Journal of Intelligent Learning Systems and Applications, 2023, 15, 108-122
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

# Social Media Cyberbullying Detection on Political Violence from Bangla Texts Using Machine Learning Algorithm

**Md. Tofael Ahmed[1,2], Almas Hossain Antar[1], Maqsudur Rahman[1], Abu Zafor Muhammad Touhidul Islam[3], Dipankar Das[2], Md. Golam Rashed[2]**

[1]Department of Information and Communication Technology, Comilla University, Comilla, Bangladesh
[2]Department of Information and Communication Engineering, University of Rajshahi, Rajshahi, Bangladesh
[3]Department of Electrical & Electronics Engineering, University of Rajshahi, Rajshahi, Bangladesh
Email: tofael@cou.ac.bd, almashossain121@gmail.com

**How to cite this paper:** Ahmed, Md.T., Antar, A.H., Rahman, M., Islam, A.Z.M.T., Das, D. and Rashed, Md.G. (2023) Social Media Cyberbullying Detection on Political Violence from Bangla Texts Using Machine Learning Algorithm. *Journal of Intelligent Learning Systems and Applications*, **15**, 108-122.
https://doi.org/10.4236/jilsa.2023.154008

**Received:** August 8, 2023
**Accepted:** October 9, 2023
**Published:** November 6, 2023



## Abstract

When someone threatens or humiliates another person online by sending those unpleasant messages or comments, this is known as Cyberbullying. Recently, Bangla text has been used much more often on social media. People communicate with others on social media through messages and comments. So bullies use social media as a rich environment to bully others, especially on political issues. Fights over Cyberbullying on political and social media posts are common today. Most of the time, it does a lot of damage. However, few works have been done for monitoring Bangla text on social media & no work has been done yet for detecting the bullying Bangla text on political issues due to the lack of annotated corpora and morphologic analyzers. In this work, we used several machine learning classifiers & a model. That will help to detect the Bangla bullying texts on social media. For this work, 11,000 Bangla texts have been collected from the comments section of political Facebook posts to make a new dataset and labelled the data as either bullied or not. This dataset has been used to train the machine learning classifier. The results indicate that Random Forest achieves superior accuracy of 91.08%.

## Keywords

Cyberbullying, Bangla Texts, Political Issues, Machine Learning, Random Forest, Social Media

## 1. Introduction

In contemporary times, individuals are dedicating a significant portion of their

time to engaging with various social media platforms. Individuals often exhibit a propensity for sharing their opinions and expressions on various social media platforms, thereby engaging in active communication with other individuals within their respective networks. The extensive utilization of social media presents numerous advantages and disadvantages. The potential negative impact of Cyberbullying is often cited as a prominent disadvantage. The phenomenon under investigation is commonly referred to as deliberate anguish experienced through electronic media. The rapid dissemination of information to a wide-ranging audience, particularly in relation to political matters, is a notable phenomenon. The impact of bullying on social media platforms can be significantly more severe compared to other forms of bullying. The utilization of the Unicode system, coupled with the widespread adoption of the internet, has led to a surge in the usage of Bangla language for political discourse on various social media platforms [1]. The phenomenon of bullying in relation to political issues has the potential to give rise to significant levels of violence within a given geographical region. Detecting and preventing political Cyberbullying in social media has emerged as a potent solution to address this issue. The current perspective in Bangladesh highlights the need for monitoring bullying posts and comments on political issues in Bangla text. However, the lack of sufficient annotated corpora, dictionaries, and morphological analyzers presents a challenge in this area of research. Further work is required to address this gap and effectively address the demands of the situation. The phenomenon of political Cyberbullying has witnessed a significant surge globally, with particular emphasis on its prevalence in Bangladesh. Bangla, also known as Bengali, holds the distinction of being the second most widely used language in the Indian subcontinent. Furthermore, it ranks as the seventh most spoken language worldwide. It is worth noting that the prevalence of Cyberbullying pertaining to political matters through the use of Bangla text is a significant concern. However, the exploration of identifying political bullying remains largely uncharted territory in the field of research.

The development of a model for the detection of political Cyberbullying is a pressing requirement in current research. According to recent studies, there is evidence to suggest that Facebook is frequently utilized as a platform for engaging in a particular form of bullying [1] [2]. According to recent research, it has been found that a significant majority, specifically over 93%, of the total Internet user population in Bangladesh, which stands at approximately 123.82 million individuals, actively engage with social media platforms. Among these platforms, Facebook emerges as the most popular choice among Bangladeshi Internet users. Facebook has been identified as a valuable platform for data collection and the creation of new datasets. Typically, when collecting data from Facebook, it is necessary to ensure that the data is in the appropriate format. According to previous research [2], it has been suggested that the conventional approach to bullying detection may prove ineffective in identifying certain types of data. Recent studies have indicated that machine learning algorithms have exhibited superior

accuracy compared to the traditional keyword search method, as demonstrated by the researcher [3] [4]. However, it is important to note that many machine learning methods are designed to be highly specialized for specific topics. The performance of content may exhibit variability as a result of linguistic disparities between non-English and English materials. In a study conducted by researchers [5], it was observed that Support Vector Machines (SVM) exhibited lower accuracy when applied to Arabic text. However, SVM demonstrated improved performance when applied to English text.

