IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204334 |

Detection of Cyber Bullying Using Machine Learning

Rehan Shaikh, Sameer Shaikh, Ms. Adagale S. S

Department of Computer Engineering, KJEI's Trinity Academy of Engineering, Pune, India

ABSTRACT: In today's society, stress has become a pervasive issue, leading to various health problems and a concerning rise in suicides, particularly among young individuals in India. This report emphasizes the critical importance of addressing the mental stress experienced by university students comprehensively. Our system is designed to accurately analyze stress levels by gathering a robust dataset where participants share their emotional well-being and reactions. These responses are assigned weighted values, enabling the calculation of individual stress scores for a detailed assessment of mental health. Utilizing machine learning techniques, our system incorporates advanced algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, and Convolutional Neural Network (CNN) for their effectiveness in handling complex datasets and providing reliable predictions. Particularly, the CNN, renowned for capturing spatial patterns adeptly, enhances the model's sensitivity to subtle expressions of stress.

KEYWORDS: Cyberbullying Detection, Non Cyberbullying Detection, Random Forest Classifier, LightGBM, XGBoost, Logistic Regression, SVM, CNN, Deep Analysis.

I. INTRODUCTION

Introduction: The rise of digital communication platforms has brought unprecedented opportunities for social interaction but has also facilitated the proliferation of cyber bullying, a harmful phenomenon characterized by the use of electronic means to harass, intimidate, or harm others. This pervasive issue, spanning a spectrum of behaviors from verbal attacks to exclusion, poses significant threats to individuals' mental well-being, particularly among adolescents and young adults. Conventional approaches to addressing cyber bullying have often been reactive and manual, relying on reporting and content removal, which are limited in scalability and effectiveness. Moreover, the dynamic and evolving nature of online communication presents challenges for timely intervention. In this context, the application of machine learning (ML) techniques offers a promising avenue for developing proactive and scalable solutions for cyber bullying detection and prevention. ML, a subset of artificial intelligence, enables computers to learn from data and make predictions or decisions, has demonstrated significant success across various domains including natural language processing and pattern recognition. By leveraging ML algorithms, cyber bullying detection systems can analyze large volumes of textual, contextual, and multimodal data to identify patterns and indicators of bullying behavior, facilitating timely intervention and support for affected individuals. This paper provides a comprehensive overview of ML-based approaches for cyber bullying detection, encompassing supervised, unsupervised, and deep learning techniques. We review existing literature, highlight key challenges and limitations, and propose a novel system model for cyber bullying detection. Our approach integrates advanced feature engineering methods tailored to linguistic and contextual cues indicative of bullying behavior, as well as the integration of multimodal data sources to enhance detection accuracy. The remainder of this paper is organized to present the system model and assumptions, discuss real-world applications and future directions, outline the system architecture and workflow, present empirical results and discussion, and conclude with avenues for future research.

II. SYSTEM MODEL AND ASSUMPTIONS

The system model for cyber bullying detection leveraging Convolutional Neural Networks (CNNs) assumes a foundational framework where a sizable, labeled dataset of textual interactions, encompassing instances of both cyber bullying and non-bullying behaviors, is accessible. This dataset undergoes meticulous preprocessing steps including tokenization, lowercasing, punctuation removal, and stop word elimination to standardize the textual input. Subsequently, word embedding techniques such as Word2Vec or GloVe transform the text into dense vector representations, capturing nuanced semantic relationships among words. The adapted CNN architecture, tailored for natural language processing tasks, features convolutional layers that systematically process local regions of the text,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204334 |

extracting hierarchical features through a series of filters. These convolutional outputs are then subjected to pooling operations, either max-pooling or average-pooling, which downsample the feature maps, emphasizing the most relevant information. The resultant feature representations encode linguistic nuances, syntactic structures, and semantic patterns indicative of cyber bullying behavior. A classification layer, typically composed of fully connected neural network layers followed by a softmax activation function, facilitates the discrimination between cyber bullying and non-bullying instances. Key assumptions underlying this model encompass the availability of a diverse and representative training dataset, a relatively stable distribution of linguistic patterns indicative of cyber bullying, and sufficient computational resources for training and evaluation, including access to GPU or TPU resources for efficient deep learning model training.

