International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V70I3P211

Volume 70 Issue 3, 95-107, March 2022
© 2022 Seventh Sense Research Group®

Review Article

A Review on Cyberstalking Detection Using Machine Learning Techniques: Current Trends and Future Direction

Arvind Kumar Gautam[1], Abhishek Bansal[2]

[1]*Research Scholar, Department of Computer Science, Indira Gandhi National Tribal University Amarkantak, Madhya Pradesh, India*
[2]*Assistant Professor, Department of Computer Science, Indira Gandhi National Tribal University Amarkantak, Madhya Pradesh, India*

[1]gautamjirewa@gmail.com, [2]abhishek.bansal@igntu.ac.in,

Abstract - *Web-based media organizations and other web applications, for example, WhatsApp, Facebook, YouTube, Instagram, Twitter, have become more well known among individuals for data sharing, live occasions, news, exposure, publicity, and cybercrimes. The utilization of online media stages additionally offers major issues through cyberstalking, cyberbullying, and different kinds of digital provocation. Cyberstalking and cyberbullying are frequently utilized reciprocally and include the utilization of the web to follow or target somebody in the web-based world. Cyberstalking is a basic worldwide issue that influences instructive foundations, casualties, and the whole human culture that should be distinguished, recognized, revealed, and controlled appropriately for the security of clients in online media. Machine learning is the most well-known method for making the cyberstalking recognition model. Researchers have recommended different recognition procedures utilizing machine learning to control and battle cyberstalking in web-based media. In this paper, the study relates to some popular features extraction methods machine learning classifiers for text classification and explores the datasets used by the researchers. The study also focuses on reasonably determining the research gaps and the scope for improving cyberstalking detection. This paper will review some cyberstalking detection techniques using machine learning, analyze the performance of popular machine learning classifiers and finally explore the issues, challenges, recent trends, and future direction for cyberstalking detection techniques.*

Keywords - *Machine learning, Cyberstalking detection, Cyberbullying, Features extraction, Word embedding.*

I. INTRODUCTION

In the era of the web world, web-based media applications and email innovation are making routinely to the academic culture. Individuals frequently utilize online media stages for different great and awful exercises, like training, business, amusement, counterfeit news promulgation, exposure, and cybercrimes. These days, individuals frequently invest energy on web-based web-based media locales like Facebook, Twitter, WhatsApp, Pinterest, message, and so forth. As indicated by [1], multiple billion individuals worldwide are utilizing web-based media applications. Consequently, numerous digital aggressors are dynamic on these stages. Although online media and other web applications have become more normal for thought expression, a few criminal clients utilize these networks in illicit and deceptive ways. Cybercriminals are likewise utilizing web-based media, and numerous cyberstalking cases are recognized every day via web-based media. Cyberstalking [2] is a not kidding digital assault in which the aggressor utilizes advanced media to bug the person in question or to gather through close-to-home assaults and the divulgence of bogus or private data among different people. Cyberstalking is a developing and critical issue chiefly among youngsters, ladies, and understudies. A few kinds of examination have shown that stalkers experience the ill effects of social and mental conditions. Cyberstalking is a pandemic and answerable for a violent and criminal society, especially in the understudies of instructive organizations. Cyberstalking casualties experience quantifiable adverse consequences identical to overcomers of injuries, for example, sexual assaults or bombing [3]. 90% of a victim of the digital following are ladies. According to the BBC [4] report, the first cyberstalking case was enlisted in 2009. The impact of cyberstalking on different web-based media stages can't be disregarded, and for this, significant consideration is needed to control cyberstalking. Cyberstalking should be contemplated in detection, counteraction, and control to diminish its destructive impact. Many types of research and approaches are suggested in the literature, mainly focused on preventing cyberstalking, called intervention and prevention approaches. Prevention approaches are not generally compelling in controlling and diminishing cyberstalking. Therefore, detection of cyberstalking is profoundly required. Researchers generally use machine learning approaches for cyberstalking detection in web-based media and other internet applications.

In this paper, the study relates to different detection techniques, features extraction techniques, machine learning classifiers for text classification and explores the datasets used by the researchers. The study also focuses on reasonably determining the research gaps and the scope for improving cyberstalking detection. This paper will review some cyberstalking detection techniques using machine learning and finally explore the issues, challenges, recent trends, and future direction for cyberstalking detection techniques. The rest of the paper is structured in a section-wise manner. Some essential background connected with machine learning and cyberstalking is explained in section II. Section III outlines some identification procedures utilized for cyberstalking and other cyberharassment detection. Section IV shows the survey of past work performed by the researchers for cyberstalking identification utilizing machine learning procedures. Section V depicts the standard methodology used for cyberstalking detection using machine learning. Experimental results by utilizing several machine learning classifiers are described in section VI. Section VII depicts the recent trend and challenges in cyberstalking detection methods. At long last, Section VII finalizes the review.

II. BACKGROUND

A. Cyberbullying and Cyberstalking

Cyberstalking and cyberbullying are often utilized reciprocally and include the utilization of the web to follow or target somebody in the internet-based world. Cyberstalking and cyberbullying use the same technology and focus on hassling web clients. Cyberbullying mostly centers around youngsters, while cyberstalking targets other gatherings of clients in the web world for online badgering. Cyberstalkers consistently use web informational collections, online media, and other web-based tools to follow, bully and undermine others. Cyberstalking is a not kidding and muddled cybercrime that affects and targets numerous people and organizations [5]. Cyberstalking, a developing worldwide issue, is often underestimated by general society, researchers, and the government. Cyberstalking is orderly, rehashed, and various digital assaults and doesn't happen on an isolated event [6]. According to the review, just about a fifth of individuals have confronted cyberstalking circumstances during the utilization of web applications [7]. According to accessible proof [8], cyberstalking cases will routinely increment surprisingly. Cyberstalking may be primarily classified into the following types based on utilization technology [9].

a) Email stalking

Email stalkers send undermining and scornful messages utilizing email. These messages may likewise contain spam or viruses.

b) Internet stalking

Internet stalkers use worldwide web-based media applications to hassle or savage other internet users.

c) Computer stalking

When a stalker hacks the victim's computer and targets it, it is called computer stalking.

d) Phone stalking

When stalkers focus on the victim utilizing rehashed, undesirable calls, messages, audio, video, and pictures through cell phones.

e) Automated stalking

Is an advanced technology utilized by stalkers to target victims utilizing mobile apps and automated computer programs constrained by dubious servers.

There are numerous instances of cyberstalking, such as making and posting a genuine or phony sexual picture of the victim to their friends and family, transferring individual data on open sites, and hacking the victim's web-based media and email account [10]. Web-based media stages are a potential hunting ground for cyberstalkers. According to the context, there are some other categories of cyberstalking. Such as trolling and flaming to someone in social media, excluding victims from any social media group, creating fake profiles (masquerading), mobbing using repeated messages, denigrating to someone, sharing and outing the private data of someone in social media, harassing regularly, and hacking the victim accounts and devices are modern categories of cyberstalking.

B. Machine Learning

Machine learning algorithms (MLA) is the most famous utilization of artificial intelligence (AI), which has the capacity concerning auto-learning and gives exact and moderate outcomes from learning experiences [11]. Machine learning uses the existing computations and classification techniques with datasets and improvement projects to offer a palatable response for the issue and use them to gain isolation proficiency. The most well-known learning method starts with insights in data, seeing the models in data and making progressed decisions, and involving them in the future, relying upon the recently recognized models. The magnificent point is to make computers adjust without human affiliation or help and change results similarly.

