Journal of Theoretical and Applied Information Technology
15th July 2021. Vol.99. No 13

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

APPROACHES TO CYBERBULLYING DETECTION ON SOCIAL NETWORKS: A SURVEY

[1] ZAINA ALSAED, [2]DERAR ELEYAN

[1]Student, Palestine Technical University- Kadoorie, Faculty of Graduate Studies, Tulkarm, Palestine

[2]Associate Professor, Palestine Technical University- Kadoorie, Department of Applied Computing, Tulkarm, Palestine

E-mail: [1]z.i.saed@students.ptuk.edu.ps, [2]d.eleyan@ptuk.edu.ps

ABSTRACT

Cyberbullying is a continuously growing issue in the insecure environment of social media networking platforms. It is common mostly among teenagers. To achieve successful cyberbullying prevention, appropriate detection of cyberbullying cases must be applied. This could be done through the application of intelligent techniques to identify mistreating behaviors. Nevertheless, automatic identification of potential online cyberbullying cases needs many requirements, especially with the huge loads of available information uploaded on the web. The primary objective of this paper is to highlight cyberbullying detection techniques so that it contributes positively to control bullying practices on social media. Its approach was reviewing existing attempts of cyberbullying detection using machine-learning algorithms and hence recap each. Overall, the outcomes are bright; however, they still have an opportunity to get better.

Keywords: *Cyberbullying, Machine learning*, *Social Networking*

1. INTRODUCTION

The accelerated growth of online information transmission platforms and the intense participation of their users have enabled peer-to-peer interaction at an unparalleled range and variety. These interaction stages, such as social media platforms and information-sharing websites (e.g. news), suggest multiple chances for intelligence distribution and belief motivation. Elsewise, they also represent a rich area for plenty of unfavorable threatening and insulting actions and expressions of cyberbullying towards targeted people due to their personalities or even when sharing their opinions.

Cyberbullying is considered one of the serious cybercrimes, as it could cause great emotional harm to the targeted person. Cyberbullying might lead to personal human-being harms extending from stress and panic to critical issues such as self-destruction and suicide. Research that was carried by the Pew Research Center [1] located that 60% of The United States' online users have encountered cyberbullying on the web, with adolescent women experiencing expressly hard sorts of such acts.

Surfing different online networking platforms recurring times every day became normal for people nowadays across the broads. Continuous-increasing quantity of users share their ideas, their feelings, and experiences amid several social media floors. Social media with a notable user base involve social media websites sites such as Twitter, Facebook, and ASKfm, social messaging applications such as WhatsApp, as well as photo and video sharing platforms such as Instagram and Snapchat, etc.

Users can now immediately join social media sites and applications at their convenience due to the universal support to the Internet and widespread wireless private telecommunication devices. Loads of content, in the class of texts, videos, and photos continue to pop up every day on common social media sites. Writers of those share views and ideas regarding a mixture of subjects and discuss shared concerns. The more users who participate in these platforms, the increased possibility of becoming valuable containers of citizens' minds and emotions concerning the facilities they use in addition to their ethical and political opinions. Consequently, these sites own a principal impact on citizen's beliefs and attitudes.

On the other hand, the obtained data represent a valuable resource for corporations, researchers, and decision-makers. While those modern information carriers, before mentioned as online social networks

expose. This critical and mostly premeditated targeting of individuals has a vital social role. In 2011, The National Crime Prevention Council announced that cyberbullying is a predicament that influences nearly half of American teenagers. The ends of cyberbullying are similar to conventional bullying and have been revealed to cause despair, low self-trust, and suicide trials [2],[3].

Nevertheless, sometimes, the results of cyberbullying could be more critical and longer-lasting because of some particular features of cyberbullying, such as the possibility of undertaking it at a rate of 24/7, and unlike regular bullying, it is not depending on the area and position [4]. Furthermore, online offenders can remain unknown [5], and being threatened by an anonymous user can be more depressing than the case of being bullied by someone familiar [12]. Besides, incognito triggers cyberbullying performance for users that might not do such act face-to-face [6].

1.1 Sentiment Analysis

Opinion or Sentiment analysis is the research of detecting and classifying views shown in a part of textual data or full-content using computers, particularly to determine whether the user's perspective towards a specific point is either neutral, positive, or negative. It is a mixture of typical language processing, text interpretation, and computational lexical. In this method, a text is recognized as a positive one if it is comprised of positive keywords, whereas it is acknowledged as a negative text if it has a negative one. This study serves to present an algorithm that can assist in the investigation of the content that may motivate and improve the process of crime detection in online networking platforms.

1.2 Motivation and Research Gap

The word “Cyberbullying” indicates usage of modern technology to hurt or irritate somebody in a preparatory, recurred, and hateful way. It is distinctive from conventional bullying since it can occur anytime and in any condition. It appears in the shape of malicious conversations, circulating lies, and distributing opprobrious forms of media on social networking platforms. The time that these abusive messages or media are published, it is so hard to remove them from social networking platforms. Therefore, to have a more reliable and more useful social platform, it is important to create a smart system that will prevent the before-mentioned action by controlling and cleaning the offensive, undesirable, and inappropriate content.

The process of cyberbullying detection is a crucial matter. Diverse problems must be fixed about the dataset, algorithm, creating a more reliable model, in terms of correctness of outcome, etc. Relating to the recent research in the area of cyberbullying detection techniques there is a hole between a large false alert and low efficiency. This gap could be defeated with the aid of the best characteristic (feature) choice, applying the most suitable machine learning techniques, and building a classifier that will give the most accurate outcomes when implemented towards the detection of cyberbullying content.

Potential recognition approaches of this offensive act are held largely by multiple studies. It is a comparably new field, yet the research development is exponentially growing because of the rising frequency of incidents of cyberbullying witnessed in online social networks, as well as the harm it is making to the community. Several studies were directed to the identification process of cyberbullying; in parallel with some researches, which have also studied the methods of cyberbullying prediction by reviewing and linking existing data on online social networks.

1.3 Paper Outlines

Our survey reports research works on automated cyberbullying detection. The study covers publications over the last decade. The range of the researches covered in the review highlights the expanding interest that cyberbullying detection and prevention means has been getting in nowadays. Supervised training and learning methods rule the systems conducted by many research papers.

