International Journal of Innovative Technology and Interdisciplinary Sciences
www.IJITIS.org
ISSN:2613-7305
Volume 5, Issue 4, pp. 1069-1080, 2022
DOI: https://doi.org/10.15157/IJITIS.2022.5.4.1069-1080
Received October 11, 2022; Accepted December 24, 2022

Cyberbullying Detection on Twitter Using Natural Language Processing and Machine Learning Techniques

Stephen Afrifa[a1,2] and Vijayakumar Varadarajan*[b3]

[1] Tianjin University, Tianjin, China

[2]University of Energy and Natural Resources, Sunyani, Ghana

[3]University of New South Wales, Sydney, Australia

[a]afrifastephen@tju.edu.cn; *[b]v.varadarajan@unsw.edu.au

ABSTRACT

People use social media to engage and debate themes ranging from entertainment to sports to politics and many others. The use of social media has also resulted in an increase in cyberbullying, which is occurring at an alarming pace. Many cyberbullying messages may be found in the comment sections of many social media platforms, including Twitter, YouTube, and others. Cyberbullying has the ability to cause stress and mental distress, which should be detected early and avoid being published on social media platforms. In this study, we provide a system for detecting cyberbullying messages in English using natural language processing (NLP) and machine learning approaches. On Twitter, a total of 16851 tweets were gathered. The dataset was applied to an NLP approach to find the most offensive terms associated with cyberbullying. Based on our NLP results, it was clear that cyberbullying happens and must be addressed as soon as possible. The dataset was also utilized to train the random forest (RF) and support vector machine (SVM) algorithms. Random forest surpassed support vector machine, which attained an accuracy of 90.5%, with 98.5%. With careful attention to data preparation, where missing and outlier values are dealt beforehand, the high percentage of the model is obtained. This method facilitates the analysis of the available data at the expense of the study's statistical power and ultimately the validity of its findings. Additionally, it aids in producing a significant bias in the outcomes and increases the effectiveness of the data. The Root mean square error and mean square error were used to analyse the results. In comparison to the support vector machine, the random forest earned the best error score. Our findings may be utilized by agencies and groups to educate individuals about the proper use of social media in order to avoid cyberbullying.

Keywords: Cyberbullying, Natural Language Processing, Machine Learning, Twitter, Random Forest, Support Vector Machine.

1. INTRODUCTION

Cyberbullying, also known as hate speech, cyberaggression, and toxic speech, is a serious social issue that affects young people who use the Internet today [1]. Cyberbullying can have serious consequences, including low self-esteem, anxiety, depression, hopelessness, and, in some cases, a lack of motivation to live, which can lead to the victim's death [2]. Cyberbullying may happen in a variety of ways, such as the spreading of offensive social media posts (text, picture, and video). An interactive platform that brings individuals together to share knowledge is social media [3]. Online social networks (OSNs) are websites that allow users to connect electronically and chat with one another. It must be emphasized that these social media sites have also fuelled online hatred, including cyberbullying. Global Internet and smartphone usage has grown [4], which has led to more people using social media platforms like Twitter, Facebook, YouTube, TikTok, and many more. The majority of cyberbullying victims are young people, typically educated people, who have reported having unpleasant and upsetting life situations [5]. The majority of social media users in Sub-Saharan Africa, including Ghana, Nigeria, and Zambia, use English [6]. Ghana's official language is English, and people use it to express themselves on various social media platforms. The goal of the current article is to identify cyberbullying in English by combining machine learning and natural language processing techniques. This study provides the current application of natural language processing (NLP) and machine learning (ML) techniques in detecting cyberbullying texts on social media platforms. This study's innovative framework may be utilized for several NLP applications, including sentiment analysis and the detection of cyberbullying as well as other types of bullying. The work is innovative in that it informs policymakers on how to effectively address the issue of hate speech on social media platforms. This will restore reason to our varied communities and create a peaceful atmosphere for everyone.

