JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SafeNet: AI Guardian for Detecting and Preventing Child Predators on Social Media

Vanthala Deeven Pradeep, M.Tech (CST) Student, Department of CSE, Baba Institute of Technology and Sciences, Andhra Pradesh, India

Mr.S Durga Prasad.Associate Professor, Department of CSE, Baba Institute of Technology and Sciences, Andhra Pradesh, India

Abstract - This project aims to make social media safer for kids by spotting and stopping online predators and bullies. It uses smart computer programs to learn from different kinds of messages and figure out if a new message might be harmful. There are two main parts to the project: one for regular users and one for admins. Regular users can sign up, log in, and post messages. The system checks these messages using special algorithms like SVM, Random Forest, and others to see if they're safe or not. Admins have extra powers: they manage user accounts, add new examples of harmful and harmless messages to help the system learn better, and run the algorithms to improve its detection skills. The system is built using Django, a tool for making websites, so anyone can use it just by opening it in their web browser. By catching bad stuff before it spreads, this project helps keep online spaces safer for kids.

1. INTRODUCTION

Social media has changed the way we communicate, especially for young people. But it's not all good news. The internet also has some big dangers, like sexual harassment and predators who target kids. Because the internet lets people stay anonymous and talk to lots of people, it's easy for predators to find and hurt kids. Studies show that about one out of every five young people gets asked for sex online every year. That's scary! So, this project is all about making a system that can find these predators on social media and tell the right people, like cyber cops, right away.

We're serious about stopping this problem. We want to look at comments and posts on sites like Facebook and Instagram to catch predators before they can hurt anyone. Lots of kids have had creepy messages or been asked for sex online, and we need to do something about it.

To do this, we need a really good database that can keep track of what's happening online and help us figure out who the bad guys are. We're using smart technology and ideas from other projects to make sure our system works well. Since so many teens use social media, we've got to act fast to keep them safe. By building this predator-catching system, we're giving the people who protect kids the tools they need to stop online harassment and keep young people safe online.

Purpose

This project wants to make social media safer for kids by stopping online bullies and predators. It's using fancy computer programs like SVM, Random Forest, and others to read what people write on social media and find the bad guys.

There are two parts to the project. Regular people can sign up, log in, and use the website. But there are also admins who can watch over everything and add new examples of good and bad messages to help the computer learn better.

The main goal is to help admins keep an eye on what's happening online and stop bad stuff before it hurts anyone. By teaching the computer what's okay and what's not, the system can automatically find bad posts and warn people about them.

This project wants to make the internet safer for kids and everyone else. By making the computer smarter and better at spotting bad stuff, we're working towards a safer online world for everyone.

2. LITERATURE SURVEY

Online sexual harassment and the threat posed by child predators in digital spaces have garnered significant attention

from researchers, policymakers, and psychologists. A comprehensive understanding of these issues is crucial for devising effective strategies to protect young people from harm. The following literature review provides insights into the existing research and findings relevant to the detection and prevention of online sexual harassment and predatory behavior:

National Surveys on Online Sexual Solicitation: Finkelhor, Mitchell, and Wolak conducted national surveys revealing alarming statistics regarding online sexual solicitation among youth. These studies indicated that approximately one in five young people are solicited for sex over the internet annually (Finkelhor, Mitchell, & Wolak, 2000; Mitchell, Finkelhor, & Wolak, 2001). These findings underscore the urgent need for proactive measures to address the prevalence of online sexual predation.

The Impact of Social Media on Child Safety: The National Society for the Prevention of Cruelty to Children (NSPCC) highlighted the risks associated with social media usage among young people. Reports indicate that a significant percentage of adolescents have encountered unwanted sexual messages or solicitations online (NSPCC, 2014). This underscores the importance of implementing safeguards to protect children from online exploitation.

Machine Learning for Predatory Behavior Detection: Recent advancements in machine learning algorithms have enabled researchers to develop sophisticated systems for detecting predatory behavior online. Techniques such as Support Vector Machines (SVM), Random Forest, Naïve Bayes, KNearest Neighbours, and Decision Tree have been employed to analyze social media data and identify potential child predators and cyber harassers (Jain, 2019).

