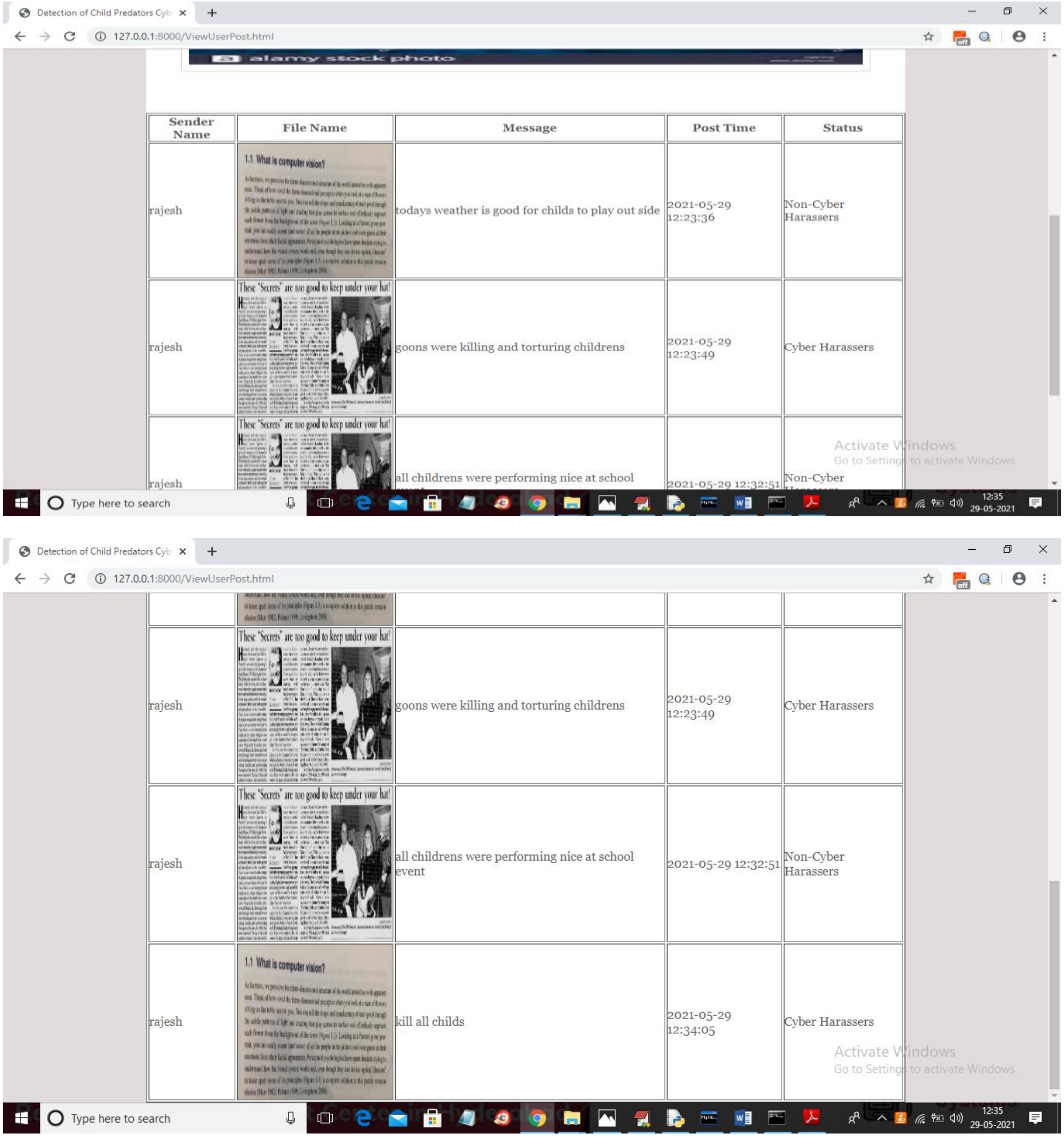
Send Post Screen

Post Message: what's wrong with you

Upload Photo: Choose File news12.jpg

Submit

Figure 14: Send post page for sending a message with a photo.



