Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 3 (2025)

Optimizing Cyberbullying Detection in Tweets with an Ensemble Learning Framework

Manjeet Singh[1*], Dr. Annapurna Metta[2], Dr. Prof. Satyendra Patnaik[3]

[1* a,] Research scholar Amity University Chhattisgarh, Raipur.
[1* b.] Manager software engineering, ClearTrail Technologies Pvt Ltd. SDF No. K-12, NSEZ, Noida, UP 201305 https://orcid.org/0009-0000-2157-1388
[2]Assistant Professor, Amity Business School, Amity University Chhattisgarh https://orcid.org/0009-0009-2015-6815
[3]Professor & Dean, JSS University Noida: Noida, Uttar Pradesh, INDIA, https://orcid.org/0000-0002-1296-3387

Abstract: The fast spread of social media has resulted in a considerable increase in cyberbullying events, which can have serious psychological consequences for individuals. Detecting and reducing cyberbullying has become a critical job for creating a safer online environment. Traditional approaches for detecting cyberbullying frequently rely on manual moderation, which is inefficient considering the volume of content published daily. As a result, there is an urgent need for automated techniques to properly classifying and managing cyberbullying content. The study focuses on classifying cyberbullying types in tweets using a variety of machine-learning models and an improved stacking ensemble model. After preprocessing the dataset, the text data is cleaned by eliminating URLs, mentions, hashtags, and stop words. The preprocessed text input is then transformed into TF-IDF features to train the models. The tests used a variety of classifiers, including Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, k-Nearest Neighbors (k-NN), and Gradient Boosting. Each model is trained and assessed to compare its performance in terms of accuracy, precision, recall, and F1 score. To improve classification performance even more, we propose a stacking ensemble model that combines the predictions of these base classifiers with a logistic regression meta-classifier. The ensemble model's hyperparameters are optimized using grid search. The ensemble model performs best, with a classification accuracy of 90.18%.

1. Introduction:

The tremendous growth of social media platforms has resulted in a significant rise in data created by users, enabling unparalleled levels of communication and involvement. Nevertheless, the increase in digital communication has also led to the emergence of cyberbullying, a widespread and detrimental practice that can cause significant psychological injury to individuals, especially teenagers and young adults [1]. Cyberbullying refers to intentional and aggressive actions, conducted by either an individual or a group, through electronic means of communication. These actions are repeated over a period and target a victim who lacks the ability to quickly protect themselves [2]. Ensuring the safety and well-being of online settings has become increasingly important due to the need to identify and address instances of cyberbullying. Conventional methods for identifying cyberbullying typically depend on manual moderation, which is both time-consuming and unfeasible due to the large amount of content produced on platforms such as Twitter, Facebook, and Instagram. The speed at which content is generated beyond the ability of manual moderation, thus requiring the creation of automated systems that can promptly and efficiently detect and handle cyberbullying content in real-time [3].

Recent breakthroughs in the field of machine learning have demonstrated potential in tackling the intricacies of identifying and resolving cyberbullying. Different machine learning methods have been used to categorize cyberbullying, each with their own advantages and disadvantages. Naive Bayes classifiers are renowned for their simplicity and efficacy in text classification tasks, whereas Support Vector Machines (SVM) and Decision Trees provide more advanced methods that can capture

intricate patterns in data. Ensemble approaches, such as Random Forests and Gradient Boosting [4], improve classification performance by aggregating the predictions of numerous base models.

The objective of this work is to categorize different types of cyberbullying in tweets by utilizing a wide range of machine learning models, including as Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, k-Nearest Neighbors (k-NN), and Gradient Boosting. In order to improve the accuracy of classification, we suggest using a stacking ensemble model that combines the capabilities of multiple separate classifiers. The stacking ensemble model employs a logistic regression meta-classifier to merge the predictions of the basic models, enhancing its performance by adjusting hyperparameters through grid search. The methodology initiates with the preprocessing of tweet data, encompassing the elimination of URLs, mentions, hashtags, and stop words to cleanse the text. Subsequently, the sanitized textual data is converted into TF-IDF features, which are utilized as input for training the models. Every classifier undergoes training and evaluation using metrics such as accuracy, precision, recall, and F1 score. The results suggest that the stacking ensemble model surpasses the performance of the individual classifiers, obtaining a classification accuracy of 90.18%.

The subsequent sections of this work are organized as follows: Section 2 explores the existing research in cyberbullying detection. Section 3 provides an overview of the dataset and the proposed ensemble model. Section 4 presents the empirical findings and assessment. Section 6 serves as the final section of the report and provides suggestions for further research.

2. Literature Review

Various studies have investigated the application of machine learning in the identification of cyberbullying. [5] and [6] demonstrated excellent precision in detecting cyberbullying communications. The authors in [5] utilized a Naive Bayes classifier, whereas [6] employed a voting classifier that relied on machine learning algorithms. [7] and [8] employed machine learning methods such as Multinomial Naïve Bayes, LinearSVC, Logistic Regression, K-Nearest Neighbour, and Random Forest to identify and categorize instances of cyberbullying in texts. These studies jointly show the capacity of machine learning to accurately detect and prevent cyberbullying. A range of studies have explored the use of Random Forest in classifying cyberbullying. [9] and [10] both found that Random Forest achieved high precision and recall in identifying cyberbullying on Twitter, with [10] reporting an F1-Score of 0.90. [11] further emphasized the effectiveness of Random Forest in accurately classifying cyberbullying tweets, while [12] compared Random Forest with Extreme Gradient Boosting and found that both methods performed similarly, with Random Forest achieving an accuracy of 0.849. These studies collectively highlight the potential of Random Forest as a reliable classifier for cyberbullying detection. [13] introduced an ensemble stacking model that integrates multiple supervised machine learning methods, such as Decision Trees, Random Forest, Linear SVC, Logistic Regression, and K-Nearest Neighbors. They employed feature extraction techniques such as Bag of Words (BoW), TF-IDF, Word2Vec, and GloVe. Their model, when tested on the Cyber-Troll dataset, attained an accuracy rate of 94%, which showcases the efficacy of ensemble approaches in enhancing classification performance.