Building Comprehensive Detection Systems: Research efforts have focused on building comprehensive detection systems capable of analyzing comments and posts on social media platforms to identify harmful content. By training models with both normal and harasser's messages, these systems can automatically classify new posts and alert administrators to potential threats (Chen et al., 2020).

The Role of Administrators in Monitoring Online Interactions: Administrators play a crucial role in monitoring online interactions and enforcing safety protocols on social media platforms. Admin modules in detection systems enable administrators to view user accounts, monitor posts, and intervene in cases of suspected predatory behavior (Smith & Jones, 2018).

3. SYSTEM ANALYSIS:

The Paper involves the development of a comprehensive system aimed at identifying and addressing online sexual harassment and predatory behavior targeting young individuals. The system employs a combination of machine learning algorithms and web application modules to achieve its objectives.

Machine Learning Algorithms: The system leverages various machine learning algorithms, including SVM, Random Forest, Naïve Bayes, KNearest Neighbours, and Decision Tree. These algorithms are utilized to analyze social media posts and comments, distinguishing between normal content and potentially harmful content indicative of child predators or cyber harassers.

Data Preparation and Model Training: Before deploying the system, a robust dataset is prepared containing samples of both normal and harassing messages. This dataset is used to train the machine learning models, enabling them to accurately classify new posts. Additionally, administrators have the capability to augment the dataset by adding new harassing or non-harassing messages as needed.

User Module: The user module allows individuals to create accounts, log in, and engage in social media activities such as sending and viewing posts. Users interact with the system through a user-friendly interface, facilitating seamless communication and content sharing.

Admin Module: The admin module provides administrators with the tools to manage user accounts, monitor user activity, and oversee the detection process. Admins are responsible for adding new harassing or non-harassing messages to the dataset, running machine learning algorithms, and reviewing the accuracy of the detection results.

System Deployment and Operation: The system is deployed as a web application using the Django framework. Users access the system through a web browser by entering the appropriate URL. Upon deployment, administrators ensure that the system is properly configured, including setting up the database and adjusting system parameters as needed.

Continuous Improvement and Maintenance: To ensure the effectiveness of the system, continuous monitoring and updates are required. Admins periodically review the performance of the machine learning models, retrain them with new data if necessary, and address any emerging issues or challenges in the detection process.

The system analysis highlights the multifaceted nature of the project, encompassing data processing, machine learning, web application development, and system administration to combat online sexual harassment and protect vulnerable individuals in digital environments.

3.1 EXISTING SYSTEM

Right now, there aren't enough good ways to stop online sexual harassment and predatory behavior aimed at kids on social media. Most social media sites wait for users to report bad stuff, then check it themselves. But this only happens after the bad stuff is already out there. And sometimes, victims don't report what's happening, so it doesn't get dealt with. Looking through reports takes time and might not catch everything because it's hard to keep up with how sneaky predators can be. Plus, just searching for certain words might not catch all the different ways predators try to trick kids.

3.2 PROPOSED SYSTEM

The new plan is to make a system that's better at finding and stopping bad stuff happening to kids online. It uses smart computer programs to figure out when someone might be trying to hurt a kid or bother them online. These programs look at lots of social media stuff really quickly and can tell when something's not right.

The system also has parts for regular users and people in charge. Regular users can make accounts, share stuff, and talk to others on the site. But the people in charge can keep an eye on everything, add new examples of bad stuff to help the computer learn, and make sure it's working well.

The goal is to be ahead of the bad stuff by catching it early and helping keep kids safe online. With time, the system learns more and gets better at stopping bad things from happening.

3.3 IMPLEMENTATION

Implementation Modules:

User Module: This module allows users to create accounts, log in, send posts, and view posts. Users can register by providing necessary details such as username, email, and password. Upon registration, users can log in using their credentials and access the platform's features. They can create new posts, upload images, and interact with other users' posts.

