

School of Computer Science Engineering and Information Systems

Winter Semester 2024-2025

Department of Information Technology

BITE498J – Project – II

Review -2

Date: 20.3.2025

**Cyber Bullying Detection System**

**21BIT0705**

**Smaran Lal Shrestha**

**Under the Guidance of**

**Prof. BHAVANI S**

**Assistant Professor Sr. Grade 1**

**SCORE**

**Guide Signature with date**

**Internal Examiner -1 Internal Examiner-2**

**Signature Signature**

**ABSTRACT:**

Cyberbullying is an escalating issue in the digital world, and addressing it requires advanced detection systems to create safer online spaces. This project focuses on developing a comprehensive solution to detect cyberbullying through both text and image analysis. For text-based detection, a hybrid approach is used, combining rule-based filtering techniques such as predefined word lists and regex patterns to capture variations of offensive language. Additionally, transformer-based models like BERT are fine-tuned using cyberbullying datasets to understand context and detect subtle nuances. Data augmentation methods, including paraphrasing and misspelling variations, are incorporated to enhance the model's accuracy and robustness.

For image-based detection, the system employs computer vision techniques to analyze harmful visual content, such as memes or inappropriate images. Pre-trained deep learning models, like Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs), are fine-tuned to identify harmful content effectively. The system integrates both text and image detection into a unified pipeline, providing a scalable solution for online platforms. The goal of this project is to help mitigate the psychological impact of cyberbullying and foster a safer and healthier online environment.

**Keywords:** Cyber Bullying, Natural Language Processing (NLP), Image Recognition, Customized Alerts

**[1.] INTRODUCTION:**

Cyberbullying has become a widespread and pressing issue in the digital age, with the rise of social media, online communication platforms, and virtual interactions. It poses significant threats to individuals' mental health, self-esteem, and overall well-being. Detecting and mitigating cyberbullying is a challenging task due to the nuanced and context-dependent nature of harmful behavior, which can manifest through text, images, or even subtle patterns like sarcasm and slang.

Traditional detection methods often fail to capture the complexity of online interactions, where language varies across cultures, and visual content can carry implicit meanings. This project aims to address these challenges by developing a comprehensive cyberbullying detection system that leverages advanced Natural Language Processing (NLP), image recognition, and multilingual support to identify harmful content effectively. By integrating these technologies into a scalable and efficient pipeline, the system aspires to provide real-time detection, empowering users and moderators to create safer digital environments.

**[2.] PROBLEM STATEMENT:**

Cyberbullying has become a pervasive issue in the digital world, posing significant threats to individuals' mental health and well-being. Detecting and addressing cyberbullying is challenging due to its nuanced nature, with harmful content often disguised through slang, sarcasm, abbreviations, or images. Existing methods are either limited in scope, fail to capture contextual subtleties, or lack the ability to analyze multimodal content such as text and images simultaneously.

The absence of scalable, efficient, and context-aware detection systems leaves online platforms struggling to mitigate the psychological and social impacts of cyberbullying. This project aims to address these challenges by leveraging advanced machine learning and deep learning techniques to develop a comprehensive system capable of accurately detecting cyberbullying and promoting safer digital spaces.

**[3.] OBJECTIVES:**

The objective of our Cyberbullying Detection System is to develop a model that can accurately detect cyberbullying in online platforms using both text and image analysis. The goals are to:

1. Detect harmful text using NLP models like BERT, classifying content as abusive or non-abusive.
2. Analyze images for offensive content using computer vision techniques.
3. Analyze text present in the images
4. Provide real-time detection and timely alerts.
5. Generate customized alerts for users and moderators based on detected content.

This system aims to create safer online environments by efficiently identifying and addressing cyberbullying in both text and visual content.

**[4.] SCOPE OF THE PROJECT:**

The primary goal of this project is to develop a system for detecting cyberbullying across online platforms using text and image analysis. This research aims to explore the effectiveness of different machine learning techniques in accurately identifying cyberbullying content. By employing advanced Natural Language Processing (NLP) models and image recognition methods, the study will compare their performance in detecting harmful behavior in both text and images. Despite the growing advancements in AI, detecting cyberbullying remains a complex challenge due to the subtle and evolving nature of abusive language and visual content. The scope of this project is to contribute to the development of more accurate, real-time, and context-aware systems that can enhance online safety and provide actionable insights for users and moderators.

**[5.] PROPOSED SYSTEM:**

The proposed system aims to detect cyberbullying in real-time across both text and image content on online platforms using advanced machine learning techniques. It will utilize transformer-based models like BERT to identify harmful language in text and Convolutional Neural Networks (CNNs) for detecting offensive visuals, such as memes or inappropriate images. The system will also support multiple languages, making it applicable in diverse cultural contexts. By providing real-time detection and customized alerts, the system will allow immediate action to be taken by users, moderators, or parents, promoting a safer and more supportive online environment.