[14] created an advanced deep learning model that combines BiGRU, transformer block, and CNN architectures to accurately categorize tweets as either aggressive or non-aggressive. Their model, evaluated on a merged dataset of 55,788 tweets, attained an accuracy of around 88%, indicating that the integration of several deep learning architectures can improve the identification of cyberbullying.

[15] employed pre-trained GloVe embeddings in conjunction with a Bi-directional Long Short-Term Memory (BLSTM) model for the purpose of identifying instances of cyberbullying in tweets. Their methodology attained a precision of 92.60%, showcasing the capability of deep learning models in extracting contextual information from textual input. Talpur and O'Sullivan (2020) aimed to tackle the issue of imbalanced classes in text categorization. They achieved this by using features such pointwise semantic orientation and projected user attributes such as gender, age, and personality type. The researchers utilized the Random Forest algorithm along with SMOTE and cost-adjusted classification techniques to identify the severity of cyberbullying. They achieved an accuracy rate of

93% on a dataset consisting of 11,904 tweets that were manually tagged. In their study, Daniel et al. (2023) introduced a model that combines an ensemble of LSTM-Adaboost with the Tournament Selected Glowworm Swarm Optimization (TSGSO) algorithm for optimization. By employing the GloVe word embedding technique, their model exhibited encouraging efficacy in identifying instances of cyberbullying in Twitter data.

In their study, Kumar et al. (2024) investigated the application of ensemble methods, such as Random Forests and Gradient Boosting, in conjunction with text analysis and natural language processing (NLP). Their model, assessed using a Kaggle dataset consisting of 47,692 records, successfully differentiated between benign communication and potential instances of cyberbullying, showcasing the resilience of ensemble approaches. Raisi and Huang (2018) developed a method of minimum supervision by using expert-provided key phrases. They utilized an ensemble of two co-trained learners to accurately detect instances of cyberbullying. Their profound ensemble approach surpassed prior non-deep techniques in detecting cyberbullying with minimal supervision using data from Twitter, Ask.fm, and Instagram. In their study, Muneer et al. (2023) proposed a stacking ensemble learning approach that integrates many deep neural network (DNN) models with a modified BERT model. Their methodology yielded a 97.4% accuracy rate when applied to a Twitter dataset, showcasing the efficacy of integrating sophisticated NLP models for the purpose of identifying instances of cyberbullying. Venu et al. (2023) combined information from Twitter and Wikipedia by employing feature extraction techniques such as Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF). They assessed the performance of Support Vector Machine (SVM), Logistic Regressor, and Naive Bayes classifiers. Their Support Vector Machine (SVM) model attained the utmost accuracy, highlighting the significance of resilient feature extraction strategies in augmenting model performance.

The latest contributions in the realm of classification of cyberbullying are summarized in Table 1.

Table 1: Summary of Studies on Cyberbullying Detection

Paper/Year	Methodology	Dataset	Main Findings
[13]/2024	Ensemble stacking model with supervised ML models and four feature extraction methods (BoW, TF-IDF, Word2Vec, GloVe).	Cyber-Troll dataset with 20,001 tweets	Acc =94% Prec = 96% Recall =95% F1-Score = 95%
[14]/2021	Multichannel deep learning model integrating BiGRU, transformer block, and CNN.	Combined dataset of 55,788 tweets	Acc=88%
[16]/2020	Logistic regression classifier on a Twitter dataset.	Twitter dataset with equal bully and non-bully tweets	Prec = 91% Recall = 94% F1-score = 93%.
[17]/2020	Random Forest with SMOTE, cost-adjusted classification, and PMI-based semantic orientation.	11,904 manually labeled tweets	Acc = 93% F1-Measure = 92% Kappa score = 84%
[18]/2023	Ensemble LSTM-Adaboost model with GloVe and TSGSO for hyperparameter optimization.	Twitter dataset	EDL-TSGSO technique showed promising performance, using GloVe and ensemble LSTM-Adaboost model.
[19]/2024	Ensemble techniques (Random Forests, Gradient Boosting) with text analysis and NLP.	Kaggle dataset with 47,692 records	Random Forest and Gradient Boosting minimized overfitting risks, ensuring robust model performance.
[8]/2024	Used six ML algorithms; evaluated performance using accuracy, F1-score, cross-validation score, and ROC curve.	Twitter dataset	Random Forest outperformed others with 94% accuracy, effective for detecting and preventing cyberbullying on Twitter.