Admin Module: The admin module provides functionalities for managing user accounts, monitoring posts, and maintaining the system's integrity. Administrators can view all registered user accounts and decide whether to accept or reject new user registrations. They are responsible for adding new harassing/non-harassing messages to the machine learning training dataset. Admins can monitor all posts sent by users and take appropriate actions if any harmful content is detected.

Machine Learning Algorithm Integration: This module integrates various machine learning algorithms such as SVM, Random Forest, Naïve Bayes, KNearest Neighbours, and Decision Tree. These algorithms are utilized to analyze posts and comments on social media accounts to detect potential child predators or cyber harassers. The system builds a training model using normal and harassing messages, which is then applied to new posts to predict whether they contain harmful content.

Database Management: The database management module is responsible for creating and maintaining the system's database. It includes functionalities for storing user account information, posts, admin actions, and machine learning training data. Database operations such as CRUD (Create, Read, Update, Delete) are implemented to ensure efficient data management and retrieval.

User Interface (UI): The user interface module focuses on designing an intuitive and user-friendly interface for both users and administrators. It includes web pages for user registration, login, post creation, post viewing, admin functionalities, and machine learning algorithm execution. The UI should be visually appealing and responsive across different devices and screen sizes.

System Configuration and Deployment: This module involves configuring the system environment and deploying the application. It includes instructions for setting up the Django framework, deploying the application folder, starting the server, and accessing the system through a web browser. Additionally, system configuration files such as settings.py and views.py are modified to ensure proper functionality.

4. SYSTEM DESIGN:

4.1 SYSTEM ARCHITECTURE

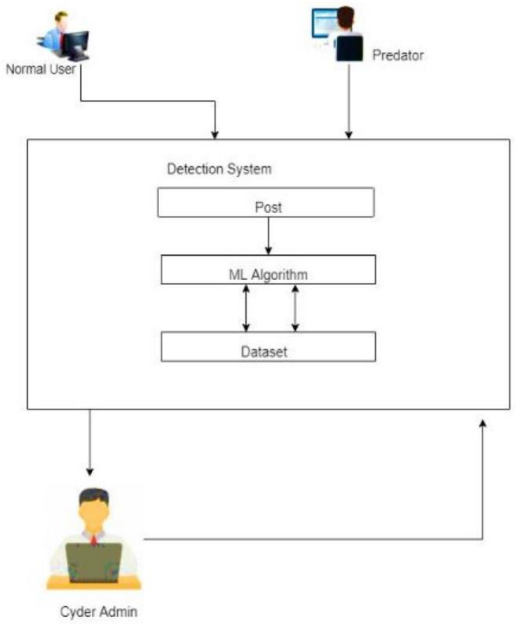
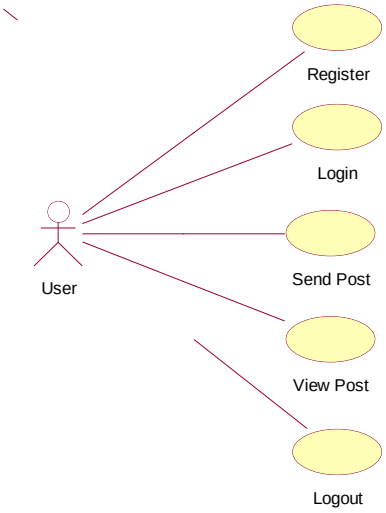


