IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN:2349-2163
Volume 13, Issue 04, April https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2026.v1304.11

Sentiment Analysis of CyberBullying Detection Using NLP and LSTM

Srinivasu Dadala
Assistant Professor, Department of CSE
Guru Nanak Institute of Technology, Hyderabad, Telangana, India
srinivasd.csegnit@gniindia.org
https://orcid.org/0009-0000-2504-2328
Baddipudi Rebecca Rejoice, Alugam Akshaya, Akepogu Bindu Priya
Students, Department of CSE
Guru Nanak Institute of Technology, Hyderabad, Telangana, India
b.rebeccarejoice@gmail.com , akshayaalagam@gmail.com, bindupriya857@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.13/Issue04/AEAP26.APAE10090
Research Article | Open Access | Double-Blind Peer-Reviewed| Article ID: IJIRAE/RS/Vol.13/Issue04/AEAP26.APAE10090
Received:02,March 2026, Revised: 29, March 2026, Accepted: 10,April 2026,Published Online: 22, April 2026.
https://www.ijirae.com/volumes/Vol13/iss-04/11.AEAP26.APAE10090.pdf
Article Citation: Dadala,Baddipudi,Alugam,Akepogu(2026),Sentiment Analysis of CyberBullying Detection Using NLP and LSTM,IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 13, Issue 04 of 2026 pages 800-806 **Doi:>** https://doi.org/10.26562/ijirae.2026.v1304.11 **BibTeX Key: Dadala@2026Sentiment**
IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2026, ISSN 2349-2163, https://doi.org/10.26562/ijirae.2026.v1304.11 The journal's official abbreviation is IJIRAE. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: Cyberbullying has become a serious issue in today's digital world, affecting both individuals and society. Detecting it on social media is important, but traditional machine learning methods often struggle with complexity and understanding subtle language patterns. This paper proposes an improved approach using Natural Language Processing (NLP) and Long Short-Term Memory (LSTM) networks to enhance detection accuracy. The system first cleans the text using steps like tokenization, stopword removal, stemming, and lemmatization, then extracts meaningful patterns and sentiment through embedding techniques. The processed data is fed into an LSTM model, which effectively understands context and sequence in language. To address imbalanced data, resampling techniques are applied, making the model more reliable and accurate in detecting cyberbullying.
Keywords: NLP; LSTM; Cyberbullying Detection; Sentiment Analysis; Emotions

I. INTRODUCTION

In today's digital world, social media and online platforms have made communication faster and easier, but they have also increased issues like cyberbullying. Cyberbullying involves using digital platforms to harm, threaten, or insult others, often leading to serious emotional and psychological effects. The anonymity of the internet makes it difficult to track and control such behavior, making it a major concern globally. Traditional detection methods using basic machine learning often struggle to understand the context and hidden meaning in online text. To overcome this, this study uses Natural Language Processing (NLP) along with LSTM models, which are effective in understanding sequences and context in language. The system preprocesses text using steps like tokenization, stop word removal, stemming, and lemmatization, and uses embeddings to capture meaning and sentiment. It also applies resampling techniques to handle imbalanced data, improving the model's accuracy and reliability in detecting cyberbullying and helping create a safer online environment.

II. LITERATURE SURVEY

Lalita Bisht, Kamika Chaudhary (2025) - This paper focuses on detecting cyberbullying using sentiment analysis and various machine learning algorithms. It highlights the growing risk of cybercrimes, especially among children, and the need for proactive detection systems[1]. The study evaluates supervised, unsupervised, and ensemble models on social media data. Among them, Random Forest achieved the highest accuracy of 87.74%, followed by SVM and KNN. However, the paper also points out challenges such as handling sarcasm, context, and evolving online language.

Ravinder Kumar, Dimpal Sharma, Ajay Kumar, Naveen Hemrajani, Ramesh Chandra Poonia (2025)-This paper proposes an efficient cyberbullying detection model using DistilBERT combined with sentiment analysis[2]. The approach improves understanding of both context and emotional tone in online text, making it more effective than traditional methods like TF-IDF and Bag of Words. DistilBERT provides a balance between speed and performance, achieving high accuracy of 93.7% and outperforming several deep learning models.

