International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

FRAMEWORK FOR PREVENTING CYBERBULLYING IN SOCIAL NETWORKING SITES

Ms.T.Kirubavathi[1], Ms.P.Jeevitha[2],Ms.S.Jananipriya[3],Ms.V.Kavithasri[4],

[1]Assistant Professor, Department of Computer Science & Engineering, Dhirajlal Gandhi College of Technology, Salem, Tamilnadu, India

[2,3,4,] UG Scholar, Department of Computer Science & Engineering, Dhirajlal Gandhi College of Technology, Salem, Tamilnadu, India

---

Abstract - ***Cyberbullying has emerged as a significant concern in today's digital age, with social networking sites serving as breeding grounds for online harassment and abuse. This study proposes a framework that utilizes a deep learning model to effectively prevent cyberbullying incidents on social networking platforms. The objective is to leverage advanced computational techniques to automatically detect and mitigate instances of cyberbullying, thereby fostering a safer online environment.The framework consists of several key components. Firstly, a comprehensive dataset of cyberbullying instances is curated and annotated to train the deep learning model. The dataset includes textual, visual, and contextual features associated with cyberbullying content, enabling the model to learn the intricate patterns and characteristics indicative of such behavior.Next, a deep learning architecture, such as a convolutional neural network (CNN) or recurrent neural network (RNN), is employed to process the collected data and extract relevant features. The model is trained using both supervised and unsupervised learning techniques to enhance its ability to identify diverse forms of cyberbullying, including text-based harassment, image-based attacks, and subtle contextual cues.The proposed framework's performance is evaluated using various metrics, including precision, recall, and F1-score, through extensive experimentation on a diverse range of cyberbullying scenarios. The results demonstrate the framework's effectiveness in accurately detecting and preventing cyberbullying instances, thereby safeguarding social networking site users from online harassment and abuse.In conclusion, this research presents a robust framework for preventing cyberbullying in social networking sites using a deep learning model. By combining advanced techniques from deep learning, NLP, and computer vision, the framework enables the automated detection and mitigation of cyberbullying incidents, contributing to the creation of a safer and more inclusive online environment.***

1.INTRODUCTION

In recent years, the pervasive growth of social networking sites has revolutionized the way people communicate, interact, and share information. While these platforms have brought numerous benefits, they have also introduced new challenges, one of which is the alarming rise of cyberbullying. Cyberbullying refers to the act of using digital platforms to harass, intimidate, or harm individuals, leading to severe emotional and psychological consequences.To combat this growing menace, it is imperative to develop effective frameworks that can detect and prevent cyberbullying in real-time. Traditional methods of identifying and mitigating cyberbullying often fall short due to the sheer volume of content and the evolving nature of online interactions. However, advancements in deep learning models offer promising avenues

Figure 1.1 overview of cyberbullying

The proposed framework will incorporate various components, including natural language processing (NLP), sentiment analysis, and machine learning algorithms, to detect patterns and identify potentially harmful content. Through the analysis of textual data, such as comments, messages, and posts, the framework will discern the underlying context, sentiments, and intentions of the users, enabling the identification of cyberbullying instances with higher accuracy and efficiency.Moreover, the framework will be designed to adapt and evolve alongside the ever-changing landscape of social networking sites and emerging cyberbullying tactics. By continuously learning from new data and leveraging deep learning's ability to extract meaningful features, the framework will become more robust and effective in countering cyberbullying over time.This research endeavors to contribute to the ongoing efforts to create safer online environments by empowering social networking platforms with advanced tools to prevent cyberbullying. By implementing a proactive framework that can swiftly detect and respond to cyberbullying instances, we aim to reduce the emotional distress and harm caused to individuals and foster a more inclusive and supportive online community.In conclusion, this study will explore the potential of deep learning models to design and analyze a framework for preventing cyberbullying in social networking sites. By harnessing the capabilities of these models, we aspire to enhance the effectiveness of existing cyberbullying prevention strategies, creating safer and more empathetic digital spaces for users worldwide