The remainder of the study is structured as follows: The literature that already exists and the technologies developed for effective cyberbullying detection are described in Section 2. The proposed work's methodology and pre-processing steps are described in Section 3. The experimental findings are reported in Section 4 along with performance comparisons, Section 5 analyse the obtained results and presents the most important recommendation for the study, and Section 6 concludes with contributions and perspectives for the future.

2. **RELATED WORKS**

Our study also considered existing literature that employed various machine learning and natural language processing techniques. To identify and promote positivity in comments, Chakravarthi [7] in a related work developed a multilingual dataset. The researcher also presented a revolutionary custom deep network architecture that leverages a concatenation of embedding from T5-Sentence. The F1-score for their model, which included a number of machine learning techniques, was 0.75 for English, 0.62 for Tamil, and 0.67 for Malayalam. In another related work by sing a variety of cutting-edge machine learning and deep learning models, Chakravarthi [8] introduced a multilingual hope speech dataset that promotes equality, diversity, and inclusion (EDI) in English, Tamil, Malayalam, and Kannada. Additionally, Albraikan et al. [4] provided a method for detecting cyberbullying in social networks called the Optimal Deep Learning-based Cyberbullying Detection and Classification (ODL-CDC). Pre-processing, prediction, and hyperparameter optimization are just a few of the activities that are included in their suggested ODL-CDC approach. Additionally, a thorough experimental examination of the benchmark dataset was conducted, and the outcomes were examined from several angles.

In a research, Mahbub et al. [9] examined the role of predatory approach terms in the identification of cyberbullying and suggested a method for creating a lexicon of these phrases. To create a lexicon of sexual approach terms, their study includes the investigation of chat logs from criminally charged individuals. Furthermore, Jones et al. [10] examined the identification of cyberbullying in text messages from mobile phones using an oxygen forensics toolbox. In order to identify cyberbullying detection, they outlined the data gathering process that involved forensics techniques, a corpus of suspicious actions including cyberbullying annotation from mobile phones, and a series of binary classification trials. Every piece of evidence helps the forensics investigation; therefore, they applied forensics techniques, machine learning (ML), and deep learning (DL) algorithms to identify suspect patterns. Machine learning and deep learning algorithms were assessed by Bharti et al. [11] to automatically identify cyberbullying from tweets. The GloVe840 word embedding approach and BLSTM, according to their findings, produced the best results on the dataset, with accuracy, precision, and F1 measures of 92.60%, 96.60%, and 94.20%, respectively. Last but not least, Sainju et al. [12] used a mixed-method approach in their study, analysing xenophobic bullying disclosures on Twitter after COVID-19 spread using both Big Data tools and qualitative analysis. According to their findings, over half of the sample was comprised of xenophobic bullying. The qualitative research also revealed that 64% of tweets about xenophobic bullying included instances that supported racial stereotypes.

3. METHODOLOGY

This section discusses a number of machine learning techniques used. Additionally, we go over our framework's suggested technique and the various pre-processing methods used in this study are outlined.

3.1 The Framework of the Study

In this article, we developed a framework for detecting cyberbullying posts on Twitter using natural language processing (NLP) and machine learning approaches. The data was pre-processed to reduce noise and produce clean data for model training. Additionally, the machine learning classifiers used are assessed using a variety of assessment measures. The Figure 1 represents the designed framework of the study. The framework is explained in depth in the next sections of this paper.

3.2 Data Collection

The Twitter data was acquired through the use of the snscrape Python package. The hashtags #shit, #flooding, #fuck, and #idiot was used to filter all of the tweets published between August 01, 2018, and September 30, 2022. Based on our search parameters, 16851 tweets were collected in total.

3.3. Data Pre-processing

Text categorization algorithms are unable to interpret raw data because they are unable to directly comprehend high-level human language. The pre-processing methods listed below were modified to assist our framework and model comprehend and learn.

