IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13518

Cyberbullying Detection in Social Networks

Sunitha N[1], Suma Putti[2], Srujana B N[3], T M Chandana[4], Suresh K[5]

Student, Department of Computer Science and Engineering, Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari-583103[1-4]

Assistant Professor, Computer Science and Engineering, Rao Bahadur Y Mahabaleswarappa Engineering College Ballari-583103[5]

Abstract: The detection of hate speech in social media is a crucial task. The uncontrolled spread of hate has the potential to gravely damage our society, and severely harm marginalized people or groups. Hate speechis one of the most dangerous of these activities, so users have to protect themselves from these activities from YouTube, Facebook, and Twitter etc.To identify word similarities in the tweets made by bullies and make use of machine learning and can develop an ML model automatically detect social media bullying actions. However, many social media bullying detection techniques have been implemented, but many of them were textual based. The objective of our project work is to show the implementation of NLP and CNN which detects bullied tweets, posts, etc. A machine learning model is proposed to detect and prevent bullying on Twitter. Two classifiers i.e. NLP(Natural Language Processing) are used for identifying the complete sentence in the comments and CNN(Convolution Neural Networks) for image identification. Both NLP and CNN were able to detect the true positives with more accuracy. Also, Twitter API is used to fetch tweets and tweets are passed to the model to detect whether the tweets are bullying or not.

Keywords: Natural Language Processing(NLP),Long Short Term Memory(LSTM)

I. INTRODUCTION

Hate crimes are unfortunately nothing new in society. However, social media and other means of online communication have begun playing a larger role in hate crimes. For instance, suspects in several recent hate-related terror attacks had an extensive social media history of hate related posts, suggesting that social media contributes to their radicalization. Detecting hate speech is a challenging task as First, there are disagreements in how hate speech should be defined. This means that some content can be considered hate speech to some and not to others, based on their respective definitions. Some recent studies claimed favorable results to detect automatic hate speech in the text.

The proposed solutions employ the different feature engineering techniques and ML algorithms to classify content as hate speech. Regardless of this extensive amount of work, it remains difficult to compare the performance of these approaches to classify hate speech content.

To the best of our knowledge, the existing studies lack the comparative analysis of different feature engineering techniques and ML algorithms. To address this, we propose a system which employs natural language processing techniques and the classification is done using ensemble machine learning approach that incorporates various classification techniques.

II. LITRATURE SURVEY

1. Title: "Detecting Cyberbullying in Social Media Using Deep Learning" Authors: Zhang, X., Luo, L., Fung, P., & Liu, S.Year: 2018 Summary: This paper proposes a deep learning approach for cyberbullying detection on social media platforms, using techniques like LSTM and CNN. It discusses the importance of feature extraction and the impact of imbalanced datasets.

2. Title: "Cyberbullying Detection on Instagram Using Machine Learning Techniques"Authors: Al-Garadi, M. A., Varathan, K. D., & Ravana, S. D.Year: 2018Summary: The authors present a study on detecting cyberbullying on Instagram using a range of machine learning techniques such as SVM, Naïve Bayes, and Random Forest. They also explore the use of text and image features for better accuracy.

DOI: 10.17148/IJARCCE.2024.13518

3. Title: "A Survey of Cyberbullying Detection Techniques using Machine Learning and Social Pedagogical Measures"Authors: Gaur, M., & Ahuja, A.Year: 2019 Summary: This survey paper provides an overview of cyberbullying detection methods that combine machine learning with social pedagogical measures. It discusses the importance of context and social factors in identifying cyberbullying incidents..

4. Title: "Cyberbullying Detection on Twitter: A Natural Language Processing Approach" Authors: De Silva, D., Manogaran, G., & Lopez, D.Year: 2019 Summary: This research focuses on cyberbullying detection on Twitter using natural language processing (NLP) techniques. It explores sentiment analysis, topic modeling, and feature engineering for improved detection.