**[6.] LITERATURE SURVEY:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.NO** | **TITLE** | **MERITS** | **DEMERITS** |
| 1 | DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform | Combines Elman type RNN and Dolphin Echolocation Algorithm (DEA) for fine-tuning parameters and reducing training time. Outperforms existing models in accuracy, precision, recall, F1-score, and specificity. Effective for short text classification in social media. | The hybrid model’s complexity could be difficult for real-time implementation due to its dependence on fine-tuning parameters. The model's performance might be dataset-dependent and may not generalize well across diverse data sources. |
| 2 | Cyber Bullying Detection Using Machine Learning | The paper proposes a new representation learning method, SMSDA, which enhances the stacked denoising auto-encoder with semantic dropout noise and sparsity constraints to improve the detection of cyberbullying. It focuses on robust numerical representation learning for text messages, which is crucial for effective detection in social media. | The method might require further real-world validation and might struggle with the diversity of social media language, such as sarcasm, emojis, or indirect bullying. |
| 3 | A Comparative Study of Machine Learning Approaches for Cyberbullying Detection in Digital Forums | The paper evaluates multiple machine learning models (LR, SVM, RF, DT, NB, XG) for detecting cyberbullying and identifies Random Forest as the most effective model. It uses two datasets (10,000 and 20,000 comments) and explores tokenization, stop words, and Bag of Words for text preprocessing. | The system may face challenges with algorithm bias, data privacy concerns, and may require improvements to handle more diverse or complex forms of bullying. |
| 4 | Bilingual Cyber Bullying Detection System: Enhancing Online Safety | The system supports both English and Hindi (transliterated into English) for detecting cyberbullying. It uses advanced data extraction and preprocessing techniques, as well as machine learning algorithms. A robust web application is created using Python, Streamlit, Pandas, and Scikit-learn for real-time text analysis. | The system may face challenges with algorithm bias, data privacy concerns, and may require improvements to handle more diverse or complex forms of bullying. |
| 5 | Enhancing Cyber Bullying Detection Using Convolutional Neural Network | The paper demonstrates the application of deep learning (CNN) and NLP to detect cyberbullying in social media, specifically focusing on Twitter. It uses Twitter API to gather data and applies CNN to classify tweets as bullying or non-bullying. | The paper might face challenges like dataset quality, false positives, and the complexity of detecting subtle forms of bullying, especially in short, ambiguous text |
| 6 | Cyber Bullying and Toxicity Detection Using Machine Learning | The paper explores a novel approach for detecting  Cyberbullying and toxicity with machine learning algorithms such as SVM, Logistic Regression, Naive Bayes, KNN, and Random Forest, achieving 90% accuracy for text and 84.5% for images using MobileNetV2. | The paper might have limitations due to the dataset size and potential challenges in addressing all forms of cyberbullying or toxicity, especially in multilingual or mixed-content scenarios. |
| 7 | Automated Multimodal Detection and Reporting of Cyber Bullying using ML and Encryption | Uses machine learning techniques for detecting cyberbullying in text, photos, and videos with high accuracy (98.5%, 88.7%, and 80.7%). Implements  cryptographic techniques (RSA, SHA-256, AES) for data security and integrity. Aims to streamline reporting, creating a safer digital environment. | May face challenges in real-time detection across all types of media and require continuous updates to adapt to new forms of cyberbullying. |
| 8 | Cyberbullying Detection using Deep Learning Models in Bengali Language | The study focuses on detecting cyberbullying in the Bengali language, a low-resource language. It uses deep learning models like Bi-LSTM and CNN, and the CNN model outperforms the Bi-LSTM model in terms of precision, recall, and F1 score. The research uses a public dataset from Kaggle. | The study is limited to the Bengali language, and its application may not be generalized to other languages. |
| 9. | Cyber-Bullying Detection Via Text Mining and Machine Learning | This paper proposes the use of machine learning and natural language processing (NLP) for detecting cyberbullying in real-time before content is uploaded. The research highlights how real-time detection can reduce the circulation of harmful messages. | Limitations of current solutions and inefficient algorithms in detecting cyberbullying. It doesn't offer a concrete framework for deployment in all online environments. |
| 11 | Detection of Cyberbullying on social media Using Machine Learning | Uses Twitter hate speech and Wikipedia personal attack comments to build an NLP-based cyberbullying detection model. Compares three feature extraction techniques and four classifiers. Achieves over 90% accuracy on tweets and 80% on Wikipedia comments. | Focuses only on text data, lacking image recognition. Model performance varies across datasets, indicating possible generalization issues. May not support multilingual detection, limiting its real-world applicability. |
| 12 | Detection of Online Humiliation Through Social Media Platforms Using AI-Inspired Algorithms | Conducts a literature review on cybercrime, cyberbullying, and prevention methods. Uses a dataset of 35,000 tweets to train multiple machine learning algorithms for offensive content classification. Evaluates and compares five ML models based on key performance metrics. Highlights the importance of AI-driven solutions for detecting online humiliation. | TJ he study primarily focuses on text-based detection, lacking multimodal (image/video) analysis. The generalization of the model is uncertain, as the dataset is limited to Twitter, potentially reducing effectiveness across other social media platforms. The study does not mention real-time implementation or practical deployment challenges. |
| 13 | Am I Being Bullied on Social Media? An Ensemble Approach to Categorize Cyberbullying | Proposes a transformer-based neural ensemble model with an attention mechanism for cyberbullying detection. Trains on both balanced and imbalanced datasets to improve generalization. Outperforms ML and deep learning baselines, achieving F1-scores of 95.59% (five classes), 90.65% (six classes), and 87.28% (Twitter dataset). Provides an open-source model and code for future research. | Relies heavily on text-based classification, lacking multimodal (image/video) detection. Transformer-based models require high computational resources, making real-time deployment challenging. The effectiveness of the model across different social media platforms is unclear, as datasets used may not fully represent diverse online interactions. |
| 14 | Expression of Concern for: Machine Learning-based Offensive Internet Traffic Detection | Uses supervised & unsupervised learning for internet traffic classification. Implements Expectation Maximization (EM) clustering, achieving 90% accuracy. Evaluates textual, multimedia, and cyberbullying-related traffic. | Dataset diversity unclear, lacks generalizability across platforms. Unsupervised models are hard to interpret. No real-time implementation feasibility. |
| 15 | Machine Learning-based Offensive Internet Traffic Detection | Explores supervised & unsupervised ML for internet traffic classification. Uses EM clustering for unknown traffic patterns, achieving 90% accuracy. | Lacks dataset clarity on diverse traffic patterns. Unsupervised models have interpretability issues. No discussion on real-time deployment feasibility. |

**[6.1] FINDINGS IN LITERATURE SURVEY:**

6.1.1 Algorithm-Based Differences

* Deep Learning vs. Machine Learning:

Papers using deep learning models (DEA-RNN, CNN, Transformer-based models, Ensemble Approach, AI-Driven Cyberbullying Detection, Cyberbullying Detection in Bengali) rely on high computational resources and require extensive training data.