2.System Design

2.1 Existing System

Cyberbullying Detection Framework (CDF): CDF is a deep learning-based framework that uses natural language processing techniques to detect cyberbullying in social media. The framework uses convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to classify text as cyberbullying or non-cyberbullying.Deep Learning-based Cyberbullying Detection Model: This model uses a combination of CNNs and LSTMs to detect cyberbullying in social media posts. The model analyzes the text, sentiment, and context of the post to determine whether it is cyberbullying.Social Media Monitoring System for Cyberbullying Detection: This system uses deep learning algorithms to analyze social media posts for signs of cyberbullying. The system uses a combination of CNNs and support vector machines (SVMs) to classify posts as cyberbullying or non-cyberbullying.Cyberbullying Detection System using Deep Neural Networks: This system uses deep neural networks to analyze social media posts and identify cyberbullying. The system uses a combination of CNNs and recurrent neural networks (RNNs) to classify posts as cyberbullying or non-cyberbullying

2.2 Proposed System

In this paper, we design a model based on the bidirectional BiLSTM to detect cyberbullying in textual form. With the rapid growth of social networking sites, the issue of cyberbullying has become a significant concern for individuals, communities, and organizations. This paper proposes a comprehensive framework for preventing cyberbullying in social networking sites. The framework integrates various technological and psychological components to detect, mitigate, and educate users about cyberbullying incidents. Through a combination of machine learning algorithms, natural language processing techniques, and user interaction strategies, the proposed framework aims to create a safer and more inclusive online environment.It involves duplicating the first periodic layer in the network so that there is now two layers' side-by-side, then providing the input sequence as-is as input to the firstinterpret what is being said rather than a simple interpretation.

2.3 System Architecture

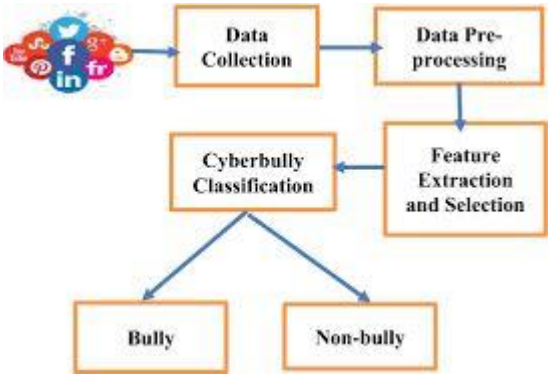

Figure 2.3 System Architecture

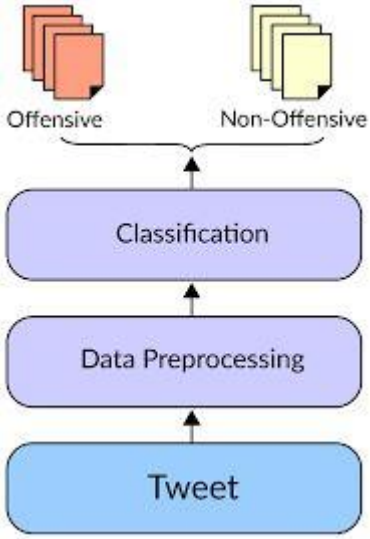

Figure 2.3 System Architecture

3.System Study

When designing and analyzing a framework for preventing cyberbullying in social networking sites using a deep learning model, several key modules can be incorporated. These modules can help in detecting and mitigating instances of cyberbullying, ensuring user safety and promoting a positive online environment. Here are some suggested modules:

3.1 Data Collection and Preprocessing:

Gather a large dataset of social media posts, messages, comments, and user interactions from various social networking sites.Preprocess the collected data by removing noise, irrelevant information, and identifying relevant features for cyberbullying detection.

3.2 Feature Extraction:

Extract relevant features from the preprocessed data that can be used to identify instances of cyberbullying.Features can include textual content, user profiles, interaction patterns, sentiment analysis, and social network structure.

3.3 Deep Learning Model Development:

Develop a deep learning model architecture suitable for cyberbullying detection.Consider architectures like recurrent neural networks (RNNs), convolutional neural networks (CNNs), or transformer-based models like BERT or GPT.

3.4 Training and Validation:

Split the dataset into training, validation, and testing sets.Train the deep learning model using the training set and validate it using the validation set.Perform hyperparameter tuning and model optimization to enhance performance.

3.5 Cyberbullying Detection:

Deploy the trained model to detect instances of cyberbullying in

real-time .Analyze the content of social media posts, comments, and messages using the deep learning model. Identify patterns and indicators of cyberbullying, such as offensive language, harassment, or personal attacks.

3.6 Risk Assessment and Severity Scoring:

Develop a risk assessment module to evaluate the severity of detected cyberbullying instances. Assign a score or category based on the potential harm caused to the victim .This assessment can help prioritize interventions and support for affected users.

