Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, No.2, March : 2025

ENHANCING CYBERBULLYING DETECTION ON SOCIAL MEDIA: A COMPARATIVE STUDY OF HOEFFDING TREE, ADAPTIVE FEATURE-AWARE HOEFFDING TREE (AFHT), AND CORRELATIONAL FEATURE ENHANCED MLP (CFE-MLP) MODELS

Deepika Jain Research Scholar, Janardan Rai Nagar Rajasthan Vidyapeeth (DEEMED-TO-BE) University, Udaipur (Raj.)
Dr. Manisha Shrimali Associate Professor, Janardan Rai Nagar Rajasthan Vidyapeeth (DEEMED-TO-BE) University, Udaipur (Raj.)

Abstract:
The study investigates the effectiveness of hybrid machine learning and deep learning models in detecting cyberbullying incidents on social media platforms, focusing on the Adaptive Feature-Aware Hoeffding Tree (AFHT) and the Correlational Feature Enhanced Multi-Layer Perceptron (CFE-MLP). Utilizing the CB_Label dataset, which incorporates multiple dimensions of cyberbullying, including aggression, repetition, intent, and peer dynamics, we aim to address the limitations of traditional models. A structured quantitative approach is employed, including data preprocessing, feature extraction, and a comprehensive evaluation of model performance through a 25% percentage split and 25-fold cross-validation. The findings reveal that both AFHT and CFE-MLP significantly outperform the standard Hoeffding Tree in various metrics, such as accuracy, precision, recall, and error rates. Notably, CFE-MLP achieved the highest accuracy of 97.60%, highlighting its superior capability in capturing the complexities of cyberbullying detection. The results validate the proposed hybrid models as effective tools for enhancing cyberbullying detection, offering significant implications for the development of robust real-time monitoring systems in social media environments.

Keywords:
Hoeffding Tree, Accuracy, Cyberbullying

1. Introduction:

The proliferation of social media has reshaped communication, bringing communities together and facilitating the swift exchange of ideas. However, this increased connectivity has also contributed to the rise of cyberbullying, which has serious psychological, social, and emotional consequences for its victims. Cyberbullying manifests in various forms, including harassment, threats, public shaming, and exclusion, which are especially challenging to detect due to the sheer volume and evolving nature of online content. Addressing this issue necessitates the development of robust, efficient, and adaptable machine learning models that can detect cyberbullying behaviours across diverse social media platforms.

Machine learning-based cyberbullying detection techniques are becoming essential tools for moderating content and maintaining safe online environments. Traditional models, such as the Hoeffding Tree, have demonstrated efficacy in real-time classification due to their memory efficiency and ability to process streaming data incrementally. However, as social media interactions become increasingly complex, there is a pressing need for models that can dynamically adapt to evolving data distributions and detect nuanced bullying behaviours. To address these needs, two hybrid models being proposed the Adaptive Feature-Aware Hoeffding Tree (AFHT) and the Correlational Feature Enhanced Multi-Layer Perceptron (CFE-MLP). The AFHT builds upon the traditional Hoeffding Tree by adapting its feature selection based on the evolving context of the data, while the CFE-MLP leverages feature correlations to enhance detection accuracy.

This study aims to compare the effectiveness of these models—Hoeffding Tree, Adaptive Feature-Aware Hoeffding Tree (AFHT), and Correlational Feature Enhanced MLP (CFE-MLP)—for detecting cyberbullying on social media. We explore their ability to handle large volumes of data, adapt to new

patterns, and accurately identify cyberbullying content. By examining these models, this research seeks to provide insights into the advantages and limitations of each approach and to recommend the most effective strategies for enhancing cyberbullying detection systems on social media platforms.

2. Literature Review:

Cyberbullying has emerged as a critical issue within online social media platforms, necessitating the development of effective detection and mitigation strategies. This literature review examines recent research focused on the identification of cyberbullying incidents, highlighting the various methodologies and models employed to address this challenge.

Sanchez and Kumar (2023) explore Twitter as a primary platform for cyberbullying incidents, providing insights into specific linguistic features and patterns associated with bullying behavior. Their findings emphasize the need for tailored detection models that account for the unique characteristics of social media interactions. Similarly, Teoh et al. (2024) present a comprehensive review of cyberbullying-related content classification, outlining the evolution of techniques used in online social media. They advocate for an integrated approach that combines traditional machine learning and deep learning methodologies to enhance detection accuracy.