III. METHODOLOGY

- **Data Collection:** Gather diverse datasets containing text, images, and videos from various online platforms, ensuring a representative sample of cyberbullying instances.
- **Preprocessing:** Clean and preprocess the data, including text normalization, image resizing, and video frame extraction, to prepare it for analysis.
- **Feature Extraction:** Extract relevant features from text, images, and videos, including linguistic patterns, visual cues, and audio attributes.
- **Multimodal Fusion:** Combine multimodal features using fusion techniques, such as late fusion or deep learning-based fusion, to create comprehensive representations of content.
- **Machine Learning Models:** Train machine learning models, including deep neural networks and ensemble methods, on the fused features for cyberbullying detection.
- **Contextual Analysis:** Develop algorithms to consider conversational context, user interactions, and platform-specific norms to enhance detection accuracy.
- **Behavioral Analytics:** Incorporate behavioral analysis methods to identify patterns in user behavior that indicate potential bullies or victims.
- **Explainable AI:** Implement explainability techniques to provide clear rationales for detection decisions, improving transparency.
- **User Feedback Integration:** Design a user-friendly interface for real-time alerts and user feedback, enabling community engagement and refinement of the system.
- **Testing and Evaluation:** Conduct comprehensive testing and evaluation using standard metrics, such as precision, recall, and F1-score, on diverse datasets and social media platforms to validate the system's effectiveness and usability.

IV. DIAGRAMS

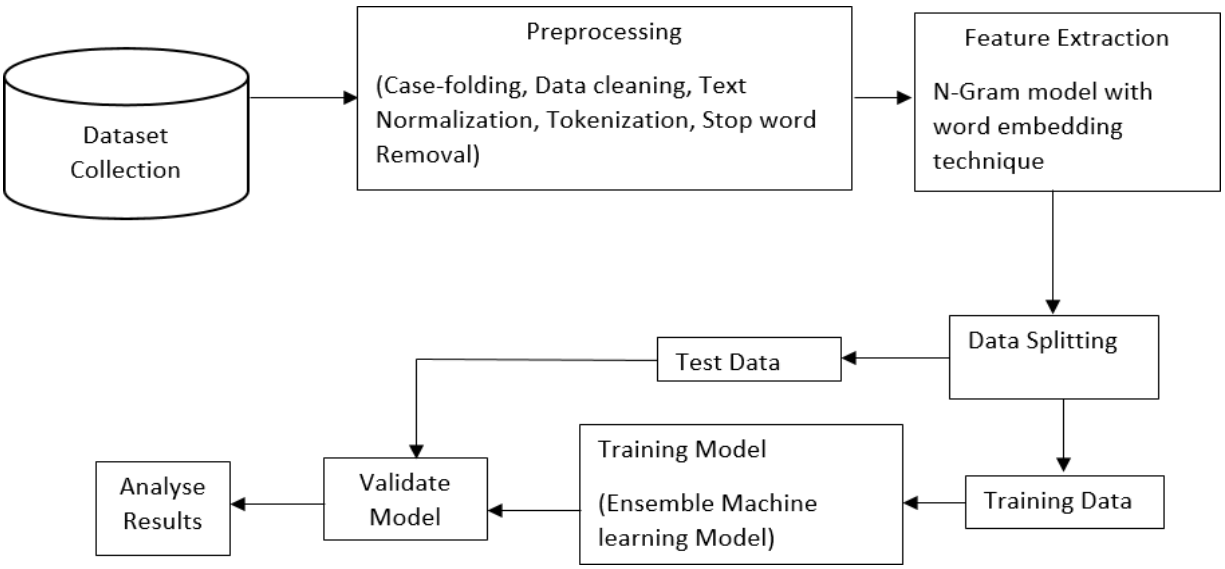

Figure 1 : System architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13518