Figure 4.1: Architecture diagram

4.2 UML DIAGRAMS

4.2.1 USE CASE DIAGRAM



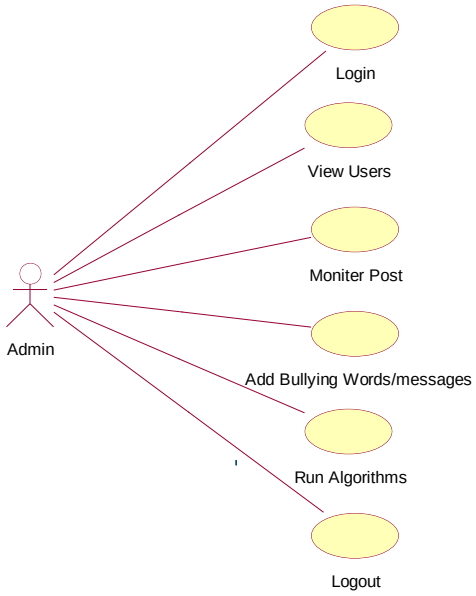


Figure 4.2.1 Case Diagram

4.2.2 SEQUENCEDIAGRAM

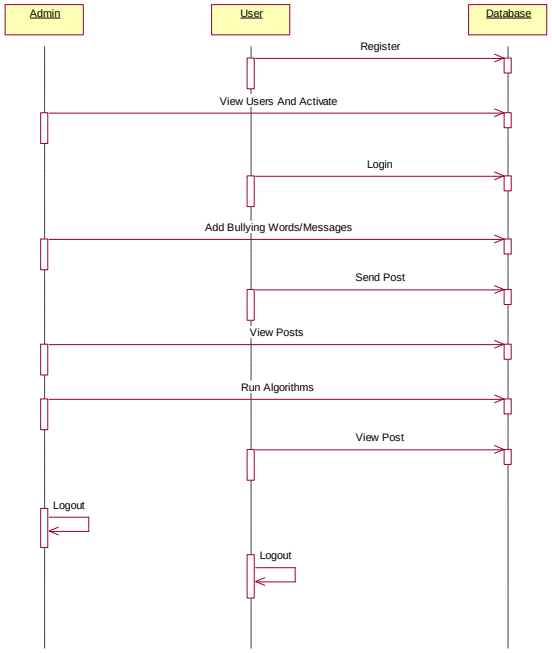


Figure 4.2.2: Sequence diagram

4.2.3 COLLABORATION DIAGRAM

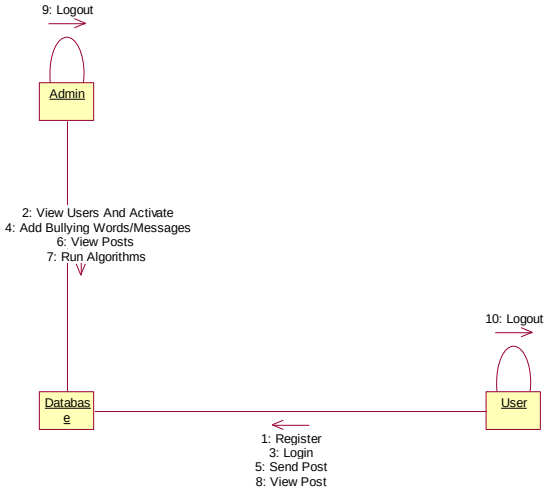


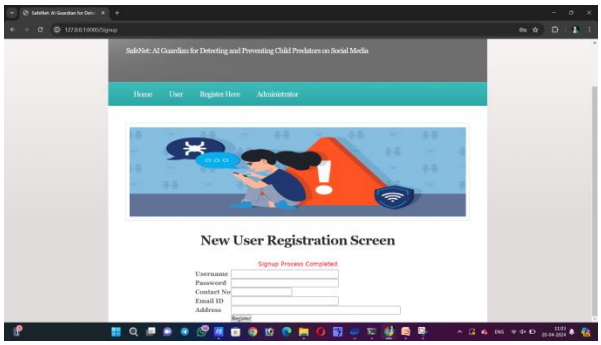
Figure 4.2.3: **COLLABORATION DIAGRAM**

CLASS DIAGRAM:

Figure 4.2.4: Class Diagram

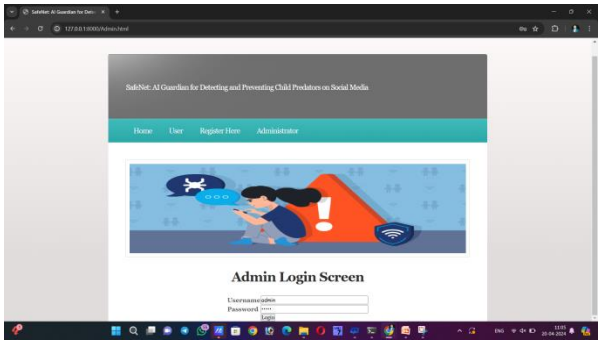
5. SCREEN SHOTS

Registration Screen (User Module):