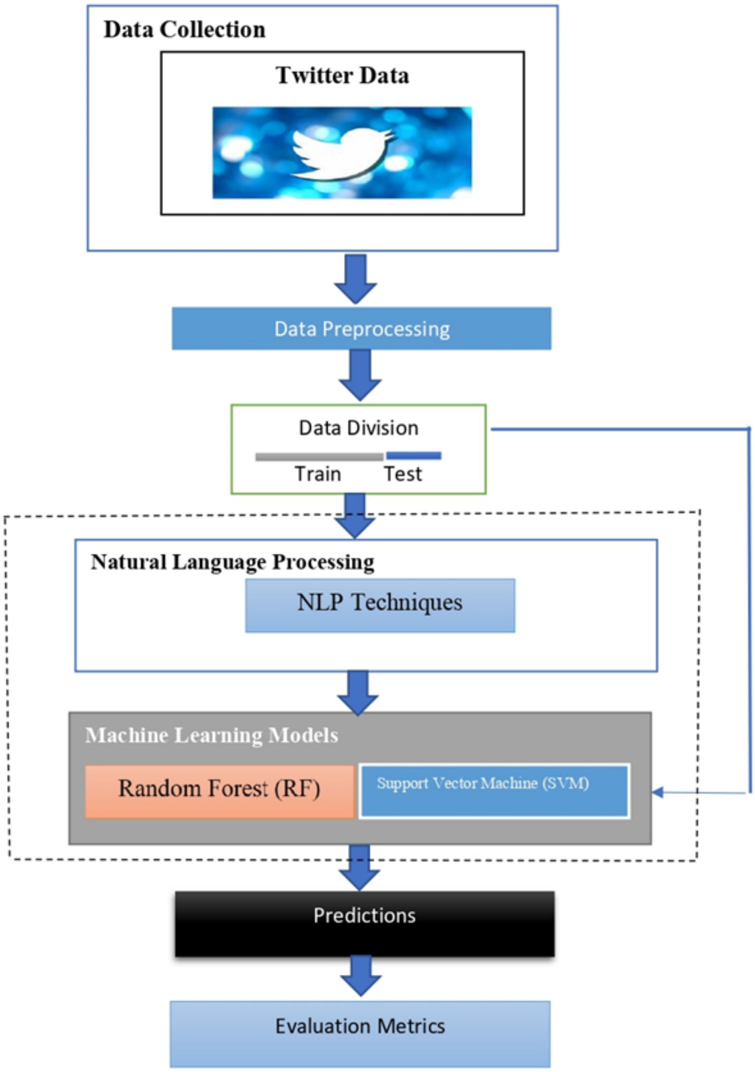

Figure 1. The proposed framework of the study

It has been decreased the text noise to enhance the performance of the classifier and speed up the classification procedure. This made machine learning and real-time natural language processing (NLP) possible. Unhelpful and distracting components like Hypertext Markup Language (HTML) tags, scripts, URL links, and ads are often used in online writing [13]. Some of the methods used to clean up online writings (tweets) include making all letters lowercase, deleting punctuation, numbers, stop words, links (HTML and URL), stemming, reducing white space, and feature selection. The equation 1 explains the pre-processed stage in the example below:

$$Y_{ij} = TX_{ik} \tag{1}$$

such that: (i) Y_{ij} preserves the "valuable information" in X_{ik}, (ii) Y_{ij} eliminates at least one of the problems in $X_{ik}$and (iii) Y_{ij} is more useful than X_{ik}. In the above relation, i = 1, . . . n where n = number of objects, j = 1, . . . m where m = number of features after preprocessing, k = 1, . . . l where l = number of attributes/features before preprocessing, and in general, m ≠ l. Additionally, we employ the Term Frequency – Inverse Document Frequency method to select the desired features. The equation 2 below shows the feature selection:

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{FF} * \log(\mathrm{N}|\mathrm{DF}) \tag{2}$$

where N indicates the number of documents, and DF is the number of documents that contain the feature. FP takes the value 0 or 1 based on the feature absent/presence in the document.