Machine learning gives more exact results rapidly by inspecting tremendous proportions of data [12]. Machine Learning techniques can be classified as supervised, unsupervised, semi-supervised, and reinforcement learning.

a) Supervised Machine Learning Algorithms

Supervised machine learning methods use classification tasks to characterize the data into labelled data. These algorithms contain the dependent data predicted from a predefined set of non-dependent data. Such types of machine learning are fundamentally utilized for regression and classification issues [11].

b) Unsupervised Machine Learning Algorithms

Any target/outcome or dependent variables are not used to predict in such ML algorithms. In such algorithms, computers are trained using the unlabeled data and mainly used for clustering the data into different groups. Descriptive modelling and Pattern detection are the main application of such algorithms [11].

c) Semi-Supervised Machine Learning Algorithms

Semi-supervised learning uses the primary benefits of supervised and unsupervised learning, and it could be involved marked and non-named information according to issues circumstances. These strategies exploit the likelihood that despite the way that the gathering interests of the unlabeled data are dark, this data passes on significant information about the gathering limits [12].

d) *Reinforcement Machine Learning Algorithms*

A trained machine model is utilized to settle on clear choices in such algorithms. The model is presented to an environment, and then it prepares automatically ceaselessly using experimentation factors. The machine gains from previous experience and attempts to catch the perfect information to make the final decision [12].

III. CYBERSTALKING DETECTION TECHNIQUES

Cyberstalkers are improving their approach and utilizing new technology to target the victims and achieve their goals. Researchers are developing the cyberstalking detection model using several detection techniques for combating stalkers. Based on the literature, the following main methods (as shown in Fig. 1) are used for cyberstalking and other cyberharassment detection [13].

A. Cyberstalking Detection using Machine Learning (ML)

Researchers generally utilize machine Learning (ML) either as a solitary methodology or a hybrid approach for cyberstalking identification. Researchers also use neural networks[14], deep learning[15], and fuzzy logic[16] techniques for better performance on massive datasets. A neural network, a subset of machine learning, is an amusement of how the human brain functions. Such type of can does getting ready and learn without any other individual ward on past data. A neural network is fit for performing calculations quicker because it works in much the same way as the human mind. Deep learning is a three-layer-based subset of machine learning. Deep learning can upgrade and refine for exactness on vast and complex datasets. Fuzzy logic is a figuring procedure that contains traditional and fuzzy sets, and it works depending on the level of truth rather than the boolean logic[17].

B. Cyberstalking Detection using Data Mining

Data mining is a more manual cycle that depends on human mediation and direction. In the beginning advance of the process in data mining, rules or examples are obscure, and it utilizes the current dataset as a data warehouse to track down designs [18].

C. Cyberstalking Detection using Statistical Method

Some statistical methods such as the CPRA-EWMA, CPRA-Shewhart, Hidden Markov Model, Bayesian learning network approach[13], and outlier detection algorithm also suggested by the researchers for cyberstalking other types of cyberharassment detection.

D. Cyberstalking Detection using Other Techniques

Some other techniques such as Cryptography, Biometric, Computer Vision, and Forensics Tools are also used by researchers for cyberstalking and different cyberharassment detection. The cyberstalking detection using the cryptography approach focused on authentication and verifying the identity of cyberstalkers and sources of cyberstalking data [19]. The biometric system is mainly helpful for face recognization, images identification, and stalkers verification[20]. Computer vision techniques [20] analyze the images and determine whether URLs are genuine or fake. Forensics tools deal with the collection of evidence from digital media, and it is helpful to track the actions of cyberstalkers and address the security issues[20].

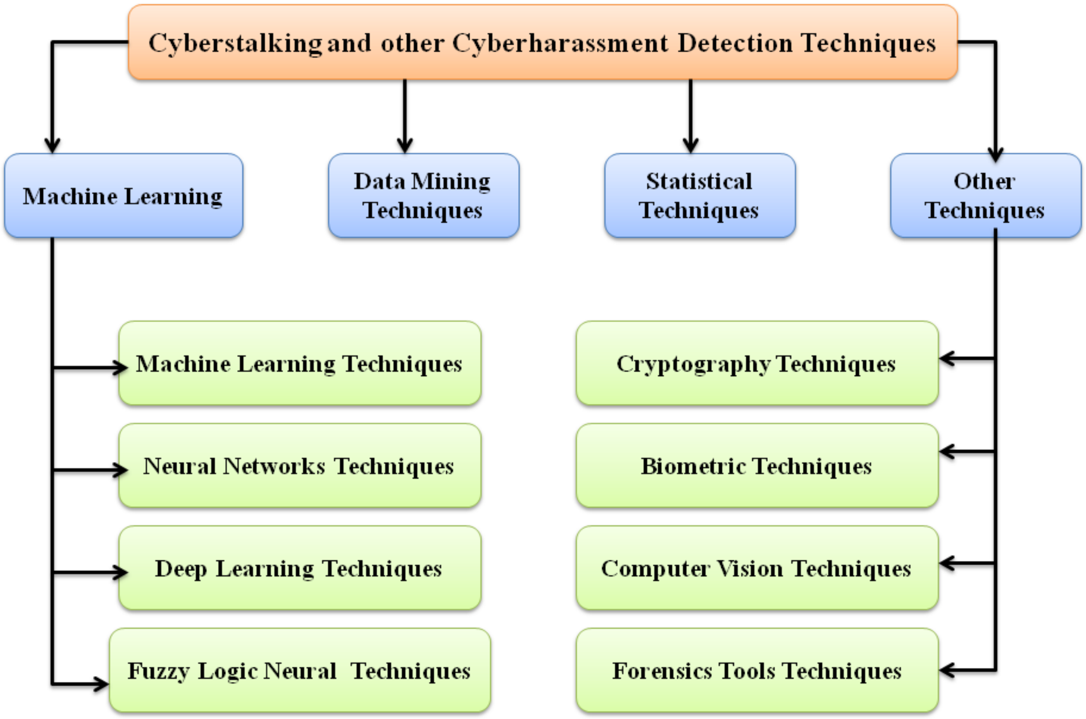
Cyberstalking and other Cyberharassment Detection Techniques
Machine Learning
Data Mining Techniques
Statistical Techniques
Other Techniques
Machine Learning Techniques
Neural Networks Techniques
Deep Learning Techniques
Fuzzy Logic Neural Techniques
Cryptography Techniques
Biometric Techniques
Computer Vision Techniques
Forensics Tools Techniques

Fig. 1 Cyberstalking detection techniques

IV. REVIEW OF LITERATURE

This section discusses and reviews some researchers' contributions that used the machine learning algorithm for cyberstalking and cyberbullying detection on social media platforms and other internet applications. Most of the studies are focused on text-based cyberstalking, while few are based on multimedia-based cyberstalking detection. In the literature, researchers have applied machine learning from the supervised and unsupervised learning algorithm and natural language processing techniques for features extraction to better the detection model's performance.

In 2011, Dinakar K. et al. [21] suggested classifying textual cyberbullying using binary and multi-class text classification techniques. The authors performed their experiment work on 4500 comments on YouTube and found that the performance of binary classifiers is better than a multi-class classifier for individual labels. Another machine learning-based model for cyberbullying detection was implemented by Kelly Reynolds et al. [22]. The authors used separate training datasets to count and normalize the information and performed questionnaires on the Formspring. Me website dataset.

