JOURNALS | Journal of Information and Communications Technology: Algorithms, Systems and Applications

Volume 1 Issue 1 (June 2025)
DOI: https://doi.org/10.64189/ict.25307
J. Inf. Commun. Technol. Algorithms, Syst. Appl., 2025, 1(1), 25307

Review Article

OPEN ACCESS

A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection

Minal Barhate,* Parikshit N. Mahalle, Vrushal Patil, Rahul Yargop, Shivani Yanpallewar, Tanmay Zade and Vedant Kothari

Department of Artificial Intelligence and Data Science, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India

*Corresponding Author
Minal Barhate
minal.barhate@vit.edu

Received: 20 April 2025
Revised: 10 June 2025
Accepted: 24 June 2025
Published Online: 25 June 2025.

Citation

M. Barhate, P. N. Mahalle, V. Patil, R. Yargop, S. Yanpallewar, T. Zade, V. Kothari, A comparative analysis of machine learning techniques for cyberbullying detection, Journal of Information and Communications Technology: Algorithms, Systems and Applications, 2025, 1(1), 25307, doi: https://doi.org/10.64189/ict.25307.

Abstract

A major worry in the current digital era is cyberbullying, which primarily affects young people who use Web 2.0-powered social media platforms. It usually entails threatening, degrading, or emotionally harming people via online tools and platforms, which frequently results in major psychological consequences like anxiety, sadness, and low self-esteem. This study uses both publicly available Twitter data and data that has been scraped from the platform to examine how machine learning algorithms can be used to detect and categorize instances of cyberbullying. After the text data is cleaned and processed using vectorization techniques, it is analyzed using a variety of supervised learning algorithms. These include Naive Bayes, Random Forest, Support Vector Machines, Logistic Regression, and ensemble models like AdaBoost and Bagging. Each model is assessed using metrics like accuracy, precision, recall, F1-score, and performance efficiency. The study highlights the importance of fine-tuning for practical application by comparing model results. Along with figuring out how to best identify cyberbullying, the goal is to promote the creation of increasingly sophisticated technologies that can help create a safer online environment. This work helps ongoing efforts to use natural language processing and machine learning to alleviate the negative effects of cyberbullying.

Keywords: Cyberbullying; Machine learning; Text classification; Social media; Twitter.

1. Introduction

The rise of digital technologies and social media has changed how people, especially young people, communicate. While platforms like Facebook, Twitter, Instagram, and YouTube help connect people worldwide, they have also become spaces where cyberbullying occurs. Cyberbullying is when someone uses phones, social networks, or messaging apps to repeatedly threaten or harass others online.[1] Serious mental health problems like anxiety, depression, and low self-esteem can result from this type of online abuse. [2-4]

As more people rely on the Internet to interact, it has become both a helpful tool and a potential danger, especially for young users. Unfortunately, traditional methods of spotting and stopping cyberbullying, like manual content review, are not enough to handle the massive amount of content posted every day.[5]

Every year the minimum age in which a child has mobile is decreasing and with the introduction to mobile in lesser age there pose a risk of the child getting exposed to cyberbullying. Introduction to bullying in its development age poses risk of child development in the wrong direction. To prevent this, we need a strong model to prevent bullying via digital services.

In this sudy, we have reviewed to identify cyberbullying through the use of machine learning (ML) and natural language processing (NLP). To determine which supervised learning model is most effective in detecting hazardous texts, we will construct and evaluate several models. Helping to safeguard users and make the internet a safer place is our aim. In addition, we are evaluating several machine learning algorithms to see which one is most appropriate for the task.

As we can clearly see in Fig. 1, every year number cyberbullying victims increases. The data shows victims of cyberbullying in percentage of internet users in age group of 10–18-year-old.

1.1 Literature survey

A major problem these days is cyberbullying, especially with the growth of social media and online messaging services. As more conversations move online, there's a growing need for systems that can automatically detect and address harmful messages. Researchers have approached this problem from several angles, exploring both how to represent text data and how to classify it effectively.[6]

One of the key challenges that we faced is how to turn raw text into something a machine can understand. Techniques similar to the Count Vectorizer and TF-IDF are commonly used to convert words into numerical features.[7,8] While both have their merits, some studies have found that Count Vectorizer can slightly outperform than TF-IDF when it comes to accuracy—especially in informal, mixed-language texts like Hinglish (a blend of Hindi and English often seen in online chats).[9] Even little gains in performance can have a significant impact when handling delicate matters like identifying bullying.[10]