Data annotation is the act of classifying, labelling, and including additional contextual information to a raw data collection so that computers can understand and utilize the data [14]. In this work, we focused on text annotation, adding labels and instructions to the tweets, to better understand how typical human sentences and other textual data are ordered for meaning and to assist our ML algorithm perceive and understand it. Machine learning (ML) can interpret texts' deeper meanings beyond their literal meanings thanks to text annotation [15]. The labelling procedure attempts to classify tweets by determining polarity. There are two columns in the dataset (tweets and results). The computation of the tweets' polarity is shown in equation 3 below.

$$\mathrm{f_m} = \begin{cases} \mathrm{f(posScore)}, & \text{if } \mathrm{f(posScore)} \leq \mathrm{f(negScore)} \\ -\mathrm{f(negScore)}, & \text{otherwise} \end{cases} \tag{3}$$

where f_m computes the absolute maximum of the two scores. It is worth noting that f(negScore) is always positive by construction. The negative sign is imposed to obtain a final prior polarity that ranges from -1 to 1. Table 1 lists sample tweets following data preparation.

Table 1. Sample tweets after data pre-processing.

Tweet samples after pre-processing
i really do not understand your point it seems that you are mixing apples and oranges
either you are fake or extremely stupid maybe both
your a retard go post your head up your
your anti semitic rants are not welcomed here you are a racist moron fu
you are a fukin moron you are just butthurt that you got rejected on wikipedia call yoursilf scolar or whatever

3.4. The Random Forest Algorithm

The machine learning technique random forest (RF) was used to identify the polarity (results) of the tweets. Each decision tree in the RF has its variable space partitioned into a smaller subspace, resulting in data that is as uniform as is practicable throughout each zone [16, 17]. The RF fits a variety of decision trees utilizing subsamples from the complete data set to improve the classification or regression accuracy [18]. Its foundation is the premise that different separate predictors each forecast insufficiently in different situations, and that by combining the independent predictors' prediction outputs, overall prediction accuracy may be raised. Equation 4 represents the RF algorithm.

$$\hat{f} = \frac{1}{N}\sum_{i-1}^{N} f_i(x') \tag{4}$$

To predict the results for x' we can average the result of all the trees fi corresponding to x'.

3.5. Support Vector Machine

Support Vector Machine (SVM) method for solving both classification and regression problems [19, 20] It is based on the idea of restricted optimization and employs the inductive principle of structural risk minimization (SRM), which results in an overall optimal answer [21]. However, to establish a relationship between the target variable and the explanatory variables, the SVR only requires a subset of the data called support vectors.

3.6. Evaluation Metrics for the Machine Learning Models

For machine learning (ML) algorithms to assess and track the effectiveness of their predictions, evaluation metrics are essential. The Root mean square error (RMSE), mean square error (MSE), and mean absolute percentage error (MAPE) were used to assess the accuracy of the random forest (RF) and support vector (SVM) techniques. The equation 5 and equation 6 below are representation of the metrics' mathematical equation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

where n is the total number of observations, $\hat{y}$ is the predicted value and y_i is the actual value.

4. RESULTS AND DISCUSSION

In this section, we describe the datasets that were used and present the results of the experiments on both the NLP and machine learning algorithms. Three metrics are used to evaluate the results: accuracy, root mean square error (RMSE), and mean square error (MSE). This section also includes graphical representations of the results.

4.1. Data Description and Results Analysis

Our datasets were divided into 80% for training and 20% for testing. We used the 20% testing sets to evaluate performance after training the classifiers with 80% of the data. The dataset is made up of English-language Twitter comments and discussions. The dataset contains 16851 tweets, of which 5392 are cyberbullying texts (Yes) and 11459 are not (No). The Figure 2 depicts the percentage representation of the dataset, i.e., cyberbullying and non-cyberbullying texts.

Furthermore, the comments on Twitter, the microblogging website, were analysed, and word clouds were provided. The word clouds for cyberbullying and non-cyberbullying texts for this dataset are shown in Figure 3 and Figure 4.

