

A Project Report

On

**“ BILLY - Buddy Against Cyber Bullying ”**

Batch Details

|  |  |  |
| --- | --- | --- |
| Sl. No. | Roll Number | Student Name |
| 1 | 20211CDV0012 | Nishad Babu Sulikeri |
| 2 | 20211CDV0011 | Ganesh V Hegde |
| 3 | 20211CDV0029 | Dinesh Kumar Reddy M |
| 4 | 20211CDV0030 | K Vamsi Krishna |
| 5 | 20211CDV0004 | Manishimha G |

**School of Computer Science,**

**Presidency University, Bengaluru.**

Under the guidance of,

Ms. Ashishika Singh

School of Computer Science,

Presidency University, Bengaluru

**CONTENTS**

1. Introduction about Project

2. Literature Review

3. Objectives

4. Methodology

5. Timeline for Execution of Project

6. Expected Outcomes

7. Conclusion

8. References

**1. INTRODUCTION**

Cyberbullying is a form of harassment on a social/online platform. It has become very common, especially among teenagers. Our Website: Provides instant help through a user-friendly chatbot (“Billy”) which not only comforts the victim but also reports the person behind, to the cyber-crime department keeping the user’s identity anonymous. It will provoke the victim to provide important information and evidences. It will also calculate the statistics of cyber-crimes in an area visible to the cyber crime department to keep a track of crimes on a regular basis. The areas with red alert will be given more focus in any part of the country. Tips and defense tactics would be mentioned on the website to guide the youth and others. All the data will be kept anonymous so that it cannot be used in a negative way

Through the integration of Natural Language Processing (NLP), sentiment analysis, and entity recognition, Billy identifies key details about bullying incidents while ensuring user anonymity. This system offers a novel approach to combating cyberbullying by not only providing psychological support but also delivering anonymized reports to law enforcement. Additionally, it tracks geographical patterns of cyberbullying incidents to assist authorities in focusing on high-risk areas.

**2. Literature Survey:**

1. **Cyberbullying Detection Using Deep Learning Techniques**  
   Authors: D. Dinakar et al. (2012)

Summary: This research leverages deep learning to automatically detect cyberbullying incidents on social media. The algorithms focus on identifying abusive language and user behavior.

Limitations: Lacks a system for providing real-time emotional support and does not offer anonymous reporting, which may deter victims from seeking help.

1. **A Framework for Anonymized Cybercrime Reporting**  
   Authors: J. Wright et al. (2014)

Summary: The study presents a framework for anonymous reporting systems, focusing on data encryption and user authentication processes to protect the identity of the reporters. Ensures user anonymity, encouraging victims to report incidents without fear of exposure.

Limitations: While effective in preserving privacy, the framework lacks features for emotional support and real-time interaction, which are crucial for helping victims cope with their experiences.

1. **Chatbots in Mental Health Support: A Case Study**  
   Authors: E. Fitzpatrick et al. (2016)

Summary: Demonstrates the potential for AI-based chatbots to provide real-time emotional support, helping users manage distress. The chatbots use NLP for real-time text analysis to detect emotional distress in users and offer pre-programmed supportive messages based on detected emotions.

Limitations: Focuses on mental health in general, lacking a specialized approach for dealing with cyberbullying or integrating reporting features.

1. **Machine Learning Algorithms for Cyberbullying Detection**  
   Authors: A. Potha et al. (2017)

Summary: The algorithms are trained to detect cyberbullying through keyword detection, frequency analysis, and sentiment analysis, classifying content as abusive or non-abusive.

Limitations: Often produces false positives and lacks emotional care and anonymity, reducing its utility in providing comprehensive support for victims.

1. **Natural Language Processing (NLP) for Emotion Detection**  
   Authors: J. Cowie et al. (2010)

Summary: The study utilizes NLP techniques to detect emotional cues in text. It focuses on extracting sentiment polarity (positive, negative, neutral) and specific emotional states (like sadness, anger, or joy) by analyzing linguistic patterns, word choices, and sentence structures in conversations.

Limitations: Limited application in real-time cyberbullying detection, focusing more on general emotion detection.