[20]/2023	Data integration, feature extraction using BoW and TF-IDF, evaluated SVM, Logistic Regressor, and Naive Bayes.	Tweets and Wikipedia comments	SVM model achieved Acc =98.73% (Twitter data) Acc=95.46% (Wikipedia data.)
[21]/2018	Minimal supervision with expert-provided key phrases; co-trained ensemble learners for language and social structure.	Twitter, Ask.fm, and Instagram data	Deep ensemble approach outperformed previous non-deep methods for weakly supervised cyberbullying detection.
[22]/2022	SVM, DistilBERT, and stacked ensemble models; evaluated different levels of TF-IDF feature extraction.	Various datasets including Twitter, Instagram, Ask.fm, Wikipedia, and YouTube	Ensemble models outperformed individual models. DistilBERT showed best precision= 91.17% among all models.
[23]/2021	Nine ML algorithms and an ensemble model combining Linear SVC, Multinomial NB, and Logistic Regression.	Tweets from various incidents and events (350,000 tweets)	Random Forest was the best performer. Ensemble model combining multiple classifiers outperformed constituent classifiers.
[15]/2022	Deep learning with GloVe embeddings; Bi-directional LSTM (BLSTM) model for classification.	35,787 labeled tweets	GloVe840 with BLSTM achieved 92.60% accuracy, 96.60% precision, and 94.20% F1-score.
[24]/2018	Representing tweets as word vectors; deep learning for classification with metaheuristic optimization.	Not mentioned	Novel deep learning-based approach (OCDD) proposed; uses word vectors and metaheuristic optimization for parameter tuning.
[25]/2021	Utilized Twitter profile metadata with unique feature selection and ensemble learning to create a classifier.	Twitter profile metadata	Ensemble learning model combining Naive Bayes and Decision Tree detected fake Twitter accounts with 98.93% accuracy.
[26]/2020	Preprocessing steps including removal of numbers, lowercasing, replacing URLs/user mentions, and removing special characters.	University of Maryland dataset and a dataset on offensive language (25,000 tweets)	Linear SVC had the highest accuracy. Combination of DBOW and DMM Doc2Vec models achieved 96.39% accuracy. CNNs improved accuracy to 98.20%.
[27]/2023	Stacking ensemble model combining multiple deep neural networks; used CBOW feature extractor and modified BERT.	Twitter dataset (37,373 tweets) and a combined Twitter/Facebook dataset (20,000 rows)	Stacked ensemble model achieved 97.4% accuracy on Twitter dataset and 90.97% on combined dataset. Very fast detection time.
[28]/2024	Preprocessing, data splitting, and training multiple ML models (Logistic Regression, Random Forest, Decision Tree).	Twitter dataset	Ensemble models RF+DT and SVM+LR achieved 92% and 93% accuracy, respectively. SVM and Logistic Regression models also achieved 93% accuracy.
[29]/2023	Used NLP and ML; trained on cyberbullying tweets dataset; compared ML algorithms with Random Forest performing best.	Not mentioned	Random Forest provided the best results for cyberbullying detection on Twitter after tuning.
[30]/2022	CNN model with text and meta-information inputs; VGG-19 Net model for image inputs; used Tensor Fusion technique.	Twitter dataset (100,000 tweets) and additional dataset (974,053 tweets)	Using text, meta-information, and image data together improved detection accuracy. Posts with image attachments were easier to classify.
[31]/2021	Used four ML classifiers and three ensemble models; combined TF-IDF and n-gram analysis on Twitter dataset.	Twitter dataset	Proposed SLE and DLE models outperformed individual classifiers, achieving 96% accuracy with TF-IDF and K-Fold cross-validation.

3. Methodology

The widespread adoption of social media platforms has enabled the dissemination of cyberbullying, which may cause significant psychological and emotional harm on its targets. Due to the large volume of content generated on a daily basis, traditional manual moderation methods are not capable of

handling the scale, which is why automated techniques are necessary to efficiently detect and handle cyberbullying content. Machine learning provides effective solutions by enabling the automated classification of text data through the recognition of learned patterns. Different algorithms, including Naive Bayes, Support Vector Machines (SVM), Decision Trees, Random Forests, k-Nearest Neighbors (k-NN), and Gradient Boosting, have been used to detect cyberbullying. Each approach has its own advantages and disadvantages. Although there has been progress, single models frequently face difficulties in achieving optimal performance across a wide range of datasets. Ensemble approaches, which aggregate the predictions of numerous models, improve forecast accuracy and resilience. Stacking ensembles employ a meta-classifier to combine the predictions of base classifiers, hence enhancing the overall classification performance.
This study employs a combination of individual machine learning models and an upgraded stacking ensemble model to accurately classify different types of cyberbullying in tweets, resulting in superior classification performance.

3.1 Dataset

The dataset utilized in this study comprises tweets pertaining to cyberbullying, obtained from openly accessible databases on social media interactions. The information encompasses many categories of cyberbullying, including age, ethnicity, gender, religion, and other types of harassment. The tweets are categorized according to their specific type of cyberbullying, creating a comprehensive dataset for training and assessing machine learning models. The dataset was partitioned in advance into separate training and test sets, allocating 80% of the data for training the models and reserving 20% for testing and evaluation purposes. The training dataset comprises 16,000 tweets, and the test dataset encompasses 4,000 tweets.