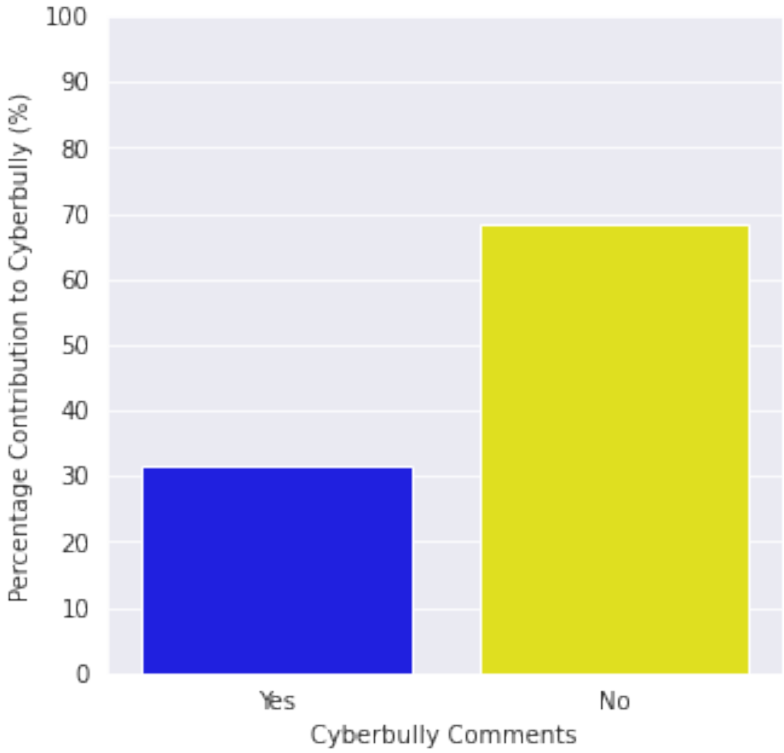

Figure 2. Percentage of cyberbullying (Yes) and non-cyberbullying (No) texts.

Figure 3. Word cloud of Twitter data with cyberbullying texts.

Figure 4. Word cloud of Twitter data with non-cyberbullying texts.

Figure 3 shows that the majority of comments contain derogatory words such as "fuck," "idiot," and "stupid." These words represent bullying, and their impact on the primary poster is enormous. These amount to stress and emotional stress on the part of the individual, particularly among young people who use social media sites. It is clear that cyberbullying on social media websites exists and must be taken seriously in order to be addressed. Furthermore, the dataset was subjected to our two machine learning classification algorithms of choice, namely, random forest (RF) and support vector machine (SVM). The outcomes of the two machine learning algorithms were compared in order to determine the best algorithm to train on our model. The random forest (RF) achieved 98.5% accuracy, while the support vector machine (SVM) achieved 90.5% accuracy. The machine learning models were then assessed using two different evaluation metrics: root mean square error and mean square error. The performance of the models is summarized in Table 2.

Table 2. Summary of model performance

Algorithm	Accuracy (%)	RMSE	MSE
RF	98.5	0.2588	0.0670
SVM	90.5	0.2644	0.0699

As can be seen, random forest achieved the highest accuracy while having the lowest RMSE (table 2). Based on our findings, the random forest algorithm performed the best when it came to Twitter data classification. The Figure 5 below depicts the pictorial view of the evaluation metrics of the RF and SVM models to facilitate more understanding of the proposed approach in this study.

