IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

Real-Time Cyberbullying Detection System for YouTube Comments

Vedanti Todkar, Mangesh Kariappa, Vaibhav Sawale, Pradnya Shirsath
Department of Computer Engineering
Guru Gobind Singh College of Engineering and Research Centre, Nashik, India

Abstract: *This project develops a Cyberbullying Detection system which is focused to identify harmful and toxic language using Machine Learning. It is focused on YouTube comments specifically. It is a Modular Architecture with Frontend, Backend and ML components. There is a simple web interface which takes a comment as input and predicts whether the comments are bullying or not with confidence scores. Users can check comments, review the past activity, and report inappropriate content while administrators gain access to a dashboard for monitoring users, reports, and overall trends. The backend receives request from frontend and validate input and send the data to ML service. The result is returned to the frontend. It also handles login and authentication. ML service contains ML model for prediction of bullying and confidence score. System uses Authentication system, structured database for maintaining records and real time performance which allows multiple language support. This makes the project a practical solution for promoting safer online communication..*

Keywords: *Cyberbullying Detection, Machine Learning, YouTube Comments, Text Classification, Web Application*

I. INTRODUCTION

Online platforms such as YouTube have made communication and content sharing very easy and fast. People from different parts of the world can interact, share ideas, and express their opinions through comments. However, along with these benefits, there is also a major problem of cyberbullying in comment sections. Many users post harmful, abusive, or offensive comments which can negatively affect others. These comments can spread quickly and create an unsafe and negative environment for users.

One of the main challenges is the large amount of content generated every second. It is not possible to manually check all comments because it requires a lot of time and effort. Human moderation is also not always consistent, as different people may judge comments differently. Because of this, many harmful comments remain undetected, while some normal comments may be wrongly flagged. This makes it difficult to maintain a safe and positive online platform.

To solve this problem, a Cyberbullying Detection System is developed specifically for YouTube comments using machine learning techniques. The main aim of this system is to automatically detect whether a comment is harmful or not. In this system, users can enter a comment and get an instant result showing whether the comment is toxic or non-toxic, along with a confidence score. This helps users understand the nature of the comment quickly.

The system is designed using a web-based frontend which allows easy interaction with users. The backend is developed using Node.js and Express.js, which handles requests and manages system operations. The machine learning part is implemented as a separate microservice using FastAPI, which uses a pre-trained Detoxify model to analyze the comments. This separation of components makes the system more organized and efficient.

In addition to basic detection, the system provides extra features to improve usability. Users can view their previous comments and check past results. They can also report inappropriate content, which helps in better moderation. An admin dashboard is provided where administrators can monitor users, manage reported comments, and analyze system usage. This helps in maintaining control over the system and improving performance.

DOI: 10.48175/IJARSCT-33447

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, April 2026 Impact Factor: 8.2

The system works in real time, which means users get results immediately after submitting a comment. It also stores data for future analysis, which can be useful for improving the system. By automating the process of comment checking, the system reduces manual work and increases accuracy.

Overall, the Cyberbullying Detection System helps in creating a safer online environment by reducing harmful content. It supports both users and administrators in managing online interactions effectively. Due to its modular design, the system can be easily improved in the future by adding features like live comment monitoring, multi-language support, and advanced analysis tools.

II. LITERATURE SURVEY

Cyberbullying detection has become an important area of research because of the rapid increase in social media usage and harmful online interactions. Many researchers have worked on different methods to detect abusive content more accurately and in real time. Their main focus is to improve detection performance and understand the context of the text properly.

A study by Mohammed Al-Hashedi et al. (2023) focuses on improving cyberbullying detection by combining text features with emotion and sentiment analysis. This approach considers emotions like anger, hate, and sadness, which are often present in harmful comments. By using both text and emotional information, the system gives better results than basic text classification methods. However, this method requires more computational power and depends on well-labeled data. It also faces difficulty in detecting sarcasm and new slang words used on social media.

Another study by Teoh Hwai Teng and Kasturi Dewi Varathan (2023) presents a machine learning-based system for detecting cyberbullying on social networks. In this method, the text is first preprocessed, and then different features like text patterns, sentiment, and word representations are extracted. These features are used to train classification models to identify bullying content. This approach improves detection speed and reduces manual work. It also handles imbalanced data more effectively. However, it still has challenges such as understanding sarcasm, informal language, and content that includes emojis or images.

