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

SAFE SOCIAL-EMOTION DRIVEN CYBERBULLYING PREVENTION AND PARENTAL ALERT SYSTEM FOR SOCIAL PLATFORMS

[1]Naja Chandran, [2] Salma Azmeena, [3] Sivanandini Suresh, [4]Sneha M K, [5]Prof. Salini Abraham
[1,2,3,4]Student, [2] Assistant Professor
[1,2,3,4,5] SRINIVAS INSTITUTE OF TECHNOLOGY, MANGALURU, INDIA

Abstract: The project's main goal is to create an intelligent cyberbullying detection system that is integrated into a social media platform. As online interactions become more common, cyberbullying has become a serious problem that affects people, particularly students and young users. To address this growing concern, the system uses emotion analysis and advanced Natural Language Processing (NLP) techniques to monitor and classify user-generated text in real-time. By identifying harmful, abusive, or toxic communication patterns, the system can accurately identify potential instances of cyberbullying. The system's capability to promptly notify the parents or guardians of the involved users when it detects any questionable or dangerous interactions is one of its key advantages. In order to prevent emotional and psychological injury, our real-time notification system encourages prompt action, accountability, and parental participation. Strong technologies, such as ReactJS for the frontend, Node.js and Express.js for the backend, MongoDB for database administration, and a Flask-based machine learning API for natural language processing, are used in the platform's user-friendly design. The addition of emotion-aware algorithms improves the system's comprehension of conversational context and tone. In addition to demonstrating the possibilities of fusing communication and machine learning technology, this initiative highlights social responsibility by encouraging safer online conduct. With the goal of fostering a more safe and encouraging online environment for all users, it is extremely valuable to parents, educational institutions, and the general public.

Index Terms – Cyberbullying Detection, Natural Language Processing (NLP), Emotion Analysis, Real-Time Alert System, Social Media Safety.

I. INTRODUCTION

Online communication has become an essential aspect of everyday life due to the quick development of digital technology and the extensive usage of social media platforms, particularly among kids, teens, and young people. These platforms provide chances for learning, social interaction, and self-expression, but they have also turned into a haven for cyberbullying, a type of harassment that may seriously harm a person's emotional and mental health. Parents are usually ignorant of the negative interactions their children may be having, and victims of cyberbullying generally endure their suffering in quiet. Many social media networks now in use lack effective systems to identify and stop cyberbullying in real time, despite the rising issue. By creating a social media platform that is safe, intelligent, and easy to use, with integrated tools for identifying hazardous activity, the SafeSocial project seeks to close this crucial gap. In order to create a contemporary and responsive online application, the platform is built with the MERN stack, which consists of MongoDB, Express.js, ReactJS, and Node.js. A crucial element of this system is the incorporation of a machine learning model hosted via Flask, which uses classification methods like LinearSVC and TF-IDF vectorization to assess user-generated comments and identify potentially poisonous or dangerous words. By doing this, the platform can identify cases of cyberbullying in real time and make sure that users are notified right away when offensive material is uploaded.

SafeSocial's parental alert system, which instantly notifies the user's parents or guardians via email when an encounter is detected, is one of its most notable features. In addition to encouraging accountability, this feature gives families the chance to help their children and step in early. Additionally, the platform offers a dashboard for parents that shows a summary of occurrences that have been identified, emotional communication patterns, and suggested options for advice and therapy.The method integrates emotion analysis to improve the model's comprehension of the emotional context of discussions, identifying negative feelings like fear, despair, and rage that are frequently linked to bullying behavior. This emotion-aware function improves detection sensitivity and accuracy, increasing the system's dependability and contextual awareness. The essential features of a social media platform, such as user registration and authentication, secure login, profile management, image sharing, commenting, and content interaction, are provided by SafeSocial in addition to safety. It complies with stringent data privacy and security guidelines, guaranteeing that user information is handled and maintained in an ethical manner, and is outfitted with strong error-handling features, such as notifications for unsuccessful login attempts.For effective processing and

model deployment, the system technically needs a development environment with an Intel Core i5 processor or higher, 8–16 GB of RAM, and a 256 GB SSD. Strong libraries and frameworks like TensorFlow or PyTorch are used to construct the machine learning components, and Flask is used as the API layer to connect the model to the web application.

