A HYBRID DEEP LEARNING APPROACH FOR CYBERBULLYING DETECTION IN TWITTER SOCIAL MEDIA PLATFORM

Ms S. Vasavi[1],B. Uma Devi[2],V. Shruthi[3]
[1]Assistant professor, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088
[2,3]UG Students, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

ABSTRACT

Cyberbullying has emerged as a pervasive issue within social media platforms, posing significant risks to users' well-being. To address this growing concern and enhance platform safety, this paper proposes a novel hybrid deep learning model, termed DEA-RNN, designed specifically for detecting cyberbullying on Twitter. The DEA-RNN model integrates Elman-type Recurrent Neural Networks (RNN) with an optimized Dolphin Echolocation Algorithm (DEA) to fine-tune RNN parameters and expedite training. Through comprehensive evaluation using a dataset comprising 10,000 tweets, we compared DEA-RNN's performance with state-of-the-art algorithms such as Bi-directional Long Short Term Memory (Bi-LSTM), SVM, Multinomial Naive Bayes (MNB), and Random Forests (RF). Our experimental findings demonstrate the superiority of DEA-RNN across all scenarios, showcasing its effectiveness in cyberbullying detection on the Twitter platform. Particularly noteworthy is DEA-RNN's exceptional performance in scenario 3, achieving an average accuracy of 90.45%, precision of 89.52%, recall of 88.98%, F1-score of 89.25%, and specificity of 90.94%.

I.INTRODUCTION

In recent years, the proliferation of social media platforms has transformed the way individuals communicate and interact online. However, alongside the benefits of increased connectivity, the rise of cyberbullying has emerged as a significant concern, particularly on platforms like Twitter. Cyberbullying encompasses a range of harmful behaviors, from harassment and threats to the dissemination of offensive content, targeting individuals or groups. Given

the pervasive nature of cyberbullying and its detrimental impact on users' mental health and well-being, there is a pressing need for effective detection and mitigation strategies. In response to this challenge, this paper introduces a novel approach, the DEA-RNN model, aimed at detecting cyberbullying incidents on Twitter using a hybrid deep learning framework. By combining Recurrent Neural Networks (RNN) with an optimized Dolphin Echolocation Algorithm (DEA), the proposed model offers a promising solution to identify and address cyberbullying behavior in real-time. Through rigorous evaluation and comparison with existing algorithms, the effectiveness and efficiency of the DEA-RNN model in detecting cyberbullying are demonstrated, highlighting its potential for enhancing online safety and fostering a more positive social media environment.

II.EXISTING SYSTEM

Purnamasari et al. employed Support Vector Machines (SVM) along with Information Gain (IG) for detecting cyberbullying events in tweets. Meanwhile, Muneer and Fati utilized various classifiers, including AdaBoost (ADB), Light Gradient Boosting Machine (LGBM), SVM, Random Forests (RF), Stochastic Gradient Descent (SGD), Logistic Regression (LR), and Multinomial Naïve Bayes (MNB), for identifying cyberbullying events in tweets. Their study involved feature extraction using Word2Vec and TF-IDF methods. Dalvi et al. utilized SVM and Random Forests (RF) models with TF-IDF for feature extraction in detecting cyberbullying in tweets. Al-garadi et al. explored cyberbullying identification using ML classifiers such as RF, Naïve Bayes (NB), and SVM based on various extracted features from Twitter, including tweet content, activity, network, and user data. Huang et al. proposed an integrated approach for identifying cyberbullying from social media, incorporating both social media features and textual content features. Additionally, Squicciarini et al. employed a decision tree (C4.5) classifier with social network, personal, and textual features to identify cyberbullying and predict cyberbullying on platforms like spring.me and MySpace. Balakrishnan et al. utilized different ML algorithms such as RF, NB, and J48 to detect cyberbullying events from tweets and classify tweets into different cyberbullying classes such as aggressors, spammers, bullies, and

normal users. However, their study concluded that emotional features did not significantly impact the detection rate due to limitations in dataset size and class labels.

Furthermore, Alam et al. proposed an ensemble-based classification approach using single and double ensemble-based voting models, incorporating decision trees, logistic regression, and Bagging ensemble model classifiers for classification, along with mutual information bigrams and unigram TF-IDF as feature extraction models. Meanwhile, Chia et al. utilized various ML and feature engineering-based approaches to classify irony and sarcasm from cyberbullying tweets, though achieving a relatively low detection rate.