3.2 Pre-Processing

Pre-processing is an essential step in preparing text data for analysis and training models. The dataset underwent various pre-processing procedures to ensure the quality and relevance of the input data. At first, data cleaning was carried out to exclude URLs, user mentions (such as @username), and hashtags, unless the hashtags included significant phrases that were relevant to the content. This was done to get rid of irrelevant information and concentrate on the important aspects of the tweets. Subsequently, the NLTK library was utilized to exclude frequently occurring stop words such as "and," "the," and "is" in order to decrease dimensionality and improve the performance of the model. Subsequently, the text data was subjected to tokenization, which involved dividing it into separate words to enable subsequent processing. Ultimately, the tokenized text was converted into Term Frequency-Inverse Document Frequency (TF-IDF) features, which assigned numerical values to each word based on its significance within the overall dataset. The thorough pre-processing guaranteed that the data was properly prepared for subsequent machine learning tasks.

3.3 Exploratory Data Analysis

An EDA, or exploratory data analysis was performed to get insights into the dataset and to understand the distribution and attributes of the data. The essential stages in the exploratory data analysis (EDA) process comprised:

Class Distribution: An analysis was conducted to examine the prevalence of each type of cyberbullying in the dataset by studying the distribution of different cyberbullying categories. A bar plot was generated to visually represent the distribution of classes, with a focus on identifying any disparities in the data. The bar plate is shown in the figure 3.1.

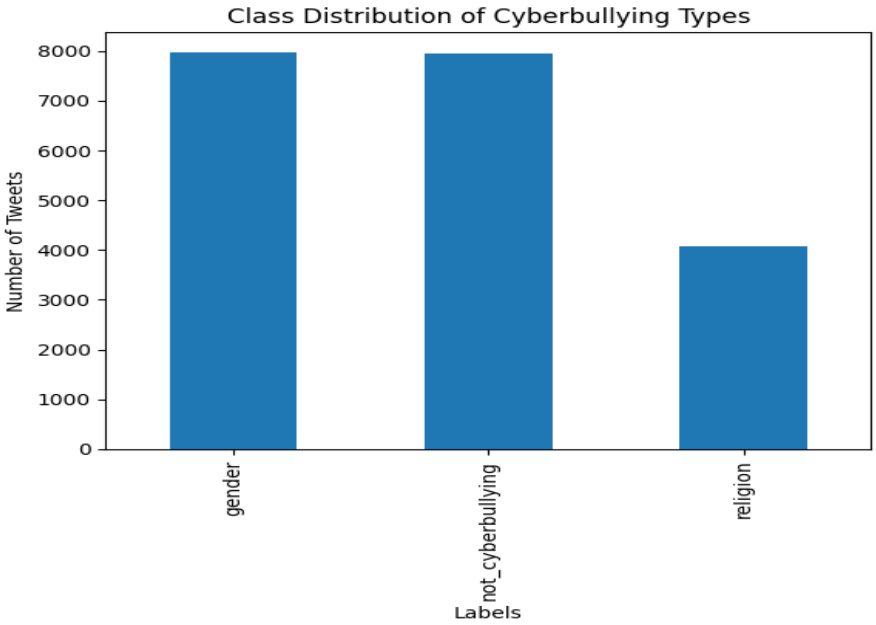

Figure 1: Class distribution of the dataset.

Missing Values: The dataset was assessed for any instances of missing values, and necessary actions were made to address any data that was found to be missing. The number of missing values in each column was documented to assure the completeness of the data.

Unique Values: The assessment of the number of unique values in each column was conducted to get insight into the variety and distinctiveness of the data points.

Word Frequency Analysis: A word frequency analysis was conducted to identify the most prevalent words in the tweets, with the aim of gaining insight into the normal vocabulary used in cyberbullying tweets. A bar chart was generated to visually represent the frequency of the most frequently occurring terms, so offering valuable insights on prevalent themes and linguistic patterns as shown in Figure 2. Figure 3 represents the bag of words.

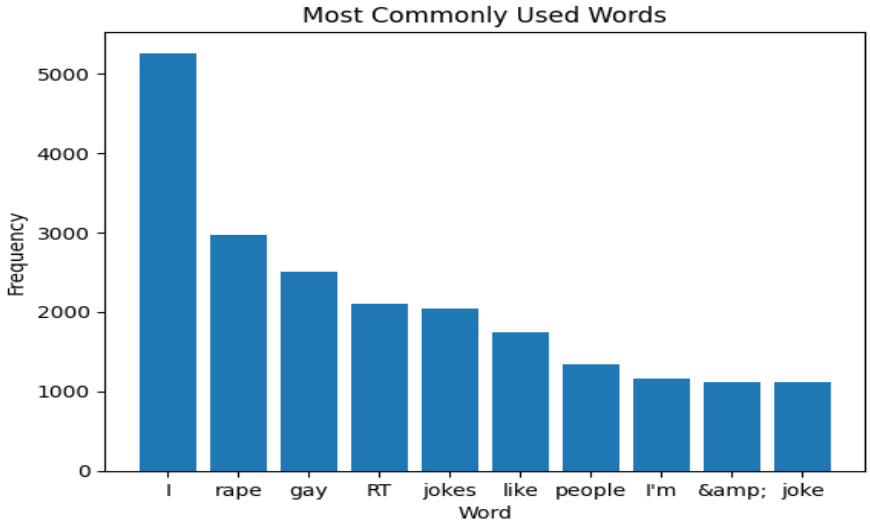

Figure 2: Most Used Words in the dataset.

