Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PERFORMANCE ANALYSIS OF SUPERVISED MACHINE LEARNING TECHNIQUES FOR CYBERSTALKING DETECTION IN SOCIAL MEDIA

ARVIND KUMAR GAUTAM[1], ABHISHEK BANSAL[2]

[1]Research Scholar, Department of Computer Science, IGNTU, Amarkantak, Madhya Pradesh, India

[2]Assistant Professor, Department of Computer Science, IGNTU, Amarkantak, Madhya Pradesh, India,

E-mail: [1]analyst.igntu@gmail.com, [2]abhishek.bansal@igntu.ac.in

ABSTRACT

In the modern days of life, people use many social media sites for information sharing among friends, relatives, and others for personal, business, and official purposes. The use of social media platforms is also raising serious issues in the form of cyberstalking. Cyberstalking has been identified as a growing anti-social problem that affects educational institutions, victims, and entire human society. An intelligent system is required to detect cyberstalking in social media. In this paper, we proposed a cyberstalking detection model and analyzed the performance of six popular supervised machine learning algorithms, namely Logistic Regression, Support Vector Machines (SVM), Random Forest, Decision Trees, K-Nearest Neighbor, and Naive Bayes. These machine learning algorithms were implemented with two feature extraction methods, Bag of Words and TF-IDF, on two datasets of different sizes and distribution containing 35734 and 70019 comments and tweets, respectively. Performance of algorithms was measured in terms of Accuracy, Precision, Recall, f-score, training time, and prediction time. Our experimental results show that Logistic Regression and Support Vector Machine were top performer algorithms for both datasets with both feature extraction methods. Logistic Regression (92.6% with BOW and 92% with TF-IDF) and Support Vector Machine (92.5% with TF-IDF and 91.9% with BOW) achieved the highest accuracy on dataset-1. Logistic Regression and Support Vector Machine also achieved the highest Precision (96.4% and 96.6% respectively) and F-Score (94.3% and 93.8% respectively), while Naïve Bayes provides the best Recall (97.6% with TF-IDF on dataset-1) for both datasets.

Keywords: *Cyberstalking Detection, Machine Learning, Features Extraction, Bag of Words, TF-IDF, Performance Metrics.*

1. INTRODUCTION

In the era of the internet world, social media applications and email technology are widely used for professional and personal communications. Nowadays, most people spend time on online social media sites like Facebook, Twitter, WhatsApp, Pinterest, telegram, etc. Therefore, many cyber attackers are active on these platforms. The use of social media platforms is also raising serious issues in the form of cyberstalking and cyberbullying. Cyberstalking [1] is a serious cyber attack in which the attacker uses digital media to harass the victim or group through personal attacks and the disclosure of false or confidential information among other persons. It may categorize as email-stalking, internet-stalking, and computer-stalking [2]. Email stalkers may send threatening and hateful messages through email. These messages may also contain spam or viruses. Internet stalkers are active on global platforms like social media apps to harass or trolling other people. Computer stalker takes unauthorized control of another computer and harassing to others without disclosing his identity. The purpose of all these types of stalkers is to harass or threaten the victim. Cyberstalking victims suffer measurable adverse effects equivalent to survivors of traumas such as sexual assaults or bombing [3]. 90% of victims of cyber-stalking are women. Therefore, they are afraid to register the cases, especially in India, due to fearing society. As per the report of BBC [4], the first case was registered in India in 2009. Cyberstalking is a unique and global cybercrime and is responsible for creating the virtual fear world. The effect of cyberstalking on various social media platforms cannot be ignored, and for this serious attention is required to control cyberstalking. According to [5] https://www.statista.com/, figure 1 shows the number of active users in different social media applications.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

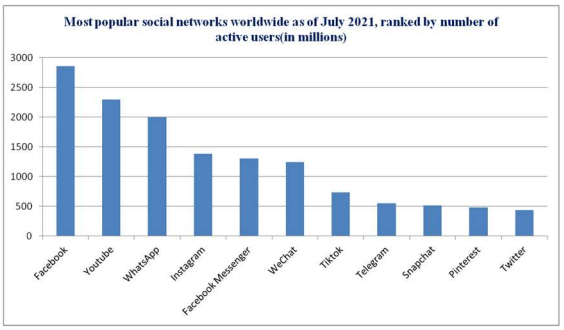

Figure 1: Most popular Social Media worldwide as of July 2021, ranked by number of active

Effective cyberstalking detection, controlling, and prevention solutions are required to tackle such type of challenging cyberstalking situation. Researchers use various supervised, semi-supervised, unsupervised machine learning algorithms and other approaches to detect and control cyberstalking. Machine learning algorithms are the most popular choice for researchers to design the model for detecting cyberstalking and other cyberharassment because machine learning techniques can classify the data and provide the result with better accuracy.

This paper focuses on reviewing the supervised machine learning algorithms used in detecting cyberstalking, cyberbullying, and other cyberharassment and analyzing the performance of classifiers using different parameters, namely accuracy, Precision, Recall, F-Score, and time complexity as performance metrics. The main contributions of this paper are-

1. We reviewed some selected quality papers between the years 2015 to 2020 to determine the popular supervised machine learning algorithms widely used in detecting cyberstalking, cyberbullying, and other cyberharassment with high performance in social media other internet applications.
2. We proposed a cyberstalking detection model using machine learning algorithms with traditional feature extraction methods, namely Bag of Words (BOW) and TF-IDF, to analyze the performance of algorithms.
3. The proposed detection model was implemented with six popular supervised machine learning algorithms on different sizes and distribution of datasets.
4. Performance analysis of algorithms in this paper will help understand the limitations and advantages of machine learning algorithms and find the significant factors such as size and distribution of datasets, pre-processing tasks, and feature extraction methods that enhance the performance of classifiers for developing cyberstalking detection models.

The proposed approach was evaluated on two different sizes of a dataset (containing Twitter tweets, Facebook comments, etc.) collected from Kaggle and other online sources. Experiments for the proposed approach were done on both datasets, and performance was measured for each machine learning classifier. Finally, we compared the outcomes of our proposed method with previous related work. The rest of the paper is organized as follows. Section 2 shows some essential background related to cyberstalking and machine learning algorithms used in this paper; Section 3 shows a literature review for various related work. Section 4 describes the proposed methodology. Section 5 shows the experimental work and results for performance analysis and comparison of outcomes with related work. Finally, Section 6 concludes the paper.