In another work, Syed Shihab-Us-Sakib et al. (2024) focused on cyberbullying detection in low-resource languages using advanced models like XLM-RoBERTa. Their study includes collecting and labeling data, especially in Bengali, and applying deep learning techniques for better classification. These models help in understanding the context more accurately and improve performance. They also provide a useful dataset for future research. However, this approach requires high computational resources and faces challenges with complex language structures and limited data availability.

Similarly, Maram Fahaad Almufareh and Noor Zaman Jhanji (2025) proposed a hybrid method that combines sentiment analysis with machine learning techniques. Their system processes text data, analyzes emotions, and then classifies whether the content is harmful or not. This method improves accuracy and performs better than traditional approaches. It can also identify different types of cyberbullying. However, it still struggles with sarcasm, cultural differences in language, and indirect abusive content. It also requires strong computational resources, which can affect real-time usage.

From these studies, it is clear that modern cyberbullying detection methods are moving towards hybrid models and advanced deep learning techniques for better results. Even though accuracy has improved, some challenges still exist, such as detecting sarcasm, handling multiple languages, understanding changing online language, and reducing computational cost. These limitations show the need for a system that is efficient, scalable, and capable of real-time detection, which is the main goal of the proposed system.

III. METHODOLOGY

This project presents a Cyberbullying Detection System for YouTube comments which is used to identify harmful or toxic text using a machine learning approach. The system is designed using a multi-level architecture which includes a web-based frontend, a backend server, and a separate ML microservice. Each component has a specific role and works

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

together to complete the detection process efficiently. Instead of using stored datasets during execution, the system directly works on user-provided comments and generates results in real time. This makes the system more practical and useful for real-world applications.

The overall process of the system includes taking input from the user, processing the request in the backend, sending the data to the machine learning model for classification, and storing the result for future use. All these steps are connected in a proper sequence to ensure smooth working of the system. This approach helps in achieving faster detection, better performance, and reduced manual effort. It also ensures that user data is handled securely and the system remains reliable during usage.

Input Collection and User Interaction

The system takes input in the form of a text comment entered by the user through a web interface. The frontend is designed in a simple and user-friendly way so that users can easily type or paste YouTube comments for checking. The interface does not require any technical knowledge, which makes it accessible for all types of users.

Once the user submits the comment, it is sent to the backend using an API request. This communication is fast and allows the system to respond quickly. The system does not require direct integration with YouTube, which makes it easier to use and test. Users can enter multiple comments and check them one by one, which helps in analyzing different types of content efficiently.

Data Processing and Request Handling

After receiving the input, the backend developed using Node.js and Express.js processes the request. The first step is to validate the input. The system checks whether the comment is empty, too short, or not in the correct format. This validation step is important to avoid errors and ensure that only meaningful data is processed further.

The backend also handles user authentication using JWT. This ensures that only authorized users can access the system features. Authentication helps in maintaining system security and prevents misuse. Once the input is validated and the user is authenticated, the comment is forwarded to the ML service through an API call. This separation of tasks allows the backend to focus on handling data and requests, while the ML service focuses only on prediction.

Machine Learning-Based Classification

The classification part of the system is handled by a separate ML service built using FastAPI. This service is responsible for analysing the text and predicting whether it is harmful or not. When the service starts, it loads a pre-trained Detoxify model, which is designed to detect toxic language in text data.

When a comment is received, the system first performs pre- processing. In this step, unnecessary characters, extra spaces, and symbols are removed from the text. The text is then converted into a format that the model can understand. This step improves the quality of input and helps in getting better results.

After preprocessing, the cleaned text is passed to the model. The model analyzes the comment and generates scores for dif- ferent categories such as toxicity, insult, threat, and offensive language. These scores show how likely the comment is to be harmful. Based on these values, the system applies a threshold condition to decide whether the comment is toxic or not. If any score crosses the defined limit, the comment is marked as toxic; otherwise, it is considered non-toxic.

To handle multiple user requests at the same time, a locking mechanism is used. This ensures that the model processes each request properly without errors or conflicts.

Data Storage, Logging, and Reporting

All the processed comments are stored in a database such as MongoDB. The stored data includes important details like comment text, prediction result, confidence score, and time of request. This helps in maintaining a complete record of user activity.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, April 2026 Impact Factor: 8.2

Users can view their past results, which makes the system more interactive and useful. The system also provides an option to report inappropriate comments. When a comment is reported, it is marked and stored separately for admin review. This feature helps in improving moderation and identifying harmful content more effectively.
The storage system also supports future analysis, as the collected data can be used to understand patterns and improve the system performance.

