Journal of Theoretical and Applied Information Technology
31st August 2020. Vol.98. No 16

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

CYBERBULLYING DETECTION: CURRENT TRENDS AND FUTURE DIRECTIONS

KAZIM RAZA TALPUR[1], SITI SOPHIAYATI YUHANIZ [2], NNILAM NUR BINTI AMIR SJARIF [3], BANDEH ALI [4], NORSHALIZA BINTI KAMARUDDIN[4]

[1]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur

[2]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur

[3]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur

[4]School of Computer Science and Statistics, Trinity College Dublin (TCD), Dublin, Ireland

[4]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur

E-mail: [1]talpurkazim@gmail.com, [2]sophia@utm.my, [3]nilamnur@utm.my, [4]bandehali@gmail.com, [4]norshaliza.k@utm.my

ABSTRACT

As we see the rapid growth of Web 2.0; online social networks-OSNs and online communications which provides platforms to connect each other all over the world and express the opinion and interests. Online users are generating big amount of data every day. As result, OSNs are providing opportunities for cybercrime and cyberbullying activities. Cyberbullying is online harassing, humiliating or insulting an online user through sending text messages of threatening or harassing using online tool of communication. This research paper provides the comprehensive overview of cyberbullying that occurs usually on OSNs websites and provides current approaches to tackle cyberbullying on OSNs. It also highlights the issues and challenges in cyberbullying detection system and outline the future direction for research in this area. The topic discussed in this paper start with introduction of OSNs, cyberbullying, types of cyberbullying, and data accessibility is reviewed. Lastly, issues and challenges concerning cyberbullying detection are highlighted.

Keywords: *Cyberbullying, Online Social Networks (OSNs)*

1. INTRODUCTION

With the emergence of big data and information communication technology (ICT), Online Social Networks (OSNs) has emerged as main part of human life. It assists humans (online users) to keep in touch with one another with use of various types of OSNs platforms such as Facebook, Twitter and application with just a swipes and taps. It is frequent source of online entertainment. Moreover, OSNs platforms have been considered as major platforms for online connecting users throughout the world. Nevertheless, it can be observed that these OSNs platform are gaining greater prominence in cyberspace.

We also observe that some online users take an advantage of the them to do positive things, however few online users conduct a different type of illegal and malicious actions on them. Cyberbullying is one of such adverse issues develop with OSNs platforms. Cyberbullying can be considered as a means of misuse with information communication of technology to with the aim to harass or harm other users (people) in deliberately, hostile, and rehashed manner [1].

The cyberbullying can be defined as the use of digital and technological advancement through mobile phones, chat rooms, emails or OSNs platforms such as Facebook and Twitter to humiliate or threaten others [2]. Most of cyberbullying incidents happen in the form of sharing text messages, rumors, tweets and sharing private pictures or embarrassing and videos on OSNs platforms [1]. Bullies have been adopted the OSNs platforms such as Facebook and Twitter etc. to perform activities of bullying behind the mobile phone or personal computer by the internet. The Pew Research Centre showed that around 59% of US teens have been harassed or bullied online[1].

[1] https://www.pewresearch.org/internet/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/

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

Additionally, research study conducted by Symantec's Norton showed that 24% of parents mentioned that their child's have been involved in cyberbullying, incidents.[2]

Most of research studies concerning cyberbullying have been conducted with focused on the effect on human and how the victims handle it psychologically [3]. Limited research has been conducted concerning cyberbullying detection and stopping before they occur utilizing machine leaning (ML) approaches.

This research paper addresses the following issues and questions;

1. Why is it important to detect cyberbullying in OSNs platforms?
2. What are the types of cyberbullying and methods to detect cyberbullying in OSNs?
3. How cyberbullying detection can be improved?

The remainder of this research paper is systematized as section 2 provides the review of cyberbullying in OSNs platforms and types, section 3 presents existing research studies in the area cyberbullying detection, Section 4 presents Data accessibility, section 5 data accessibility challenges in cyberbullying detection, section 6 representation of data challenge in cyberbullying detection, section 6 open research issues and finally conclusion is given for the study.