2. BACKGROUND

2.1 Cyberbullying and Cyberstalking

Cyberstalking and Cyberbullying are often used interchangeably and involve using the internet to stalk or target someone in the online world. Cyberstalking and Cyberbullying both use the same technology and target to harass internet users. Cyberbullying mainly focuses on teenagers, while cyberstalking targets other groups of users in the internet world for online harassment. Cyberstalkers regularly use web data sets, social media, and other internet-based technology to follow, bully and undermine others. Cyberstalking is a serious and complicated cybercrime that affects and targets many persons and institutions [6]. Cyberstalking, a growing global issue, is often underestimated by the public, researchers, and government. Cyberstalking is systematic, repeated, and numerous cyber-attacks and does not occur on a single occurrence [7]. As per the survey, almost twenty percent of people have faced cyberstalking situations during the use of internet applications [8]. As per available evidence [9], cyberstalking cases will regularly increase in an unexpected way. There are many examples of cyberstalking, like making and posting a real or fake sexual image of the victim to their loved ones, uploading personal information on public websites and Twitter, and hacking the victim's social media account [10]. Social media platforms are a potential hunting ground for cyber-stalker. Several types of research have shown that cyber-stalkers suffer from

social and psychological conditions also. Therefore, they use the different types of cyberstalking to target the victim. They are –

1. **Trolling and flaming**: posting rude and angry messages on social media
2. **Excluding**: Remove the victim from any social media network.
3. **Masquerading**: Creating fake profiles in social media to spoil the reputation and personality of the victim.
4. **Mobbing**: Sending repeated messages regularly by a gang of stalkers with the same goal and agenda to target the victim and victim relative.
5. **Denigrating**: Posting or sending some malicious and uncomfortable data to others for debasing the victim in others' view.
6. **Outing**: Sharing and posting the victim's personal information with others without the victim's permission.
7. **Harassing**: regularly sending unnecessary messages to the victim.

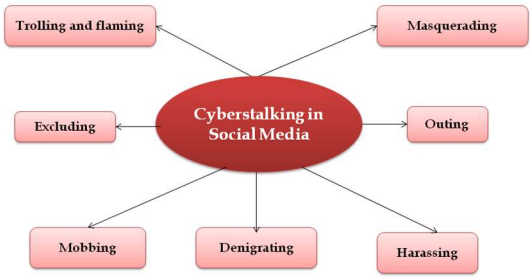

Figure 2: Cyberstalking to target the victim

2.1 Machine Learning

Machine learning (ML) is the most popular application of artificial intelligence (AI), which has the capacity for automatic learning and provides accurate and progressive results from experiences [11].

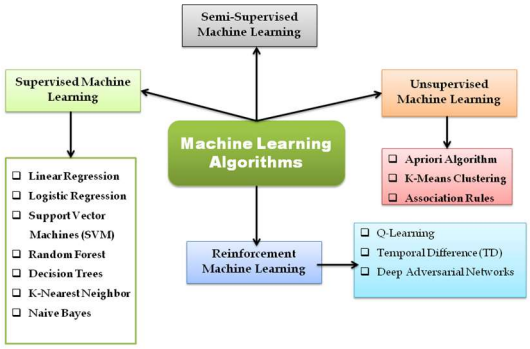

Figure 3: Example of Machine Learning Algorithms

Machine learning utilizes the existing calculations and classification techniques with datasets and development programs to give a satisfactory answer for the issue and use them to learn all alone. Machine learning gives more precise outcomes more quickly by examining vast measures of information. Machine Learning algorithms can be classified into several types [12], namely Supervised Machine Learning Algorithms, Unsupervised Machine Learning Algorithms, Semi-supervised Machine Learning Algorithms, and Reinforcement Machine Learning Algorithms.

2.1.1 Supervised machine learning algorithms

Supervised machine learning techniques use classification tasks to classify the data into labeled data. These algorithms consist of the dependent variable, which is to be predicted from a predefined set of independent variables. Using a set of dependent and independent variables, we use a function to map the inputs to desired outputs as a training process until the desired accuracy is not achieved on the training data. Such types of machine learning are mainly used for regression and classification problems [12]. The proposed detection model has implemented the following six popular supervised machine learning algorithms to analyze the performance.

1) Logistic Regression

Logistic Regression is well known Machine Learning algorithm which makes the separate hyper-plane between two datasets by using the logistic function [13]. This algorithm uses the features as inputs and provides the result based on the probability of an appropriate class for the input.

2) K-Nearest Neighbor

This algorithm is one of the simplest, non-parametric, and lazy learning that works based on instance learning and is mainly used for multi-class problems [13]. In this algorithm, a new sample is classified using the distance from its neighbor. Further, K-nearest neighbors are found from the training dataset and objects placed into the most frequent class among neighbors.

3) Support Vector Machines (SVM)

This is the most popular algorithm among researchers which can be used for classification and regression purposes [14]. It can differentiate the classes individually in n-dimensional and provide a more accurate prediction than other machine learning classifiers.

4) Decision Trees

This algorithm provides help to take and represent the decision, and it can be used for both regression and classification purposes [15]. In the Decision Tree algorithm, node and leaf as tree structures are used. The internal node indicates the condition, while the

leaf node is used for representing the decision.

5) Random Forest

This algorithm is an enhanced form of decision tree algorithm which contains multiple decision tree classifiers [16]. In this algorithm, every individual tree indicates the prediction class while maximum numbers of the prediction class represent the final result. This classifier uses the various decision trees as merged trees.

6) Naive Bayes

The functioning of this machine learning algorithm is based on Bayes Theorem and predicts the result using the probability of any object [17]. Naïve Bayes algorithm is a very efficient technique to solve the problem of binary and multi-class classification.

2.1.2 Unsupervised machine learning

In such machine learning algorithms, any target/outcome or dependent variables are not used to predict. In such algorithms, computers are trained using the unlabeled data and primarily used for clustering the data into different groups. Descriptive modeling and Pattern detection are the main application of such types of algorithms [11].

2.1.3 Semi-supervised machine learning

Semi-supervised learning falls between supervised and unsupervised learning, and it may be used labeled and non-labeled data as per problems situation. Thus, without even a trace of marks in most perceptions yet present in hardly any, semi-supervised learning techniques are the best possibility for the model structure. These techniques exploit the possibility that even though the gathering participations of the unlabeled information are obscure, this information conveys important data about the gathering boundaries [12].

2.1.4 Reinforcement machine learning

In such types of algorithms, a trained machine model is used to make specific decisions. The machine is exposed to an environment, and then it trains itself continually using trial and error factors. The machine learns from experience and captures the best possible knowledge to make accurate business decisions [12].

