Conference Proceeding Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges

[1]P.Priya (M.Sc) and [2]Prof K. R. Aruna (Msc., M.Phil., M.Tech., B.Ed.),
[1,2]Department of Computer Science, Kamban College of Arts and Science, for Women, Tiruvannamalai, India

Abstract: Prior to the innovation of Information Communication Technologies (ICT), social interactions evolved within small cultural boundaries, such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications. These social technologies have created a revolution in user-generated information, online human networks, and rich human behaviorrelated data. However, the misuse of social technologies, such as social media (SM) platforms, has introduced a new form of aggression and violence that occurs exclusively online. A new means of demonstrating aggressive behavior in SM websites is highlighted in this work. The motivations for the construction of prediction models to fight aggressive behavior in SM are also outlined. We comprehensively review cyberbullying prediction models and identify the main issues related to the construction of cyberbullying prediction models in SM. This paper provides insights on the overall process for cyberbullying detection and most importantly overviews the methodology. Though data collection and feature engineering process has been elaborated, yet most of the emphasis is on feature selection algorithms and then using various machine learning algorithms for prediction of cyberbullying behaviors. Finally, issues and challenges have been highlighted as well, which present new research directions for researchers to explore.

I. INTRODUCTION

Uncover hidden knowledge through deep Machine or deep learning algorithms help learning from raw data . Big data analytics researchers understand big data . Abundant has improved several applications, and information on humans and their societies forecasting the future has even become can be obtained in this big data era, but this possible through the combination of big data acquisition was previously impossible .One and machine learning algorithms .of the main sources of human-related data is An insightful analysis of data on human social media (SM). By applying machine behavior and interaction to detect and learning algorithms to SM data, we can restrain aggressive behavior involves exploit historical data to predict the future of multifaceted angles and aspects and the a wide range of applications. Machine merging of theorems and techniques from learning algorithms provide an opportunity multidisciplinary and interdisciplinary to effectively predict and detect negative fields. The accessibility of large-scale data forms of human behavior, such as produces new research questions, novel cyberbullying . Big data analysis can computational methods, interdisciplinary approaches, and outstanding opportunities to discover several vital inquiries quantitatively. However, using traditional methods (statistical methods) in this context is challenging in terms of scale and accuracy. These methods are commonly based on organized data on human behavior and small-scale human networks (traditional social networks).

OSNs provide criminals with tools to perform aggressive actions and networks to commit misconduct. Therefore, methods that address both aspects (content and network) should be optimized to detect and restrain aggressive behavior in complex systems.

A. Rise of Aggressive Behavior

SM Prior to the innovation of communication technologies, social interaction evolved within small cultural boundaries, such as locations and families .The recent development of communication technologies exceptionally transcends the temporal and spatial limitations of traditional communication. In the last few years, online communication has shifted toward userdriven technologies, such as SM websites, blogs, online virtual communities, and online sharing platforms. SM websites have become dynamic social communication websites for millions of users worldwide. Data in the form of ideas, opinions, preferences, views, and discussions are spread among users rapidly through online social communication. The online interactions of SM users generate a huge volume of data that can be utilized to study human behavioral patterns .SM websites also provide an exceptional opportunity to analyze patterns of social interactions among populations at a scale that is much larger than before. Aside from renovating the means through which people are influenced, SM websites provide a place for a severe form of misbehavior among users. Online complex networks, such as SM websites, changed substantially in the last decade, and this change was stimulated by the popularity of online communication through SM websites. Online communication has become an entertainment tool, rather than serving only to communicate and interact with known and unknown users.

B. Motivations for Predicting Aggressive Behavior on SM Websites

Many studies have been conducted on the contribution of machine learning algorithms to OSN content analysis in the last few years. Machine learning research has become crucial in numerous areas and successfully produced many models, tools, and algorithms for handling large amounts of data to solve real-world problems, Machine learning algorithms have been used extensively to analyze SM website content for spam, phishing, and cyberbullying prediction. Aggressive behavior includes spam propagatioN, phishing, malware spread, and cyberbullying. Textual cyberbullying has become the dominant aggressive behavior in SM websites because these websites give users full freedom to post on their platforms.