For cyberbullying detection systems to be effective, selecting the right algorithm is very essential.[11] After comparing all the techniques most effective one's are linear support vector (LSV) classifier and stochastic gradient descent (SGD) classifier.[12] These effective techniques are ideal for real world challenges where speed and scalability are crucial because they provide high accuracy along with that they run rapidly. By combining various approaches, some studies have investigated more innovative solutions. Combining fuzzy logic with the multinomial Naïve Bayes classifier is one of the prime examples.[13,14] Naïve Bayes is a straightforward

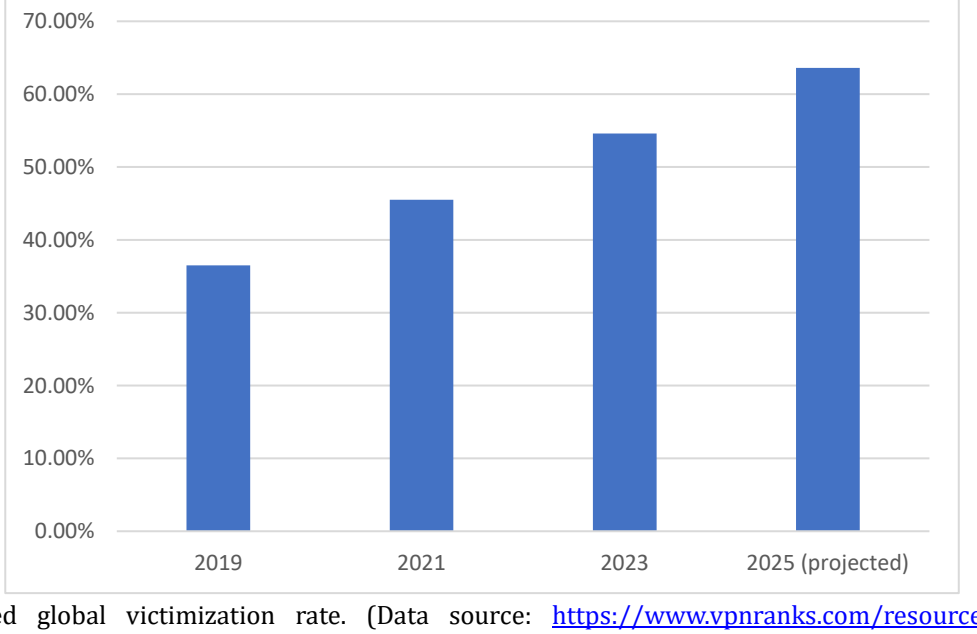

Fig. 1: Estimated global victimization rate. (Data source: https://www.vpnranks.com/resources/cyberbullying-statistics/)

but reliable technique for text classification,[15] while fuzzy logic enhances the system's comprehension of ambiguous and might be emotionally charged content.[16] The hybrid approach where we evaluate a message severity level or emotional concern along with it whether it constitutes bullying which provides deep comprehension of the content.[17]

Efficient SVM and probabilistic models like Naive Bayes had been combined in few of the more advance cyberbullying detection software.[18-20] This strategy of combing both this technique provides important features of both the techniques like SVM's accuracy and Naïve Bayes's speed and simplicity. The final product is a framework that can detect cyberbullying more accurately. Statistics from various surveys shows that higher number of people are cyberbullied as compared to in-person bullying, which is one of the key reasons behind developing such systems. As a result, efforts are made to create such system specially for dynamic environments like chat application and game development too. While using classifiers like Naïve Bayes to identify potentially harmful messages based on previously learned patterns, these tools usually start by preprocessing the input text-removing noise like special characters, duplicates, or irrelevant content.[21,22]

Finally, starting from basic text classification to more complex that considers the subtleties of language, context and user behaviour too. The development of such safe, cyberbullying free environments by deploying more intelligent, accurate AI-driven technologies is highly promising if this field of study is pursued further.

2. Methodology

Creating an efficient detection system has become more crucial as digital platforms continue to expand, given the rise in cyberbullying, which primarily targets younger users. To tackle this, we started by utilizing a substantial dataset of 35,787 tweets that were obtained from Kaggle in CSV format. The text of the tweet was included in each entry, along with a label designating it as either offensive (1) or non-offensive (0). After performing analysis on available dataset, we found out that 23,547 tweets were recognized as offensive along with these 12,239 tweets were recognized as non-offensive.

