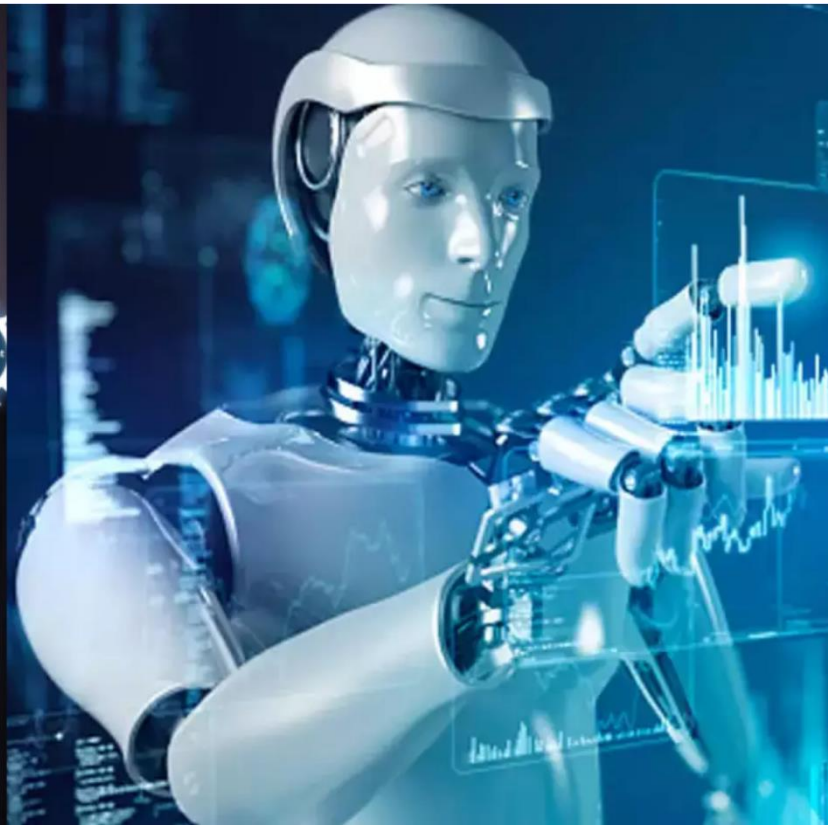
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503053|

Multilingual AI Framework for Cyberbullying Detection and Online Safety

Mrs. T Swathi, J[1]. Srija, K[2]. Hari Priya[2], S. Charan[2], S. Nikhil[2]

Assistant Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India[1]

IV B.Tech, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India[2]

ABSTRACT: The Multilingual AI Framework for Cyberbullying Detection and Online Safety is designed to identify harmful and abusive text in online communication platforms. Cyberbullying has become a major issue on social media, chat applications, and online forums, affecting users emotionally and psychologically. This system uses Natural Language Processing (NLP) and Machine Learning techniques to detect cyberbullying accurately. It supports multiple languages including English, Telugu, and code-mixed text, making it suitable for diverse users. The input text is preprocessed and converted into numerical features using TF-IDF. A Linear Support Vector Machine (Linear SVM) model is used to classify messages as Cyberbullying or Non-Bullying. The system provides real-time detection and generates alerts when harmful content is identified. It also maintains logs and statistics for monitoring purposes. The framework is scalable, efficient, and easy to integrate with web applications. Overall, the system helps promote safe and respectful online communication.

I. INTRODUCTION

The rapid expansion of social media and online platforms has led to a significant rise in cyberbullying and digital harassment. Monitoring harmful content manually is difficult due to the large volume of multilingual and code-mixed text shared every day. Traditional systems mainly rely on keyword filtering and single-language models, which often fail to understand context and sarcasm.

This project proposes a Multilingual AI Framework for Cyberbullying Detection using Natural Language Processing and machine learning techniques. The system uses TF-IDF for feature extraction and Linear SVM for classification. It provides real-time detection of harmful messages to improve online safety and promote respectful communication.

II. EXISTING SYSTEM

Current systems for detecting online bullying are weak. They only look for specific words and don't understand context or slang. They also only work for one language and can't keep up with new words people invent. This makes them often wrong and not very helpful for keeping people safe online. several challenges persist:

1. Focus on only one or two languages**:** Most systems support limited languages, making them ineffective in multilingual and code-mixed communication environments.
2. Low accuracy in detecting slang, sarcasm, and indirect bullying: They fail to understand context and hidden meanings, leading to incorrect or missed detections.
3. Limited adaptability across different platforms: Existing models are often designed for specific platforms and do not perform well in varied online environments.
4. Inability to provide real-time detection and response: Many systems analyze data after posting, which delays action against harmful content.