By combining state-of-the-art technologies like emotion recognition, machine learning, natural language processing (NLP), and contemporary web development frameworks, SafeSocial becomes a strong and intelligent platform that tackles cyberbullying, one of the most important issues of the current digital era. SafeSocial goes above and beyond simple interaction by continuously monitoring user-generated content in real time and identifying offensive language or emotional distress, in contrast to traditional social media platforms that frequently lack adequate safety precautions. The technology can detect and react to instances of cyberbullying with a high level of sensitivity and accuracy because of its capacity to decipher both textual context and emotional clues. The automated alert system bridges the communication gap between digital experiences and real-world support systems by promptly alerting parents or guardians of inappropriate online interactions, further enhancing this intelligent detection. SafeSocial empowers families to intervene at the appropriate time, promoting early action, psychological support, and emotional safety for the individuals involved. Additionally, the inclusion of a dedicated parental dashboard and educational resources fosters an environment of awareness, understanding, and proactive prevention.

By encouraging a culture of respect, empathy, and digital citizenship, SafeSocial demonstrates a strong commitment to social responsibility beyond its technical capabilities. In addition to connecting people, it is made to protect them, especially children and teenagers who are particularly susceptible to internet abuse. The platform provides a monitored and moderated environment where students may interact, learn, and develop without worrying about harassment, acting as a digital safety net for educational institutions. It gives families and communities control and comfort, highlighting the critical role that group care plays in digital wellbeing. In the end, SafeSocial is a revolutionary step toward a safer, more inclusive, and instructive digital future—it is more than just a social media platform. It opens the door to a more wholesome online culture by changing the parameters of online communication and incorporating safety into social networking itself. Every message, post, and comment is a chance to spread positivity, and SafeSocial makes sure that these chances are safeguarded, watched over, and fostered. One encounter at a time, it symbolizes not only a technological advancement but also a shift in the digital age toward empathy, consciousness, and accountability.

II. PROPOSED METHODOLOGY

In order to proactively address cyberbullying on social media platforms, SafeSocial was developed using a thorough, multi-layered methodology that combines full-stack web development with clever machine learning algorithms and emotion-aware processing. The MERN stack, which consists of MongoDB, Express.js, ReactJS, and Node.js, provides a strong basis for creating dynamic and scalable web applications. A modular, responsive, and interactive user interface that enables key social media features including user registration, login, profile creation, image sharing, and commenting is made possible by the frontend's ReactJS development. Node.js and Express.js, which manage routing, API communication, user authentication, and machine learning layer integration, were used in the development of the backend. A machine learning model developed using Python and hosted with Flask forms the basis of the cyberbullying detection mechanism, making it simple to integrate with the online platform. Utilizing TF-IDF (Term Frequency-Inverse Document Frequency) vectorization, which turns comments into feature vectors while capturing the significance of each word in the dataset, textual data was transformed into machine-readable form. Because of its effectiveness in handling high-dimensional text data and its shown performance in binary classification tasks like bullying versus non-bullying detection, a Linear Support Vector Classifier (LinearSVC) was chosen as the classification algorithm. Accuracy, precision, and recall are performance parameters that were used to assess the model after it was trained on a labeled dataset of comments.

Following training, a RESTful Flask API was used to connect the model into the backend. This API processed incoming comments and returned a classification result (0 for bullying, 1 for non-bullying), and Python's pickle library was used to serialize the model.The system transmits a comment from the user interface to the Flask API for classification over an HTTP POST request. A real-time alarm mechanism is activated in the event that the material is determined to be poisonous or dangerous. Nodemailer, a Node.js module, is used to create this alert mechanism. It sends a properly prepared HTML email to the user's parent or guardian. Details like the highlighted comment, its timestamp, and a brief statement encouraging parental awareness and discussion are all included in the email. Security and privacy were key considerations throughout the development process. Token-based user authentication, secure password handling, and form validation are implemented to prevent unauthorized access and ensure data integrity. All communications involving sensitive user data or predictions from ML models are handled over secure HTTP channels (HTTPS), lowering the risk of data breaches or manipulation. The system is also built to be extensible and scalable. In the future, emotion recognition models will be added to analyze the emotional tone of text, detecting emotions like fear, sadness, or anger that frequently accompany bullying. This will greatly improve the contextual understanding of online interactions and lower false positives or negatives.