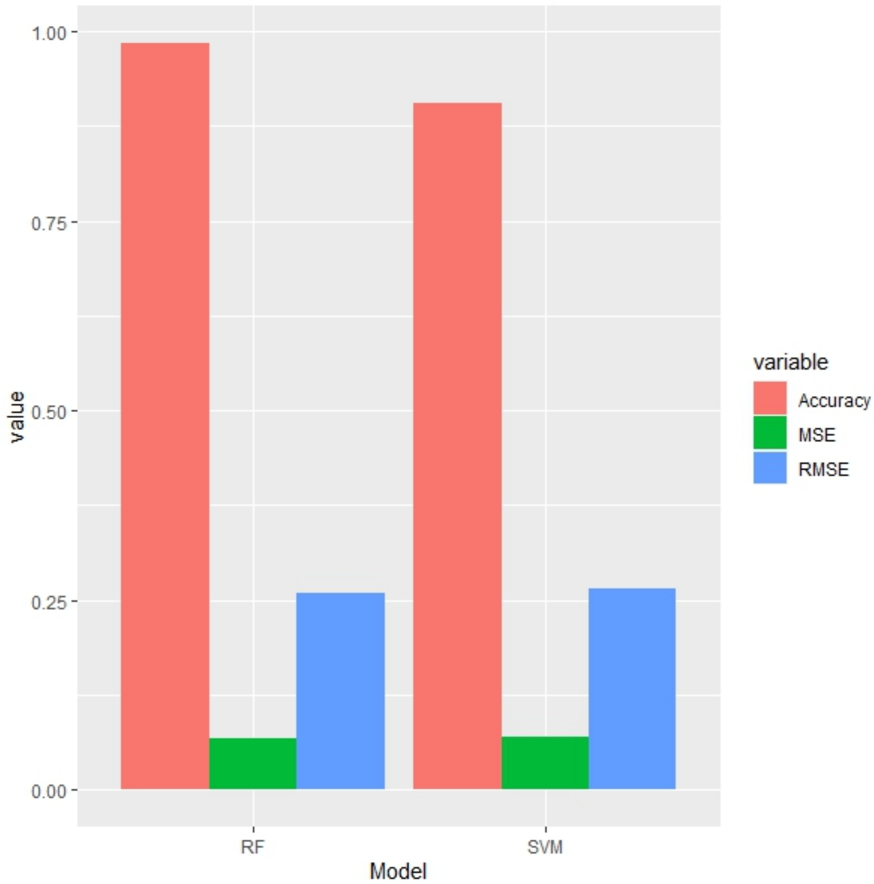

Figure 5. Pictorial view of the evaluation metrics of the models.

5. ANALYSING

The rapid evolution of technology is having a huge influence on the world. Today, a lot of individuals share their ideas on laws, entertainment, taxation, and other topics on social networking sites like Facebook and Twitter. It is important to note that although young people utilize these platforms primarily, certain comments published on them are offensive and so constitute cyberbullying. Cyberbullying has repercussions, and one of those repercussions is the victim committing suicide. This study offers a framework to identify cyberbullying texts using machine learning (ML) and natural language processing (NLP) methods. The dataset used to train the algorithm contained messages that our approach was able to identify as cyberbullying. Our findings demonstrated the existence of cyberbullying on social media platforms, which calls for immediate attention. Additionally, the ML methods produced the best outcomes on the dataset. Support vector machine (SVM) accuracy was 90.5%, while random forest (RF) accuracy was 98.5%. The RMSE and MSE evaluation metrics were used to evaluate these machine learning approaches. Machine learning evaluation metrics make it possible to determine how well a model fits a dataset. The study's best machine learning (ML) method was the random forest (RF) algorithm, which had the lowest error scores. To demonstrate the reliability of our study, the outcomes of our study were also compared to data from other current literature. Our findings outperformed those of the published papers. Table 3 compares our results with those from previous research that have been published.

Table 3. Comparison between obtained results with existing published studies

Published studies	Year of publication	Method and results of the published study	Our approach and obtained results
Raj et al. [22]	2021	Support Vector Machine (89.4%)	**Support Vector Machine (90.5%)**
Chakravarthi [7]	2022	Support Vector Machine (90.1%)	
Ahmed et al. [23]	2022	Support Vector Machine (76.5%)	
Al-Marghilani [24]	2022	Support Vector Machine (78.05%)	
Jones et al. [10]	2021	Support Vector Machine (82%)	
Jones et al. [10]	2021	Random Forest (87%)	**Random Forest (98.5%)**
Alhashmi and Darem [25]	2022	Random Forest (87.24%)	
Rahayu et al. [26]	2022	Random Forest (88%)	
Bharti et al. [11]	2021	Random Forest (91.35%)	
(Gomez et al. [27]	2022	Random Forest (96%)	

It is clear from table 3 above that our proposed model outperformed all published works that were compared against. The results are intriguing since the research that were examined are recent articles that were released in 2021 and 2022. Last but not least, our suggested model uses support vector machines (SVM) and random forest (RF) to deliver the most recent and effective results for identifying cyberbullying texts. This demonstrates the robustness of our study.