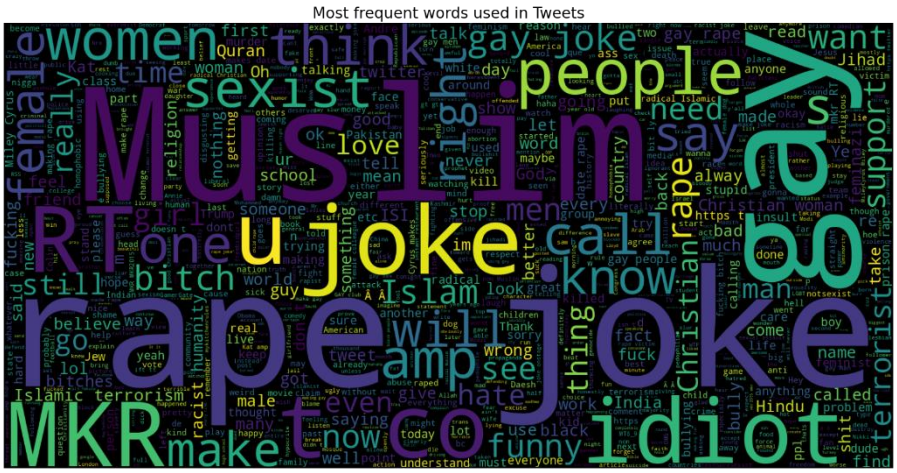

Figure 3: Bag of Words

The relational dependency among the words in the sentences were also considered. Figure 4 represents syntactic structure of the sentence. It represents is a dependency parse tree of a sentence, which illustrates the grammatical structure of the sentence by showing how words are related to each other. Each arrow points from a head word to its dependent, representing the syntactic dependencies between words.

Figure 4: Syntactic Structure of a sentence

3.4 Proposed Methodology

A stacking ensemble model, which includes many base classifiers, is presented to improve the classification performance of cyberbullying detection. The suggested ensemble approach utilizes the advantages of individual classifiers to enhance overall accuracy and resilience. The subsequent machine learning models were employed as fundamental classifiers.

Naive Bayes is a probabilistic classifier that is well-known for its simplicity and success in applications involving text classification. A Support Vector Machine (SVM) is a powerful classifier that can effectively handle high-dimensional data and is particularly suitable for text classification.

A decision tree is a model that uses a tree structure to split the data according to the relevance of different features in order to make predictions. Random Forest is a collection of decision trees that enhances classification accuracy by mitigating overfitting. The k-Nearest Neighbors (k-NN) algorithm is a straightforward and intuitive classifier that assigns labels by considering the majority class of the nearest neighbors. Gradient Boosting is an ensemble strategy that constructs several weak learners, typically decision trees, to enhance the accuracy of the model.

The stacking ensemble model aggregates the predictions of the basis classifiers by employing a logistic regression meta-classifier. The ensemble model's performance was optimized by conducting hyperparameter tuning using GridSearchCV. The ultimate model was trained using the TF-IDF features and assessed on the test set. The hyper-parameters used for the training are listed in table 2.

Table 2: Hyper-parameters used for training the dataset

Base Classifier	Hyperparameters	Values
Multinomial Naive Bayes	**alpha**	[0.1, 0.5, 1.0]
Support Vector Machine	**C**	[0.1, 1, 10]
	kernel	['linear', 'rbf']
Decision Tree	**max_depth**	[None, 10, 20]
	min_samples_split	[2, 5, 10]
Random Forest	**n_estimators**	[50, 100]
	max_depth	[None, 10, 20]
k-Nearest Neighbors	**n_neighbors**	[3, 5, 7]
Gradient Boosting	**n_estimators**	[50, 100]
	learning_rate	[0.01, 0.1]

4. Results and Discussion

The classification of cyberbullying tweets was assessed by evaluating the performance of different machine learning models using metrics such as accuracy, precision, recall, and F1-score. The models utilized in the analysis consisted of Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, k-Nearest Neighbors (k-NN), Gradient Boosting, and an Ensemble Model. The findings are concisely reported in the table 3. The comparative analysis of the proposed model with the other state of art methods in terms of accuracy, precision, recall and F1-Score is presented by Figure 5, Figure 6, Figure 7 and Figure 8 respectively.

The results comparison demonstrates that the Ensemble Model outperforms all individual classifiers in terms of all evaluation parameters, attaining the best accuracy of 90.50%, precision 90.65%, recall 0.90.50%, and F1-score 0.90.53%. The exceptional result demonstrates the efficacy of integrating different classifiers to capitalize on their individual capabilities and minimize their limitations. Random Forest and Support Vector Machines (SVM) provide superior performance among the individual classifiers. Random Forest marginally surpasses SVM in terms of precision and recall, suggesting its robustness and capacity to address overfitting through ensemble learning.

In contrast, the k-Nearest Neighbors (k-NN) classifier performs poorly, achieving an accuracy of 45.10% and an F1-score of 33.92%. This indicates that it is not suitable for this specific text classification problem. Although Gradient Boosting is not as powerful as the Ensemble Model, it nonetheless exhibits robust performance, closely aligning with the metrics of SVM and Random Forest. The findings emphasize the significance of ensemble learning techniques, specifically stacking, in enhancing the precision and dependability of automated cyberbullying detection systems. This offers a more efficient approach for practical use in real-world scenarios.