2. CYBERBULLYING IN OSN PLATFORMS

A. Definitions of cyberbullying

Cyberbullying can take place via text messaging or on online platforms such as Facebook and Twitter, or online gaming where online users can post or share their feelings and thought in bad manner. In concise, OSNs are being utilized through bullies to threaten and harass online users (people).

Some of the most famous definitions of cyberbullying as follows:

[4] describe a cyberbullying, An aggressive, intentional act carried out by group or individual, using electronic forms of contact, repeatedly and overtime against a victim who cannot easily defend him or herself.[5] defined cyberbullying is when someone repeatedly make fun of another person online or repeatedly picks on another person through email or text message or when someone is posts something online about another person that they don't like. In addition Pachin and Hinduja describe cyberbullying as "willful and repeated harm inflicted by the medium of electronic text [6].

Cyberbullying is an injustice happening in virtually utilizing online devices such as mobile phones, computers and tablets. Unlikely conventional bullying, which was limited to schools' yards, college yards and streets, the large amount digital tools and devices frequently used in daily lives has brought cyberbullying into homes and bedrooms. Traditional bullying is commonly considered as a subpart of arrogant behaviour [7]. It is defined by characterized over the time period a discrepancy of authority between bully and targeted victims [6], [7].

Types of cyberbullying

There are several types of cyberbullying which can be recognized in [8], [9]. These types are such as flooding, masquerade, flaming, harassment, denigration, outing and trickery, exclusion cyberstalking and cyberthreats. These types of cyberbullying as follows. as shown in figure showing in Figure1.

Flooding: consists of the bullying monopolizing the OSNs/chat groups so that victim cannot post a message or no control to stop it. Commonly "flooding" is found in online chatrooms and online groups [8], [9].

Masquerade: this word is also referred to as *Impersonation*, consist in breach of OSNs victims. To take control of their personal identity, and to provide their online users with falsified and rumours news of the victims [10].

Furthermore, Masquerade term involves the bully logging into a OSNs, online chatroom group and using his/her user's screen name to either bully a targeted victim directly or damage the credibility or reputation [8], [9].

Flaming or bashing: this term describes that sending aggressive text messages to the victims in group of chat rooms in online games. Nevertheless, cases of flaming can be found in OSNs where hate speech is taking roots. Generally, hate speech victims are famous actors, actress, politicians,

[2] https://www.pcmag.com/archive/study-a-quarter-of-parents-say-their-child-involved-in-cyberbullying-266888

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

singers and athletes etc., however that always keep calm and don't respond and react. Rather, their fans and followers respond or react and stick together for their idol and in this way a so-called Flame war [10].

Flaming is involving two or more online users attacking each other. Flaming commonly include rude, offensive, insult, using vulgar language and sometimes threaten to other. Moreover, flaming generally found in public communication platforms[8], [9].

Harassment: meant that sending insulting or abusing text messages in a quick fire. Here, persistence is recognizable, as compared to Flaming.

Denigration: the main aim is to spread untrue rumours online gossip about an individual person or victims. Therefore, victims are socially isolated because of their wrecked credibility and reputation. Denigration is form of cyberbullying most commonly utilized by students against schools teachers, employees [8]–[10].

Outing: outing is openly sharing of images or personal conversations, specially that involve private information. Outing commonly occurs when a cyberbullying receives an email from a target that contains intimate private information and then forward messages to others users [8], [9].