In this paper, we offer a refined survey of cyberbullying recognition and detection ways. As well as analyzing such methods using multiple determinants employed for evaluation. The rest of the survey is outlined as follows:

Section 2 demonstrates the definition, types, and background of cyberbullying. Section 3 is covering the relevant research performed for cyberbullying detection methods. Section 4 is a performance comparison between the findings of reviewed papers. A discussion of open research challenges is provided in Section 5. In Section 6 we provide some future directions for upcoming research areas and finally, we conclude in Section 7.

2 LITERATURE REVIEW

Proposed work done in the area of cyberbullying can be classified into some categories, they are the definition of cyberbullying, and the latest research topics carried out concerning cyberbullying in online social networks, the diverse wide-scale scope of cyberbullying detection tools and techniques. Such of each category are discussed and reviewed as follows.

2.1 Social Media Networks

Online Social Networks (OSNs) are nowadays the modern age of connection. The decision-makers use such platforms to obtain ideas of the public concerning a particular subject [45], these online websites have now considered the most broadly adopted platforms to carry computerized cooperation projects. As stated by smallbiztrends.com, the popular OSNs are Facebook, Twitter, and Instagram, which many users are utilizing in addition to corporations and communities [46]. With wide online content to share utilizing these platforms, teenagers are likely to become either a cyberbullying victim or even a perpetrator of such act.

2.2 Definition and Types of Cyberbullying

2.1.1 Cyberbullying definition

When it comes to understanding the term "Cyberbullying", the main question that comes up and should be answered is: should cyberbullying be considered as a subcategory of traditional bullying, or does it represent a distinct phenomenon with its independent, special characteristics that diverges partially from traditional bullying?

Through the past 5-10 years, researchers have conducted a vast number of publications to explain and answer this question, suggesting various definitions of the cyberbullying concept from different perspectives. From Olweus point of view, traditional Bullying is defined as *"Intentional aggression carried out repeatedly by one individual or a group of individuals towards a person who is unable to easily defend him or herself"* [7].

Typically, to classify an abusive act as bullying, some terms should exist such as unevenness of power between the victim and the offender [8]. Hence, an extended definition of bullying was proposed, so that it makes it obvious that usually the act of bullying may be considered as a shape of peer abuse. Mainly, this definition is based on: (1) A hateful, offensive behavior that (2) suggests a pattern of iterative behavior, and (3) takes place in an interpersonal relationship described in terms of favoring the perpetrator(s) and power imbalance [9].

After defining the traditional form of bullying act, the advent of cyberbullying became prevalent side by side with the rapid development of online, digital means of communication, so it is very important to define cyberbullying in the context of traditional bullying. Cyberbullying is known as an aggressive, frequent, premeditated act that is done by a group or an individual, applied using means of digital, multi-modal types of communication, against a victim who could not defend her/himself [10]. One of the biggest variations between traditional bullying and cyberbullying is that the offender of cyberbullying intends to hurt the feelings of the victim [11], exploiting the easy use of today's means of digital communication, in addition to exploiting the three terms mentioned in the traditional definition of bullying.

Despite all of the efforts made to conceptualize cyberbullying overlapping with traditional bullying and suggest it as a distinct phenomenon, some definitions consider cyberbullying as an electronic form of face-to-face bullying only [12]. However, such consideration may disregard the complexities of abusive behavior, such as aggression recurrence in an electronic context and power imbalance. A closer look at some of these intricacies like recurrence and repetition in cyberbullying shows that it is problematic to contextualize such, as it is hard to estimate the difference between the perpetrator and the victim when it comes to counting the number of incidents occurrences and thus their consequences. This could be explained briefly through an example of a single aggressive incident when a victim uploads an embarrassing picture on one of the social media platforms, it could lead to a large-scale, continuous mockery and abasement for the victim. Here, the abusive act has ended, but the consequences of this act have resulted in extended, elongated humiliation to the victim.

2.1.2 Types of cyberbullying

Cyberbullying has two types: direct cyberbullying in which occurs among a couple of people only, the bully and a victim, and indirect cyberbullying, which is more dangerous, mainly differs from direct one, that a group of participants can take part in cyberbullying actions [26]. Indirect cyberbullying on social media can occur by making

fun of someone on a post with harassing comments and shares from several people.

Maher [13] has come up with around eight forms of cyberbullying behaviors in his research, four of them are:

1. Harassment: Sending insulting messages to the victim privately.
2. Masquerade: Based on pretending to imitate or impersonate the victim.
3. Exclusion: Excluding someone premeditatedly from an online group. Usually, the exclusion is prevalent among teenagers.
4. Flooding: Sending frequent frivolous messages/comments/posts to prevent someone from participating in the conversation.

2.2 Cyberbullying Research on Online Social Networks

A broad spectrum of researches and publications on cyberbullying has been fulfilled and conducted from various disciplines, as well as proposing a wide variety of Cyberbullying detection tools and techniques. Psychological and sociological researchers have done plenty of studies to highlight and discuss possible strategies of cyberbullying intervention and prevention by studying the personality and the motivations of bullies [16],[17]. This area of studies was mainly concerned with evaluating the impact of authority responsibilities and peer roles on cyberbullying behaviors encouragement or even mitigation, which represents the basis of motivating and enhancing the development of modern approaches and techniques of automated detection of cyberbullying in online social networks.

Within computer science, Hosseinmardi et al. [18] have attempted to explore the correlation between cyberbullying and anonymity in Ask.fm semi-anonymous online platform. First, they collected 30K user profiles in Ask.fm social Network, they used snowball sampling techniques in specific [19]. The following step was the analysis of these profiles, which was done using interaction and word graphs, network characteristics, the impact of negativity on in-degree and out-degree in addition to some frequency distributions. Finally, the research came up with a result that the least active users on social networking are most likely to be vulnerable. However, the analysis scope of this research was limited to public comments/posts, which overlooks the issue of private messaging harassment incidents.

That leads to Kontostathis et al.'s [20] research experiment of 29 transcripts on 288 chat logs which were gathered from a project's website to trap potential sexual predators by pretending that the project's volunteers are teenagers. Then they classified categories of frequently used phrases by predators into approach and grooming, false trust development, and isolation. Experimental methods achieved an accuracy of 93%.

Framing the problem slightly differently, Yin et al. [21] grouped online social networks into:

1. Discussion style: Built on several numbers threads, containing multiple posts to a predefined topic. Users have the option of joining a thread or starting a new one, either by comments or by posts. MySpace platform was used as a discussion environment data collection resource.
2. Chat style: In this style, ongoing conversations are unrestrained, and typically, each message consists of a few words with little information. Kongregate platform was used to collect data.

