INTERNATIONAL JOURNAL OF
TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

PREDICTING CYBERBULLYING ON SOCIAL MEDIA IN THE BIG DATA ERA USING MACHINE LEARNING ALGORITHMS

[1]G.Rajasekharam, Associate Professor, Head of Department

[2]A.Mahesh [3]K.Madhu Babu [4]M.Durga Prasad [5]B.Neelima

Miracle Educational Group of Institutions, Vizianagaram, A.P, India

ABSTRACT

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 behaviour-related data. The fast growing use of social networking sites among the teens have made them vulnerable to get exposed to bullying. Cyberbullying is the use of computers and mobiles for bullying activities. Comments containing abusive words effect psychology of teens and demoralize them. The scourge of cyberbullying has assumed alarming proportions with an ever-increasing number of adolescents admitting to having dealt with it either as a victim or as a bystander. Anonymity and the lack of meaningful supervision in the electronic medium are two factors that have exacerbated this social menace. Comments or posts involving sensitive topics that are personal to an individual are more likely to be internalized by a victim, often resulting in tragic outcomes. Our initial experiments show that using features from our hypotheses in addition to traditional feature extraction techniques like TF – IDF and N – gram increases the accuracy of the system

1. INTRODUCTION

Machine or deep learning algorithms help researchers understand big data [1]. Abundant information on humans and their societies can be obtained in this big data era, but this acquisition was previously impossible [2]. One of the main sources of human-related data is social media (SM). By applying machine learning algorithms to SM data, we can exploit historical data to predict the future of a wide range of applications. Machine learning algorithms provide an opportunity to effectively predict and detect negative forms of human behavior, such as cyberbullying [3]. Big data analysis can uncover hidden knowledge through deep learning from raw data [1]. Big data analytics has improved several applications, and forecasting the future has even become possible through the combination of big data and machine learning algorithms [4].

An insightful analysis of data on human behavior and interaction to detect and restrain aggressive behavior involves multifaceted angles and aspects and the merging of theorems and techniques from multidisciplinary and interdisciplinary. The accessibility of large-scale data produces new research questions, novel 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). Applying these

ISSN: 2057-5688

methods to large online social networks (OSNs) in terms of scale and extent causes several issues. On the one hand, the explosive growth of OSNs enhances and disseminates aggressive forms of behavior by providing platforms and networks to commit and propagate such behavior. On the other hand, OSNs offer important data for exploring human behavior and interaction at a large scale, and these data can be used by researchers to develop effective methods of detecting and restraining misbehavior and/or aggressive behavior. 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

2. LITERATURESURVEY

Machine or deep learning algorithms help researchers understand big data [1]. Abundant information on humans and their societies can be obtained in this big data era, but this acquisition was previously impossible [2]. One of the main sources of human-related data is social media (SM). By applying machine learning algorithms to SM data, we can exploit historical data to predict the future of a wide range of applications. Machine learning algorithms provide an opportunity to effectively predict and detect negative forms of human behavior, such as cyberbullying [3]. Big data analysis can uncover hidden knowledge through deep learning from raw data [1]. Big data analytics has improved several applications, and forecasting the future has even become possible through the combination of big data and machine learning algorithms [4]. An insightful analysis of data on human behavior and interaction to detect and restrain aggressive behavior involves multifaceted angles and aspects and the merging of theorems and techniques from multidisciplinary and interdisciplinary fields. The accessibility of large-scale data produces new research questions, novel 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) Applying these methods to large online social networks (OSNs) in terms of scale and extent causes several issues. On the one hand, the explosive growth of OSNs enhances and disseminates aggressive forms of behavior by providing platforms and networks to commit and propagate such behavior. On the other hand, OSNs offer important data for exploring human behavior and interaction at a large scale, and these data can be used by researchers to develop effective methods of detecting and restraining misbehavior and/or aggressive behavior. 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. The remainder of this paper is organized as follows. Subsection I.A presents an overview of aggressive behavior in SM, and a new means in which SM websites are utilized by users to commit aggressive behavior is highlighted. I.B summarizes the motivations for constructing prediction modelsto combat aggressive behavior in SM. I.C highlight the importance of constructing cyberbullying prediction models. I.D, provide the methodology followed in this paper. Section 2 presents a comprehensive review of cyberbullying prediction models for SM websites from data collection to evaluation. Section 3 discusses the main issues related to

