**A**

**PROJECT REPORT**

**ON**

**“CYBERBULLY DETECTION”**

**SUBMITTED TO**

**SHIVAJI UNIVERSITY, KOLHAPUR**

**IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT**

**FOR THE AWARD OF DEGREE**

**BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**

**SUBMITTED BY**

|  |  |  |
| --- | --- | --- |
| **MISS.** | **BHARATI BHASKAR PATIL** | **22UAD006** |
|  |  |  |

**UNDER THE GUIDANCE OF**

**Mr. S. P. Pise**

 



**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE ENGINEERING**

**DKTE SOCIETY’S TEXTILE AND ENGINEERING INSTITUTE, ICHALKARANJI**

**(AN EMPOWERED AUTONOUMOUS INSTITUTE)**

**2024-2025**

**D.K.T.E. SOCIETY’S**

**TEXTILE AND ENGINEERING INSTITUTE, ICHALKARANJI**

**(AN EMPOWERED AUTONOUMOUS INSTITUTE)**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE ENGINEERING**



**CERTIFICATE**

**This is to certify that, project work entitled**

**“CYBERBULLY DETECTION”**

**is a bonafide record of project work carried out in this college by**

|  |  |  |
| --- | --- | --- |
| **MISS.** | **BHARATI BHASKAR PATIL** | **22UAD006** |
|  |  |  |
|  |  |  |
|  |  |  |

**is in the partial fulfillment of award of degree Bachelor of Technology in Artificial Intelligence and Data Science Engineering prescribed by Shivaji University, Kolhapur for the academic year 2024-2025.**

**MR. S. P. PISE**

**(PROJECT GUIDE)**

**PROF. (DR.) T. I. BAGBAN PROF.(DR.) L.S.ADMUTHE**

**(HOD AI & DS DEPT.) (DIRECTOR)**

**EXAMINER: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DECLARATION**

We hereby declare that, the project work report entitled “Cyberbully Detection” which is being submitted to D.K.T.E. Society’s Textile and Engineering Institute Ichalkaranji, affiliated to Shivaji University, Kolhapur is in partial fulfillment of degree B.Tech.(AI & DS). It is a bonafide report of the work carried out by us. The material contained in this report has not been submitted to any university or institution for the award of any degree. Further, we declare that we have not violated any of the provisions under Copyright and Piracy / Cyber / IPR Act amended from time to time.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Title** | **Name of the Student** | **PRN** | | **Signature** |
| MISS. | BHARATI BHASKAR PATIL | 22UAD006 | <<SIGNATURE>> | |
|  |  |  |  | |
|  |  |  |  | |
|  |  |  |  | |

**ACKNOWLEDGEMENT**

With great pleasure we wish to express our deep sense of gratitude to Mr. S. P. Pise for his valuable guidance, support, and encouragement in the completion of this project report.

Also, we would like to take the opportunity to thank our head of department Dr. T. I. Bagban for his cooperation in preparing this project report.

We feel gratified to record our cordial thanks to other staff members of the Artificial Intelligence and Data Science Department for their support, help, and assistance which they extended as and when required.

Thank you,

|  |  |  |
| --- | --- | --- |
| **Title** | **Name of the Student** | **PRN** |
| MISS. | BHARATI BHASKAR PATIL | 22UAD006 |
|  |  |  |
|  |  |  |
|  |  |  |

**ABSTRACT**

Cyberbullying has become a pressing issue with the widespread use of digital communication platforms, posing serious threats to the mental health of individuals, particularly teenagers and young adults. This project presents the design and development of a Cyberbullying Detection System aimed at automatically analyzing user-input text to identify instances of offensive or bullying language. Leveraging Natural Language Processing (NLP) techniques, the system classifies textual content as either bullying or non-bullying, thereby enabling timely detection of harmful interactions. In cases where bullying is detected, the system provides users with the option to report the message for further review. The project follows the Agile Incremental Development Model, allowing for modular development, continuous testing, and iterative refinement. Developed as an individual effort, the project demonstrates disciplined task management and consistent self-evaluation. The solution is designed to be scalable and extensible, with potential future enhancements including multilingual support, real-time content monitoring, integration with social media platforms, and advanced user management functionalities. Overall, the system contributes to fostering a safer and more respectful online environment through the strategic application of technology to combat cyberbullying.