6. CONCLUSION AND FUTURE WORKS

People are interacting with each other more frequently as the number of social media sites grows, which has led to an increase in cyberbullying. We developed a framework based on natural language processing (NLP) and machine learning techniques to detect cyberbullying texts. Twitter, the microblogging website, yielded a total of 16851 tweets. Our models were trained using the dataset. The dataset included 5392 texts about cyberbullying and 11459 texts about non-cyberbullying. Unpleasant words such as "stupid," "idiot," and "fuck" were seen in cyberbullying tests, according to our word cloud. These texts are unpleasant and can cause stress and emotional distress in posters, which must be addressed. Cyberbullying education must be implemented effectively by institutions, agencies, non-governmental organizations (NGOs), and governmental bodies. After dividing the dataset into training and testing sets, the random forest (RF) performed best with an accuracy of 98.5% and an RMSE of 0.2588, while the support vector machine (SVM) achieved an accuracy of 90.5% and an RMSE of 02644. The trained models can detect cyberbullying texts on social media platforms such as Twitter, Facebook, TikTok, and YouTube. These modes can also be used to quantitatively predict the influence of the text using the word cloud generated by NLP. Individuals and organizations can use the model's analysis to provide education on cyberbullying and its impact on the primary poster, particularly young people who use social media sites. In the future, we hope to collect as much information as possible from various social media platforms such as Facebook, TikTok, and YouTube. Deep learning algorithms will be used to train the dataset in order to generate decision-making systems. Furthermore, high performance quantum machine learning models will be utilized to support various machine learning algorithms such as artificial neural networks and ensemble models.

ACKNOWLEDGEMENT

The authors are grateful to Adwoa Afriyie for her encouragement and advise provided throughout the study.

CONFLICT OF INTERESTS

The authors would like to confirm that there is no conflict of interests associated with this publication and there is no financial fund for this work that can affect the research outcomes.

REFERENCES

[1] Azeez A.N., Misra S., OLawal I.O. and Oluranti J. Identification and Detection of Cyberbullying on Facebook Using Machine Learning Algorithms, *J. Cases Inf. Technol.*, 2021; 23(4); 1-21.

[2] Ali W.H.N.W., Fauzi F. and Mohd M. Identification of Profane Words in Cyberbullying Incidents within Social Networks, *J. Inf. Sci. Theory Pract.*, 2021;

9(1); 24-34.

[3] Mangaonkar A., Pawar R., Chowdhury S.N., and Raje R.R. Enhancing collaborative detection of cyberbullying behavior in Twitter data, *Cluster Comput.*, 2022; 25(2); 1263-1277.

[4] Albraikan A.A., Hassine S.B.H., Fati S.M., Al-Wesabi F.N., Hilal A.M., Motwakel A., Hamza M.A., Al Duhayyim M. "Optimal Deep Learning-based Cyberattack Detection and Classification Technique on Social Networks," *Comput. Mater. Contin.*, 2022; 72(1); 907-923.

[5] T. Ahmed, S. Ivan, M. Kabir, H. Mahmud, and K. Hasan, "Performance analysis of transformer-based architectures and their ensembles to detect trait-based cyberbullying," *Soc. Netw. Anal. Min.*, 2022; 12(1); 1-10.

[6] Graphic.com.gh, "Literacy rate now 69.8 per cent - Graphic Online,". Available at https://www.graphic.com.gh/news/general-news/literacy-rate-now-69-8-per-cent.html?msclkid=54c9645fd0b711ec81576ad7c0ee2fdd. Accessed on 11 May 2022.

[7] Chakravarthi R.B. Hope speech detection in YouTube comments, *Soc. Netw. Anal. Min.*, 2022; 12(1); 1-19.

[8] Chakravarthi R.B. Multilingual hope speech detection in English and Dravidian languages, *Int. J. Data Sci. Anal.*, 2022; 14(4); 389-406.

[9] Mahbub S., Pardede E. and Kayes M.S.A. Detection of Harassment Type of Cyberbullying: A Dictionary of Approach Words and Its Impact," *Secur. Commun. Networks*, 2021; 2021; 1-12.