After finishing data collection, the supervised classifier was trained through topical and sentimental features to detect Bullying and harassment. Hence, a vast number of researches in cyberbullying analysis and detection area was done, among these was Al-Garadi et al. research [15], which was directed to detect cyberbullying on Twitter by following opinion mining and sentiment analysis techniques with a reported result of 67.3%. The classifier was built and evaluated by using the Amazon Mechanical Turk platform, which was also used to help in labeling the Tweets. Therefore, they have listed a group of negative words; to arrange and label tweets containing them. Then the sentiment classifier streamlines the tweets into four groups:

1. Negative content with bullying intentions.
2. Negative content without bullying intentions.
3. Positive content.
4. Neutral.

Although this approach provides a feasible detection technique of cyberbullying, nevertheless, its main shortcomings are that the process of

labeling data is costly and is limited on tweet analysis regardless of the other content.

Further cyberbullying insights would help in improving the detection process, specifically in multi-modal social networks such as Instagram, as the user have an open option of uploading videos and images in addition to textual content too. In contrast with textual cyberbullying, this sort of platform provides a prospect of abuse and harassment through social networking platforms for offenders so they could post or upload harmful content (videos and images) instead of using inappropriate comments only.

Additionally, well comprehension of cyberbullying behavior in online social networks could be done through a deep analysis of the engagement among the media contents and cyberbullying behavior. At long last, diving into the subtleties of cyber-aggression and cyberbullying and examining the potential distinctive variables between these two practices are additionally some undiscovered zones of future examination and investigation.

3. CYBERBULLYING DETECTION TOOLS AND TECHNIQUES

The widespread use of social media platforms such as LinkedIn, Instagram, Facebook, Twitter, YouTube, Pinterest, etc. is very challenging, as it brought many benefits, but at the same time, it has many drawbacks. The top five Social networking platforms with the highest records of cyberbullying detected incidents and experiences were: Facebook, Twitter, YouTube, Ask.fm, and Instagram [22].

A massive quantity of data is served every hour on different social networking platforms, with an exponential rate of growth. Such data can come in a form of textual comments and posts, photos, videos, and hashtags. For instance, the Facebook site is comprised of a combination of text, photos, and videos. Instagram is used for posting images and videos, while Twitter is used for sharing textual data with a limited number of characters, named tweets. The research area concerning cyberbullying detection is often focused on analyzing textual content rather than image or combined content because of its relative ease.

Detection of cyberbullying incidents, which is achieved by analyzing social media content, is a basic measure towards the prevention and defense of such acts. The majority of the researchers who tend to classify this content into bullying and non-bullying content are using machine-learning approaches. That is because hand-operated analysis of information and relationships between numerous information are inclined to errors and sometimes may lead even to blunders. AI can address such difficulties and can be effectively applied to these issues. However, one of the main challenges that face researchers is that the most considerable requirement is the availability of datasets to perform the process of training and hence testing the machine. Mahlangu et al. [23] have reviewed this issue and came up with a result that most of the researchers have either created their datasets or have scrawled websites.

To apply AI calculations an information dataset is made including occurrences illustrated by a bunch of features. Such of these could be binary, continuous, and categorical. Moreover, machine learning algorithms can also be categorized either as supervised machine learning, in which data instances are associated with labels [24], or unsupervised machine learning, where data instances are unlabeled [25], which is mainly used to discuss how to categorize and group data relations into clusters and inter-cluster ones. As its name suggests, the learner is not supervised, thus, any actions that provide the best result must be explored without any help or any type of guidance, so that useful classes and groups of items can be obtained. In contrast, supervised machine learning algorithms differentiate that they are used to observe data classification whether it is done properly or not, or appointed moderately high probabilities of having a place with the specific classification. This section outlines several previous contributions discussing cyberbullying detection by machine learning techniques and tools.

Mainly, the basis of cyberbullying research is textual cyberbullying modeling [27]. According to Homa et al. [18], analyzing textual data is not limited to extracting insulting or offensive words only. Therefore, the authors stated that cyberbullying incidents identification is not based on bad word presence, but also an offensive behavior has to be checked whether it is repeated regularly or not before considering it as bullying.

Taking the Twitter platform as a model to apply textual cyberbullying detection techniques, Zhao, Zhou, and Mao [27] have tried to divide the problem into a sub-problem concerning the process of detecting threads containing sensitive topics and content, and hence the classification of this textual content. Sensitive topics include sexual-related topics, caste/racism-related topics, and intelligence-related topics. As claimed by their research, the last

step after data collection of such sensitive topics was to determine the impropriety of these comments so that a cyberbullying act is detected.

Some contributions tried to discover various ways of establishing machine learning classifiers for cyberbullying and evaluating them [28], these classifiers were human expert systems, supervised machine learning algorithms, and a hybrid model comprising of both afore-mentioned systems. To evaluate such systems, the resource of retrieved labeled data was the YouTube site. After the comparison, the outcome result was that the hybrid model had the best performance out of the two remaining models. The sensitivity to the class skew of the dataset (10% bullying and 90% non-bullying) has resulted in the machine learning models' reported relatively under-performance.

Kowalski et al. [12] have adopted different strategies, for example, building query terms of expressions relating to cyberbullying have been created in the past to identify the occurrence of cyberbullying incidents. This approach was done based on using labeled data from FormSpring.me, and then they proceeded to generate query terms investigating both machine and language learning. The result was that machine-learning outcomes outperformed the language learning's produced terms with higher accuracy and better recall.

Beginning work in cyberbullying identification methods has generally focused on studying the chats' content. However, they did not take care of the importance of attributes of the parties of the cyberbullying act. Social investigations exhibited that bullying ways differ between males and females. For instance, women will, in general, utilize forceful correspondence styles, for example, barring somebody from a gathering of connivance against them while men tend to use more words and expressions threatening insult. Lee and Ma [29] detailed that pronouns like "I", "you", "she", and so forth are utilized more by females, and thing specifiers, for example, "a", "the", "that" are utilized conspicuously by men. These discoveries inspired a few cyberbullying specialists to incorporate sex explicit data in cyberbullying location procedures. Similarly, Sex explicit data in online informal organizations has been accounted for to be valuable in improving the exhibition of a cyberbullying identification framework [30] with an out-degree centrality score of 0.571 versus 0.33.