The raw data from tweet needed to be cleaned and prepared before we move to next steps. To confirm that the dataset was prepared for precise analysis, this preprocessing step was very important. We used NLP technique like NLTK, which allowed us to transform the unstructured data more user-friendly format. This involved breaking up the tweets into smaller, more manageable chunks using tokenization and eliminating frequently used, semantically weak words (also called stop words) that don't add much to the study.

We came to conclusion to use Count Vectorizer to convert the processed text into numerical features that our ML model could learn more easily. This tool was specially designed to deal with the casual and frequently erratic language found on social media. We initiated by cleaning the text, removing stop words and converting all characters to lowercase in order to get the data ready for model training. This made feature extraction more consistent and helped standardize the input. Using PunktSentenceTokenizer, we first divided each tweet into sentences as part of our preprocessing. Whitespace, WordPunct, TreebankWord, and PunktWord tokenizers were the four methods we used to tokenize these sentences into words. After tokenization, converting all text to lowercase was a crucial normalization step that guaranteed consistency and increased the dependability of the features used to train our models.

We used 75% of the dataset for training and the remaining 25% for testing after it had undergone extensive preprocessing. Using this configuration, we trained a range of machine learning models and assessed their performance to find out how well they could identify offensive or non-offensive tweets.

2.1 Machine learning models used

2.1.1 Bagging Classifier (BC)

Bagging is type of ensemble learning method, which is intended to increase the stability and accuracy of models. A number of weaker models are trained on random subsets of dataset, and their predictions are combined-usually by voting in classification tasks or averaging in case of regression tasks. This method reduces overfitting and improves the final model's stability and dependability of cyberbullying text classification. Below mentioned equations are bagging prediction for regression task and classification task.

$$\widehat{y_{\text{bag}}}(x) = \frac{1}{M}\sum_{m=1}^{M} h_m(x) \quad (1)$$

$$H_{\text{bag}}(x) = \text{sign}\left(\sum_{m=1}^{M} h_m(x)\right) \quad (2)$$

2.2.2 Stochastic Gradient Descent (SGD) Classifier

A very effective optimization technique that works particularly well for large-scale learning applications is stochastic gradient descent (SGD). Unlike batch methods that process the entire dataset at once, SGD updates the model parameters one data point at a time. This incremental approach allows for quicker convergence and makes it ideal for handling massive datasets. The update rule,

$$\theta_{t+1} = \theta_t - \eta\, \nabla\theta L(\theta_t; w_i, w_i) \quad (3)$$

effectively handles vast and sparse feature spaces, which are typical in text classification issues like cyberbullying detection, allowing for scalable optimization.

2.2.3 Logistic Regression Classifier (LRC)

One of the fundamental algorithms in statistical learning, logistic regression is frequently applied to challenges involving binary categorization. It learns a decision boundary (also known as a hyperplane) that efficiently divides the two classes and estimates the likelihood of an outcome using the logistic (sigmoid) function.

$$p(y = 1 \mid x) = \frac{1}{1+e^{-z}}, \quad \text{where } z = w^{\top}x + b \tag{4}$$

$$\ln \frac{p(y=1|x)}{p(y=0|x)} = w^{\top}x + b \tag{5}$$

$$L(\theta) = -\sum_{i=1}^{n}[y_i \log p_i + (1 - y_i)\log(1 - p_i)] \tag{6}$$

The logistic function estimates the probability of a text being cyberbullying, enabling binary classification with interpretable outputs.

2.2.4 Decision Tree Classifier (DTC)

Decision trees are easy-to-understand models that classify data by recursively splitting it into branches based on feature values. They are versatile, performing well even with missing data, and can work with both categorical and numerical inputs. They are particularly helpful in the context of text classification because they can reveal intricate decision patterns. For instance, when detecting cyberbullying, the tree can split based on the presence of specific keywords or linguistic cues. Measures like the Gini Index

$$I_G = 1 - \sum_{i=1}^{K} p_i^2 \tag{7}$$

and Entropy

$$H = -\sum_{i=1}^{K} p_i \log_2 p_i \tag{8}$$

guide these splits, helping the model focus on the most informative features. By repeatedly splitting on word-based features, decision trees can effectively learn patterns common in abusive or harmful language.

2.2.5 Linear Support Vector Classifier (LSVC)

Support Vector Machines (SVMs) and Linear Support Vector Classifiers (LSVC) perform very well for linear classification problems. We incorporated a bi-gram tokenizer along with a minimum word frequency threshold of 2, which boosted model's recall by more than 70%. Due to this modification LSVC became the strong contender for precisely detecting cyberbullying. The function which decides whether a text is offensive or non-offensive, works by drawing a optimal margin separation between them is mentioned below:

$$\text{H(x)="sign" (w^⊤ x+b)}, \tag{9}$$

We can achieve better generalization by maximizing the separation between these two classes, which provides more accuracy for classifying abusive content.