Chinni Roshini Durga, Sandeep Vemuri, Veeranki Kavya Lahari (2024) This paper presents a system for detecting negative comments on social media using NLP and the Random Forest algorithm[3]. The model classifies text into hate speech, offensive, or normal categories and also includes sentiment analysis to convert negative comments into positive ones. Additionally, it tracks user behavior and restricts users who repeatedly post harmful content, promoting healthier online communication. The system is designed to handle large volumes of data efficiently, making it suitable for real-time applications. It also considers contextual meaning to improve classification accuracy and reduce false detection.

By combining machine learning with sentiment analysis, the model enhances its ability to understand user intent. Overall, this approach supports the creation of a safer and more respectful digital environment. R Elankavi, B. Geethavani, S.C. Meghana, Vemula Rekha, J. Sai Priya, Kommaddi Mohamad Lookmaan (2024) This paper introduces that the internet connects billions of users, enabling fast communication and knowledge sharing, but it has also increased online abuse, harassment, and cyberbullying[4]. These behaviors can lead to serious mental health issues, especially among young people, making cyberbullying a critical concern. To address this, the CBSA (Cyberbullying with Sentiment Analysis) model combines Natural Language Processing (NLP) with sentiment analysis to understand both message content and emotional intent. This approach goes beyond simple keyword detection to identify harmful communication more effectively.

III. EXISTING SYSTEM

Existing cyberbullying detection systems mainly use traditional machine learning models like SVM, Naive Bayes, and Random Forest, which rely on manually created features such as n-grams and word frequencies[5,6,7]. Although these methods provide decent accuracy, they often fail to understand deeper meaning, context, and complex language patterns like sarcasm or slang. They also face issues with imbalanced data and noisy inputs, which affect performance. Additionally, these models cannot effectively capture the sequential nature of language, making them less suitable for advanced detection. Therefore, more intelligent and adaptive approaches are needed for better results.

Existing System Disadvantages

- High Computational Overhead
- Limited Sequential Context Understanding
- Dependence on Manual Feature Engineering

Proposed System

The proposed system uses a deep learning approach that combines NLP techniques with an LSTM model to improve cyberbullying detection [8,9]. It first preprocesses the text using steps like tokenization, stopword removal, stemming, and lemmatization to clean the data. The text is then converted into word embeddings to capture meaning and context. LSTM helps in understanding the sequence and flow of language, making it effective for detecting harmful patterns in conversations. The system also handles imbalanced data using resampling techniques. Overall, it provides more accurate and context-aware detection compared to traditional methods [10,11].

Proposed System Advantages:

- Effective Sequential Data Processing with LSTM
- Improved Contextual Understanding via NLP Techniques
- Automated Feature Extraction
- Enhanced Accuracy in Cyberbullying Detection