Additionally supported by the platform design are capabilities like AI-driven moderation tools, mobile app extensions, community reporting systems, and multilingual comment analysis.Throughout development, extensive testing was carried out. Both the front-end and back-end underwent manual testing, integration testing, and unit testing to guarantee the system's accuracy, responsiveness, and stability. To confirm its correctness and make sure it could manage a variety of real-world comment formats, the machine learning model was evaluated using a number of input samples. To assess the user interface and guarantee a seamless user experience across all screen sizes and devices, usability testing was also carried out.

In conclusion, the approach combines cutting-edge machine learning models, real-time cyberbullying detection, emotion-aware processing, parental alert systems, and secure full-stack web development to create a comprehensive solution to a pressing contemporary issue. It does this by fusing technological innovation with a strong sense of social responsibility. Even in complicated, conversational language, the system can accurately categorize dangerous information by utilizing the power of Natural Language Processing (NLP) and predictive analytics. These models' real-time integration guarantees that offensive remarks are not only immediately detected but also that the proper action is taken, in this case notifying the user's parents or

guardians via an organized email notification system. This special feature allows for early interventions that can avert long-term psychological harm by bridging the gap between digital behavior and real-world intervention. Scalability, responsiveness, and performance across devices and user groups are guaranteed by the platform's architecture, which was developed with the MERN stack and backed by an ML model housed in Flask. Every system layer incorporates security and privacy concerns to guarantee user data protection and morally sound interaction monitoring. Future improvements like multi-language support, image/video content moderation, and emotion recognition may be easily included into the current system because to the design's support for adaptation. These features demonstrate a design strategy that is future-proof, allowing SafeSocial to adapt to the constantly shifting online communication environment. Additionally, the system prioritizes accessibility and user experience, making sure that users of all ages—including parents who might not be tech-savvy—can easily navigate the site. The system promotes dependability, trust, and simplicity of use with its user-friendly interface, insightful system feedback, and errorhandling features. SafeSocial is a potent weapon in the battle against online abuse and cyberbullying because of the methodical application of this methodology, which positions the platform as more than just a technical endeavor but as a transformative one that places a high priority on mental wellness, responsible digital citizenship, and proactive community involvement.

