IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review Of Machine Learning Algorithms For Detection Of Cyberbullying On Social Media Networks

1Ankita V. Rachh, 2Dr. Yagnesh Shukla

1Research Scholar, 2Dean, Faculty of Engineering and Technology

1Atmiya University, Rajkot,

2Atmiya University

Abstract:

Cyberbullying, the use of electronic devices to bully others, has become a prevalent issue globally. Identifying and preventing cyberbullying is crucial to protect individuals from its harmful effects. Cyberbullying is a pervasive and harmful phenomenon, causing significant emotional distress and psychological damage to victims. Cyberbullying has become a major issue in today's society, especially among young individuals who are constantly using social media platforms. Machine learning (ML) algorithms offer a powerful tool for automating cyberbullying detection to mitigate this issue effectively. This paper explores the potential of machine learning algorithms to automatically detect cyberbullying in online platforms. In order to address this problem, this paper proposes a machine learning-based approach for detecting instances of cyberbullying in online platforms. By leveraging the power of machine learning algorithms, we aim to accurately identify and classify cyberbullying behaviour, leading to more effective mitigation strategies and interventions. We discuss various algorithms and their applications in text analysis, focusing on their strengths and weaknesses in identifying cyberbullying content. We also delve into the challenges and ethical considerations associated with employing machine learning for this purpose. We propose a method for detecting cyberbullying using machine learning algorithms. We discuss the challenges associated with accurately identifying cyberbullying behaviour and how machine learning can be leveraged to effectively detect and prevent such behaviour. We present an experimental evaluation of our proposed method and demonstrate its effectiveness in detecting cyberbullying.

Keywords: Cyberbullying, SVM, NB, Random Forest, Machine Learning

I. Introduction:

Cyberbullying is a growing concern in today's digital age, with individuals facing harassment, threats, and other forms of abuse on social media platforms. Traditional methods of detecting cyberbullying rely on manual intervention, which is both time-consuming and prone to errors. Machine learning algorithms offer a more efficient and accurate approach to identifying cyberbullying behaviour.

Cyberbullying has become a prevalent issue in today's digital age, with the rise of social media platforms and online communication channels. To combat this problem, researchers and technologists have turned to

machine learning algorithms as a potential solution for detecting and preventing cyberbullying behaviours. We will delve into the application of machine learning algorithms for cyberbullying detection, the key features and data collection methods used in this process, and the evaluation metrics employed to assess the effectiveness of cyberbullying detection systems.

Cyberbullying, defined as the use of electronic communication to harass, intimidate, or threaten individuals, has become a growing concern in the digital age. The anonymity and reach of online platforms have made it easier for individuals to engage in harmful behaviour, often without facing the consequences of their actions. As a result, cyberbullying has been linked to various negative outcomes, including depression, anxiety, and even suicide.

In order to combat cyberbullying, it is crucial to develop effective detection mechanisms that can identify instances of harmful behaviour in online spaces. Traditional methods of manual monitoring and reporting are often ineffective and time-consuming, requiring significant human resources to review and assess online content. Machine learning algorithms, on the other hand, offer a more efficient and scalable solution to detecting cyberbullying behaviour, by leveraging large datasets to train models that can automatically identify and classify instances of cyberbullying.

The rise of social media and online platforms has provided a fertile ground for cyberbullying. The anonymity and distance offered by the internet empower bullies to engage in harmful behaviours without fear of immediate consequences. The detection and prevention of cyberbullying are critical to foster a safe and inclusive online environment.

Traditional methods of detection, such as user reporting and manual moderation, are often inadequate. Human moderators struggle to keep up with the immense volume of content and may miss subtle forms of bullying. This necessitates the development of automated solutions using machine learning algorithms.

II. Literature Review:

Various studies have explored the use of machine learning algorithms for detecting cyberbullying, with promising results. These studies demonstrate the potential of machine learning algorithms in effectively identifying cyberbullying behaviour.

(1) Towards comprehensive cyber bullying detection: A dataset incorporating aggressive texts, repetitions, peerness and intent to harm, Naveed Ejaz, Fakhra Razi, Salimur Chaudhary (2024) [1]

Text messages are sourced from real dataset and user's data is generated synthetically.