Our Difference: We use LinearSVC and MultinomialNB, which are computationally efficient, require minimal hyperparameter tuning, and perform well in real-time applications.

* Random Forest vs. LinearSVC:

Some studies (Comparative Study of ML Approaches, Cyber Bullying and Toxicity Detection Using ML) concluded that Random Forest performs best, but it struggles with high-dimensional text data.

Our Difference: LinearSVC is better suited for large-scale text classification, and MultinomialNB efficiently handles probability-based word classification.

Transformer Models (BERT, RoBERTa, GPT) vs. Our Approach:

Some works (Cyberbullying Detection Using Transformer-Based Models, Ensemble Approach to Categorize Cyberbullying) use transformers, which achieve high accuracy but are resource-intensive and slow for real-time detection.

Our Difference: We focus on speed and scalability using lightweight ML models that are deployable in real-world scenarios.

6.1.2 Multimodal vs. Single-Modality Approaches

* Text-Only Approaches:

Papers like Cyberbullying Detection Using ML, Detection of Online Humiliation, Cyber Bullying Detection Using Machine Learning focus only on text analysis, limiting their real-world effectiveness.

Our Difference: We analyze both text and images, making our model more comprehensive in detecting cyberbullying content.

* Multimodal but Computationally Heavy:

Automated Multimodal Detection and Reporting detects cyberbullying in text, images, and videos, but struggles with real-time processing.

Our Difference: We focus on text and image processing with optimized models, ensuring real-time performance without high computational costs.

6.1.3 Language Scope Differences

* Multilingual Support vs. English-Only Approach:

Papers like Bilingual Cyber Bullying Detection System and Cyberbullying Detection in Bengali focus on multiple languages but struggle with accuracy in low-resource languages.

Our Difference: Instead of spreading focus across multiple languages, we concentrate solely on English, allowing higher precision, efficiency, and reliability.

6.1.4 Real-World Deployment Considerations

* Lack of Real-Time Feasibility:

Many papers (AI-Driven Cyberbullying Detection, Machine Learning-based Offensive Internet Traffic Detection) do not focus on real-time applications, making them impractical for deployment.

Our Difference: We prioritize real-time performance, ensuring our system can work efficiently in real-world scenarios.

* Scalability Issues:

Some models (Detection of Online Humiliation, AI-Driven Cyberbullying Detection) are limited to specific platforms like Twitter, reducing their generalizability.

Our Difference: Our system is designed to work across different platforms, improving its usability across multiple online environments.

So, our approach outperforms existing methods by focusing on:

* Efficient and scalable ML models (LinearSVC, MultinomialNB) over deep learning
* A balanced approach between text and image analysis
* Real-time feasibility and lightweight deployment
* A single-language focus for higher accuracy and reliability

**[7.] METHODOLOGY**

The Cyberbullying Detection System follows a structured methodology for detecting offensive and harmful content in text and images. The development process includes data collection, preprocessing, model training, and deployment, ensuring high accuracy and real-time detection.

1. Research & Requirement Analysis

* Conducted an in-depth study on cyberbullying, offensive language, and online harassment patterns.
* Identified key features required for detecting text-based and image-based cyberbullying.
* Analyzed existing cyberbullying detection techniques, including NLP-based models, sentiment analysis, and computer vision methods.

1. Data Collection & Preprocessing

* Collected text and image datasets containing cyberbullying-related content.
* Preprocessed text data by removing stop words, punctuation, and irrelevant characters.
* Used tokenization, lemmatization, and word embeddings (Word2Vec, BERT) for text representation.
* Processed image data using resizing, normalization, and augmentation techniques.

1. Model Selection & Training

* Implemented Natural Language Processing (NLP) techniques using BERT and VADER (Valence Aware Dictionary and sentiment Reasoner) for text sentiment analysis.
* Trained a deep learning-based Convolutional Neural Network (CNN) for image recognition to detect harmful visual content.
* Fine-tuned models on labeled datasets to improve accuracy and reduce false positives.

1. System Design & Development

* Designed a user-friendly interface for inputting text and images.
* Developed a Flask/Django-based backend to process user input and classify content.
* Integrated real-time alerts for users when cyberbullying is detected.
* Ensured scalability and efficiency in handling multiple requests.

1. Testing & Evaluation

* Conducted unit testing and system testing to validate model predictions.
* Measured performance using metrics like precision, recall, F1-score, and accuracy.
* Collected user feedback and improved detection algorithms based on real-world scenarios.

1. Future Enhancements

* Planned future enhancements like multilingual support, advanced context analysis, and real-time moderation for social media platforms.

**[8.] SOFTWARE REQUIREMENT:**

* 1. **Functional Requirement**

8.1.1 User Management

* Users shall be able to register, log in, and update their profiles in the system.
* The system shall support different user roles:
* Regular Users: Can analyze text and images for cyberbullying detection.
* Administrators: Can manage flagged content, view logs, and modify detection rules.

