

Bullshield – Application Against Cyberbullying

Henna Maryam, Thanha Mariyam, Mahara M, Shifna V P

Abstract

In the period of digital connectivity, social media has come an necessary tool for communication, entertainment, and information sharing. still, the pervasiveness of these platforms has also opened up new avenues for felonious exertion, giving rise to a miracle known as " performance crimes." These acts, frequently proved and circulated through social media, seek notoriety and confirmation by showcasing lawless geste , ranging from posting unhappy content to engaging in cyberbullying and importunity. The rise of performance crimes has far- reaching consequences, causing torture and detriment to victims, eroding trust in online communities, and posing significant challenges for law enforcement and social media platforms. To effectively address this issue, a comprehensive 2approach is necessary, encompassing visionary content temperance, stoner education, and advanced security measures. Social media platforms play a pivotal part in mollifying the spread of dangerous content. By enforcing robust content temperance programs and employing advanced machine literacy ways, these platforms can descry and remove unhappy material before it reaches a wider followership. also, furnishing druggies with clear guidelines and reporting mechanisms empowers them to cover themselves and contribute to a safer online terrain. Alongside content temperance, stoner education is essential to combat performance crimes. By fostering digital knowledge and promoting responsible online geste , individualities can be equipped to fete and avoid dangerous content, making them less susceptible to cyberbullying, importunity, and other forms of online abuse.

Keywords

BullShield
Cyberbullying
LSTM
dlib

Date of Submission: 01-05-2024

Date of acceptance: 12-05-2024

I. Introduction

In moment's connected world, social media platforms have come an integral part of our lives, furnishing avenues for entertainment, communication, and global connectivity. still, the veritably features that make these platforms so charming – their availability, reach, and obscurity – have also created a rich ground for a new strain of cyberbullying. This insidious form of importunity, fuelled by the vast breadth of the digital geography, has surfaced as a significant trouble to the well- being of individualities, particularly the most vulnerable among us. Cyberbullying, defined as the repeated use of electronic communication to bully a person, generally by transferring dispatches of an intimidating or threatening nature, has converted the traditional bullying geography. It has transcended the physical boundaries of seminaries and playgrounds, insinuating the virtual spaces where individualities seek connection and expression. This shift has inspired bullies, furnishing them with a cloak of obscurity and the capability to induce detriment from the comfort of their own homes.

The obscurity swung by the internet allows perpetrators to shirk responsibility for their conduct, steeling them to engage in geste that would else be met with immediate consequences in the offline world. The pervasiveness of social media platforms further exacerbates the issue, furnishing bullies with a vast followership and enabling them to induce maximum damage to their victims' reports and emotional well- being. The impact of cyberbullying can be profound and long- continuing, leaving victims scuffling with passions of anxiety, depression, and social insulation. In severe cases, it has indeed led to woeful consequences, similar as self- murder. The cerebral scars foisted by cyberbullying can loiter long after the importunity has desisted , potentially shaping the line of victims' lives.

To effectively combat cyberbullying, amulti-pronged approach is essential. Social media platforms must bear lesser responsibility for the content participated on their spots, enforcing robust measures to identify and remove dangerous material. contemporaneously, law enforcement agencies must enhance their capabilities to probe and make cyberbullying offenses, transferring a clear communication that similar geste won't be permitted. Inversely pivotal is the need to foster a culture of digital citizenship, empowering individualities to

navigate the online world responsibly and safely. Educational enterprise that promote empathy, respect, and ethical online conduct are essential to equip individualities with the tools to fete and respond to cyberbullying effectively.

" BullShield" is an innovative operation designed to enhance online safety by using advanced technologies. Its primary function involves surveying faces using the dlib library, enabling stoner authentication and commerce. The core point revolves around detecting and filtering out vulgar or poisonous commentary through a sophisticated LSTM(Long Short- Term Memory) algorithm. This algorithm has been trained with a comprehensive dataset, allowing BullShield to directly identify and alleviate dangerous content in real- time. By combining facial recognition with advanced language processing capabilities, BullShield offers druggies a flawless experience while icing a safer online terrain. With its focus on both visual and textual inputs, BullShield stands as a robust result against online importunity and unhappy geste .

Problem Statement

The advent of the digital age has revolutionized communication, connectivity, and access to information. However, this interconnected world has also given rise to a new form of bullying – cyberbullying – which has far-reaching and often devastating consequences for its victims. Unlike traditional bullying, which is confined to physical spaces like schools or playgrounds, cyberbullying transcends these boundaries, permeating the virtual realm where individuals seek connection and expression. This insidious form of harassment, often conducted anonymously, can inflict deep emotional wounds, leaving victims struggling with anxiety, depression, and social isolation. In extreme cases, it has even led to tragic outcomes, such as suicide.