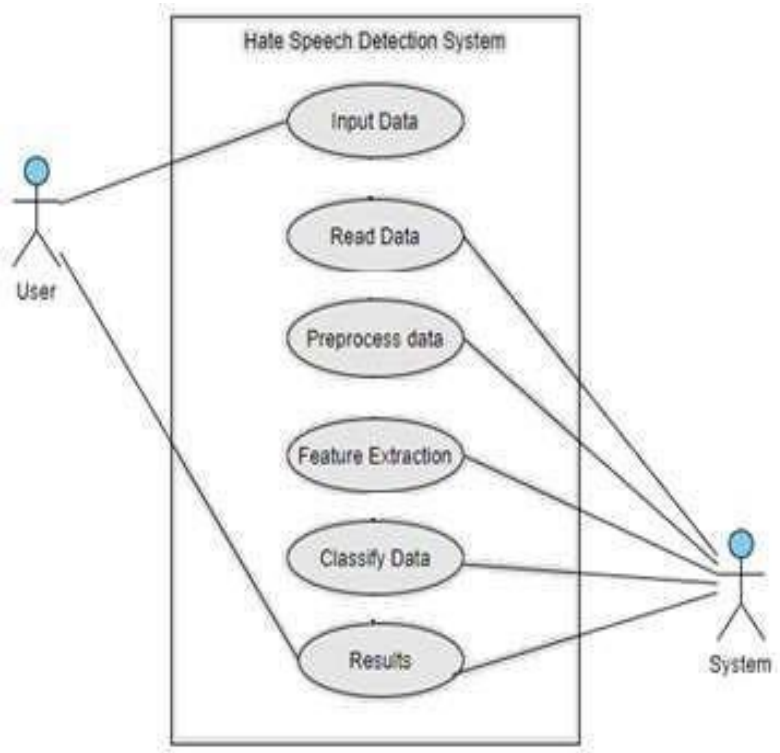

Figure 2: Use case diagram

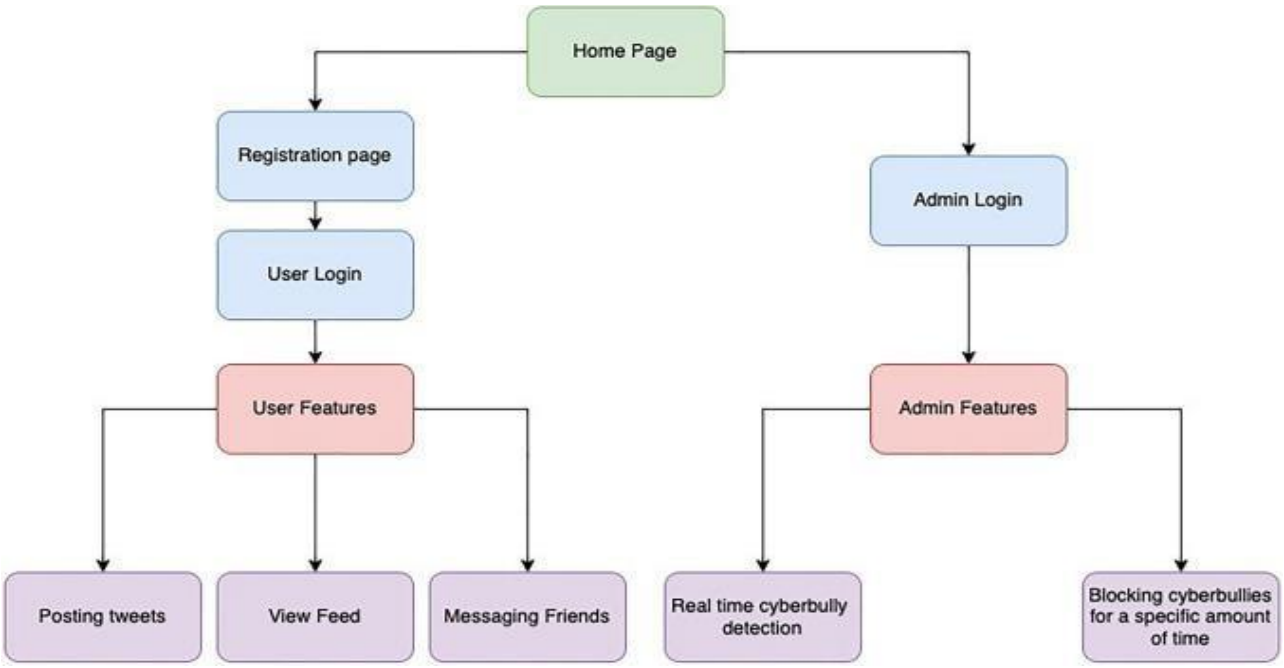

Figure 3: WebApp architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13518