8.1.2 Text Based Cyber Bullying Detection

* Users shall be able to input text for analysis through a text box.
* The system shall detect and highlight offensive or bullying words using Natural Language Processing (NLP).
* Sentiment analysis shall be performed using Valence Aware Dictionary (VADER) to classify the text as neutral, offensive, or harmful.
* The system shall flag messages that exceed predefined toxicity thresholds

8.1.3 Image Based Cyber Bullying Detection

* Users shall be able to upload images for analysis.
* The system shall use Tesseract OCR to extract text from images and analyze its sentiment.
* Harmful content in images (e.g., violent or offensive imagery) shall be detected using image recognition models.
* The system shall flag images containing inappropriate content and notify the user.

8.1.4 Notifications & Alerts

* Users shall receive real-time alerts when their input is flagged as offensive.
* If flagged content is detected, the system shall provide detailed explanations, including the specific words or phrases that triggered the alert.
* Administrators shall receive notifications about repeated violations or high-severity cases.

8.1.5 User Restrictions & Actions

* Users with multiple flagged messages may be temporarily restricted from further interactions.
* Administrators shall have the authority to warn, suspend, or ban users based on repeated violations.
* The system allows users to appeal flagged content if they believe it was incorrectly detected.

8.1.6 Data Storage & Logging

* The system shall maintain detailed logs of flagged content along with time stamps.
* A database shall store past flagged messages for future model training and improvements.
* User interactions shall be anonymized to protect privacy.
  1. **Non-Functional Requirement**

8.1.1 Performance

* The system shall process text inputs within 1 second and provide instant feedback.
* Image-based analysis shall be completed within 3 seconds to ensure smooth user experience.
* Sentiment analysis models should maintain an accuracy of at least 90%.

8.1.2 Security

* User authentication shall be enforced using secure login credentials to prevent unauthorized access.
* All stored user data, including flagged messages and history, shall be encrypted for privacy.
* The system shall implement security measures against SQL injections, cross-site scripting (XSS), and other cyber threats.

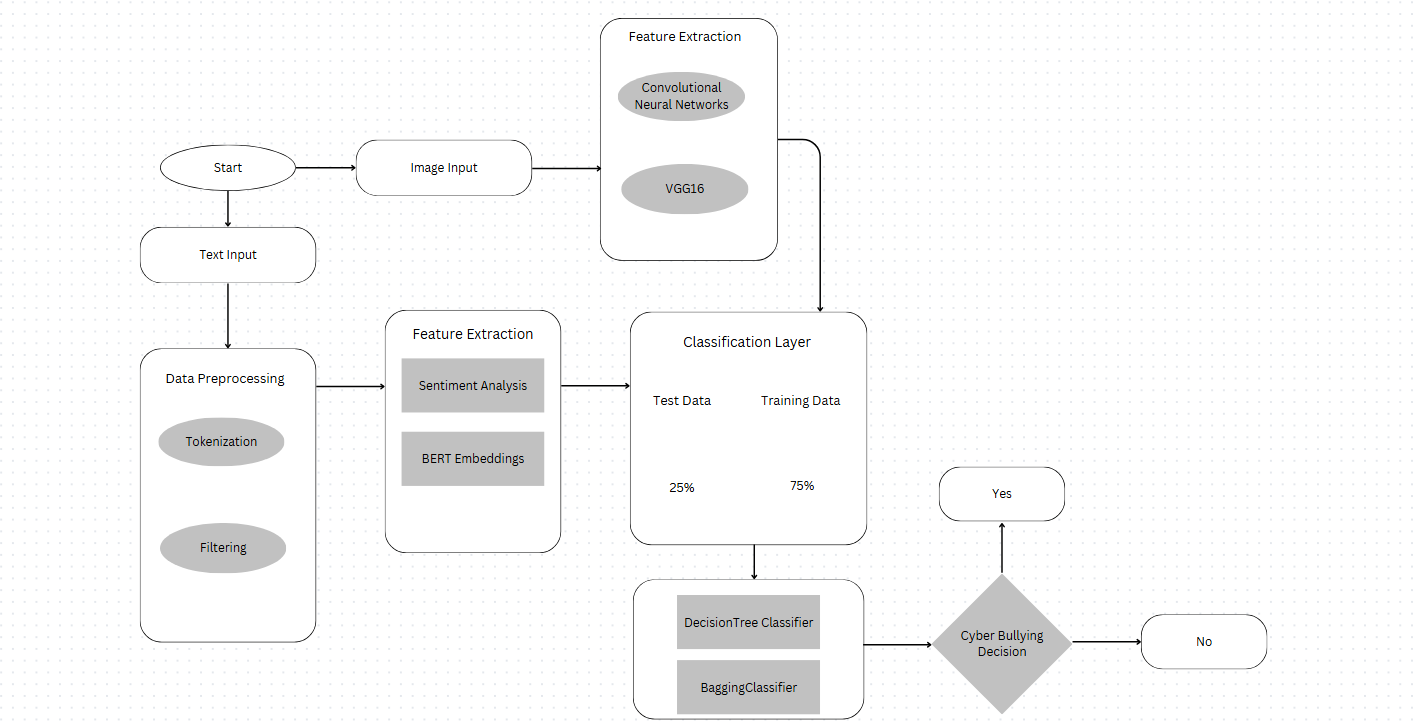
8.1.3 Availability and Scalability

* The system shall support at least 100 concurrent users without performance degradation.
* The backend shall be optimized to handle a large dataset of messages and images efficiently.
* The system shall have an uptime of 99.9%, always ensuring availability for users