Hence, the main goal is to create political Cyberbullying detection and monitoring system for Bangla text on social media networks. So, the targets of this paper:

• Since we did not get any existing dataset for our work, we need to build an entirely new dataset that is robust, clean, structured, and labelled correctly.

• Develop a model for analyzing political Bangla text on social media by fusing text analytics and machine learning techniques.

• Compare the machine learning classifier performance, and find the best one which provides the best performance for detecting political bullying from Bangla text.

## 2. Background

Significant research efforts have been dedicated to text categorization and cyberbully detection in English text, while comparatively less attention has been given to detecting Bangla text in these domains. However, further analysis is required to investigate the detection and prevention of bullying in the context of Bangla text related to political issues on social media platforms. In a study conducted by Reichart, Dinakar, and Lieberman, a comprehensive analysis was performed to compare different supervised methods for text classification. The researchers aimed to evaluate the effectiveness and performance of these techniques in order to identify the most suitable approach for this task. By examining various supervised approaches, the study sought to provide valuable insights into the strengths and weaknesses of each method, ultimately contributing to the advancement of text classification research. In the study conducted by Z. Xue, Yin, and Hong [3], supervised learning was employed for text classification.

The researchers labelled the texts using the Ngrams methodology and applied a weighting technique known as TF-IDF. However, it should be noted that the dataset in question is relatively small in size. In a study conducted by Kelly Reynolds, the Decision tree (J48) and k-nearest neighbour (KNN) algorithms were utilised [6]. In order to compile a comprehensive dataset, researchers gather YouTube video comments as a primary source of information. In this study, the data was manually labelled to ensure accuracy and reliability. The research employed a multiclass classification approach, which allowed for the categorization of the data into multiple distinct classes. The researchers employed a limited number of classifiers, specifically two, in their study. In a previous study the application of Support Vector Machines (SVM) for text classification was explored.

However, the results indicated that SVM did not exhibit superior performance in this context. In their study, Zhijie *et al.* [7] conducted a comparison between NB, KNN, and SVM. The results indicated that SVM outperformed the other two methods in terms of performance. The research conducted focuses on the analysis of English text, with a particular emphasis on the limited size of the dataset used.

The performance and accuracy of the algorithm may exhibit variability as a result of the linguistic disparities between English and non-English content. In the context of Indian text analysis, it has been observed that the Naive Bayes (NB) algorithm yields superior results. Additionally, a combination of Ontology-based classification and NB has been found to produce even better outcomes specifically for Panjabi text. In a study conducted by researchers, it was found that Support Vector Machines (SVM) outperformed Naive Bayes (NB) in the context of Urdu text analysis [8]. According to previous research [9], it has been found that the utilisation of the Artificial Neural Network (ANN) model-based approach for Tamil text yields improved performance. In a previous study conducted on Indian text classification, researchers successfully employed the NB classifier. However, the specific quantity of data utilised in their experiment was not explicitly mentioned. In a study conducted by B Nandhini *et al.* [10], a Naïve Bayes machine learning approach was proposed. The researchers reported a notable accuracy rate of 91% in their findings.

The dataset utilised in the study was obtained from MySpace.com. Subsequently, an additional model was proposed, namely the integration of NB (Naive Bayes) and FuzGen (Fuzzy Genetic Algorithm), yielding a commendable accuracy of 87%. The researchers utilise an established dataset rather than generating a novel one. In their study, Shane Murnion *et al.* [11], gathered data from the chat logs of War of Tanks games and conducted manual classification of the collected data. Subsequently, a comparison between the obtained outcome and the Naïve classification shall be conducted. The obtained outcome, unfortunately, exhibits a significant lack of quality. In a study conducted by Walisa Romsaiyud *et al.* [12], the authors introduced the Naive Bayes (NB) algorithm as a solution to achieve a high accuracy rate of 95.79% for a dataset sourced from Slashdot, Kongregate, and MySpace.

However, a limitation of their approach is that the cluster processes do not operate concurrently, posing a challenge in terms of efficiency and parallelization. In their study, Noviantho *et al.* [13] obtained a dataset from Kaggle and employed Support Vector Machines (SVM) and Naive Bayes (NB) algorithms for their analysis. The obtained accuracy for Support Vector Machines (SVM) was 97.11%, while Naive Bayes (NB) achieved an accuracy of 92.81%. However, it is important to note that the user did not provide information regarding the size of the dataset used for training or testing. Consequently, the reliability of these results may be compromised due to the lack of specificity regarding the dataset proportions. In their study, Maroun Chamoun *et al.* [14], introduced a

novel model for the detection and identification of cyberbullying specifically in Arabic text. In the conducted study, the researchers employed the Naïve Bayes algorithm, which yielded a precision rate of 90.85%. Subsequently, they utilised the Support Vector Machine (SVM) algorithm, resulting in a precision rate of 94.1%. However, it is worth noting that the SVM algorithm exhibited a relatively high proportion of false positives. The research team achieved impressive results in terms of accuracy, with a score of 96%. However, it is important to note that their dataset suffered from an imbalance issue.