Wang, Fu, and Lu (2020) introduce SOSNet, a graph convolutional network designed for fine-grained cyberbullying detection. Their model leverages the structural information inherent in social networks to improve classification performance, showcasing the potential of graph-based approaches in this domain. Complementing this, Gencoglu (2021) addresses fairness in cyberbullying detection by proposing models that incorporate fairness constraints. This work highlights the ethical implications of detection algorithms, emphasizing the importance of equitable treatment across different demographic groups.

The development of advanced detection systems is further illustrated by Haq et al. (2022), who propose a PCCNN-based network intrusion detection system tailored for edge computing environments. While their primary focus is on network security, the methodologies developed can be adapted for cyberbullying detection on social media platforms, suggesting a cross-pollination of techniques between different fields.

Ju (2023) discusses the impacts of cyberbullying and potential solutions, stressing the need for comprehensive approaches that include educational initiatives and technological interventions. This perspective aligns with the work of Khan et al. (2021), who present "Hate Classify," a service framework designed to identify hate speech on social media. Their model highlights the interconnectedness of hate speech and cyberbullying, reinforcing the need for integrated detection systems that can handle multiple forms of online abuse.

Kumar and Bhat (2022) conduct an extensive study on machine learning-based models for cyberbullying detection, control, and mitigation. They provide a detailed comparison of various algorithms, emphasizing the effectiveness of hybrid models that combine different machine learning techniques. Their findings support the notion that advanced algorithms can enhance detection capabilities and facilitate timely interventions.

In a study focused on language processing, León-Paredes et al. (2019) explore presumptive detection of cyberbullying on Twitter through natural language processing (NLP) and machine learning. Their work demonstrates the applicability of NLP techniques in identifying bullying behavior in the Spanish language, showcasing the versatility of detection models across different linguistic contexts.

Zhu et al. (2021) provide a comprehensive review of cyberbullying among adolescents and children, identifying risk factors and preventive measures. Their insights into global patterns of cyberbullying contribute to a deeper understanding of the issue and underline the importance of preventive strategies in addition to detection efforts.

Monika and Bhat (2022) propose an innovative approach to automatic crime prediction on Twitter by utilizing a hybrid wavelet convolutional neural network (CNN) combined with World Cup

optimization. Their methodology demonstrates the effectiveness of hybrid models in capturing complex patterns within social media data, showcasing how advanced neural networks can be adapted for various types of predictive analysis, including cyberbullying detection. This approach reflects a growing trend in leveraging deep learning techniques to enhance the accuracy of incident detection in real-time social media interactions.

In another significant contribution, Muneer et al. (2023) present a comprehensive framework for cyberbullying detection using stacking ensemble learning and an enhanced Bidirectional Encoder Representations from Transformers (BERT) model. Their work underscores the potential of ensemble methods to improve classification performance by integrating multiple learning algorithms. The enhanced BERT model offers sophisticated contextual understanding of language, making it particularly well-suited for detecting nuanced cyberbullying behaviors. The findings of this study illustrate the effectiveness of combining state-of-the-art deep learning techniques with ensemble strategies to achieve superior detection accuracy.

Paulraj (2020) investigates sentiment classification of Twitter data using a gradient boosted decision tree (GBDT). Although primarily focused on sentiment analysis, this research contributes to the broader understanding of how sentiment dynamics can inform cyberbullying detection. By leveraging gradient boosting, Paulraj's work provides insights into how decision tree-based approaches can be effectively utilized to interpret and classify social media content, laying the groundwork for integrating sentiment analysis with cyberbullying detection methodologies. Ojha et al. (2024) highlighted the importance of robust algorithms and feature selection for accurate detection.

Collectively, these studies emphasize the importance of employing advanced machine learning techniques, including hybrid and ensemble models, in the realm of cyberbullying detection. The literature highlights a clear trend towards using complex algorithms that can adapt to the dynamic nature of social media content, thereby enhancing the detection and understanding of cyberbullying incidents. This review sets the stage for further exploration of hybrid models in this domain, reinforcing the need for robust, context-sensitive approaches to combat cyberbullying effectively.