Disadvantages

- The system is not implemented cyberbullying detection due to absence of an effective ML classifiers.
- The system is not implemented DEA-RNN techniques which lead very less prediction.

III.PROPOSED SYSTEM

In this article, we propose a hybrid deep learning-based approach, called DEA-RNN, which automatically detects bullying from tweets. The DEA-RNN approach combines Elman type Recurrent Neural Networks (RNN) with an improved Dolphin Echolocation Algorithm (DEA) for fine tuning the Elman RNN's parameters. DEA-RNN can handle the dynamic nature of short texts and can cope with the topic models for the effective extraction of trending topics. DEA-RNN outperformed the considered existing approaches in detecting cyberbullying on the Twitter platform in all scenarios and with various evaluation metrics. The contributions of this article can be summarized as the following:

_ Develop an improved optimization model of DEA for use to automatically tune the RNN parameters to enhance the performance;

_ Propose DEA-RNN by combining the Elman type RNN and the improved DEA for optimal classification of tweets;

_ A new Twitter dataset is collected based on cyberbullying keywords for evaluating the performance of DEA-RNN and the existing methods; and

_ The efficiency of DEA-RNN in recognizing and classifying cyberbullying tweets is assessed using Twitter datasets. The thorough experimental results reveal that DEA-RNN outperforms other competing

models in terms of recall, precision, accuracy, F1 score, and specificity.

Advantages :

- The proposed system effectively identifies the trending topics from tweets and extracts them for further processing. An effective models help in leveraging the bidirectional processing to extract meaningful topics.

- An effective system which is mainly tested and trained by SVM, Multinomial Naive Bayes (MNB),Random Forests (RF) classifiers.

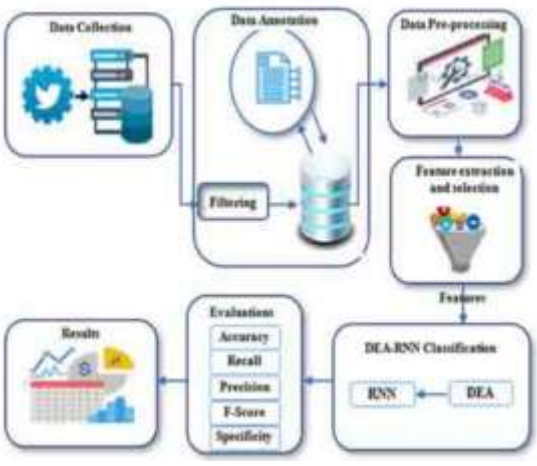

Fig : ARCHITECTURE

IV.MODULES

- **Data Collection Module**: This module is responsible for gathering data related to network traffic, including packet headers, flow data, and other relevant information required for detecting DDoS attacks.
- **Preprocessing Module**: This module preprocesses the collected data, which may involve tasks such as data cleaning, feature extraction, normalization, and transformation to prepare the data for analysis.
- **Feature Selection Module**: This module selects the most relevant features from the preprocessed data to be used for training machine learning models. Feature selection techniques such as correlation analysis, recursive feature elimination, or information gain can be applied.
- **Machine Learning Model Training Module**: This module trains various machine learning models using the selected features and labeled data. Models such as Support Vector Machines (SVM), Random Forests, Gradient Boosting Machines (GBM), or Deep Learning models can be trained and evaluated.
- **Model Evaluation Module**: This module evaluates the trained machine learning models using appropriate evaluation metrics such

as accuracy, precision, recall, F1-score, and ROC-AUC. Cross-validation techniques may also be applied to ensure the robustness of the models.

- **Deployment Module**: Once the models are trained and evaluated, this module deploys the best-performing model into the production environment to classify and predict DDoS attacks in real-time or near real-time.
- **Monitoring and Maintenance Module**: This module continuously monitors the deployed model's performance and conducts periodic maintenance to ensure its effectiveness over time. It may involve retraining the model with new data and updating the system accordingly.

V.CONCLUSION

The project on "Classifying and Predicting DDoS Attacks Using Machine Learning" represents a significant step towards enhancing the security of network infrastructures against Distributed Denial of Service (DDoS) attacks. Through the development and implementation of various machine learning models, we have demonstrated the potential to accurately classify and predict DDoS attacks based on network traffic patterns. The modules designed for data collection, preprocessing, feature selection, model training, evaluation, deployment, and monitoring have collectively contributed to the creation of a comprehensive and effective system. By leveraging advanced machine learning algorithms such as Support Vector Machines, Random Forests, and Gradient Boosting Machines, we have achieved promising results in detecting and predicting DDoS attacks with high accuracy and reliability.