This imbalance had a significant impact on their precision score, which was 56%. Additionally, the misleading findings resulting from the dataset's imbalance further underscore the importance of addressing this issue in future research. The researchers utilised the WEKA machine learning software in conjunction with a dataset acquired from the popular social networking platform, Myspace. In a study conducted by Xiang Zhang *et al.* [15], a novel approach was proposed to tackle the challenges of limited data availability, class imbalance, and the presence of noise and bullying in the dataset. The researchers introduced a pronunciation-based convolutional neural network (PCNN) as a potential solution to these issues. In a recent study, it was found that Twitter, a popular social media platform, sent a total of 13,313 messages. Additionally, another social media platform called Formspring.me was observed to have sent out 13,000 messages. These findings highlight the significant volume of messages being exchanged on these platforms, indicating the widespread usage and engagement of users. The calculation of accuracy was impeded by the uneven nature of the Twitter dataset.

Franciska De Jong *et al.* [16] [17] [18] [19] used a Support Vector Machine classifier in their first and second papers. The main distinction between the two publications is that the second research incorporated gender information in categorization, resulting in results of 43% precision, 16% recall, and no mention of accuracy. Furthermore, they collected 4626 comments from 3858 users for their second article. Those comments are labelled bullying and non-bullying manually. They used the SVM classifier, and their findings ranged up to 78% accuracy and 55% recall. They can use more classifiers. Md. Tofael Ahmed, Maqsudur Rahman *et al.* [20] used three datasets, and SVM provided the best performance (76% accuracy) for the first dataset and Multinomial Naive Bayes outperformed for the second (84%) and third (80%) datasets. The size of the first dataset is small. Md. Tofael Ahmed, Maqsudur Rahman *et al.* [21] proposed a model to detect bullying in Bangle and Romanized Bangla text. They use three datasets one for Bangla, another is Romanized Bangla, and the last combines the previous two datasets. They collected the data from the Youtube comments and preprocessed the data using machine learning and deep learning algorithms. CNN performs best for Bangla text and MNB for the other two. Their first dataset could be better. Md. Tofael Ahmed, Maqsudur Rahman *et al.* [22] introduced the PMI-SO model to detect cyberbullying in Bangla text. They use two datasets. The first contains 5000 Bangla texts collected from Social Media, and the last is the PMI dataset, containing 10,277 Bangla texts. XGBoost provides the best ac-

curacy to them (93%), whereas SVM delivers 85%, the lowest accuracy. Pushpita Shil, Umma Saima *et al.* [23] proposed a model for detecting spam from Bangla text. For this, they collect 2759 Facebook comments. Use KNN, SVC, and Gradient Boosting & MNB for classification.

However, their dataset contains a small amount of data. It could have been bigger. Arnisha Akther *et al.* [24] proposed a robust hybrid ML model for detecting cyberbullying in the Bengali language on social media. They use 44,001 available Bangla texts for their dataset and get 98.57% and 98.82% in binary and multilevel classification, respectively. Priya and Sachin Gupta [25] collect data from Twitter to identify political hate speech with the help of a machine learning algorithm. They did not specify the number of pieces of data they used. Pranav Kompally *et al.* [26] discussed a decentralized deep-learning approach called Malang to detect abusive textual content. It has two levels called System and cloud. The system level reads the user message and classifies the text. After that, it sends the abused text to the cloud. Cloud Level used deep learning to determine the toxic content categories; its success percentage is 98.2%.

This study makes use of a recently obtained new dataset that has been carefully balanced to ensure equal representation of various factors. The majority of individuals utilize a limited selection of classifiers. The potential for discovering the optimal classifier is constrained. But in this study, a total of seven classifiers were employed. The outcomes generated by this phenomenon exhibit a considerable degree of variability, indicating that they do not fall within a narrow range. In this research work, it is necessary to increase the magnitude of the result difference observed in the machine learning classifier.

## 3. Research Methodology

### 3.1. Proposed Model

First, we manually collect the Bangla text from the political Facebook posts and use our developed Python scrapper. Store the collected data in the CSV file. The data collected from Facebook are too noisy, multilingual, unstructured, and emoji mix-up with the content. It requires clean-up for better results. For this, we perform data cleaning and preprocessing operations. Then extract the feature of the data. Use TF-IDF to remove the textual feature. After that, we divided the dataset into two parts. Training data and test data. Train data have been used to train the classifiers and test data to analyze the proposed model performance. In this work, seven machine learning classifiers are used. The proposed models consist of the components shown in **Figure 1**.

### 3.2. Data Set

Bullying on political issues mainly occurs on Facebook because a maximum number of people in Bangladesh use it. Now, it has made a habit for people to use Facebook to share their thoughts and express their opinions. So, Facebook is the best social media to collect such political bullying data. However, the problem