ISSN: 2057-5688

the construction of cyberbullying prediction models. Research challenges, which present new research directions, are discussed in Section 4, and the paper is concluded in Section 5. A. Rise of Aggressive Behavior on SM Prior to the innovation of communication technologies, social interaction evolved within small cultural boundaries, such as locations and families [5]. 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 user-driven technologies, such as SM websites, blogs, online virtual communities, and online sharing platforms. New forms of aggression and violence emerge exclusively online [6]. The dramatic increase in negative human behavior on SM, with high increments in aggressive behavior, presents a new challenge [6, 7]. The advent of Web 2.0 technologies, including SM websites that are often accessed through mobile devices, has completely transformed functionality on the side of users [8]. SM characteristics, such as accessibility, flexibility, being free, and having well-connected social networks, provide users with liberty and flexibility to post and write on their platforms. Therefore, users can easily demonstrate aggressive behavior [9, 10]. 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 [11]. 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. Although SM websites provide many benefits to users, cyber criminals can use these websites to commit different types of misbehavior and/or aggressive behavior. The common forms of misbehavior and/or aggressive behavior on OSN sites include cyberbullying [3], phishing [12], spam distribution [13], malware spreading [14], and cyberbullying [15]. Users utilize SM websites to demonstrate different types of aggressive behavior. The main involvement of SM websites in aggressive behavior can be summarized in two points [9, 15]. I. OSN communication is a revolutionary trend that exploits Web 2.0. Web 2.0 has new features that allow users to create profiles and pages, which, in turn, make users active. Unlike Web 1.0 that limits users to being passive readers of content only, Web 2.0 has expanded capabilities that allow users to be active as they post and write their thoughts. SM websites have four particular features, namely, collaboration, participation, empowerment, and timeliness [16]. These characteristics enable criminals to use SM websites as a platform to commit aggressive behavior without confronting victims [9, 15]. Examples of aggressive behavior are committing cyberbullying [17-19] and financial fraud [20], using malicious applications [21], and implementing social engineering and phishing [12]. II. SM websites are structures that enable information exchange and dissemination. They allow users to effortlessly share information, such as messages, links, photos, and videos [22]. However, because SM websites connect billions of users, they have become delivery mechanisms for different forms of aggressive behavior at an extraordinary scale. SM websites help cybercriminals reach many users [23]. 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 [24, 25]. Machine learning algorithms have been used extensively to analyze SM website content for spam [26-28], phishing [29], and cyberbullying prediction [19, 30]. Aggressive behavior includes spam propagation [13, 31-34], phishing [12], malware spread [14], and cyberbullying [15]. Textual cyberbullying has become the dominant aggressive behavior in SM websites because these websites give users full freedom to post on their platforms[17, 35-39]. 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 [40]. 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 behavior [41]. Developing an effective prediction model for predicting cyberbullying is therefore of practical significance. With all these considerations, this work performs a content-based analysis for predicting textual cyberbullying on SM websites. The motivation of this review is explained in the following section. 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 [42] and has been documented as a serious national health problem [43] 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 [44]. Studies have also shown that cyberbullying victims incur a high risk of suicidal ideation [45, 46]. Other studies [45, 46] 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 [47]. 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 [48]. A well-known 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 [9]. The effects of public cyberbullying are worse than those of private ones, and anonymous scenarios of cyberbullying are worse than non-anonymous cases [49, 50]. 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 [43]. Recent research has indicated that most experts favor the automatic monitoring of cyberbullying [51]. A study that examined 14 groups of adolescents confirmed the urgent need for automatic monitoring and prediction models for cyberbullying [52] 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 [53]. 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 [54, 55]. 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 connected to the Internet to perform misbehavior without confronting the victims [40]. 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 [48]. 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 [41]. The nature of SM websites allows cyberbullying to occur secretly, spread rapidly, and continue easily [54]. 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, DBLP Computer Science Bibliography, ACM Digital Library, ScienceDirect, SpringerLink, and IEEE Xplore. The major keywords used for the literature search were coined in relation to social media as follows: cyberbullying, aggressive behavior, big data, and cyberbullying models. The second phase involved searching for literature through Qatar University's digital library. The articles retrieved from the search were scrutinized to ensure that the articles met the inclusion criteria. According to the inclusion criteria, for an article to be selected for the survey, it must report an empirical study describing the prediction of cyberbullying on SM sites. Otherwise, the article would be excluded in the selection. Many articles were rejected based on titles. The abstract and conclusion sections were examined to ensure that articles satisfied the screening criteria, and those that did not satisfy the criteria were excluded from the survey

3. PROBLEM STATEMENT

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 learning-based approaches in terms of precision . Dadvar *et al.* 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.The study concluded that the percentage of ``bad" words in a text is indicative of cyberbullying. Another research expanded a list of pre-defined profane words and allocated different weights to create bullying features. These features were concatenated with bag-of-words and latent semantic features and used as a feature input for a machine learning algorithm.

3.1 LIMITATIONS OF SYSTEM

The System is not much affective due to Semi supervised machine learning techniques. The system doesn't have sentiment classification for cyberbullying.