1. **A Hybrid Model for Cyberbullying Reporting**  
   Authors: M. Baldry et al. (2018)

Summary: This hybrid model combines automated content detection tools with human moderation. The system first uses keyword-based filtering and sentiment analysis to flag potential incidents of cyberbullying, which are then reviewed by human moderators for confirmation and further action

Limitations: Lacks real-time engagement and emotional support for victims, relying heavily on human intervention for resolving cases.

1. **Entity Recognition for Cybercrime Reporting**  
   Authors: T. Lappas et al. (2019)

Summary: The paper uses entity recognition techniques, leveraging NLP and machine learning to extract key details (e.g., names, dates, locations) from cybercrime reports. These entities are used to enrich the report and provide actionable intelligence to law enforcement.

Limitations: Does not integrate emotional support for victims, focusing solely on gathering evidence.

1. **Interactive Chatbots for Social Good**  
   Authors: A. Poria et al. (2015)

Summary: Highlights the potential for chatbots to assist users in various social contexts, using AI to engage and help.The chatbots use simple decision trees, pre-programmed dialogues, and NLP for better user interaction. They focus on tasks like providing information, connecting users to services, and offering basic emotional support.

Limitations: Lacks a focus on sensitive issues like cyberbullying and does not provide anonymous reporting or emotional support.

1. **NLP in Chatbots for Social Good**  
   Authors: A. Rajeswari et al. (2015)

Summary: This study explores the application of natural language processing (NLP) in chatbots designed to serve social causes, such as mental health support and crisis intervention. It leverages machine learning and NLP algorithms to process user inputs, detect emotional states, and provide personalized responses.

Limitations: Lacks domain-specific focus necessary to address sensitive issues like cyberbullying. Furthermore, the chatbots in this study do not include mechanisms for secure reporting or anonymity.

1. **Real-time Detection and Reporting of Cyberbullying**  
   Authors: P. Kumar et al. (2020)

Summary: Proposes a real-time detection system that monitors online behaviors to report cyberbullying incidents. By continuously analyzing user behavior and flagged content, the system generates reports when predefined cyberbullying patterns are detected. These reports are then forwarded to moderators or law enforcement agencies.

Limitations: Lacks focus on protecting the emotional well-being of victims and does not provide anonymity.

**3. Objectives:**

1. **Provide Emotional Support**: Develop a chatbot that offers real-time, empathetic support to victims of cyberbullying, addressing the emotional distress caused by the incidents.
2. **Anonymous Reporting**: Ensure complete anonymity for victims reporting cyberbullying incidents to the relevant authorities.
3. **Improve Accuracy in Cyberbullying Detection**: Utilize NLP and machine learning to accurately detect cyberbullying incidents, reducing false positives and negatives.
4. **Generate Regional Crime Statistics**: Provide law enforcement with real-time statistics on cyberbullying incidents, enabling them to focus on high-risk areas.

**Existing Methods - Drawbacks:**

1. **Manual Reporting Systems**
   * **Drawback**: Often cumbersome and time-consuming. Victims may hesitate due to lack of anonymity.
2. **AI-based Moderation Systems**
   * **Drawback**: These systems focus on content moderation but don’t provide emotional support or guidance to victims.
3. **Social Media Keyword Filtering**
   * **Drawback**: Inefficient due to high rates of false positives, especially when detecting subtle forms of bullying.
4. **Anonymous Reporting Tools**
   * **Drawback**: Lack of integration with emotional support and poor guidance in the reporting process.

4. **Methodology:**

**Hardware and Software Details:**

* **Hardware**:
* Any system capable of supporting AI-based software.
* **Software**:
* **Botpress**: For developing the chatbot’s conversational flows.
* **Python**: For scripting machine learning algorithms and NLP tasks.
* **NLP**: for entity recognition and intent detection.
* **Database**: MySQL or MongoDB for storing anonymized user reports.

**Design Procedure:**

* **Website Design**: Create a website using React and tailwind CSS, where the chatbot will be implemented and all other modules will be added.
* **Chatbot Design**: Create conversational flows using Botpress, where "Billy" greets the user, provides emotional support, and prompts for incident details.
* **NLP Integration**: Use NLP to detect user intent (emotional support or incident reporting) and extract key entities (time, place, persons involved).
* **Anonymization**: Implement a system to anonymize all user data before storing it or sending reports to law enforcement.
* **Real-time Reporting**: Send anonymized reports to the cyber-crime department through secure APIs.