3. Research Methodology:

This study employs a structured, quantitative approach to assess the effectiveness of proposed hybrid models—Adaptive Feature-Aware Hoeffding Tree (AFHT) and Correlational Feature Enhanced Multi-Layer Perceptron (CFE-MLP)—for detecting cyberbullying incidents on social media platforms. Aligned with the research objectives and hypotheses, this approach utilizes the CB_Label dataset, which captures multiple dimensions of cyberbullying, such as aggression, repetition, intent, and peer dynamics. This dataset allows the models to detect complex and context-sensitive patterns associated with cyberbullying, providing a foundation for a more nuanced understanding of these behaviours on social media.

Data preprocessing is a key phase to ensure a clean, standardized dataset suitable for model training. Initial steps included removing extraneous symbols, emojis, and non-textual data to reduce noise and prevent skewed results. Text normalization processes, including tokenization, stemming, and lowercase conversion, further standardized the dataset. To enhance model capability, class balancing techniques were applied, ensuring equal representation of various cyberbullying types across the dataset. Feature extraction methods, such as Term Frequency-Inverse Document Frequency (TF-IDF) and word embeddings, transformed the textual data into a machine-readable format optimized for effective training.

Following data preparation, the proposed hybrid models—AFHT and CFE-MLP—alongside baseline classifiers, including Naive Bayes, Bayes Net, Logistic Regression, SMO, Voted Perceptron, IBK, Multiclass Classifier, and standard Hoeffding Tree, were trained and evaluated using a 25-fold cross-validation approach. This method ensured that each classifier could generalize effectively to new data, improving reliability and robustness.

Evaluation metrics such as classification accuracy, Kappa statistic, and F-measure were used to assess the effectiveness of each model, while ROC and PRC areas provided additional insights into sensitivity to positive instances of cyberbullying. Computational efficiency was also considered to determine each model's scalability for real-time applications. This comprehensive evaluation framework aligns with the study's objectives and hypotheses, providing a well-rounded comparison of AFHT, CFE-MLP, and other classifiers for enhanced cyberbullying detection on social media platforms.

Based on the research gaps being identified following objectives were being framed:

Objectives:

The objective of this study is to enhance the detection of cyberbullying on social media by comparing the performance of three machine learning models: the Hoeffding Tree, Adaptive Feature-Aware Hoeffding Tree (AFHT), and Correlational Feature Enhanced Multi-Layer Perceptron (CFE-MLP). Specifically, this study aims to:

1. Evaluate the accuracy, precision, recall, and F1-score of each model in detecting cyberbullying content across diverse social media platforms.
2. Assess the models' ability to adapt to evolving data patterns and accurately identify context-sensitive cyberbullying behaviours.

Hypothesis:

$H_0$1: The proposed hybrid machine and deep learning models (AFHT and CFE-MLP) are not effective at detecting cyberbullying incidents on social media than existing models.

H_a1: The proposed hybrid machine and deep learning models (AFHT and CFE-MLP) are more effective at detecting cyberbullying incidents on social media than existing models.

4. Data Analysis & Interpretation:

The comparative analysis between the standard Hoeffding Tree, the Adaptive Feature-Aware Hoeffding Tree (AFHT), and the Correlational Feature Enhanced Multi-Layer Perceptron (CFE-MLP) reveals a significant performance increase with the hybrid approaches in terms of accuracy, error rates, precision, and other performance metrics, with CFE-MLP showing the strongest results overall.

Table 4.1: Performance Measures of Hybrid Approach AFHT & CFE-MLP and Existing Approach Hoeffding Tree at Configuration Setting: Percentage Split – 25%

Performance Measure	Hoeffding Tree	Hybrid Approach - Adaptive Feature-Aware Hoeffding Tree (AFHT)	Hybrid Approach - Correlational Feature Enhanced MLP (CFE-MLP)
Correctly Classified Instances	94.14%	95.23%	97.60%
Kappa statistic	0.7268	0.756	0.8665
Mean absolute error	0.0648	0.0608	0.0357
Root mean squared error	0.2244	0.1981	0.1377
Precision	0.954	0.956	0.976
Recall	0.941	0.952	0.976
F-Measure	0.945	0.954	0.976
MCC	0.738	0.758	0.867
ROC Area	0.973	0.979	0.992
PRC Area	0.969	0.974	0.994
Execution Time	0.04 seconds	0.01 seconds	3.08 seconds