4. PROPOSED SYSTEM

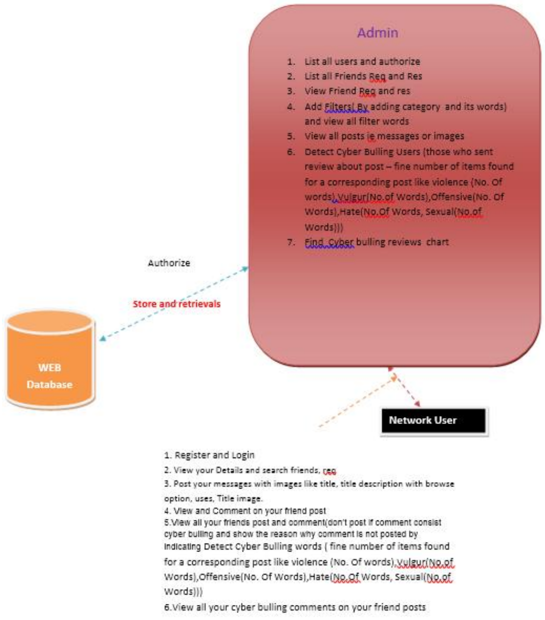
The proposed system is constructing cyberbullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labeled text instances. Another means is to use a lexicon-based model that involves computing orientation for a document from the semantic orientation of words or phrases in the document. Generally, the lexicon in lexicon-based models can be constructed manually or automatically by using seed words to expand the list of words. However, cyberbullying prediction using the lexicon-based 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, lexicon based 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

4.1 Advantages Of system

ISSN: 2057-5688

The system is more effective due to LOGISTIC REGRESSION CLASSIFICATION and UNSUPERVISED MACHINE LEARNING. An effective cyberbullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labeled text instance and also 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

5. SYSTEM ARCHITECTUR



1. List all users and authorize
2. List all Friends Req and Res
3. View Friend Req and res
4. Add Filters(By adding category and its words) and view all filter words
5. View all posts ie messages or images
6. Detect Cyber Bulling Users (those who sent review about post – fine number of items found for a corresponding post like violence (No. Of words),Vulgur(No.of Words),Offensive(No. Of Words),Hate(No.Of Words, Sexual(No.of Words)))
7. Find Cyber bulling reviews chart

1. Register and Login
2. View your Details and search friends, req
3. Post your messages with images like title, title description with browse option, uses, Title image.
4. View and Comment on your friend post
5.View all your friends post and comment(don't post if comment consist cyber bulling and show the reason why comment is not posted by indicating Detect Cyber Bulling words (fine number of items found for a corresponding post like violence (No. Of words),Vulgur(No.of Words),Offensive(No. Of Words),Hate(No.Of Words, Sexual(No.of Words)))
6.View all your cyber bulling comments on your friend posts

6.IMPLEMENTATION

6.1 Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as view and authorize users, view all friends request and responses, Add and View Filters, View all posts, Detect Cyber Bullying Users, Find Cyber Bullying Reviews Chart.

Viewing and Authorizing Users

In this module, the admin views all users details and authorize them for login permission. User Details such as User Name, Address, Email Id, Mobile Number.

Viewing all Friends Request and Response

In this module, the admin can see all the friends' requests and response history. Details such as Requested User Name and Image, and Requested to User Name and Image, status and date.

Add and View Filters

In this module, the admin can add filters (like Violence, Vulgar, Offensive, Hate, and Sexual) as Categories with the words those related to corresponding filters.

View all posts

In this module, the admin can see all the posts added by the users with post details like post name, description and post image.

Detect Cyber Bullying Users

In this module, the admin can see all the Cyber Bullying Users (The users who had posted a comment on posts using cyber bullying words which are all listed by the admin to detect and filter). In this, the results shown as, Number of items found for a corresponding post like Violence (no. of words belongs to Violence Filter used in comments by the users), Vulgar (no. of words belongs to Vulgar Filter used in comments by the users), Offensive (no. of words belongs to Offensive Filter used in comments by the users), Hate (no. of words belongs to Hate Filter used in comments by the users), Sexual (no. of words belongs to

Sexual Filter used in comments by the users).

Find Cyber Bullying Reviews Chart

In this module, the admin can see all the posts with number of cyber bullying comments posted by users for particular post.

6.2 User

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like viewing their profile details, searching for friends and sending friend requests, Posting Your Messages as Posts by giving details, View and Comment on Friend Posts, viewing all friends posts and comment, view all your cyber bullying comments on your friend posts.

Viewing Profile Details, Search and Request Friends

In this module, the user can see their own profile details, such as their address, email, mobile number, profile Image.

The user can search for friends and can send friend requests or can accept friend requests.

Add Posts

In this, the user can add their own posts by giving post details such as, post title, description, uses, and image of post.

View and Comment on Your Friends Post

In this, the user can see his entire friend's post details (post title, description, uses, creator and image of post) and can comment on posts.

View all Friends Posts and Comment (Cyber bullying Related)

In this, the user can see his all friend's post details (post title, description, uses, creator and image of post) and can comment on posts.