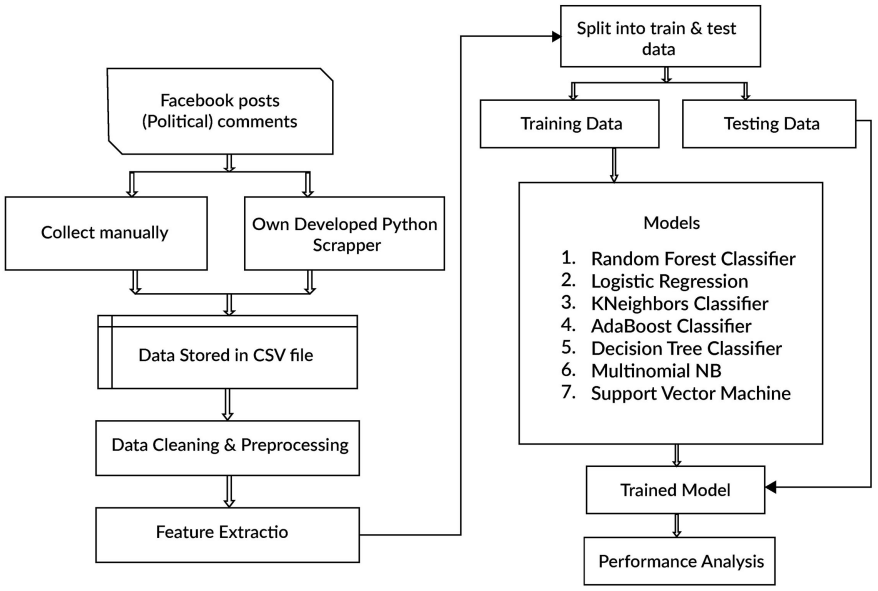


**Figure 1.** Proposed model for bullying detection.

is that collecting data from Facebook is a challenging work. Facebook does not provide the comments data of specific posts. As a result, a substantial portion of the data (6500) had to be manually collected from the comments section of political Facebook post and an in-house Python scraper was developed to automatically gather the remaining data (4500). A total of 11,000 Facebook comments from the political posts are collected to construct the dataset, then labelled the data manually, either bullied or not-bullied. The dataset contained 49% bullying text, and two experts manually verified the dataset. Then dataset is divided into two parts training data and test data. Eight thousand two hundred fifty records are utilised for training, and 2750 records are utilised for testing.

### 3.3. Data Preprocessing & Feature Extraction

The data acquired from Facebook needs to be louder, and it has to be more uniform, unstructured, and brief across many languages. To ensure that the dataset accurately reflects the information being sought, punctuation, Unicode emotions, special characters, stop words, emoji, white space, undesired words, and precise words were eliminated by hand. In addition, special characters are becoming obsolete as a result of the fact that there is no longer a demand for them. The subsequent stage is the extraction of features. In order to extract features from the dataset, we utilize a tool called term frequency-inverse term frequency (TF-IDF). The TF-IDF algorithm is a sophisticated feature extraction approach that searches through textual data for meaningful words. It converts the text into a numbers vector so that machine learning classifiers may use it. Find the frequency of occurrence (TF) of a specific word in the document; the result is the number of times the word appears divided by the total number of words in the document. We annotated all the datasets into two categories such as bullying and non-bullying. Some of the annotated data is presented in **Figure 2**.

| Texts | Label |
|---|---|
| [illegible] ধান কেটে দিবে | 1 (Bullying) |
| [illegible] ভালো মানুষ | 0 (Non-Bullying) |
| উনার স্ত্রীর সাথে এম সি কলেজের ঘটনা ঘটাইতে চাইছিল [illegible] | 1 (Bullying) |
| আগে [illegible] কে ধরা হোক | 1 (Bullying) |
| এটা পুরাতন কথা [illegible] | 1 (Bullying) |
| এইটা [illegible] দের কাজ | 1 (Bullying) |

**Figure 2.** Sample dataset for proposed model.

$$\mathrm{TF} = \frac{\text{Number of repetance of a word in the documents}}{\text{Total number of words in the documents}} \tag{1}$$

IDF determines the importance of each word in a document.

$$\mathrm{IDF} = \log_2 \frac{\text{Total number of documents}}{\text{Number of documents containing pairticular terms}} \tag{2}$$

To obtain TF-IDF, both TF and IDF need to be multiplied. The result of this multiplication proved the normalized weights.

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF} * \mathrm{IDF} \tag{3}$$

### 3.4. Machine Learning Classifier

The utilization of machine learning classifiers for forecasting categorical data is a widely adopted practice in various domains. Machine learning encompasses three primary learning methods: supervised, unsupervised, and reinforcement. These methods play a crucial role in the field of machine learning, enabling the development of intelligent systems capable of acquiring knowledge and making informed decisions. By understanding and utilizing these learning methods, researchers and practitioners can leverage the power of machine learning to solve complex problems and enhance various applications across different domains. In this study, a total of seven supervised machine learning classifiers were utilized. Decision Tree (DT), Multinomial naive Bayes (MNB), AdaBoost Classifier (AB), and Support Vector Machine (SVM) are among the well-known supervised learning methods frequently utilized in various domains of research and application. These methods have gained significant attention due to their effectiveness in solving classification problems. The information provided proved to be valuable in influencing the development of the ultimate model.

#### 3.4.1. Random Forest Classifier

In this study, an ensemble classifier was implemented, utilizing the Decision Tree algorithm in a randomized manner. The purpose of employing an ensemble approach is to enhance the overall predictive performance of the classifier. By randomly selecting subsets of the training data and constructing decision trees based on these subsets, the classifier aims to capture different aspects of the data and reduce the risk of over fitting. This randomized approach allows for increased diversity among the individual decision trees, leading to a more robust and accurate ensemble classifier. In the Random Forest algorithm, a Bootstrap dataset is generated by randomly selecting documents from the original dataset.