SM websites contain large amounts of text and/or non-text content and other information related to aggressive behavior. In this work, a content analysis of SM websites is performed to

predict aggressive behavior. Such an analysis is limited to textual OSN content for predicting cyberbullying behavior. Given that cyberbullying can be easily committed, it is considered a dangerous and fast-spreading aggressive behavior. Bullies only require willingness and a laptop or cell phone with Internet connection to perform misbehavior without confronting victims .The popularity and proliferation of SM websites have increased online bullying activities. Cyberbullying in SM websites is rampant due to the structural characteristics of SM websites. Cyberbullying in traditional platforms, such as emails or phone text messages, is performed on a limited number of people. SM websites allow users to create profiles for establishing friendships and communicating with other users regardless of geographic location, thus expanding cyberbullying beyond physical location. Anonymous users may also exist on SM websites, and this has been confirmed to be a primary cause for increased aggressive user behaviour.

C. Why Constructing Cyberbullying Prediction Models Is important

The motivations for carrying out this review for predicting cyberbullying on SM websites are discussed as follows. Cyberbullying is a major problem and has been documented as a serious national health problem due to the recent growth of online communication and SM websites. Research has shown that cyberbullying exerts negative effects on the psychological and physical health and academic performance of people . Studies have also shown that cyberbullying victims incur a high risk of suicidal ideation .Other studies reported an association between cyberbullying victimization and suicidal ideation risk. Consequently, developing a cyberbullying prediction model that detects aggressive behavior that is related to the security of human beings is more important than developing a prediction model for aggressive behavior related to the security of machines.

Cyberbullying can be committed anywhere and anytime. Escaping from cyberbullying is difficult because cyberbullying can reach victims anywhere and anytime. It can be committed by posting comments and statuses for a large potential audience. The victims cannot stop the spread of such activities .Although SM websites have become an integral part of users' lives, a study found that SM websites are the most common platforms for cyberbullying victimization. A wellknown characteristic of SM websites, such as Twitter, is that they allow users to publicly express and spread their posts to a large audience while remaining anonymous. The effects of public cyberbullying are worse than those of private ones, and anonymous scenarios of cyberbullying are worse than nonanonymouscases. Consequently, the severity of cyberbullying has increased on SM websites, which support public and anonymous scenarios of cyberbullying. These characteristics make SM websites, such as Twitter, a dangerous platform for committing cyberbullying. Recent research has indicated that most experts favor the automatic monitoring of cyberbullying. A study that examined 14 groups of adolescents confirmed the urgent need for automatic monitoring and prediction models for cyberbullying because traditional strategies for coping with cyberbullying in the era of big data and networks do not work well. Moreover, analyzing large amounts of complex data requires machine learning-based automatic monitoring.

1) Cyberbullying on SM Websites Most researchers define cyberbullying as using electronic communication technologies to bully people .Cyberbullying may exist in different types or forms, such as writing aggressive posts, harassing or bullying a victim, making hateful posts, or insulting the victim. Given that cyberbullying can be easily committed, it is considered a dangerous and fastspreading aggressive behavior. Bullies only require willingness and a laptop or cell phone connected to the Internet to perform misbehavior without confronting the victims . The popularity and proliferation of SM websites have increased online bullying activities. Cyberbullying on SM websites is performed on a large number of users due to the structural characteristics of SM websites.

Cyberbullying in traditional platforms, such as emails or phone text messages, is committed on a limited number of people. SM websites allow users to create profiles for establishing friendships and interacting with other online users regardless of geographic location, thus expanding cyberbullying beyond physical location. Moreover, anonymous users may exist on SM websites, and this has been confirmed to be a primary cause of increased aggressive user behavior . The nature of SM websites allows cyberbullying to occur secretly, spread rapidly, and continue easily .Consequently, developing an effective prediction model for predicting cyberbullying is of practical significance. SM websites contain large amounts of text and/or non-text content and information related to aggressive behavior.