2.2.6 Random Forest Classifier (RFC)

Using random samples and features, Random Forest constructs several decision trees, then uses majority voting to aggregate the results. This ensemble method balances bias and variance and is known for its robustness and high accuracy.

$$H_{\text{forest}}(x) = \text{sign}(\textstyle\sum_{m=1}^{M} h_m(x)) \tag{10}$$

Ensemble of decision trees trained on different word subsets increases accuracy and stability in detecting toxic online content.

2.2.7 AdaBoost Classifier (ABC)

AdaBoost, a popular boosting algorithm, works by sequentially training classifiers and adjusting the weights of instances based on their classification accuracy. Misclassified instances are given higher weight in subsequent rounds, making the model focus more on difficult examples.

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{11}$$

$$w_i^{(t+1)} = w_i^{(t)} \cdot \exp\left(-\alpha_t y_i h_t(x_i)\right) \tag{12}$$

$$H_{\text{AdaBoost}}(x) = \text{sign}(\textstyle\sum_{t=1}^{T} \alpha_t h_t(x)) \tag{13}$$

Adaptive weighting of misclassified cyberbullying texts enhances model focus on challenging abusive examples.

2.2.8 Multinomial Naïve Bayes Classifier

Jobs requiring the classification of texts and documents commonly employ MNBC. It assumes feature independence and calculates the probability of a class given the words in the text. Its simplicity and speed make it especially popular in cyberbullying detection tasks.

$$P(c \mid x) \propto P(c)\prod_{i=1}^{n} P(x_i \mid c) \tag{14}$$

$$\log P(c \mid x) = \log P(c) + \textstyle\sum_{i=1}^{n} x_i \log P(i \mid c) + \text{const} \tag{15}$$

Word frequency probabilities enable fast and effective identification of harmful language using the Naïve Bayes assumption.

2.3 Evaluation metrics

Metrics for Evaluation We employed Python-based machine learning tools to test each model's performance, and the results were analyzed using a confusion matrix, which provides the following information:

- True Negatives (TN): Non-offensive tweets were correctly identified.
- False Positives (FP): Inaccurately labeling non-offensive tweets as offensive.
- False Negatives (FN): Derogatory tweets that the model failed to detect.
- True Positives (TP): recognized abusive tweets with

accuracy.

We obtained a number of important performance measures from these values:

Accuracy shows the proportion of tweets that were deemed offensive that were truly offensive:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (16)$$

Precision calculates the percentage of tweets that were successfully identified out of all the samples:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (17)$$

Recall Reflects how well the model captured all actual offensive tweets:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (18)$$

F1-score provides a single statistic that considers both precision as well as recall in a balanced way:

$$\text{F1-score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}) \quad (19)$$

3. Results and discussions

For identifying and classifying instances of cyberbullying in online chatting, this study applies a number of ML algorithm. Various preprocessing techniques were applied on dataset like tokenization, stop word removal and stemming to increase the model performance. After this, we trained and evaluated a variety of classifiers to see how well they identify harmful content.

The SGD Classifier performed the best overall out of all the models that were tested, earning high ratings for accuracy, precision, recall, and F1-score. The SGD model consistently performed better than other algorithms, as shown in the confusion matrix (Fig. 2). It worked especially well for categorizing content about bullying across delicate categories like gender, sexual orientation, and religion. The confusion matrix provided additional evidence of the model's ability to lower false positives and false negatives. In contrast, the Naive Bayes classifier did worse, especially when it came to multi-class classification tasks.

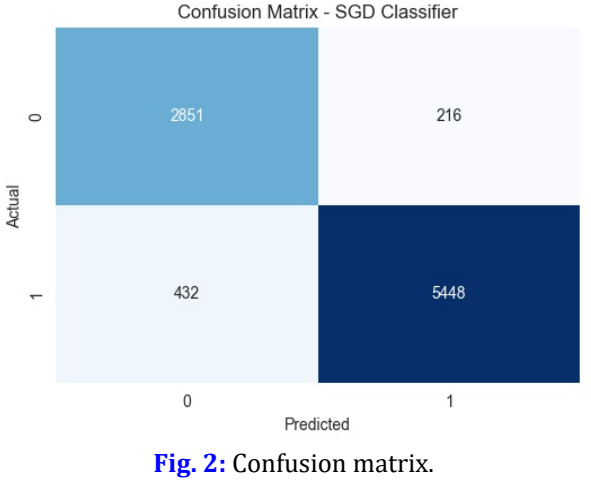

Fig. 2: Confusion matrix.