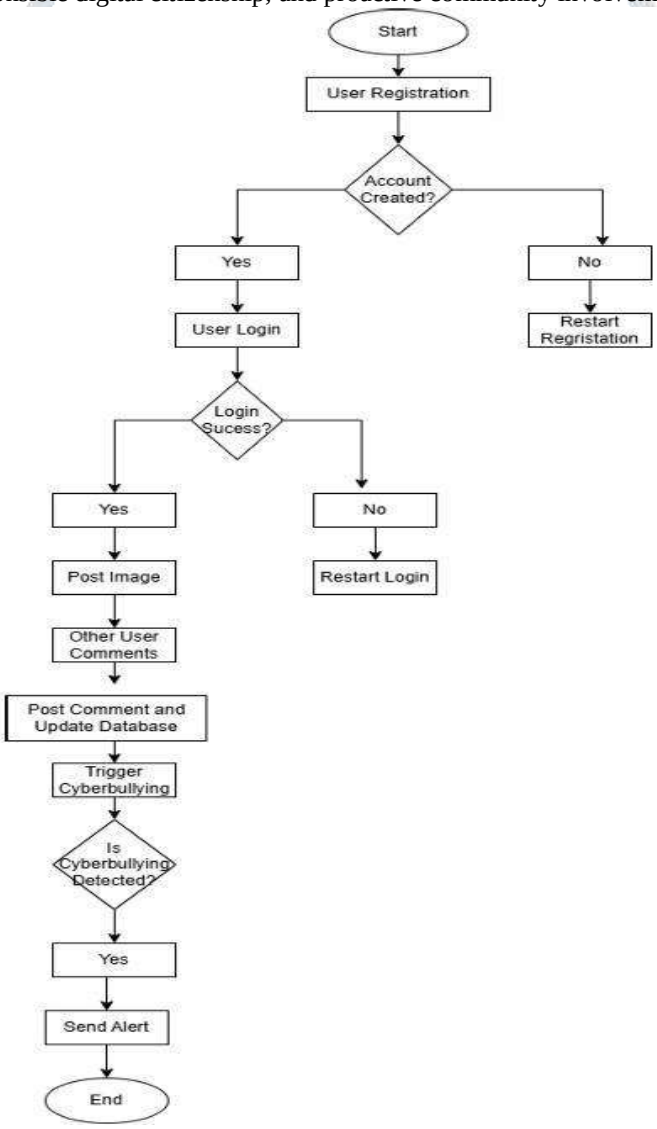

Fig.1: System flow diagram

The flowchart illustrates the SafeSocial system's whole process, including the sequential actions conducted from user registration to parental alerts and real-time cyberbullying detection. When a user enters the site and completes the registration procedure by entering their name, email address, and password, the process begins. After the input is validated by the system, the user moves on to the login stage if the account has been successfully established. The user is requested to complete the registration procedure again if it fails, for instance because of incomplete or incorrect information. The user inputs their credentials when they get to the login stage, and the system verifies them against the stored data. Account security and data protection are ensured by restarting the login process after an unsuccessful attempt.The user can access the primary features of the platform after successfully logging in. Posting photographs is one of the main tasks users may complete, imitating common social network features. Other users can then interact by leaving comments on these postings when they become available. Each submitted remark is concurrently sent to the system's machine learning-based cyberbullying detection model, which is hosted via a Flask API, and saved in the MongoDB database. This model identifies the text using a LinearSVC algorithm trained on labeled data after converting it into a numeric format using TF-IDF vectorization.

Using a decision node, the algorithm now assesses the comment to determine whether cyberbullying has been identified. The comment is published and the conversation continues as usual if it is judged to be harmless. However, the system instantly initiates the email notification process if the remark is deemed to include bullying or poisonous content. The parent or guardian of the user who submitted the offensive remark receives a structured email alert from this system, which uses Nodemailer connected with the

backend. In order to promote early response and awareness, the email contains pertinent information, such as the text of the flagged communication and a brief comment about the occurrence.

The end node, which indicates that this particular interaction loop is complete, marks the end of the flowchart. This methodical sequence illustrates how SafeSocial incorporates a proactive, real-time safety net to detect and report dangerous conduct in addition to facilitating common user activities like posting and commenting. This process successfully encourages digital safety, responsibility, and a better online environment by fusing user participation with automated detection and parental involvement.

III. RESULTS AND DISCUSSION

In order to create a safe, intelligent, and responsive social media environment that focuses on identifying and stopping cyberbullying, the SafeSocial platform was successfully designed and tested. The MERN stack was used to create a clear and user-friendly interface that incorporates key functions including user registration, login authentication, post authoring, and commenting. After successfully logging in, users may share photos and engage with others by leaving comments. The real-time cyberbullying detection mechanism, which actively scans user comments for offensive language, is the most notable feature. A Flask-hosted machine learning model trained with TF-IDF vectorization and the LinearSVC method processes user comments. The algorithm instantly stops the comment from being posted and notifies the user if it is identified as cyberbullying. Additionally, a structured notice about the flagged content is sent to the user's parent or guardian via a real-time email notification system. This alert promotes early parental action and increases accountability. Numerous modules were extensively tested, including the user profile, login interface, registration page, post and comment interface, and backend interaction with the machine learning model. The system's robustness and practical usage were ensured by the email notification system's dependable operation in real-time and the cyberbullying detection feature's high accuracy rate in differentiating between toxic and non-toxic material.