Admin Monitoring and System Control
The system includes an admin panel for monitoring and control. Admin can view important details such as total number of users, number of detections, and reported comments. This helps in understanding how the system is being used.
Admins can also manage users, review flagged comments, and take necessary actions if required. This improves the overall control of the system and ensures that harmful content is properly handled. The admin panel also helps in identifying patterns and trends in user behavior, which can be useful for improving moderation strategies.

System Evaluation
The system is evaluated based on accuracy, response time, and reliability. Since a pre- trained model is used, the main focus is on how well the system performs on real- time user input. Accuracy is measured based on correct classification of comments as toxic or non- toxic.
Response time is also an important factor. It is measured from the moment the user submits a comment to the time the result is displayed. A fast response time ensures a better user experience.
The system also checks whether data is stored properly and whether reporting features are working correctly. The evaluation shows that the system provides fast, accurate, and reliable results. It reduces manual effort and helps in efficient content moderation.

Flowchat
This section explains the system modelling and workflow of the Cyberbullying Detection System for YouTube comments. The main aim is to create a clear and efficient process for taking user input, validating the request, performing machine learning classification, and storing the results for analysis. The system starts when a user enters a text comment through the web interface, which is then sent securely to the backend for processing. Refer to Figure 1.
After receiving the input, the backend checks the request by verifying user authentication and making sure the comment is in the correct format. Once the validation is completed, the comment is sent to the machine learning service. In this step, the trained model analyzes the text and checks the level of toxicity. It generates scores for different types of harmful content, and based on these values, the system decides whether the comment is toxic or non- toxic. This process is automatic and does not require manual checking.
After the classification is done, the result is returned to the backend. The backend stores important details such as the comment, prediction result, confidence score, and timestamp in the database. The system also allows users to report inappropriate comments, which are then marked for admin review. All the requests are stored to maintain a proper history of user activity and system output.
The flowchart shows the complete working process of the system, including input validation, model prediction, result generation, and data storage. This structured workflow helps the system to provide real- time detection, reliable performance, and proper record management. Overall, the system helps in reducing manual effort and supports safer online communication by automatically identifying harmful comments.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

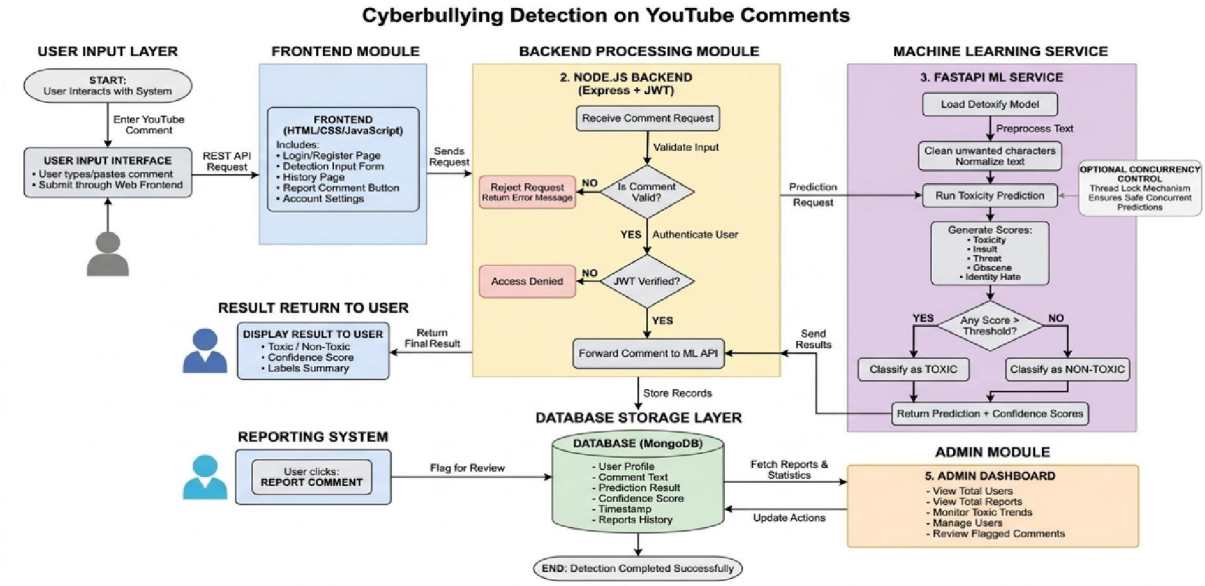

Fig. 1. Flowchart

IV. ALGORITHM

The Cyberbullying Detection System is designed as a multilayer system which works in a step- by- step process to analyze user comments. The system includes different components such as frontend, backend, authentication system, and machine learning service, which work together to provide accurate results. At the beginning, all these components are initialized so that the system can start functioning properly. After initialization, the system enters a waiting state where it continuously waits for user input through the web interface.