The findings suggest that text-based feature engineering, combined with robust classification models, plays a crucial role in detecting cyberbullying. The results highlight the importance of selecting suitable algorithms and tuning hyperparameters to improve detection accuracy. These insights can guide the future development of real-time cyberbullying detection systems on digital platforms.

Our other aim was to compare multiple algorithms on various parameters. Fig. 3 compares algorithms based on accuracy, recall, precision, and f1 score. As we can see all algorithms were par with everyone but adaboost classifier lacked precision so it was very much unsuitable for the task. Fig. 4 shows time complexity of algorithms and we can see BG and RFC were very much time intensive algorithms. Hence these three were already out of considerations. The detailed scores of SGD Classifier are as follows:

Accuracy: 92.81%

Precision: 96.97%

Recall: 91.94%

F1-Score: 94.39%

After this we condensed a sav file on this model and stores the vectorizer data in pickle file. These model and vector files were used to predict any sentence as offensive or non-offensive. We then build a chatting webapp and deployed this model and using rest api we were able to replicate cyberbullying detection in real time chatting.

3.1 Issues and challenges

Through the course of study, we faced many challenges but two of them were notable. The first and foremost issue we faced while conducting the study was data collection, data cleaning, and labelling. As we worked on more than 35000 sentences for out model, labelling such huge data was a tedious task. We took help from automated AI bots like ChatGPT but human intervention was necessary. Many times, the LLM model was hallucinating by giving false positives, due to lack of understanding of conversational slang. So manually verifying all the single line tweets took a generous amount of time. After this we embedded the data for better model feeding. Any model works better with numbers rather than text. Embedding the text opened up various things we can do like clustering the data and figuring out which cluster have most no of cyberbullying violations and tracing it to actual text.

The second major challenge was while proposing a practical application, we need to take in consideration of user privacy for detecting cyberbullying in chats or messages. To tackle this, we trained our model on one liner data so that while detecting cyberbullying in real life scenario, the model will not be requiring complete chatting context to detect cyberbullying. Instead, it can