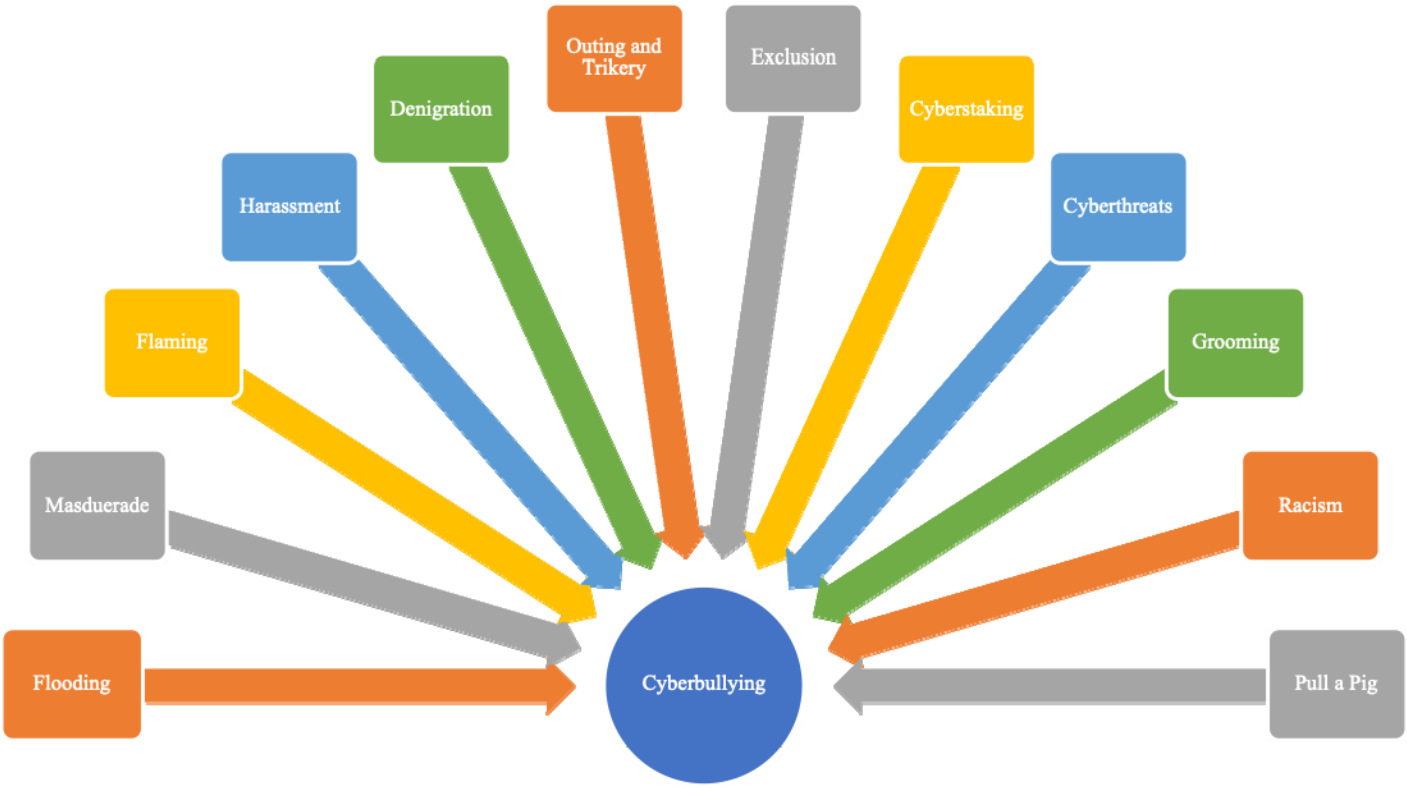

Figure 1. Cyberbullying types

In generally, sex is always popular subject of text messages of this category [10].

The term can be defined, is repeated ongoing sending of offensive text messages to an individual user (people) target. Harassment text messages are commonly sent by online communication platforms such as, e-mail, text messaging [8], [9].

Trickery: at the very beginning, the cyberbully stands as equal of the victims, who share trust and confidence with the cyberbully. Feeling comfortable with her/him. In just a short time, the cyberbully starts threatening the victim to share or post online confidence, when he or she is not giving favours such as sending nude photos or nude videos to the cyberbully [10].

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

This type of cyberbullying can also happen as part of outing. An innocent targeted user can be tricked into thinking that a sending videos or images and communication is private, when the cyberbully intends to trick the target user into disclosing or communication something awkward that will be dispersed to other online users as threat [8], [9].

For instance, a teen boy may engage to teen girl in sexually oriented communication via online text messaging or instant messaging, with a group of his friends surrounding his personal computer. Or teen girl communicating through instant text messaging may encourage another girl to disclose her on the basis of personal interest in student under the promise that the personal information data will remain "secret" then circulate this communication to other users [8], [9].

Exclusion: is type of cyberbullying that deals with designation of who is member of the ingroup and who is an outcast or excluding someone from engage in an online group. Exclusion can be occurs in an online gaming platform, online group blogging platform [8], [9].