III. PROPOSED SYSTEM

The proposed system is a Multilingual AI-based Cyberbullying Detection Framework designed to identify harmful and abusive text in online platforms. It uses Natural Language Processing (NLP) with Machine Learning models like TF-IDF and Linear SVM for accurate classification. The system provides real-time detection and supports multiple languages including regional and code-mixed text.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503053|

ADVANTAGES OF THE PROPOSED SYSTEM

1. **Multilingual Support:** The system can detect cyberbullying in English, regional languages, and code-mixed text. This makes it suitable for diverse and multilingual online users.
2. **Higher Accuracy:** It uses context-aware machine learning models instead of simple keyword matching. This reduces incorrect predictions and improves detection performance.s.
3. **Real-Time Detection:** Messages are analyzed instantly after submission. Harmful content triggers immediate alerts or warnings.
4. **Platform Adaptability:** The framework can be integrated with social media, chat applications, and online forums. This ensures flexibility across different digital platforms.
5. **Improved Online Safety:** Early detection of abusive messages helps prevent harassment. It promotes respectful and healthy online communication.
6. **Scalability:** The system can handle large volumes of text data efficiently. It can be expanded to support more users and platforms in the future.

IV. METHODOLOGY

4.1 Data Collection

Gather multilingual text data related to cyberbullying detection from social media platforms, chat applications, and publicly available datasets. The data includes input features (user messages or comments) and labels (Bullying or Non-Bullying). Datasets are collected from sources such as Kaggle and other open repositories.

o English Cyberbullying Dataset – Kaggle

o Social Media Comments Dataset – Public sources

4.2 Data Preprocessing

Clean the collected text data to remove noise such as special characters, URLs, emojis, and extra spaces. Convert text to lowercase and perform tokenization to break sentences into words. Stopwords are removed to improve model efficiency.

Datasets may be imbalanced, so techniques like SMOTE or oversampling are applied to balance bullying and non-bullying classes. Missing or inconsistent entries are handled appropriately, and text is transformed into numerical form using TF-IDF for further processing.

4.3 Model Training

This step involves selecting an appropriate machine learning algorithm and training it on the processed data. The dataset is split into two parts – training data and testing data (based on required proportion). The training data is used to build the model using Linear SVM, while the testing data is used to evaluate performance. The trained model learns patterns to classify messages as Cyberbullying or Non-Bullying accurately.

V. SYSTEM ARCHITECTURE

System architecture is the high-level design or blueprint of a system. It defines how different components are structured and organized. It explains how modules interact and communicate with each other. The architecture shows how data flows from input to processing and finally to output. It includes elements such as user interface, processing units, databases, and communication layers.

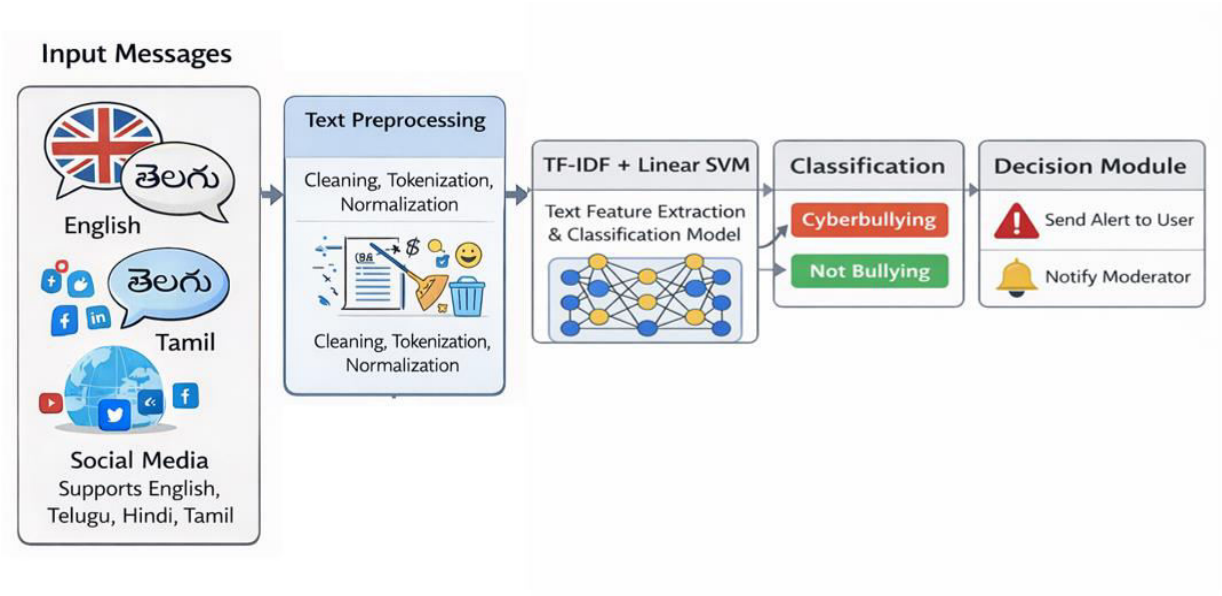
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503053|