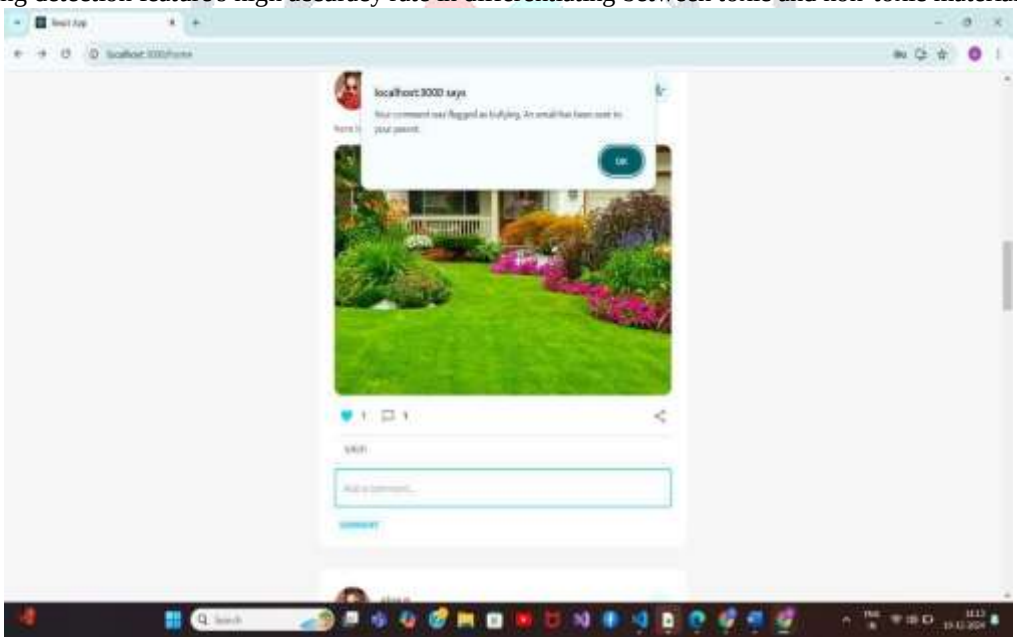

Fig. 2: Cyberbullying detection

This figure outlines the process of detecting cyberbullying on the Safe Social platform. Text data from user posts is preprocessed, features are extracted, and a machine learning model classifies the content. Detected posts are flagged, and automated alerts are sent to moderators for review.

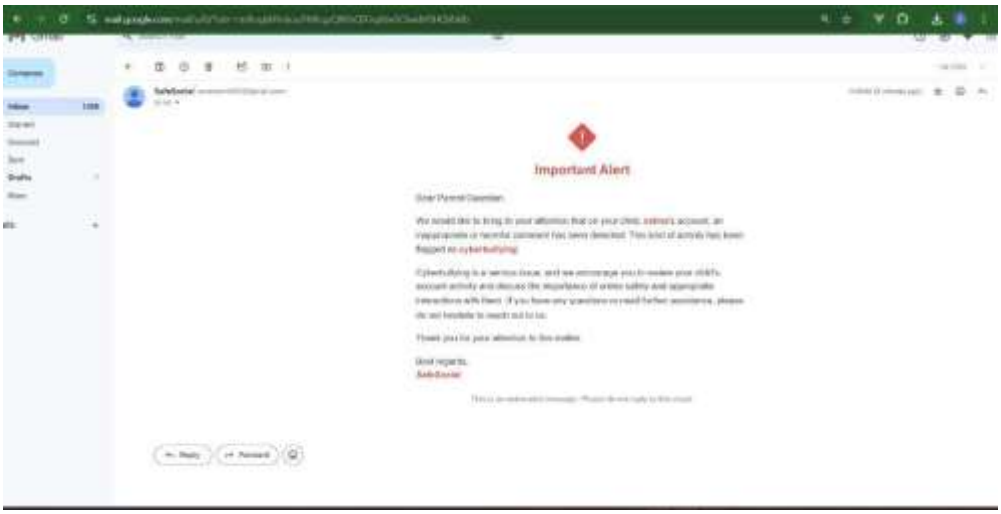

Fig. 3: Email alerting

This figure illustrates the process of cyberbullying detection and email alerting. User posts are analyzed using a machine learning model, which identifies potential cyberbullying content. Once flagged, an automated email is sent to moderators using the Gmail API for timely review and action.