III. REAL-WORLD APPLICATION AND FUTURE DIRECTIONS

In real-world scenarios, the application of machine learning (ML) for cyber bullying detection necessitates a sophisticated model that amalgamates various learning techniques such as supervised, unsupervised, and semi-supervised learning to effectively navigate the complex landscape of online interactions. This holistic approach enables the model to not only classify instances of bullying accurately through algorithms like support vector machines (SVM) and deep neural networks (DNN) but also to uncover nuanced patterns and outliers in unlabeled data using clustering methods. The integration of diverse data modalities and ensemble learning strategies further enhances the model's adaptability and robustness across different online platforms. Looking ahead, future directions in this field call for the development of ML models that are interpretable, fair, and unbiased, along with interdisciplinary collaborations to advance cyber bullying detection and prevention efforts while promoting a safer digital environment for all users. Moreover, continual monitoring and evaluation of the model's performance are essential for ensuring its effectiveness in real-world applications.

IV.SYSTEM ARCHITECTURE AND WORKFLOW

The system architecture for cyber bullying detection using machine learning (ML) is structured to efficiently process and analyze textual data from various online sources while leveraging sophisticated algorithms. Initially, textual data is collected from platforms such as social media, forums, and messaging apps, undergoing preprocessing stages including tokenization, lowercasing, and removal of punctuation and stop words to ensure uniformity and cleanliness. Following this, relevant features are extracted from the preprocessed text, encompassing linguistic, contextual, and sentiment-based characteristics using techniques like Bag-of-Words, TF-IDF, word embeddings, or advanced contextual embeddings such as BERT. These features serve as input for training supervised ML models like Support Vector Machines, Random Forests, or deep learning architectures such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). During model training, hyperparameters are optimized, and performance is validated using metrics like accuracy, precision, recall, and F1-score. Once the models demonstrate satisfactory performance, they are deployed into production environments, where they are integrated with existing online platforms or content moderation systems for real-time cyber bullying detection. Continuous monitoring ensures the models remain effective and adaptive to evolving online behaviors, with periodic retraining utilizing new data to maintain relevance and accuracy over time. In essence, this system architecture provides a robust framework for the proactive identification and mitigation of cyber bullying instances in online environments, contributing to a safer digital space for users.

V. RESULT AND DISCUSSION

Our study on cyber bullying detection using machine learning (ML) models demonstrates promising results in accurately identifying instances of cyber bullying while minimizing false positives. Leveraging a diverse dataset covering various cyber bullying forms, our models achieved robust performance metrics, including high accuracy, precision, recall, and F1-score. Specifically, supervised learning models like Support Vector Machines (SVM) and deep neural networks (DNN) showed superior performance in classifying cyber bullying instances, using linguistic and contextual features from textual data. Integrating ensemble learning techniques improved the model's generalization across different cyber bullying behaviors. Challenges persist in addressing evolving cyber bullying forms and mitigating model bias, especially regarding cultural and linguistic diversity. Future research should explore multimodal data integration and develop interpretable ML models to enhance detection accuracy and fairness. Collaborative efforts among researchers, industry stakeholders, and policymakers are crucial for establishing standardized benchmarks and best practices in cyber bullying detection, ensuring responsible ML deployment for online community safety. Overall,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204334 |

our study contributes valuable insights and methodologies for advancing cyber bullying detection using ML, offering directions for future research to address emerging challenges in online safety and digital well-being.