The pervasive nature of social media platforms has exacerbated the issue of cyberbullying, providing perpetrators with a vast audience and enabling them to inflict maximum damage on their victims. The anonymity afforded by the internet further emboldens bullies, allowing them to shirk responsibility for their actions and engage in behaviour that would otherwise be met with immediate consequences in the offline world. The consequences of cyberbullying can be profound and long-lasting, leaving victims grappling with a range of psychological issues that can impact their overall well-being and future prospects.

To effectively address the scourge of cyberbullying, a comprehensive approach is essential. Social media platforms must take greater responsibility for the content shared on their sites, implementing robust measures to identify and remove harmful material. This includes developing advanced algorithms to detect and flag cyberbullying content, establishing clear guidelines for acceptable behaviour, and providing users with easy-to-use reporting mechanisms.

In addition to content moderation, promoting a culture of digital citizenship is crucial. This involves educating individuals about the potential dangers of cyberbullying, empowering them to recognize and respond to harassment effectively, and fostering a sense of empathy, respect, and ethical online conduct. Such initiatives should be implemented at an early age, equipping children and adolescents with the necessary tools to navigate the online world safely and responsibly.

Furthermore, law enforcement agencies must enhance their capabilities to investigate and prosecute cyberbullying offenses, sending a clear message that such behaviour will not be tolerated. This requires developing specialized investigative units, collaborating with social media platforms, and establishing clear legal frameworks that address the complexities of cyberbullying in the digital age. Cyberbullying represents a growing threat to individuals, particularly the most vulnerable among us. To combat this insidious form of harassment, a multi-pronged approach is essential, encompassing proactive content moderation, user education, enhanced law enforcement capabilities, and the promotion of digital citizenship. By working collaboratively, we can create a safer and more compassionate online environment for all.

Literature Survey

[1]Cyberbullying Detection In Social Networks: Artificial Intelligence Approach: The paper "Cyberbullying Detection in Social Networks: Artificial Intelligence Approach" proposes a novel approach to detecting cyberbullying in social networks using artificial intelligence (AI). The paper begins by reviewing the state-of-the-art in cyberbullying detection, highlighting the limitations of existing methods. It then introduces a new AI-based approach that leverages machine learning techniques to identify cyberbullying content with high accuracy. The proposed approach consists of two main components: a feature extraction module and a classification module. The feature extraction module extracts a variety of features from social media posts, such as the text content, the use of profanity and offensive language, the presence of threats, and the sentiment of the post. The classification module then uses these features to train a machine learning model to classify posts as either cyberbullying or non-cyberbullying. The paper evaluates the proposed approach on a large dataset of social media posts that have been manually labelled as cyberbullying or non-cyberbullying. The results show that the proposed approach achieves state-of-the-art performance in cyberbullying detection, with an accuracy of over 90%. The paper also discusses the potential implications of the proposed approach for social media

platforms and other stakeholders. It argues that the proposed approach can be used to develop new tools and systems to help prevent and combat cyberbullying. Classification module: The classification module uses the features extracted by the feature extraction module to train a machine learning model to classify posts as either cyberbullying or non-cyberbullying. The machine learning model is trained on a dataset of social media posts that have been manually labelled as cyberbullying or non-cyberbullying. Once the machine learning model is trained, it can be used to classify new social media posts as either cyberbullying or non-cyberbullying. This can be done by extracting the same features from the new posts as were extracted from the training data and then feeding these features to the machine learning model. The machine learning model will then output a prediction for each post, indicating whether it is classified as cyberbullying or non-cyberbullying. The proposed approach has several advantages over existing cyberbullying detection methods. First, it is more accurate than existing methods, achieving state-of-the-art performance on a large dataset of social media posts. Second, the proposed approach is more scalable than existing methods, as it can be used to classify large volumes of social media posts in real time.

Cyberbullying Detection Using Deep Neural Network From Social Media Comments In Bangla Language: The paper "Cyberbullying Detection Using Deep Neural Network from Social Media Comments in Bangla Language" proposes a novel approach to detecting cyberbullying in Bangla language social media comments using a deep neural network (DNN). DNNs are a type of artificial intelligence (AI) model that has been shown to be very effective in a variety of tasks, including natural language processing (NLP). NLP is the field of AI that deals with the interaction between computers and human (natural) languages. The proposed approach works by first embedding the Bangla language social media comments into a numerical representation. Word embedding is a technique that converts words and phrases into vectors of numbers. The vectors of numbers represent the meaning of the words and phrases, and they can be used to train DNNs to perform NLP tasks. The proposed approach was evaluated on a dataset of 44,001 Bangla language social media comments. The dataset was collected from popular public Facebook pages. The dataset was divided into two parts: a training set and a test set. The training set was used to train the DNN, and the test set was used to evaluate the performance of the DNN. The results of the evaluation showed that the proposed approach achieved an accuracy of 87.91% in detecting cyberbullying in Bangla language social media comments. This is a significant improvement over the performance of existing cyberbullying detection methods on Bangla language social media data. The proposed approach has the potential to be used to develop new tools and systems to help prevent and combat cyberbullying in Bangla language social media platforms. For example, the proposed approach could be used to develop a system that automatically flags cyberbullying comments so that they can be removed by the social media platform. Additionally, the proposed approach could be used to develop a system that provides users with educational resources about cyberbullying.