IV. SYSTEM ARCHITECTURE

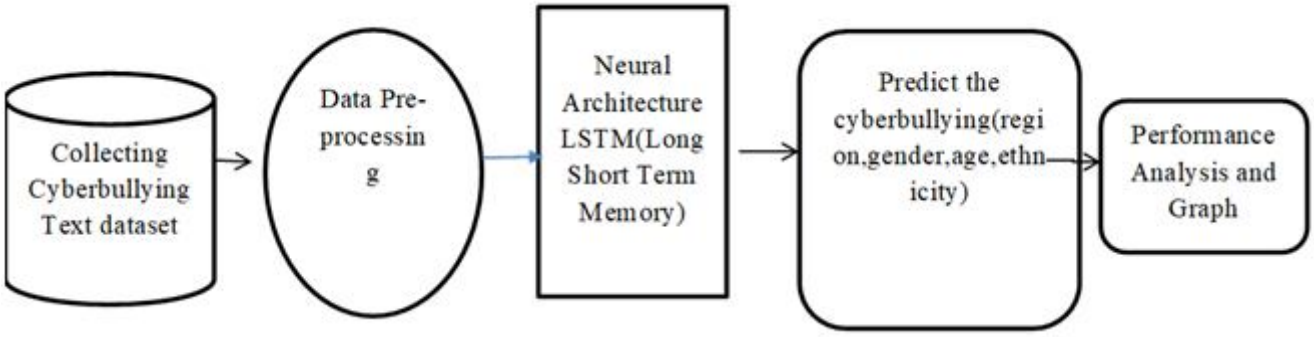

Fig 1: System Architecture

The system architecture for Sentiment Analysis of Cyberbullying Detection using NLP and LSTM consists of several key components that work together to identify and classify abusive text [12,13,14]. The architecture begins with data collection from social media platforms, followed by preprocessing using NLP techniques such as tokenization, stop-word removal, and lemmatization. The cleaned text is then converted into numerical vectors using word embeddings like Word2Vec or GloVe. These vectors are passed to an LSTM-based deep learning model, which analyzes the sequential patterns and sentiment of the text to detect cyberbullying[15]. The output is classified as bullying or non-bullying, and flagged posts are stored in a database for review. The system can be deployed as a web or API-based application, allowing real-time monitoring and moderator feedback for continuous model improvement [16].

Modules Name:

- Data Collection Module
- Dataset Module
- Data Preparation Module
- Model Selection Module
- Analyze and Prediction Module
- Accuracy on test set Module
- Saving the Trained Module

1.Data Collection Module: In this module, social media text data such as comments, posts, and messages are collected from various online platforms and publicly available datasets[17]. The collected data includes both bullying and non-bullying samples to ensure balanced representation. The focus is on gathering real-world data containing diverse linguistic expressions, slang, and emojis commonly used in online communication. Data is sourced ethically and anonymized to protect user privacy. This forms the foundation for effective model training and evaluation.

2.Dataset Module: The dataset consists of labeled text samples categorized into different classes such as hate speech, harassment, threats, and neutral content. It includes multiple features like message content, sentiment polarity, and contextual cues. The dataset is pre-split into training, validation, and testing subsets to facilitate model evaluation. Proper labeling ensures that the model learns distinct language patterns associated with cyberbullying. The dataset's quality and diversity directly influence the accuracy and generalization of the detection model.

3.Data Preparation Module: This module focuses on cleaning and preprocessing the collected text to remove unwanted noise such as URLs, special characters, and punctuation. Techniques like tokenization, stopword removal, stemming, and lemmatization are applied to standardize the textual data. Word embeddings such as Word2Vec or GloVe are used to convert words into meaningful numerical vectors. The goal is to preserve semantic information while reducing data complexity. This ensures that the model receives high-quality, structured input for effective learning.

4.Model Selection Module: The Long Short-Term Memory (LSTM) model is chosen for this project due to its ability to capture sequential and contextual relationships in text. LSTM effectively handles long-term dependencies, making it suitable for understanding the emotional tone and structure of cyberbullying language[18]. The model architecture includes input, hidden, and output layers optimized for text classification. Parameters such as learning rate, batch size, and dropout are fine-tuned to achieve optimal results.

5. Analyze and Prediction Module: In this stage, the preprocessed data is fed into the trained LSTM model to analyze textual patterns and predict whether a given message contains cyberbullying content. The model evaluates the emotional sentiment, context, and intensity of words to make predictions. Real time text inputs can also be processed for instant detection. The system's output provides a binary or multi-class label indicating the likelihood of cyberbullying. Visualization tools can be integrated to interpret prediction results and understand model decisions.

6. Accuracy on Test Set Module: After training, the model's performance is tested using unseen data to measure its accuracy, precision, recall, and F1-score. This module ensures that the system generalizes well and performs consistently across different types of input text. Confusion matrices and performance metrics are analysed to assess model strengths and weaknesses. High accuracy on the test set indicates that the model can effectively distinguish between bullying and non-bullying content. The results validate the robustness and reliability of the system.

7. Saving the Trained Module: Once the LSTM model achieves satisfactory performance, it is saved for future use without retraining. The trained model is serialized using formats like .h5 or .pkl for easy deployment. This allows integration into real-time applications, chat moderation tools, or web-based platforms. Saving the model ensures scalability and enables further fine-tuning or retraining when new data becomes available. It also supports efficient reuse for continuous improvement in cyberbullying detection accuracy.