**INDEX**

1. **Introduction 1** 
   1. Problem definition 1
   2. Aim and objective of the project 1
   3. Scope and limitation of the project 2
2. **Background study and literature overview 3**
   1. Literature overview 3
   2. Investigation of current project and related work 4
3. **Requirement analysis 5**

a. Requirement Gathering 6

b. Requirement Specification 6

c. Use case Diagram 7

1. **System design 8**
   1. Architectural Design 8
   2. Flow Chart 9
   3. System Modeling 10
      1. Dataflow Diagram 10
2. **Implementation 11**

a. Agile Methodologies 11

b. Development Model 12

1. **Future Scope 13**
2. **References (Public repository GitHub source code links) 14**

**1. Introduction**

**a. Problem definition**

With the widespread use of digital platforms, users—especially students and young adults—are frequently exposed to harmful messages in the form of cyberbullying. Offensive content shared over social media and chat platforms can deeply affect mental health and emotional well-being. Traditional reporting mechanisms are often delayed or ineffective. There is a growing need for an intelligent, real-time solution that can detect harmful messages and enable users to take immediate action.  
This project addresses the problem by providing an automated cyberbullying detection system where users can log in, analyze messages, and report harmful content. It ensures a faster response by forwarding reported messages to the admin for review and necessary action. Currently, the system supports English language text only.

**b. Aim and objective of the project**

**Aim:**  
To develop a web-based system that allows users to analyze messages for cyberbullying, report offensive content, and enable admin intervention for timely action.

**Objectives:**

* To implement a message analysis module using a machine learning model trained on English cyberbullying data.
* To classify input text as bullying or non-bullying.
* To allow users to report messages that are flagged or perceived as harmful.
* To create an admin panel to review reported messages and take necessary actions.
* To ensure the system works efficiently for English-language input.

**c. Scope and limitation of the project**

Scope:

* Can be deployed in educational institutions and online platforms to reduce the impact of cyberbullying.
* Provides a streamlined process: login → message analysis → reporting → admin review.
* Enables a feedback mechanism where users actively participate in moderating harmful content.

Limitations:

* The system supports only English language inputs; other languages are not handled.
* It focuses solely on text-based content and does not process images, videos, or voice messages.
* The accuracy of the machine learning model is limited by the quality and size of the training dataset.
* The system does not automatically block users or take punitive actions—it only flags and forwards content to the admin.

**2. Background study and literature overview**

1. **Literature overview**

With the proliferation of social media and instant messaging platforms, cyberbullying has become a significant concern, especially among younger demographics. Early detection systems were largely rule-based, relying on predefined keyword lists to identify harmful language. However, these approaches often struggled with interpreting context, sarcasm, and evolving slang, resulting in limited accuracy and high false positives. To overcome these limitations, more recent research has leveraged machine learning and deep learning techniques to improve detection capabilities.

For instance, **Dinakar et al. (2011)** proposed a framework that utilized a combination of topic modeling and classification techniques to detect cyberbullying in online conversations. While their approach marked an important shift from rule-based to data-driven methods, it still relied on relatively shallow contextual analysis and was limited in handling nuanced or evolving language patterns. Similarly, **Badjatiya et al. (2017)** introduced deep learning models, specifically using word embeddings and recurrent neural networks, to detect hate speech in tweets. Their work demonstrated significant improvements in capturing semantic context and language complexity.

In comparison, the current project adopts Natural Language Processing (NLP) techniques to classify text input as bullying or non-bullying and provides an integrated user-reporting feature to enhance moderation. Unlike the earlier models which primarily focused on detection accuracy, this project emphasizes user engagement and ease of use, making it more suitable for real-time deployment in public platforms. However, gaps remain in scalability, language diversity, and real-time monitoring capabilities, which are areas targeted for future enhancements.