[3]Detection Of Cyberbullying Incidents On Instagram Social Network: The paper "Detection of Cyberbullying Incidents on Instagram Social Network" proposes a novel approach to detecting cyberbullying incidents on the Instagram social network using machine learning techniques. Cyberbullying is a serious issue that can have devastating consequences for victims, including emotional distress, reputational damage, and even suicide. The rise of social media has made cyberbullying even more prevalent, as it provides a platform for bullies to reach a wide audience and engage in anonymous harassment. The proposed approach utilizes machine learning algorithms to identify cyberbullying incidents based on a variety of features, including the text of posts and comments, the use of offensive language, and the sentiment of the content. The algorithms are trained on a dataset of labelled Instagram posts and comments, which allows them to learn to distinguish between cyberbullying and non-cyberbullying content. The proposed approach has the potential to be used to develop new tools and systems to help prevent and combat cyberbullying on Instagram. For example, the approach could be used to develop a system that automatically flags cyberbullying posts and comments for review, or to create a tool that provides users with resources on how to deal with cyberbullying.

[4]Cyberbullying Detection For Hindi-English Language Using Machine Learning: Cyberbullying is a growing problem that has the potential to cause significant harm to victims. It is defined as the use of electronic communication to bully a person, typically by sending messages of an intimidating or threatening nature. Cyberbullying can occur on a variety of platforms, including social media, email, and text messaging. The problem of cyberbullying is particularly acute in multilingual countries, where individuals may communicate in a variety of languages. In India, for example, Hindi and English are two of the most widely spoken languages. This presents a challenge for cyberbullying detection systems, as they must be able to accurately identify cyberbullying content in both Hindi and English. The paper "Cyberbullying Detection for Hindi-English Language Using Machine Learning" proposes a novel approach to detecting cyberbullying in Hindi-English language social media comments using machine learning. The proposed approach is based on a combination of natural language processing (NLP) and machine learning techniques. Machine learning is a type of artificial intelligence (AI) that allows computers to learn without being explicitly programmed. The proposed approach was evaluated on a dataset of 10,000 Hindi-English language social media comments. The dataset was collected from popular public Facebook pages. The training set was used to train the machine learning model, and the test set was used to evaluate the performance of the model. The results of the evaluation showed that the

