AOI: 10.11234.2024.00016
ACT2024: Special Proceedings of Applied Computer Technology

Digital harassment Detection using Multi-Model Supervised Methods on X-Data

[1]Anbarasi, [2]V Allen Jerome, [3]Ruthwik Reddy Baini, [4]N.Deepa *, [5]Susheela Vishnoi, [6]Sushama
[1,2,3,4,]SRM Institute of Science and Technology, School of Computing, Kattankulathur, Chennai,India
[5,6] Manipal University Jaipur, Jaipur, Rajasthan

ABSTRACT
The increasing frequency of negative online behaviour has drawn a lot of interest to predictive analysis of cyberbullying on Twitter data. This research suggests a unique method for predicting incidents of cyberbullying on Twitter by using a multi-model supervised strategy. The proposed approach aims to enhance efficacy and enhance the precision of cyberbullying detection through the integration of textual, social, and network attributes. The models are trained and assessed using Twitter data sets that include both cyberbullying and non- cyberbullying events. Sentiment analysis, bag-of-words, and semantic similarity are examples of textual features; follower count and account age are examples of social features. Analysing the user's interaction patterns and network structure is part of network features. The models are created and assessed using a variety of machine learning algorithms, including support vector machines (SVM), random forests (RF), and neural networks (NN). The outcomes of the experiments show that the combined strategy outperforms the individual models in terms of predictive performance. The significance of feature selection in enhancing model accuracy is further emphasised by the study. This research helps establish practical tactics and countermeasures to lessen the negative impacts of cyberbullying by precisely detecting incidences of cyberbullying on Twitter.

Keywords: Support Vector Machines, Random Forests, neural networks, textual features, social features, network features, cyberbullying prediction, Twitter data, multi- model supervised approach, and feature selection.

I. I. INTRODUCTION

In the current digital era, cyberbullying has grown in importance, and social media sites like Twitter have become hubs for this destructive activity. Therefore, it is crucial to provide methods and resources that are efficient in detecting and stopping cyberbullying. This article employs a multi-model supervised strategy to introduce a predictive analysis method for identifying cyberbullying through Twitter data. Our objective is to utilize natural language processing and machine learning for automated identification of cyberbullying occurrences within tweets. Utilising extensive Twitter data analysis, our methodology aims to offer significant perspectives and forecasts concerning instances of cyberbullying. Given the size and complexity of social media platforms, we think an automated approach is essential because manual monitoring and intervention can be ineffective and time-consuming.

Our methodology is based on the use of several supervised models that have been trained with labelled data. The purpose of these models is to identify relevant characteristics in tweets, like sentiment, linguistic patterns, and lexical clues that point to cyberbullying. Our goal is to increase our cyberbullying detection system's accuracy and dependability by merging the predictions of several models.

We have gathered a sizable dataset of actual tweets from Twitter, encompassing both cases of cyberbullying and non-cyberbullying, in order to validate our methodology. Our supervised models use this dataset as training data so they may get better at making predictions by seeing a wide variety of examples. A subset of the data has also been personally labelled by us as a gold standard for assessing how well our predictive models work.

We use a mix of ML algorithms, including SVM, RF's, and NN, to develop our multi-model supervised methodology. Every model is trained using distinct feature sets and generates unique predictions using test data. A voting process is used to aggregate the various forecasts and establish the final prediction, resulting in a thorough and reliable study.

Our study's findings present encouraging opportunities for identifying and stopping cyberbullying on Twitter. Social media companies may prevent cyberbullying and create a safer online environment by promptly detecting cases of cyberbullying. Furthermore, by using our predictive analysis to shed light on the frequency, patterns, and trends of cyberbullying, politicians, educators, and academics will be better able to develop focused interventions and tactics. Using multi-model supervised approaches, our research concludes with a novel way to predictive analysis of cyberbullying using Twitter data. Our goal is to create a strong tool that can automatically identify instances of cyber bullying by utilising machine learning and natural language processing. This will help to mitigate and avoid this widespread problem that occurs online.