3. REVIEW OF LITERATURE

The literature study will focus on cyberstalking and cyberbullying detection across various social media applications using machine learning techniques performed by many researchers. Many researchers conducted text-based cyberstalking detection, while few researchers focused the study on multimedia content for cyberstalking detection.

In 2015, Ghasem Z. et al. [18] presented machine learning solutions for controlling cyber-bullying and cyber-stalking. This approach is mainly focused on automatic detection and evidence documentation of email-based cyber-stalking. Authors used machine learning, text mining, statistical analysis, and email forensics to detect and mitigate email-based cyberstalking. Several feature selection methods such as Chi-Square (chi), Information Gain (IG), Odd Ratio (OR), Mutual Information (MI), Deviation from Poisson Distribution (PDM), Class Discriminating Measure (CDM), and Gini index (GI) were used by the authors. The authors experimented using Support Vector Machine and Neural network techniques in 5172 email datasets containing spam and genuine email. Nandhini et al. [19] have proposed a framework using Naive Bayes machine learning technique on myspace.com dataset and claimed that they achieved 91% accuracy. Chavan et al. [20] also performed the experimental work on a dataset from Kaggle for their proposed approach using Logistic Regression and Support Vector Machine classifier. Authors claimed that their proposed model achieved 73.76% accuracy using Logistic Regression while 77.65% accuracy was achieved using the Support Vector Machine. S.Vidhya et al. [21] have proposed a framework for Feature Extraction for Document Classification. The authors used a term frequency (TF) with a stemmer-based feature extraction algorithm, and the performance of the approach was tested using various classifiers. The authors claimed that the proposed method produced a better result than other methods. In 2016, Frommholz I. et al. [22] proposed a detection framework, "Anti cyberstalking Text-based System (ACTS)" for Textual Analysis and Cyberstalking Detection using machine learning algorithms which were mainly focused on author identification, text classification, personalization, and digital text forensics.

In 2017, Ganesan et al. [23] proposed an approach to analyzing the cybercrime data from the web pages database by using the unpredicted patterns. Authors claimed that this model would be capable of categorizing the cybercrime offenses as violent or non-violent and also able to categorize the types of cybercrimes such as cyber terrorism, cyberstalking, cyber fraud, and cyber theft. Romsaiyud et al. [24] have proposed an enhanced framework using Naive Bayes machine learning and achieved 95.79% accuracy. For evaluation work, the authors used multiple datasets from MySpace, Slashdot, and Kongregate. Lsa et al. [25] proposed another approach using the Support Vector Machine (SVM) and Naive Bayes classifier. The authors evaluated their experimental work on a dataset from Kaggle and

claimed that SVM produces 97.11% accuracy while 92.81% accuracy was achieved using the Naive Bayes classifier. In 2018, Hitesh Kumar et al. [26] proposed a framework using Natural Language Processing and Machine Learning techniques to detect insulting and offensive comments on different social media networks. The authors used support vector machine, Logistic Regression, Random Forest, and Gradient Boost machine learning algorithm to implement the model and found good results.

In 2019, Amanpreet Singh et al. [27] have reviewed and compared previous research works related to machine learning techniques, pre-processing methods, and the performance of machine learning algorithms. The authors discussed the methodology, datasets, and findings of various previous research works and found that Most researchers used support Vector Machine (SVM) algorithms for cyberbullying and cyberstalking detection. J.I. Sheeba et al. [28] proposed a Bystander Intervention Model using a random forest classifier for Identification and Classification of Cyberbullying. The authors used the Latent Semantic Analysis and Random Forest Classifier to identify and categorize cyberbullying into various subcategories for experimental work. John Hani et al. [29] proposed a model for cyberbullying detection in social media. The authors used Neural Networks and SVM classification models to detect and prevent cyberbullying in social media. For experimental work, authors use the sentiment analysis and TFIDF algorithms on the Kaggle dataset. Authors have utilized the various classifiers as supervised machine learning algorithms to train and detect cyberbullying and cyberstalking. The authors achieved better accuracy when they performed the experimental work on the same dataset. The authors claimed that their proposed approach achieved 92.8% and 90.3% accuracy for Neural Network and Support Vector Machine. Ravinder Ahujaa et al. [30] have proposed a model to measure the Impact of Features Extraction on the Sentiment Analysis. TF-IDF and N-Grams were used for features extraction on the SS-Tweet dataset. The proposed approach applied six classification algorithms: Decision Tree, Support Vector Machine, K-Nearest Neighbour, Random Forest, Logistic Regression, and Naive Bayes. After doing sentiment analysis on the SS-Tweet dataset, the authors found that TF-IDF features are giving better results (3-4%) than N-Gram features, while logistic Regression produced best predictions on both feature extraction methods of sentiments with maximum output for all performance metrics, namely accuracy, Recall, Precision, and f-score.

In 2020 Manowarul Islam et al. [31] has proposed a framework using a supervised machine learning approach for improving the accuracy of cyberbullying and cyberstalking detection on social media networks. The authors evaluated their proposed approach on Decision Tree, Random Forest, Naive Bayes, and Support Vector Machine classifier. Bag-of-Words (BoW) and TF-IDF (Term Frequency-Inverse Document Frequency) were used for feature extraction by the authors. The authors claimed that they had achieved better accuracy. Hoyeon Park et al. [32] have measured the Impact of Word Embedding Methods on the Performance of Sentiment Analysis with Machine Learning Techniques. The authors applied different machine learning classifiers, namely Naïve Bayes, support vector machine, random forest, gradient boosting, and XGBoost, to compare the performance of BoW, TF-IDF, and Word2Vec features extraction techniques. Authors claimed that TF-IDF provided a better result, 84.27%, than Word2Vec (79.8%). The authors also found that vector modeling in word embedding of sentiment analysis is more suitable for machine learning than sequential modeling. Amgad Muneer et al. [33] proposed a framework using a machine learning approach to improve cyberbullying detection accuracy on Twitter. The authors evaluated their proposed system using Decision Logistic Regression, LGBM Classifier, SGD Classifier, Random Forest, AdaBoost Classifier, Naive Bayes, and Support Vector Machine. For feature extraction, WordtoVec and TF-IDF (Term Frequency-Inverse Document Frequency) were used. The authors measured the performance of algorithms and achieved better accuracy.