D. Methodology

This section presents the methodology used in this work for a literature search. Two phases were employed to retrieve published papers on cyberbullying prediction models. The first phase included searching for reputable academic databases and search engines. The search engines and academic databases used for the retrieval of relevant papers were as follows: Scopus, Clarivate Analytics' Web of Science.

II. PREDICTING CYBERBULLYINGON SOCIAL MEDIA IN THE BIG DATA ERA USING MACHINE LEARNING ALGORITHMS

Our world is currently in the big data era because quintillion bytes of data are generated daily . Organizations continuously generate large-scale data. These large-scale datasets are generated from different sources, including the World Wide Web, social networks, and sensor networks. Big data have nine characteristics, namely, volume, variety, variability and complexity, velocity, veracity, value, validity, verdict, and visibility .For example, Flickr generates almost 3.6 TB of data, Google is believed to process almost 20,000 TB of data per day, and the Internet gathers an estimated 1.8 PB of data daily. SM is an online platform that provides users an opportunity to create an online community, share information, and exchange content. SM users and the interaction among organizations, people, and products are responsible for the massive amount of data generated on SM platforms. SM platforms, such as Facebook, YouTube, blogs, Instagram, Wikipedia, and Twitter, are of different types. The data generated by SM outlets can be structured or unstructured in form. SM analytics is the analysis of structured and unstructured data generated by SM outlets. SM analytics can be in any of the following forms: link prediction, community, content, social influence, structured, and unstructured. SM is now in the big data era. For example, Facebook stores 260 billion photographs in over 20 PB of storage space, and up to one million pictures are processed per second. YouTube receives 100 hours of downloaded videos in each minute.

The most common means of constructing cyberbullying prediction models is to use a text classification approach that

involves the construction of machine learning classifiers from labeled text instances. Another means is to use a lexicon-based model that involves computing orientation for a document from the semantic orientation of words or phrases in the document. Generally, the lexicon in lexicon-based models can be constructed manually (similar to the approaches used in) or automatically by using seed words to expand the list of words. However, cyberbullying prediction using the lexiconbased approach is rare in literature. The primary reason is that the texts on SM websites are written in an unstructured manner, thus making it difficult for the lexicon-based approach to detect cyberbullying based only on lexicons. However, lexicons are used to extract features, which are often utilized as inputs to machine learning algorithms. For example, lexiconbased approaches, such as using a profane-based dictionary to detect the number of profane words in a post, are adopted as profane features to machine learning models .The key to effective cyberbullying prediction is to have a set of features that are extracted and engineered. Features and their combinations are crucial in the construction of effective cyberbullying prediction models. Most studies on cyberbullying prediction used machine learning algorithms to construct cyberbullying prediction models. Machine learning-based models exhibit decent performance in cyberbullying prediction Consequently, this work reviews the construction of cyberbullying prediction models based on machine learning. The machine learning field focuses on the development and application of computer algorithms that improve with experience. The objective of machine learning is to identify and define the patterns and correlations between data. The importance of analyzing big data lies in discovering hidden knowledge through deep learning from raw data.

A. Data Collection

Data are important components of all machine learning-based prediction models.

However, data (even "Big Data") are useless on their own until knowledge or implications are extracted from them. Data extracted from SM websites are used to select training and testing datasets.

Supervised prediction models aim to provide computer techniques to enhance prediction performance in defined tasks on the basis of observed instances (labeled data). Machine learning models for a certain task primarily aim to generalize; a successful model should not be limited to examples in a training dataset only but must include unlabeled real data. Data quantity is inconsequential; what is crucial is whether or not the extracted data represent activities on SM websites well. The main data collection strategies in previous cyberbullying prediction studies on SM websites can be categorized into data extracted from SM websites by using either keywords, that is, words, phrases, or hashtags or by using user profiles . The issues in these data collection strategies and their effects on the performance of machine learning algorithms are highlighted in the Data Collection section (related issues).