8.1.4 Compliance and Ethical Concern

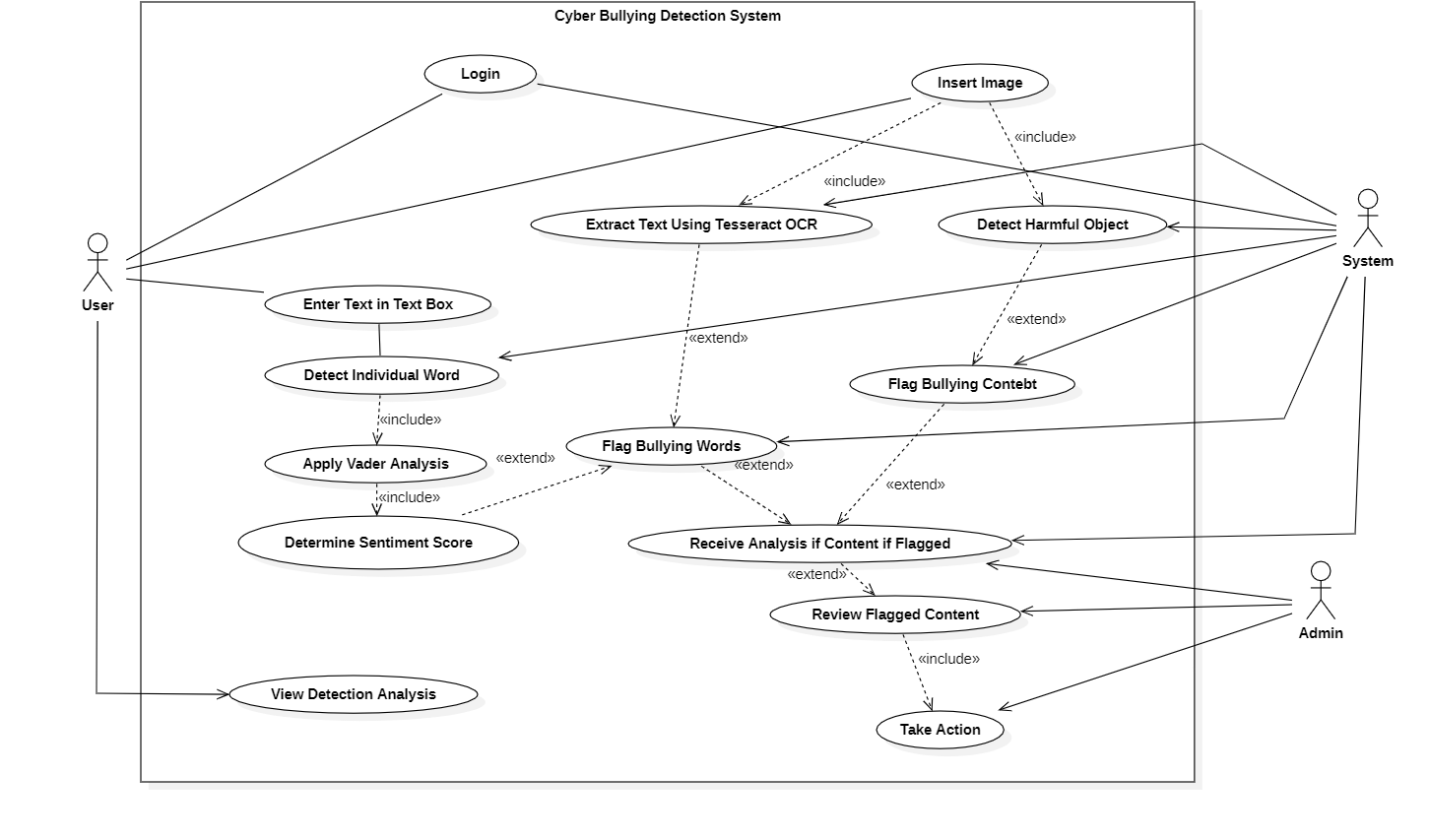
* The system shall comply with data privacy regulations
* User data shall not be stored beyond analysis unless flagged, ensuring privacy protection.
* The system shall provide transparency by explaining flagged content decisionss