In the literature review, several related research papers between the years 2015 to 2020 were selected to find the popular supervised machine learning techniques and contributions of previous work performed by researchers to detect cyberbullying, cyberstalking, and other cyberharassment using machine learning techniques. As per the literature review, several social media networks and other online applications such as Twitter, Facebook, Youtube, Snapchat, Instagram, and emails are often used by cyberstalkers through text and multimedia content. Mainly, cyberstalking detection framework focused on content in the English language, although researchers are also showing their interest in the content of other languages for cyberstalking detection. Based on the study and review, the following are the summarized findings that give new directions towards research.

1. Even if many researchers have worked to detect cyberbullying, cyberstalking, and other cyber harassment, more enhanced techniques are required to control the stalking downright.
2. Features such as chat, speech, profile, user conversation, participant's interaction on social media platforms, and sentiment analysis for determining the different meanings of comments will be more beneficial for detecting cyberstalking and other cyberharassment.
3. The selection of machine learning techniques should be based on the proper performance analysis according to datasets sizes and features extraction methods.
4. Finally, we can say that there is rich literature growing on machine learning techniques for developing detection models, and there is much scope for improvement in this area.

Inspired by authors at [26, 29, 33], our proposed framework was evaluated using six machine learning classifiers and measured performance.

4. PROPOSED METHODOLOGY

This proposed methodology section described our proposed cyberstalking detection framework using machine learning algorithms. The proposed machine learning framework works on textual data and consists of four main phases for cyberstalking detection: pre-processing, features extraction, text classification, and cyberstalking detection. The proposed machine learning framework is shown in figure-4.

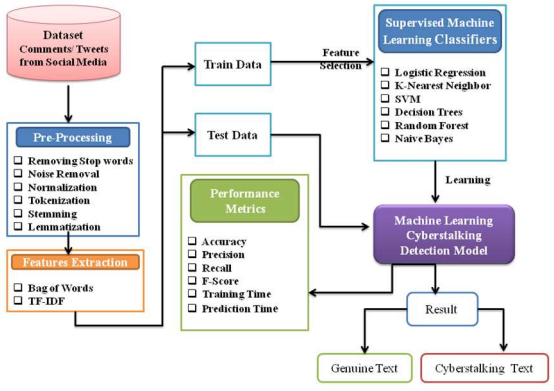

Figure 4: Proposed Machine Learning Model for Cyberstalking Detection in Social Media

In the first part, messages, tweets, comments, and posts from different social media networks were collected for making the dataset. After that, in the pre-processing phase, data was cleaned and prepared using Natural Language Processing. In the feature extraction phase, Bag of Word (BOW) and TF-IDF were applied separately with word-level. After that, the dataset was split for training and testing purposes for machine learning classifiers. Furthermore, data were classified as cyberstalking or non-cyberstalking text.

4.1 Dataset

We have collected the datasets from Kaggle and other sources and made a mixed dataset containing Twitter tweets and comments from Facebook and YouTube. Text of datasets was classified as non-cyberstalking text and cyberstalking text. Dataset-1 includes a total of 35734, while dataset-2 contains a total of 70019 unique records. Table-1,figure-5, and figure-6 show the size of datasets and distribution of rows for the training set, test set, cyberstalking text, and non-cyberstalking text.

Table 1: Size of and distribution dataset

Dataset-1
Number of unique rows in the total set: 35734
Number of rows in the training set: 26800 (74.9%)
Number of rows in the test set: 8934 (25.1%)
Number of rows for Non-Cyberstalking text: 12257 (34.3%)
Number of rows for Cyberstalking text: 23477 (65.7%)
Dataset-2
Number of unique rows in the total set: 70019
Number of rows in the training set: 52514 (74.9%)
Number of rows in the test set: 17505 (25.1%)
Number of rows for Non-Cyberstalking text: 38020 (54.3%)
Number of rows for Cyberstalking text: 31999 (45.7%)

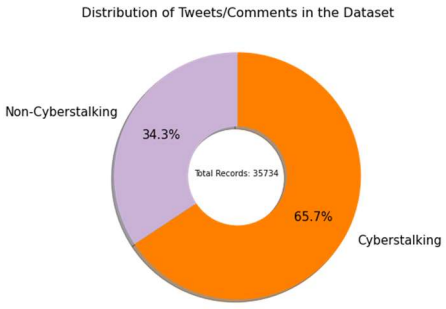

Figure 5: Dataset 1 - Total records 35734

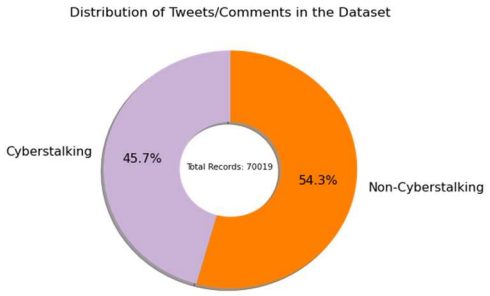

Figure 6: Dataset 2 - Total records 70019

4.2 Pre-Processing

Dataset collected from various sources of social media often contains different unnecessary characters or text. We need to clean and prepare the data for the detection phase before the evaluation of the machine learning algorithms. In this phase, using the collection of keywords, the text of the datasets is filtered and normalized to a specific format. In the pre-processing stage, normally following pre-processing tasks are performed using Natural Language Processing (NLP)

1) Noise Removal

The dataset collected from various sources contains unnecessary noise data in the form of special characters, digits, pieces of text, repeated words, punctuation marks, and white spaces [34]. These noise data are required to remove from the dataset using natural language processing.

2) Removing Stop words

The most common words in datasets such as articles, prepositions, pronouns, etc., those that do not give the meaning and do not determine syntactic, semantic, and sentiment meaning are called stop words [35]. These stop words are required to be removed using natural language processing.

3) Tokenization

Converting sentences into words is called tokenization [35]. In tokenization, the sentence is divided into separated words and added to the list.

4) Normalization

Normalization is used for uniformity of pre-processing on each text [35]. Several tasks such as converting all text to either upper or lower case and transforming numbers to their equivalent words are performed concurrently. Normalization is highly required because words with the same meaning as 'girl' and 'GIRL' will represent non-identical words in the. In this paper, all texts are converted into lower case letters.

5) Stemming

A process of transforming different tenses of words to their root form is called stemming [36]. Stemming provides necessary support to eliminate unwanted computation of words from the list. For example, 'lose', 'losing' and 'lost' will be converted into 'lose' using streaming methods.

6) Lemmatization

Lemmatization is a process of merging two or more words into a single word to reduce the words to a word existing in the language using synonyms [36]. In this step, synonyms of each word are merged into one word.