B. Feature Engineering

Feature is a measurable property of a task that is being observed. The main purpose of engineering feature vectors is to provide machine learning algorithms with a set of learning vectors through which these algorithms learn how to discriminate between different types of classes. Feature engineering is a key factor behind the success and failure of most machine learning models. The success and failure of prediction may be based on several elements. The most significant element is the features used to train the model. Most of the effort in constructing cyberbullying prediction models using learning algorithms is devoted to this task. In this context, the design of the input space (i.e., features and their combinations that are provided as an input to the classifier) is vital.

Proposing a set of discriminative features, which are used as inputs to the machine learning classifier, is the main step toward constructing an effective classifier in many applications. Feature sets can be created based on human-engineered observations, which rely on how features correlate with the occurrences of classes. For example, recent cyberbullying studies established the correlation between different variables, such as age, gender, and user personality, and cyberbullying occurrence. These observations can be engineered into a practical form (feature) to allow the classifier to discriminate between cyberbullying and noncyberbullying and can thus be used to develop effective cyberbullying prediction models. Proposing features is an important step toward improving the discrimination power of prediction models. Similarly, proposing a set of significant features of cyberbullying engagement on SM websites is important in developing effective prediction models based on machine learning algorithms. State-of-the-art research has developed features to improve the performance of cyberbullying prediction. For example, a lexical syntactic feature has been proposed to deal with the prediction of offensive language; this method is better than traditional learningbased approaches in terms of precision. Dadvar et al. (2012) examined gender information from profile information and developed a gender-based approach for cyberbullying prediction by using datasets from Myspace as a basis. The gender feature was selected to improve the discrimination capability of a classifier. Age and gender were included as features in other studies, but these features are limited to the information provided by users in their online profiles. Several studies focused on cyberbullying prediction based on profane words as a feature. Similarly, a lexicon of profane words was constructed to indicate bullying, and these words were used as features for input to machine learning algorithms. Using profane words as features demonstrates a significant improvement in model performance. For example, the number of "bad" words and the density of "bad" words were proposed as features for input to machine learning in a previous work.

1. Chi-Square Test

Another common feature selection model is the chisquare test. This test is used in statistics, among other variables, to test the independence of two occurrences. In feature selection, chi-square is used to test whether the occurrences of a feature and class are independent. Thus, the following quantity is assumed for each feature, and they are ranked by their score.

C. Machine Learning Algorithms

Many types of machine learning algorithms exist, but nearly all studies on cyberbullying prediction in SM websites used the most established and widely used type, that is, supervised machine learning algorithms

.The accomplishment of machine learning algorithms is determined by the degree to which the model accurately converts various types of prior observation or knowledge about the task. Much of the practical application of machine learning considers the details of a particular problem. Then, an algorithmic model that allows for the accurate encoding of the

facts is selected. However, no optimal machine learning algorithm works best for all problems. Therefore, most researchers selected and compared many supervised classifiers to determine the ideal ones for their problem. Classifier selection is generally based on the most commonly used classifiers in the field and the data features available for experiments. However, researchers can only decide which algorithms to adopt for constructing a cyberbullying prediction model by performing a comprehensive practical experiment as a basis. Table 2 summarizes the commonly used machine learning algorithms for constructing cyberbullying prediction models.

III. ISSUES RELATED TO CONSTRUCTING CYBERBULLYING PREDICTION MODELS

In this section, the issues identified from the reviewed studies are discussed. The main issues related to cyberbullying definition, data collection feature engineering, and evaluation metric selection are identified and discussed in following subsections.

A. Issues Related to Cyberbullying

Definition Traditional bullying is generally defined as "intentional behavior to harm another, repeatedly, where it is difficult for the victim to defend himself or herself" [127]. By extending the definition of traditional bullying, cyberbullying has been defined [90] as "an aggressive behavior that is achieved using electronicplatforms by a group or an individual repeatedly and over time against a victim who cannot easily defend him or herself. These issues have been discussed by researchers to simplify the concept of cyberbullying in the online context. First, the concept of repetitive act in cyberbullying is not as straightforward as that in SM [47]. For example, SM websites can provide cyberbullies a medium to propagate cyberbullying posts for a large population. Consequently, a single act by one committer may become repetitive over time [47]. Second, power imbalance is presented in different forms in online communication. Researchers [128] have suggested that the content in online environments is difficult to eliminate or avoid, thus making a victim powerless.