**Modules:**

1. **User Interaction Module**
   * Handles the chat interface and emotional support messages based on NLP-driven intent detection.
2. **Incident Reporting Module**
   * Gathers and anonymizes details about the incident, such as time, place, and involved individuals.
3. **Anonymization Module**
   * Strips identifying information from reports before sending them to authorities.
4. **Data Analysis Module**
   * Analyzes reported incidents for trends and generates red-alert regions based on cyberbullying frequency.
5. **API Integration Module**
   * Interfaces with law enforcement systems to forward reports and statistical data securely.

**5. OUTCOMES**

1. **Increased Incident Reporting**: Anonymity and emotional support will encourage victims to report cyberbullying incidents more frequently.
2. **Accurate Detection of Cyberbullying**: The integration of NLP and machine learning will significantly reduce false positives, ensuring that only genuine incidents are flagged.
3. **Geographical Crime Data**: The chatbot will help law enforcement agencies by providing real-time statistics, highlighting regions where cyberbullying is prevalent.
4. **Psychological Well-being of Victims**: Providing immediate emotional support will contribute to improved mental health outcomes for victims of cyberbullying.

6. **Timeline (Gantt Chart):**

**7. CONCLUSION**

The proposed system, "Billy," provides a novel solution to combat cyberbullying by combining emotional support with an anonymous reporting system. By leveraging AI, NLP, and anonymization, this project addresses the shortcomings of existing cyberbullying detection methods, ensuring victims' safety and encouraging higher reporting rates. The statistical analysis module offers valuable insights for law enforcement, helping them respond more effectively to cyber-crime.

**References:**

1. Dinakar, D., Jones, B., & Havasi, C. (2012). "Cyberbullying Detection Using Deep Learning Techniques." *International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*.  
   DOI: 10.1109/WI-IAT.2012.71
2. Wright, J., Martinez, L., & O'Brien, K. (2014). *A Framework for Anonymized Cybercrime Reporting*. Journal of Information Security and Applications, 19(2), 91-99. DOI: 10.1016/j.jisa.2014.01.001
3. Fitzpatrick, E., Dwyer, C., & Stokes, N. (2016). *Chatbots in Mental Health Support: A Case Study*. Proceedings of the International Conference on Health Informatics. DOI: 10.5220/0005632402830290
4. Potha, A., & Maragoudakis, M. (2017). *Machine Learning Algorithms for Cyberbullying Detection*. International Journal of Advanced Computer Science and Applications, 8(1), 240-246. DOI: 10.14569/IJACSA.2017.081240
5. Cowie, J., & Cornelius, N. (2010). *Natural Language Processing for Emotion Detection*. IEEE Transactions on Affective Computing, 1(1), 97-107. DOI: 10.1109/TAFFC.2010.15
6. Baldry, M., & Jansen, R. (2018). *A Hybrid Model for Cyberbullying Reporting*. Computers in Human Behavior, 89, 289-294. DOI: 10.1016/j.chb.2018.02.001
7. Lappas, T., & Mustafaraj, E. (2019). *Entity Recognition for Cybercrime Reporting*. IEEE Transactions on Information Forensics and Security, 14(8), 2100-2109. DOI: 10.1109/TIFS.2019.2892318
8. Poria, A., Cambria, E., & Gelbukh, A. (2015). *Interactive Chatbots for Social Good*. Proceedings of the International Conference on Computational Linguistics (COLING), 181-187. DOI: 10.5555/2858916.2859016
9. Patchin, C., & Hinduja, S. (2014). *Cyberbullying Prevention Tools in Schools*. Journal of School Violence, 13(1), 1-19. DOI: 10.1080/15388220.2014.893857
10. Kumar, P., & Joshi, P. (2020). *Real-time Detection and Reporting of Cyberbullying*. IEEE Access, 8, 72942-72951. DOI: 10.1109/ACCESS.2020.2987565