System architecture helps developers understand the overall working of the system before implementation. It ensures that the system is scalable, reliable, and efficient. A well-designed architecture improves performance and simplifies maintenance. It also supports future upgrades and integration with other systems. Overall, system architecture provides a clear and structured framework for building complex applications.

VI. RESULTS AND OUTPUT

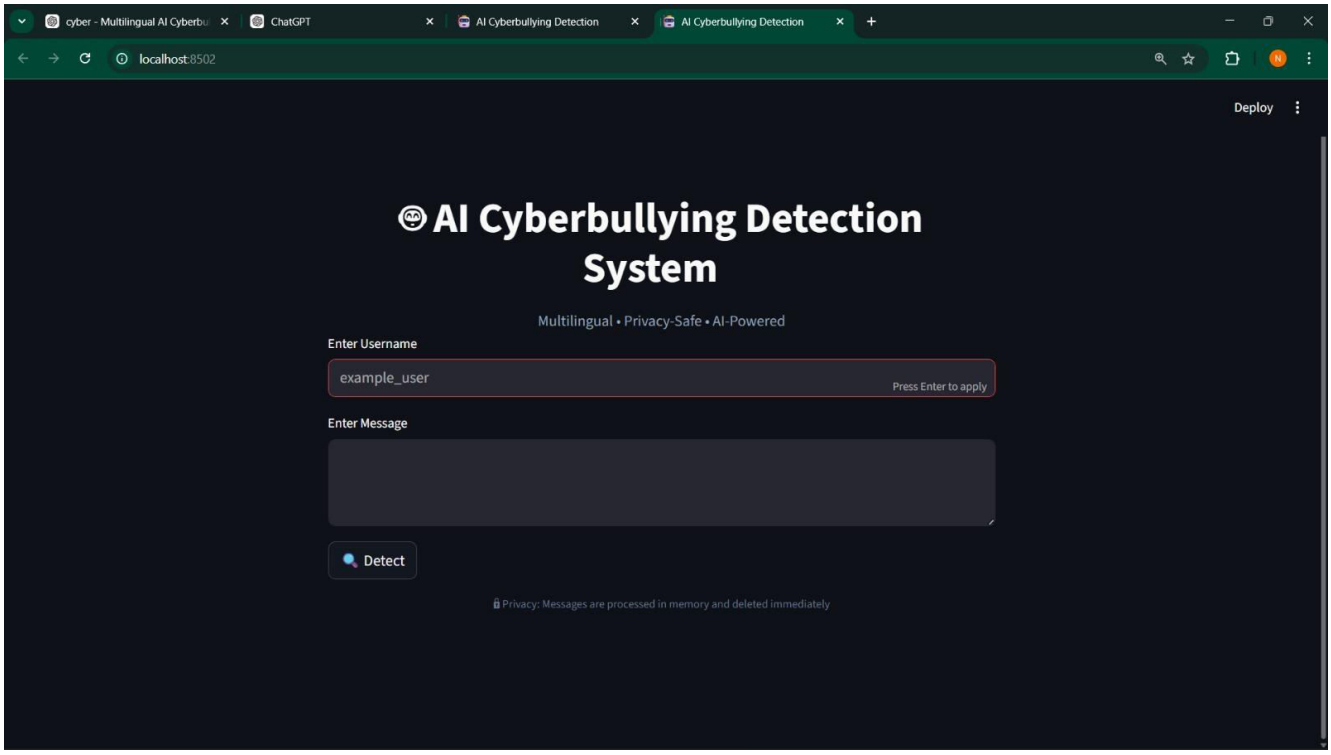


Fig: User Interface

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503053|

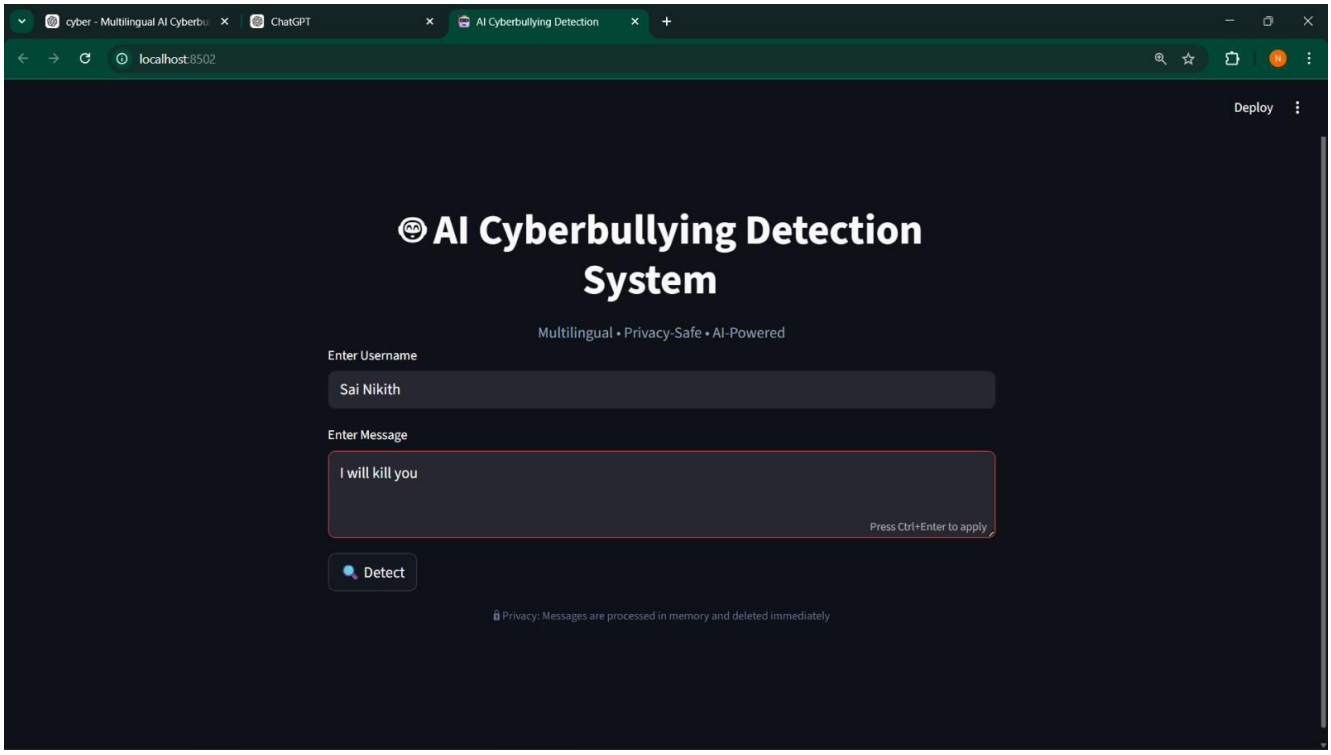


Fig: Enter input value

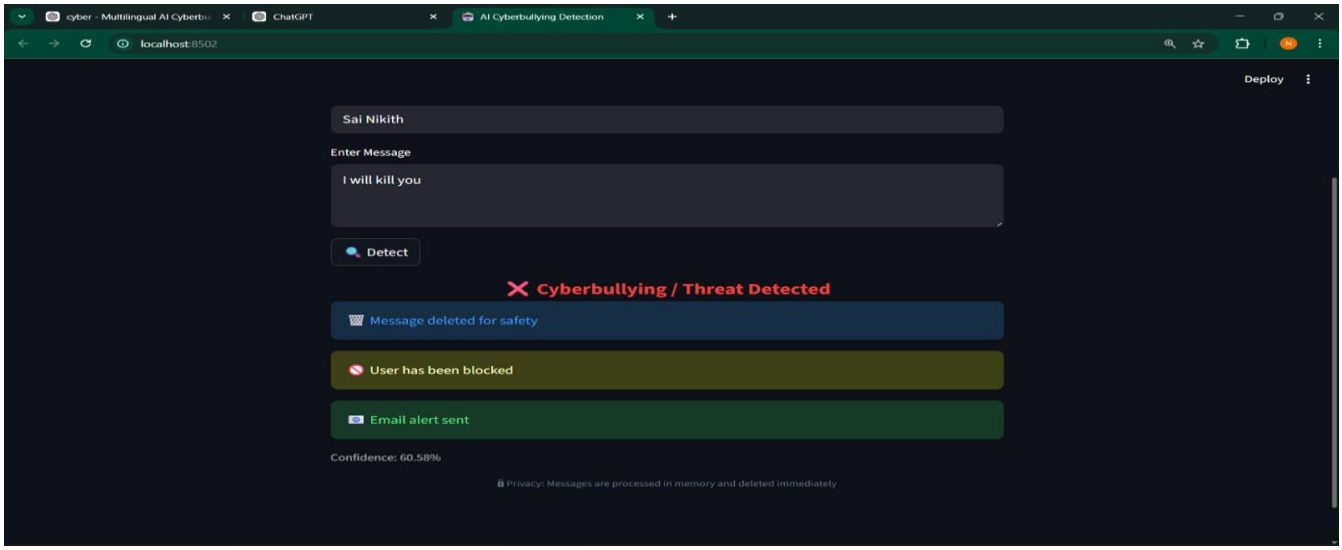


Fig: Cyberbullying Detection System Output Interface

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503053|

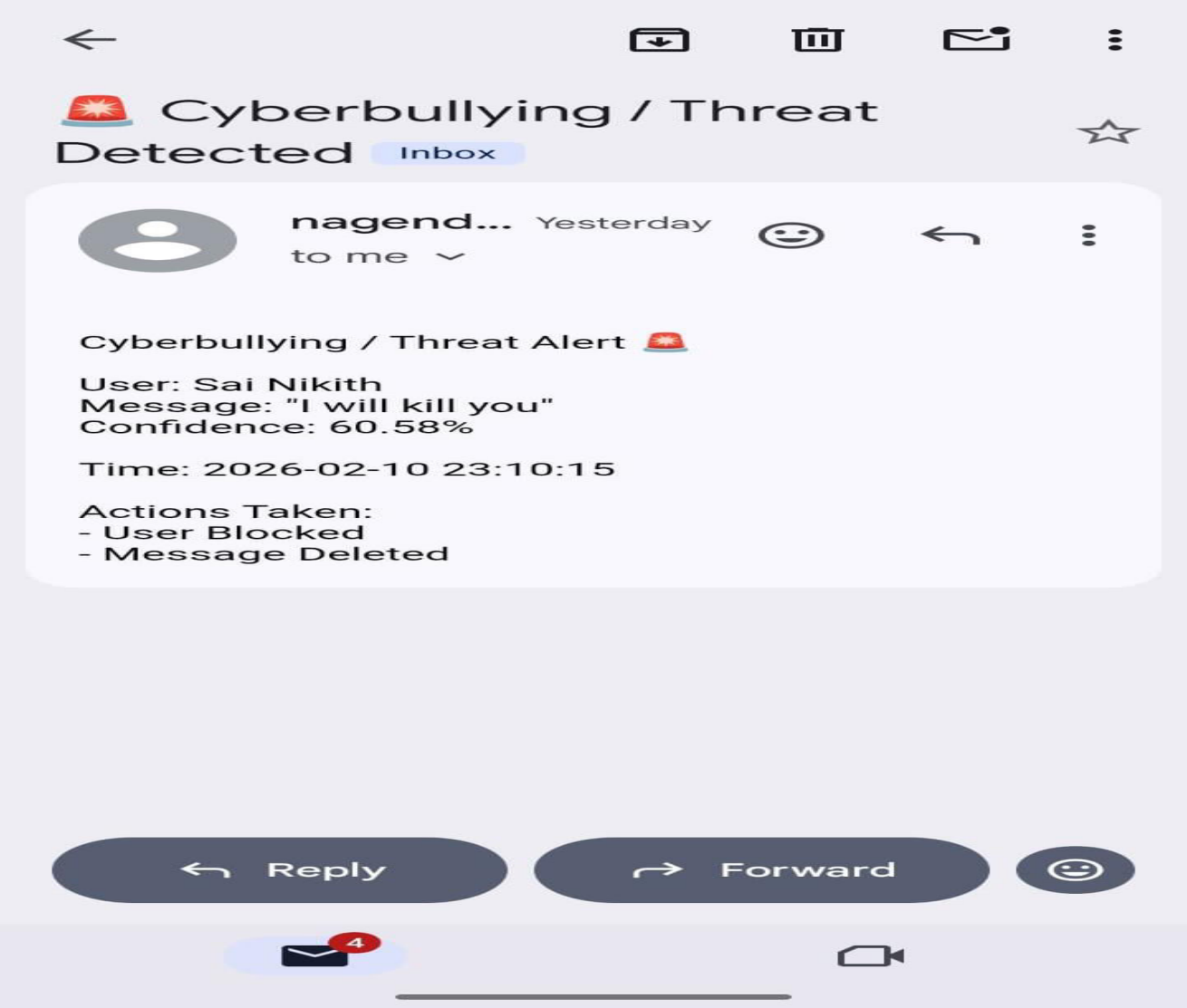


Fig: Email Notification for Detected Cyberbullying Threat

VII. CONCLUSION

The proposed Multilingual Cyberbullying Detection System provides a reliable and efficient approach for identifying harmful and abusive text in online communication. The system is developed using Natural Language Processing techniques, where