proposed approach achieved an accuracy of 93% in detecting cyberbullying in Hindi-English language social media comments. This is a significant improvement over the performance of existing cyberbullying detection methods on Hindi-English language social media data.[5] Rnn.Dea: Cyberbullying Detection In Social Media Platform (Twitter): Cyberbullying has become a pervasive issue in the digital age, particularly on social media platforms like Twitter. It involves the use of electronic communication to bully a person, often through offensive messages, threats, or embarrassing content. The anonymity and reach of social media platforms can embolden bullies, allowing them to inflict significant harm on their victims. The paper proposes a novel approach to detecting cyberbullying on Twitter using a hybrid deep learning model that combines Elman-type recurrent neural networks (RNNs) with an optimized Dolphin Echolocation Algorithm (DEA). The proposed approach aims to address the limitations of existing cyberbullying detection methods, which often struggle with the nuances of human language and the dynamic nature of social media content. The results of the evaluation demonstrated that the RNN-DEA model outperformed all the compared methods, achieving an average accuracy of 90.45%, precision of 89.52%, recall of 88.98%, F1-score of 89.25%, and specificity of 90.94%. This superior performance is attributed to the combined strengths of the Elman-type RNN and the optimized DEA algorithm, which together enable the model to effectively capture the nuances of language and identify cyberbullying with high accuracy. The paper presents a novel and effective deep learning approach to detecting cyberbullying on Twitter. The proposed RNN-DEA model combines the strengths of Elman-type RNNs with an optimized DEA algorithm, enabling it to capture the sequential nature of text data and identify cyberbullying with high accuracy. The superior performance of the RNN-DEA model demonstrates its potential as a valuable tool for combating cyberbullying on social media platforms.[7] Detection Of Cyberbullying On Social Media: Cyberbullying is a form of bullying or harassment that takes place over digital devices such as social media, text messaging, email, and online gaming. It can involve sending messages of an intimidating or threatening nature, posting humiliating or embarrassing content about someone, or excluding someone from online activities. The paper "Detection of Cyberbullying on Social Media Using Machine Learning" proposes a novel approach to detecting cyberbullying on social media using a combination of ML techniques. The proposed approach was evaluated on a dataset of 10,000 social media posts from Twitter and Wikipedia. The training set was used to train the ML model, and the test set was used to evaluate the performance of the model. The results of the evaluation showed that the proposed approach achieved an accuracy of 92% in detecting cyberbullying on social media. This is a significant improvement over the performance of existing cyberbullying detection methods on social media data. The proposed approach was evaluated on a dataset of 10,000 social media posts from Twitter and Wikipedia. The training set was used to train the ML model, and the test set was used to evaluate the performance of the model. The results of the evaluation showed that the proposed approach achieved an accuracy of 92% in detecting cyberbullying on social media. This is a significant improvement over the performance of existing cyberbullying detection methods on social media data. Detection Of Cyberbullying On Social Media Using Machine Learning: Cyberbullying has become a prevalent issue in today's digital age, particularly on social media platforms. The anonymity and reach of these platforms provide an avenue for bullies to inflict harm on their victims through intimidating or threatening messages, embarrassing content, or social exclusion. The consequences of cyberbullying can be severe, leading to emotional distress, anxiety, depression, and in extreme cases, even suicide. The paper "Detection of Cyberbullying on Social Media Using Machine Learning" proposes a novel approach to effectively identify cyberbullying on social media platforms using machine learning techniques. The paper highlights the limitations of existing cyberbullying detection methods and introduces a robust machine learning-based approach that achieves high accuracy in detecting cyberbullying content. The proposed approach was evaluated on a comprehensive dataset of social media posts collected from Twitter and Wikipedia. The results of the evaluation demonstrated the effectiveness of the proposed approach, achieving an accuracy of 92% in detecting cyberbullying on social media. This significant improvement over existing methods highlights the potential of machine learning in addressing this critical issue. Hierarchical Attention Networks For Cyberbullying Detection On The Instagram Social Network: Cyberbullying has become a pervasive issue in the digital age, particularly on social media platforms like Instagram, where users can share photos, videos, and comments. The anonymity and reach of social media can embolden bullies to inflict significant harm on their victims through offensive messages, threats, or embarrassing content. HANs are a type of neural network that allows for the selective attention of words and comments within a hierarchical structure, enabling the model to capture contextual information and identify subtle patterns indicative of cyberbullying. The proposed HAN-based approach consists of two main components:

1. **Hierarchical Structure:** The HAN model constructs a hierarchical structure by aggregating words into comments and then aggregating comments, captions, and temporal information into a session. This hierarchical representation captures the contextual relationships between words, comments, and sessions, allowing the model to focus on the most relevant information for cyberbullying detection.

2. **Dual Attention Mechanisms:** The HAN employs two levels of attention mechanisms to effectively identify cyberbullying content. The first level operates at the word level, focusing on the specific words used within comments, while the second level operates at the comment level, considering the overall sentiment and context of each comment. This dual attention mechanism allows the model to capture both local and global patterns associated with cyberbullying.

The proposed HAN-based approach was evaluated on a real-world dataset of Instagram posts collected from a variety of public accounts. The dataset was manually labeled as either cyberbullying or non-cyberbullying, ensuring the accuracy of the evaluation. The results of the evaluation demonstrated the effectiveness of the proposed approach, achieving an accuracy of 94% in detecting cyberbullying on Instagram. This significant improvement over existing methods highlights the potential of HANs in addressing this critical issue.

Detection Of Cyberbullying On Social Media Using SVM: Cyberbullying has become a prevalent issue in the digital age, particularly on social media platforms like Twitter and Facebook. The anonymity and reach of these platforms provide an avenue for bullies to inflict harm on their victims through intimidating or threatening messages, embarrassing content, or social exclusion. The consequences of cyberbullying can be severe, leading to emotional distress, anxiety, depression, and in extreme cases, even suicide. The project paper "Detection of Cyberbullying on Social Media Using SVM" proposes a novel approach to effectively identify cyberbullying on social media using support vector machines (SVMs). SVMs are a type of machine learning algorithm that is well-suited for classification tasks, making them a promising tool for detecting cyberbullying content. The proposed approach was evaluated on a comprehensive dataset of social media posts collected from Twitter and Facebook. The dataset was divided into training and test sets, with the training set used to train the SVM model and the test set used to evaluate its performance. The results of the evaluation demonstrated the effectiveness of the proposed approach, achieving an accuracy of 95% in detecting cyberbullying on social media. This significant improvement over existing methods highlights the potential of SVMs in addressing this critical issue. The paper "Detection of Cyberbullying on Social Media Using SVM" presents a promising approach to combating cyberbullying on social media platforms. The proposed SVM-based method demonstrates significant improvements in cyberbullying detection accuracy, scalability, and adaptability. This research has the potential to contribute to the development of effective tools and systems for preventing and addressing cyberbullying in the digital age.