Cyberstalking: In this type of cyberbullying victims concerned to their lives, because online stalkers forcing them to send information or photos and threaten to kill them if they don't follow. In addition, this type of cyberbullying also happens when ex-wives are humiliated by their ex-husbands or vice-versa. It is worthy of note that in 2017 about 65% of victims of cyberstalking in Italy were men [10].

Moreover, cyberstalking term can be defined as recurrent sending of hurtful, threatening, and destructive text messages that involve extortion or extremely aggressive. Cyberstalkers may also try to denigrate their target and destroy their reputations and friendships [8], [9].

Cyberstalkers is commonly linked with the termination of or problems within, online sexual relationship or in person. In simple way Cyberstalking describe, sending online intimidating and threating text messages to targeted users in online platforms.

Cyberthreats: Cyberthreats can be categorize as follows:

- Direct threats are statements intending to commit suicide or harm someone else. It usually involves information regarding an actual planned event.
- Distressing material is online content that intimate the clues that certain person is emotionally distressed and may be considering self-harm or suicide [8], [9].

Cyberthreats is very similar to cyberstalking and in concise way we define, the online user who involves sending text messaging which includes threats to another online user or groups.

Grooming: it signifies the practice to groom children online. Initially cyberbullies approach to the victims with a polite, fatherly or friendly conduct, and intended to gain the trust of the victims (Typically posing like police officer, a doctor, a teacher, specifically as somebody that would never hurt or harm to children) when the trust is won, the victims become sexually trapped using for instance online video conversation or webcams [10].

Racism: this terms applies to the discrimination due to different ethnicity, nationality, race or religion [10].

Pull a Pig: it refers to the victims of *Pull a Pig* cases are young girls considered as fat and ugly. Commonly cyberbullies sending flattering text messages to approach the victims and that they're trying to seduce them and the helpless victims are going for it and afterwards cyberbullies share or post their conversation on OSNs with the goal of mocking [10].

B. Impact of Cyberbullying

This section finds the influence of cyberbullying on victims across various research studies. impact on cyberbullying has been widely discussed in the area of adolescent's mental health concern issue. In their study, researchers have investigated the relationship between involvement with cyberbullying and adolescents tendency to internalize issues for instance the development of negative affective disorders, depression, suicidal ideation, somatic symptoms, loneliness and the research study showed that about 32% of cyberbullying targets experienced at least one symptoms of stress [11]. Youngster cyberbullying victims experience a lot of same damage as the person experiencing traditional

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

bullying. These affect which includes drops in grades and low self-esteem, and depression[3].

[12] conveyed that cyberbullying has become an issue, not only for Western nations but also for underdeveloped nations and results shows about 80% of participants in this research study had be frequently experienced cyberbullying and cyberbullying is considered as a stressful life event.

[13] in their study show that how cyberbullying linked to the mental health problems on students of high school. Logistic regression (LR) tests were performed to identify significant relationship between cyberbullying.

3. EXISTING RESEARCH STUDIES IN CYBERBULLYING DETECTION

Yin et al., [14] applied supervised machine learning method to classify online harassment. In this study they were tested approach on Myspace, Slashdot, Kongregate dataset and three features which includes, local features, sentiment features and contextual features. However, in this approach predators and victims were not recognized.

Dinakar et al.,[15] presented supervised machine learning approach, and compare the performance of binary and multiclassification for the variety of classifiers used which includes Naïve Bayes (NB), C4.5, JRip and SVM to detect in YouTube comments, therefore in this approach large dataset were not used and predator and victims were not recognized.

Reynolds et al.,[16] they used supervised machine learning approach combined with labeled data were used to learn the system and detect bullying content. Nevertheless, in this research conversation pragmatics not considered.

Özel et al.,[17] proposed supervised machine learning approach to detect Turkish language messages in OSNs and they created own dataset from Twitter and Instagram. However, in their study they used small dataset and random forest never considered.

Huang et al.,[18] presented a technique through integrating OSN features with text of features to detect cyberbullying. Nonetheless, they never concentrate on indirect cyberbullying.