the input text is processed through data preprocessing, feature extraction using TF-IDF, and classification using a Linear Support Vector Machine (Linear SVM). This combination ensures accurate detection with faster processing time and low computational complexity.

One of the major strengths of the system is its ability to handle multilingual and mixed-language inputs. Users can provide messages in any language, and the system processes them using a unified model without requiring separate language-specific models. This makes the system more practical and suitable for real-world environments where users communicate in multiple languages.

The system is designed to work in real time, allowing immediate classification of messages as Cyberbullying or Non-Bullying. When harmful content is detected, the system generates alerts or warnings, which can help moderators or

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503053|

administrators take necessary action. This helps in reducing the spread of abusive content and promotes a safer digital environment.

The architecture is modular, simple, and scalable, making it easy to integrate into web applications, social media platforms, online forums, or chat systems. The use of efficient machine learning techniques ensures good performance even with large volumes of text data.

Overall, the proposed system contributes to improving online safety by detecting and controlling cyberbullying effectively. It encourages responsible communication and supports the creation of a healthy and secure online community.

VIII. FUTURE SCOPE

1. Advanced Deep Learning Models
Implementing models like BERT or LSTM can improve the system's ability to understand context and complex language patterns. This will increase detection accuracy and reduce false predictions.
2. Automatic Language Detection
Adding a language identification module can automatically detect the input language. This helps apply suitable preprocessing techniques for better performance**.**
3. Multimedia Cyberbullying Detection
The system can be extended to detect harmful content in images, audio, and videos. This will make the framework more comprehensive and effective.
4. Real-Time Social Media Integration
Direct integration with social media platforms can enable continuous monitoring of user content. It ensures instant detection and moderation of harmful messages.
5. Cloud-Based Deployment and Scalability
Deploying the system on cloud infrastructure allows handling large-scale user data efficiently. It also improves system performance and availability.

REFERENCES

1. P. Nandhini and J. Sheeba, "Online Social Network Bullying Detection Using Intelligence Techniques," Procedia Computer Science, vol. 45, pp. 485–492, 2015.
2. S. Agrawal and A. Awekar, "Deep Learning for Detecting Cyberbullying Across Multiple Social Media Platforms," in Proc. European Conference on Information Retrieval (ECIR), 2018, pp. 141–153.
3. R. Rosa et al., "Automatic Cyberbullying Detection: A Systematic Review," Computers in Human Behavior Reports, vol. 2, 2020.
4. Z. Zhang, A. Robinson, and J. Tepper, "Cyberbullying Detection with Word Embedding and Neural Networks," IEEE International Conference on Advances in Computing, Communications and Informatics, 2016.
5. A. K. Jain and B. Kumar, "Natural Language Processing for Multilingual Text Classification," IEEE Access, vol. 8, pp. 13682–13692, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details