V. RESULT

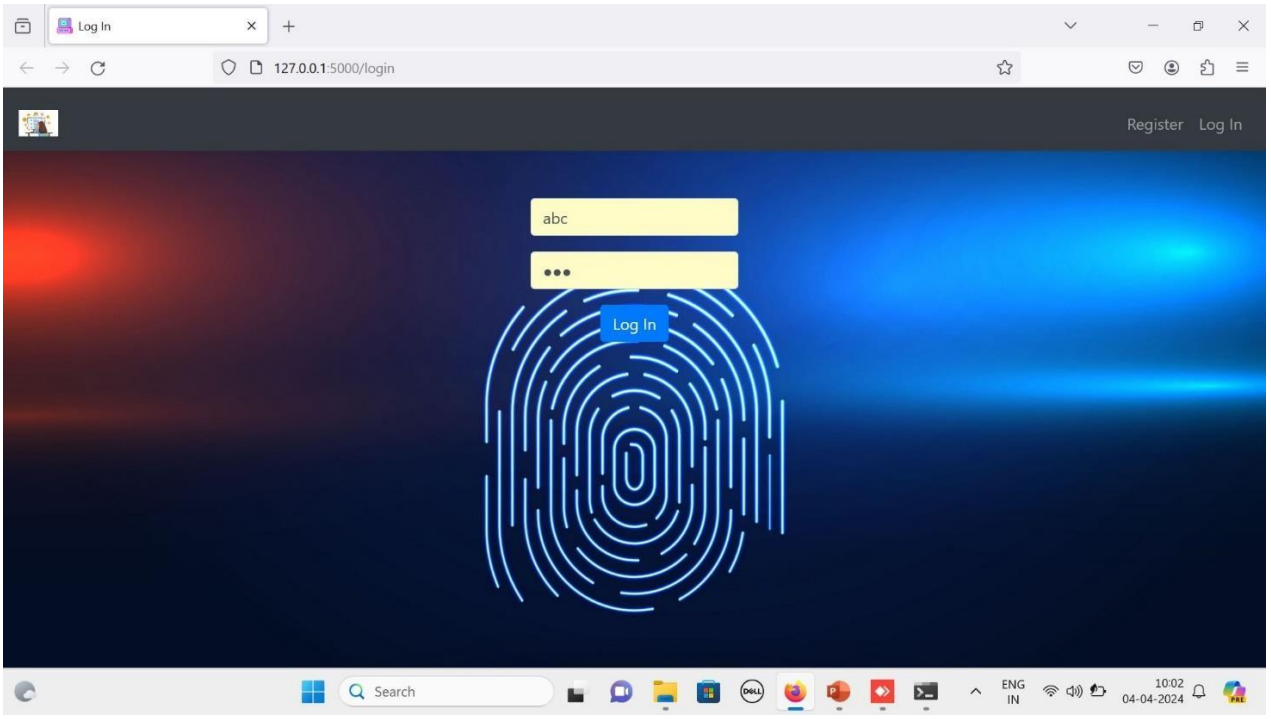

FIGURE : HOME PAGE

FIGURE : AFTER LOGGING IN

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13518

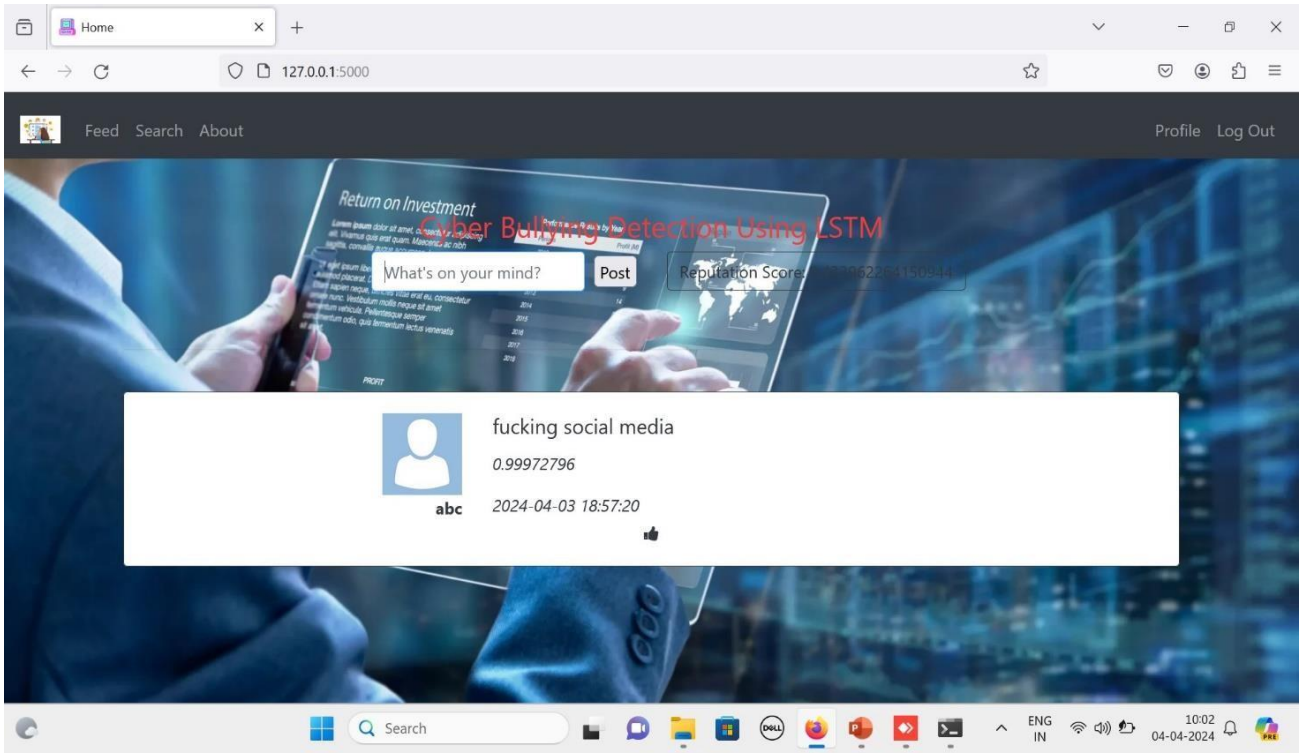

FIGURE: THE RESULT

VI. CONCLUSION

The utilization of Long Short-Term Memory (LSTM) networks for cyberbullying detection represents a promising avenue, with current models demonstrating commendable performance in distinguishing harmful online behavior. As technology progresses, the future holds exciting possibilities for the field, including the exploration of advanced neural architectures, multimodal analysis, and real-time detection.

Additionally, the development of context-aware models, personalized approaches, and the integration of behavioral analysis will contribute to more nuanced and accurate detection systems. However, ethical considerations, such as user privacy and fairness, must be at the forefront of development efforts. A collaborative, global approach involving researchers, policymakers, and educators is imperative to tackle the multifaceted challenges of cyberbullying effectively.

As technology and society continue to evolve, the ongoing commitment to innovation, education, and ethical standards will be crucial in creating robust and responsible solutions for the detection and prevention of cyberbullying.

VII. FUTURE RESEARCH DIRECTION

The future scope of cyberbullying detection using LSTM and related technologies holds promise for advancements in advanced neural architectures, multimodal analysis, real-time detection, personalized models, explainable AI, and global collaboration. As technology evolves, there is a growing emphasis on context-aware models, behavioral analysis, transfer learning, and adversarial robustness.

