ABSTRACT

The project "**Automatic Detection of Cyberbullying**" develops a machine learning-based web application for identifying bullying content in text data from social media and online platforms. By preprocessing text using techniques such as lemmatization, stopword removal, and punctuation elimination, the system utilizes multiple models—Logistic Regression, Random Forest, SVM, and Naive Bayes—to classify text as bullying or non-bullying.

Using the TF-IDF vectorizer for feature extraction, the best-performing model is selected based on accuracy and saved for future predictions. The application, built with Flask, allows users to input text and receive real-time bullying predictions, along with a performance overview of each model.

Future Enhancements include advanced model tuning, deep learning integration (e.g., LSTM, BERT), real-time social media integration, multilingual support, and a user feedback loop for continuous learning. The system will also incorporate sentiment and emotion analysis, offer interactive visualizations of model performance, and ensure privacy and fairness in the detection process. These updates will enhance the system's ability to detect cyberbullying more accurately and in real-time, providing valuable tools for online platforms to combat harmful content.