The resulting dataset exchanged messages randomly between users.The intent of harm is quantified as a numeric value using the ratios of repetition and aggregration.

(2) Machine Learning based Intelligent Cyber bullying Avoidance System, D Dhanlaxmi, Deepika Rani (2023)[2]

This system detects the statements by comparing the tweets with a dataset of offensive words. If the tweet contains any bullied word available in the dataset, the tweet is detected as a bully statement otherwise non bullying statement.It uses three algorithms: Naïve Bayes, SVM, Neural Network.

(3) Cyber bullying detection and machine learning: a systematic literature review , Vimala Balakrisnan, Mohammed Kaity (2023)[3]

This review focused on three key aspects, namely, machine learning algorithms used to detect cyber bullying, features and performance measures and further supported with classification roles, language of study, data source and type of media. This review paper includes 68 articles.

(4) Cyber bullying Detection on Social Media using Machine Learning, B Bokolo, Q Liu (2023)[4]
This study compares three machine learning algorithms, Support Vector Machine (SVM), Naive Bayes and a Bidirectional Long Short-Term Memory (Bi-LSTM) on a cyber bullying Twitter dataset. Bi-LSTM model performs the best, achieving 98% accuracy, followed by SVM with 97% accuracy and Naive Bayes with 85%.

(5) A Review of Deep Learning Models for detecting Cyber bullying on Social Media Networks, J Batani, E Mbunge (2022)[5]
This study uses twitter text dataset with long short-term memory(LSTM), bidirectional LSTM, recurrent neural networks and bidirectional gated recurrent unit are predominantly used to detect different forms of cyberbullying such as hate speech, harassment, sexism, bullying among others. The study also revealed that cyberbullying causes psychological effects such as stress, anxiety, depression etc.

(6) Analyzing Machine Learning Techniques for Cyber bullying Detection: A Review Study, J Batani, E Mbunge (2022)[6]
These studies used supervised and unsupervised techniques to identify cyber bullying characteristics by matching text-based information with distinguished characteristics. The paper also summarized the type of features and their combinations used to detect cyber bullying behaviours.

(7) Approaches to Automated Detection of Cyber bullying: A Survey, S Salavu, Y he (2020)[7]
This paper categorize existing approaches into 4 main classes, namely supervised learning, lexicon-based, rule-based and mixed-initiative approaches. Supervised learning use classifiers such as SVM and Naive Bayes to develop predictive models for cyber bullying detection. Lexicon-based utilize word lists and use the presence of words within the lists to detect cyber bullying. Rule-based match text to predefined rules to identify bullying and Mixed-initiatives combine human-based reasoning with combination of approaches.

(8) A Study of Cyber bullying Detection Using Machine Learning Techniques, S Kargutkar, V Chitre (2020)[8]
A system is proposed to give a double characterization of cyber bullying.
This technique utilizes an inventive idea of CNN for content examination anyway the current strategies utilize a guileless way to deal with furnish the arrangement with less precision. A current dataset is utilized for experimentation and system is proposed with other existing methods and is found to give better precision and grouping.

(9) Cyber bullying Detection on Social Networks Using Machine Learning Approaches, M Islam, A Uddin (2020)[9]
The purpose of this research is to design and develop an effective technique to detect online abusive and bullying messages by merging natural language processing and machine learning. Two distinct features, namely Bag- of -Words (BoW) and term frequency-inverse text frequency (TF-IDF) are used to analyse the accuracy level of four distinct machine learning algorithms.