Don't Post If the comment consists of Cyber bullying words and Shows the reason why comment is not posted by indicating Detected Cyber Bullying Words like Numbers of Cyber Bullying words Related to Filter Violence found in comment, Numbers of Cyber Bullying words Related to Filter Vulgar found in comment, Numbers of Cyber Bullying words Related to Offensive found in comment, Numbers of Cyber Bullying words Related to Hate found in comment, Numbers of Cyber Bullying words Related to Sexual found in comment,

View all Your Cyber bullying comments on your friend posts

The user can see all his posted cyber bullying comments on their friend created posts ,

7. ALOGORITHMS USED

7.1 Feature Selection Algorithms

Feature selection algorithms were rarely adopted in state-of-the-art research to perform cyberbullying prediction on SM websites via machine learning (all extracted features are used to train the classifiers). Most of the examined studies (e.g., [18, 61, 68, 70-72, 85, 95, 97, 100])

ISSN: 2057-5688

did not use feature selection to decide which features are important in training machine learning algorithms. Two studies [19, 62] used chi-square and PCA to select a significant feature from extracted features. These feature selection algorithms are briefly discussed in following subsections.

1) Information Gain Information gain is the estimated decrease in entropy produced by separating examples based on specified features. Entropy is a well-known concept in information theory; it describes the (im)purity of an arbitrary collection of examples [101]. Information gain is used to calculate the strength or importance of features in a classification model according to the class attribute. Information gain [102] evaluates how well a specified feature divides training datasets with respect to class labels, as explained in the following equations. Given a training dataset (Tr) , the entropy of (Tr) is defined as. $I(Tr) = -\sum Pn \log2 Pn$, (1) where Pnis the probability that Trbelongs to class n. For attributeAttdatasets, the expected entropy is calculated as

$I(Att) = \sum ($

TrAtt

$Tr) \times I(TrAtt)$. (2)

The information gain of attribute Attdatasets is

$IG(Att) = I(Tr)- I(Att)$ (3)

2) Pearson Correlation

Correlation-based feature selection is commonly used in reducing feature dimensionality and evaluating the discrimination power of a feature in classification models . It is also a straightforward model for selecting significant features. Pearson correlation measures the relevance of a feature by computing the Pearson correlation between it and a class. The Pearson correlation coefficient measures the linear correlation between two attributes [103]. The subsequent value lies between −1 and +1, with −1 implying absolute negative correlation (as one attribute increases, the other decreases), +1 denoting absolute positive correlation (as one attribute increases, the other also increases), and 0 denoting the absence of any linear correlation between the two attributes. For two attributes or features X and Y, the Pearson correlation coefficient measures the correlation [104] as follows:

$rxy = \sum(xi-x)(yi-y) (n-1)SxSy$, (4)

7.2 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 [67, 100]. 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 [73, 106, 107]. 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. The following sections describe the machine learning algorithms commonly used for constructing cyberbullying prediction models (Table 2).

7.3 Support vector machine in cyberbullying

Support vector machine (SVM) is a supervised machine learning classifier that is commonly used in text classification [108]. SVM is constructed by generating a separating hyperplane in the feature attributes of two classes, in which the distance between the hyperplane and the adjacent data point of each class is maximized [109]. Theoretically, SVM was developed from statistical learning theory [110]. In the SVM algorithm, the optimal separation hyperplane pertains to the separating hyperplane that minimizes misclassifications that is achieved in the training step. The approach is based on minimized classification risks [107, 111]. SVM was initially established to classify linearly separable classes. A 2D plane comprises linearly separable objects from different classes (e.g., positive or negative). SVM aims to separate the two classes effectively. SVM identifies the exceptional hyperplane that provides the maximum margin by maximizing the distance between the hyperplane and the nearest data point of each class.

In real-time applications, precisely determining the separating hyperplane is difficult and nearly impossible in several cases. SVM was developed to adapt to these cases and can now be used as a classifier for non-separable classes. SVM is a capable classification algorithm because of its characteristics. Specifically, SVM can powerfully separate non-linearly divisible features by converting them to a high-dimensional space using the kernel model [112].

The advantage of SVM is its high speed, scalability, capability to predict intrusions in real time, and update training patterns dynamically. SVM has been used to develop cyberbullying prediction models and found to be effective and efficient. For example, Chen et al. (2012) [18] applied SVM to construct a cyberbullying prediction model for the detection of offensive content in SM. SM content with potential cyberbullying were extracted, and the SVM cyberbullying prediction model was applied to detect offensive content. The result showed that SVM is more accurate in detecting user offensiveness than naïve Bayes (NB). However, NB is faster than SVM. Chavan and Shylaja (2015) [19] proposed the use of SVM to build a classifier for the detection of cyberbullying in social networking sites. Data containing offensive words were extracted from social networking sites and utilized to build a cyberbullying SVM prediction model. The SVM classifier detected cyberbullying more accurately than LR did. Dadvar et al. (2012) [61] used SVM to build a gender specific cyberbullying prediction model. An SVM text classifier was created with gender specific characteristics.