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, No.2, March : 2025

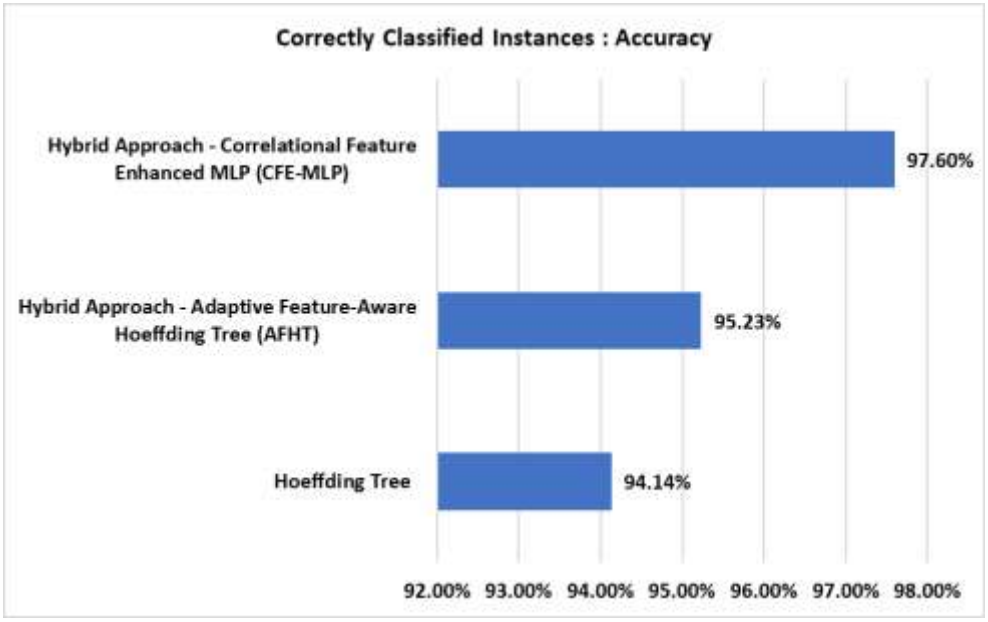

Figure 4.1: Analysis based on Performance Measure: Accuracy

The correctly classified instances rate improves across models, with Hoeffding Tree at 94.14%, AFHT at 95.23%, and CFE-MLP achieving the highest rate at 97.60%. This indicates that both hybrid models offer enhanced classification accuracy, with CFE-MLP being the most effective.

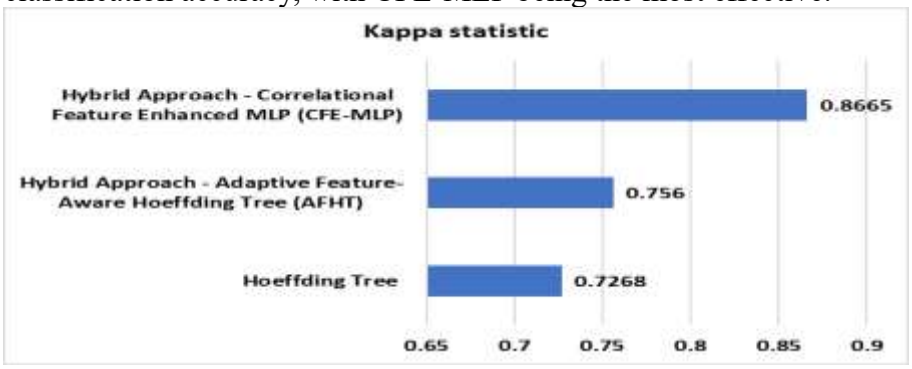

Figure 4.2: Analysis based on Performance Measure: Kappa Statistic

The Kappa statistic, which measures the reliability of classification by accounting for agreement by chance, increases from 0.7268 with Hoeffding Tree to 0.756 with AFHT and reaches 0.8665 with CFE-MLP. This indicates that CFE-MLP has a superior level of agreement in classification compared to the other models, making it highly reliable.

