# Cyberbullying Detection And Prevention In Social Network

**Mrs. I. Joy Sinthia M.E[1], Sham Kumar T[2], Raghul R[3], Ramamoorthy K[4], Sudharsan S[5]**
[1]HOD, Dept of Computer Science and Engineering
[2, 3, 4, 5] Dept of Computer Science and Engineering
[1, 2, 3, 4, 5] M.A.M College of Engineering ,Tiruchirappalli, Tamil Nadu, India.

***Abstract-*** *The rapid growth of social media platforms has significantly increased the spread of harmful online content, including cyberbullying and hate speech. These forms of communication negatively impact individuals and communities, often leading to psychological distress and social conflicts. Existing content filtering systems are limited in their ability to effectively detect and prevent such behavior in real time. This paper proposes an intelligent cyberbullying detection and prevention system using Natural Language Processing (NLP) and the VADER sentiment analysis technique. The system analyzes user-generated content, classifies sentiments, and filters offensive messages based on predefined rules. A blacklist mechanism is implemented to identify repeat offenders, while real-time alerts notify users and administrators of harmful activity. The proposed system enhances online safety by providing an adaptive, efficient, and user-friendly solution for managing social media interactions.*

***Keywords:*** Cyberbullying, NLP, VADER, Sentiment Analysis, Machine Learning, Social Media, Hate Speech Detection

## I. INTRODUCTION

Social media platforms have transformed communication by enabling users to share ideas and interact globally. However, this rapid expansion has also led to increased misuse in the form of cyberbullying and hate speech. These harmful activities target individuals based on personal attributes such as race, religion, gender, and nationality, leading to serious psychological and social consequences.

Traditional filtering systems are insufficient in handling dynamic and short-text content commonly found on social media. To address these challenges, advanced techniques such as Natural Language Processing (NLP) and sentiment analysis are used. This paper presents a system that leverages VADER-based sentiment analysis to detect and prevent cyberbullying in real time.

## II. RELATED WORK

Several studies have explored cyberbullying detection using machine learning and deep learning approaches. Supervised learning models such as Support Vector Machines and Random Forest have shown promising results in classifying offensive content. Deep learning models, including LSTM and transformer-based architectures like RoBERTa, provide improved contextual understanding.

Word embedding techniques such as Word2Vec, GloVe, and fastText have also been used to enhance text representation. Despite these advancements, many systems struggle with real-time detection, adaptability, and handling informal social media language.

## III. PROBLEM STATEMENT

The increasing volume of user-generated content on social media has made it difficult to monitor and control harmful interactions. Existing systems fail to:

- Detect cyberbullying in real time
- Adapt to user-specific filtering preferences
- Handle short and informal text effectively
- Provide immediate alerts to users

There is a need for an intelligent system capable of detecting, filtering, and preventing harmful content efficiently.

## IV. PROPOSED SYSTEM

The proposed system uses NLP and VADER sentiment analysis to classify user comments as positive, neutral, or negative. Negative content is filtered or blocked automatically.

A rule-based mechanism allows users to customize filtering criteria, while a blacklist system tracks repeat offenders. The system also provides real-time alerts to users and administrators.

Key Features:

- Sentiment classification using VADER
- Real-time comment processing
- Rule-based filtering
- Automatic blocking mechanism
- Alert and notification system

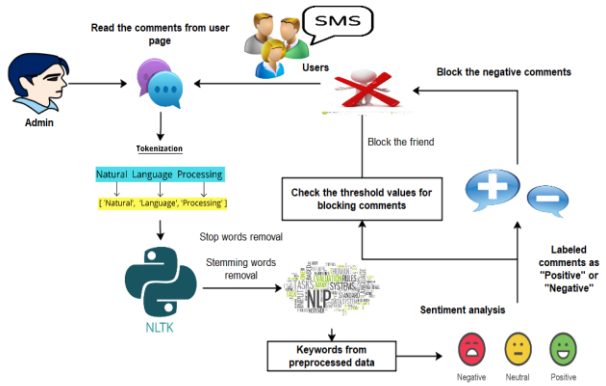


**Fig.1: System Architecture**

## V. SYSTEM REQUIREMENTS

The system requires minimal hardware and software resources:

- Processor: Dual-core 2.6 GHz
- RAM: 4 GB
- Software: Python, HTML, CSS, JavaScript
- Database: MySQL

The system is implemented using Flask for backend processing and integrates NLP libraries for text analysis.

## VI. IMPLEMENTATION AND RESULTS

The system processes user comments in real time and applies sentiment analysis to classify them. NLP techniques such as tokenization and stop-word removal are used for preprocessing.

The VADER algorithm calculates sentiment scores and determines whether content is harmful. The system successfully filters negative content and triggers alerts when necessary. Experimental results show improved accuracy in detecting cyberbullying compared to traditional filtering methods.


**Fig.2: Home Page/User Interface**

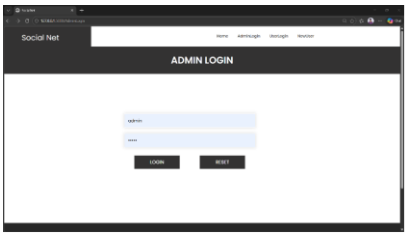


**Fig. 3: Admin Login Interface**

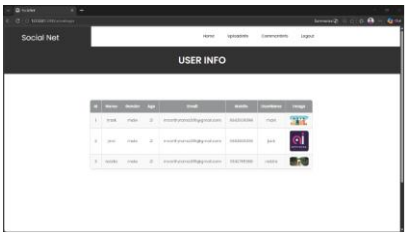


**Fig. 4: User Info**

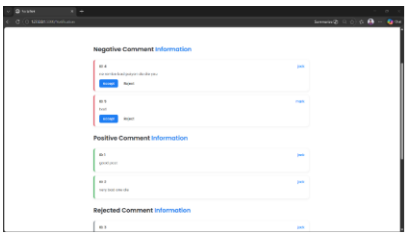


**Fig. 5: Alert Notification**

## VII. CONCLUSION

This paper presents an effective cyberbullying detection and prevention system using NLP and VADER. The system provides real-time filtering, customizable rules, and automated alerts, ensuring a safer online environment.

## VIII. FUTURE WORK

- Future improvements include:
- Integration of image and video analysis
- Use of advanced transformer models (BERT, GPT)
- Multi-language support
- Enhanced adaptive learning mechanisms

## REFERENCES

[1] B. R. Chakravarthi , "Detecting abusive comments," NLP Journal, 2023.
[2] A. Faraj , "Word embeddings analysis," UHD Journal, 2024.
[3] M. Hietanen ., "Hate speech definition," Journal of Communication, 2023.
[4] A. Jamjoom , "Robertanet model," IEEE Access, 2024.
[5] S. Kagi, "Cyberbullying detection," Journal of Technology, 2025.