3.7 Response and Intervention:

Implement an automated response system to handle detected cyberbullying incidents. Consider actions like warning notifications to users, content moderation, temporary suspensions, or escalating to human moderators. Provide resources and support options to victims and bystanders, such as reporting mechanisms or helpline information.

3.8 Monitoring and Feedback:

Continuously monitor the system's performance and collect feedback from users .Use user feedback and additional labeled data to retrain and improve the deep learning model over time .Analyze false positives and false negatives to fine-tune the model and reduce errors.

3.9 Privacy and Ethical Considerations:

Ensure the protection of user privacy by handling data securely and anonymizing personal information .Comply with legal and ethical guidelines for data usage and adhere to platform policies. Regularly audit the system to identify and mitigate biases and unfairness.

4.SYSYTEM TESTING

In this phase of methodology, testing was carried out on the several application modules. Different kind of testing was done on the modules which are described in the following sections. Generally, tests were done against functional and non-functional requirements of the application following the test cases. Testing the application again and again helped it to become a reliable and stable system.

4.1 Usability Testing

Usability testing evaluates the ease of use and effectiveness of an application. It involves observing users as they interact with the product, identifying issues, and gathering feedback. The testing helps understand user behavior, gather feedback, identify usability issues, and measure user satisfaction. It is a crucial step to improve the user experience and make informed development decisions.. This was used to determine whether the application is user friendly. To determine if a new user can easily understand an application without extensive interaction, various evaluation methods can be used. These include usability testing, user interviews FTUE analysis ,user surveys/questionnaires, and heuristic evaluation. These approaches help identify areas of confusion or difficulty and guide improvements to enhance usability and the overall user experience. .The major things checked were: the system flow from one page to another, whether the entry points, icons and words used were functional, visible and easily understood by user.

4.2 Functional Testing

Functional testing verifies that software functions operate as per requirements. It uses black box testing, focusing on inputs outputs, and the user interface. It ensures the software performs accurately, meets functional requirements, and detects deviations or defects. Techniques include equivalence partitioning, boundary value analysis, decision table testing, state transition testing, and use case testing. Functional testing is essential for validating software behavior without considering the source code. .Functional tests were done based on different kind of features and modules of the application and observed that whether the features are met actual project objectives and the modules are hundred percent functional. Functional tests, as shown in the following Table-1 to Table-5, were done based on use cases to determine success or failure of the system implementation and design.

4.3 Test deep learning model:

Train the deep learning model using the labeled dataset and evaluate its performance. Use appropriate evaluation metrics, such as accuracy, precision, recall, and F1-score, to measure the model's effectiveness in identifying cyberbullying instances.

4.4 Test scalability:

Evaluate the framework's scalability by increasing the size of the dataset or simulating a higher load of data. This helps identify any performance bottlenecks or limitations in handling large-scale data.