Authors in [43], [44] chose a related path on MySpace platform by providing an SVM classifier on posts classified by the authors' sex. They observed that cyberbullying detection progress was considerably increased on the gender divided posts when weighed against outcomes acquired training similar classifier on a non-separated one.

Likewise, Graph models, another approach to understanding cyberbullying cases in social media platforms; have been actively employed in cyberbullying research. The contribution of Hosseinmardi et al. [18] introduced a graph model to derive a cyberbullying network. This then headed to recognizing the common live offenders and targeted victims through a grading algorithm. They developed the ranking method by implementing a weighted TF-IDF function, the main methodology was scaling features similar to bullying by a factor of two.

From this point, the detection of cyber bullies and cyber predators has been studied in some of the past work [31]. A suggested definition of a cyber-predator is that he or she could be a person who exploits the Internet to seek vulnerable victims to avail from them in many perspectives, inclusive of financial, emotional, sexual, or psychological exploitation.

Cyber predators grasp the way of manipulating youngsters, building fake credence, and confidence [10]. Accordingly, studying online sexual predators was a critical issue to discuss to improve the procedure of distinguishing predators and victims by investigating the one-to-one talks [32], hence identification of text-mining and communication procedures.

Rezvan et al. [31] divided the online predator discovery matter into two sub-problems, particularly recognizing predators and approving predator's communication techniques/lines for naming them. They stated three levels (stages) of this procedure:

1. Pre-filtering step.
2. Feature extraction step, using:
    a. Behavioral features: based on the number of subjects and questions discussed, intention, and purpose to seize the action of the users.
    b. Lexical features: e.g. bigrams and unigrams [12], and the quantity of emoji used in the online discussion between the victim and the likely predator. [15]
3. Grouping stage, done through:

a. Decision trees [18].
b. Maximum-Entropy [27].
c. Neural Network [12].

In the same context, Andriansyah et al. [33] carried out an analysis study towards the problem of assorting comments on Instagram using Support Vector Machine (SVM), to determine whether they should be considered as cyberbullying or not. To gain the training dataset, approximately 1K comments were used and 34 comments as a test dataset. The resources of such datasets were comments from Instagram profiles of Indonesian celebrities, particularly, Karin Novilda and Samuel Alexandar. The next step of this process was the implementation stage, primarily, they formed a text expression model with R language to generate the SVM model. Once the progress of the SVM method is built, they applied it to forecast if a comment is classified as cyberbullying or not. The outcome has reached a correctness of 79.41 %.

Eshan and Hasan [32] went into studying the use of machine learning to identify offensive Bangla topics and texts. They examine several machine-learning algorithms and differentiate which one is more suitable. Their tests involve algorithms such as:

a. Support Vector Machine.
b. Multinomial Naïve Bayes (MNB).
c. Random Forest (RF).

To complete the training process of the dataset, they gathered data from some of the Facebook accounts of Bangladeshi famous people. All special characters similar to @, - etc. were excluded, taking Bengali Unicode only into consideration. Cross-validation on the 10 folds method was used to prove the effectiveness of this approach.

Adopting this process, they could to discover 50% of the insulting statements. Besides, the trials were carried with three kinds of string features: trigram, unigram, and bigram. Later, these features are selected from all of the obtained comments, hence vectorized utilizing CountVectorizer and TfidfVectorizer. Finally, they came up with a decision that features of trigram TF-IDF Vectorizer beside SVM linear kernel returns the greater certainty of 82% outperforming all of the applied algorithms.

A supplement to what previous researchers did, Noviantho et al. [34] assembled a ranking approach using SVM, combined with diverse kernels and Naïve Bayes. They measured their system with the study of Reynolds et al. (2011) who handled decision trees and K-NN. Mainly, conversation information collected from the Kaggle (www.kaggle.com) was the resource of processed data. Then they moved on to the stages of data pre-processing, selection, ranking, and evaluation. They split the data into 2, 4, and 11 levels. After performing text extraction, they sorted it through Naïve Bayes, SVM in addition to linear, Poly, RBF, and sigmoid kernels.

For the evaluation stage, the efficiency rate was measured using the confusion matrix method. According to this model, SVM provided the most reliable result of 91.95% whereas SVM-RBF produced the wickedest average result of 86.73% for 11 classes. For the use of Ngrams, the best average result achieved by n-gram was 92.75% while the worst was 89.05%.

Continuing the research of cyberbullying detection using SVM, Nurrahmi and Nurjanah [35] have also attempted to do so. This was done by choosing posts from Twitter as a dataset, which were collected by using a web scraper tool Selenium, which used chrome driver and then opened the URL for doing inquiries for twitter login, then asked for data in the structure of the HTML form, and parsed it to prepare the needed data.

Following that, the pre-processing step is applied to the gathered data. This involves eliminating special characters and URLs from posts on Twitter. At the end of applying this method, they got:

1. 301 cyberbullying tweets.
2. 399 non-cyber bullying tweets.
3. 2,053 negative words.
4. 129 swear words.

SVM and K-NN were used to classify cyberbullying. SVM achieved the highest F1-score of 67%.

Ozel et al. [36] proposed the first research to detect cyberbullying of Turkish texts. They constructed a dataset from Instagram and Twitter messages and implemented machine-learning methods such as SVM, Decision Tree, MNB, and K-NN to identify and group cyberbullying.

The dataset was created manually comprised of 900 Twitter and Instagram messages. A number of 450 messages were classified as cyberbullying ones

and the rest were considered as cyberbullying irrelevant content. Male users sent a number of 225 messages of cyberbullying-classified content (Half of them) whereas female users sent the others. Two common feature collection methods: Chi-Square and Information Gain were tested to determine whether they enhance the ranking accuracy or not. Then they implemented machine-learning classifiers to each turn for both datasets, measured the F-measure rates, and then got the average of them for five turns to give a starting point for comparing results. Accordingly, there were two samples of datasets; the first one was with emoticons and the other one was without emoticons. After comparing both of them, the emoticons dataset had a more reliable classification exactness. The feature selection methods both presented alike results, but the Information Gain method had somewhat a better result.

In terms of exactness, Naïve Bayes was the best in case of not applying features, while K-NN gave the best records of accuracy when features were used. However, the Decision Tree method had the least accuracy of all classifiers. The accuracy of SVM was nether than Naïve Bayes and k- Nearest Neighbor in the majority of cases due to the non-optimization of parameters. After measuring running time, Naïve Bayes was considered as the best classifier, based on its performance in both training and testing interval with 0.37 seconds, SVM classifier was the next best with a record of 0.75 seconds.