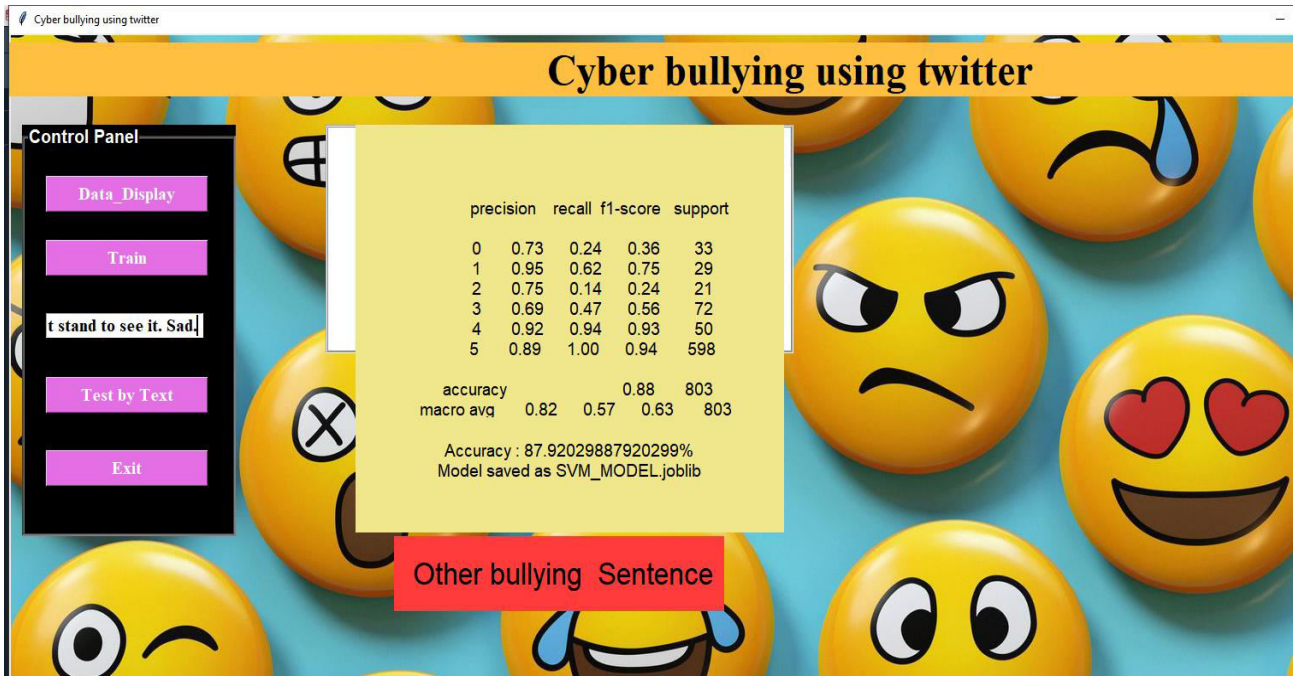

Fig. 1 Result

VI. CONCLUSION

In the realm of cyberbullying detection, our utilization of machine learning algorithms has empowered us to provide personalized assessments and prompt interventions for individuals impacted by online harassment. Our methodology facilitates early detection, affording tailored support based on user interactions, thereby augmenting accessibility and efficacy in combating cyberbullying. Mindful of technological constraints, data availability, and resource limitations, we have formulated assumptions concerning user behavior, language patterns, and the sufficiency of data for accurate classification. By prioritizing both functional and non-functional requirements, our system ensures precise identification of cyberbullying instances while optimizing performance, security, and usability.