[10] Jones M.G., Winster G.S. and Valarmathie P. Integrated Approach to Detect Cyberbullying Text: Mobile Device Forensics Data, *Comput. Syst. Sci. Eng.*, 2021; 40(3); 963-978.

[11] Bharti S., Yadav K.A., Kumar M. and Yadav D. Cyberbullying detection from tweets using deep learning, *Kybernetes*, 2021; 51(9); 2695–2711.

[12] Sainju D.K., Zaidi H., Mishra N. and Kuffour A. Xenophobic Bullying and COVID-19: An Exploration Using Big Data and Qualitative Analysis. *Int. J. Environ. Res. Public Health*, 2022; 19(8); 4824.

[13] Deepak Chowdary E., Venkatramaphanikumar S. and Venkata Krishna Kishore K. "Aspect-level sentiment analysis on goods and services tax tweets with dropout DNN," *Int. J. Bus. Inf. Syst.*, 2020; 35(2); 239-264.

[14] Bazzan C.L.A., Engel M.P., Schroeder F.L. and Da Silva C.S. Automated annotation of keywords for proteins related to mycoplasmataceae using machine learning techniques, *Bioinformatics*, 2002; 18(2); 35-43.

[15] Montoyo A., Martínez-Barco P. and Balahur A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, *Decis. Support Syst.*, 2012; 53(4); 675-679.

[16] Grandgirard J., Poinsot D., Krespi L., Nénon, P.J. and Cortesero M.A. Breiman and Cutler's Random Forests for Classification andRegression, *Entomol. Exp. Appl.*, 2002; 103(3); 239-248.

[17] Afiza N., Razali M., Malizan A.N., Hasbullah A.N. and Wook M. Opinion mining for national security : techniques , domain applications , challenges and

research opportunities, *J. Big Data*, 2021; 8; 1-46.

[18] Afrifa S., Zhang T., Appiahene P. and Vijayakumar V. Mathematical and Machine Learning Models for Groundwater Level Changes : A Systematic Review and Bibliographic Analysis, *Futur. Internet*, 2022; 14(9); 259.

[19] Schug F., Okujeni A., Hauer J., Hostert P., Nielsen J. and van der Linden S. “Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series,” *Remote Sens. Environ.*, 2018; 210; 217-228.

[20] Appiahene P., Missah M.Y. and Najim U. Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks, *Adv. Fuzzy Syst.*, 2020; 2020; 1-12.

[21] Aghelpour P., Mohammadi B. and Biazar M.S. Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA, *Theor. Appl. Climatol.*, 2019; 138(3-4); 1471-1480.

[22] Raj C., Agarwal A., Bharathy G., Narayan B. and Prasad M. Cyberbullying detection: Hybrid models based on machine learning and natural language processing techniques, *Electron.*, 2021; 10(22); 1-20.

[23] Ahmed T.M., Rahman M., Nur S., Islam T.M.Z.A. and Das D. “Natural language processing and machine learning based cyberbullying detection for Bangla and Romanized Bangla texts,” *Telkomnika (Telecommunication Comput. Electron. Control.*, 2022; 20(1); 89-97.

[24] Al-Marghilani A. Artificial Intelligence-Enabled Cyberbullying-Free Online Social Networks in Smart Cities. *Int. J. Comput. Intell. Syst.*, 2022; 15(1); 1-13.

[25] Alhashmi A.A. and Darem A.A. Consensus-based ensemble model for arabic cyberbullying detection, *Comput. Syst. Sci. Eng.*, 2022; 41(1); 241–254.

[26] Rahayu C., Lucky H. and Suhartono D. Cyberbullying Detection Using Word Embedding Fast Text, *ICIC Express Lett. Part B Appl.*, 2022; 13(6); 655–662.

[27] Gomez E.C., Sztainberg O.M. and Trana E.R. Curating Cyberbullying Datasets: a Human-AI Collaborative Approach. *Int. J. Bullying Prev.*, 2022; 4(1); 35–46.