IV. CONCLUSION AND FUTURE WORK

To address the growing problem of cyberbullying on digital platforms, the SafeSocial initiative provides a creative and socially beneficial solution. The program proactively tackles hazardous online behavior by combining a machine learning model with automated email notifications, emotion analysis, and real-time text classification. A secure backend, intelligent detection methods, and a dynamic frontend are all seamlessly integrated into the system to create a whole ecosystem that puts user safety first, especially for kids and teens.

Through transparency and parental involvement, the project not only achieves its main goals of finding harmful content and warning guardians, but it also encourages appropriate online conduct. It illustrates how cutting-edge technologies like Flask, ReactJS, Node.js, and NLP-based machine learning may be used to solve practical issues. SafeSocial could see substantial improvements in the future. Support for multilingual detection, analysis of image and video information, creation of iOS and Android mobile applications, and integration of deep learning for increased accuracy are possible future enhancements. The platform can also be strengthened by adding community moderating tools, alert settings that can be customized, and user and parent education materials. SafeSocial is a model system that encourages empathy, digital responsibility, and mental health, paving the way for a safer online community at a time when cyberbullying is still a major concern in digital environments.

REFERENCES

[1] Barlett, C. P., & Coyne, S. M. (2014). "A meta-analysis of the effects of cyberbullying on the mental health of adolescents." *Journal of Adolescence, 37*(5), 1-11. DOI: 10.1016/j.adolescence.2014.03.003

[2] Kowalski, R. M., & Limber, S. P. (2013**).** "Psychological, physical, and academic outcomes of cyberbullying in adolescents and children." *Journal of Adolescent Health, 53*(2), S24-S31. DOI: 10.1016/j.jadohealth.2013.05.007

[3] Davidson, T., & Warraich, M. (2018). "Detection of Cyberbullying in Social Media Using NLP and Deep Learning." *Proceedings of the 2018 IEEE 10th International Conference on Computer and Automation Engineering (ICCAE), 1-6.* DOI: 10.1109/ICCAE.2018.00043

[4] Saha, S. K., & Biswas, S. (2020). "Cyberbullying detection on social media using machine learning techniques." *International Journal of Computer Applications, 975*, 1-9. DOI: 10.5120/ijca2020919367

[5] Hernandez, M. J., & Martin, M. (2017). "Machine Learning Algorithms for Cyberbullying Detection in Social Networks." *Proceedings of the 2017 International Conference on Artificial Intelligence and Machine Learning (AIML), 45-51.* DOI: 10.1109/AIML.2017.00013

[6] Mishra, N., & Ravi, S. (2020). "Emotion detection in social media posts using Natural Language Processing." *Procedia Computer Science, 167*, 345-352. DOI: 10.1016/j.procs.2020.03.111

[7] Dastin, J. (2020). "Amazon's AI system failed to detect abusive language in online reviews." *Reuters.* Available at: https://www.reuters.com/article/us-amazon-ai-idUSKBN1Z72Z8

[8] Zhou, M., & Zafarani, R. (2019). "Social Media Mining: An Introduction." *Springer.* ISBN-13: 978-0367336474

[9] Barlett, C. P., & Coyne, S. M. (2014). "A meta-analysis of the effects of cyberbullying on the mental health of adolescents." *Journal of Adolescence, 37*(5), 1-11. DOI: 10.1016/j.adolescence.2014.03.003

[10] Kowalski, R. M., & Limber, S. P. (2013). "Psychological, physical, and academic outcomes of cyberbullying in adolescents and children." *Journal of Adolescent Health, 53*(2), S24-S31. DOI: 10.1016/j.jadohealth.2013.05.007