In the frame of multilingual cyberbullying detection, Haidar et al. [37] worked on Arabic text cyberbullying detection. They have explained how NLP and several machine learning techniques, including SVM, Naïve Bayes, Decision Tree, and K-NN operate to recognize cyberbullying. Data sets were built from data available on Facebook and Twitter and then listed data with ML algorithms. To measure the performance of the classifiers, they offered to reassemble, accuracy, and F-measure to accomplish a method with the best execution. The gap here was that they did not use an actual methodology to detect cyberbullying. They suggested the aforementioned methods only.

Likewise, Del Vigna et al. [38] produced a customized abusive words classifier for the Italian Language. They developed a frame of comments mainly gathered from public pages of Italian newspapers on Facebook, political leaders, actors, musicians, etc. They processed 99 posts from these pages and gained about 17.5K comments. Some of these were considered as one among the three stages of malice (hate):

1. Non-hate.
2. Slight hate.
3. Extreme hate.

The remaining comments were interpreted as one or both forms of hate: hate and non-hate. They examined these datasets with two machine-learning techniques: Recurrent Neural Network called Long Short-Term Memory (LSTM) and SVM. Then they performed a 10-fold cross-validation method for each of the gathered datasets. After applying such on the three-class dataset, LSTM and SVM classifiers presented 60.50% and 64.61% of precision for each sequentially. For the two-class dataset, SVM and LSTM achieved correctness of 80.60% and 79.81% for each. It is noticed that they performed better with the SVM classifier. But the results of the three-class dataset were not adequate when applying each of them.

Researchers in [41, 42] , formulated abuse dictionaries (lexions) handling speech records selected by the authors or gained from external sources (e.g. and urbandictionary.com and noswearing). By comparing the existance of swearing or profanity to cyberbullying, the usage of such just lexicons neglects other important features of cyberbullying like recurrance and the existance of a unequal power.

Prediction of cyberbullying incidents in media-based social networks was also studied, Rezvan et al. [31] attempted to predict cyberbullying incidents of a posted picture data which usually comes with a text caption, including the comments posted on the photo too, using American social networking profiles, and taking 25,000 public accounts on the Instagram platform as their dataset. Toward labeling, they utilized a reference with profane terms. To create and prepare the classifier, a fivefold cross-validation technique was involved. Also, logistic retraction was implemented to train the prediction classifier. Set 0 has recorded 98% of cyberbullying actions. This attested that cyberbullying occurrences can be guessed with 0.99 renderings for Set 0. According to a ridge recession classifier, The most valid false positive percentage across Set 0 is 3%, using only the contents of the photo, media, and user description.

Zhao et al.'s [27] study was directed to detect cyberbullying incidents on Twitter. They followed a full novel approach named the embedding-enhanced Bag of Words model (EBoW). As for the

dataset, they adopted textual data or posts on the Twitter platform. Next, performing EBoW required the description of a list of offending words according to expert experience and lexical sources, moreover, they enlarged the abusive words to clarify bullying features. Diverse measurements were ascribed to bullying features referring to the cosine correlation linking words and EBoW.

Subsequently, according to the previous measurements, they ranked the severity of cyberbullying. They caught 684 bullying posts out of a total number of 1,762 sample posts. Training and testing were applied among 5-fold associating with LDA, BoW, LSA, sBoW. The performance of EBoW got out the best result of all. With an F1 Score of 78.0%, an accuracy of 76.8%, and 79.4% of Recall.

For the enhancement of cyberbully detection, Mangaonkar et al. [39] adopted cooperative computing. Their conclusion shows an advancement in terms of time and accuracy of the detection mechanism compared with the standalone model. They generated two datasets, both gained from tweets from Twitter. An equivalent dataset managing 170 bullying content, and similar non-bullying ones. The extra dataset was unequal utilizing 177 bullying content and 1163 non-bullying content. Then they implemented Logistic Regression, NB, and SVM methods of the machine learning, by text and bigram tokenizers parameter contexts. With the customized (Balanced) dataset, Logistics Regression made a slightly better performance compared with the others, with a percentage of higher than 60% accuracy-recall, and precision. The next position was occupied by Naïve Bayes which was alike to Logistic regression and SVM, with an improved recall but poor accuracy and precision. As for unequal datasets, Logistics Regression was repeatedly executed with more further than 30% accurate forecasts on average, while in Naïve Bayes, the values had declined and SVM was rejected due to its failure. Following that collaboration systems, i.e, AND likeness, OR likeness, and Random 2 Or parallelism were done to decide whether any better change exists in terms of recall, accuracy, and precision. Within the procedures approached, AND parallelism gave the highest accuracy and OR parallelism recorded the greatest recall, and 7 among 15 cases adopting these techniques worked quite better compared with their steady equivalent.

This study delivered some new prospects on how to enhance the result of practicing collaboration methods after using machine learning algorithms; to perform cyberbullying detection. However, they stated that the outcomes reached were out of any checking of the algorithms applied, which means that a little adjustment of the algorithms may give much better results. One effective plan of this study was that the background of two Twitter profiles was not acknowledged, which represents an effective function in discovering cyberbullying. Besides, a concern of this study was that the SVM classifier worked inadequately, whereas yet in the majority of proposed papers SVM was described as the most reliable method, so if SVM was well-adjusted, it could have had much more favored performance.

Gorro et al. [40] intended to identify cyberbullying doers in Facebook on a textual basis and the trustworthiness judgment of users and additionally inform them concerning the wickedness of cyberbullying. Datasets were gathered by a custom-made web scraper tool. Labeling data was performed by formulating a web-based mechanism, which includes a listing of associates, appending non-positive and abusive words, weighing labeling grade and refreshing frame, and lastly specified the tweet either within a negative word context or an abusive word context. Later, the dataset is preprocessed by excluding symbols, tokenizing, characters, etc. Next, features are extracted and the outcome of this level is tabulated. Finally, to generate SVM and KNN, the data is trained. By exposing cyberbullying using both aforementioned models, they determined that SVM combined with RBF kernel (c=4) gives the best f1-score with a percentage of 67%. Nevertheless, using SVM with linear kernel and KNN is more restricted than using them with RBF kernel. Through implementing the feature extraction stage, they marked the reliability of users and discovered 257 ordinary users, 45 dangerous bullying users, 53 bullying performers, and 6 potential bullying doers.