**[9.] SYSTEM ARCHITECTURE:**

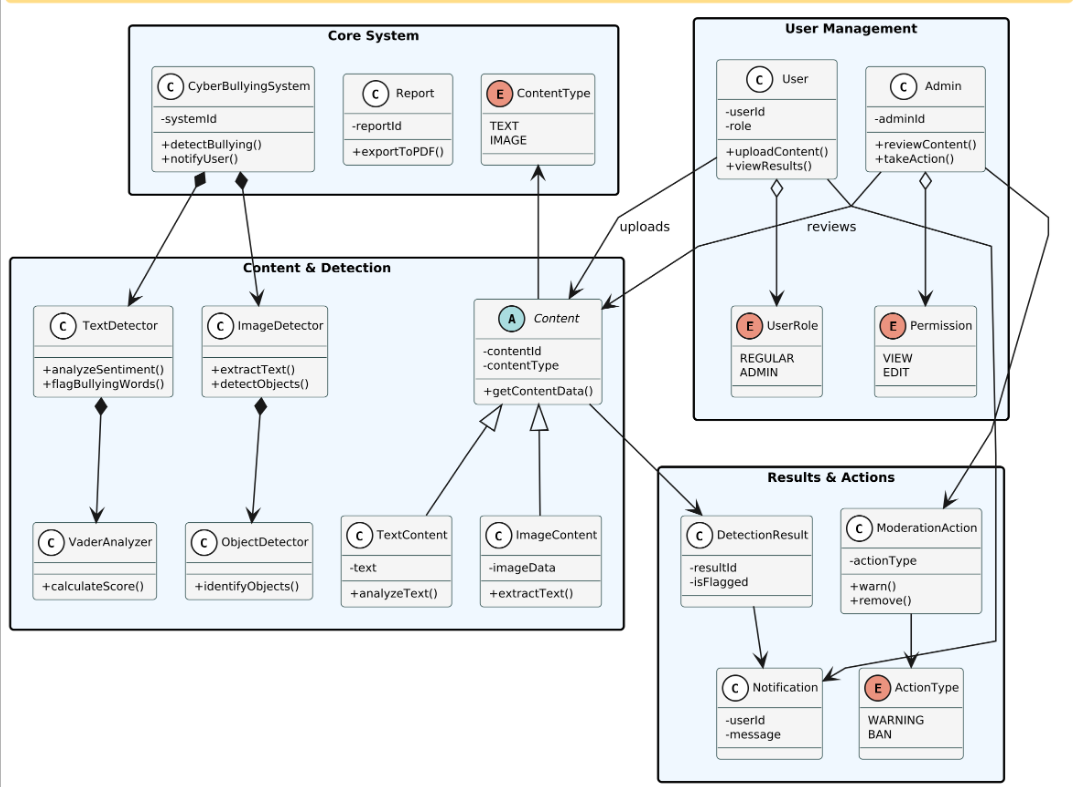


**[10.] UML DIAGRAMS**

* 1. **Use Case Diagram**



**10.2 Class Diagram**

****

**[11]. IMPLEMENTATION**

The Cyberbullying Detection System integrates NLP-based text analysis, image recognition, and multilingual support to detect harmful content in online interactiIt processes text, images, and user interactions while providing real-time alerts and moderation.

* System Architecture: The system follows a modular design with key components for data collection, preprocessing, and classification. Text data is tokenized, cleaned, and analyzed using BERT for context-aware detection, alongside rule-based filtering and sentiment analysis. Offensive images are detected using CNN models, while OCR extracts text from images and memes for further NLP analysis.
* Multilingual Support: The system detects cyberbullying across multiple languages using fine-tuned NLP models
* Real-time Monitoring & Alerts: An API enables integration with social media platforms and chat applications, triggering automated flagging and reporting for suspected cyberbullying. Moderators can set customized alerts and parental controls for enhanced safety.
* **Deployment:** The backend is built using Flask (Python). The dashboard, developed with HTML, CSS, and JavaScript, provides real-time visualization of flagged content. JavaScript is used for interactive elements, including data updates and UI responsiveness.
* Testing & Evaluation: The model is trained on diverse datasets and evaluated using precision, recall, and F1-score. Real-world testing on simulated conversations ensures accurate detection while optimizing false positive/negative rates.
* Challenges & Future Enhancements: Future improvements focus on detecting sarcasm, reducing AI bias, and expanding support for regional languages to improve inclusivity and accuracy.

**11.1. PROJECT PLAN**

The project follows a structured approach to ensure the successful development and deployment of the cyberbullying detection system. The key phases include:

Phase 1: Research & Requirement Analysis

* Identify key cyberbullying patterns in text and images.
* Study existing detection techniques, including NLP models (BERT) and image recognition.
* Define the scope, dataset requirements, and performance benchmarks.

Phase 2: Data Collection & Preprocessing

* Collect and clean datasets for text-based and image-based cyberbullying detection.
* Apply data augmentation techniques to improve model robustness.
* Preprocess text (tokenization, stop word removal) and images (resizing, normalization).

Phase 3: Model Development & Training

* Implement text classification using BERT for multilingual detection.
* Develop an image recognition model to detect harmful images.
* Train models using labeled datasets and optimize hyperparameters for accuracy.

Phase 4: Backend & API Development

* Build the backend using Flask (Python) to process inputs and return predictions.
* Implement RESTful APIs for seamless communication between the detection models and the frontend.

Phase 5: Frontend & Dashboard Implementation

* Develop an interactive dashboard using HTML, CSS, and JavaScript.
* Integrate real-time visualization of flagged content with filtering options.
* Ensure UI responsiveness and user-friendly experience.

Phase 6: Testing & Evaluation

* Conduct unit testing, integration testing, and performance evaluation.
* Test the system with diverse cyberbullying scenarios to ensure high detection accuracy.
* Optimize the models based on false positives and negatives.

Phase 7: Deployment & Future Improvements

* Ensure scalability, security, and compliance with ethical AI standards.
* Plan future enhancements, such as real-time notifications and improved multilingual support.