V.IMPLEMENTATION

This project is a web-based application designed to detect cyberbullying using sentiment analysis with NLP and an LSTM model. When the application runs, it loads a pre-trained LSTM model along with a tokenizer to process user input effectively. When a user enters text, the system first cleans it by removing links, special characters, and common stopwords to keep only meaningful content. The cleaned text is then converted into numerical form using the tokenizer, where each word is assigned a unique number. To ensure consistency, the input is adjusted to a fixed length before being passed to the model. The LSTM model then analyzes the sequence of words to understand context and sentiment, rather than just identifying offensive keywords. This helps in detecting both direct and subtle forms of cyberbullying. Based on learned patterns, the model predicts whether the text contains cyberbullying and also identifies its type, such as gender-based, religion-based, age-based, or ethnicity-based. The final result is simplified into an easy-to-understand label. This output is then displayed clearly on the web interface for the user.

Existing Algorithm

The existing approach for cyberbullying detection mainly uses traditional machine learning algorithms such as SVM, Naive Bayes, and Random Forest. These models rely on basic NLP techniques like tokenization and feature extraction methods such as n-grams and word frequency. After preprocessing, the extracted features are used to train classifiers that predict whether a text is abusive or not. However, these methods focus more on individual words rather than overall context. They often fail to understand complex language patterns like sarcasm or hidden meanings. As a result, their performance is limited in handling real-world social media data.

Proposed Algorithm

NLP and LSTM:

The proposed algorithm uses a deep learning approach that combines NLP techniques with an LSTM model to detect cyberbullying in text. The input text is first preprocessed by removing unwanted elements, converting it to lowercase, and eliminating stopwords to retain meaningful content. It is then converted into numerical form using a tokenizer, followed by padding to maintain a fixed input length. The processed data is passed into the LSTM model, which analyzes the sequence and context of words to understand the intent. Based on learned patterns, the model predicts whether the text is cyberbullying and identifies its category. The final output is displayed as a simple and understandable label for the user.

IJIRAE::International Journal of Innovative Research in Advanced Engineering ISSN:2349-2163
Volume 13, Issue 04, April https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2026.v1304.11

VI. EXPERIMENTAL RESULTS

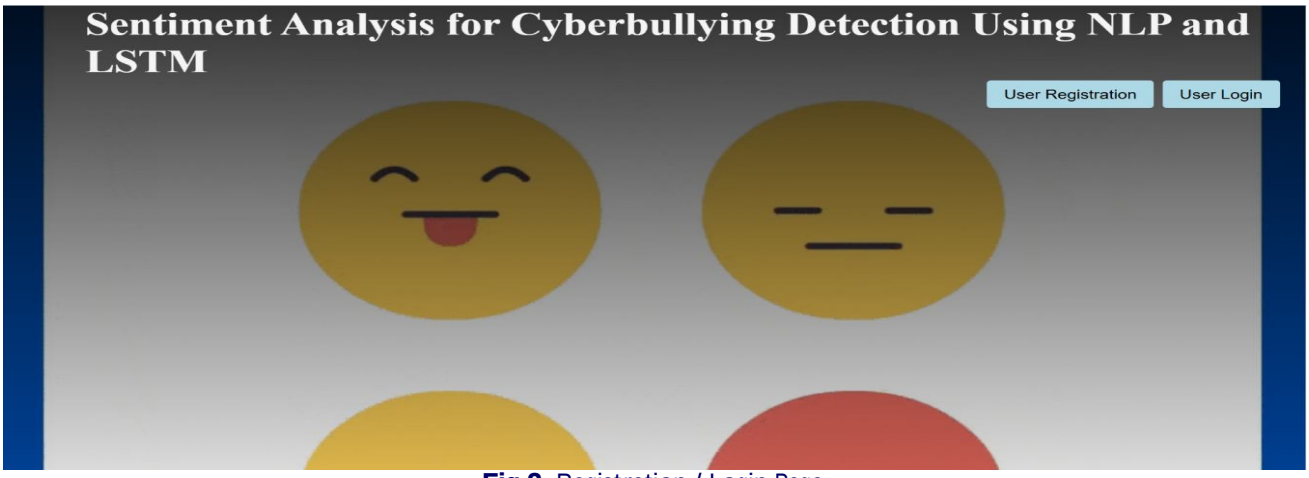

Fig 2: Registration / Login Page

The Registration / Login page shows a clean and simple login page for a cyberbullying detection system based on sentiment analysis. It includes options for user registration and login, making it easy for both new and existing users to access the system. Emoji faces are used to represent different sentiments, giving a quick understanding of the project's purpose. Overall, the interface is user-friendly and designed for easy navigation and accessibility.