In 2012, Dadvar M. et al. [23] suggested an enhanced model for cyberbullying detection utilizing the substance, sentiment, and logical elements as a hybrid approach. The authors claimed that their detection model is equipped for observing web-based bullies and stalkers. They got high accuracy involving supervised machine learning procedures for distinguishing. Again Dadvar M. et al. [24] proposed another embraced model using the gender-based way to deal with cyberbullying detection via web-based media networks. For the assessment of work, a support vector machine was utilized. Authors observed that guys use more offensive and negative words contrasted with females. They utilize the MySpace dataset for trial work containing a few remarks focusing on teenagers and ladies. Vinita Nahar et al. [25] have proposed a model for cyberbullying detection utilizing two feature selection techniques. The BOW technique was used for standard features extraction from offensive and non-offensive comments. Interestingly, the probabilistic dormant semantic examination strategy performed feeling highlights extraction from just oppressive messages. The authors asserted that they accomplished high exactness using a support vector machine for text classification.

In 2013, Vinita Nahar et al. [26] developed a detection model using ensemble machine learning classifiers on a small set of labelled data and a massive set of non-labelled data. The authors used the concept of link analysis to determine the role of victims and predators and claimed that ensemble machine learning classifiers outperformed with better accuracy. Dadvar M. et al. [27] again implemented a feature-based detection model. They showed that detection performance could be improved using offensive-specific features such as user's profile and comments history. The experiment was performed on a Youtube dataset containing

4626 comments from 3858 users. The authors applied the Bag of Words technique for features extraction and support vector machine algorithm for text classification.

In 2014, Vivek K. Singh et al.[28] have proposed a detection model utilizing social and textual analysis. Social highlights were utilized to show the connection among clients and the ID of sender and receiver. The authors utilized the Twitter dataset containing around 900,000 posts of 27135 users. Machine learning classifiers like Bagging, J48, Dagging, and Naive Bayes were utilized for the text classification. Cicero Nogueira et al. [29] proposed a model for sentiment analysis utilizing the deep learning neural network method. The authors performed test work on Twitter tweets and accomplished with an accuracy of 86.4%.

In 2015, Ghasem Z. et al. [30] introduced a machine learning-based solution for controlling cyberbullying and cyberstalking. This methodology is predominantly centered around automatic identification and proof documentation of email-based cyberstalking. Authors utilized machine learning, text mining, factual analysis, and email crime scene investigation to distinguish and relieve cyberstalking. The authors performed the experimental work involving support vector machine and neural network procedures in 5172 email datasets containing spam and certified email. Another machine learning-based detection model was implemented by Nandhini et al. [31]. The authors used the Naive Bayes machine learning method on the myspace.com dataset and asserted that they accomplished 91% accuracy. Chavan et al. [32] likewise played out the exploratory work on a dataset from Kaggle for their proposed approach utilizing logistic regression and support vector machine classifier. Authors claimed that their proposed model accomplished 73.76% accuracy using logistic regression, while 77.65% accuracy was achieved using the support vector machine.

In 2016, Frommholz I. et al. [33] proposed a detection model for textual analysis and cyberstalking detection utilizing machine learning algorithms which were essentially centered around creator ID, message classification, personalization, and digital text forensics. Michele Di Capua [34] has proposed a detection model utilizing an unsupervised machine learning approach via online media stages. The authors performed experimental work on a few datasets from Twitter, YouTube, and Formspring involving the support vector machine for text classification. Syntax, semantic, sentiment, and social features were utilized for feature extraction. Vivek K. Singh et al. [35] proposed a detection system using a social and textual feature to predict users' roles as predators or victims. The proposed framework was assessed on the equivalent datasets utilized by past researchers and claimed that it accomplished better accuracy.

In 2017, Ganesan et al. [36] suggested analyzing the cybercrime data from the web pages database using unpredicted patterns and data mining techniques. Based on the experiment, the authors claimed that this model would categorize the cybercrime offences as violent or non-violent and categorize the types of cybercrimes such as cyber terrorism, cyberstalking, cyber fraud, and cyber theft. Romsaiyud et al. [37] have proposed an enhanced framework using Naive Bayes machine learning and achieved 95.79% accuracy. The authors used multiple datasets from MySpace, Slashdot, and Kongregate for evaluation work. Lsa et al. [38] proposed another approach using the SVM and NB 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. [39] proposed a framework using NLP and ML models to detect insulting and offensive comments on different social media networks. The authors claimed that their proposed framework would enhance the detection performance for offensive comments in social media networks. Sweta A. et al.[40] proposed the detection model on different social media platforms that can be transferred to another platform for offensive comments detection. Experimental work was performed on datasets from Formspring, Wikipedia, and Twitter using four different Deep Learning techniques. User profiles and social graphs were used to better the model's performance.

In 2019, Amanpreet Singh et al. [41] reviewed and compared the various 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. JI Sheeba et al. [42] proposed a bystander intervention model involving a random forest classifier to identify and classify cyberbullying. The authors utilized the latent semantic analysis and random forest classifier to categorize cyberbullying into different subcategories. John Hani et al. [43] implemented a model for cyberbullying detection in online media utilizing neural networks and SVM classification. Experimental work was performed using the sentiment analysis and TF-IDF technique on the Kaggle dataset. Authors have used the different classifiers as supervised machine learning algorithms for the training and testing. The authors accomplished better accuracy when they performed the experimental work on the same dataset that past researchers utilized. The authors claimed that their proposed approach achieved 92.8% and 90.3% accuracy for neural networks and SVM, respectively.

In 2020, V. Balakrishnan et al. [44] implemented a framework utilizing ML algorithms for automated detection of cyberbullying in Twitter tweets. This method groups the tweets as bully tweets, aggressor tweets, spammer tweets, and valid tweets with psychological characters, sentiment, and feelings. The experiment was performed utilizing NB, RF, and J48 ML algorithms on a dataset containing 5453 tweets. Manowarul Islam et al. [45]developed a supervised ML system to improve cyberbullying and cyberstalking detection accuracy via web-based media networks. The authors assessed their proposed framework on DT, RF, NB, and SVM ML-classifier. BOW and TF-IDF were utilized for features extraction. The authors claimed that they had accomplished better accuracy. Amgad Muneer et al. [46] have proposed a machine learning framework to enhance the detection accuracy on Twitter. The authors assessed their proposed framework utilizing DT, LR, LGBM, SGD, RF, AdaBoost, NB, and SVM classifiers. Word2Vec and TF-IDF strategies were used for features extraction. Anant Khandelwal et al. [47] have proposed a unified system for aggression identification on English-Hindi blended comments utilizing distinctive deep learning models like Deep Pyramid CNN, Disconnected RNN, and Pooled BiLSTM. The authors used features extraction strategies, specifically emotion sensor feature, parts-of-speech(pos), sentiment analysis, topic signals from the message, and TF-IDF emoticon feature. The authors performed the test work on TRAC 2018 Dataset and Kaggle Dataset.