4. PERFORMANCE COMPARISON

Accuracy, Precision, Recall, and F-measure are the primary metrics of evaluating Classifiers. Straightforward differentiation of the researches based on these values provided by them is difficult. Mainly because the datasets utilized by them will affect the outcomes.

Without applying the tests on the identical dataset, a judgment of the obtained metrics' rates is pointless. Yet investigations that employed a

similar dataset tend to examine distinct selections from inside the dataset.

The most primary and frequently used measurements are:

- Recall (Detection Rate)
- True Positive Rate (TPR): Determined from True Positive (TP).
- False Positive Rate (FPR): Determined from False Positive (FP).
- Precision
- Accuracy
- True Negative Rate (TNR): Determined from True Negative (TN).
- False Negative Rate (FNR): Determined from False
- Negative (FN)
- F1-Score

Among the reviewed studies, it is ambiguous when researchers claimed “accuracy”. Whether they intended the numerical arithmetical accuracy, or if they were applying the word "accuracy" wrongly. Considering this, Table 1 shows the summary of the reviewed researches in this survey.

We noticed that many of the prominent high rates attained are by those that utilizing datasets that come in the websites and forums. These may not be potential illustrative platforms of cyberbullying and because of that, the records obtained utilizing such corpora are not comparable upon those accomplished using a more indicative sample of data like those that are available in the social media platforms.

For instance, In [41], researchers scored a 92%

Research	Social Media Platform	Dataset size	Applied Algorithm	Outcome	Research gap
SVM Classifier on Indonesian Selebgram [33]	Instagram	1000 comments	SVM	79.4% Accuracy	Help of combining kernels could provide more reliable outcomes
Bangali Abusive Text Detector [32]	Facebook	7500 comments	MNB, RF, SVM	82% Accuracy of SVM with tigram	Lack of implementation the spelling checking techniques
Text mining-based Classifier [34]	Kaggle	12,729 data	SVM, Naïve Bayes	97% Accuracy of SVM with poly kernel	Using abbreviated words, spelling checker was not implemented
Text Classifier [35]	Twitter	700 tweets	SVM, K-NN	67% F1-score for SVM	Stemming check was not implemented, Male and female partitioning was useless
Social and Textual Cyberbullying Detector [30]	Twitter	900,000 tweets	Bagging,J48,SMO , Dagging, Naïve Bayes, ZeroR	RoC of 0.755	The classifier was not mentioned
Turkish Abusive Text Detector [36]	Instagram and Twitter	900 messages	SVM, NB, Decision Tree, K-NN	NVB highest record of 0.81 F-score	No implementation
Twitter Cyberbullying Detector [39]	Twitter	1510 tweets	Naïve Bayes, SVM and Logistic Regression	Logistic Regression has above 60% precision recall, and accuracy	The outcome of three-class dataset is not adequate
Bullying features-based Detector [27]	Twitter	1762 tweets	SVM with (EBoW)	EboW Precision was 76.8%, Recall 79.4%, F1 score of 78.0%	No dataset classification
Abusive Words Classifier for the Italian Language [38]	Facebook	17500 comments	Selenium scrapper tool, SVM	SVM for two-class (80.60%) and three-class (64.61%)	Did not apply any other models which might provide much better result
Selenium and SVM Classifier [40]	Facebook	1200 posts	SVM	Precision 88%, Recall 87%	Dataset was too small.

Table 1 Comparison between reviewed papers

F-measure rate on a dataset from the Kongregate website, which is dedicated to video games with a low likelihood of cyberbullying, by applying an SVM classifier. While research's experiments results in [44] utilizing an SVM classifier on MySpace generated a 28% F-measure record. MySpace is a social media platform in which cyberbullying is expected to be more common.

Upon assessing the raw scores recorded by every study, we pick ones with the high records of the F-measure per each cyberbullying detection test and display the raw values of the scores for these. When the F-measure record is not provided, we take the value of Accuracy, Precision, and Recall. This outcome is displayed in Table 1. Researchers can implement this outcome as a key to future analyses using similar datasets.

5. OPEN RESEARCH ISSUES AND LIMITATIONS OF CURRENT WORK

Two fundamental research problems are encountering cyberbullying detection research, first is the shortage of a generally chosen description of the cyberbullying term for recognition objectives and the second is a lack of big identified cyberbullying corpora.

5.1 Non-comprehensive Understanding of Cyberbullying features

Notwithstanding that the preponderance of authors agrees on the description of the term cyberbullying to cover the essential principles of recurrence, intention to abuse, and power differentiation, it was that some studies are considering cyberbullying in such a less comprehensive way. Some studies usually relate the discovery of any sort of offensive and insulting posts or any form of content to the identification of cyberbullying with small or no try to verify a plan to make abuse, a power mismatch, or the repeated environment of the assaulting actions. To improve such researchers need to include the general description of cyberbullying.

5.2 Insufficiency and Shortage of Cyberbullying Datasets

The obstacle caused by the need for simply available labeled collections is marked by the fact that many studies reported a few distinct openly accessible datasets. Social media floors, especially those that are Messaging-focused like Facebook, Instagram, and Whatsapp are under-described in such datasets, and collections of datasets based on these OSNs will be addressed and welcomed by the researchers in the area of cyberbullying research.

Datasets can cover complete online conversations between many people and highlight various commentary schemes. To clarify, explanations can be either by involved users associated with roles of each. This can be done by allowing classifiers to be instructed to identify the several forms of cyberbullying or annotation by conversations (i.e. specifying the link that takes place among users according to the nature of messages transferred), by bullying kind (direct/indirect).

6. FUTURE DIRECTIONS

Upon reviewing recent literature, we suggest 33some future tips to encourage cyberbullying identification and detection research.

6.1 Advancement of Cyberbullying Detection of Non-Textual Content

While the center of the researches in our survey has mostly been on textual cyberbullying, online content of photos or videos can similarly be utilized as facilitating methods for online bullying and their influence can be more harmful. Furthermore, when OSNs advance their capability to identify and recognize textual cyberbullying, offenders may exploit the usage of different forms of data to avoid antibullying countermeasures. New improvements in Optical Character Recognition side by side to image processing techniques aid the attempts of cyberbullying detection in different media sorts.