Title	Authors	Algorithms	Strength	Weakness
Towards comprehensive cyber bullying detection: A dataset incorporating aggressive texts, repetitions,	Naveed Ejaz, Fakhra Razi, Salimur Chaudhary	Logistic Regression, Support Vector Machine, Shallow Neural Network, Multinomial Naïve Bayes	Comparison of all traditional Machine Learning methods and check the difference between all classification	Text messages chosen randomly from database of aggressive and non-aggressive messages, so messages are disconnected

peerness and intent to harm				
Machine Learning based Intelligent Cyber bullying Avoidance System	D Dhanlaxmi, Deepika Rani	Naïve Bayes, Support Vector Machine (SVM), Neural Network (NN)	Accuracy of Neural Network is higher than NB and SVM (98%)	Time consuming for data collection and data analysis.
Cyber bullying detection and machine learning: a systematic literature review	Vimala Balakrishnan, Mohammed Kaity	Naïve Bayes, Support Vector Machine(SVM) , Random Forest	Text, audio, Image , Video features comparison of all articles with features, performance measures	Focused on supervised, semi supervised and un supervised only not used any non English articles.
Cyber bullying Detection on Social Media using Machine Learning.	B Bokolo, Q Liu	SVM, Naïve Bayes, Bidirectional long short term memory (Bi-LSTM)	SVM and Bi-LSTM have highest accuracy than Naïve Bayes	Only twitter content is used as dataset, other social media platforms like facebook, youtube, Instagram not used
A Review of Deep Learning Models for detecting Cyber bullying on Social Media Networks	J Batani, E Mbunge	LSTM, Bi-LSTM , Recurrent neural network	Uses Text database of tweeter for review of papers of 2016 to 2021	Not applicable for multimedia data (image, audio, video)
Analyzing Machine Learning Techniques for Cyber bullying Detection: A Review Study	J Batani, E Mbunge	Naïve Bayes, KNN, Support Vector Machine (SVM)	Various feature abstraction techniques used for cyber bullying identification	Online identification of bullying needed
Approaches to Automated Detection of Cyber bullying: A Survey	S Salavu, Y he	SVM, Naïve Bayes	Binary classification of messages and bully victim identification	Applied to use database only, not real data

A Study of Cyber bullying Detection Using Machine Learning Techniques.	S Kargutkar, V Chitre	CNN (Convolution Neural Network)	Tweets are remarked with bully or not	Used very large database for detection
Cyber bullying Detection on Social Networks Using Machine Learning Approaches	M Islam, A Uddin	Decision tree, Random Forest, Naïve Bayes, SVM	BOW(Bag of Words) and TF-IDF (Term frequency inverse text frequency) features are used for identification	Automatic detection and classification not possible

Table 1: Comparison of Cyberbullying Detection Methods

III. Methodology

Machine learning algorithms excel at analysing large data sets and identifying patterns. This makes them well-suited to tackle the challenges of cyberbullying detection.