In 2021, A. Asante et al. [48] proposed a content-based technical solution for cyberstalking detection. The proposed model utilized a few modules: message identification, filtering, detection (content detection and profiling offender), and evidence modules. Authors utilized machine learning, data mining strategies, digital forensics, and profiling to investigate text, picture, and media substance, gather proof, and profile offenders. N. Dughyala et al. [49] have proposed a model for automating the detection of cyberstalking utilizing machine learning and natural language processing methods. The authors claimed that their proposed system would automatically detect cyberstalking and recognize the stalker on the web. BP. Doppala et al. [50] have proposed a machine learning approach to automatically determine harassment in social networks. The authors performed the test work utilizing different machine learning classifiers and claimed that their proposed structure delivered better accuracy and automatically detected cyber harassment in the social network. Jain et al. [51] developed a cyberbullying detection model by utilizing the machine learning techniques with BOW, TF-IDF, and Word2Vec as feature extraction methods. After the experimental results, the authors found that machine learning techniques performed better with BOW and TF-IDF than the Word2Vec model. Pericherla et al. [52] proposed a machine learning model to analyze the performance of various feature extraction methods for cyberbullying detection. The authors utilized the logistic regression and LightGBM machine learning algorithms for classification in the experimental works. BOW, TF-IDF, Word2Vec, FastText, Glove, ALBERT(A Lite version of BERT), ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately), XLNet, RoBERTa, and GPT-2were applied for feature extraction. The authors claimed that GPT-2 and RoBERTa performed better with machine learning than other features extraction methods. Raj et al. [53] suggested a hybrid model for cyberbullying detection utilizing NLP and ML algorithms. Authors used several machine learning and deep learning algorithms (XG Boost, SVM NB, LR, CNN, GRU, BiGRU, LSTM, BiLSTM, and CNN-BiLSTM) for classification tasks while TF-IDF, BOW, GloVe, and FastText were applied as feature extraction methods. The authors observed that deep learning performed better compared to machine learning techniques based on the experimental results, although both provided almost similar results. The authors also observed that machine learning performed better with TF-IDF while GloVe enhanced the performance of deep learning algorithms.

V. STANDARD MACHINE LEARNING METHODOLOGY FOR CYBERSTALKING DETECTION

Generally, researchers utilize a multistage system for cyberstalking detection using machine learning techniques. As shown in Fig. 2, the following main phase is used for cyberstalking detection.

A. Data Pre-Processing

Dataset collected from various social media sources and other internet applications often contains different unnecessary characters or text. Before evaluating the machine learning algorithms, clean and prepared data are mandatory for the ML classifier in the detection phase. Data of the datasets are filtered and normalized to a specific format using keywords. In the pre-processing stage, usually, several tasks are performed using Natural Language Processing (NLP). Data pre-processing uses stop words removal task, noise removal, normalization, tokenization, stemming, and lemmatization[54],[55],[56]. After performing the pre-processing task using Natural Language Processing, clean data are sent to the next phase for feature extraction.

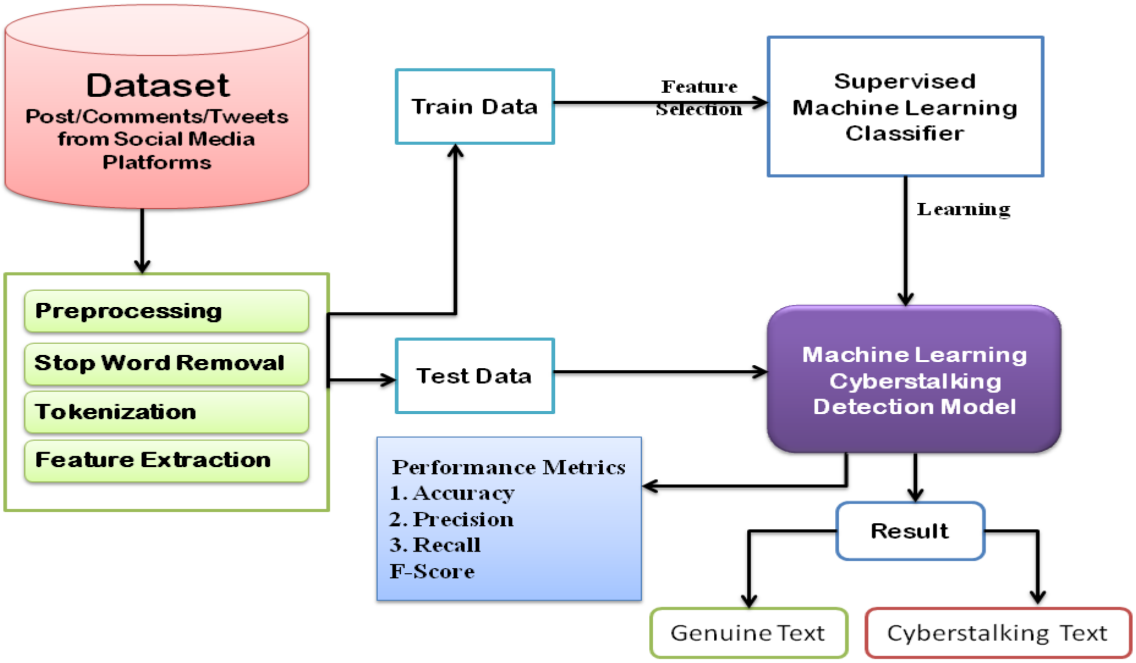

Fig. 2 Basic layout of cyberstalking detection techniques using machine learning

B. Feature Extraction

In this stage, feature vectors are calculated using feature extraction techniques. Text information is changed over into numbers because the machine learning classifier can't understand the information as text form. BOW [57], Word2Vec [58], and TF-IDF [59] are broadly used to change the texts into numbers. TF-IDF (Term Frequency-Inverse Document Frequency) is an accurate estimation that can quantify the significance of any word of records in an assortment of the corpus. In a BOW, each word is given equal importance. In TF-IDF, the consistently happening words should be more critical because common occurred words are more significant for classification. Word2vec produces one vector for each word, and its vectors address each word's novel situation. The researchers also broadly use n-gram (unigram, bigram, trigram, and so forth) methods alongside BOW, Word2Vec, and TF-IDF. With the growth of natural language processing, researchers utilize several advanced word embedding-based feature extraction for better performance. GloVe[60], FastText[61], ELMo[62], BERT[63], ALBERT[64], ELECTRA[65], GPT-2[66], XL-NET[67], RoBERTa[68], SBERT[69], Doc2VEC[70], InferSent[71], and Universal Sentence Encoder [72] are some popular and advanced examples of feature extraction methods.