#### 3.4.2. Logistic Regression

The calculation or prediction of the probability of a binary event is a common application in various fields of research. The sigmoid function was employed for the purpose of data classification. This function assigns a probabilistic value within the range of 0 to 1 [27].

#### 3.4.3. KNeighbor Classifier

The utilization of the distance metric to measure the dissimilarity or similarity between data instances is a common approach in various research domains. This technique allows for the quantification of the proximity or separation between data points, enabling the analysis and comparison of their characteristics. By employing distance-based methods, researchers can effectively assess the relationships and patterns within datasets the calculation of distance in this study employed the Euclidean distance metric.

#### 3.4.4. AdaBoost Classifier

The proposed approach involves the use of a meta-estimator that first trains a classifier on the original dataset. Subsequently, additional copies of the trained classifier are created and applied to the same dataset. This technique aims to enhance the performance of the classifier by leveraging the collective decision-making capabilities of multiple instances of the classifier. Within this context, it is evident that a multitude of base learners are present. The calculation of the sample weight for each feature is performed. The sample weight is updated based on the observed performance.

#### 3.4.5. Decision Tree Classifier

In a decision tree classifier, there are two types of nodes: the Decision node, also known as the Test node, and the Leaf node, often referred to as the Classification node. The process begins with the entire dataset being positioned at the initial Decision node. Based on the outcomes of the applied tests, the dataset is then divided recursively, and this process continues iteratively [28].

#### 3.4.6. Multinomial Naive Bayes

In the realm of text classification, there exists a widely utilized approach known as supervised learning. This algorithmic technique involves training a model using a labelled dataset, where each text sample is associated with a predefined class or category. By leveraging this labelled data, the supervised algorithm is able to learn patterns and the probability calculation is performed using the Bayes theorem, as described in reference [29]. The determination of the most probable value is achieved by calculating the probability for a specific sample.

#### 3.4.7. Support Vector Machine (SVM)

It categorizes data by establishing a decision boundary or hyperplane that separates the target class from the opposing class within an n-dimensional space [30]. To determine the optimal hyperplane, identifying the margin with the highest value is crucial. This margin is defined as the gap between the last data

point of the target class and the nearest data point of the opposing class.

## 4. Result and Discussion

The evaluation of performance is a crucial aspect in assessing the effectiveness of machine learning classifiers. In evaluating the performance of the classifiers, various metrics were taken into consideration. These metrics included accuracy, precision, recall, F1-score, and the ROC curve. The purpose of this study is to evaluate the performance of the classifier and present the corresponding results.

The dataset has been partitioned into two distinct segments. In this study, a standard practice was followed where 75% of the available data was allocated for training purposes, while the remaining 25% was reserved for testing the developed model. The feature extraction process involved utilizing the training data to train multiple machine learning classifiers. These classifiers were then used to construct our political bullying detection model. The evaluation of algorithm accuracy and analysis of model performance involve utilizing features extracted from the testing dataset.

In the analysis presented in Figure 3, it is observed that the Random Forest classifier demonstrated accurate identification of 10,019 instances. In the conducted research, the K-Nearest Neighbours (KNN) algorithm successfully identified a total of 8707 cases accurately. In the conducted study, Support Vector Machines (SVM) and AdaBoost algorithms were employed to accurately classify a dataset consisting of 11,000 instances. The SVM algorithm successfully identified 9779 instances correctly, while the AdaBoost algorithm achieved a slightly lower accuracy by correctly identifying 9778 instances. The Random Forest Classifier exhibited superior performance compared to other algorithms as it achieved the highest number of accurately classified instances. In Figure 4, the data depicts the count of instances that have been classified incorrectly. In this study, it was observed that the Random Forest algorithm exhibited a failure to accurately identify the lowest number of instances.

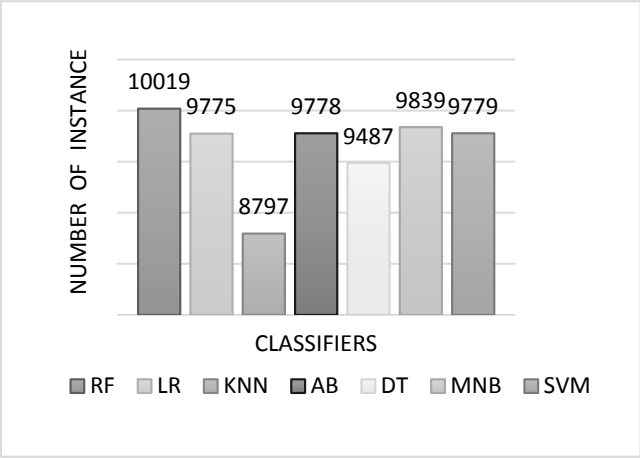


Figure 3. Number of correctly classified instances.