With available online trends every day (e.g. memes and videos) enhancing widely prevalent in the current age, perpetrators to commit cyberbullying can readily exploit these. We, thus, envision that improving means responsible for detecting cyberbullying cases among multimedia data is a fundamental space for upcoming research proposals.

6.2 Real-time Detection of Cyberbullying incidents

Our study showed that the conventional procedure in bully discovery investigation is to instruct and assess classifiers on inactive data handled at a moment in time. The outcomes declared for these trials, nevertheless, give no sign of efficiency of real-time detection of cyberbullying of classifiers especially when talking about capability to deal with streaming data speed of detection.

For instance, take messaging networks like Whatsapp, to be valid on such, a robust

cyberbullying detection method needs to be ready to list messages conveniently as they are transferred between the users. A classifier can be assessed on wherewith fast it can recognize cyberbullying incidents while they happen in the stream by utilizing APIs (e.g. Twitter Streaming API) which transmits consecutive data streams.

6.3 Extending Conventional Cyberbullying Role Identification

If cyberbullying happens, there are usually many parts at action besides the commonly involved parties (bullies and victims).

These involve supporters, instigators, and witnesses. Upcoming proposed researches would consider detection criteria to outline these extra roles and record the possibility of how could individuals adapt or choose such roles. For instance, there are some issues yet unnoticed by researchers, such as do watchers (witnesses) ultimately convert to bullies or become supporters? What is the way organized incidents including many bullies arranged, whether they are prepared, and do they agree on the details preceding a crime? And if this case can be associated with a comment posted by the victim?

7. CONCLUSION

Upon analyzing the modern literature in detecting cyberbullying automatically, we noted that the majority of the studies are directed toward OSNs in which datasets could be obtained easily. Common platforms of OSNs are yet not defendable in terms of cyberbullying. The accuracy of detection outcomes can be enhanced by taking various factors, which are related to cyberbullying, into account. One issue associated with some researches is that the amount of openly obtainable cyberbullying datasets is not enough.

Moreover, several of such available datasets are old. It is explaining that powerful social media networks are yet depending on "Report Abuse" reports in combating cyberbullying at the time of applying AI applications (e.g. face identification and music suggestions) are now well-known hallmarks of daily life.

Judging the process of a cyberbullying detection operation is a very important matter. Numerous present metrics test the execution performance of such an operation. The confusion matrix is considered one of the most essential and regularly used techniques, which is mainly a specially designed table that enables the conception of the performance of a tested algorithm. In the case of balanced datasets presence (datasets containing equal proportions of bullying content vs. non-bullying ones), accuracy is a fair metric.

Cyberbullying is a matter of high significance, one that transforms the lives of youth and may lead to catastrophic consequences. The present status of cases for cyberbullying control, therefore, demands instant consideration and enhancement. This change is only probable if the researchers, along with instructional establishments, software businesspeople, social media networks, and law enforcement perform mindful and combined works to aid the distribution of experience in all ways. By doing such, the aim of achieving applicable cyberbullying detection approaches can improve research borders worldly.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of Palestine Technical University-Kadoorie PTUK for this research work as a part of PTUK research fund.

REFERENCES:

[1] A. Geiger, "How and why we studied teens and cyberbullying,"Pew Research Center, 2018. [Online]. Available: http://www.pewresearch.org/fact-tank/2018/09/27/qa-how-andwhy-westudied-teens-and-cyberbullying/.

[2] Dehue, F., Bolman, C. & V¨ollink, T. 2008. Cyberbullying: Youngsters' experiences and parental perception. CyberPsychology & Behavior, 11, 217-223.

[3] R., Broeren, S., Van De Looij–Jansen, P. M., De Waart, F. G. & Raat, H. 2014. Cyber and Traditional Bullying Victimization as a Risk Factor for Mental Health Problems and Suicidal Ideation in Adolescents. PloS one, 9, e94026.

[4] Shariff, S. & Patchin, J. W. 2009. Confronting cyber-bullying, Cambridge University Press.

[5] Shariff, S. 2008. Cyber-bullying: Issues and solutions for the school, the classroom and the home, Routledge.

[6] Campbell, M. A. 2005. Cyber bullying: An old problem in a new guise? Australian Journal of Guidance and Counselling, 15, 68-76.

[7] S. Salawu, Y. He, and J. Lumsden, "Approaches to Automated Detection of Cyberbullying: A Survey," IEEE Trans.

Affective Comput., vol. 11, no. 1, pp. 3–24, Jan. 2020, doi: 10.1109/taffc.2017.2761757.

[8] D. Olweus: Sweden. In The Nature of School Bullying. Edited by Smith PK, Morita Y, Junger-Tas-J, Catalano R, Slee P. Routledge; 1999:7-27.

[9] D. Olweus and S. P. Limber, "Some problems with cyberbullying research," Current Opinion in Psychology, vol. 19, pp. 139–143, Feb. 2018, doi: 10.1016/j.copsyc.2017.04.012.

[10] P. K. Smith, J. Mahdavi, M. Carvalho, S. Fisher, S. Russell, and N. Tippett, "Cyberbullying: its nature and impact in secondary school pupils," J. Child Psychol. Psychiatry, vol. 49, no. 4, pp. 376–385, Apr.2008.

[11] H. Vandebosch and K. Van Cleemput, "Defining Cyberbullying: A Qualitative Research into the Perceptions of Youngsters," CyberPsychology Behav., vol. 11, no. 4, pp. 499–503, Aug. 2008.

[12] R. M. Kowalski, S. Limber, and P. W. Agatston, Cyberbullying: bullying in the digital age. Wiley-Blackwell, 2012.

[13] D. Maher, "Cyberbullying: an ethnographic case study of one Australian upper primary school class," Youth Stud. Aust., vol. 27, no. 4, pp. 50–58, Dec. 2008.

[14] I. Alanazi and J. Alves-Foss, "Cyber Bullying and Machine

Learning: A Survey," Nov. 2020, doi: 10.5281/ZENODO.4249341.

[15] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," Comput. Human Behav., vol. 63, pp. 433–443, Oct. 2016.

[16] I. R. Berson, M. J. Berson, and J. M. Ferron, "Emerging Risks of Violence in the Digital Age," Journal of School Violence, vol.1, no. 2, pp. 51–71, Mar. 2002, doi: 10.1300/j202v01n02 04.