II. RELATED WORKS

1. FAEO-ECNN: Cyber bullying Detection in Social Media Platforms Using Topic Modelling and Deep L[Murshed et al., 2023]:

In this research, the authors introduce FAEO-ECNN, a technique for detecting cyberbullying on social media platforms. This method merges deep learning methods and topic modelling to pinpoint cyberbullying content. The FAEO-ECNN model exhibits encouraging outcomes in effectively identifying instances of cyberbullying.

2. Email-Based Cyberstalking Detection On Textual Data Using Multi-Model Soft Voting Technique of Machine Learning Approach [Gautam & Bansal, 2023]:

Gautam and Bansal present a deep learning approach for the identification of cyberstalking via email communication. They make use of a soft voting approach with multiple models to classify textual data and identify instances of cyberstalking. The proposed method shows effectiveness in detecting cyberstalking incidents in emails.

3. Cyberbullying Detection Using Weakly Supervised and Fully Supervised Learning [Abhishek, 2022]:

Abhishek delves into weakly and fully supervised learning strategies for detecting cyberbullying. The research centers on harnessing extensive social media datasets and illustrates the efficacy of both approaches in precisely identifying cyberbullying instances.

4. Cyberbullying and Cyber Violence Detection: A Triangular User-Activity-Content View [Wang et al., 2022]:

Wang and colleagues introduce a triangular perspective on user engagement and content interaction for identification of online harassment and cyber aggression. The approach considers user behaviour, activity patterns, and content analysis to identify instances of online harassment and cyber aggression. The study highlights the significance of considering multiple perspectives for accurate detection.

5. Cyberbullying Detection: An Ensemble Learning Approach [Roy et al., 2022]:

Roy et al. proposed a collaborative learning method for identifying cyberbullying. The study combines multiple learning algorithms to enhance the effectiveness of models for detecting cyberbullying. The results show that the ensemble approach outperforms individual algorithms in accurately identifying cyberbullying incidents.

6. Performance Analysis of Annotation Detection

Techniques for Cyber-Bullying Messages Using Word- Embedded Deep Neural Networks [Giri & Banerjee, 2023]:

Giri and Banerjee examine various methods for identifying cyberbullying in messages by employing deep neural networks with word embeddings. The study evaluates the performance of these techniques and provides insights into the effectiveness of different approaches in detecting cyberbullying content.

7. Improving Cyberbullying Detection with User Interaction [Ge et al., 2021]:

Ge et al. propose a method to improve cyberbullying detection by incorporating user interaction. The study explores the use of user feedback to enhance the accuracy of cyberbullying detection algorithms. The results indicate that user interaction can effectively improve the performance of cyberbullying detection systems.

8. Analysis of Deep Learning-Based Approaches for Spam Bots and Cyberbullying Detection in Online Social Networks [Kumar et al., 2024]:

Kumar et al. analyse various deep learning methods for identifying fraudulent accounts and online harassment in social media platforms. The research assesses the effectiveness. of different models and the efficacy of deep learning techniques in identifying these malicious activities.

9. Detection of Types of Cyber-Bullying Using Fuzzy C- Means Clustering and XGBoost Ensemble Algorithm [Süzen & Duman, 2021]:

Süzen and Duman propose a method for the detection of different types of cyberbullying using fuzzy c-means clustering and an XGBoost ensemble algorithm. The study demonstrates the effectiveness of the proposed approach in accurately classifying various types of cyberbullying instances.

10. A Review on Deep-Learning-Based Cyberbullying Detection [Hasan et al., 2023]:

Hasan et al. provide an analysis of deep learning methodologies for identifying cyberbullying. The study covers various deep learning models, techniques, and evaluation methods used in cyberbullying detection. The review highlights the advancements and challenges in the field of deep-learning-based cyberbullying detection.

III. EXISTING SYSTEM

The existing system for Predictive analysis of cyberbullying on Twitter data using supervised learning with different techniques has several disadvantages.