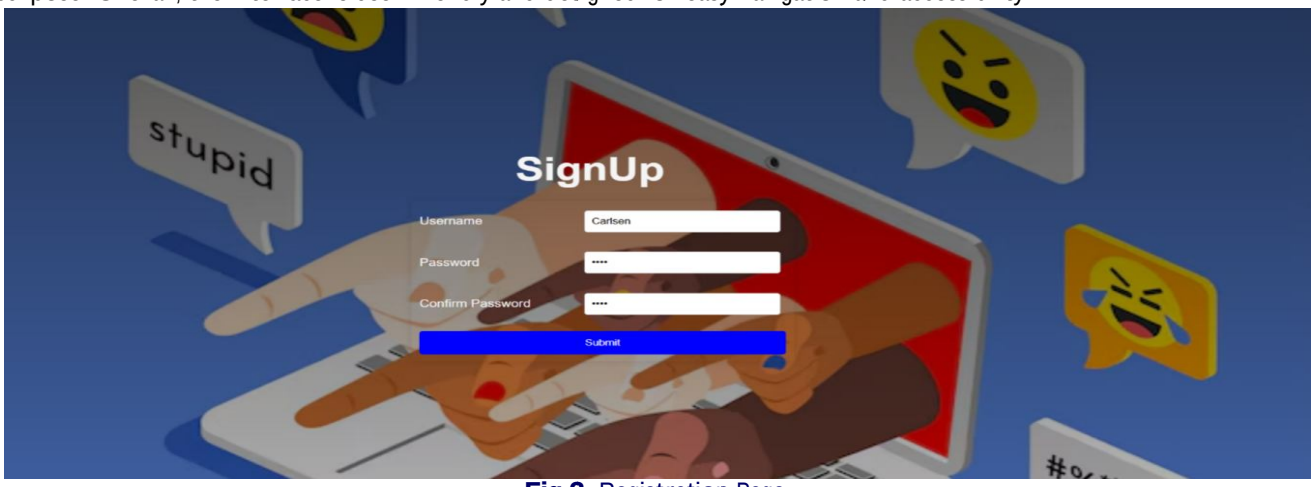

Fig 3: Registration Page

The Registration Page shows the sign-up page of a cyberbullying detection system based on sentiment analysis. It provides a simple form for users to enter a username, password, and confirm password to create an account. The background with messages and emojis reflects the cyberbullying theme and purpose of the project. Overall, the page is user-friendly, secure, and designed for easy registration and access.

Fig 4: Login Page

The Login Page shows the login page displayed after successful user registration in the cyberbullying detection system. It allows users to enter their username and password to securely access the platform. The background highlights different forms of cyberbullying, reflecting the purpose of the project. Overall, the page is simple, user-friendly, and ensures safe and easy access to the system.