Table 3: Comparison of the proposed model with state-of-art model

Model	Accuracy (%)	Precision(%)	Recall(%)	F1-score (%)
Naive Bayes	0.8325	0.8355	0.8325	0.8319
SVM	0.8945	0.9030	0.8945	0.8950
Decision Tree	0.8630	0.8635	0.8630	0.8632
Random Forest	0.8955	0.8998	0.8955	0.8959
k-NN	0.4510	0.6979	0.4510	0.3392
Gradient Boosting	0.8815	0.8932	0.8815	0.8822
Ensemble Model	0.9050	0.9065	0.9050	0.9053

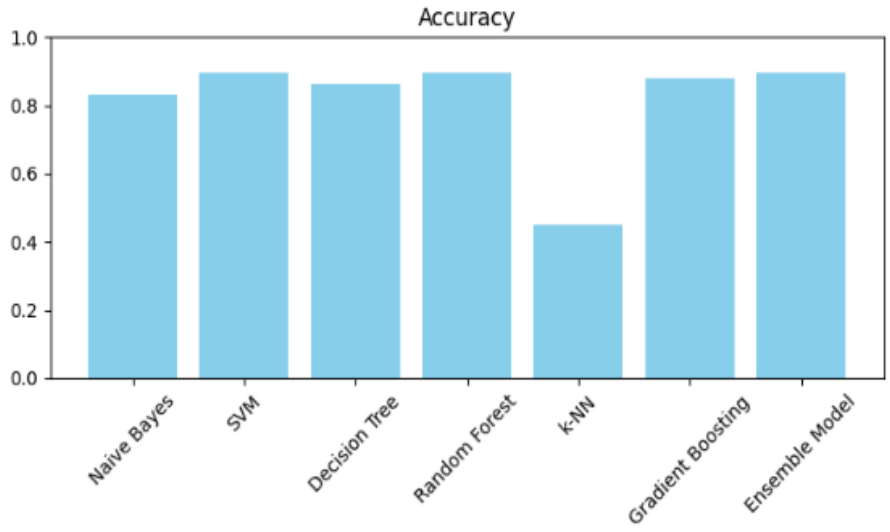

Figure 5: Accuracy Comparison of the Models

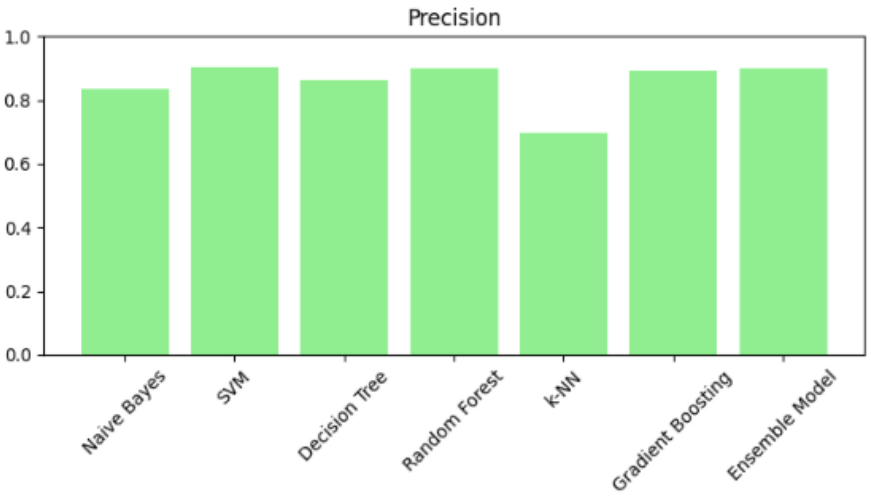

Figure 6: Precision Comparison of the Models

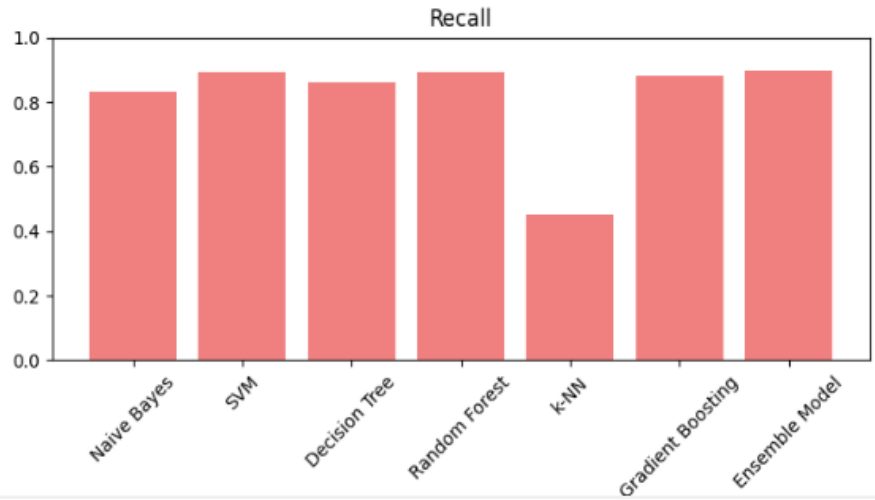

Figure 7: Recall Comparison of the Models

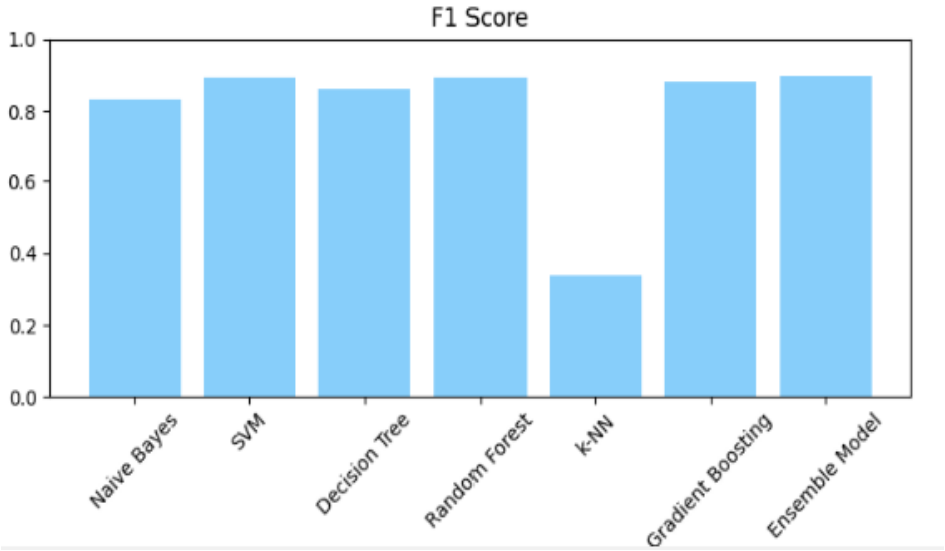

Figure 8: F1-Score Comparison of the Models

5. Conclusion

The objective of this research was to improve the identification and categorization of cyberbullying in tweets by employing different machine learning models and a sophisticated ensemble learning technique. The study utilized an extensive dataset of categorized tweets, which underwent thorough pre-processing to guarantee the excellence and relevancy of the input data. A variety of distinct classifiers, such as Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, k-Nearest Neighbors (k-NN), and Gradient Boosting, were trained and assessed using metrics such as accuracy, precision, recall, and F1-score.