Firstly, one of the main drawbacks of the system is its heavy reliance on supervised techniques. While supervised learning algorithms are effective in predicting cyberbullying based on labelled training data, they require a large amount of manually labelled data to be trained. This process can be time-consuming and labor-intensive, as it involves human annotators manually identifying instances of cyberbullying in the data. Furthermore, the reliance on pre-labeled data can limit. The system's capacity to adjust to emerging or evolving cyberbullying forms , as it may not have sufficient training data for these emerging patterns.

Another disadvantage is the challenge of dealing with the dynamic and ever-changing nature of Twitter data. The system may struggle to keep up with the rapid pace at which new tweets are being posted, making it difficult to provide real-time analysis and prediction of cyberbullying incidents. Additionally, Twitter is known for its brevity, with users often communicating through short and cryptic messages. This can make it challenging to accurately interpret the meaning and context of tweets, which is crucial for effective cyberbullying detection.

Furthermore, the reliance on Twitter data alone may limit the system's effectiveness. Cyberbullying can occur on various social media platforms and online forums, and focusing solely on Twitter data may lead to an incomplete understanding of the overall cyberbullying landscape. Incorporating data from other platforms and sources could provide a more comprehensive analysis and improve the accuracy of predictions.

Lastly, the existing system may raise privacy concerns. In order to perform predictive analysis, the system needs access to a large amount of user data, including personal information and private messages. This raises ethical questions regarding the collection, storage, and use of user data, which must be carefully addressed to ensure user privacy and consent.

In conclusion, the existing system for cyberbullying predictive analysis on Twitter data has several disadvantages, including the heavy reliance on supervised techniques, challenges in dealing with dynamic data, limitations in interpreting context, the focus on a single platform, and potential privacy concerns. Overcoming these drawbacks is crucial for the development of a comprehensive and effective system for cyberbullying detection and prevention.

IV. PROPOSED SYSTEM

The proposed work aims to develop a predictive analysis model for cyberbullying detection on Twitter using a multi-model supervised technique. Cyberbullying has become a prevalent issue in the digital age, negatively impacting individuals' mental health and overall well-being. Social media platforms, especially Twitter, have become breeding grounds for cyberbullying due to their widespread usage and the ease of spreading harmful content anonymously. To address this problem, this research proposes a predictive analysis model that combines multiple machine learning algorithms to accurately detect instances of cyberbullying on Twitter.

The model will leverage a multi-model approach, which involves training and combining multiple supervised learning algorithms. This method enhances accuracy by integrating the strengths of various algorithms and alleviates the limitations of any one algorithm. By using a diverse range of methods, the model will be equipped to handle the multidimensional aspects of cyberbullying, such as textual analysis, sentiment analysis, and content classification.

The proposed work will utilise a large dataset of Twitter data, which includes both cyberbullying instances and non-bullying instances, to train and evaluate the effectiveness of the model. Various features such as textual content, user profiles, and interaction patterns will be extracted and used as input for the model. The dataset undergoes preprocessing to eliminate noise, manage missing data, and equalize class distribution, ensuring impartial training.

To assess the model's effectiveness, various metrics including accuracy, precision, recall, and F1 score will be employed. Cross-validation methods will gauge the model's resilience and applicability. It is expected that the proposed model will achieve high accuracy in detecting instances of cyberbullying on Twitter, thereby aiding in addressing and alleviating cyberbullying occurrences.

Ultimately, this research aims to contribute to the ongoing efforts in combating cyberbullying by developing an effective predictive analysis model. By accurately identifying instances of cyberbullying on Twitter, this model will provide valuable insights to help social media platforms, organisations, and individuals take proactive measures to address the issue and create a safer online environment.