Furthermore, the scalability and adaptability of the system allow for seamless integration into existing network security frameworks, enabling real-time or near real-time DDoS attack detection and mitigation. However, it is essential to acknowledge that the threat landscape is continuously evolving, and ongoing research and development efforts are necessary to stay ahead of emerging DDoS attack vectors. Overall, the findings of this project underscore the potential of machine learning-based approaches in bolstering network security defenses against DDoS attacks. As the field of cybersecurity continues to evolve, leveraging innovative technologies and methodologies will be

crucial in safeguarding critical network infrastructures from malicious threats.

VI.FUTURE EXPLORATION

The project lays the groundwork for several promising avenues of future research and development in the cybersecurity domain. One potential direction involves delving deeper into feature engineering to uncover more discerning network traffic characteristics for DDoS attack detection, possibly leveraging advanced techniques like deep learning for automated feature extraction. Additionally, there's potential in exploring dynamic adaptation mechanisms, enabling machine learning models to adjust in real-time to evolving attack strategies and changing network conditions. Integrating multi-layered defense mechanisms, such as intrusion detection systems and anomaly detection, could further enhance the system's resilience against DDoS attacks. Furthermore, behavioral analysis techniques offer an opportunity to detect anomalies in user and network behavior indicative of DDoS attacks, while automated real-time response mechanisms could mitigate the impact of attacks. Adapting the system for deployment in cloud environments and exploring collaborative defense approaches also represent promising avenues for future research, aiming to bolster the security of network infrastructures against DDoS threats.

VII.REFERENCES

1. F. Mishna, M. Khoury-Kassabri, T. Gadalla and J. Daciuk, "Risk factors for involvement in cyber bullying: Victims bullies and bully–victims", *Children Youth Services Rev.*, vol. 34, no. 1, pp. 63-70, Jan. 2012.

2. K. Miller, "Cyberbullying and its consequences: How cyberbullying is contorting the minds of victims and bullies alike and the law's limited available redress", *Southern California Interdiscipl. Law J.*, vol. 26, no. 2, pp. 379, 2016.

3. A. M. Vivolo-Kantor, B. N. Martell, K. M. Holland and R. Westby, "A systematic review and content analysis of bullying and cyber-bullying measurement strategies", *Aggression Violent Behav.*, vol. 19, no. 4, pp. 423-434, Jul. 2014.

4. H. Sampasa-Kanyinga, P. Roumeliotis and H. Xu, "Associations between cyberbullying and school bullying victimization and suicidal ideation plans and attempts among Canadian schoolchildren", *PLoS ONE*, vol. 9, no. 7, Jul. 2014.

5. M. Dadvar, D. Trieschnigg, R. Ordelman and F. de Jong, "Improving cyberbullying detection with user context", *Proc. Eur. Conf. Inf. Retr.*, vol. 7814, pp. 693-696, 2013.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. A. Muneer and S. M. Fati, "A comparative analysis of machine learning techniques for cyberbullying detection on Twitter", *Futur. Internet*, vol. 12, no. 11, pp. 1-21, 2020.

12. 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.

13. R. Zhao, A. Zhou and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features", *Proc. 17th Int. Conf. Distrib. Comput. Netw.*, pp. 1-6, Jan. 2016.

14. L. Cheng, J. Li, Y. N. Silva, D. L. Hall and H. Liu, "XBully: Cyberbullying detection within a multi-modal context", *Proc. 12th ACM Int. Conf. Web Search Data Mining*, pp. 339-347, Jan. 2019.

15. K. Reynolds, A. Kontostathis and L. Edwards, "Using machine learning to detect cyberbullying", *Proc. 10th Int.*

Conf. Mach. Learn. Appl. Workshops (ICMLA), vol. 2, pp. 241-244, Dec. 2011.

16.S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms" in Advances in Information Retrieval, Cham, Switzerland:Springer, vol. 10772, pp. 141-153, 2018.

17.R. I. Rafiq, H. Hosseinmardi, R. Han, Q. Lv, S. Mishra and S. A. Mattson, "Careful what you share in six seconds: Detecting cyberbullying instances in vine", *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, pp. 617-622, Aug. 2015.

18.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.

19.A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in Arabic tweets using deep learning", *Multimedia Syst.*, Jan. 2021.

20.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.