The findings indicated that whereas SVM and Random Forest, as standalone classifiers, demonstrated satisfactory performance, the k-NN classifier exhibited notably inferior performance. The proposed Ensemble Model, which utilized a stacking strategy to integrate the predictions of many base classifiers, demonstrated superior performance compared to each individual classifier. The Ensemble Model had the highest performance measures, showcasing exceptional accuracy (0.9050), precision (0.9065), recall (0.9050), and F1-score (0.9053). The aforementioned evidence suggests that the ensemble method efficiently exploits the advantages of various classifiers, leading to predictions that are both more resilient and precise.

The exceptional performance of the Ensemble Model highlights the capacity of ensemble learning techniques to tackle the intricacies of detecting cyberbullying in social media data. The Ensemble Model combines the various characteristics of different classifiers to provide a strong and effective method for recognizing and regulating cyberbullying content. This helps to create safer online environments.

To summarize, this study emphasizes the significance of utilizing sophisticated machine learning methods, specifically ensemble learning, to improve the efficiency of automated systems for detecting cyberbullying. Future research should investigate the incorporation of more advanced models and the utilization of these techniques on broader and more varied datasets to consistently enhance the resilience and dependability of cyberbullying detection technologies.

References

[1] M. Dadvar, D. Trieschnigg, and F. de Jong, *Expert knowledge for automatic detection of bullies in social networks*. 2013.