1. **Investigation of current project and related work**

This project centers on the development of a real-time Cyberbullying Detection System that utilizes Natural Language Processing (NLP) and machine learning to classify user-submitted text as either bullying or non-bullying. The system not only performs automated detection using a Linear Support Vector Classifier (LinearSVC) with TF-IDF vectorization, but also enables users to report offensive content, which is then directed to an administrative dashboard for further action. The architecture is designed to be lightweight, scalable, and adaptable for integration into web-based platforms, educational tools, and chat interfaces.

Related Work  
Previous research has contributed significantly to the evolution of cyberbullying detection methodologies. Notably:

* Dinakar et al. (2011) proposed a framework that combined topic modeling and classification techniques to detect cyberbullying in online conversations. Their work marked an important shift from rule-based systems to data-driven approaches, offering improved flexibility in identifying bullying language. However, their model provided limited contextual understanding, which restricted its ability to interpret nuanced, evolving, or sarcastic language commonly found in user-generated content.
* Badjatiya et al. (2017) advanced the field by introducing deep learning techniques, particularly the use of word embeddings and recurrent neural networks (RNNs) for hate speech detection on Twitter. Their models demonstrated superior performance in capturing semantic meaning and contextual relationships between words. Despite their effectiveness, these models required substantial computational resources and were better suited for large-scale deployments rather than real-time lightweight systems.

While many of these systems show promise, a majority focus on large-scale platforms and require substantial computational resources. In contrast, the current project is designed to be **lightweight, adaptable, and user-friendly**, suitable for integration into smaller applications or as a standalone reporting tool.

**3. Requirement analysis**

**a. Requirement Gathering**

### **Stakeholders Identification**

* **End Users**:
  + Individuals affected by cyberbullying.
  + Users seeking feedback on their own language use.
* **Moderators/Administrators**:
  + Review reported messages and take action.
* **Development Team**:
  + Implements frontend (HTML/CSS), backend (Flask), ML model, and database.
* **Legal and Policy Makers**:
  + Ensure compliance with legal and ethical standards.

### **User Needs**

* Real-time detection of bullying behavior.
* User-friendly and responsive web interface.
* Ability to report detected bullying content.
* Ensured privacy and data protection.
* Cross-platform accessibility (desktop/mobile browsers).

### **Functional Requirements**

* Text input box for users to enter/paste content.
* Machine learning classification of text using trained SVM model.
* Option to report bullying messages.
* Clear feedback on classification result.

### **Non-Functional Requirements**

* Response time within 3–5 seconds.
* Scalable to support high traffic.
* Model accuracy of at least 90%.
* Secure storage and access to user data.
* Easy and intuitive user experience.

## **b. Requirement Specification**

### **Functional Specifications**

* **Text Input**:
  + Text entered via a web form built with HTML/CSS.
* **Classification**:
  + Text processed using an SVM model trained on bullying datasets.
  + NLP techniques (tokenization, TF-IDF) applied for feature extraction.
* **Reporting**:
  + If bullying is detected, users can report the content.
  + Each report includes text, timestamp, and user session info.
* **Feedback**:
  + Instant classification result shown to the user with a report option.

### **Non-Functional Specifications**

* **Performance**:
  + Fast response time (3–5 seconds).
* **Scalability**:
  + Support for thousands of messages and reports.
* **Accuracy**:
  + Maintain ≥ 90% classification accuracy.
* **Security**:
  + Encrypted data handling with access controls in MongoDB.
* **Usability**:
  + Simple and intuitive interface for users and admins.

## **Technical Stack**

* **Frontend**:
  + HTML
  + CSS
  + JavaScript (optional for enhancements)
* **Backend**:
  + Python
  + Flask (REST API & server-side logic)
* **Machine Learning**:
  + Scikit-learn
  + SVM for classification
  + TF-IDF for feature extraction
  + NLP with spaCy or NLTK
* **Database**:
  + MongoDB (NoSQL, document-based storage for reports)

**c. Use case Diagram**

A diagram of a cyberbullying system

Description automatically generated

**4. System design**

**a. Architectural Design**

System architecture is a conceptual model that describes the structure and behaviour of multiple components and subsystems.