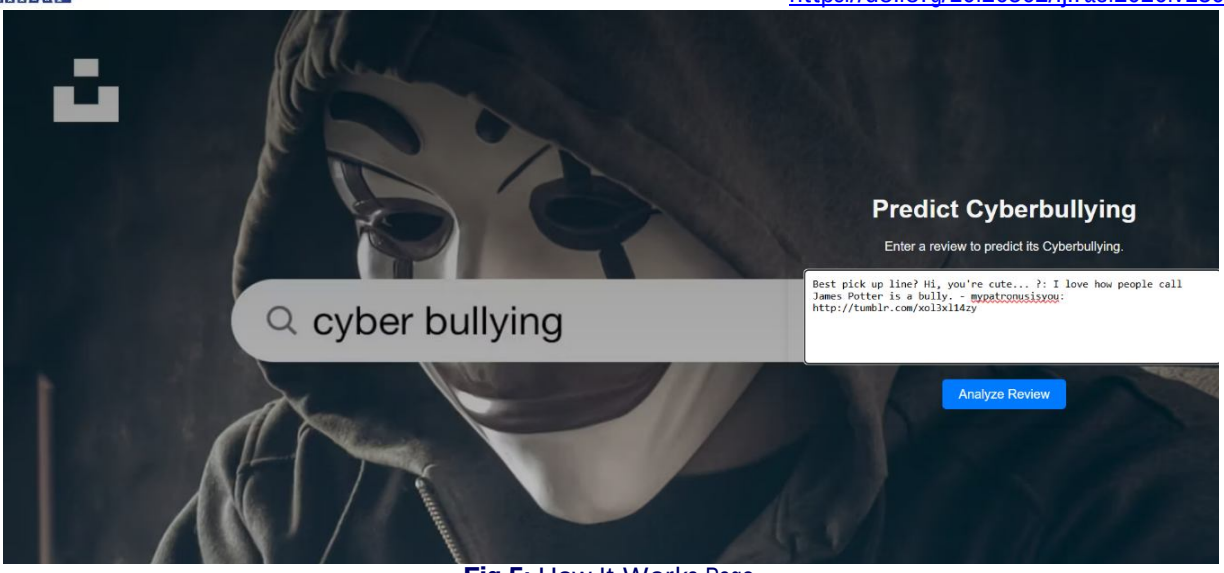

Fig 5: How It Works Page

The How It Works Page shows the main prediction page of the cyberbullying detection system after user login. It provides a simple text box where users can enter or paste content to check for cyberbullying. By clicking the "Analyze Review" button, the system processes the text and gives a prediction. Overall, the page is user-friendly and serves as the core feature for real-time analysis.

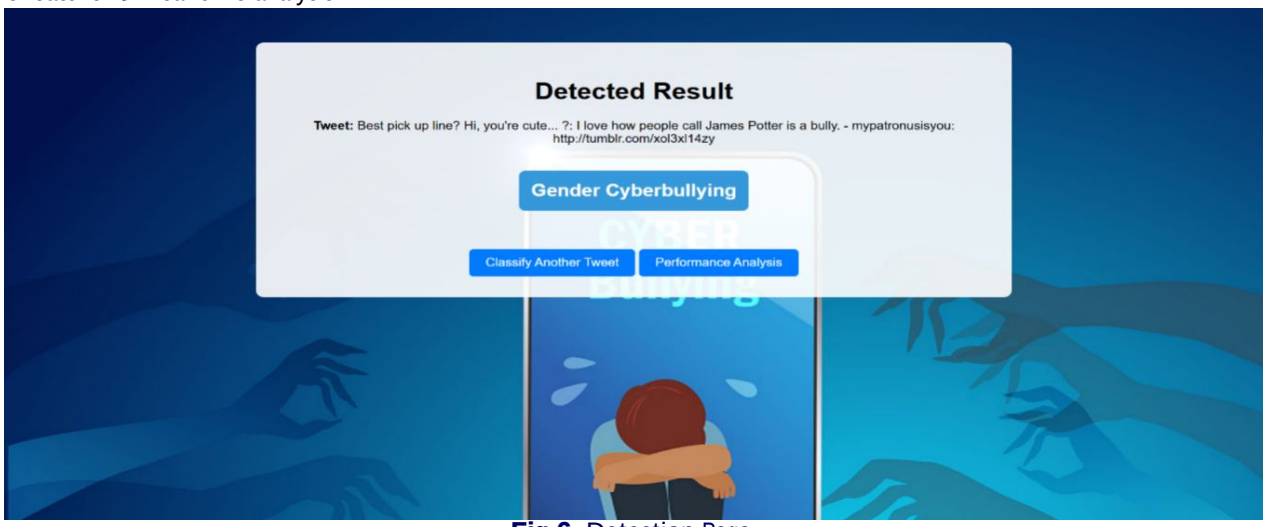

Fig 6: Detection Page

The Detection Page shows the result page displayed after analyzing the input text in the cyberbullying detection system. It presents the entered text along with the detected result, indicating whether it is cyberbullying, such as gender-based bullying in this case. The outcome is clearly highlighted for easy understanding. Overall, the page is simple and focuses on delivering the final prediction effectively.