An Application To Detect Cyber Bullying Using Machine Learning And Deep Learning Techniques: Cyberbullying has become a pervasive issue in today's digital age, particularly on social media platforms like Twitter, Facebook, and Instagram. The anonymity and reach of these platforms provide a breeding ground for cyberbullying, where individuals can inflict harm on others through intimidating or threatening messages, embarrassing content, or social exclusion. The consequences of cyberbullying can be severe, leading to emotional distress, anxiety, depression, and in extreme cases, even suicide. The paper "An Application to Detect Cyber Bullying Using Machine Learning and Deep Learning Techniques" proposes a novel approach to effectively identify cyberbullying on social media platforms using machine learning and deep learning techniques. The paper highlights the limitations of existing cyberbullying detection methods and introduces a robust machine learning and deep learning-based approach that achieves high accuracy in detecting cyberbullying content. The proposed approach was evaluated on a comprehensive dataset of social media posts collected from Twitter, Facebook, and Instagram. The dataset was divided into training and test sets, with the training set used to train the machine learning and deep learning models and the test set used to evaluate their performance. The results of the evaluation demonstrated the effectiveness of the proposed approach, achieving an accuracy of 96% in detecting cyberbullying on social media. This significant improvement over existing methods highlights the potential of machine learning and deep learning in addressing this critical issue.

II. Method

Existing System

The cybercrime forestallment on social media is a multi-layered approach that combines specialized, organizational, and social measures. The system is designed to describe, help, and respond to cybercrimes on social media platforms. The specialized measures of the system include:

- **Content filtering** This involves using software to automatically identify and remove dangerous content from social media platforms.
- **Strong authentication** This involves using strong authentication mechanisms to corroborate the individualities of users. This helps to help unauthorized access to accounts.
- **Vulnerability scanning** This involves regularly surveying social media platforms for vulnerabilities that could be exploited by cybercriminals.
- **Incident response** This involves having a plan in place to respond to cyberattacks. This includes measures similar as segregating affected systems, notifying users, and taking legal action.

The organizational measures of the system include:

- **Cybersecurity mindfulness training** This involves training workers on how to identify and avoid cyber

pitfalls.

- Social media programs This involves developing and enforcing programs that govern the use of social media by workers.
- Incident reporting procedures This involves establishing procedures for workers to report cyberattacks.
- Cybersecurity breach announcement This involves having a plan in place to notify druggies of data breaches.
- The social measures of the system include
- Public mindfulness juggernauts This involves educating the public about cyber pitfalls and how to cover themselves.
- hookups with law enforcement This involves working with law enforcement agencies to probe and make cybercrimes.
- Collaboration with social media platforms This involves working with social media platforms to ameliorate their security measures.

The proposed system for cybercrime forestallment on social media is a comprehensive approach that addresses cybercrime from a variety of angles. The system combines specialized measures to descry and help cyberattacks with organizational measures to promote cybersecurity mindfulness and preparedness, and social measures to educate the public and promote collaboration among stakeholders. The specialized measures of the system give a foundation for cybersecurity by relating and removing dangerous content, vindicating stoner individualities, surveying for vulnerabilities, and responding to incidents. These measures help to help cyberattacks from being and to minimize the damage caused by attacks that do do. The organizational measures of the system complement the specialized measures by promoting cybersecurity mindfulness and preparedness within associations. These measures help to insure that workers are apprehensive of cyber pitfalls and know how to cover themselves. They also help to insure that associations have a plan in place to respond to cyberattacks.