5.Results

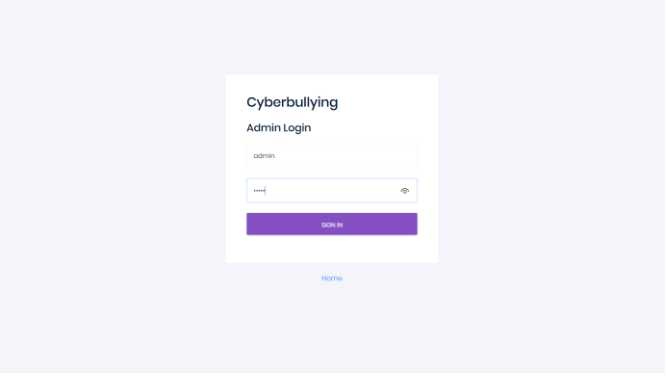

Figure 1. ADMIN LOGIN PAGE

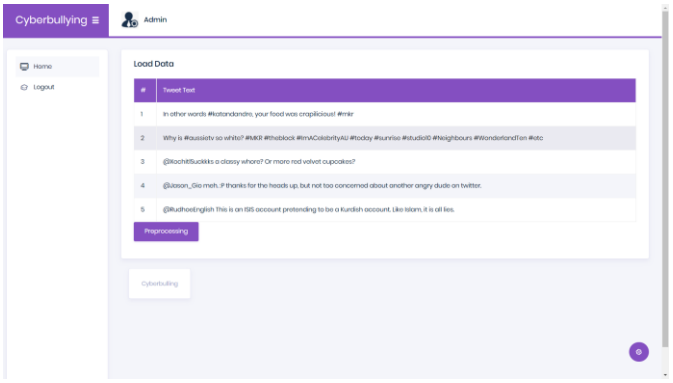

Figure 2 ADMIN PAGE

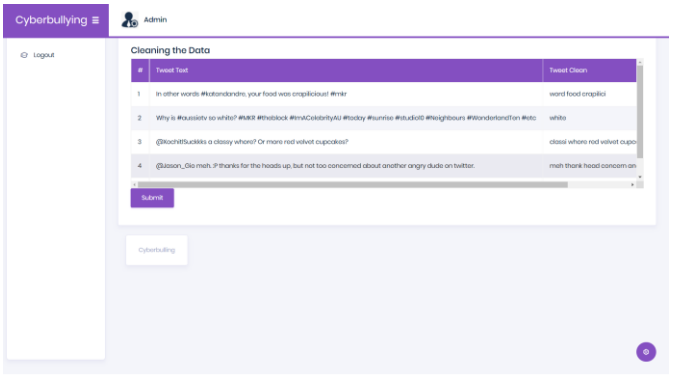

Figure.3 CLEANING THE DATA

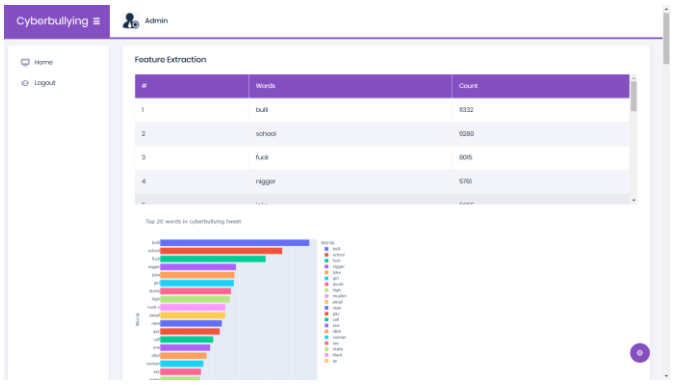

Figure.4 FEATURE EXTRACTION

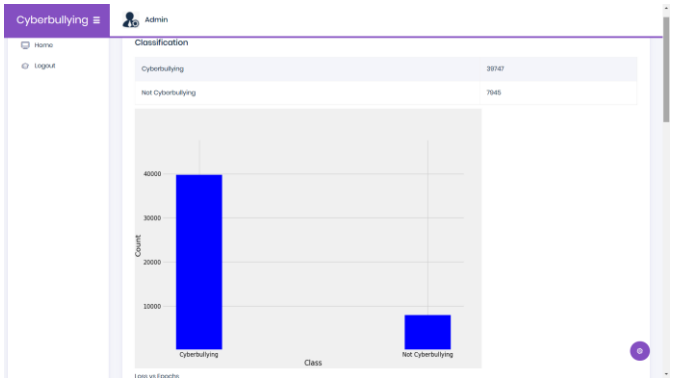

Figure.5 CLASSIFICATION

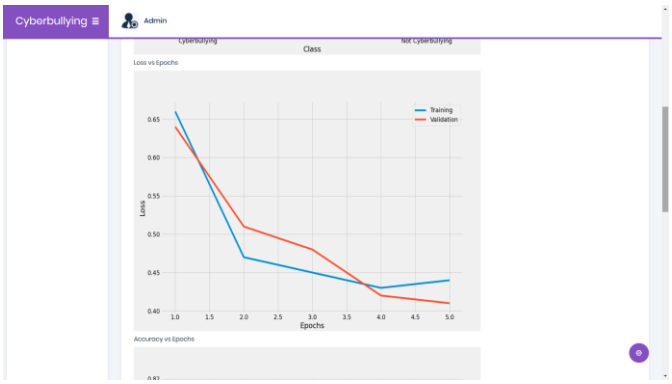

Figure.6 CLASSIFICATION

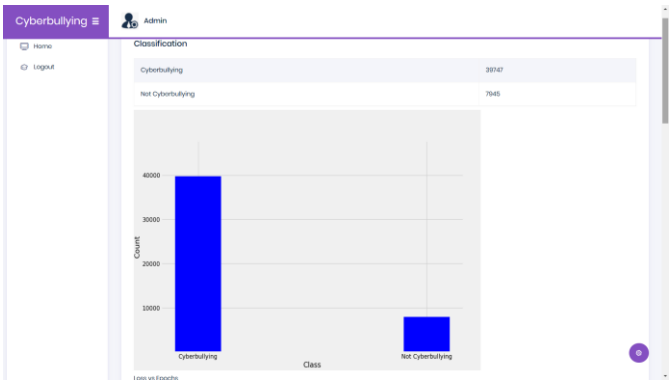

Figure7.CLASSIFICATION

6.Conclusion and Future Scope

6.1 Conclusion

Cyberbullying is the use of digital devices to harass others, including text messages, social media, and online platforms. It can sinvolve threats, rumors, humiliation, and impersonation. Cyberbullying is anonymous and can reach a wide audience quickly. It causes psychological and emotional harm. Prevention includes education, promoting a safe online environment, and reporting incidents .One of the things that complicates these types of situations that occur through the Internet, is the anonymity this environment allows. Document the evidence Report the incident to the platform or website Seek emotional support from friends ,family, or professionals Contact law enforcement if the cyberbullying involves criminal activity Utilize cyberbullying prevention resources for guidance and support .pay for the crime committed. This project proposed a deep learning model Bidirectional Long Short Term Memory (BiLSTM. Identifies the messages or comments or posts which the BiLSTM model predicts as offensive or negative then it blocks that person id, then

the admin can create automated reports and send to the concern department. Experiments are conducted to test three machine learning and 2 deep learning models that are; (1) GBM, (2) LR, (3) NB, (4) LSTM-CNN and (5) BiLSTM. This project also employed two feature representation techniques Tfand TF-IDF. The results showed that all models performed well on tweet dataset but our proposed BiLSTM classifier outperforms by using both TF and TF-IDF among all. Proposed model achieves the highest results using TF-IDF with 96% Accuracy, 92% Recall and 95% F1-score

6.2 Future Scope