These definitional aspects are under intense debate, but to simplify the definition of cyberbullying and make this definition applicable to a wide range of applications.

B. Data Collection

Many cyberbullying prediction studies extracted their datasets by using specific keywords or profile IDs. Nevertheless, by simply tracking posts that have particular keywords, these researches may have presented potential sampling bias, limited the prediction to posts that contain the predefined keywords, and overlooked many other posts relevant to cyberbullying. Such data collection methods limit the prediction model of cyberbullying to specified keywords. The identification of keywords for extracting posts is also subject to the author's understanding of cyberbullying. Extracting wellrepresentative data from SM is the first step toward building effective machine learning prediction models. However, SM websites' public application program interface (API) only allows the extraction of a small sample of all relevant data and thus poses a potential for sampling bias. With these points in mind, researchers should ensure minimum bias as much as possible when they extract data to guarantee that the examples selected to be represented in training data are generalized and provide an effective model when applied to testing data. Bias in data collection can impose bias in the selected training dataset based on specific keywords or users, and such a bias consequently introduces overfitting issues that affect the capability of a machine learning model to make reliable predictions on untrained data.

C. Feature Engineering

Features are vital components in improving the effectiveness of machine learning prediction models [79].Most of the discussed studies attempted to provide effective machine learning solutions to cyberbullying on SM websites by providing significant features. However, these studies overlooked other important features. For example, online cyberbullies may dynamically change the way they use words and acronyms. SM websites help create cyberbullying acronyms that have not been commonly used in committing traditional bullying or are beyond SM norms. Recent survey response studies (questionnaire-based studies) have reported positive correlations between different variables, such as personality and sociability of a user in an online environment, and cyberbullying occurrences. The observations of these studies are important in understanding such behavior in online environments. However, these observations are yet to be used as features with machine learning algorithms to provide significant models. These observations can be useful when transformed to a practical form (features) that can be employed to develop effective machine learning prediction models for cyberbullying on SM websites. The abundant information provided by SM websites should be utilized to convert observations into a set of features. For example, two studies [17, 96] attempted to improve machine learning classifier performance by including features, such as age and gender, that show improvement in classifier performance, but these features are extracted from direct user details mentioned in the online profiles of users. However, most studies found that only a few users provide complete details in their online. A significant correlation has also been found between sociability of a user and cyberbullying engagement in online environments [131]. Users who are highly active in online environments are likely to engage in cyberbullying [134]. According to these observations, SM websites possess features that can be used as signals to measure the sociability of a user, such as number of friends, number of posts, URLs in posts, hashtags in posts, and number of users engaged in conversations (mentioned). The combination of these features with traditionally used ones, such as profanity features, can provide comprehensive discriminative features. The reviewed studies (Table 1) focused on using either a traditional feature model (e.g., bag-of-words) or information (e.g., age or gender) limited to user profile information (information written by users in their profile). Given that such information is limited, comprehensive features should be proposed to improve classifier performance.

D. Machine Learning Algorithm Selection

A machine learning algorithm is selected to be trained on proposed features. However, deciding which classifier performs best for a specific dataset is difficult. More than one machine learning algorithm should be tested to determine the best machine learning algorithm for a specific dataset. Three points may be used as guide to narrow the selection of machine learning algorithms to be tested. First, a specific literature on machine learning for cyberbullying detection is important in selecting a specified classifier.

The pre-eminence of the classifier may be circumscribed to a given domain [135]. Therefore, general previous research and findings on machine learning can used as a guide to select a

machine learning algorithm. Second, a literature review of text mining .

E. Imbalanced Class Distribution

In many cases of real data, datasets naturally have imbalanced classes in which the normal class has a large number of instances and the abnormal class has a small number of instances in the dataset. Abnormal class instances are rare and difficult to be collected from realworld applications. Examples of imbalanced data applications are fraud detection, instruction detection, and medical diagnosis. Similarly, the number of cyberbullying posts is expected to be much less than the number of non-cyberbullying posts, and this assumption generates an imbalanced class distribution in the dataset in which the instances of non-cyberbullying contain much more posts than those of cyberbullying. Such cases can prevent the model from correctly classifying the examples. Many methods have been proposed to solve this issue.