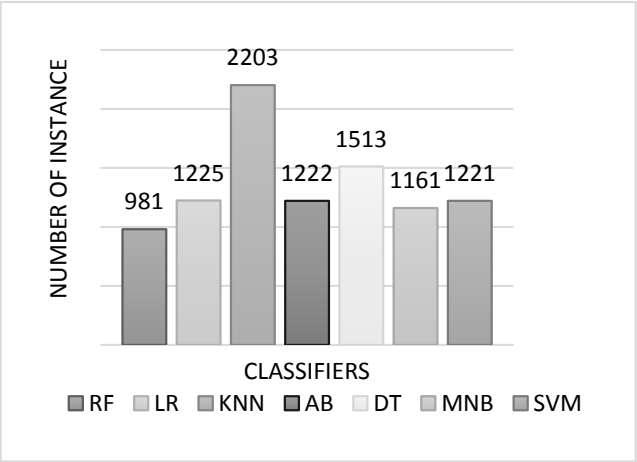


**Figure 4.** Number of incorrectly classified instances.

According to the findings presented in Table 1, the Random Forest algorithm demonstrates notable performance in terms of precision, recall, and F1-score, achieving values of 0.78, 0.82, and 0.80, respectively. Additionally, it attains an overall accuracy rate of 91.08%, which surpasses the accuracy rates of other algorithms examined in the study. The results of the evaluation indicate that the system achieved the highest scores in terms of accuracy, recall, and f1-score. The KNB demonstrates a commendable accuracy rate of 89.44%, positioning it as one of the top performers in terms of accuracy. Additionally, the precision, recall, and F1-score metrics exhibit values of 0.78, 0.72, and 0.76 respectively. The aforementioned options are also deemed to be favorable. According to recent research, the widely used machine learning classifier Support Vector Machine (SVM) has demonstrated an accuracy rate of 88.90%. The precession, recall, and F1-score were found to be 0.78, 0.72, and 0.75, respectively.

The final parameter to be considered in the performance analysis is the area under the Receiver Operating Characteristic (ROC) curve. When the area under the receiver operating characteristic (ROC) curve increases, it indicates that the model has improved in its ability to accurately identify instances. Figure 5 displays the Receiver Operating Characteristic (ROC) curve for all algorithms. According to our research findings, the Random Forest algorithm exhibits a larger area under the Receiver Operating Characteristic (ROC) curve compared to other algorithms. According to research, Random Forest has been found to possess a higher accuracy in identifying instances. The Random Forest Classifier has been found to exhibit superior performance across various evaluation metrics such as accuracy, recall, F1-Score, and ROC curve. Additionally, it has been observed to possess greater significance compared to all other classifiers ranked in the second-best position in terms of precision. According to research, Random Forest has been identified as a highly effective classifier for detecting instances of political bullying text specifically in the Bangla language.

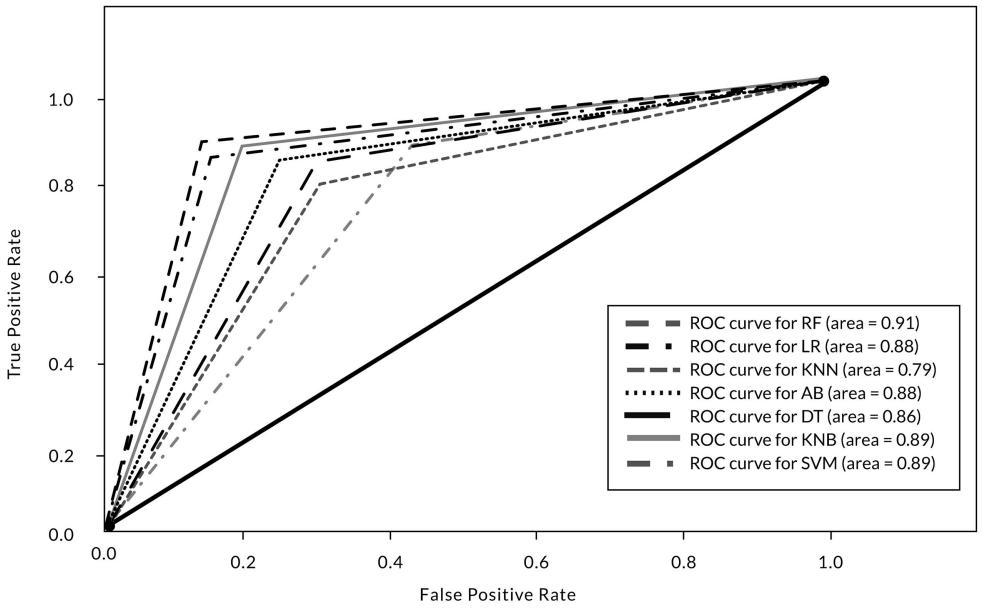


**Figure 5.** ROC curve of all algorithms.

**Table 1.** Result of all classifier model.

| Classifier Name | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Random Forest | 91.08% | 0.78 | 0.82 | 0.80 |
| Logistic Regression | 88.68% | 0.77 | 0.72 | 0.75 |
| KNeighbor | 79.97% | 0.67 | 0.72 | 0.70 |
| AdaBoost | 88.89% | 0.87 | 0.53 | 0.65 |
| Decision Tree | 86.24% | 0.74 | 0.81 | 0.77 |
| Multinomial naive Bayes | 89.44% | 0.78 | 0.73 | 0.76 |
| SVM | 88.90% | 0.78 | 0.72 | 0.75 |