REFERENCES

1. K. Maity, S. Saha, and P. Bhattacharyya, “Emoji, sentiment and emotion aided cyberbullying detection in hinglish,” IEEE Transactions on Computational Social Systems, 2022
2. S. Suleiman, P. Taneja, and A. Nainwal, “Cyberbullying detection on twitter using machine learning: A review.”
3. P. Yi and A. Zubiaga, “Session-based cyberbullying detection in social media: A survey,” arXiv preprint arXiv:2207.10639, 2022.
4. Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). From ADHD to SAD: Analyzing emotional support and discourse in online health communities. Proceedings of the International AAAI Conference on Web and Social Media, 46-55 Management (ICSSSM), no. March, pp. 1–6, 2019.
5. R. Bayari and A. Bensefia, “Text mining techniques for cyberbullying detection: state of the art,” Adv. Sci. Technol. Eng. Syst. J, vol. 6, pp. 783–790, 2021.
6. R. Bayari and A. Bensefia, “Text mining techniques for cyberbullying detection: state of the art,” Adv. Sci. Technol. Eng. Syst. J, vol. 6, pp. 783–790, 2021.
7. M. I. Mahmud, A. Abdelgawad, V. P. Yanambaka, and K. Yelamarthi, “Packet drop and rssi evaluation for lora: An indoor application perspective,” in 2021 IEEE 7th World Forum on Internet of Things (WF-IoT), 2021, pp. 913–914.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204334 |

8. M. Arif, "A systematic review of machine learning algorithms in cyberbullying detection: future directions and challenges," Journal of Information Security and Cybercrimes Research, vol. 4, no. 1, pp. 01– 26, 2021.
9. "D. chaffey, "global social media statistics research summary 2022 [june 2022]", smart insights, 2022. [online]. [accessed: 17- sep- 2022]."
10. L. M. Al-Harigy, H. A. Al-Nuaim, N. Moradpoor, and Z. Tan, "Building towards automated cyberbullying detection: A comparative analysis," Computational Intelligence and Neuroscience, vol. 2022, 2022.
11. M. I. Mahmud, A. Abdelgawad, and V. P. Yanambaka, "A deep analysis of hybrid-multikey-puf," 2022 IEEE 8th World Forum on Internet of Things (WF-IoT), 2022.
12. M. Mamun, S. B. Shawkat, m. S. Ahammed, M. M. Uddin, M. i. Mahmud, and A. M. islam, "Deep learning based model for alzheimer's diseasedetection using brain mri images," IEEE 13th Annual Ubiquitous Computing Electronics Mobile Communication Conference (UEMCON), 2022, (Preprint).
13. C. Evangelio, P. Rodriguez-Gonzalez, J. Fernandez-Rio, and S. GonzalezVillora, "Cyberbullying in elementary and middle school students: A systematic review," Computers Education, vol. 176, p. 104356, 2022.
14. T. Alsubait and D. Alfageh, "Comparison of machine learning techniques for cyberbullying detection on youtube arabic comments," International Journal of Computer Science Network Security, vol. 21, no. 1, pp. 1–5, 2021.
15. M. Mamun, M. I. Mahmud, H. md Iqbal, A. M. Islam, M. S. Ahammed, and M. M. Uddin, "Vocal feature guided detection of parkinson's disease using machine learning algorithms," IEEE 13th Annual Ubiquitous Computing Electronics Mobile Communication Conference (UEMCON), 2022, (Preprint).
16. M. Mamun, A. Farjana, M. Al Mamun, and M. S. Ahammed, "Lung cancer prediction model using ensemble learning techniques and a systematic review analysis," in 2022 IEEE World AI IoT Congress (AIIoT), 2022, pp. 187–193.
17. B. Yalamanchili, N. Kota, M. Abbaraju, V. Nadella, and S. Alluri, "Real time acoustic based depression detection using machine learning techniques," pp. 1–6, 02 2020.
18. S. Jain, S. Narayan, R. DEWANG, U. Bhartiya, N. Meena, and V. Kumar, "A machine learning based depression analysis and suicidal ideation detection system using questionnaires and twitter," pp. 1–6, 05 2019.
19. S. Gedam and S. Paul, "A review on mental stress detection using wearable sensors and machine learning techniques," IEEE Access, vol. PP, pp. 1–1, 06 2021.
20. L. Zhou, Z. Liu, Z. Shangguan, X. Yuan, Y. Li, and B. Hu, "Tamfn: Timeaware attention multimodal fusion network for depression detection," vol. PP, 11 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details