F. Evaluation Metric Selection

Accuracy, precision, recall, and AUC are commonly used as evaluation metrics [19, 38]. Evaluation metric selection is important. The selection is based on the nature of manually labeled data. Selecting an inappropriate evaluation metric may result in better performance according to the selected evaluation metric. Then, the researcher may find the results to be significantly improved, although an investigation of how the machine learning model is evaluated may produce contradicting results and may not truly reflect the improvement of performance. For example, cyberbullying posts are commonly considered abnormal cases, whereas noncyberbullying posts are considered normal cases. The ratio between cyberbullying and non-cyberbullying is normally large. Generally, non-cyberbullying posts comprise a large portion. For example, 1000 posts are manually labeled as cyberbullying and non-cyberbullying. The non-cyberbullying posts are 900, and the remaining 100 posts are cyberbullying. If a machine learning classifier classifies all 1000 posts as non cyberbullying and is unable to classify any posts as cyberbullying, then this classifier is considered impractical. By contrast, if researchers use accuracy as the main evaluation metric, then the accuracy of this classifier calculated as mentioned in the accuracy equation will yield a high accuracy percentage.

In the example, the classifier fails to classify any cyberbullying posts but obtains a high accuracy percentage. Knowing the nature of manually labeled data is important in selecting an evaluation metric. In cases where data are imbalanced, researchers may need to select AUC as the main evaluation metric. In class-imbalance situations, AUC is more robust than other performance metrics [141]. Cyberbullying and non-cyberbullying data are commonly imbalanced datasets (non-cyberbullying posts outnumber the cyberbullying ones) that closely represent the reallife data that machine learning algorithms need to train on. Accordingly, the learning performance of these algorithms is independent of data skewness [73]. Special care should be taken in selecting the main evaluation metric to avoid uncertain results and appropriately evaluate the performance of machine learning algorithms.

IV. ISSUES AND CHALLENGES

This section presents the issues and challenges while guiding future researchers to explore the domain of sentiment analysis through leveraging machine learning algorithms and models for detecting cyberbullying through social media.

A. Human Data Characteristics

Although SM big data provide insights into large human behavior data, in reality, the analysis of such big data remains subjective [142]. Building human prediction systems involves steps where subjectivity about human behavior does exist. For example, when creating a manually labeled dataset to train a machine learning algorithm to predict cyberbullying posts, human bias may exist based on how cyberbullying is being defined and the criteria used to categorize the text as cyberbullying text.

Predicting human behavior is crucial but complex. To achieve an effective prediction of human behavior, the patterns that exist and are used for constructing the prediction model should also exist in the future input data. The patterns should clearly represent features that occur in current and future data to retain the context of the model. Given that big data are not generic and dynamic in nature, the context of these data is difficult to understand in terms of scale and even more difficult to maintain when data are reduced to fit into a machine learning model. Handling context of big data is challenging and has been presented as an important future direction.

B. Culture Effect

What was considered cyberbullying yesterday might not be considered cyberbullying today, and what was previously considered cyberbullying may not be considered cyberbullying now due to the introduction of OSNs. OSNs have a globalized culture. However, machine learning always learns from the examples provided. Consequently, designing different examples that represent a different culture remains to be defined, and robust work from different disciplines is required. For this purpose, cross disciplinary coordination is highly desirable.

C. Language Dynamics

Language is quickly changing, particularly among the young generation. New slang is regularly integrated into the language culture. Therefore, researchers are invited to propose dynamic algorithms to detect new slang and abbreviations related to cyberbullying behavior on SM websites and keep updating the training processes of machine learning algorithms by using newly introduced words.