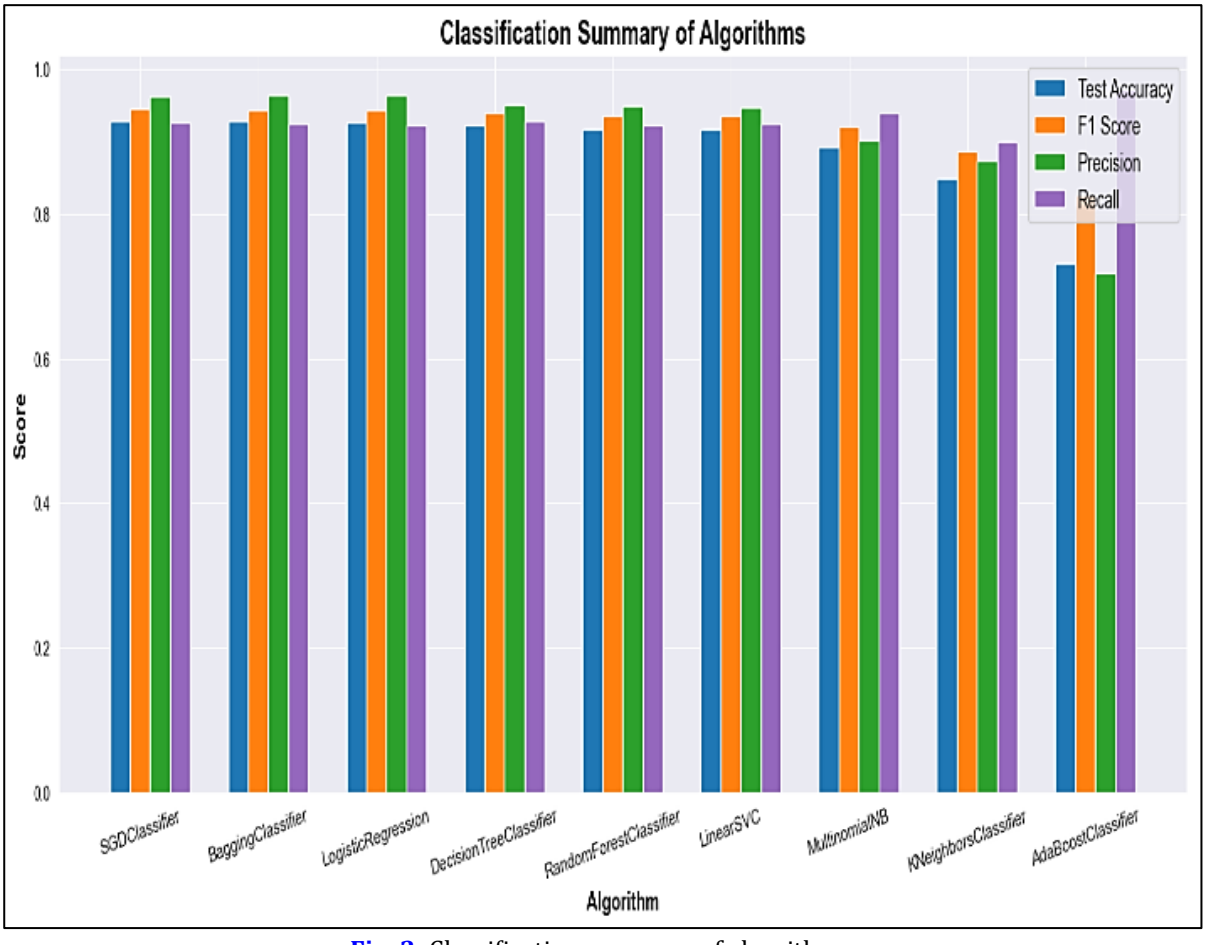

Fig. 3: Classification summary of algorithms.