The SVM cyberbullying prediction model enhanced the detection of cyberbullying in SM. Hee et al. (2015) [72] developed an SVM-based cyberbullying detection model to detect cyberbullying in a social network site. The SVM-based model was trained using data containing cyberbullying extracted from the social network site. The researchers found that that the SVM-based cyberbullying model effectively detected cyberbullying. Dinakar et al. (2015) [73] constructed an SVM-based cyberbullying detection model for YouTube. Data were collected from YouTube comments on videos posted on the site. The data were used to train

SVM and construct a cyberbullying detection model, which was then used to detect cyberbullying. The results suggested that the SVM-based cyberbullying model is more reliable but not as accurate as rule-based Jrip. However, the SVM-based cyberbullying model is more accurate than NB and tree-based J48. Mangaonkar et al. (2015) [95] proposed the use of SVM for the detection of cyberbullying in Twitter. An SVM-based cyberbullying model was constructed from data extracted from Twitter. The SVM-based cyberbullying prediction model was applied to detect cyberbullying in Twitter. SVM detected cyberbullying better than NB- and LR-based cyberbullying detection models did.

7.4 NB algorithm

NB was used to construct cyberbullying prediction models in [18, 38, 73, 74, 95]. NB classifiers were constructed by applying Bayes' theorem between features. Bayesian learning is commonly used for text classification. This model assumes that the text is generated by a parametric model and utilizes training data to compute Bayes-optimal estimates of the model parameters. It categorizes generated test data with these approximations [113]. NB classifiers can deal with an arbitrary number of continuous or categorical independent features [107]. By using the assumption that the features are independent, a high- dimensional density estimation task is reduced to one-

dimensional kernel density estimation [107].

The NB algorithm is a learning algorithm that is grounded on the use of Bayes theorem with strong(naive) independence assumptions. This method was discussed in detail in [114]. The NB algorithm is one of themost commonly used machine learning algorithms [115], and it has been constructed as a machine learning classifier in numerous social media based studies [116-118].

7.5 Random forest

Random forest (RF) was used in the construction of cyberbullying prediction models in [72, 86]. RF is a machine-learning model that combines decision trees and ensemble learning [119]. This model fits several classification trees to a dataset then combines the predictions from all the trees [120]. Therefore, RF consists of many trees that are used randomly to select feature variables for the classifier input. The construction of RF is achieved in the following simplified steps.

1. The number of examples (cases) in training data is set to N, and the number of attributes in the classifier is M.

2. A number of random decision tress is created by selecting attributes randomly. A training set is selected for each tree by choosing n times from all N existing instances. The rest of the instances in the training set are used to approximate the error of the tree by forecasting their classes.

3. For each tree's nodes, m random variables are selected on which to base the decision at that node. The finest split is computed using these m attributes in the training set. Each tree is completely built and is not pruned, as can be done in building a normal tree classifier.

4. A large number of trees are thus created. These decision trees vote for the most popular class. These processes are called RFs [119].

RF constructs a model that comprises a group of tree-structured classifiers, in which each tree votes for the most popular class [119]. The most highly voted class is the selected as the output.

7.6 Decision tree

Decision tree classifiers were used in construction of cyberbullying prediction models in [38, 95]. Decision trees are easy to understand and interpret; hence, the decision tree algorithm can be used to analyze data and build a graphic model for classification. The most commonly improved version of decision tree algorithms used for cyberbullying prediction is C.45 [38, 70, 95]. C4.5 can be explained as follows. Given N number of examples, C4.5 first produces an initial tree through the divide-and-conquer algorithm as

follows [121]: If all examples in N belong to the same class or N is small, the tree is a leaf labeled with the most frequent class in N. Otherwise, a test is selected based on, for example, the mostly used information gain test on a single attribute with two or more outputs. Considering that the test is the root of the tree creation partition of N into subsets N1 ,N2 ,N3 .... regarding the outputs for each example, the same procedure is applied recursively to each subset [121].

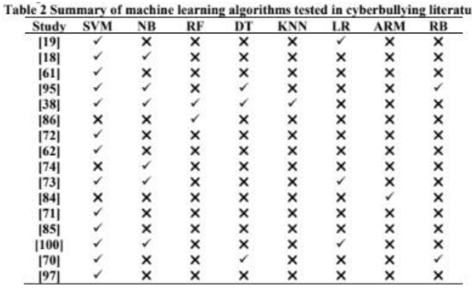
7.7 K-nearest neighbor

K-nearest neighbor (KNN) is a nonparametric technique that decidesthe KNNs of X0 and uses a majority vote to calculate the class label of X0 . The KNN classifier often uses Euclidean distances as the distance metric [122]. To demonstrate a KNN classification, classifying new input posts (from a testing set) is considered by using a number of known manually labeled posts. The main task of KNN is to classify the unknown example based on a nominated number of its nearest neighbors, that is, to finalize the class of unknown examples as either a positive or negative class. KNN classifies the class of unknown examples by using majority votes for the nearest neighbors of the unknown classes. For example, if KNN is one nearest neighbor [estimating the class of an unknown example using the one nearest neighbor vote (k = 1)], then KNN will classify the class of the unknown example as positive (because the closest point is positive). For two nearest neighbors (estimating the class of an unknown example using the two nearest neighbor vote), KNN is unable to classify the class of the unknown example because the second closest point is negative (positive and negative votes are equal). For four nearest neighbors (estimating the class of an unknown example using the four nearest neighbor vote), KNN classifies the class of the unknown example as positive (because the three closest points are positive and only one vote is negative). The KNN algorithm is one of the simplest classification algorithms, but despite its simplicity, it can provide competitive results [123]. KNN was used in the construction of cyberbullying prediction models in [38].