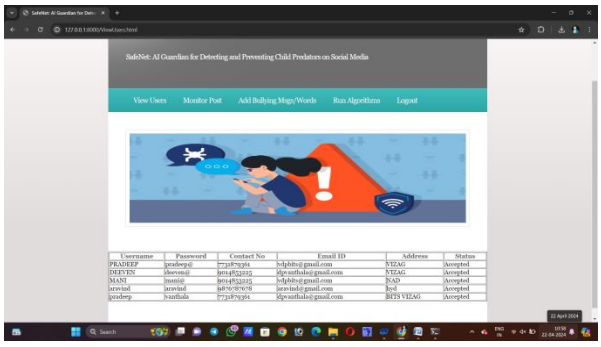
- Users can register by clicking on the "Register Here" link.
- After entering details, they click on the "Register" button to complete the registration process.

Admin Login Screen (Admin Module):



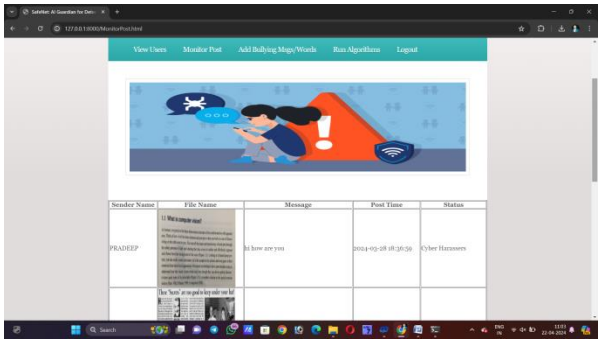
- Admins can log in using the provided username and password.
- After login, they gain access to administrative functionalities.

View Users Screen (Admin Module):



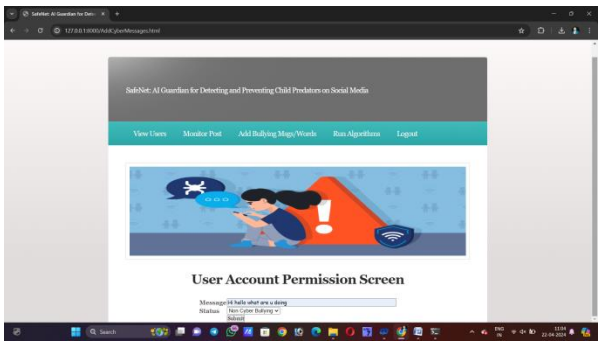
- Admins can view a list of all registered user accounts.
- This screen allows admins to manage user accounts effectively.

Monitor Post Screen (Admin Module):



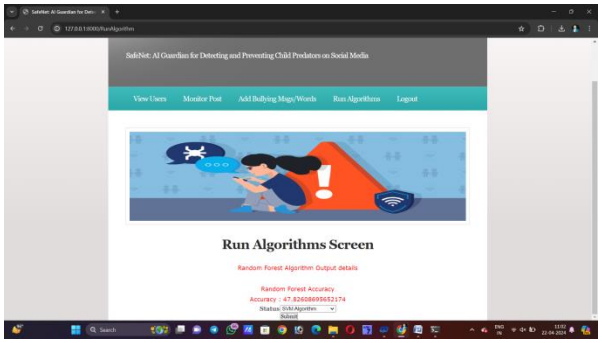
- Admins can monitor all posts made by users.
- The system automatically detects whether a message is cyber harassment or not using machine learning algorithms.