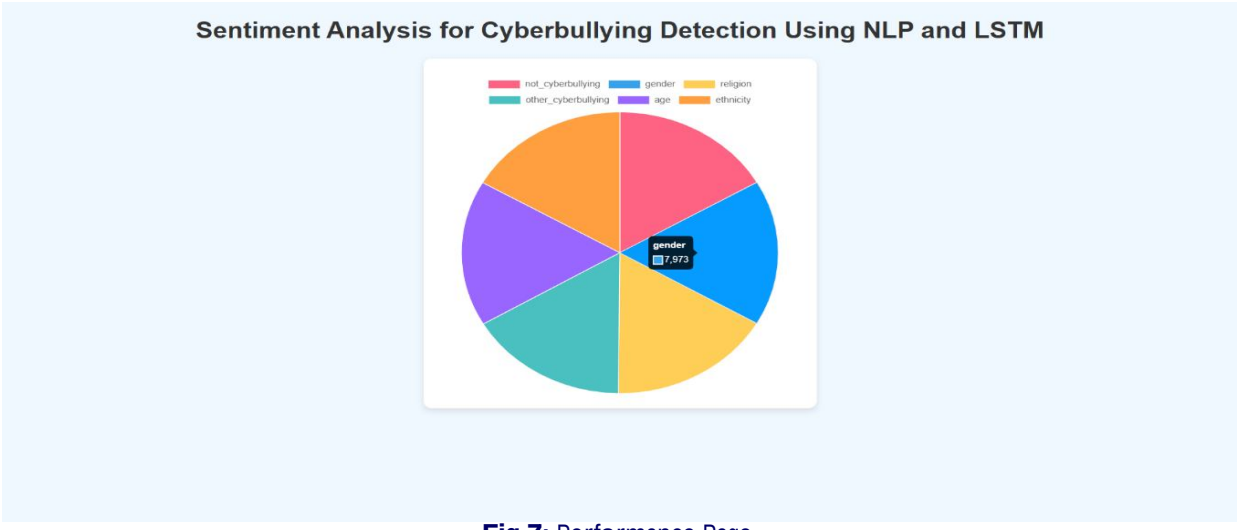

Fig 7: Performance Page

The image shows the performance analysis page of the cyberbullying detection system. It displays a pie chart representing different categories like gender, religion, age, and non-cyberbullying data. This visual makes it easy to understand the distribution and model performance. Overall, the page provides a clear and simple way to analyze the effectiveness of the system.

VII. CONCLUSION

In Conclusion, the proposed cyberbullying detection system combines NLP techniques with an LSTM model to identify harmful online content more effectively. It uses preprocessing steps like tokenization, stemming, lemmatization, and stopword removal to clean and prepare the data. Word embeddings are used to capture semantic meaning and context, while the LSTM model understands the sequence and emotional tone of text. This helps in detecting both direct and indirect forms of cyberbullying. The model achieves better accuracy compared to traditional machine learning approaches. It also handles class imbalance using resampling techniques, improving fairness and reliability. Additionally, the system can be integrated into social media platforms for real-time monitoring and moderation. Overall, it contributes to creating a safer and more positive online environment. The framework can also be extended to handle multiple languages and different types of online platforms. With further improvements, the system's performance and scalability can be enhanced for wider real-world applications.

VIII. FUTURE ENHANCEMENT

In the future, the system can be enhanced to support multiple languages, cross-platform analysis, and real-time integration with social media applications. Advanced models like BERT and multimedia detection (text, audio, image) can improve accuracy and coverage. Features such as explainable AI, user behaviour analysis, and adaptive learning can further increase transparency, fairness, and overall effectiveness.