All pre-processing tasks (Noise Removal, Removing Stop words, Tokenization, Normalization, Stemming, and Lemmatization) are not always required because pre-processing tasks are also responsible for increasing or decreasing the performance of classifiers. The selection of pre-processing tasks should be according to datasets and features extraction models. After performing the pre-processing task, formatted data was sent to the next phase of the proposed framework for Feature Extraction.

4.3 Feature Extraction:

In this phase, a feature dictionary and feature vectors are prepared. The text data were converted into numbers using feature extraction because the machine learning classifier cannot understand the data as raw text. Feature extraction methods often play a crucial role in enhancing the performance of machine learning classification. Feature extraction methods compute the weights of the words in the text and then create a feature vector based on a group of predefined keywords. By using the different features, extraction approaches such as filtration, fusion, mapping, and clustering, several word-level, sentence-level, and n-gram level features extraction methods are used for feature extraction. Bag of Words, TF-IDF, Word2Vec, GloVe, FastText, ELMo, BERT, ALBERT, ELECTRA, GPT-2, XL-NET, and Roberta, SBERT, Doc2VEC, InferSent, and Universal Sentence Encoder are some popular methods for feature extraction.

The Bag of Words (BOW) is the simplest feature extraction method that represents the text into numbers. The Bag of words does not count the positioning, grammar, and structure of the words in the text. It just counts the frequencies of words in the target text and puts those words into a bag. Bag of Words uses a vocabulary of known words and measures available words' presence for features extraction. Each word count as a feature and each word is given equivalent significance [37]. Bag of words collects the data, designs a vocabulary, and finally scores the words to create the vectors. In the Bag of Words, each feature is matched with the input data. If the feature occurs in the input data, then feature frequency is represented by value 1; otherwise, 0 is used to describe the feature frequency.

TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical calculation that can measure the relevance of any word of documents in a collection of documents. In TF-IDF, the regularly occurring words should be given more significance because frequently occurring words are more valuable for the classification [38]. Term Frequency

(TF) of any term is calculated based on the number of occurrences in the document to the total words in that document, while Inverse Document Frequency (IDF) is used to determine the importance of any term in the document.

In this paper, our experimental work applied Bag of Words (BOW) and TF-IDF (Term Frequency-Inverse Document Frequency), both feature extraction methods separately with word-level in the proposed machine learning framework to analyze the performance of classifiers with both feature extraction models. Further, Feature dictionaries and feature vectors were used as input by machine learning techniques for training and testing the data classifier.

4.4 Data Classification using Machine Learning

In this phase, training and testing datasets were sent to the machine learning classifier to train and test the detection model. We used six supervised machine learning algorithms, namely Logistic Regression, Support Vector Machines (SVM), Random Forest, Decision Trees, K-Nearest Neighbor, and Naive Bayes algorithms for classification. After classification, data was sent to the Machine Learning Cyberstalking Detection model for further action.

4.5 Machine Learning Cyberstalking Detection

In this phase, the Machine Learning classifier detection modal detected the cyberstalking data. With the help of the machine learning classifier and data dictionary, data were classified into genuine posts or cyberstalking posts.

4.6 Performance Metrics

Parameters used to monitor, measure, and analysis the performance of a model during training and testing time are called performance metrics. We used a confusion matrix for obtaining the performance metrics. A confusion matrix is a table of a "N x N" matrix that contains True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) based on actual and predicted values. Where N represents the number of target classes. In the confusion matrix, columns represent the actual values of the target variable while rows represent the predicted values of the target variable, and the value of the target variable may be either Positive value (1) or Negative value (0). When the predicted value matches the actual value, and both actual and predicted value is "1" it is called True Positives (TP). When the actual value was negative (0), and the model predicted a negative (0) value, then it is called True Negative (TN). In other cases, when the actual value was negative (0) but the model predicted a positive value (1), then it is called False Positives (FP). In the case of False Negatives (FN), the actual value was positive (1), but the model predicted a negative value (0).

In this paper, based on the confusion metrics, several parameters of performance metrics such as Accuracy, Precision, F-Score, Recall, Training Time, and Prediction Time were used to measure and analyze the performance of machine learning classifiers in the cyberstalking detection model.

1. Accuracy

Accuracy is the number of correct predictions predicted by the machine learning model. We use the following formula to calculate the accuracy of a model.

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (1)$$

2. Precision

Precision measures the ratio between the True Positives and all the other Positives values, which the machine learning model predicts. Precision is measured using the following formula.

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

3. Recall

Recall shows the detection rate and measures the ratio of true positive prediction to total positive.

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

4. F-Score

F-Score indicates the harmonic average between Precision and Recall. F-Score provides combined trends about Precision (P) and Recall (R). When Precision and Recall both give equal value, then F-Score produces maximum value.

$$\text{F-Score} = (2 * P * R) / (P + R) \quad (4)$$

5. EXPERIMENTAL RESULT AND DISCUSSION

This section will discuss experimental work, results, performance metrics, and time complexity to analyze the performance of supervised machine learning algorithms in the proposed detection model. Further, the result of our detection model will be compared with the outcomes of related works. We performed the experimental work on two different sizes and distribution of datasets using six machine learning algorithms namely Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Support Vector Machines (SVM), and K-Nearest Neighbor (KNN) with both feature extraction methods BOW and TF-IDF. The experiment used python language with Scikit Learn and other machine learning library packages to implement the detection model.

Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.1 Performance of Algorithms

In the first experiment, machine learning classifiers were implemented with the Bag of Words (BOW) feature extraction method. The performance of algorithms with BOW for Dataset-1 is shown in Table-2 and Figure-7. As per the experimental result, it was found that Logistic Regression produces better accuracy (92.6%), F-Score (94.3%), and Precision(96.4%), while Naïve Bayes provides better Recall(93.9%) than other machine learning classifiers. Accuracy of Support Vector Machine (91.9%), Decision Tree (91.9%), and Random Forest (91%) were very close to Logistic Regression.