C. Data Classification using Machine Learning Algorithms

In this phase, datasets are split into the training and testing part. Generally, 80% of data are utilized to prepare the ML model, while the leftover 20% of data are utilized to test the model before application in a real scenario. Initially, the machine learning model is trained with accurate and relevant data, and after that, the trained Ml model is applied to testing the data or any real-time data. A trained ML model calculates the prediction or probability score of data based on the learning experience. Based on the text's prediction or probability score, the trained model classifies the text into cyberstalking or non-cyberstalking. Researchers use various popular machine learning algorithms for classification. Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), K-Nearest Neighbor (KNN), Naive Bayes (NB), Bagging classifier, Gradient Boosting algorithm, XGBoost, and AdaBoosting algorithm are the most popular algorithms for text classification. SVM finds a hyperplane in an N-layered space that distinctly characterizes the elements [73],[74]. The component of the hyperplane relies on the number of features. The SVM supports several kernels, namely polynomial, sigmoid, Radial Basis Function, linear and nonlinear kernels with different mathematical functions. Logistic regression makes the different hyper-plane among datasets and accepts the elements as input to give the outcomes in the form of probability[75]. Logistic regression is classified as binomial, multinomial and ordinal. K-Nearest Neighbor is lethargic and occurrence realizing which characterized the new example in light of the separation from its neighbour [75]. Naïve Bayes utilize the Bayes Theorem to predict the outcome using the probability of any object [76]. Naïve Bayes are classified into multinomial NB, Gaussian NB and Bernoulli NB. Decision Trees make nodes and leaves as tree constructions to take and address the prediction [77]. Random Forest is an advanced form of decision trees classifier that uses multiple decision trees [74], [78]. A bagging classifier is an ensemble classifier that fits base classifiers on random subsets of the dataset and afterwards totals their singular predictions to frame actual predictions [79]. AdaBoost is the first boosting algorithm that can be applied in classification and regression assignments. AdaBoost works by placing more weight on challenging to arrange occasions and less on those all around dealt with well and has the ability to combine the numerous feeble classifiers into a solitary solid classifier[73]. Gradient boosting is an efficient algorithm for regression and classification that can limit the inclination mistake of the model [80]. A subsample of the training data is drawn aimlessly without substituting the full preparation dataset at every iteration in the Stochastic Gradient Boosting algorithm. The arbitrarily chosen subsample is utilized rather than the entire sample to fit the base learner [81]. XGBoost is a more efficient ensemble Machine Learning algorithm that uses a gradient boosting framework based on a decision tree approach [82].

D. Performance Metrics

After classifying data using machine learning-based classifiers, several parameters are used to measure the machine learning model's performance. Generally, a confusion matrix is used to calculate the performance parameters. In the case of cyberstalking detection (binary classification), the confusion matrix is a tabular representation of a "2x2" truth table that contains values of TP, FP, TN, and FN. TP (true-positive) shows the total correctly predicted as cyberstalking text, while FP (false-positive) indicates the total wrongly predicted as cyberstalking text. TN (true-negative) shows the total correctly predicted as non-cyberstalking text, while FN (false-negative) presents the total wrongly predicted as non-cyberstalking text. Several performance metrics such as accuracy, precision, f-score, recall, AUC, mean absolute error, mean squared error, specificity, training time, prediction time are used to measure the performance of cyberstalking detection model.

VI. EXPERIMENTAL RESULTS BASED ON A STANDARD METHODOLOGY

Experimental works were performed as per the standard methodology discussed in section V using python language. For training and testing purposes mixed dataset from Kaggle[83], [84], [85], [86], [87] was utilized. As mentioned in Fig. 3, the dataset contained a total of 35734 unique tweets/comments from social media, with 66% of cyberstalking rows and 34% of non-cyberstalking rows. Total 26800 rows were used for the training set, while testing of machine learning models was performed on 26800 rows.

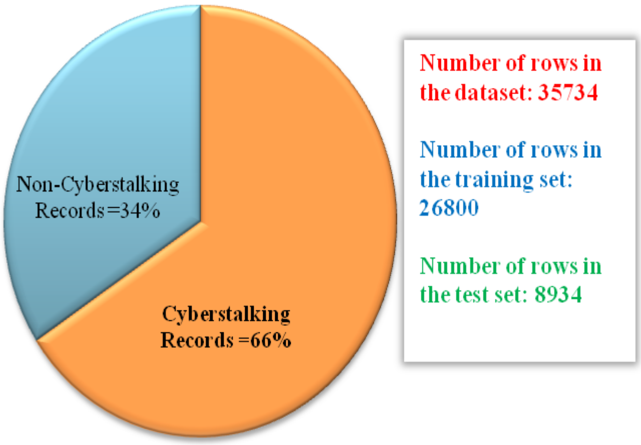

Fig. 3 Distribution of dataset for training and testing

The TF-IDF technique as traditional feature extraction and BERT as advanced word embedding techniques were both used separately for feature extraction. For text classification, top 11 popular machine learning algorithms, namely Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, Decision Tree(DT), Stochastic Gradient Descent (SGD), AdaBoost, Gradient Boosting(GB), Random Forest(RF), Bagging classifier, Naive Baye (NB), K-Nearest Neighbor (KNN) were utilized in the experiment. The performance of several utilized machine learning classifiers is mentioned in Table 1, Table 2, Fig. 4, and Fig. 5.

Experimental results described in Table 1 and Fig. 4 show that SVM outperformed other machine learning classifiers with TF-IDF feature extraction. SVM provided the highest accuracy of 92.5% and f-score of 94.3%, while GB achieved the highest precision of 98.7%, and NB achieved the highest recall of 97.6%. Bagging classifier (92.4%), LR (92%), SGD (92%), and XGBoost(91.7%) obtained the next highest accuracy respectively and were very near to the accuracy of SVM. KNN (81.7%) and NB (83.6%) received the lowest accuracy, respectively. The performance of machine learning classifiers with the BERT model is described in Table 2 and Fig. 5. As per experimental results, LR achieved the highest accuracy of 87.5%, the highest f-score of 90.4%, while SGD provided the highest precision of 91.6%, and KNN provided the highest recall of 93.7%. SVM (87.4), SGD (87.2%), GB (85.3%), and XGBoost (85.4%) obtained the next highest accuracy, respectively, and very near to the accuracy of LR. DT (73.6%) and NB (79.3%) obtained the lowest accuracy, respectively. The performance of machine learning classifiers with BERT word embedding was also satisfactory, although machine learning classifiers performed better with the TF-IDF method.

Table 1. Performance of machine learning algorithms with TF-IDF feature extraction method

ML Algorithms	Accuracy	Precision	Recall	F-Score
SVM	0.9253	0.9473	0.9380	0.9426
Bagging	0.9244	0.9576	0.9253	0.9412
LR	0.9202	0.9500	0.9267	0.9382
SGD	0.9207	0.9686	0.9082	0.9374
XGBoost	0.9174	0.9763	0.8954	0.9341
Decision Tree	0.9129	0.9390	0.9271	0.9330
Random Forest	0.9092	0.9335	0.9272	0.9303
AdaBoost	0.8993	0.9632	0.8795	0.9194
GB	0.8970	0.9873	0.8534	0.9155
Naive Bayes	0.8356	0.8108	0.9764	0.8859
KNN	0.8170	0.8592	0.8611	0.8602

Table 2. Performance of machine learning algorithms with BERT feature extraction method

ML Algorithms	Accuracy	Precision	Recall	F-Score
LR	0.8746	0.9021	0.9067	0.9044
SVM	0.8741	0.9025	0.9051	0.9038
SGD	0.8725	0.9158	0.8865	0.9009
GB	0.8528	0.8743	0.9050	0.8894
XGBoost	0.8541	0.8825	0.8962	0.8893
Random Forest	0.8443	0.8600	0.9099	0.8843
KNN	0.8256	0.8212	0.9373	0.8754
AdaBoost	0.8244	0.8556	0.8798	0.8676
Bagging	0.8116	0.8596	0.8509	0.8552
Naive Bayes	0.7934	0.8905	0.7798	0.8315
Decision Tree	0.7359	0.7966	0.8005	0.7986