[17] S. Hinduja and J. W. Patchin, "Social Influences on Cyberbullying Behaviors Among Middle and High School Students," J Youth Adolescence, vol. 42, no. 5, pp. 711–722, Jan. 2013, doi:10.1007/s10964-012-9902-4.

[18] H. Hosseinmardi, A. Ghasemianlangroodi, R. Han, Q. Lv, and S. Mishra, "Towards understanding cyberbullying behavior in a semianonymous social network," in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), 2014, pp. 244–252.

[19] F. Baltar and I. Brunet, "Social research 2.0: virtual snowball sampling method using Facebook," Internet Res., vol. 22, no. 1, pp. 57– 74, Jan. 2012.

[20] A. Kontostathis and A. Kontostathis, "ChatCoder: Toward the Tracking and Categorization of Internet Predators," PROC. TEXT Min. Work. 2009 HELD CONJUNCTION WITH NINTH SIAM Int. Conf. DATA Min. (SDM 2009). SPARKS, NV. MAY 2009., 2009.

[21] D. Yin, B. D. Davison, Z. Xue, L. Hong, A. Kontostathis, and L. Edwards, "Detection of Harassment on Web 2.0," PROCEEDINGS OF THE CONTENT ANALYSIS IN THE WEB 2.0 (CAW2.0) WORKSHOP AT WWW2009. .

[22] Ditch the Label Anti Bullying Charity, The annual cyberbullying survey 2013 (2013), http:// www.ditchthelabel.org/annualcyber-bullying-survey-cyber-bullying-statistics/10

[23] T. Mahlangu, C. Tu, P. Owolawi, A review of automated detectionmethods for cyberbullying, in International Conference on Intelligent and Innovative Computing Applications (ICONIC) (IEEE, 2018)

[24] T. Hastie, J. Friedman, and R. Tibshirani, "Overview of Supervised Learning," Springer, New York, NY, 2001, pp. 9–40.

[25] H. B. Barlow, "Unsupervised Learning," Neural Comput., vol. 1, no. 3, pp. 295–311, Sep. 1989.

[26] C. Langos, "Cyberbullying: The Challenge to Define," Cyberpsychology, Behavior, and Social Networking, vol. 15, no. 6, pp. 285–289, Jun. 2012, doi: 10.1089/cyber.2011.0588.

[27] R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," in Proceedings of the 17th International Conference on Distributed Computing and Networking - ICDCN '16, 2016, pp.1–6.

[28] S. A. Ozel, E. Sarac, S. Akdemir, and H. Aksu, "Detection of cyberbullying on social media messages in Turkish," in 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 366–370.

[29] C. S. Lee and L. Ma, "News sharing in social media: The effect of gratifications and prior experience," Comput. Human Behav., vol. 28, no. 2, pp. 331–339, Mar. 2012.

[30] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber Bullying Detection Using Social and Textual Analysis," in Proceedings of the 3rd International Workshop on Socially-Aware Multimedia SAM '14, 2014, pp. 3–6.

[31] M. Rezvan, S. Shekarpour, L. Balasuriya, K. Thirunarayan, V. L. Shalin, and A. Sheth, "A Quality Type-aware Annotated Corpus and Lexicon for Harassment Research," in Proceedings of the 10th ACM Conference on Web Science - WebSci '18, 2018, pp. 33–36.

[32] S. C. Eshan and M. S. Hasan, "An application of machine learning to detect abusive Bengali text," in 2017 20th International Conference of Computer and Information Technology (ICCIT), 2017, pp. 1–6.

[33] M. Andriansyah et al., "Cyberbullying comment classification on Indonesian Selebgram using support vector machine method," in 2017 Second International Conference on Informatics and Computing (ICIC), 2017, pp. 1–5.

[34] Noviantho, S. M. Isa, and L. Ashianti, "Cyberbullying classification using text mining," in 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), 2017, pp.241–246.

[35] H. Nurrahmi and D. Nurjanah, "Indonesian Twitter Cyberbullying Detection using Text Classification and User Credibility," 2018 Int. Conf. Inf. Commun. Technol. ICOIACT 2018, vol. 2018–Janua, pp. 543–548, 2018.

[36] S. A. Ozel, E. Sarac, S. Akdemir, and H. Aksu, "Detection of cyberbullying on social media messages in Turkish," in 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 366–370.

[37] A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of cyberbullying behavior in Twitter data," in 2015 IEEE International Conference on Electro/Information Technology (EIT), 2015, pp. 611–616.

[38] K. D. Gorro, M. J. G. Sabellano, K. Gorro, C. Maderazo, and K. Capao, "Classification of Cyberbullying in Facebook Using Selenium and SVM," in 2018 3rd International Conference on Computer and Communication Systems (ICCCS), 2018, pp. 183–186.

[39] A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of cyberbullying behavior in Twitter data," in 2015 IEEE International Conference on Electro/Information Technology (EIT), 2015, pp. 611–616.

[40] K. D. Gorro, M. J. G. Sabellano, K. Gorro, C. Maderazo, and K. Capao, "Classification of Cyberbullying in Facebook Using Selenium and SVM," in 2018 3rd International Conference on Computer and Communication Systems (ICCCS), 2018, pp. 183–186.

[41] Nahar, V., Li, X. and Pang, C.. An Effective Approach for Cyberbullying Detection. Communications in Information Science and Management Engineering, 3(5), 2013, p.238.

[42] Bretschneider, U., Wöhner, T., and Peters, R. (2014). Detecting Online Harassment in Social Networks [online]. Available from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1003&context=icis2014 [Accessed 25th January 2021]

[43] Dadvar, M. and De Jong, F. . Cyberbullying detection: A Step toward a Safer Internet Yard. IN: International conference companion on World Wide Web. 21st. Lyon, April 16 - 20, 2012. London: ACM, 121-126.

[44] Dadvar, M., De Jong, F.M.G., Ordelman, R. J. F. and Trieschnigg, R. B. (2012a). Improved Cyberbullying Detection Using Gender Information [online]. Available from http://eprints.eemcs.utwente.nl/21608/01/DIR12_reviewed04.pd f [Accessed 5th January 2021].

[45] Boudjelida, A., Mellouli, S., & Lee, J. (2016, March). Electronic citizens participation: Systematic review. In Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance (pp. 31-39). ACM.

[46] Maina, A. "20 Popular Social Media Sites Right Now." Small Business Trends, February 15, 2017. Retrieved from https://smallbiztrends.com/2016/05/popular-social-media-sites.html