Table 2: Performance Metrics of Machine Learning Algorithms with BOW for Dataset-1

Dataset 1 : Total Tweets/Comments = 35734				
Algorithm	**Accuracy**	**Precision**	**Recall**	**F-Score**
Logistic Regression	0.926	0.964	0.922	0.943
SVM	0.919	0.946	0.929	0.938
Decision Tree	0.919	0.953	0.922	0.937
Random Forest	0.910	0.934	0.928	0.931
Naive Bayes	0.887	0.893	0.939	0.916
K-Nearest Neighbor	0.864	0.923	0.865	0.893

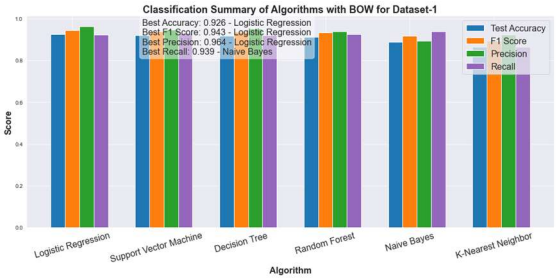

Figure 7: Classification Summary of Algorithms with BOW for Dataset-1

The second experiment applied machine learning algorithms on the same dataset-1 with TF-IDF features extraction method. In Table-3 and Figure-8, the performance of algorithms with TF-IDF is shown. As per the experimental result, it was found that the Support Vector machine produces better accuracy(92.5%), F-Score(94.3%), while Logistic Regression provides better Precision(95%) and Naïve Bayes provides better Recall(97.6) than other machine learning classifiers when TF-IDF was used for features extraction. Accuracy of Logistic Regression (92%), Decision Tree (91.3%), and Random Forest (90.5%) were very close to Support Vector Machine. Comparative results of algorithms with BOW and TF-IDF for Dataset-1 show that performances of algorithms are almost the same with both feature extraction.

Table 3: Performance Metrics of Machine Learning Algorithms with TF-IDF for Dataset-1

Dataset 1 : Total Tweets/Comments = 35734				
Algorithm	**Accuracy**	**Precision**	**Recall**	**F-Score**
SVM	0.925	0.947	0.938	0.943
Logistic Regression	0.920	0.950	0.927	0.938
Decision Tree	0.913	0.940	0.926	0.933
Random Forest	0.905	0.928	0.926	0.927
Naive Bayes	0.836	0.810	0.976	0.886
K-Nearest Neighbor	0.817	0.859	0.861	0.860

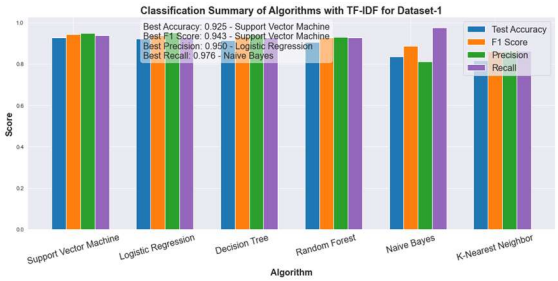

Figure 8: Classification Summary of Algorithms with TF-IDF for Dataset-1

In the third experiment, machine learning classifiers were implemented with BOW on Dataset-2. The performance of algorithms with BOW for Dataset-2 is shown in Table-4 and Figure-9. As per the experimental result, it was found that again Logistic Regression produces better accuracy (85%), F-Score (82.5%), and Precision (88.2%), while Naïve Bayes provides better Recall (78.8%) than other machine learning classifiers. Accuracy of Support Vector Machine (82.7%), Random Forest (82.1%), and Decision Tree (81.9%) were close to Logistic Regression.

Table 4: Performance Metrics of machine Learning Algorithms with BOW for Dataset-2

Dataset 2 : Total Tweets/Comments = 70019				
Algorithm	**Accuracy**	**Precision**	**Recall**	**F-Score**
Logistic Regression	0.850	0.882	0.773	0.825
SVM	0.827	0.842	0.765	0.801
Decision Tree	0.819	0.811	0.771	0.799
Random Forest	0.821	0.822	0.775	0.798
Naive Bayes	0.812	0.807	0.788	0.789
K-Nearest Neighbor	0.744	0.852	0.531	0.655

Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

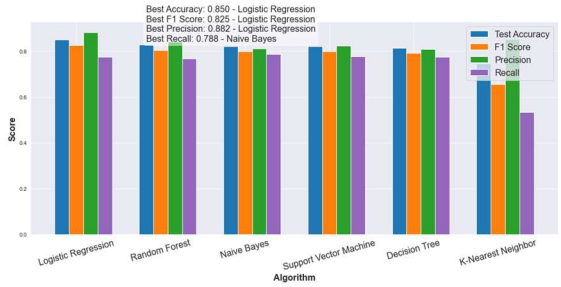

Figure 9: Classification Summary of Algorithms with BOW for Dataset-2

In the fourth experiment, machine learning classifiers were implemented with TF-IDF on Dataset-2. The performance of algorithms with TF-IDF for Dataset-2 is shown in Table-5 and Figure-10. As per the experimental result, it was found that again Logistic Regression produces better accuracy (85%), F-Score (82.2%), and Precision (89.6%) while Support Vector Machine provide better Recall (78.4%) than other machine learning classifiers. Accuracy of Support Vector Machine (83.7%), Random Forest (82.7%), and Decision Tree (81%) were close to Logistic Regression. Comparative results of algorithms with BOW and TF-IDF for Dataset-2 show that algorithms' performances are almost identical with both feature extraction. In both datasets and with both BOW and TF-IDF, Logistic Regression and Support Vector Machine are top performer algorithms with while K-Nearest Neighbor classifier obtained the lowest accuracy. Algorithms outperformed on dataset-1 than dataset-2.

Table 5: Performance Metrics of machine Learning Algorithms with TF-IDF for Dataset-2

Dataset 2 : Total Tweets/Comments = 70019				
Algorithm	**Accuracy**	**Precision**	**Recall**	**F-Score**
Logistic Regression	0.850	0.896	0.759	0.822
SVM	0.837	0.847	0.784	0.814
Random Forest	0.827	0.840	0.768	0.802
Decision Tree	0.810	0.803	0.774	0.788
Naive Bayes	0.817	0.845	0.735	0.786
K-Nearest Neighbor	0.734	0.841	0.633	0.657

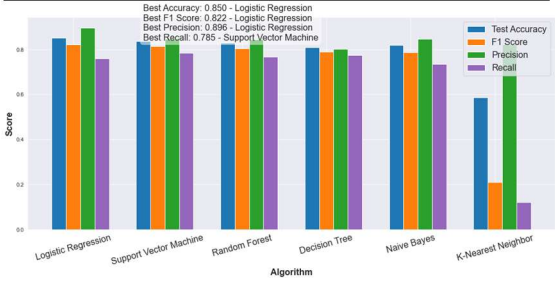

Figure 10: Classification Summary of Algorithms with TF-IDF for Dataset-2

5.2 Time complexity of Algorithms

The time complexity of algorithms with BOW and TF-IDF for both datasets is shown in Table-6 and Table-7, respectively. As per the experimental result, it was found that K-Nearest Neighbor has achieved the best training time, Random Forest has obtained the worst training time, Logistic Regression and Support Vector Machine have achieved the best prediction time while K-Nearest Neighbor has obtained the worst prediction time.