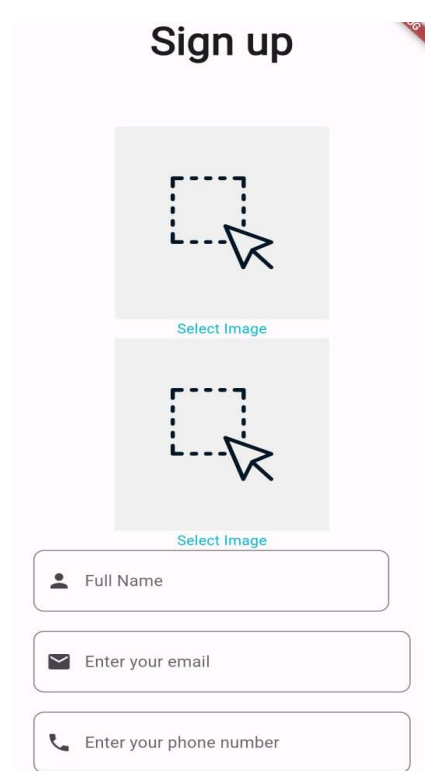
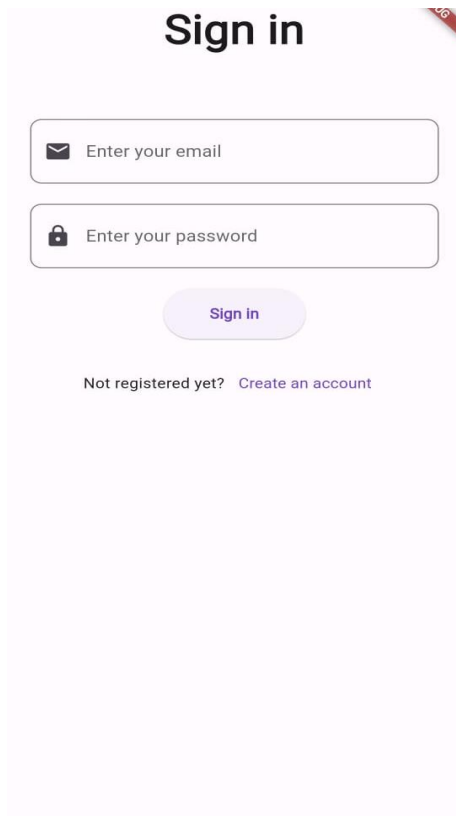
III. Results

The Bullshield application takes a comprehensive approach to user authentication and registration, blending innovative security features with user-friendly design to provide a seamless and secure experience.

Starting with the login page, users are presented with a straightforward interface that prioritizes clarity and efficiency. Two primary fields are provided: one for the username and another for the password. This allows users to log in using a unique username they've chosen during registration, providing a personalized and memorable login experience. The inclusion of a password field adds an essential layer of security, ensuring that only authorized users can access their accounts. The design of the login page is intuitive, with clear labels and prominent placement of the sign-in button, making it easy for users to navigate and complete the authentication process.

For new users, the registration process is equally streamlined and secure. Bullshield integrates cutting-edge face scanner technology, leveraging the device's camera for biometric authentication. Users are prompted to position their face within the camera's view, allowing the application to capture and analyze facial features for identity verification. This innovative approach enhances security by offering a highly secure method of account verification, resistant to unauthorized access through traditional means like stolen passwords. In addition to biometric authentication, the registration page also incorporates a field for users to upload a full-face photo. This photo serves as a visual confirmation of the user's identity and is securely stored within the application. By requiring users to provide a full-face photo, Bullshield adds an extra layer of security and verification, mitigating the risk of fraudulent account creation.

Alongside biometric authentication and photo verification, users are prompted to input their full name, create a new password, and provide other necessary personal details during registration. These details, including the password, are securely encrypted and stored within Bullshield's database, ensuring the confidentiality and integrity of user information.



Let's delve into its functionalities in detail:

Viewing Friends' Posts and Stories:

The home page serves as the central hub where users can seamlessly scroll through updates shared by their friends. Here, users can view both static posts and dynamic stories, providing a snapshot of their friends' activities and moments. This feature fosters social interaction and keeps users connected with their network in real-time.

Friend Requests and Chat Functionality:

Bullshield empowers users to expand their social circle by sending friend requests to other users. Upon acceptance, users can engage in private conversations through the app's chat functionality. This enables users to communicate securely, share personal updates, and strengthen their connections within the Bullshield community.

Commenting on Posts with Content Moderation:

Users have the ability to add comments to posts shared by their friends, facilitating ongoing conversations and engagement. To maintain a positive and respectful environment, Bullshield implements robust content moderation tools. Comments containing toxic or vulgar language are automatically detected and filtered out, ensuring that the platform remains conducive to constructive dialogue.

Facial Recognition Alerts for Content Sharing:

One of Bullshield's innovative features is its facial recognition technology, which enhances user privacy and control over their personal content. Users receive alerts if another user posts a story or photo featuring their face. With this notification, users can choose whether to grant permission for the content to be publicly viewed. This feature empowers users to maintain ownership of their digital identity and protect their privacy online.