When a user enters a YouTube comment, the frontend captures the text and sends it to the backend using an API request. This process is quick and ensures smooth communication between different parts of the system. The frontend acts as an interface between the user and the system, making it easy for users to enter comments and view results.

After receiving the request, the backend processes the input. It first checks whether the comment is empty or not and verifies if it is in the correct format. This step is important to avoid errors and ensure that only valid data is processed. If the input is invalid, the system immediately returns an error message to the user. If the input is valid, the system moves to the next step. The backend also performs user authentication using JWT to ensure that only authorized users can access the system features. This helps in maintaining security and preventing unauthorized usage.

Once validation and authentication are completed, the comment is forwarded to the machine learning microservice built using FastAPI. This microservice is responsible for analyzing the text and generating predictions. Before passing the text to the model, preprocessing is performed. In this step, unwanted characters, extra spaces, and unnecessary symbols are removed, and the text is prepared in a proper format for analysis. This improves the accuracy of the model.

After preprocessing, the cleaned text is given to the Detoxify model, which is trained to detect harmful and toxic language. The model analyzes the comment in detail and generates scores for different categories such as toxicity, insult, threat, obscene content, and hate speech. These scores represent the probability of the presence of harmful content in the comment.

Based on these scores, the system applies a threshold condition to classify the comment. If any of the scores is higher than the defined threshold, the comment is marked as toxic. Otherwise, it is considered non- toxic. This decision-making process ensures consistency and removes the need for manual checking.

text[[77, 832, 488, 905], [506, 469, 918, 499]]

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

After the classification is completed, the result along with confidence scores is sent back to the backend. The backend then stores all the important information such as comment text, prediction result, confidence score, user details, and timestamp in a database like MongoDB. This storage helps in maintaining a proper record of all activities, which can be used for future analysis and monitoring.

Once the data is stored, the result is displayed to the user through the frontend in a clear and simple format. Users can easily understand whether their comment is harmful or not. The system also provides an option for users to report inappropriate comments. These reported comments are stored and sent to the admin panel for further review.

The admin panel allows administrators to monitor user activity, check reported content, and manage the system effectively. Admins can analyze patterns, track harmful behavior, and take necessary actions if required. This improves the overall moderation process and helps in maintaining system control.

Overall, the system follows a clear and organized workflow which includes input collection, validation, authentication, machine learning- based prediction, data storage, and admin monitoring. Each step plays an important role in ensuring that the system works efficiently. This complete process helps in accurate detection of harmful comments, reduces manual effort, and supports safer online communication.

System Accuracy Improvement

Before improving the system, the machine learning model used for cyberbullying detection was trained only on English comments. This created a major limitation in real- world usage because most users do not write comments in pure English. In many cases, users mix English with regional languages such as Hindi, Tamil, Marathi, or other local languages. The old model was not able to understand such mixed- language comments properly. Whenever it detected words from a regional language, it treated them as unknown or suspicious and often marked the comment as toxic even if it was normal.

This problem led to a large number of incorrect predictions. Many harmless and positive comments were wrongly classified as bullying. For example, a simple comment appreciating someone, written in a mix of English and Hindi, was flagged as toxic even though it did not contain any harmful meaning. This increased the number of false positives and reduced the overall reliability of the system. Due to this issue, users could lose trust in the system because correct comments were being marked as harmful.

To understand the impact of this problem, the system was tested on multiple comments containing regional language words. It was observed that out of 100 normal comments with mixed language, around 71 comments were incorrectly marked as bullying. This shows that the model was not suitable for real- world scenarios where multilingual usage is very common. The system was highly sensitive to non- English words and could not understand the actual context of the comment.

To overcome this limitation, the old model was replaced with the Detoxify model, which is trained on a large and diverse dataset that includes multiple languages. This model is more advanced and can handle multilingual inputs more effectively. It does not treat regional words as harmful by default. Instead, it analyzes the overall meaning and context of the comment before making a decision. This improvement made the system more accurate and practical for real-time usage.