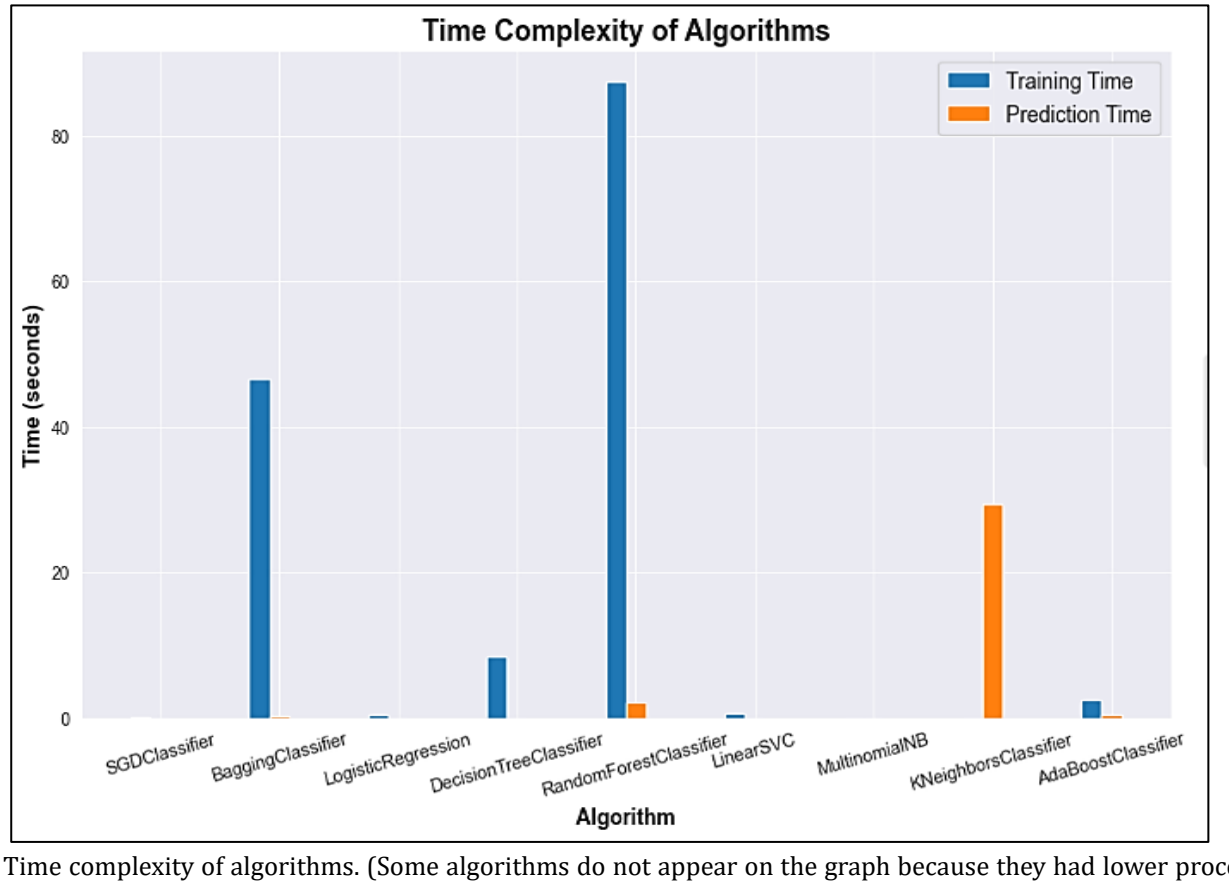

Fig. 4: Time complexity of algorithms. (Some algorithms do not appear on the graph because they had lower processing times)

look up on single message itself. We even tried working on cyberbullying detection on encrypted data but didn't achieve any conclusion because of lack of data and time. Using encrypted data for detection further enhances its purpose while keeping real life scenarios in mind.

3.2 Future scope

The fight against cyberbullying is ongoing, and this study opens several paths for future improvement. One promising direction is using deep learning models like LSTM or BERT, which understand the meaning and context of words better than traditional machine learning methods.

Utilizing information from many social media sites, like Facebook, Instagram, and Reddit, would be an additional enhancement. This would help the model handle different ways people communicate across platforms and make it more accurate for different age groups and communities.

Adding real-time detection through streaming APIs can allow the system to act quickly, even before serious harm occurs. Support for multiple languages is also important, since cyberbullying happens in many languages, not just English.

Linking the detection system with mental health support services or automated chatbots could help victims get help immediately. Finally, creating user-friendly tools—like browser extensions or mobile apps-would make the system easy to use and more accessible to everyone, helping to create a safer online space.

This proposed system could be furthermore developed in fashion where we don't even need to read textual content. We can train models on text embedding and convert the text data in embeddings. These embeddings could work like encryption, where the model and our system can't read actual message but still can process the cyberbullying detection function. These models could be deployed in a cloud database and using Rest Api integrated with major online platforms like twitter, Instagram *etc*. LLMs could also be used, e.g. in every textbox of online platform, in real time the LLM can process and warn the user about potential cyberbullying he/she may be performing.

4. Conclusion

This study shows that machine learning classifiers are effective tools for detecting cyberbullying. By analyzing both the content and context of online messages, these models can accurately identify harmful behavior. After testing several machine learning models on our dataset, we discovered that the Stochastic Gradient Descent Classifier (SGDC) worked best. It showed good accuracy and reliability, especially when tested using k-fold cross-validation, which ensures the model performs well over numerous data samples. Overall, our system has shown a strong ability to detect cyberbullying and offers a solid starting point for further development. With future enhancements like deep learning, multilingual support, real-time detection, and integration with support tools, this work lays the foundation for building safer, more respectful online communities.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The datasets generated and/or analyzed during the current study that support the findings are available from the corresponding author upon reasonable request.

Conflict of Interest

There is no conflict of interest.

Artificial Intelligence (AI) Use Disclosure

The authors declare that artificial intelligence (AI)-assisted tools were used only for language refinement, grammar improvement, and manuscript structuring purposes during the preparation of this work. All technical content, experimental implementation, results, and interpretations were independently developed and verified by the authors.

Supporting Information

Not applicable

References

[1] D. Marín Suelves, A. Rodríguez Guimeráns, M. Mercedes Romero Rodrigo, S. López Gómez, Cyberbullying: education research, *Education Sciences*, 2023, **13**, 763, doi: 10.3390/educsci13080763.

[2] S. I. Ali, N. B. Shahbuddin, The relationship between cyberbullying and mental health among university students, *Sustainability*, 2022, **14**, 6881, doi: 10.3390/su14116881.

[3] Z. F. Albikawi, Anxiety, Depression, Self-esteem, internet addiction and predictors of cyberbullying and cybervictimization among female nursing university students: a cross-sectional study, *International Journal of Environmental Research and Public Health*, 2023, **20**, 4293, doi: 10.3390/ijerph20054293.

[4] C. Li, P. Wang, M. Martin-Moratinos, M. Bella-Fernández, H. Blasco-Fontecilla, Traditional bullying and cyberbullying in the digital age and its associated mental health problems in children and adolescents: a meta-analysis, *European Child*

& Adolescent Psychiatry, 2024, **33**, 2895–2909, doi: 10.1007/s00787-022-02128-x.