REFERENCES

1. R. Gün and G. G. Akduman, "What is cyberbullying?" in Bullying in Media and beyond. Turkey: IGI Global, pp. 473-485, https://doi.org/10.4018/978-1 6684-5426-8.CH028
2. Y. Hu, E. M. Clancy, and B. Klettke, "Understanding the vicious cycle: Relationships between non consensual sexting behaviours and cyberbullying perpetration," Sexes, vol. 4, no. 1, pp. 155—166, Feb. 2023, https://doi.org/10.3390/sexes4010013
3. E.I.Galyashina and V.D.Nikishin, "The concepts of aggressive information impact through the lens of internet users' worldview security," J. Siberian Federal Univ. Humanities Social Sci., vol. 14, no. II, pp. 1660-1673, Nov. 2021, https://doi.org/10.17516/1997-1370-0848.
4. S.Joshi, H. G. Nagariya, N. Dhanotiya, and S. Jain, "Identifying fake profile in online social network: An overview and survey," Commun. Comput. Inf sci., vol. 1240, pp. 17-28, Jan. 2020, https://doi.org/10.1007/978-98115-6315-72
5. E.Vogels, Teens and Cyberbullying 2022, Pew Research Center, Dec. 23, 2024. This report examines trends in teen cyberbullying and online harassment, providing statistical insights into exposure and frequency. It emphasizes the role of social media in facilitating bullying and the psychological impact on victims.
6. S.Cook,Cyberbullying Statistics and Facts for 2024, Dec. 23, 2024. This online article provides a comprehensive overview of cyberbullying prevalence, victim demographics, and platform-specific risks. It highlights the increasing trend of harassment on popular social media networks.
7. L.H.Collantes, Y.Martafian, S.N.Khofifah, T.K.Fajarwati, N.T.Lassela, and M.Khairunnisa (2020), The Impact of Cyberbullying on Mental Health of the Victims, Proc. 4th Int. Conf. Vocational Educ. Training (ICOVET), pp. 30–35. This study investigates how cyberbullying negatively affects mental health, identifying stress, anxiety, and social withdrawal as common consequences.
8. S.Unnava and S.R.Parasana (2024), A Study of Cyberbullying Detection and Classification Techniques: A Machine Learning Approach, Eng. Technol. Appl. Sci. Res., vol. 14, no. 4, pp. 15607–15613. The paper evaluates multiple ML techniques for automated cyberbullying detection, comparing SVM, Random Forest, and Naïve Bayes classifiers.
9. R.Endsuy(2021), Sentiment Analysis Between VADER and EDA for the U.S. Presidential Election 2020 on Twitter Datasets, J. Appl. Data Sci., vol. 2, no. 1, pp. 8–18. The study explores sentiment analysis models for detecting polarizing language, which can also inform cyberbullying detection strategies.
10. L.Grunin, G.Yu, and S.S.Cohen (2021), The Relationship Between Youth Cyberbullying Behaviors and Their Perceptions of Parental Emotional Support, Int. J. Bullying Prevention, vol. 3, no. 3, pp. 227–239. This research highlights the protective effect of parental support against cyberbullying incidents.
11. Ali and N. Hameed (2017), Hybrid Tools and Techniques for Sentiment Analysis: A Review, Int. J. Multidisciplinary Sci. Eng., vol. 8, no. 4, pp. 28–33. The paper reviews hybrid sentiment analysis methods and their relevance for identifying harmful online content.
12. J.O.Atoum (2020), Cyberbullying Detection Through Sentiment Analysis, Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI), pp. 292–297. The study applies sentiment analysis techniques to detect offensive messages on social media platforms.
13. P.Yi and A.Zubiaga (2023), Session-Based Cyberbullying Detection in Social Media: A Survey, Online Social Netw. Media, vol. 36, Art. no. 100250. This survey categorizes session-based detection approaches and identifies research gaps for real-time detection systems.
14. T.Ahmed,S.Ivan, M.Kabir, H.Mahmud, and K.Hasan (2022), Performance Analysis of Transformer-Based Architectures and Their Ensembles to Detect Trait-Based Cyberbullying, Social Netw. Anal. Mining, vol. 12, no. 1, p. 99. The paper evaluates transformer models for identifying cyberbullying traits in social media text.

15. T.Ahmed, M.Kabir, S.Ivan, H.Mahmud, and K.Hasan (2021), Am I Being Bullied on Social Media? An Ensemble Approach to Categorize Cyberbullying, Proc. IEEE Int. Conf. Big Data (Big Data), pp. 2442–2453. This research proposes an ensemble of ML models to improve cyberbullying detection accuracy.
16. Kumar, B. S., Geetha, M. P., Padmapriya, G., & Premkumar, M. (2020). An approach for improving the labelling in a text corpora using sentiment analysis. Advances in Mathematics: Scientific Journal, 9(10), 8165-8174.
17. S.Muppidi, B.S.Kumar and K.P.Kumar, "Sentiment Analysis of Citation Sentences using Machine Learning Techniques," 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2021, pp. 1-5, doi: 10.1109/i-PACT52855.2021.9696703.
18. Palagati, S.K.Balan, S.Arun Joe Babulo, L. Raja, K.K.Natarajan and R.Kalimuthu, "Comparative Analysis of Machine Learning Algorithms and Datasets for Detecting Cyberbullying on Social Media Platforms," 2024 International Conference on Computing and Intelligent Reality Technologies (ICCIRT), Coimbatore, India, 2024, pp. 391-396, https://doi.org/10.1109/ICCIRT59484.2024.10922033