Table 6: Time Complexity of Machine Learning Algorithms with BOW and TF-IDF for Dataset-1

S. No	ML Algorithm	Prediction Time with BOW	Training Time with BOW	Prediction Time with TF-IDF	Training Time with TF-IDF
1	Logistic Regression	0.001	0.667	0.001	0.566
2	Support Vector Machine	0.015	0.270	0.002	0.066
3	Decision Tree	0.015	3.224	0.016	3.676
4	Random Forest	2.459	35.439	2.368	27.872
5	Naive Bayes	0.002	0.006	0.003	0.006
6	K-Nearest Neighbor	15.365	0.015	17.677	0.001
	Best Training Time :	0.015- K-Nearest Neighbor		0.001- K-Nearest Neighbor	
	Worst Training Time :	35.439- Random Forest		27.872- Random Forest	
	Best Prediction Time :	0.001- Logistic Regression		0.001- Logistic Regression	
	Worst Prediction Time :	15.365- K-Nearest Neighbor		17.677- K-Nearest Neighbor	

Table 7: Time Complexity of Machine Learning Algorithms with BOW and TF-IDF for Dataset-2

S. No	ML Algorithm	Prediction Time with BOW	Training Time with BOW	Prediction Time with TF-IDF	Training Time with TF-IDF
1	Logistic Regression	0.008	3.862	0.003	3.467
2	Support Vector Machine	0.012	25.433	0.002	0.763
3	Decision Tree	0.345	84.838	0.418	99.896
4	Random Forest	30.938	1144.595	37.062	1088.051
5	Naive Bayes	0.009	0.022	0.0181	0.024
6	K-Nearest Neighbor	191.055	0.009	179.691	0.011
Best Training Time :		0.009- K-Nearest Neighbor		0.011- K-Nearest Neighbor	
Worst Training Time :		1144.595- Random Forest		1088.051- Random Forest	
Best Prediction Time :		0.008- Logistic Regression		0.002- Support Vector Machine	
Worst Prediction Time :		191.055- K-Nearest Neighbor		179.691- K-Nearest Neighbor	

Table 8: Comparative outcomes with Related Work

Authors	Year	Classifier	Accuracy	Precision	Recall	F-Score
Vikas S Chavan [26]	2018	Logistic Regression	73.76%	0.644	0.614	0.629
		Support Vector Machine	77.65%	0.702	0.582	0.637
John Hani [29]	2019	Neural Network	91.76%	0.924	0.917	0.919
		Support Vector Machine	89.87%	0.896	0.901	0.898
Amgad Muneer [33]	2020	Logistic Regression	90.57%	0.952	0.905	0.928
		LGBM Classifier	90.55%	0.961	0.895	0.927
		SGD Classifier	90.6%	0.968	0.889	0.927
		Random Forest	89.84%	0.934	0.913	0.923
		AdaBoost Classifier	89.30%	0.962	0.875	0.916
		Naive Bayes	81.39%	0.795	0.973	0.875
		Support Vector Machine	67.13%	0.671	1.000	0.803
Our Results with BOW for Dataset-1	2021	Logistic Regression	**92.6%**	0.964	0.922	**0.943**
		Support Vector Machine	**91.9%**	0.946	0.929	**0.938**
		Decision Tree	**91.9%**	0.953	0.922	**0.937**
		Random Forest	91%	0.934	0.928	**0.931**
		Naive Bayes	88.7%	0.893	0.939	0.916
		K-Nearest Neighbor	86.4%	0.923	0.865	0.893
Our Results with TF-IDF for Dataset-1		Support Vector Machine	**92.5%**	0.947	0.938	**0.943**
		Logistic Regression	**92%**	0.950	0.927	**0.938**
		Decision Tree	91.3%	0.940	0.926	**0.933**
		Random Forest	90.5%	0.928	0.926	0.927
		Naive Bayes	83.6%	0.810	0.976	0.886
		K-Nearest Neighbor	81.7%	0.859	0.861	0.860

5.3 Comparison of Results with Related Work

The performance results of our proposed approach were also compared with the previous experiments performed by the researchers of [26, 29, 30]. The comparative outcomes are shown in Table-8. Our proposed model provided better results for dataset-1 as to the Accuracy and F-score, while Precision and Recall were almost equal and similar. Experimental result on dataset-2 was not used for the comparison.

6. CONCLUSION

Cyberstalking is a rising and challenging category of cybercrime that creates a fear situation for users of internet applications. In this paper, we explored the various machine learning techniques used by the researchers for cyberstalking and cyberbullying detection in social media networks and proposed a machine learning framework for cyberstalking detection. We experimented on two different sizes and distribution of datasets using six supervised machine learning algorithms based on BOW and TF-IDF feature extraction methods. The Accuracy, Precision, Recall, F-Score, Prediction time, and Training time were measured to analysis, the performance of algorithms in the proposed detection model. As per our experimental results, Logistic Regression (92.6% with BOW and 92% with TF-IDF) and Support

Vector Machine (92.5% with TF-IDF and 91.9% with BOW) achieved the highest accuracy while K-Nearest Neighbor obtained the lowest accuracy (86.4% with BOW and 81.7% with TF-IDF) for dataset-1. Logistic Regression and Support Vector Machine also achieved the highest Precision and F-Score, while Naïve Bayes provides the best Recall for both datasets. Logistic Regression and Support Vector Machine was top performer algorithms for both datasets. Machine learning classifiers performed better for dataset-1 (65.7% cyberstalking rows) than dataset-2 (45.7% cyberstalking rows).

In our experimental work, we found that the performance of algorithms is also dependent on the size and distribution of the dataset and feature extraction methods. Performance analysis of machine learning algorithms in this paper will surely help others to select the appropriate machine learning classifier for cyberstalking and other cyberharassment detection. However, there is no best algorithm for all cases, and the selection of a classifier should be according to the problem and datasets.

REFERENCES

[1] P. E. Mullen, M. Pathé, R. Purcell, "Stalking: New constructions of human behavior", *Australian and New Zealand Journal of Psychiatry*, 35, 9–16, 2001.

[2] E. Ogilvie, "Cyberstalking. Trends and Issues in Crime and Criminal Justice", *Australian Institute of Criminology*, (166), 1,2000.

[3] E. Short, T. Stanley, M., Baldwin, G. G. Scott, "Behaving Badly Online: Establishing Norms of Unacceptable Behaviours", *Studies in Media and Communication*, *3*(1), 2015,pp. 1-10.