A diagram of a computer program

Description automatically generated

**b. Flow Chart**

A diagram of a computer program

Description automatically generated

**c. System Modeling**

1. **Dataflow Diagram**

A diagram of a data analysis process

Description automatically generated

A diagram of a process

Description automatically generated

**5. Implementation**

1. **Agile Methodologies**

In this project, a research-driven adaptation of the **Agile Incremental Development Model** was employed to support individual work, promote continuous improvement, and maintain flexibility throughout the system's development. The methodology was tailored specifically for solo development in a research context.

* **Iterative Development:**  
  The system was incrementally built by breaking down the project into smaller functional components—such as the text input interface, NLP-based cyberbullying detection module, report handling logic, and result display. Each component was researched, developed, and tested independently before integration.
* **Research-Driven Sprints:**  
  Sprint cycles were planned around both implementation and exploration tasks. This included experimenting with different NLP models, evaluating classification techniques (like Naive Bayes), and testing various data preprocessing methods.
* **Personal Sprint Planning:**  
  Weekly goals were outlined using sprint boards and timelines. These sprints allowed for focused task execution while accommodating the trial-and-error nature of research and model fine-tuning.
* **Self-Evaluation and Reflection:**  
  At the end of each sprint, completed modules were reviewed for performance and usability. Research notes and logs were maintained to capture insights and evaluate alternate strategies for improvements.
* **User Feedback Integration:**  
  Prototype versions were shared with peers and academic mentors. The received feedback guided adjustments in both algorithmic performance and user interface design, ensuring the solution remained user-centered.
* **Documentation & Flexibility:**  
  Progress was tracked through personal journals and digital task boards. This lightweight documentation allowed flexibility to revisit earlier decisions, pivot approaches based on new research findings, and refine the system without disrupting overall flow.

This agile research approach ensured that the system evolved through continuous learning, experimentation, and structured personal development—key traits in both solo project execution and applied machine learning research.

1. **Development Model**

A diagram of a process

Description automatically generated

For this project, the **Agile Incremental Development Model** was adopted, as it aligns well with the solo development process while ensuring structured progress and continuous enhancement.

* **Modular Development:** The system was divided into independent modules such as user input interface, text analysis module, result display, and reporting mechanism. This approach ensured clarity and easier debugging during development.
* **Sequential and Iterative Progress:** Each module was planned, developed, tested, and refined in a sequential yet iterative manner. Adjustments were made based on observations and test outcomes at every stage.
* **Self-Driven Prototyping:** Early versions of the system were created and tested by me to validate functionality. Based on the outcomes, necessary refinements were incorporated.
* **Reduced Risk through Continuous Testing:** Each component was tested independently during its development phase, minimizing the chance of major errors later in the project.
* **Scalability and Extendibility:** The modular design allows the system to be expanded in the future. Features such as user authentication, sentiment analysis, or a dashboard for moderators can be integrated without major changes to the existing structure.

**6. Future Scope**

The Cyberbullying Detection System holds significant potential for future enhancement and real-world application. As technology advances and online interactions increase, this project can be extended and scaled in various meaningful ways:

* 1. Integration with Social Media Platforms:  
  The system can be integrated with platforms like Facebook, Instagram, or Twitter to automatically flag and filter bullying content in real-time.
* 2. Multilingual Support:  
  Currently supporting English, the model can be expanded to detect cyberbullying in regional and international languages using NLP-based translation and sentiment detection techniques.
* 3. Advanced Machine Learning Models:  
  Incorporating deep learning techniques such as LSTM or BERT can improve the accuracy and contextual understanding of bullying phrases.
* 4. User Authentication and Role Management:  
  A full-fledged user system with login, roles (user/moderator), and access control can be added to make the system more secure and personalized.
* 5. Real-time Monitoring and Alerts:  
  The system can be enhanced to monitor chat rooms, forums, and comment sections live and issue instant alerts or warnings to users or administrators.
* 6. Reporting Dashboard:  
  A dashboard for moderators or administrators to view reported messages, track patterns, and take necessary action can be developed.
* 7. Mobile App Development:  
  Creating a cross-platform mobile application will increase accessibility and allow users to report and detect cyberbullying directly from their smartphones.

**7. References (public repository GitHub source code links)**

[**https://github.com/bhartii23/cyberbully-detection**](https://github.com/bhartii23/cyberbully-detection)