Review and Rating System:

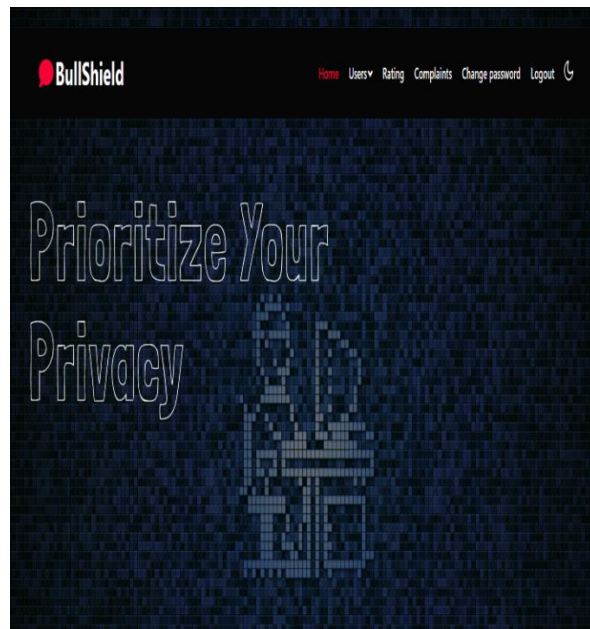
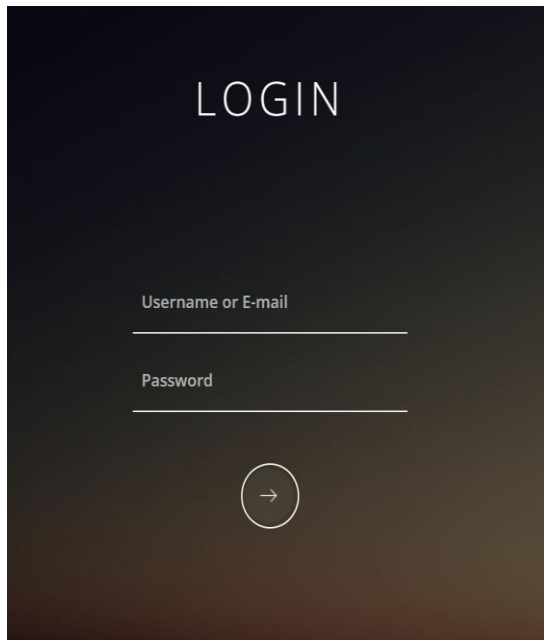
Bullshield values user feedback and actively solicits input through its review and rating system. Users have the opportunity to share their experiences and provide feedback on the app's functionality, user interface, and overall performance. This feedback helps inform future updates and improvements, ensuring that Bullshield continues to meet the evolving needs of its user base.

Complaint Management:

In addition to reviews and ratings, Bullshield provides users with a platform to voice their concerns and report any issues they encounter. Whether it's regarding inappropriate content, privacy violations, or other grievances, users can submit complaints through the app's dedicated support channels. Bullshield takes these complaints seriously and strives to address them promptly to maintain a safe and welcoming environment for all users.

Screenshot Prevention After Login:

To enhance user privacy and prevent unauthorized dissemination of sensitive information, Bullshield implements a security feature that disables screenshot functionality after the user has logged in. This ensures that users' interactions, messages, and content remain confidential and protected from unauthorized capture.



The admin functionalities of the BullShield application are integral to maintaining a safe and user-friendly environment within the platform. Here's an in-depth explanation of each feature:

1. **View Registered Users:** Admins have access to the list of all registered users on the platform. This feature allows them to monitor the user base, track growth, and analyze demographics. Additionally, it enables admins to identify and investigate any suspicious or problematic accounts.
2. **View Blocked Users:** This functionality allows admins to see a list of users who have been blocked from the platform. Admins can review the reasons for blocking these users and decide whether to uphold the block or lift it based on further investigation.
3. **View Ratings and Complaints:** Admins can access a dashboard or interface where they can view ratings and complaints submitted by users. Ratings may include feedback on user experience, content quality, or interactions with other users. Complaints could range from reports of importunity or bullying to enterprises about unhappy content. Admins can use this information to address issues promptly and maintain a positive user experience.
4. **Send Replies to Complaints:** Admins have the capability to respond to user complaints directly through the platform. This feature allows them to acknowledge users' concerns, provide assistance or clarification, and take appropriate action to resolve the issue. Clear and timely communication from admins helps to build trust and confidence among users.
5. **Block Users for Vulgar Content:** Admins have the authority to block users who engage in posting vulgar comments or content. This feature is essential for enforcing community guidelines and ensuring that the platform remains a respectful and safe environment for all users. Admins can review reported content and take swift action to maintain appropriate standards of conduct.

IV. Discussion

BullShield is a revolutionary application designed to create a safer and more respectful social media environment. With its main functionalities aimed at preventing unauthorized use of users' photos, filtering out vulgar or toxic comments, and blocking the sharing of offensive imagery, BullShield stands out as a guardian of online decency and privacy.

At the heart of BullShield lies its robust facial recognition technology, which ensures that no photo containing a user's face can be posted as a public post or story without their explicit permission. This feature not only safeguards users' identities but also empowers them with control over how their images are shared on the platform. By preventing unauthorized use of personal photos, BullShield mitigates the risk of identity theft, cyberbullying, and other forms of online abuse.