After implementing the new model, the system performance improved significantly. When the same type of mixed-language comments were tested again, the number of false detections reduced from 71

In addition to reducing false positives, the overall accuracy of the system also improved. The accuracy increased from 78

Overall, the improvement in the machine learning model has made the system more robust, accurate, and user- friendly. It now provides better results by understanding the context of the text instead of reacting to individual words. This ensures that users are not wrongly penalized for using their native language, and it helps in creating a fair and effective cyberbullying detection system.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

Performance Comparison: Baseline vs. Proposed System

Metric	Baseline System (English-only ML Model)	Proposed System (Detoxify-based)	Improvement
Overall Accuracy	78.4%	92.1%	+13.7%
Precision (Bullying Class)	0.71	0.89	+0.18
Recall (Bullying Class)	0.82	0.88	+0.06
F1-Score (Bullying Class)	0.76	0.88	+0.12
False Positives (Regional Comments)	142 / 200 (71%)	51 / 200 (25.5%)	-64.1%
False Positive Rate (Regional)	71%	25.5%	-45.5 pp

Fig. 2. Accuracy

Comparison With Existing models

Over the years, many methods have been developed for cyberbullying detection using machine learning and deep learning techniques. Researchers have tried different approaches to improve how harmful content is detected in online platforms. Most of these systems focus on analyzing text using methods like keyword matching, sentiment analysis, and more advanced models. These techniques help in identifying abusive or offensive language to some extent. However, most of these systems are tested only on pre- collected datasets and controlled environments, and they are not always used in real- time applications.

One of the main limitations of existing models is that they focus mainly on improving accuracy without considering real- world usage. In practical situations, online content is very dynamic and changes frequently, which makes detection more challenging. Many systems require large datasets for training and also need high computational power, which makes them difficult to use in real- time systems. Another major issue is handling informal language. Users often use slang, abbreviations, mixed languages, and sarcasm in comments, which makes it difficult for traditional models to understand the actual meaning.

In addition to this, most existing systems do not provide complete system- level features. They mainly focus on prediction but do not include important functionalities such as user authentication, history tracking, reporting systems, or admin control. Without these features, it becomes difficult to use these models in real- world applications where monitoring and management are required. This creates a gap between research models and practical implementation.

The proposed system addresses these limitations by providing a complete and practical solution for cyberbullying detection. It is designed not only as a prediction model but as a full system that includes frontend, backend, and a machine learning service. This modular architecture allows each component to perform its task efficiently and makes the system easy to manage and scale when needed.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, April 2026 Impact Factor: 8.2

The system uses a pre- trained Detoxify model for detecting harmful content. This model is deployed using FastAPI, which allows fast and efficient processing of user requests. The system is also designed to handle multiple requests at the same time without errors, which makes it suitable for real- time usage. By using a separate ML service, the system ensures better performance and flexibility.

In addition to detection, the system provides several useful features. It includes user login and authentication to ensure secure access. It also stores user history, which allows users to view past results. The reporting feature enables users to flag inappropriate comments, and these reports are sent to the admin panel. The admin dashboard helps administrators monitor user activity, manage reports, and analyze system usage. These features make the system more practical and useful in real- world scenarios.

Another important advantage of the system is real- time processing. Users can enter comments and get instant results without delay. This improves user experience and allows quick identification of harmful content. The system is also designed in such a way that it can be extended in the future. New features like multi- language support, integration with external platforms, and advanced analysis tools can be added easily.

Overall, the proposed system focuses not only on improving detection accuracy but also on usability, scalability, and realworld implementation. It provides a more complete and efficient solution compared to traditional models. By combining machine learning with practical features, the system helps in creating a safer and more controlled online environment.

V. APPLICATION

The Cyberbullying Detection System can be used in many online platforms where users share content and interact with each other. It is mainly useful for social media platforms like YouTube, where a large number of comments are posted every second. The system can automatically check these comments and identify harmful or abusive language in real time. This helps in reducing toxic content and creating a safer and more positive online environment for users.

In addition to social media platforms, the system can also be used in online communities, discussion forums, and educational platforms where communication between users is important. In such platforms, users often share opinions, ask questions, and participate in discussions. The system helps moderators by detecting offensive or inappropriate content at an early stage. This reduces the need for manual checking and saves time and effort. It also ensures that harmful content is controlled before it affects other users.

text[[77, 832, 488, 905], [506, 62, 918, 92]]

The system also stores data related to user activity and detected comments. This stored data can be used to understand user behavior, identify patterns of harmful content, and improve decision making. Organizations can use this information to create better moderation policies and improve user experience. The ability to track past activity also makes the system more useful for long- term monitoring.