Detection of Child Predators Cyb

127.0.0.1:8000/ViewUserPost.html

Sender Name	File Name	Message	Post Time	Status
rajesh		todays weather is good for childs to play out side	2021-05-29 12:23:36	Non-Cyber Harassers
rajesh		goons were killing and torturing childrens	2021-05-29 12:23:49	Cyber Harassers
rajesh		all childrens were performing nice at school	2021-05-29 12:32:51	Non-Cyber

127.0.0.1:8000/ViewUserPost.html

rajesh		goons were killing and torturing childrens	2021-05-29 12:23:49	Cyber Harassers
rajesh		all childrens were performing nice at school event	2021-05-29 12:32:51	Non-Cyber Harassers
rajesh		kill all childs	2021-05-29 12:34:05	Cyber Harassers

Figure 15: View post page for viewing all the messages with uploaded photos posted by users.

5. **CONCLUSIONS**

This research presents the backend logic of a web application focused on user management and cyberbullying detection. It embodies a comprehensive approach to tackling issues related to cyberbullying and enhancing user experiences. First and foremost, the application offers robust user management capabilities. Users can register, log in, and view their profiles. User data, including usernames, passwords (which should be further secured using techniques like hashing and salting), contact information, and status, are meticulously stored within a MySQL database. This feature forms the foundation of the user experience and provides the basis for managing user interactions. One of the standout features of the application is its cyberbullying detection functionality. It leverages a diverse set of machine learning algorithms, including SVM, Decision Trees, KNN, Random Forest, and Naive Bayes, to classify user-submitted text messages or "posts" as either "Cyber Harassers" or "Non-Cyber Harassers." The application extracts relevant features from these text messages and maintains them in a dataset for training and inference. The application enables users to submit posts that encompass sender names, filenames (for attached images), messages, timestamps, and statuses. These posts are meticulously recorded within a database, forming a comprehensive record of user interactions. In terms of user interaction, the application offers a suite of user-friendly web pages. Users can register, log in, view profiles, add words associated with cyberbullying to the dataset, submit posts, run machine learning algorithms for detection, and monitor posts. This intuitive and accessible user interface facilitates smooth user engagement.

REFERENCES

[1] Wachs S, Wolf KD, Pan C. Cyber grooming: Risk factors, coping strategies and associations with cyberbullying. Psicothema 2012;24(4):628–33.

[2] KELLER, N.B.a.M.H. video games and online chats are ‘‘hunting grounds” for sexual predators. Available from https://www.nytimes.com/interactive/2019/ 12/07/us/video-games-child-sex-abuse.html [Accessed DEC. 7, 2019]. [3] Amer, N. Arabic-sexual-harassment-dataset. Available from https:// github.com/Nooramer8/Arabic-sexual-harassment-dataset. [Accessed 09-10- 2023].

[4] Nandhini BS, Sheeba J. Online social network bullying detection using intelligence techniques. Proc Comput Sci 2015;45:485–92. doi: https://doi. org/10.1016/j.procs.2015.03.085.

[5] Al-Katheri ASA, Siraj MM. Classification of sexual harassment on Facebook using term weighting schemes. Internat J Innov Comput 2018;8(1):15–9. doi: https://doi.org/10.11113/ijic.v8n1.157.

[6] M.Saeidi, S.Sousa, E.Milios, N.Zeh, L.Berton. Categorizing online harassment on Twitter. in Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2019, 3, 283- 297. https://doi.org/10.1007/978-3-030 43887-6_22.

[7] Liu, D., C.Y. Suen, and O. Ormandjieva. A novel way of identifying cyber predators. 2017, 1712.03903,1-6. https://doi.org/10.48550/arXiv.1712.03903

[8] Pandey, R., et al. Distributional semantics approach to detect intent in Twitter conversations on sexual assaults. in 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). 2018, 1 270-277. https://doi.org/ 10.1109/wi.2018.00-80.

[9] S.Karlekar, and M. Bansal.. Safecity: Understanding diverse forms of sexual harassment personal stories, arXiv preprint arXiv. 2018, 2,1-7. https://doi.org/ 10.18653/v1/d18-1303.

[10] Espinoza I, Weiss F. Detection of harassment on Twitter with deep learning techniques. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2019;1168:307–13. doi: https://doi.org/10.1007/978- 3-030-43887-6.

[11] Liu Y et al. Sexual harassment story classification and key information identification. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. p. 2385–8. https://doi.org/10.1145/ 3357384.3358146.

[12] Kim, J., et al. Analysis of online conversations to detect cyberpredators using recurrent neural networks. in Proceedings for the First International Workshop on Social Threats in Online Conversations: Understanding and Management. 2020,1,15-20. https://www.aclweb.org/anthology/2020.stoc-1.3.

[13] Hamzah NA, Dhannoon BN. The detection of sexual harassment and chat predators using artificial neural network. Karbala Int J Mod Sci 2021;7 (4):6–20.