## 5. Conclusion

Extensive research has been conducted to investigate the occurrence and characteristics of Cyberbullying in English texts. However, it is worth noting that there is a dearth of studies examining the phenomenon of political bullying specifically in Bangla texts. In this study, an investigation was conducted to assess the classification performance of various supervised machine learning algorithms, namely Support Vector Machines (SVM), Multinomial Naive Bayes (MNB), k-Nearest Neighbors (KNN), Random Forest, Decision Tree, Logistic Regression, and AdaBoost. The evaluation was carried out using Bangla text data, and the empirical results were analyzed. Based on empirical evidence, it has been observed that the Random Forest algorithm demonstrates superior accuracy and performance in comparison to alternative classification algorithms. This is evident through measures such as recall, F1-score, and the area under the ROC curve. The Random Forest Classifier was employed in order to achieve an accuracy rate of 91.08%. The obtained accuracy is the highest recorded in our

research. The acquisition of such a high level of accuracy in our research will undoubtedly contribute to the advancement of identifying instances of political bullying in Bangla text, thereby improving the overall safety of social media usage for individuals. The identification of Cyberbullying is often hindered by the limited size of training data. In order to enhance performance, a substantial amount of data is necessary. The enhancement of the process can be further achieved by leveraging the importance of specific attributes.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Rice, E., *et al.* (2015) Cyberbullying Perpetration and Victimization among Middle-School Students. AJPH, Washington DC, e66-e72. https://doi.org/10.2105/AJPH.2014.302393

[2] Harsh, Liu, H., Li, J.D., *et al.* (2017) Sentiment Informed Cyberbullying Detection in Social Media. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, Cham, 52-67.

[3] Xue, Z., Hong, L., Yin, D. and Davison, B.D. (2009) Detection of Harassment on Web 2.0., in 1st Content Analysis in Web 2.0 (CAW 2.0), Madrid, Spain.

[4] Huang, Q.J., Singh, V.K. and Atrey, P.K. (2014) Cyberbullying Detection Using Social and Textual Analysis. *Proceedings of the* 3*rd International Workshop on Socially-Aware Multimedia*, Orlando, 7 November 2014, 3-6. https://doi.org/10.1145/2661126.2661133

[5] Wahbeh, Al-Kabi, M. and Abdullah, H. (2012) Comparative Assessment of the Performance of Three WEKA Text Classifiers Applied to Arabic Text. *Abhath Al-Yarmouk* (*Basic Sciences and Engineering*), **21**, 15-28.

[6] Ding, S., Zhu, H., Liu, X.-L. and Zhang, L. (2010) An Overview on Semi-Supervised Support Vector Machine. *Neural Computing and Applications*, **28**, 969-978. https://doi.org/10.1007/s00521-015-2113-7

[7] Liu, Z.J., *et al.* (2010) Study on SVM Compared with the Other Text Classification Methods. 2010 2*nd International Workshop on Education Technology and Computer Science* (*ETCS*), Wuhan, 6-7 March 2010, 219-222.

[8] Gogoi, M. and Sarma, S.K. (2015) Document Classification of Assamese Text Using Naïve Bayes Approach. *International Journal of Computer Trends and Technology*, **30**, 1-5.

[9] Rajan, K., Ramalingam, V., Palaniappan, B., Ganesan, M. and Palanivel, S. (2009) Automatic Classification of Tamil Documents Using Vector Space Model and Artificial Neural Network. *Expert System with Applications*, **36**, 10914-10918. https://doi.org/10.1016/j.eswa.2009.02.010

[10] Nandhini, B. and Sheeba, J.I. (2015) Cyberbullying Detection and Classification Using Information Retrieval Algorithm. *Proceedings of the* 2015 *International Conference on Advanced Research in Computer Science Engineering & Technology* (*ICARCSET* 2015), Unnao, 6-7 March 2015, 20. https://doi.org/10.1145/2743065.2743085

[11] Murnion, S., Buchanan, W.J., Smales, A. and Russell, G. (2018) Machine Learning and Semantic Analysis of In-Game Chat for Cyberbullying. *Computers & Security*,

**76**, 197-213. https://doi.org/10.1016/j.cose.2018.02.016

[12] Romsaiyud, W., Nakornphanom, K., Prasertsilp, P., Konglerd, P. and Nurarak, P. (2017) Automated Cyberbullying Detection Using Clustering Appearance Patterns. 2017 9*th International Conference on Knowledge and Smart Technology* (*KST*), Chonburi, 1-4 February 2017, 242-247. https://doi.org/10.1109/KST.2017.7886127

[13] Noviantho, Ashianti, L. and Isa, S.M. (2017) Cyberbullying Classification Using Text Mining. 2017 1*st International Conference on Informatics and Computational Sciences* (*ICICoS*), Semarang, 15-16 November 2017, 241-246. https://doi.org/10.1109/ICICOS.2017.8276369

[14] Chamoun, M., Serhrouchni, A. and Haidar, B. (2017) A Multilingual System for Cyberbullying Detection: Arabic Content Detection Using Machine Learning. *Advances in Science, Technology and Engineering Systems Journal*, **2**, 275-284. https://doi.org/10.25046/aj020634

[15] Zhang, X., Tong, J., Vishwamitra, N., Dillon, E., Macbeth, J., Hu, H.X., Whittaker, E., Mazer, J.P. and Kowalski, R. (2016) Cyberbullying Detection with a Pronunciation-Based Convolutional Neural Network. 2016 15*th IEEE International Conference on Machine Learning and Applications* (*ICMLA*), Anaheim, 18-20 December 2016, 740-745. https://doi.org/10.1109/ICMLA.2016.0132

[16] De Jong, F. and Dadvar, M. (2012) Cyberbullying Detection: A Step toward a Safer Internet Yard. *Proceedings of the* 21*st International Conference on World Wide Web*, Lyon, 16-20 April 2012, 121-126.