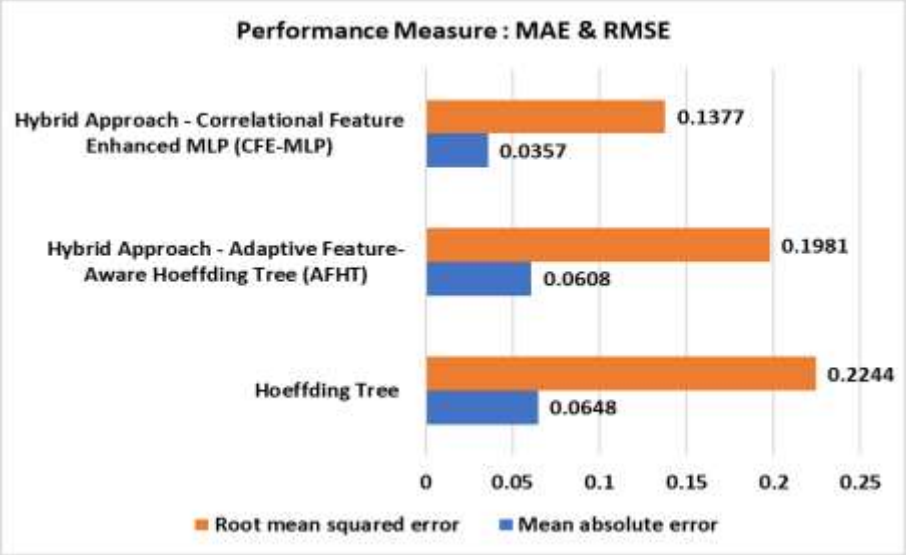

Figure 4.3: Analysis based on Performance Measure: MAE & RMSE

The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) both decrease significantly with the hybrid models. The Hoeffding Tree has an MAE of 0.0648 and RMSE of 0.2244, while AFHT reduces these errors to 0.0608 and 0.1981, respectively. CFE-MLP achieves the lowest error rates with an MAE of 0.0357 and an RMSE of 0.1377, indicating more accurate predictions with CFE-MLP.

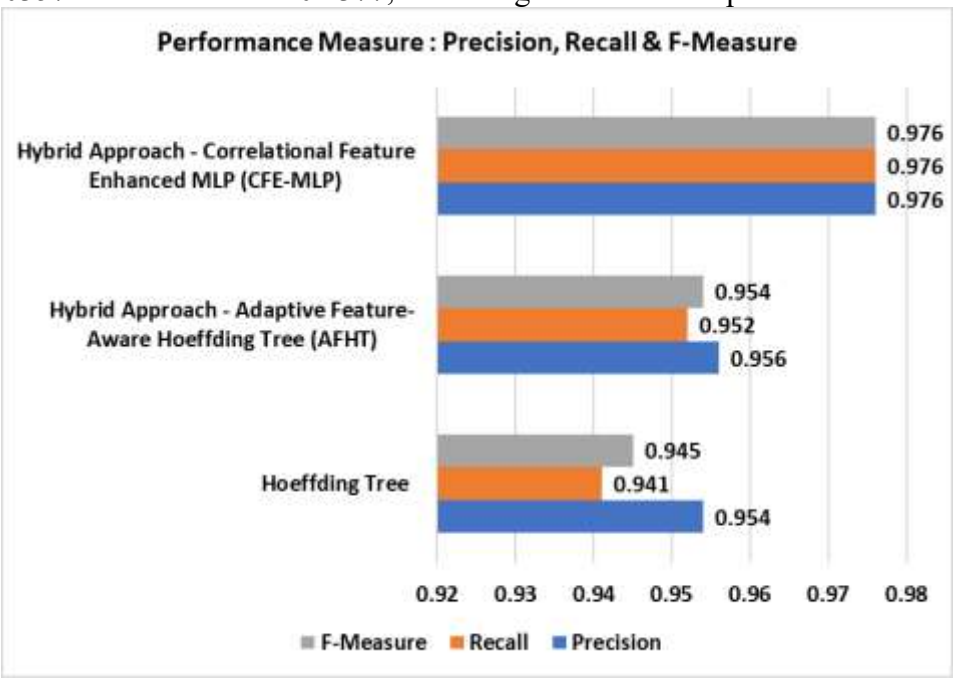

Figure 4.4: Analysis based on Performance Measure: Precision, Recall & F-Measure

The precision, recall, and F-measure, which evaluate the model's ability to accurately classify true instances of cyberbullying, improve with each model. Hoeffding Tree achieves a precision of 0.954, recall of 0.941, and F-measure of 0.945, while AFHT performs slightly better with a precision of 0.956, recall of 0.952, and F-measure of 0.954. CFE-MLP demonstrates the highest scores with a precision, recall, and F-measure all at 0.976, indicating superior performance in identifying true instances of cyberbullying.