V. SYSTEM ARCHITECTURE

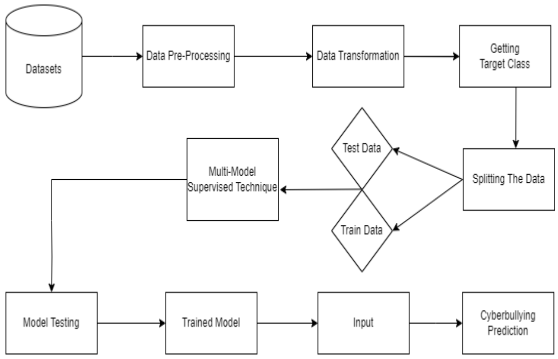

Fig. 1. System Architecture

VI. METHODOLOGY

1. Data Gathering and Preprocessing:

In this module, the first step is to collect Twitter data related to cyberbullying. This can be done using the Twitter API, which allows us to access public tweets. The collected data may include tweets, user profiles, and other related information. Once the data is obtained, preprocessing techniques are used for Data cleansing and

normalization. It includes filtering out irrelevant content like retweets and duplicates, as well as managing noise, spelling discrepancies, and abbreviations. Furthermore, text preprocessing methods like tokenization, stemming, and eliminating stop words are employed to transform the unprocessed text into a suitable format for subsequent analysis.

2. Feature Extraction and Selection:

In this module, features are extracted from the preprocessed data to represent the characteristics of cyberbullying. These features can include linguistic, semantic, and syntactic aspects of the tweets, as well as user-related attributes like follower count and tweet volume. Methods like chi-square test, mutual information, and correlation analysis are utilized for feature selection to identify pertinent and distinguishing attributes. This process diminishes data complexity, enhancing both model efficiency and predictive accuracy.

3. Multi-Model Supervised Techniques:

In this module, multiple supervised machine learning models are employed to predict cyberbullying on Twitter. These models include decision trees, support vector machines (SVM), naive Bayes, random forests, and neural networks. Each model is trained using preprocessed data and selected features from the previous modules. The training dataset is divided into training and validation subsets to assess model performance. Employing ensemble methods like bagging and boosting integrations diverse models, enhancing predictive precision. These finalized models are primed for deployment, categorizing fresh tweets as cyberbullying or non-cyberbullying.

By implementing these three modules, a comprehensive cyberbullying predictive analysis system can be built to identify and combat cyberbullying on Twitter. The system utilizes data gathering and preprocessing techniques, feature extraction and selection methods, as well as multi-model supervised methods to precisely forecast the occurrence of cyberbullying within tweets. Such a system can significantly contribute to promoting safe and respectful online interactions.