[17] Dadvar, M., Trieschnigg, D., De Jong, F. and Ordelman, R. (2012) Improved Cyberbullying Detection Using Gender Information. *Proceedings of the* 12*th Dutch-Belgian Information Retrieval Workshop* (*DIR* 2012), Ghent, 24 February 2012, 23-25.

[18] Dadvar, M., Trieschnigg, D., De Jong, F. and Ordelman, R. (2013) Improving Cyberbullying Detection with User Context. In: *European Conference on Information Retrieval*, Springer, Berlin, 693-696. https://doi.org/10.1007/978-3-642-36973-5_62

[19] De Jong, F. and Dadvar, M. (2014) Experts and Machines against Bullies: A Hybrid Approach to Detecting Cyberbullies. In: *Canadian Conference on Artificial Intelligence*, Springer, Berlin, 275-281. https://doi.org/10.1007/978-3-319-06483-3_25

[20] Ahmed, M.T., Rahman, M., Nur, S., Islam, A.Z.M.T. and Das, D. (2022) Natural Language Processing and Machine Learning Based Cyberbullying Detection for Bangla and Romanized Bangla Texts. *TELKOMNIKA Telecommunication Computing Electronics and Control*, **20**, 89-97. https://doi.org/10.12928/telkomnika.v20i1.18630

[21] Ahmed, M.T., Rahman, M., Nur, S., Islam, A., Islam, M.T. and Das, D. (2021) Deployment of Machine Learning and Deep Learning Algorithms in Detecting Cyberbullying in Bangla and Romanized Bangla Text: A Comparative Study. 2021 *International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies* (*ICAECT*), Bhilai, 19-20 February 2021, 1-10. https://doi.org/10.1109/ICAECT49130.2021.9392608

[22] Ahmed, M.T., Rahman, M., Nur, S., Islam, A.M.T. and Das, D. (2021) Introduction of PMI-So Integrated with Predictive and Lexicon Based Features to Detect Cyberbullying in Bangle Text Using Machine Learning. *Proceedings of* 2*nd International Conference on Artificial Intelligence: Advances and Applications*, Jaipur, 27-28 March 2021, 685-697. https://doi.org/10.1007/978-981-16-6332-1_56

[23] Shil, P., Saima, U., Rahman, R.M. and Islam, M.S. (2021) An Approach for Detecting Bangla Spam Comments on Facebook. 2021 *International Conference on Electronics, Communications and Information Technology* (*ICECIT*), Khulna, 14-16

September 2021, 1-4. https://doi.org/10.1109/ICECIT54077.2021.9641358

[24] Akther, A., Acharjee, U.K., Talukder, M.A., Islam, M. and Uddin, M.A. (2023) A Robust Hybrid Machine Learning Model for Bengali Cyber Bullying Detection in Social Media. https://doi.org/10.31224/3124

[25] Priya and Gupta, S. (2022) Identification of Political Hate Speech Using Machine Learning-Based Text Toxicity Analysis. In: Tuba, M., Akashe, S. and Joshi, A., Eds., *ICT Systems and Sustainability*, Springer, Berlin, 217-236. https://doi.org/10.1007/978-981-19-5221-0_22

[26] Kompally, P., Chakkarvarthy, S., Walczak, S. and Johnson, S. (2021) MaLang: A Decentralized Deep Learning Approach for Detecting Abusive Textual Content. *Applied Science*, **11**, Article No. 8701. https://doi.org/10.3390/app11188701

[27] Wright, R.E. (1995) Logistic Regression. American Psychological Association, Washington DC, 217-244.

[28] Hossain, M.I., Rahman, M., Ahmed, T. and Touhidul Islam, A.Z.M. (2021) Forecast the Rating of Online Products from Customer Text Review Based on Machine Learning Algorithms. *International Conference on Information and Communication Technology for Sustainable Development* (*ICICT4SD*), Dhaka, 27-28 February 2021, 6-10. https://doi.org/10.1109/ICICT4SD50815.2021.9396822

[29] Xu, S., Li, Y. and Wang, Z. (2017) Bayesian Multinomial Naïve Bayes Classifier to Text Classification. In: Park, J.J., Chen, S.-C. and Choo, K.-K.R., Eds., *Advanced Multimedia and Ubiquitous Engineering*, Springer, Berlin, 347-352. https://doi.org/10.1007/978-981-10-5041-1_57

[30] Wu, Q. and Zhou, D.-X. (2006) Analysis of Support Vector Machine Classification. *Journal of Computational Analysis & Applications*, **8**, 99-119.