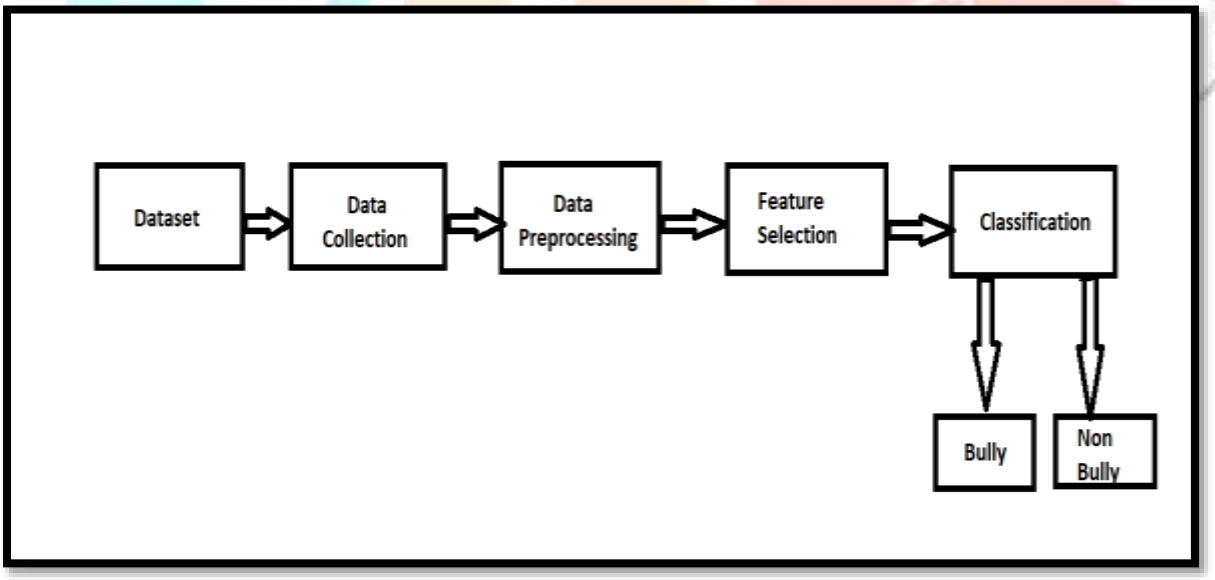

Figure 1: Steps of Cyberbullying Detection

Figure 1 shows the steps of Cyberbullying detection.

a. Dataset
Dataset consist of data from various social media platform like Twitter, Facebook, Instagram, WhatsApp, Youtube etc.

b. Data Collection
Data can be collected from dataset for further processing.

c. Data Preprocessing
Data Preprocessing is process of generating new data after cleaning of data and make it suitable for machine learning model.

d. Feature Selection

Feature selection is a process that chooses a subset of features from the original features so that the feature space is optimally reduced according to a certain criterion.

e. Classification

Classification is a supervised machine learning process that predicts the class of input data based on the algorithms training data.

Machine learning algorithms play a crucial role in the detection of cyberbullying incidents on various online platforms. Two commonly utilized algorithms for this purpose are Support Vector Machines (SVM) and Random Forest. SVM is a supervised learning model that analyses data for classification and regression analysis. In cyberbullying detection, SVM can be trained on labelled data to categorize incoming messages or interactions as either benign or malicious. On the other hand, Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode of the classes as the prediction. This algorithm is favoured for its ability to handle large datasets efficiently and to reduce overfitting in the model.

In cyberbullying detection, the selection of relevant features and effective data collection methods are essential for the accuracy of machine learning models. Key features used in this context include textual features, such as sentiment analysis and the frequency of specific words associated with negative or aggressive language. By analysing the language used in messages or comments, machine learning models can identify potentially harmful content. Additionally, user interaction features, such as the frequency of interactions between users and the connections within a social network, can provide valuable insights into the dynamics of cyberbullying behaviours. Collecting diverse and comprehensive data sets that encompass various online platforms and communication channels is crucial for training robust cyberbullying detection models.

To evaluate the performance of machine learning models in cyberbullying detection, specific metrics are employed to measure their effectiveness. Common evaluation metrics include Accuracy, Precision, Recall, and F1 Score. Accuracy represents the proportion of correctly classified instances, while Precision measures the ratio of correctly predicted positive observations to the total predicted positives. Recall, also known as sensitivity, calculates the proportion of correctly predicted positive instances out of the actual positives. The F1 Score is the harmonic mean of Precision and Recall, providing a balance between the two metrics. Additionally, the Area under the ROC Curve (AUC) is used to assess the model's ability to distinguish between classes and is particularly useful for imbalanced datasets commonly encountered in cyberbullying detection scenarios.

IV. Conclusion:

The application of machine learning algorithms for cyberbullying detection holds significant promise in combating online harassment and promoting a safer digital environment. By leveraging the power of SVM, Random Forest, and other advanced algorithms, along with incorporating key features and utilizing appropriate evaluation metrics, researchers and practitioners can develop effective cyberbullying detection systems. Through continued research and innovation in this field, we can strive towards a more inclusive and secure online community for all users. Machine learning offers promising solutions for detecting cyberbullying in online platforms. By leveraging NLP techniques, supervised learning algorithms, and multimodal analysis, we can develop effective systems to identify and mitigate this harmful behaviour. However, ethical considerations, data bias, and continuous adaptation are crucial in ensuring the responsible and effective implementation of these technologies. Future research should focus on improving accuracy, mitigating bias, and fostering ethical development of machine learning solutions for cyberbullying detection.