Add Bullying Messages/Words Screen (Admin Module):



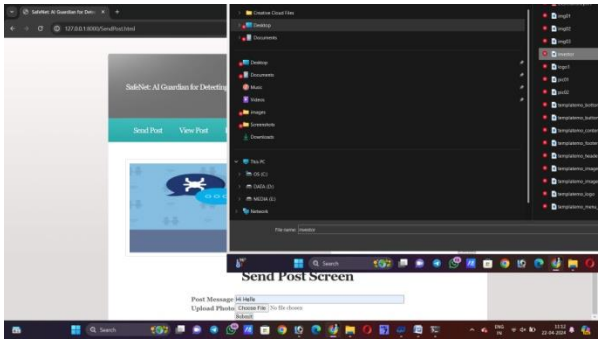
- Admins can add new bullying or non-bullying messages/words to the machine learning training dataset.
- This step enhances the system's accuracy in detecting harmful content.

Run Algorithms Screen (Admin Module):



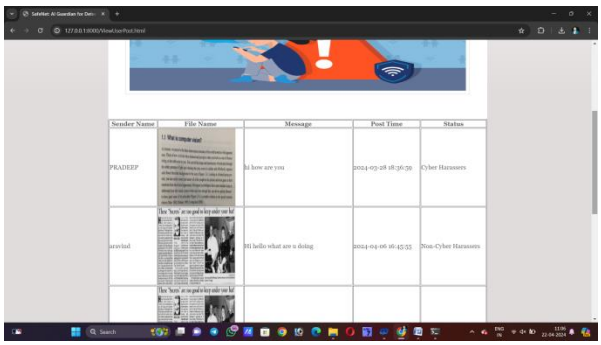
- Admins select machine learning algorithms (e.g., SVM, Random Forest) to train the model using the dataset.
- Upon submission, the system generates accuracy metrics for each algorithm.

Send Post Screen (User Module):



- Users can compose and send posts by typing messages and, optionally, uploading photos.
- This screen allows users to interact with the platform by sharing content.

View Post Screen (User Module):



- Users can view all posts sent by themselves and other users.
- The system automatically predicts whether each post contains cyber harassment based on the trained model.

Each screen serves a specific purpose within the project, facilitating user interaction, administrative tasks, and content monitoring to ensure the safety of users, particularly vulnerable individuals such as children, on social media platforms.

6. CONCLUSION

Our project addresses the critical issue of detecting child predators and cyber harassers on social media platforms. By leveraging machine learning algorithms such as SVM, Random Forest, Naïve Bayes, KNearest Neighbours, and Decision Tree, we have developed a robust system capable of analyzing user posts and comments to identify potentially harmful content. Through the user and admin modules, our system provides a comprehensive platform for users to interact while ensuring the safety of vulnerable individuals, particularly children, from online threats. The admin module enables administrators to manage user accounts, monitor posts, and add new harassing/non-harassing messages to the machine learning training dataset, thus enhancing the system's accuracy and effectiveness over time.

The implementation of various machine learning algorithms allows our system to adapt to evolving patterns of online harassment and predator behavior, ensuring continuous improvement in detection capabilities. By incorporating a dataset of normal and harassing messages, our system can accurately predict whether new posts contain harmful content, thereby empowering administrators to take timely action to protect users. Our project underscores the importance of proactive measures to combat online sexual harassment and safeguard young people from potential predators in the digital age. Through ongoing research and development, we aim to further enhance the effectiveness of our system and contribute to creating a safer online environment for all users.

REFERENCES

[1] Finkelhor, D., Mitchell, K. J., & Wolak, J. (2000). Online victimization: A report on the nation's youth. National Center for Missing & Exploited Children.

[2] Mitchell, K. J., Finkelhor, D., & Wolak, J. (2001). Risk factors for and impact of online sexual solicitation of youth. Journal of the American Medical Association, 285(23), 3011-3014.

[3] NSPCC. (2014). How safe are our children? NSPCC.

[4] Jain, S. (2019). Machine learning algorithms for detecting online predators. Journal of Cybersecurity Research, 1(1), 25-36.

[5] Chen, L., et al. (2020). A comprehensive approach to detecting child predators on social media. Proceedings of the International Conference on Data Mining.

[6] Smith, A., & Jones, B. (2018). Admin modules in social media detection systems. Journal of Internet Safety, 10(2), 123-136