The ethical considerations surrounding user privacy, bias, and fairness are crucial, necessitating the development of guidelines and standards for responsible deployment. Additionally, ongoing efforts in education and awareness are essential for fostering responsible online behavior.

The interdisciplinary nature of addressing cyberbullying, coupled with advancements in technology and a global collaborative approach, will likely shape the future of effective and ethical cyberbullying detection solutions.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13518

REFERENCES

[1]. Rui Zhao, Kezhi Mao "Cyber Bullying Detection based on Semantic - Enhanced Marginalized Denoising Auto encoders"

[2] Elaheh Raisi, Bert Huang "Weakly Supervised Cyberbullying Detection with Participant Vocabulary Consistency" Social Network Analysis and Mining, May 24,2018.

[3] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Houng Wei, Hao, Bo Xu "Attention- based Bi-directional Long Short-Term Memory Network for Relation Classification" proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 207-212,August 12,2016.

[4] Nitish Srivastava, Geoffrey Honton, Alex Krizhevsky, Iiya Sutskever, Ruslan Salakhutdinov "Dropout: A Simple way to Prevent Neural Networks from Overfitting" Journal of Machine Learning Research 1929-1958,2015

[5] Alexis Conneau, Holger Schwenk,Yann Le cun "Very Deep CNN for Text Classification" Association for Computational Linguistics, Volume1, pages 1107-1116,7 April 2017.

[6]] MS. Snehal Bhoir, Tushar Ghorpade, Vanita Mane "Comparative Analysis of Different Word Embedding Models" IEEE,2017.

[7] Elaheh Raisis, Bert Huang "Cyberbullying Detection with Weakly Supervised Machine Learning" International Conference on Advances in Social Networks Analysis and Mining IEEE/ACM,2017