V. References:

1. Naveed Ejaz,Fakhra Razi,Salimur Chaudhary (2024) Towards comprehensive cyberbullying detection A dataset incorporating aggressive texts, repetitions, peerness and intent to harm. In Computers in human behaviour. Elsevier
2. D Dhanlaxmi, Deepika Rani (2023) Machine Learning based Intelligent Cyberbullying Avoidance System. In Proceedings of the International Conference on Sustainable Computing and Smart Systems. IEEE Xplore
3. Vimala Balakrisnan, Mohammed Kaity (2023) Cyberbullying detection and machine learning: a systematic literature review. In Artificial Intelligence Review. Springer Nature
4. B Bokolo, Q Liu (2023) Cyber bullying Detection on Social Media using Machine Learning. In Security, Privacy, and Digital Forensics of Mobile Systems and Networks. IEEE Xplore
5. J Batani, E Mbunge (2022) A Review of Deep Learning Models for detecting Cyber bullying on Social Media Networks. In Springer Nature.
6. S Aziz, M Usman (2022) Analyzing Machine Learning Techniques for Cyber bullying Detection: A Review Study . In International Conference on Emerging Technologies. IEEE Xplore
7. S Salavu, Y he (2020) Approaches to Automated Detection of Cyber bullying: A Survey. In IEEE Transactions On Affective Computing. IEEE Xplore
8. S Kargutkar, V Chitre (2020) A Study of Cyber bullying Detection Using Machine Learning Techniques. In Proceedings of the Fourth International Conference on Computing Methodologies and Communication. IEEE Xplore
9. M Islam, A Uddin (2020) Cyber bullying Detection on Social Networks Using Machine Learning Approaches. In IEEE Asia-Pacific Conference on Computer Science and Data Engineering. IEEE Xplore
10. ,Muhammad Arif (2021) , A Systematic Review of Machine Learning Algorithms in Cyberbullying Detection: Future Directions and Challenges. In Journal of Information Security & Cybercrimes Research.
11. Peiling Yia,, Arkaitz Zubiagaa (2022) .Session-based Cyberbullying Detection in Social Media: A Survey. In Elsevier
12. Bandari Saichandana , Dr. Pille Kamakshi (2023), Classification of Cyberbullying Detection in Social Networking with Audio using Machine Learning Approach. In International Journal on Recent and Innovation Trends in Computing and Communication.
13. Dipali Pacharane, Rutuja Pujari, Niam Sandbhor, Sharvari Shinde, Dheeraj Patil, Chandrakant Kokane (2023) , Detection of Cyberbullying Using Machine Learning and Deep Learning Algorithms. In International Journal of Scientific Research in Science and Technology.
14. Mrs. K.Rajeswari , Mushruf Basha M, Praveen S , Ranjith S R ,Sandeap V (2022), Prevention and Suppression of Cyberbullying Using Machine Learning. In International Journal of Research in Engineering and Science.
15. Hadiya E M (2022), Cyber Bullying Detection in Twitter using Machine Learning Algorithms, In International Journal of Advances in Engineering and Management.
16. Bandeh Ali Talpur ID, Declan O'Sullivan (2020), Cyberbullying severity detection: A machine learning approach. In PLOS ONE.
17. Vedadri Yoganand Bharadwaj, Vasamsetti Likhitha, Vootnoori Vardhini , Adari Uma Sree Asritha, Saurabh Dhyani, M. Lakshmi Kanth (2023), Automated Cyberbullying Activity Detection using Machine Learning Algorithm. In Advancement in Image Processing and Pattern Recognition.
18. Nivethitha Rand Dr.L.S.Jayashree (2021), Cyberbullying Detection in Social Networks using Machine Learning Models, In E3S Web of Conferences.
19. Miss. Jafri Sayeedaaliza Abutorab, Miss. Wagh Roshani Balasaheb, Miss. Gaikwad Vaishnavi Subodh, Miss. Sonawane Ujjwala Dattu (2022), Detection of Cyberbullying on Social Media Using Machine Learning. In ICCAP.
20. Venkatesh, M. Abdul Malik, K. Hermus Isitore, S. Sriram (2022), Detection of Cyberbullying on Social Media Using Machine Learning. In International Research Journal of Modernization in Engineering Technology and Science.