7.8 Logistic regression classification

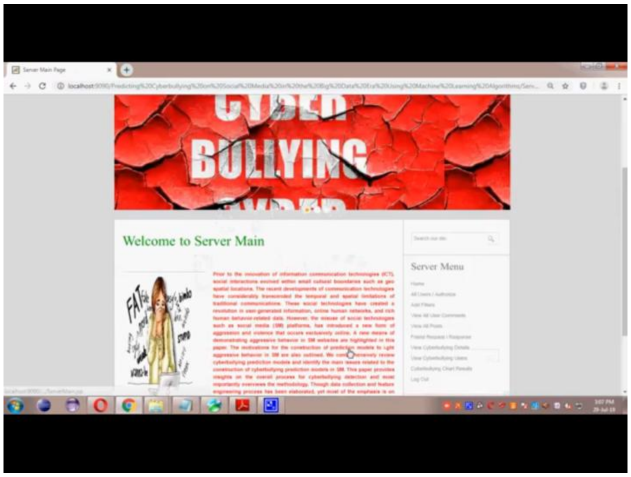
Logistic regression is one of the common techniques imported by machine learning from the statistics field. Logistic regression is an algorithm that builds a separating hyperplane between two datasets by means of the logistic function [124]. The logistic regression algorithm takes inputs (features) and generates a forecast according to the probability of the input being appropriate for a class. For example, if the probability is >0.5, the classification of the instance will be a positive class; otherwise, the prediction is for the other class (negative class) [125]. Logistic regression was used in the construction of cyberbullying prediction models in [19, 73].

Table 2 Summary of machine learning algorithms tested in cyberbullying literature



Study	SVM	NB	RF	DT	KNN	LR	ARM	RB
[19]	✓	✕	✕	✕	✕	✓	✕	✕
[18]	✓	✓	✕	✕	✕	✕	✕	✕
[61]	✓	✕	✕	✕	✕	✕	✕	✕
[95]	✓	✓	✕	✓	✕	✕	✕	✓
[38]	✓	✓	✓	✓	✓	✕	✕	✕
[86]	✕	✕	✓	✕	✕	✕	✕	✕
[72]	✓	✕	✕	✕	✕	✕	✕	✕
[62]	✓	✕	✕	✕	✕	✕	✕	✕
[74]	✕	✓	✕	✕	✕	✕	✕	✕
[73]	✓	✓	✕	✕	✕	✓	✕	✕
[84]	✕	✕	✕	✕	✕	✕	✓	✕
[71]	✓	✕	✕	✕	✕	✕	✕	✕
[85]	✓	✕	✕	✕	✕	✕	✕	✕
[100]	✓	✓	✕	✕	✕	✓	✕	✕
[70]	✓	✕	✕	✓	✕	✕	✕	✓
[97]	✓	✕	✕	✕	✕	✕	✕	✕

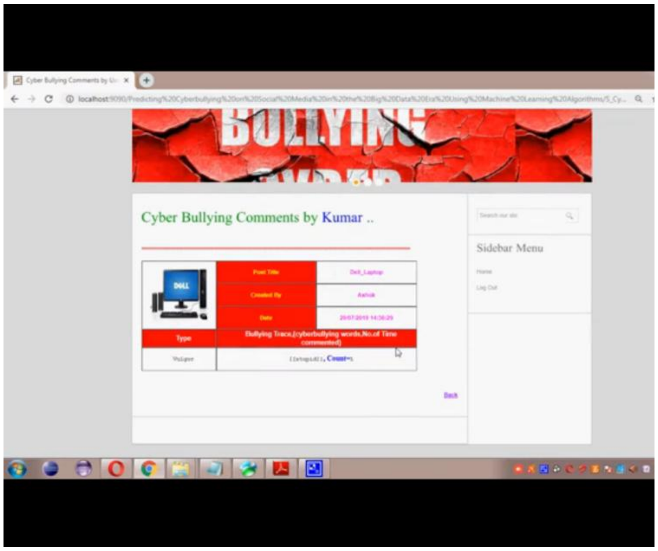
** SVM = support vector machine family, NB = naïve Bayes, RF = random forest, DT = decision tree family, KNN = K-nearest neighbor, LR = log regression, ARM = association rule mining, RB = rule-based algorithms