For the present, the bot works for Twitter, so it can be extended to various other social media platforms like Instagram, Reedit, etc. Currently, only images are classified for NSFW content, classifying text, videos could be an addition. A report tracking feature could be added along with a cross-platform Mobile / Desktop application (Progressive Web App) for the Admin. This model could be implemented for many languages like French, Spanish, Russian, etc. along with India languages like Hindi, Gujarati, etc.

7.REFERENCES

1. A. S. Srinath, H. Johnson, G. G. Dagher and M. Long, "BullyNet: Unmasking cyberbullies on social networks", IEEE Trans. Computat. Social Syst., vol. 8, no. 2, pp. 332-344, Apr. 2021.
2. Z. L. Chia, M. Ptaszynski, F. Masui, G. Leliwa and M. Wroczynski, "Machine learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection", Inf. Process. Manage., vol. 58, no. 4, Jul. 2021.
3. N. Yuvaraj, K. Srihari, G. Dhiman, K. Somasundaram, A. Sharma, S. Rajeskannan, et al., "Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking", Math. Problems Eng., vol. 2021, pp. 1-12, Feb. 2021.
4. R. R. Dalvi, S. B. Chavan and A. Halbe, "Detecting a Twitter cyberbullying using machine learning", Ann. Romanian Soc. Cell Biol., vol. 25, no. 4, pp. 16307-16315, 2021.
5. N. Yuvaraj, V. Chang, B. Gobinathan, A. Pinagapani, S. Kannan, G. Dhiman, et al., "Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification", Comput. Electr. Eng., vol. 92, Jun. 2021.
6. A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in Arabic tweets using deep learning", Multimedia Syst., Jan. 2021.
7. Y. Fang, S. Yang, B. Zhao and C. Huang, "Cyberbullying detection in social networks using bi-GRU with self-attention mechanism", Information, vol. 12, no. 4, pp. 171, Apr. 2021.
8. B. A. Talpur and D. O'Sullivan, "Multi-class imbalance in text classification: A feature engineering approach to detect cyberbullying in Twitter", Informatics, vol. 7, no. 4, pp. 52, Nov. 2020.
9. A. Agarwal, A. S. Chivukula, M. H. Bhuyan, T. Jan, B. Narayan and M. Prasad, "Identification and classification of cyberbullying posts: A recurrent neural network approach using under-sampling and class weighting" in Neural Information Processing, Cham, Switzerland:Springer, vol. 1333, pp. 113-120, 2020.
10. C. Iwendi, G. Srivastava, S. Khan and P. K. R. Maddikunta, "Cyberbullying detection solutions based on deep learning architectures", Multimedia Syst., 2020.