[5] M. Patil, R. Kawatagi, R. Shreya, S. Zakkam, V. Kadapure, The study of methods for detection and prevention of cyberbullying, *International Journal of Research Publication and Reviews*, 2023, **4**, 929-935.

[6] A. Mangaonkar, A. Hayrapetian, R. Raje, Collaborative detection of cyberbullying behavior in Twitter data, 2015 IEEE International Conference on Electro/Information Technology (EIT), Dekalb, IL, USA, 2015, 611-616, doi: 10.1109/EIT.2015.7293405.

[7] A. Mangaonkar, R. Pawar, N. S. Chowdhury, R. R. Raje, Enhancing collaborative detection of cyberbullying behavior in Twitter data, *Cluster Computing*, 2022, **25**, 1263-1277, doi: 10.1007/s10586-021-03483-1.

[8] J. S. Abutorab, R. B. Wagh, V. S. Gaikwad U. D. Sonawane, A. I. Waghmare, Detection of cyberbullying on social media using machine learning, *International Research Journal of Modernization in Engineering Technology and Science*, 2022, 04, 4526-4534.

[9] K. Shah, C. Phadtare, K. Rajpara, Cyber-Bullying Detection in Hinglish languages using machine learning. International Journal of Engineering and Technical Research, 2012, **11**, 439, doi: 10.17577/IJERTV11IS050318.

[10] R. Kumar, S. Parakh, C. N. S. Vinoth Kumar, Cyberbullying detection system using machine learning, *Turkish Journal of Computer and Mathematics Education*, 2021, **1**, 656-661.

[11] Sahana V., Anil Kumar K. M., A. A. Darem, A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on FormSpring in Textual Modality, *International Journal Of Computer Network And Information Securit*y, 2023, **4**, 36-47, doi: 10.5815/ijcnis.2023.04.04.

[12] A. Muneer, S. M. Fati, A Comparative analysis of machine learning techniques for cyberbullying detection on twitter, *Future Internet*, 2020, **12**, 187, doi: 10.3390/fi12110187.

[13] A. Akhter, U. K. Acharjee, M. Masbaul A. Polash, Cyber bullying detection and classification using multinomial naïve bayes and fuzzy logic, *International Journal of Mathematical Sciences and Computing*, 2019, **5**, 1-12, doi: 10.5815/ijmsc.2019.04.01.

[14] L. Ketsbaia, B. Issac, X. Chen, S. M. Jacob, A multi-stage machine learning and fuzzy approach to cyber-hate detection, *IEEE Access*, 2023, **11**, 56046-56065, doi: 10.1109/ACCESS.2023.3282834.

[15] J. Chen, H. Huang, S. Tian, Y. Qu, Feature selection for text classification with Naïve Bayes, *Expert Systems with Applications*, 2009, **36**, 5432-5435, doi: 10.1016/j.eswa.2008.06.054.

[16] B. Sri Nandhini, J. I. Sheeba, Online social network bullying detection using intelligence techniques, *Procedia Computer Science*, 2015, **45**, 485-492, doi: 10.1016/j.procs.2015.03.085.

[17] M. Patidar, M. Lathi, M. Dhakad, Y. Barge, Cyber bullying detection for twitter using ml classification and algorithms, *International Journal for Research in Applied Science & Engineering Technology*, 2021, **9**, 24-29, doi: 10.22214/ijraset.2021.38701.

[18] H. Saini, H. Mehra, R. Rani, Garima Jaiswal, A. Sharma, A. Dev, Enhancing cyberbullying detection: a comparative study of ensemble CNN–SVM and BERT models, *Social Network Analysis and Mining*, 2024, **14**, doi: 10.1007/s13278-023-01158-w.

[19] R. R. Dalvi, S. B. Chavan, A. Halbe, Detecting a twitter cyberbullying using machine learning, 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, 297-301, doi: 10.1109/ICICCS48265.2020.9120893.

[20] V. S. Chavan, S. S. Shylaja, Machine learning approach for detection of cyber-aggressive comments by peers on social media network, 2015 International Conference on Advances in Computing, Communications and Informatics, 2015.

[21] K. L. Tan, C. P. Lee, K. M. Lim, A survey of sentiment analysis: Approaches, datasets, and future research, *Applied Sciences*, 2023, **13**, doi: 10.3390/app13074550.

[22] K. K. Kiilu, G. Okeyo, R. Rimiru, K. Ogada, Using Naïve Bayes Algorithm in detection of hate Tweets, *International Journal of Scientific and Research Publications*, 2018, **8**, 99-107, doi: 10.29322/IJSRP.8.3. 2018.p7517.