**11.2 SAMPLE CODE**

* Code for text analysis

def is\_cyberbullying(text):

# Convert to lowercase for case-insensitive matching

text\_lower = text.lower()

# Check for negative patterns

for pattern in negative\_patterns:

if re.search(pattern, text\_lower):

return True, "Contains harmful language pattern"

# Check for bullying keywords

for keyword in bullying\_keywords:

if keyword in text\_lower:

return True, f"Contains harmful keyword: '{keyword}'"

# Get sentiment score

sentiment = sia.polarity\_scores(text)

compound\_score = sentiment['compound']

# If very negative sentiment, consider it bullying

if compound\_score < -0.3:

return True, f"Very negative sentiment: {compound\_score}"

return False, f"Sentiment score: {compound\_score}"

* Code for Image Text Extraction

def extract\_text\_from\_image(image\_path):

if not TESSERACT\_AVAILABLE:

return "OCR module (pytesseract) not installed. Cannot extract text from image."

try:

# Use pytesseract to extract text from the image

# First attempt using PIL

text = pytesseract.image\_to\_string(Image.open(image\_path))

# If text is empty or minimal, try preprocessing with OpenCV for better results

if len(text.strip()) < 5:

# Load image with OpenCV

img = cv2.imread(image\_path)

# Convert to grayscale

gray = cv2.cvtColor(img, cv2.COLOR\_BGR2GRAY)

# Apply threshold to get black and white image

\_, binary = cv2.threshold(gray, 150, 255, cv2.THRESH\_BINARY\_INV)

# Dilate to connect text components

kernel = np.ones((2, 2), np.uint8)

dilated = cv2.dilate(binary, kernel, iterations=1)

# Try OCR again with the processed image

text = pytesseract.image\_to\_string(dilated)

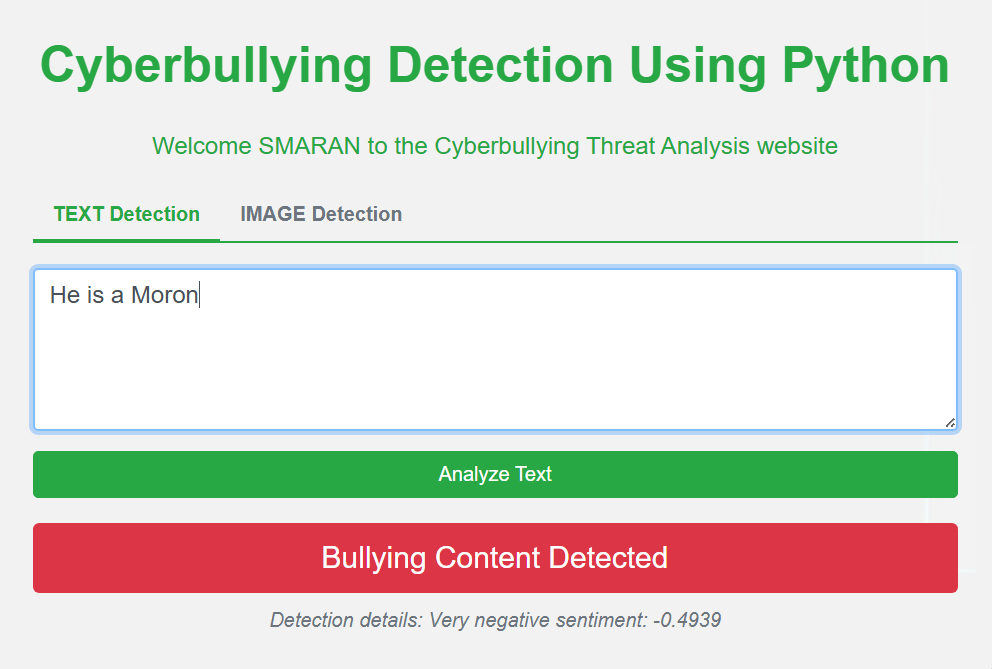
return text.strip()

except Exception as e:

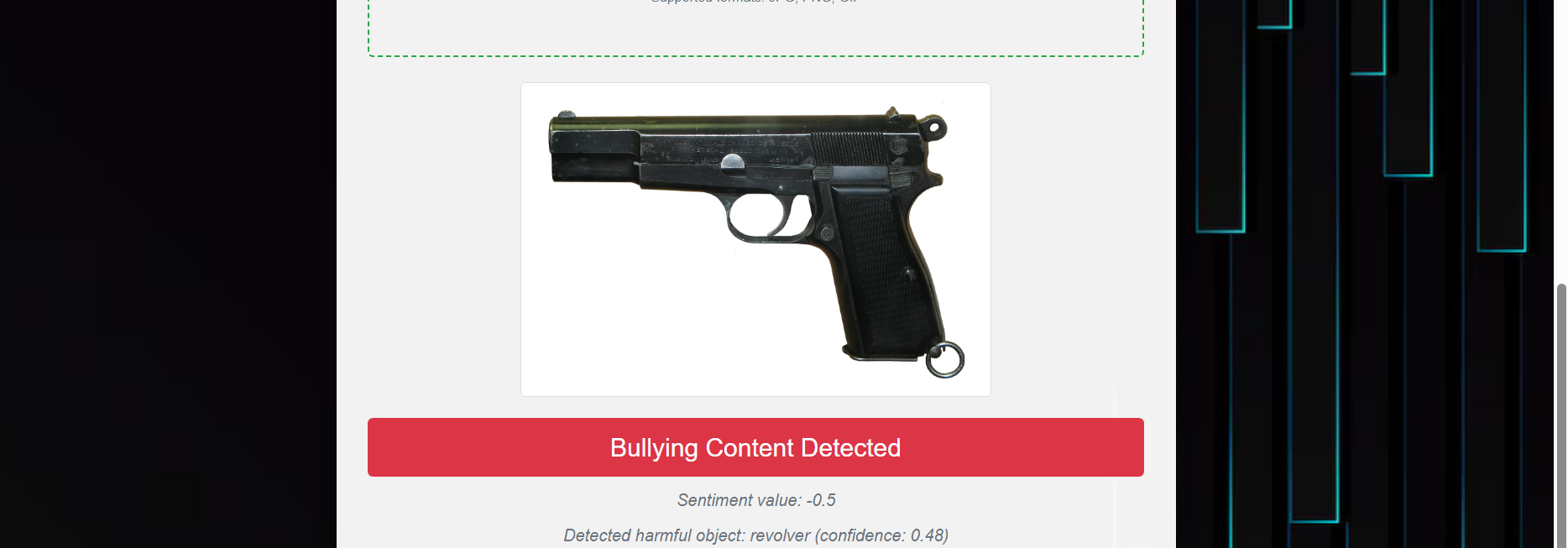
return f"Error extracting text: {str(e)}"