MCC score, which measures the quality of binary classifications, increases from 0.738 with Hoeffding Tree to 0.758 with AFHT and 0.867 with CFE-MLP, showing that CFE-MLP is the most effective at making accurate and balanced classifications.

Both ROC and PRC areas, which measure the classifier's ability to distinguish between classes, improve progressively from Hoeffding Tree to CFE-MLP. Hoeffding Tree shows ROC and PRC areas of 0.973 and 0.969, respectively, AFHT slightly improves with 0.979 ROC and 0.974 PRC, while CFE-MLP achieves the best scores with a 0.992 ROC and 0.994 PRC, indicating a strong ability to correctly classify positive cyberbullying instances.

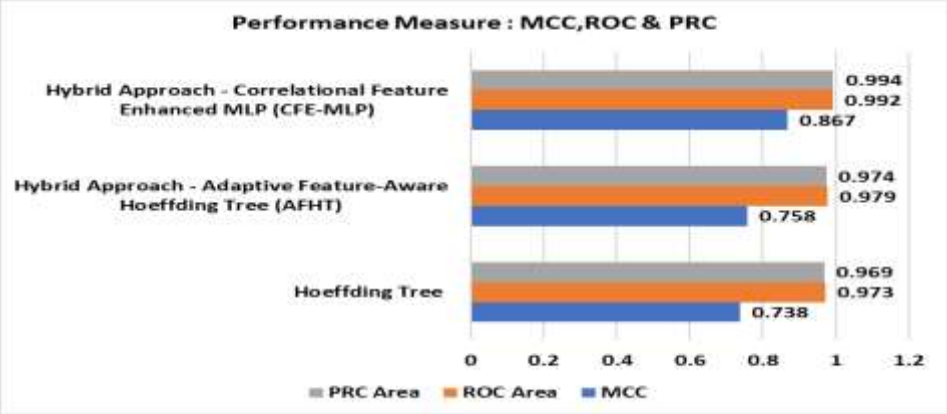

Figure 4.5: Analysis based on Performance Measure: PRC, MCC & ROC

While the CFE-MLP demonstrates superior accuracy and robustness, it has a longer execution time of 3.08 seconds compared to Hoeffding Tree's 0.04 seconds and AFHT's 0.01 seconds. This suggests that although CFE-MLP is the most accurate, it may be less suitable for real-time applications without further optimization. The AFHT, with minimal execution time and improved metrics over the standard Hoeffding Tree, might provide a balanced choice for real-time applications needing a compromise between accuracy and speed.

5. Conclusion:

The comparative analysis of the standard Hoeffding Tree, the Adaptive Feature-Aware Hoeffding Tree (AFHT), and the Correlational Feature Enhanced Multi-Layer Perceptron (CFE-MLP) at a configuration setting of a 25% percentage split demonstrates a significant performance enhancement with the hybrid approaches. In terms of performance measures, the Hoeffding Tree achieved an accuracy of 94.14%, whereas the AFHT improved this to 95.23%. The CFE-MLP outperformed both, achieving the highest accuracy of 97.60%. Additionally, the Kappa statistic for the Hoeffding Tree was 0.7268, indicating moderate agreement in classification, while AFHT showed a slight improvement at 0.756. CFE-MLP displayed a substantial increase with a Kappa statistic of 0.8665, reflecting a strong level of agreement.