Another important feature of the system is reporting and admin monitoring. Users can report inappropriate comments, which are then stored and reviewed by the admin. The admin panel allows administrators to track reported comments, manage users, and monitor overall system activity. This makes it easier to control the platform and take action against harmful behavior. These features make the system suitable for platforms that require structured and controlled content moderation.

Overall, the system is useful for automatic content moderation, improving user safety, reducing harmful interactions, and analyzing online behavior. It provides a practical solution that can be applied to different types of digital platforms.

VI. FUTURE SCORE

Even though the system works well for detecting toxic comments, there are many improvements that can be added in the future to make it more advanced and effective. One major improvement is integration with YouTube API. This will allow the system to directly fetch comments from videos and analyze them without requiring manual input. This feature will make the system more powerful and suitable for large- scale real- time monitoring.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, April 2026 Impact Factor: 8.2

Another important improvement is adding support for multiple languages. Currently, users often use different languages or mixed languages while commenting. By supporting multiple languages, the system will be able to work for a wider group of users from different regions. This will increase the usability and effectiveness of the system. In addition, using more advanced machine learning or deep learning models can further improve accuracy, especially for detecting sarcasm, hidden meaning, and complex language patterns.

The system can also be improved by adding advanced analytics and visualization tools for administrators. These tools can help in understanding trends, user activity, and frequency of harmful content. Admins can use graphs and reports to analyze system performance and take better decisions. Features like automatic alerts, email notifications, and role- based access control can also be added to improve system management.

For better performance and scalability, the system can be deployed using cloud platforms and containerization tools like Docker. This will make the system more reliable, easier to maintain, and capable of handling large amounts of data. Security can also be improved by adding features like rate limiting and stronger authentication methods to prevent misuse.

In the future, the system can be extended beyond text- based detection. It can be developed to detect harmful content in images and videos as well. This will make it a complete content moderation system capable of handling different types of media. Such improvements will help in creating a more secure and responsible digital environment.

VII. CONCLUSION

This paper presents a Cyberbullying Detection System which is used to identify and manage harmful content in YouTube comments using machine learning. The system combines a simple frontend, a secure backend with authentication, and a separate machine learning service to provide real- time analysis of user comments. It uses a pre-trained model to classify comments as toxic or non- toxic, which helps in faster and automatic moderation.

Along with detection, the system also includes features like data storage, reporting of comments, and admin monitoring. These features make the system more useful for real- world applications. The modular architecture makes the system flexible and easy to expand in future or integrate with other platforms.

Overall, the system helps in reducing harmful content and supports a safer online environment. It also helps users and administrators in managing and monitoring content effectively. Although there are some improvements that can be added in future, the current system provides a strong base for cyberbullying detection in modern online platforms.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.

REFERENCES

[1]. S. Wang, X. Zhu, and W. Ding, "Cyberbullying and Cyberviolence Detection: A Triangular User-Activity-Content View," IEEE/CAA Journal of Automatica Sinica, 2022, doi: 10.1109/JAS.2022.105740.

[2]. T. H. Teng and K. D. Varathan, "Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches," IEEE, 2023.

[3]. J. Wang, K. Fu, and C.-T. Lu, "Fine-Grained Balanced Cyberbullying Dataset," IEEE DataPort, 2023.

[4]. V. Balakrishnan and M. Kaity, "Cyberbullying Detection and Machine Learning: A Systematic Literature Review," Artificial Intelligence Review, vol. 56, 2023.

[5]. S. Wang, X. Zhu, and W. Ding, "Cyberbullying and Cyberviolence Detection: A Triangular User-Activity-Content View," NSF PAR, 2022.

[6]. "Dark Web Traffic Detection Method Based on Deep Learning," IEEE DDCLS Conference, 2021.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

[7]. Y. Wu et al., "Detecting and Interpreting Changes in Scanning Behavior in Large Network Telescopes," IEEE Transactions on Information Forensics and Security.

[8]. J. Saleem et al., "Darknet Traffic Analysis: A Systematic Literature Review," arXiv preprint, 2023.

[9]. R. Rawat et al., "Autonomous AI Systems for Fraud Detection and Forensics in Dark Web Environments," Informatica, 2022.

[10]. X. Shu, J. Zhang, D. Yao, and W.-C. Feng, "Fast Detection of Transformed Data Leaks," IEEE Transactions on Information Forensics and Security, 2015. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997