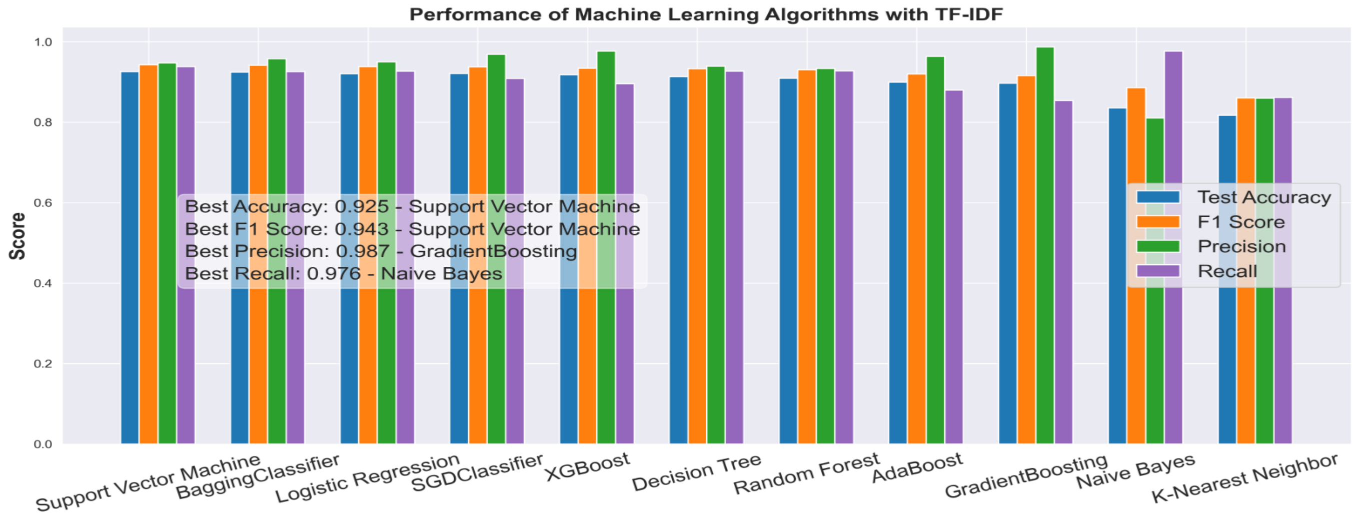

Fig. 4 Performance of machine learning algorithms with TF-IDF

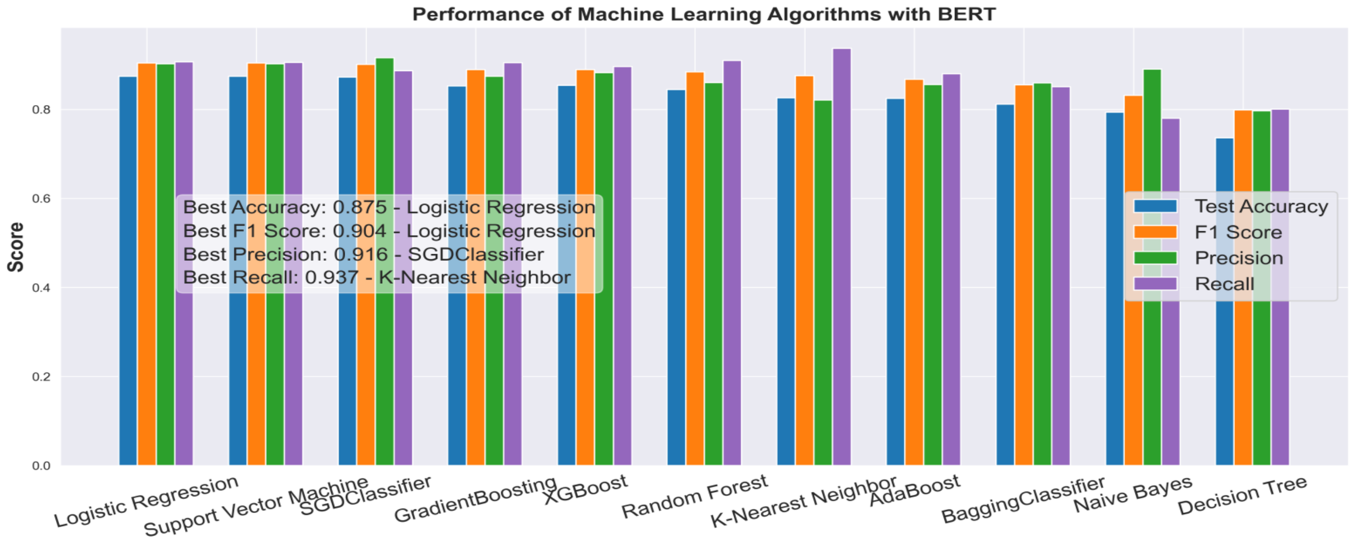

Fig. 5 Performance of machine learning algorithms with BERT

VII. CURRENT TRENDS AND FUTURE DIRECTION

Several related research papers between the year 2011 to 2021 were studied and reviewed to find the trends, methodology, and contributions of previous work performed by researchers to detect cyberbullying, cyberstalking, and other cyberharassment using machine learning techniques. As per the literature review in section IV, several social media networks and other online applications such as Twitter, Facebook, Youtube, Snapchat, Instagram, emails, and online conferencing tools 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 contents of other languages for cyberstalking detection. Twitter datasets are used by most researchers for cyberstalking detection, although other dataset sources, Formspring, YouTube, Instagram, Vine, Wikipedia, Myspace, Kongregate, and Kaggle, are also used. Machine learning techniques, especially SVM, LR, and boosting algorithms, are broadly utilized for cyberstalking and other cyberharassment detection. Experimental results from this paper and the literature show that the performance of machine learning algorithms is surpassing in cyberstalking detection. Machine learning algorithms outperform more conveniently with BOW, TF-IDF, and Word2Vec and additionally also can work with advanced word embedding-based feature extraction with better results. Researchers are additionally interested in unsupervised learning, reinforcement learning, deep learning, neural organization, and information mining procedures. Deep Learning is becoming more famous for the perplexing and enormous size of datasets. The researchers generally use BOW and TF-IDF for feature extraction, yet researchers are interested in other sentiment features, semantic features, and content, social features for legitimate detection. Several advanced and latest word embeddings-based and language model-based feature extraction methods such as GloVe, FastText, InferSent, ELMo, BERT, GPT-2, ALBERT are also popular among the researchers. Based on the review and survey, coming up next are the summed-up finding that gives new direction and scope towards research.

- Even if many researchers have worked to detect cyberbullying, cyberstalking, and other cyber harassment, more enhanced techniques are required to control the stalking downright.
- Many researchers focused on cyberstalking detection for textual data. Stalking through images, audio, and videos is becoming popular among adolescents. Cyberbullying, cyberstalking, and other cyberharassment detections on multimedia data(audio, video, image) would be another crucial research area.
- There is a high possibility for cyberstalking and other cyber harassment through audio, computerized programs, and mobile apps that would be another crucial research area that is still underestimated.
- Many researchers proposed models to detect cyberstalking and other cyberharassment after its occurrence; only a few studies focused on automatic detection and control of cyberstalking. Enhanced techniques are highly needed for automated detection and controlling of cyberstalking.
- Cyberstalking and other cyberharassment detection approach mainly focus on textual data in English and other foreign languages. Only very few studies for the cyberstalking detection on textual data in Indian languages such as Hindi, which have a vast number of users in the online world.
- 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.
- Behavioural patterns and meanings of stalker comments should be determined using semantic features and deep sentiment analysis for robust cyberstalking detection.
- Cyberstalking detection should be in a real-time manner. However, data gathering in real-time is challenging because various social media networks often do not provide data due to privacy policy.
- An email is a growing tool for cyberstalking and other cybercrimes. Often researchers focus only on phishing mail filtration and spam classification, but cyberstalkers are also harassing the victims using non-spam emails. More attention and research are required for cyberstalking through non-spam emails.
- Identification of cyberstalkers and fake social media accounts is a challenging task in cyberstalking. A global dictionary should be available online to store and access information that can help cyberstalking detection before its occurrence.