VII. RESULT AND DISCUSSION

Table.1. Performance Metrics

Accuracy	Precision	Recall	F1 score
96.8	98.4	96.3	97.7

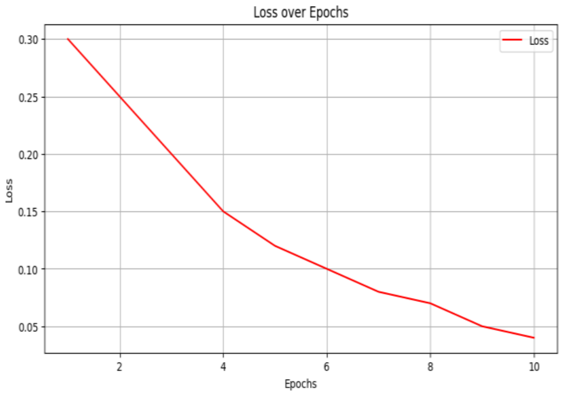

Fig.2. Loss graph

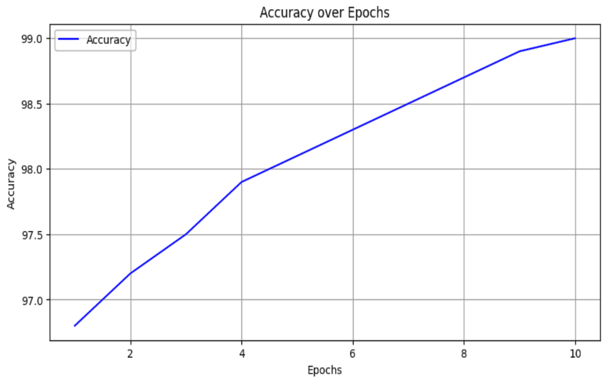

Fig.3. Accuracy graph

The platform that employs multi-model supervised methods for predictive analysis of cyberbullying using Twitter data is designed to detect and prevent instances of cyberbullying on the popular social media platform. Utilizing sophisticated machine learning algorithms, this system scrutinizes extensive Twitter data to promptly detect potential instances of cyberbullying. The multi- model supervised technique employed in this system combines the power of multiple models, enabling more accurate predictions and reducing false positives.

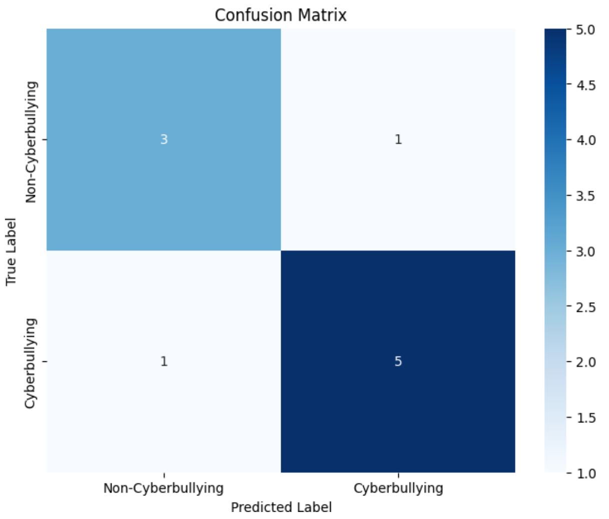

Fig.4. Confusion Matrix

The System begins by collecting a large dataset of tweets and labels them as either cyberbullying or non-cyberbullying. The annotated dataset is subsequently employed to educate diverse machine learning models. Each model is trained to recognize different patterns and features associated with cyberbullying, such as aggressive language, personal attacks, and threats.

During the prediction phase, incoming tweets are analyzed by each model simultaneously. The predictions of each model are then combined, and a final prediction is made, taking into consideration the outputs of all models. This approach ensures that the system makes informed and accurate predictions, while reducing the chances of misclassified tweets.

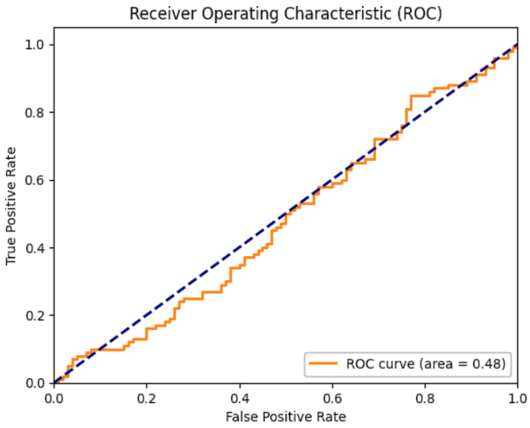

Fig.5. ROC Curve

Once a potential case of cyberbullying is detected, the system can take several actions to prevent further harm. These actions may include automatically flagging the tweet for manual review by a human moderator, blocking or reporting the user responsible for the cyberbullying, or providing users with resources and support to deal with the situation.

In essence, the multi-model supervised approach for predictive analysis of cyberbullying on Twitter presents a potent solution, fostering a safer and more inclusive digital environment. Predictive analysis of online harassment using diverse supervised models on Twitter data is an effective tool in combating cyberbullying, helping to create a safer and more inclusive online environment.

VIII. CONCLUSION

In summary, the multi-model supervised system for predictive analysis of cyberbullying on Twitter is an effective and efficient tool in identifying and predicting instances of cyberbullying on the platform. By leveraging a combination of machine learning algorithms and classifiers, this system is able to analyse large volumes of Twitter data and accurately categorise tweets as either cyberbullying or non-cyberbullying. The use of multiple models allows for enhanced accuracy and robustness, ensuring that instances of cyberbullying are captured and addressed promptly.