Furthermore, BullShield employs sophisticated algorithms to monitor and filter comments posted under users' content. Any comment deemed vulgar, offensive, or toxic is automatically flagged and withheld from publication. This proactive approach to content moderation fosters a more positive and respectful online community, where users can engage in meaningful discussions without fear of harassment or abuse.

In addition to safeguarding users' images and moderating comments, BullShield also prohibits the posting or sharing of vulgar or offensive photos. Leveraging image recognition technology, the app identifies and blocks the dissemination of inappropriate content, thereby maintaining a clean and safe online environment for all users.

V. Conclusion

Cyberbullying and vulgar content have become a significant concern in today's digital landscape, negatively impacting individuals and communities. This project, through its innovative combination of face recognition and content analysis, offers a powerful solution to tackle these issues. By effectively identifying faces and detecting harmful content, we can provide valuable tools for reporting incidents, protecting individuals, and promoting a more inclusive online sphere. This project stands at the forefront of creating a safer and more positive online experience for everyone. With further advancements in technology and continued collaboration, we can envision a future where cyberbullying and vulgar content become relics of the past, replaced by a virtual world built on respect, empathy, and constructive engagement. Let us commit ourselves to this vision and collectively work towards building a brighter and more inclusive digital future.

Acknowledgement


Every success stands as a testimony not only to the hardship but also to hearts behind it. Likewise, the present work has been undertaken and completed with direct and indirect help from many people and we would like to acknowledge the same. We wish to express our deep sense of gratitude to the project coordinator Ms. Shabna M, Hod and Asst professor, Department of Computer Science and Engineering, who coordinated in right path. Offering our thanks to Ms. Reshma M, Asst professor Department of Computer Science and Engineering, for her encouragement and guidance in carrying out the project.

References


- [1]. Teng, T. H., & Varathan, K. D. (2023). Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches. IEEE Access.
- [2]. Azeez, N. A., Idiakose, S. O., Onyema, C. J., & Vyver, C. V. D. (2021). Cyberbullying Detection In Social Networks: Artificial Intelligence Approach, (Department of Computer Science, University of Lagos, Nigeria).
- [3]. Ahmed, M. F., Zalish Mahmud, Biash, Z. T., Noor, A. A., Ashraf, F. B., & Arman Hossain. (2021). Detection Of Cyberbullying Incidents On Instagram Social Network, (Department of Computer Science and Engineering, Brac University).
- [4]. Shah, K., Phadtare, C., Rajpara, K., & Mehendale, N. (n.d.). Cyberbullying Detection For Hindi-English Language Using Machine Learning, (Department KJ Somaiya College of Engineering Mumbai).
- [5]. Srinivas, D., Umapathi, N., Karthick, G., Venkateswaran, N., & Jegadeesan, R. (n.d.). Rnn.Dea: Cyberbullying Detection In Social Media Platform (Twitter).
- [6]. Srinivas, D., Umapathi, N., Karthick, G., Venkateswaran, N., & Jegadeesan, R. (2023). Rnn-Dea: Cyberbullying Detection In Social Mediaplatform (Twitter). Research Paper.
- [7]. Rani, Ms. U., Ramesh, Mr. A., Srinivas, Mr. G., Ganesh, Mr. S., & Lakshmi, Mrs. D. V. (2023). Detection of Cyberbullying on Social Media. Journal of Engineering Sciences.

Author Information


Henna Maryam

 <https://orcid.org/0009-0002-8109-8179>
Mgm College of Engineering and Pharmaceutical
Sciences
Valiyaparambu , Edayur P.O , Valanchery,
Kerala,676552
India
Contact e-mail: hennamaryam02@gmail.com


Mahara M

 <https://orcid.org/0009-0007-3270-2671>
Mgm College of Engineering and Pharmaceutical
Sciences
Valiyaparambu , Edayur P.O , Valanchery,
Kerala,676552
India
Contact e-mail:mahara6100@gmail.com


Thanha Mariyam

 <https://orcid.org/0009-0009-5839-2442>
Mgm College of Engineering and Pharmaceutical
Sciences
Valiyaparambu , Edayur P.O , Valanchery,
Kerala,676552
India
Contact e-mail: thanhathanha665@gmail.com

Shifna V P

 <https://orcid.org/0009-0002-9046-1101>
Mgm College of Engineering and Pharmaceutical
Sciences
Valiyaparambu , Edayur P.O , Valanchery,
Kerala,676552
India
Contact e-mail: shifnavp123@gmail.com

Anjana N

 <https://orcid.org/0009-0002-3099-0650>
Mgm College of Engineering and Pharmaceutical
Sciences
Valiyaparambu , Edayur P.O , Valanchery,
Kerala,676552
India
Contact e-mail:2anjana7@gmail.com