Mangaonkar et al.,[1] proposed collaborative paradigm and utilized various machine learning approaches for classification of bully or non-bully data. wherefore lack of very less features were used to train.

Van Hee et al., [19] proposed a supervised machine learning approach for detecting cyberbullying and as specific of cybervictimization. they presented construction and annotation (labelled) corpus of Dutch social media posts and explored the feasibility of automatic cyberbullying detection. however, in this study never applied few features for instance syntactic patterns, semantic information.

Galán-García et al.,[20] proposed supervised machine learning method for detect and associate fake ID profiles on Twitter social media network which are utilized for defamatory activities to read ID profiles within the same network through analyzing comments content generated by bother ID profiles.

Whereas, more Natural Language Processing (NLP) can be applied to increase the efficiency and also dataset from other OSN can be utilized.

4. DATA ACCESSIBILITY

In this section we provide an overview of the collection of data which were applied in previous research studies in area of cyberbullying detection., There are no standard which dataset utilized for cyberbullying detection. However, most of research studies used same OSNs in order to collect data for instance (YouTube, Twitter, Formspring etc.)

Table 1: Research studies in cyberbullying detection and datasets used.

Studies	Author	Year	OSNs Platform	Language Focused
1	Khandelwal et al. [21]	2020	Kaggle and OSN website TRAC	English
2	Yao et al. [22]	2019	Instagram	English
3	Chelmis et al. [23]	2019	Instagram	English
4	Cheng et al. [24]	2019	Instagram and Vine	English

[3] https://www.ncpc.org/resources/cyberbullying/what-is-cyberbullying/

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

5	Cheng et al. [25]	2019	Formspring and Twitter	English
6	Zois et al. [26]	2018	Twitter	English
7	Tahmasbi et al. [27]	2018	Twitter	English
8	Tomkins et al. [28]	2018	Twitter	English
9	Yao et al. [29]	2018	Instagram	English
10	Rafiq et al. [30]	2018	Vine	English
11	Agrawal et al. [31]	2018	Formspring, Twitter, Wikipedia	English
12	Van Hee et al. [19]	2018	ASKfm	English and Dutch
13	Singh et al.	2017	Instagram	English
14	Raisi et al. [32]	2017	ASKfm and Instagram	English
15	Dani et al. [33]	2017	Twitter and MySpace	English
16	Galán-García et al. [20]	2016	Twitter	English
17	Zhao et al. [34]	2016	Twitter and MySpace	English
18	Hosseinmardi et al. [35]	2015	Instagram	English
19	Rafiq et al. [36]	2015	Vine	English
20	Nahar et al. [37]	2014	MySpace, Kongregate, Slashdot	English
21	Huang et al. [18]	2014	Twitter	English
22	Dadvar et al. [38]	2013	YouTube	English

5. DATA ACCESSIBILITY CHALLENGE

In cyberbullying detection one of main challenge is collecting the data from OSNs, because it is not easy process since it related to data of privacy and OSNs and other online platforms websites don't disclose personal private information. However, this may cause of lack information such as friends of list can be retrieved. Moreover, data labelling or annotation is one of hardest part because it needs interference from professional experts to label the dataset as studied by [39]. If there were researcher who can share publicly accessible for other researchers that they have utilized, it would be an important research contribution all for all over the world.

6. REPRESENTATION OF DATA CHALLENGE

Many researchers only perform research related to online bullying words in telecommunication. Whereas, extracting the features which includes content text message have their own challenge. If any account of users does not provide personal information which includes, gender, location, gender performance in detection of cyberbullying may be disintegrate [40].
Language analyzing used by online users in order to describe range of age. It might take some to time to detecting word utilized in dataset that related to age. For instance, word 'study' may associate with online users with range 13 years between 18 years. In comprehensively define, to develop an adequate and proper system of cyberbullying detection is not easy since it is not easy to apply, because it involves human behavior actions and cyberbullying phenomenon or nature, that is hard to interpret.

7. OPEN RESEARCH ISSUES IN CYBERBULLYING DETECTION AND CHALLENGES

As a result of our research study of existing literature, we suggested some future research directions for advancing work on cyberbullying detection.