[4] https://www.bbc.com/news/world-asia-india-33532706

[5] https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[6] M. Baer, "Cyberstalking and the Internet Landscape We Have Constructed", *Virginia Journal of Law & Technology*, 15,154, 2020 pp. 153-227.

[7] J. L. Truman, "Examining intimate partner stalking and use of technology in stalking victimization". *Ph.D. thesis, University of Central Florida Orlando, Florida*, 2010.

[8] D. A. Jurgens, P. D. Turney, and K. J. Holyoak. "SemEval-2012 Task 2: Measuring Degrees of Relational Similarity", *First Joint Conference on Lexical and Computational Semantics*, 1, 2012, pp. 356–364.

[9] N. Parsons-pollard and L. J. Moriarty, "Cyberstalking: Utilizing What We do Know", *Victims an Offenders,* 4(4), 2009, pp. 435–441.

[10] N. M. Zainudin, K.H. Zainal, N. A. Hasbullah, N. A. Wahab, S. Ramli, "A review on cyberbullying in Malaysia from digital forensic perspective", *International Conference on Information and Communication Technology (ICICTM), IEEE*, 2016, pp. 246-250.

[11] https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

[12] https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861

[13] Osisanwo FY, Akinsola JE, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J. "Supervised machine learning algorithms: classification and comparison", *International Journal of Computer Trends and Technology (IJCTT)*, 48, 3, 2017, pp. 128-138.

[14] S. Ray, "A Quick Review of Machine Learning Algorithms," *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 35-39.

[15] Mahesh, Batta. "Machine Learning Algorithms-A Review", *International Journal of Science and Research (IJSR).[Internet],* 9, 2020, pp. 381-386.

[16] Breiman, L. "Random Forests", *Machine Learning* 45, 2001, pp. 5–32. https://doi.org/10.1023/A:1010933404324

[17] Rish, "An empirical study of the naive bayes classifier", *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.

[18] Z. Ghasem, I. Frommholz, and C. Maple, "Machine learning solutions for controlling cyberbullying and cyberstalking", *International Journal of Information Security*, 6, 2, 2015, pp.55-64.

[19] B. Nandhini and JI Sheeba, "Cyberbullying detection and classification using information retrieval algorithm*", International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015),* ACM, 2015, pp. 20.

[20] Vikas S Chavan and SS Shylaja, "Machine learning approach for detection of cyber-aggressive comments by peers on social media network", *In Advance in computing communications and informatics (ICACCI), 2015 International Conference, IEEE,* 2015,pp. 2354–2358.

[21] S.Vidhya, D.Asir Antony Gnana Singh, E.Jebamalar Leavline, "Feature Extraction for Document Classification", *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 4, Special Issue 6, 2015.

[22] Ingo Frommholz, Haider M. al-Khateeb, Martin Potthast, Zinnar Ghasem, Mitul Shukla , Emma Short, "On Textual Analysis and Machine Learning for Cyberstalking Detection", *Datenbank Spektrum* 16, 2016, pp. 127–135.

[23] M. Ganesan, P. Mayilvahanan, "Cyber Crime Analysis in Social Media Using Data Mining Technique", *International Journal of Pure and Applied Mathematics*, 116, 22, 2017, pp. 413–424.

[24] Walisa Romsaiyud, Kodchakorn na Nakornphanom, Pimpaka Prasertsilp, Piyaporn Nurarak, and Pirom Konglerd, "Automated cyberbullying detection using clustering appearance patterns", *In Knowledge and Smart Technology (KST), 2017 9th International Conference*, IEEE, 2017, pp. 242–247.

[25] Sani Muhamad Isa, Livia Ashianti, "Cyberbullying classification using text mining", *In Informatics and Computational Sciences (ICICoS), 2017 1st International Conference,* IEEE, 2017, pp. 241–246.

[26] Hitesh Kumar Sharma, K Kshitiz, Shailendra, "NLP and Machine Learning Techniques for Detecting Insulting Comments on Social Networking Platforms", *Proceedings of the International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France*, 2018.

[27] Amanpreet Singh, Maninder Kaur, "Content-based Cybercrime Detection: A Concise Review", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-8 Issue-8, 2019, pp. 1193-1207.

[28] J.I. Sheeba, S. Pradeep Devaneyan , Revathy Cadiravane, "Identification and Classification of Cyberbully Incidents using Bystander Intervention Model", *International Journal of Recent Technology and Engineering (IJRTE)* 8,254, 2019.

[29] John Hani Mounir, Mohamed Nashaat, Mostafaa Ahmed, Eslam A. Amer, "Social Media Cyberbullying Detection using Machine Learning", *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 5, 2019.

[30] Ravinder Ahujaa , Aakarsha Chuga , Shruti Kohlia , Shaurya Guptaa , Pratyush Ahuja, "The Impact of Features Extraction on the Sentiment Analysis", *Procedia Computer Science*, Volume 152, 2019, pp. 341-348.

[31] Md Manowarul Islam, Selina Sharmin, "Cyberbullying Detection on Social Networks Using Machine Learning Approaches", *IEEE Asia-Pacific Conference on Computer Science and Data Engineering(CSDE)*,978-1-6654-1974-1/20,IEEE, 2020, DOI:10.1109/CSDE50874.2020.9411601

[32] Hoyeon Park, Kyoung-jae Kim, "Impact of Word Embedding Methods on Performance of Sentiment Analysis with Machine Learning Techniques", *Journal of The Korea Society of Computer and Information* Vol. 25 No. 8, 2020, pp. 181-188.

[33] Amgad Muneer, Suliman Mohamed Fati, "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter", *Future Internet*, 12, 187, 2020.

[34] Vijayarani, S., Ms J. Ilamathi, and Ms Nithya. "Pre-processing techniques for text mining-an overview." *International Journal of Computer Science & Communication Networks* 5.1, 2015, pp. 7-16.

[35] https://towardsdatascience.com/all-you-need-to-know-about-text-preprocessing-for-nlp-and-machine-learning-bc1c5765ff67.

[36] Kadhim, Ammar Ismael, "An evaluation of pre-processing techniques for text classification", *International Journal of Computer Science and Information Security (IJCSIS),* 16.6, 2018, pp. 22-32.

[37] Rui, Weikang, Kai Xing, and Yawei Jia, "BOWL: Bag of word clusters text representation using word embedding", *International Conference on Knowledge Science, Engineering and Management, Springer, KSEM*, 2016.

[38] Das B, Chakraborty S. "An improved text sentiment classification model using TF-IDF and next word negation", *arXiv preprint*, 2018, arXiv: 1806.06407.