This system holds great potential in improving online safety and preventing cyberbullying ultimately creating a safer and more inclusive digital environment for all users.

IX. FUTURE WORK

Future efforts in advancing the system for predictive analysis of cyberbullying on Twitter data utilizing a multi-model supervised methodology. will concentrate on improving detection accuracy and efficiency. Firstly, integrating sophisticated natural language processing methods like word embeddings, sentiment analysis, and named entity recognition can enhance the identification of subtle cyberbullying cases. Secondly, integrating multiple ML algorithms can improve the system's ability to identify different types of cyber bullying behaviour. Thirdly, exploring additional features from user profiles, such as the number of followers, verified status, and historical activities, can provide better context for detecting cyberbullying patterns. Moreover, future endeavors may entail assessing the system's efficacy on expanded and varied datasets to guarantee its scalability and resilience. Finally, considering ethical concerns, such as preserving user privacy and reducing biases in the data, should be an integral part of the future work. Overall, these advancements will contribute to the development of a comprehensive system that can efficiently predict and mitigate cyberbullying on Twitter.

REFERENCES

[1] Murshed,B. A.H., Suresha, Abawajy, J.,Saif,M. A.

[2] N., Abdulwahab, H. M., & Ghanem, F. A. (2023). FAEO- ECNN: cyberbullying detection in social mediaplatforms using topic modelling and deep learning. Multimedia Tools and Applications, 1-40.

[3] Gautam, A. K., & Bansal, A. (2023). Email-Based Cyberstalking Detection On Textual Data Using Multi-Model Soft Voting Technique Of Machine Learning Approach. Journal of Computer Information Systems, 1- 20.

[4] Abhishek, A. (2022). Cyberbullying Detection Using Weakly Supervised And Fully Supervised Learning.

[5] Wang, S., Zhu, X., Ding, W., & Yengejeh, A. A. (2022). Cyberbullying and cyberviolence detection: A triangular user-activity-content view. IEEE/CAA Journal of Automatica Sinica, 9(8), 1384-1405.

[6] Roy, P. K., Singh, A., Tripathy, A. K., & Das, T. K. (2022). Cyberbullying detection: an ensemble learning approach. International Journal of Computational Science and Engineering, 25(3), 315-324.

[7] Giri, S., & Banerjee, S. (2023). Performance analysis of annotation detection techniques for cyber-bullying messages using word-embedded deep neural networks. Social Network Analysis and Mining, 13(1), 23.

[8] Ge, S., Cheng, L., & Liu, H. (2021, April). Improving cyberbullying detection with user interaction. In Proceedings of the Web Conference 2021 (pp. 496-506).

[9] Kumar, A. S., Kumar, N. S., Devi, R. K., & Muthukannan, M. (2024). Analysis of Deep Learning- Based Approaches for Spam Bots and Cyberbullying Detection in Online Social Networks. AI-Centric Modeling and Analytics, 324-361.

[10] Süzen, A. A., & Duman, B. (2021). Detection of types cyber-bullying using fuzzy c-means clustering and xgboost ensemble algorithm. CRJ, (1), 27-34.

[11] Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., & Islam, S. (2023). A Review on Deep-Learning-Based Cyberbullying Detection. Future Internet, 15(5), 179.

[12] Deepa N, Naresh R, Anitha S, Suguna R (2023), “A novel SVMA and K-NN classifier based optical ML technique for seizure detection”, in Optical and Quantum Electronics, Vol.55, PP: 1083 Impact Factor: 3.0. (2023), https://doi.org/10.1007/s11082-023-05406-3

[13] Achyut Shankar, Pandiaraja Perumal, Deepa N, Vaisali R Kulkarni, (2023)“An Intelligent Recommendation System in E-Commerce Using Ensemble Learning”, has been accepted for publication in Multimedia Tools and Applications, https://doi.org/10.1007/s11042-023-17415-1