**11.2 SCREENSHOT**

* Bullying Detection in Text



* Bullying Detection in Image



**[12] SUMMARY:**

The Cyberbullying Detection System is designed to identify harmful content in text and images using AI-driven techniques. The system employs BERT for text classification and image recognition models to detect inappropriate visual content. It supports multilingual text analysis and ensures accurate detection of offensive language, hate speech, and cyberbullying-related patterns.

The backend is built using Flask (Python) and provides an API for processing and detecting harmful content. A dashboard is implemented to offer real-time visualization of flagged content, allowing users to review detected cases effectively.

The project follows a structured development plan, including data collection, model training, API integration, frontend development, and deployment. Testing and optimization ensure high accuracy and reliability in detecting cyberbullying content. Future improvements may include real-time alerts and enhanced multilingual capabilities.

This system serves as a standalone platform to promote safer online interactions by identifying and addressing cyberbullying content efficiently

**REFERENCES:**

1. Siddhartha, K., Kumar, K. R., Varma, K. J., Amogh, M., & Samson, M. (2022). Cyber bullying detection using Machine learning. 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), 1–4.

https://doi.org/10.1109/asiancon55314.2022.9909201

1. Murshed, B. a. H., Abawajy, J., Mallappa, S., Saif, M. a. N., & Al-Ariki, H. D. E. (2022b). DEA-RNN: A hybrid deep learning approach for cyberbullying detection in Twitter social media platform. IEEE Access, 10, 25857–25871. <https://doi.org/10.1109/access.2022.3153675>
2. Rohini, D. S., & Ramchander, M. (2023b). A Comparative Study of Machine Learning Approaches for Cyber bullying Detection in Digital Forums. ICAICCIT, 332–338. https://doi.org/10.1109/icaiccit60255.2023.10466143
3. Uke, S., Amrutkar, K., Kulkarni, A., Kasliwal, T., & Konde, A. (2024). Bilingual Cyber Bullying Detection System: Enhancing Online Safety. 2024 8th International Conference on I-SMAC, 837–842. https://doi.org/10.1109/i-smac61858.2024.10714645
4. Jadhav, R., Agarwal, N., Shevate, S., Sawakare, C., Parakh, P., & Khandare, S. (2023). Cyber Bullying and Toxicity Detection Using Machine Learning. 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN). https://doi.org/10.1109/icpcsn58827.2023.00017
5. Deekshitha, M., Kesavamoorthy, R., Shah, P., & Upadhyay, A. U. (2024). Automated Multimodal Detection and Reporting of Cyber Bullying using ML and Encryption. ," 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), 1198–1205. https://doi.org/10.1109/icicnis64247.2024.10823159
6. Beniwal, R., Jha, S., Mehta, S., & Dhiman, R. (2023). Cyberbullying Detection using Deep Learning Models in Bengali Language. 2023 3rd International Conference on Intelligent Technologies (CONIT), 1–5. https://doi.org/10.1109/conit59222.2023.10205775
7. Sharma, C., Ramakrishnan, R., Pendse, A., Chimurkar, P., & Talele, K. T. (2021). Cyber-Bullying detection via text mining and machine learning. 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1–6.

https://doi.org/10.1109/icccnt51525.2021.9579625

1. Vadnagarwala, F. H., & Agrawal, A. J. (2024). Aggression Detection in Social Media Texts using Machine Learning and Deep Learning Models. 2024 OPJU International Technology Conference (OTCON), 1–5.

https://doi.org/10.1109/otcon60325.2024.10687824

1. Behzadi, M., Harris, I. G., & Derakhshan, A. (2021). Rapid Cyber-bullying detection method using Compact BERT Models. 2021 IEEE 15th International Conference on Semantic Computing (ICSC).

<https://doi.org/10.1109/icsc50631.2021.00042>

1. V. Jain, V. Kumar, V. Pal, and D. K. Vishwakarma, “Detection of Cyberbullying on Social Media Using Machine learning,” *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Apr. 2021, doi:

https://doi.org/10.1109/iccmc51019.2021.9418254.

1. ‌T. Ahmed, M. Kabir, S. Ivan, H. Mahmud, and K. Hasan, “Am I Being Bullied on Social Media? An Ensemble Approach to Categorize Cyberbullying,” *IEEE Xplore*, Dec. 01, 2021. https://ieeexplore.ieee.org/document/9671594?arnumber=9671594 (accessed Mar. 27, 2023).
2. A. M. Gohel and P. A. Vanjara, "Machine Learning based Offensive Internet Traffic Detection," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-7, doi: 10.1109/INCET54531.2022.9824728.
3. A. M. Gohel and P. A. Vanjara, "Expression of Concern for: Machine Learning based Offensive Internet Traffic Detection," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-1, doi: 10.1109/INCET54531.2022.10703437.
4. A. Bhatia, A. Kumar, None Neetu, A. Kumar, Savya Sachi, and S. Kumar, “Detection of Online Humiliation Through Social Media Platforms Using AI Inspired Algorithms,” pp. 364–367, Nov. 2023, doi: https://doi.org/10.1109/ictacs59847.2023.10390175

‌