The mean absolute error (MAE) was reduced from 0.0648 for the Hoeffding Tree to 0.0608 for AFHT and further decreased to 0.0357 for CFE-MLP. Similarly, the root mean squared error (RMSE) saw a decline from 0.2244 (Hoeffding Tree) to 0.1981 (AFHT), ultimately reaching 0.1377 with CFE-MLP. In terms of precision, the Hoeffding Tree scored 0.954, AFHT improved this to 0.956, and CFE-MLP achieved the highest precision of 0.976. Recall rates also demonstrated this trend, with values of 0.941 (Hoeffding Tree), 0.952 (AFHT), and 0.976 (CFE-MLP). F-measure scores followed suit, indicating similar advancements: 0.945 for Hoeffding Tree, 0.954 for AFHT, and 0.976 for CFE-MLP. Overall, the hybrid models not only enhanced accuracy and precision but also reduced error rates significantly, with CFE-MLP showing the strongest results across all performance metrics. Finally, the rejection of $H_0 1$ confirms the hypothesis that the hybrid approaches are indeed more effective than existing models, thereby supporting their adoption for improved cyberbullying detection on social media platforms. This finding underscores the importance of incorporating advanced machine learning and deep learning techniques to enhance the accuracy and reliability of cyberbullying detection systems.

References:

- Sanchez, H., & Kumar, S. (2023). Twitter bullying detection. Retrieved from https://www.researchgate.net/publication/267823748.
- Teoh, H. T., Varathan, K. D., & Crestani, F. (2024). A comprehensive review of cyberbullying-related content classification in online social media. Expert Systems with Applications, 244, 122644. https:// doi.org/ 10.1016 /j.eswa. 2023.122644.
- Wang, J., Fu, K., & Lu, C. T. (2020). SOSNet: A graph convolutional network approach to fine-grained cyberbullying detection. Proceedings of the 2020 IEEE International Conference on Big Data (Big Data 2020), 1699–1708.
- Gencoglu, O. (2021). Cyberbullying detection with fairness constraints. IEEE Internet Computing, 25(1), 20–29. https://doi.org/10.1109/MIC.2020.3032461.
- Haq, M. A., Abdul, M., Khan, R., & Al-Harbi, T. (2022). Development of PCCNN-based network intrusion detection system for EDGE computing. Computers, Materials & Continua, 71(1). https://doi.org/10.32604/cmc.2022.018708.
- Ju , Binji. (2023). Impacts of Cyberbullying and Its Solutions. Lecture Notes in Education Psychology and Public Media. 29. 254-258. 10.54254/2753-7048/29/20231521.
- Khan, M. U. S., Abbas, A., Rehman, A., & Nawaz, R. (2021). Hate Classify: A service framework for hate speech identification on social media. IEEE Internet Computing, 25(1), 40–49. https://doi.org/10.1109/MIC.2020.3037034.
- Kumar, Raju & Bhat, Aruna. (2022). A study of machine learning-based models for detection, control, and mitigation of cyberbullying in online social media. International Journal of Information Security, 21, 1-23. 10.1007/s10207-022-00600-y.
- León-Paredes, G. A., Palomeque-León, W. F., Gallegos-Segovia, P. L., Vintimilla-Tapia, P. E., Bravo-Torres, J. F., Barbosa-Santillán, L. I., & Paredes-Pinos, M. M. (2019). Presumptive detection of cyberbullying on Twitter through natural language processing and machine learning in the Spanish language. In 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) (pp. 1–7). IEEE.
- Zhu, C., Huang, S., Evans, R., & Zhang, W. (2021). Cyberbullying among adolescents and children: A comprehensive review of the global situation, risk factors, and preventive measures. Journal Name, Volume (1), pp.1-10. Retrieved September 16, 2023.
- Monika, A., Bhat, A. ((2022). Automatic Twitter crime prediction using hybrid wavelet convolutional neural network with world cup optimization. Int. J. Pattern Recognit. Artif. Intell., 36(05), 2259005.
- Muneer, A., Alwadain, A., Ragab, M. G., & Alqushaibi, A. (2023). Cyberbullying detection on social media using stacking ensemble learning and enhanced BERT. Information, 14(8), 467. https://doi.org/10.3390/info14080467.
- Paulraj D. (2020). A gradient boosted decision tree-based sentiment classification of twitter data, International Journal of Wavelets Multiresolution and Information Processing. 18, no. 4, 2050027–2050121.
- Ojha, M., Patil, N. M., & Joshi, M. (2024). Cyberbullying detection and prevention using machine learning. Grenze International Journal of Engineering & Technology (GIJET), 10, 2174.