VIII. CONCLUSION

With the development in Internet innovation, online media and other web applications have become quickly well known among individuals. Cyberstalkers and other cybercriminals are making the negative and dread face of online media and other applications of virtual worlds. In this study, several related papers between the year 2011 to 2021 were selected to study and review previous research done in the field of cyberstalking, cyberbullying, and other cyber harassment. This study explored the different datasets, methodology, features extraction techniques, and classifiers techniques used by the researchers. In this paper, several popular machine learning algorithms were implemented with TF-IDF and BERT feature extraction methods based on the standard methodology of cyberstalking detection, and performance was measured. As per experimental results, it was found that machine learning classifies performed conveniently with basic and advanced feature extraction methods and provided better results. However, the performance of machine learning classifiers with TF-IDF was superior to the performance of ML classifiers with BERT. In the case of ML classifiers with TF-IDF, SVM provided the highest accuracy of 92.5%, the highest f-score of 94.3%, while GB achieved the highest precision of 98.7%, and NB achieved the highest recall of 97.6%. Bagging classifier (92.4%), LR (92%), SGD (92%), and XGBoost(91.7%) successfully provided the next highest accuracy respectively. In the case of ML classifiers with BERT, LR achieved the highest accuracy of 87.5%, the highest f-score of 90.4%, while SGD provided the highest precision of 91.6%, and KNN provided the highest recall of 93.7%. SVM (87.4) and SGD (87.2) provided the next highest accuracy, respectively. Literature survey and experimental results show that machine learning techniques are very relevant and capable of performing efficiently with better results in cyberstalking detection. This study will surely guide further research in this area and the scope of improvement. Research and review through this paper found that researchers have proposed various significant cyberstalking detection techniques, but existing approaches are not sufficient to completely detect cyberstalking cases, especially for multimedia content-based cyberstalking, real-time detection, and automatically cyberstalking detection. The future battle between cyberstalkers and researchers will be exciting and challenging for the safe use of social media and other internet applications.

REFERENCES

[1] (2021) The Statistics website [Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[2] P. E. Mullen, M. Pathé, R. Purcell, Stalking: New constructions of human behaviour, Australian and New Zealand Journal of Psychiatry, 35 (2021) 9–16.

[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) 1-10.

[4] (2015) The BBC News Website. [Online]. Available: https://www.bbc.com/news/world-asia-india-33532706

[5] M. Baer, Cyberstalking and the Internet Landscape We Have Constructed, Virginia Journal of Law & Technology, 154(15) (2020) 153-227.

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

[7] 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, paper 1 (2012) 356–364.

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

[9] Gautam, Arvind Kumar, and Abhishek Bansal. Performance Analysis of Supervised Machine Learning Techniques For Cyberstalking Detection In Social Media, Journal of Theoretical and Applied Information Technology 100(2) (2022).

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

[11] (2017) The Analytics Vidhya website. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

[12] (2017) The Towards Data Science website. [Online] Available: https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861

[13] W. A. Al-Khater, S. Al-Maadeed, A. A. Ahmed, A. S. Sadiq and M. K. Khan, Comprehensive Review of Cybercrime Detection Techniques, in IEEE Access, 8 (2020) 137293-137311.

[14] Tolba, Marwa, Salima Ouadfel, and Souham Meshoul, Hybrid ensemble approaches to online harassment detection in highly imbalanced data, Expert Systems with Applications 175 (2021) 114751.

[15] Sadiq, S., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S., and B. W. Aggression detection through the deep neural model on Twitter, Future Generation Computer Systems, 114, (2021) 120-129.

[16] Ayo, F. E., Folorunso, O., Ibharalu, F. T., Osinuga, I. A., and Abayomi-Alli, A, A probabilistic clustering model for hate speech classification in Twitter, Expert Systems with Applications, 173, (2021) 114762.

[17] Bini, Stefano A. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? The Journal of arthroplasty 33(8) (2018) 2358-2361.

[18] Singhal, Paridhi, and Ashish Bansal Improved textual cyberbullying detection using data mining, International Journal of Information and Computation Technology 3(6) (2013) 569-576.

[19] A. Derhab, A. Bouras, F. B. muhaya, M. K. Khan and Y. Xiang, Spam trapping system: Novel security framework to fight against spam botnets, Proc. 21st Int. Conf. Telecommun. (ICT), paper (2014) 467-471.

[20] J. F. Peters, Foundations of Computer Vision: Computational Geometry Visual Image Structures and Object Shape Detection, Berlin, Germany: Springer, (2017).

[21] K. Dinakar, Modeling the Detection of Textual Cyberbullying, Proceedings of the International AAAI Conference on Web and Social Media, paper 5(3) (2011) 11–17.

[22] Kelly Reynolds, April Kontostathis, Lynne Edwards, Using Machine Learning to Detect Cyberbullying, IEEE 10th International Conference on Machine Learning and Applications, paper 2 (2011) 241–244.

[23] M. Dadvar, R. Ordelman, F.D. Jong, D.Trieschnigg, Towards User Modelling in the Combat against Cyberbullying, in Natural Language Processing and Information Systems, Springer-Verlag Berlin Heidelberg, paper (2012) 277–283.

[24] Maral Dadvar, Franciska de Jong, Roeland Ordelman, Dolf Trieschnigg, Improved Cyberbullying Detection Using Gender Information, 12th -Dutch-Belgian Information Retrieval Workshop, paper (2012) 693–696.

[25] Vinita Nahar, Sayan Unankard, Xue Li, Caoyi Pang, Sentiment Analysis for Effective Detection of Cyber Bullying, Proceedings of the 14th Asia-Pacific international conference on Web Technologies and Applications, paper (2012).

[26] Vinita Nahar, Xue Li, Chaoyi Pang, Yang Zhang, Cyberbullying Detection based on Text-Stream Classification, Proceedings of the 11th Australasian Data Mining Conference, Canberra, Australia, paper (2013).

[27] Maral Dadvar, Dolf Trieschnigg, Roeland Ordelman, and Franciska de Jong, Improving cyberbullying detection with user context, Proceedings of the European Conference on Information Retrieval, paper (2013) 693-696.

[28] Vivek K. Singh, Qianjia Huang, Pradeep K. Atrey, Cyber Bullying Detection Using Social and Textual Analysis, 3rd International Workshop on Socially-Aware Multimedia - SAM, paper (2014).

[29] Cicero Nogueira, Dos Santos, MairaGatti, Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts, International Conference on Computational Linguistics, paper (2014).

[30] Z. Ghasem, I. Frommholz, and C. Maple, Machine learning solutions for controlling cyberbullying and cyberstalking, International Journal of Security, 6(2) (2015) 55-64.

[31] 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), paper (2015) 20.

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

[33] 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) 127–135.

[34] Michele Di Capua, Emanuel Di Nardo, Alfredo Petrosino, Unsupervised Cyber Bullying Detection in Social Networks, Proceedings of the 23rd International Conference on Pattern Recognition (ICPR) Cancún Center, México, paper (2016).