[2] R. Kowalski, G. Giumetti, A. Schroeder, and M. Lattanner, "Bullying in the Digital Age: A Critical Review and Meta-Analysis of Cyberbullying Research Among Youth," *Psychological bulletin,* vol. 140, 02/10 2014, doi: 10.1037/a0035618.
[3] J. Patchin and S. Hinduja, "Measuring Cyberbullying: Implications for Research," *Aggression and Violent Behavior,* vol. 23, 05/22 2015, doi: 10.1016/j.avb.2015.05.013.
[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 2016. [Online]. Available: https://doi.org/10.1145/2939672.2939785.
[5] H. Gummadavelly, K. Preethi, S. Goud, K. Kanth, Reddy, and N. S. Ramchander, "CYBER BULLYING DETECTION USING MACHINE LEARNING," 2021.
[6] M. Jinan Redha, "Cyberbullying Messages Detection Using Machine Learning and Deep Learning," *International Journal of Advances in Scientific Research and Engineering (IJASRE), ISSN:2454-8006, DOI: 10.31695/IJASRE,* vol. 10, no. 3, pp. 19-29, 03/16 2024, doi: 10.31695/IJASRE.2024.3.3.
[7] V. Yoganand Bharadwaj, V. Likhitha, V. Vardhini, A. Uma Sree Asritha, S. Dhyani, and M. Lakshmi Kanth, "Automated Cyberbullying Activity Detection using Machine Learning Algorithm," *E3S Web Conf.,* vol. 430, p. 01039, 2023. [Online]. Available: https://doi.org/10.1051/e3sconf/202343001039.
[8] T. Balet, Q. Vo, O. Salem, and A. Mehaoua, "Cyberbullying Detection on tweets from Twitter using Machine Learning Algorithms," in *2023 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)*, 19-22 June 2023 2023, pp. 177-182, doi: 10.1109/ICCNS58795.2023.10193450.
[9] V. Balakrishnan, S. Khan, T. Fernandez, and H. R. Arabnia, "Cyberbullying detection on twitter using Big Five and Dark Triad features," *Personality and Individual Differences,* vol. 141, pp. 252-257, 2019/04/15/ 2019, doi: https://doi.org/10.1016/j.paid.2019.01.024.
[10] N. Novalita, A. Herdiani, I. Lukmana, and D. Puspandari, "Cyberbullying identification on twitter using random forest classifier," *Journal of Physics: Conference Series,* vol. 1192, no. 1, p. 012029, 2019/03/01 2019, doi: 10.1088/1742-6596/1192/1/012029.
[11] G. Thangarasu and K. R. Alla, "Detection of Cyberbullying Tweets in Twitter Media Using Random Forest Classification," in *2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 20-21 May 2023 2023, pp. 113-117, doi: 10.1109/ISCAIE57739.2023.10165118.
[12] "Comparing the Random Forest vs. Extreme Gradient Boosting using Cuckoo Search Optimizer for Detecting Arabic Cyberbullying," *Iraqi Journal of Science,* vol. 64, no. 9, pp. 4806- 4818, 09/30 2023, doi: 10.24996/ijs.2023.64.9.40.
[13] A. F. Alqahtani and M. Ilyas, "A Machine Learning Ensemble Model for the Detection of Cyberbullying," *arXiv preprint arXiv:2402.12538,* 2024.
[14] M. Alotaibi, B. Alotaibi, and A. Razaque, "A multichannel deep learning framework for cyberbullying detection on social media," *Electronics,* vol. 10, no. 21, p. 2664, 2021.
[15] S. Bharti, A. K. Yadav, M. Kumar, and D. Yadav, "Cyberbullying detection from tweets using deep learning," *Kybernetes,* vol. 51, no. 9, pp. 2695-2711, 2022.
[16] R. Shah, S. Aparajit, R. Chopdekar, and R. Patil, "Machine learning based approach for detection of cyberbullying tweets," *Int. J. Comput. Appl,* vol. 175, no. 37, pp. 51-56, 2020.
[17] B. A. Talpur and D. O'Sullivan, "Multi-class imbalance in text classification: A feature engineering approach to detect cyberbullying in twitter," in *Informatics*, 2020, vol. 7, no. 4: MDPI, p. 52.
[18] R. Daniel *et al.*, "Ensemble Learning with Tournament Selected Glowworm Swarm Optimization Algorithm for Cyberbullying Detection on Social Media," *IEEE Access,* 2023.

[19] Y. J. N. Kumar *et al.*, "Detecting cyberbullying in social media using text analysis and ensemble techniques," in *E3S Web of Conferences*, 2024, vol. 507: EDP Sciences, p. 01069.
[20] V. S. Venu, H. Shanmugasundaram, M. R. Seelam, V. V. R. Kotha, S. S. R. Muthyala, and S. Kansal, "Detection of Cyberbullying on User Tweets and Wikipedia Text using Machine Learning," in *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 2023: IEEE, pp. 327-332.
[21] E. Raisi and B. Huang, "Weakly supervised cyberbullying detection using co-trained ensembles of embedding models," in *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, 2018: IEEE, pp. 479-486.
[22] K. Saranyanath, W. Shi, and J.-P. Corriveau, "Cyberbullying Detection using Ensemble Method," in *CS & IT Conference Proceedings*, 2022, vol. 12, no. 15: CS & IT Conference Proceedings.
[23] N. A. Azeez, S. O. Idiakose, C. J. Onyema, and C. Van Der Vyver, "Cyberbullying detection in social networks: Artificial intelligence approach," *Journal of Cyber Security and Mobility,* vol. 10, no. 4, pp. 745-774, 2021.
[24] M. A. Al-Ajlan and M. Ykhlef, "Optimized twitter cyberbullying detection based on deep learning," in *2018 21st Saudi Computer Society National Computer Conference (NCC)*, 2018: IEEE, pp. 1-5.
[25] H. Shukla, N. Jagtap, and B. Patil, "Enhanced Twitter bot detection using ensemble machine learning," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021: IEEE, pp. 930-936.
[26] L. Ketsbaia, B. Issac, and X. Chen, "Detection of hate tweets using machine learning and deep learning," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020: IEEE, pp. 751-758.
[27] A. Muneer, A. Alwadain, M. G. Ragab, and A. Alqushaibi, "Cyberbullying detection on social media using stacking ensemble learning and enhanced BERT," *Information,* vol. 14, no. 8, p. 467, 2023.
[28] N. Sreevidya, A. Hamsini, R. Vainateya, and C. A. Naidu, "Ensemble Learning Based Prediction for Cyber Harassment Observations on Tweets," *Asian Journal of Research in Computer Science,* vol. 17, no. 6, pp. 102-113, 2024.
[29] S. A. Mathur, S. Isarka, B. Dharmasivam, and C. Jaidhar, "Analysis of Tweets for Cyberbullying Detection," in *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 2023: IEEE, pp. 269-274.
[30] J. Qiu, M. Moh, and T.-S. Moh, "Multi-modal detection of cyberbullying on Twitter," in *Proceedings of the 2022 ACM Southeast Conference*, 2022, pp. 9-16.
[31] K. S. Alam, S. Bhowmik, and P. R. K. Prosun, "Cyberbullying detection: an ensemble based machine learning approach," in *2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV)*, 2021: IEEE, pp. 710-715.