Welcome to Server Main

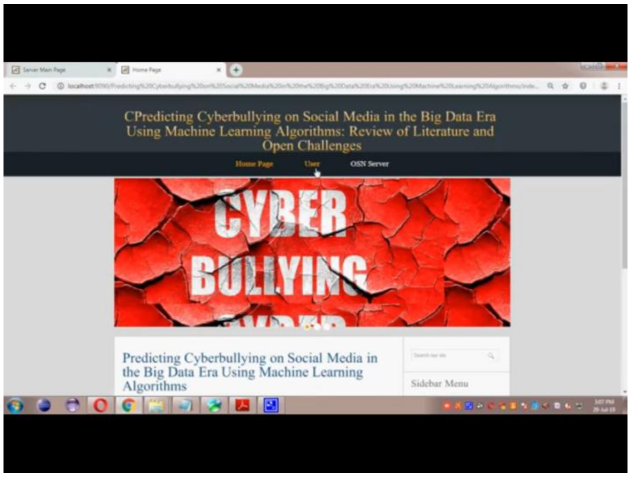
Prior to the innovation of information communication technologies (ICT), social interactions evolved within small cultural boundaries such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications. These social technologies have created a revolution in user-generated information, online human networks, and rich human behavior-related 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 are highlighted in this paper. 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

8. OUTPUT RESULTS



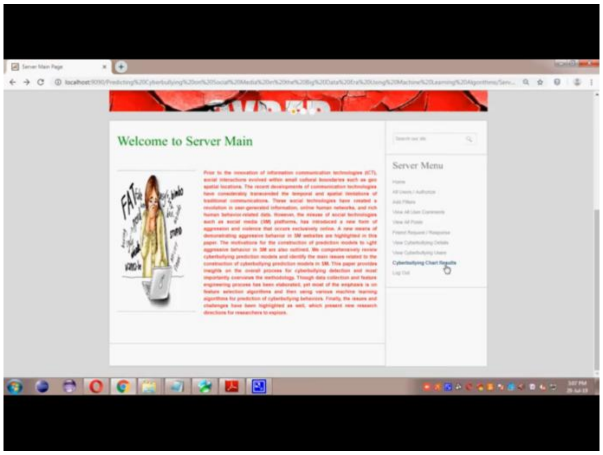
Cyber Bullying Comments by Kumar ..

	Post Title	Dell_Laptop
	Created By	Ashok
	Date	20/07/2019 14:36:29
Type	Bullying Trace,(cyberbullying words,No.of Time commented)	
Vulgar	[(stupid)], Count=1	



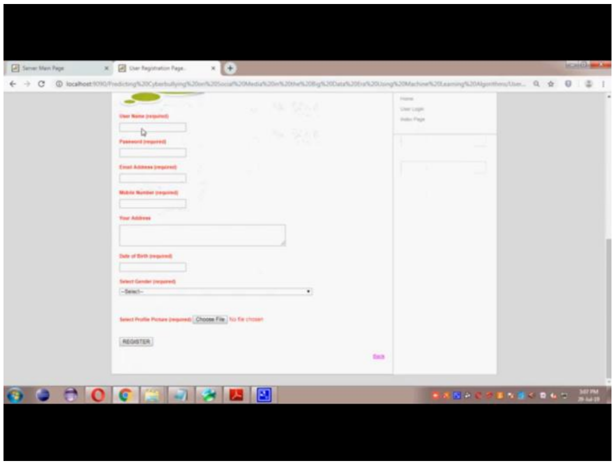
CPredicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges

Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms



Welcome to Server Main

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 behavior-related 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 are highlighted in this paper. 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, the issues and challenges have been highlighted as well, which present new research directions for researchers to explore.



User Name (required)

Password (required)

Email Address (required)

Mobile Number (required)

Your Address

Date of Birth (required)

Select Gender (required) --Select--

Select Profile Picture (required) Choose File No file chosen

9. 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, precision recall and f-measure which gives us the area under the curve function for modeling the behaviors in cyberbullying. Finally, the main issues and open research challenges were described and discussed.

10. 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, precision recall and f-measure which gives us the area under the curve function for modeling the behaviors in cyberbullying. Finally, the main issues and open research challenges were described and discussed

11. References

[1] V. Subrahmanian and S. Kumar, ``Predicting human behavior: The next frontiers,'' *Science*, vol. 355, no. 6324, p. 489, 2017.

[2] H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, ``Homophily in the digital world: A LiveJournal case study,'' *IEEE Internet Comput.*, vol. 14, no. 2, pp. 15_23, Mar./Apr. 2010.

[3] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, ``Cybercrime detection n online communications: The experimental case of cyberbullying detection in the Twitter network,'' *Comput. Hum. Behav.*, vol. 63, pp. 433_443, Oct. 2016.

[4] L. Phillips, C. Dowling, K. Shaffer, N. Hodas, and S. Volkova, ``Using social media to predict the future: A systematic literature review,'' 2017, *arXiv:1706.06134*.

[Online]. Available: https://arxiv.org/abs/1706.06134

[5] H. Quan, J. Wu, and Y. Shi, ``Online social networks & social network services: A technical survey,'' in *Pervasive Communication Handbook*. Boca Raton, FL, USA: CRC Press, 2011, p. 4.

[6] J. K. Peterson and J. Densley, ``Is social media a gang? Toward a selection, facilitation, or enhancement explanation of cyber violence,'' *Aggression Violent Behav.*, 2016.