[35] Vivek K. Singh, Qianjia Huang, Pradeep K. Atrey, Cyberbullying detection using probabilistic socio-textual information fusion, Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASNAM), paper (2016).

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

[37] Walisa Romsaiyud, Kodchakorna Nakornphanom, Pimpaka Prasertsilp, Piyaporn Nurarak, and Pirom Konglerd, Automated cyberbullying detection using clustering appearance patterns, In Knowledge and Smart Technology (KST), IEEE 9th International Conference on, paper (2017) 242–247.

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

[39] 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, paper (2018).

[40] Sweta Agrawal, Amit Awekar, Deep Learning for Detecting Cyberbullying Across Multiple Social Media Platforms, Springer International Publishing AG, part of Springer Nature, (2018) 141–153.

[41] Amanpreet Singh, Maninder Kaur, Content-based Cybercrime Detection: A Concise Review, International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8(8) (2019) 1193-1207.

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

[43] 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, 10(5) (2019).

[44] V. Balakrishnan, S. Khan, H.R. Arabnia, Improving cyberbullying detection using Twitter users' psychological features and machine learning, Science Direct, ELSEVIER, Computer & Security, 90 (2020) 101710.

[45] Manowarul Islam, Selina Sharmin, Cyberbullying Detection on Social Networks Using Machine Learning Approaches, IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), paper (2020).

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

[47] Anant Khandelwal and N. Kumar, A Unified System for Aggression Identification in English Code-Mixed and Uni-Lingual Texts, in Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, paper (2020) 55–64.

[48] A. Asante and X. Feng, Content-Based Technical Solution for Cyberstalking Detection, 3rd International Conference on Computer Communication and the Internet (ICCCI), paper (2021) 89-95.

[49] N. Dughyala, S. Potluri, S. KJ and V. Pavithran, Automating the Detection of Cyberstalking, 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), paper (2021) 887-892.

[50] Doppala B.P., NagaMallik Raj S., Stephen Neal Joshua E., Thirupathi Rao N., Automatic Determination of Harassment in Social Network Using Machine Learning. In: Saha S.K., Pang P.S., Bhattacharyya D., Smart Technologies in Data Science and Communication, Lecture Notes in Networks and Systems, Springer, Singapore, paper 210 (2021).

[51] Jain, V., Kumar, V., Pal, V., & Vishwakarma, D. K., Detection of Cyberbullying on Social Media Using Machine learning, 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, paper (2021) 1091-1096.

[52] Pericherla, Subbaraju, and E. Ilavarasan, Performance analysis of Word Embeddings for Cyberbullying Detection. IOP Conference Series: Materials Science and Engineering, paper 1085(1) (2021).

[53] Raj, Chahat, et al. Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques, Electronics 22(10) (2021) 2810.

[54] 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) 7-16.

[55] (2018) The Towards Data Science website. [Online] Available: https://towardsdatascience.com/all-you-need-to-know-about-text-preprocessing-for-nlp-and-machine-learning-bc1c5765ff67.

[56] Kadhim, Ammar Ismael, An evaluation of pre-processing techniques for text classification, International Journal of Computer Science and Information Security (IJCSIS) 16(6) (2018) 22-32.

[57] Rui, Weikang, Kai Xing, and Yawei Jia, BOWL: Bag of word clusters text representation using word embeddings, International Conference on Knowledge Science, Engineering and Management. Springer, Cham, paper (2016).

[58] (2020) Medium website. [Online]. Available: https://medium.com/@kashyapkathrani/all-about-embeddings-829c8ff0bf5b

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

[60] Anand, Mukul, and R. Eswari, Classification of abusive comments in social media using deep learning, 3rd international conference on computing methodologies and communication (ICCMC). IEEE, paper (2019).

[61] K. Wang, Y. Cui, J. Hu, Y. Zhang, W. Zhao, & L. Feng, Cyberbullying detection based on the fast text and word similarity schemes, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), paper 20(1) (2020) 1-15.

[62] Swamy, Steve Durairaj, Anupam Jamatia, and Björn Gambäck. Studying generalisability across abusive language detection datasets. Proceedings of the 23rd conference on computational natural language learning (CoNLL), paper (2019).

[63] Hoang Tran, Loc, Tuan Tran, and An Mai, Text classification problems via BERT embedding method and graph convolutional neural network, arXiv e-prints, (2021) 2111.

[64] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, & R. Soricut, Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942 (2019).

[65] K. Clark, M.T. Luong, Q.V. Le, & C. D. Manning, Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555 (2020).

[66] Ethayarajh, Kawin. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings, arXiv preprint arXiv:1909.00512 (2019).

[67] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, & Q.V. Le, Xlnet: Generalized autoregressive pre-training for language understanding, Advances in neural information processing systems, 32 (2019).

[68] Liu, Yinhan, et al. Roberta: A robustly optimized bert pre-training approach. arXiv preprint arXiv: 1907.11692 (2019).

[69] Reimers, Nils, and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv: 1908.10084 (2019).

[70] S. Tomkins, L. Getoor, Y. Chen, & Y. Zhang, A socio-linguistic model for cyberbullying detection, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, (2018) 53-60.

[71] Modha, Sandip, Prasenjit Majumder, and Thomas Mandl, An empirical evaluation of text representation schemes to filter the social media stream, Journal of Experimental & Theoretical Artificial Intelligence, (2021) 1-27.

[72] Paul, Sayanta, Sriparna Saha, and Mohammed Hasanuzzaman. Identification of cyberbullying: A deep learning-based multimodal approach. Multimedia Tools and applications (2020) 1-20.

[73] Samir Kumar Bandyopadhyay, Payal Bose, Amiya Bhaumik, Sandeep Poddar, Machine Learning and Deep Learning Integration for Skin Diseases Prediction International Journal of Engineering Trends and Technology 70(2) (2022) 11-18.

[74] Parita Shah, Priya Swaminarayan, Maitri Patel, Nimisha Patel, Sentiment Analysis on Movie Reviews in Regional Language Gujarati Using Machine Learning Algorithm, International Journal of Engineering Trends and Technology 70(1) (2022) 313-326.

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

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

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

[78] Breiman L. Random forests. Machine learning, 45(1) (2001) 5-32

[79] Govindan, Vithyatheri, and Vimala Balakrishnan, A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection, Journal of King Saud University-Computer and Information Sciences (2022).

[80] Gumaei, Abdu, et al. An effective approach for rumour detection of Arabic tweets using extreme gradient boosting method. Transactions on Asian and Low-Resource Language Information Processing 21(1) (2022) 1-16.

[81] Li, Bin, Qingzhao Yu, and Lu Peng. Ensemble of fast learning stochastic gradient boosting, Communications in Statistics-Simulation and Computation 51(1) (2022) 40-52.

[82] C. Raj, A. Agarwal, G. Bharathy, B. Narayan, & M. Prasad, Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques, Electronics, 22(10) (2021) 2810.

[83] (2020) Cyberbullying datasets - Mendeley website. [Online]. Available: https://data.mendeley.com/datasets/jf4pzyvnpj/1

[84] (2020) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/mrmorj/hate-speech-and-offensive-language-dataset

[85] (2022) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/andrewmvd/cyberbullying-classification

[86] (2021) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/sanamps/toxiccommentclassification

[87] (2014) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/c/detecting-insults-in-social-commentary/data