D. Prediction of Cyberbullying Severity

The level of cyberbullying severity should be determined. The effect of cyberbullying is proportional to its severity and spread. Predicting different levels of cyberbullying severity does not only require machine learning understanding but also a comprehensive investigation to define and categorize the level of cyberbullying severity from social and psychological perceptions. Efforts from different disciplines are required to define and identify the levels of severity then introduce related factors that can be converted into features to build multiclassifier machine learning for classifying cyberbullying severity into different levels as opposed to a binary classifier that only detects whether an instance is cyberbullying or not.

E. Unsupervised Machine Learning

Human learning is essentially unsupervised. The structure of the world was discovered by observing it and not by being told the name of every objective. Nevertheless, unsupervised machine learning has been overshadowed by the success of supervised learning [143]. This gap in literature may be caused by the fact that nearly all current studies rely on

manually labeled data as the input to supervised algorithms for classifying classes. Thus, finding patterns between two classes by using unsupervised grouping remains difficult. Intensive research is required to develop unsupervised algorithms that can detect effective patterns from data. Traditional machine learning algorithms lack the capability to handle cyberbullying big data.

The traditional machine learning algorithms pointed out in this survey lacks the capability to process big data in a standalone format. Big data have rendered traditional machine learning algorithms impotent. Cyberbullying big data generated from SM require advanced technology for the processing of the generated data to gain insights and help in making intelligent decisions.

CONCLUSION

This study reviewed existing literature to detect aggressive behavior on SM websites by using machine learning approaches. We specifically reviewed four aspects of detecting cyberbullying messages by using machine learning approaches, namely, data collection, feature engineering, construction of cyberbullying detection model, and evaluation of constructed cyberbullying detection models. Several types of discriminative features that were used to detect cyberbullying in online social networking sites were also summarized. In addition, the most effective supervised machine learning classifiers for classifying cyberbullying messages in online social networking sites were identified. One of the main contributions of current paper is the definition of evaluation metrics to successfully identify the significant parameter so the various machine learning algorithms can be evaluated against each other. Most importantly we summarized and identified the important factors for detecting cyberbullying through machine learning techniques specially supervised learning. For this purpose, we have used accuracy.

REFERENCES

[1] Subrahmanian, V. and S. Kumar, *Predicting human behavior: The next frontiers.* Science, 2017. 355(6324): p. 489-489.

[2] Lauw, H., et al., *Homophily in the digital world: A LiveJournal case study.* Internet Computing, IEEE, 2010. 14(2): p. 15-23.

[3] Al-garadi, M.A., K.D. Varathan, and S.D. Ravana, *Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network.* Computers in Human Behavior, 2016. 63: p. 433-443.

[4] Phillips, L., et al., *Using Social Media to Predict the Future: A Systematic Literature Review.* arXiv preprint arXiv:1706.06134, 2017.

[5] Quan, H., J. Wu, and J. Shi, *Online social networks & social network services: A technical survey.* Pervasive Communication Handbook. CRC, 2011: p. 4.

[6] Peterson, J.K. and J. Densley, *Is Social Media a Gang? Toward a Selection, Facilitation, or Enhancement Explanation of Cyber Violence.* Aggression and Violent Behavior, 2016.

[7] Watters, P.A. and N. Phair, *Detecting illicit drugs on social media using automated social media intelligence analysis (ASMIA)*, in *Cyberspace safety and security*. 2012, Springer. p. 6676.

[8] Fire, M., R. Goldschmidt, and Y. Elovici, *Online social networks: threats and solutions.* IEEE Communications Surveys & Tutorials, 2014. 16(4): p. 2019-2036.

[9] Shekokar, N.M. and K.B. Kansara. *Security against sybil attack in social network*. in*Information Communication and Embedded Systems (ICICES), 2016 International Conference on*. 2016. IEEE.

[10] Ratkiewicz, J., et al. *Detecting and Tracking Political Abuse in Social Media*. in*ICWSM*. 2011.

[11] Aggarwal, A., A. Rajadesingan, and P. Kumaraguru. *PhishAri: Automatic realtime phishing detection on twitter*. in*eCrime Researchers Summit (eCrime), 2012*. 2012. IEEE.