[7] BBC. (2012). *Huge Rise in Social Media*. [Online]. Available: http://www.bbc.com/news/uk-20851797

[8] P. A.Watters and N. Phair, ``Detecting illicit drugs on social media using automated social media intelligence analysis (ASMIA),'' in *Cyberspace Safety and Security*. Berlin, Germany: Springer, 2012, pp. 66_76.

[9] M. Fire, R. Goldschmidt, and Y. Elovici, ``Online social networks: Threats and solutions,'' *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2019_2036, 4th Quart., 2014.

[10] N. M. Shekokar and K. B. Kansara, ``Security against sybil attack in social network,'' in *Proc. Int. Conf. Inf. Commun. Embedded Syst. (ICICES)*, 2016, pp. 1_5.

[11] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, ``Detecting and tracking political abuse in social media,'' in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 297_304.

[12] A. Aggarwal, A. Rajadesingan, and P. Kumaraguru, ``PhishAri: Automatic realtime phishing detection on Twitter,'' in *Proc. eCrime Res. Summit (eCrime)*, Oct. 2012, pp. 1_12.

[13] S. Yardi *et al.*, ``Detecting spam in a Twitter network,'' *First Monday*, Jan. 2009. [Online]. Available: https://_rstmonday.org/ article/view/2793/2431

[14] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, ``Analyzing spammers' social networks for fun and pro_t: A case study of cyber criminal ecosystem on twitter,'' in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 71_80.

[15] G. R. S.Weir, F. Toolan, and D. Smeed, ``The threats of social networking: Old wine in new bottles?'' *Inf. Secur. Tech. Rep.*, vol. 16, no. 2, pp. 38_43, 2011.

[16] M. J. Magro, ``A review of social media use in e-government,'' *Administ. Sci.*, vol. 2, no. 2, pp. 148_161, 2012.

[17] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, ``Improving cyberbullying detection with user context,'' in *Advances in Information Retrieval*. Berlin, Germany: Springer, 2013, pp. 693_696.

[18] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, ``Detecting offensive language in social media to protect adolescent online safety,'' in *Proc. Int. Conf Privacy, Secur., Risk Trust (PASSAT)*, Sep. 2012, pp. 71_80.

[19] V. S. Chavan and S. S. Shylaja, ``Machine learning approach for detection of cyber-aggressive comments by peers on social media network,'' in *Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI)*, Aug. 2015, pp. 2354_2358.

[20] W. Dong, S. S. Liao, Y. Xu, and X. Feng, ``Leading effect of social media for _nancial fraud disclosure: A text mining

based analytics,'' in *Proc. AMCIS*, San Diego, CA, USA, 2016.

[21] M. S. Rahman, T.-K. Huang, H. V. Madhyastha, and M. Faloutsos, ``FRAppE: Detecting malicious Facebook applications,'' in *Proc. 8th Int. Conf. Emerg. Netw. Exp. Technol.*, 2012, pp. 313_324.

[22] S. Abu-Nimeh, T. Chen, and O. Alzubi, ``Malicious and spam posts in online social networks,'' *Computer*, vol. 44, no. 9, pp. 23_28, Sep. 2011.

[23] B. Doerr, M. Fouz, and T. Friedrich, ``Why rumors spread so quickly in social networks,'' *Commun. ACM*, vol. 55, no. 6, pp. 70_75, Jun. 2012.

[24] J. W. Patchin and S. Hinduja, *Words Wound: Delete Cyberbullying and Make Kindness Go Viral*. Golden Valley, MN, USA: Free Spirit Publishing, 2013.

[25] J. Cheng, C. Danescu-Niculescu-Mizil, and J. Leskovec, ``Antisocial behavior in online discussion communities,'' in *Proc. 9th Int. AAAI Conf. Web Social Media*, Apr. 2015.

[26] S. Liu, J. Zhang, and Y. Xiang, ``Statistical detection of online drifting Twitter spam: Invited paper,'' in *Proc. 11th ACM Asia Conf. ComputCommun. Secur.*, 2016, pp. 1_10.

[27] Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A. H. Wang, ``Twitter spammer detection using data stream clustering,'' *Inf. Sci.*, vol. 260, p. 64_73, Mar. 2014.

[28] M. Jiang, S. Kumar, V. S. Subrahmanian, and C. Faloutsos, ``KDD 2017 tutorial: Data-driven approaches towards malicious behavior modeling,'' *Dimensions*, vol. 19, p. 42, 2017.

[29] S. Y. Jeong, Y. S. Koh, and G. Dobbie, ``Phishing detection on Twitter streams,'' in *Proc. Paci_c_Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2016, pp. 141_153.

[30] I. Frommholz, H. M. Al-Khateeb, M. Potthast, Z. Ghasem, M. Shukla, and E. Short, ``On textual analysis and machine learning for cyberstalking detection,'' *Datenbank-Spektrum*, vol. 16, no. 2, pp. 127_135, 2016.