Haider et al., [41] proposed a Multilingual system for cyberbullying detection in the content of Arabic language. in their system they utilized two machine learning classification techniques such as SVM and Naïve Bayes on WEKA. However, in their research work the Multilingual system is limited to Arabic language and employed only two classification techniques.

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

Pawar et al., [42] presented multilingual cyberbullying detection model which is only focused on two famous Indian languages such as Marathi and Hindi. In their research they were applied supervised machine learning techniques on different manually labelled datasets of Marathi and Hindi languages and experiments results shows that F1 archived about 96% and accuracy archived nearly 97%. However, in this research work the they utilized small datasets.

Dadvar et al., [38] proposed a technique to detect cyberbullying on YouTube comments (online video sharing website), which is combination of features; such as online user based features, cyberbullying specific features and content based features cyberbullying of detection in social network videos sharing YouTube user comments.

León-Paredes et al., [43] proposed a system for Spanish language to detecting cyberbullying on Twitter network. The Spanish cyberbullying prevention system (SPC) which were based on Natural Language Processing (NLP) and machine learning techniques such as Logistic Regression, Naïve Bayes and Support Vector Machine classification. In their research the dataset collected from Twitter OSN which were 960,578 Spanish tweet text messages. In their research study the experimental results show that maximum accuracy achieved about 93% by utilizing SVM algorithm. However, this is research study only limited on Spanish language and they utilized limited features.

Akhter et al., [44] proposed a supervised machine learning method for detecting cyberbullying in Bangla text. In their research study they employed four machine learning classifiers such as KNN, SVM, J48 and NB. In addition, they collected data from social media posts which is written in Bangla language text (dataset comprised 2,400 posts). The experimental results show that SVM algorithm achieved accuracy 97%. However, in their research, they were tested on small dataset and this research is only focused on Bangla text messages and utilized limited only on user's information features which is collected from social media.

Lu et al., [45] presented multilingual Char-CNNS (Character-level Convolutional Neural Network with shortcuts) model which is detect cyberbullying in Chinese language and English language. in their research study the dataset collected from Chinese Weibo social networking website (Chinese comments) and Twitter OSN (for English language tweets messages). The experimental results show that F-measure achieved 71.6. However, this model only limited on two international languages and utilized char-level of features.

Mouheb et al., [46] proposed a system for detecting cyberbullying specifically in the context of Arabic language. In their research they collected data from Twitter OSN. In addition, the experimental analysis show that proposed system efficiently detecting the cyberbullying tweet messages. However, in their research study proposed system only limited on monitoring child activity and filter-based tweet posts which contain offensive words in Arabic language text.

Nurrahmi et al., [47] presented a system for detecting cyberbullying in Indonesian language tweets messages. In their research they detect actors and cyberbullying text. they were collected dataset from Twitter OSN and developed labelled dataset and classified into cyberbullying and non-cyberbullying. They utilized two machine learning classifiers such as KNN and SVM.

Furthermore, the experimental analysis shows the SVM classifier F1 score achieved 67%. However, in this research only limited on Indonesian text and used manually small dataset, this system can be utilized POST tagger tool for increasing classification of accuracy and feature extraction results.

Dalvi et al., [48] presented a machine learning model for to detecting cyberbullying on Twitter OSN. In their research study they collected dataset from different websites such as GitHub and Kaggle etc. and applied two machine learning classifiers techniques which includes Naïve Bayes and SVM. The experimental results show that SVM achieved 71.25 accuracy for detecting content of cyberbullying. Consequently, in their study limited text messages and not included other features which is commonly used in twitter network and other OSNs platform. In addition, this model can be applied in online gaming zone platforms such as PubG and counter strike. Moreover, in this model can be utilized other machine learning classifiers with large dataset.

Zhang et al., [49] proposed a method to automatically detect cyberbullying in Japanese language text from Twitter OSN. In their research study they utilized multiple textual features and used machine learning algorithms which includes

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

Random forest, decision tree, Gradient boosting regression, logistic regression, linear support vector machine and deep learning model. The experimental result achieved over 90%. However, in their research study were limited in textual features and Japanese language text. For further research in Japanese language extracting updated bullying phrases and words to detect cyberbullying in other OSNs.

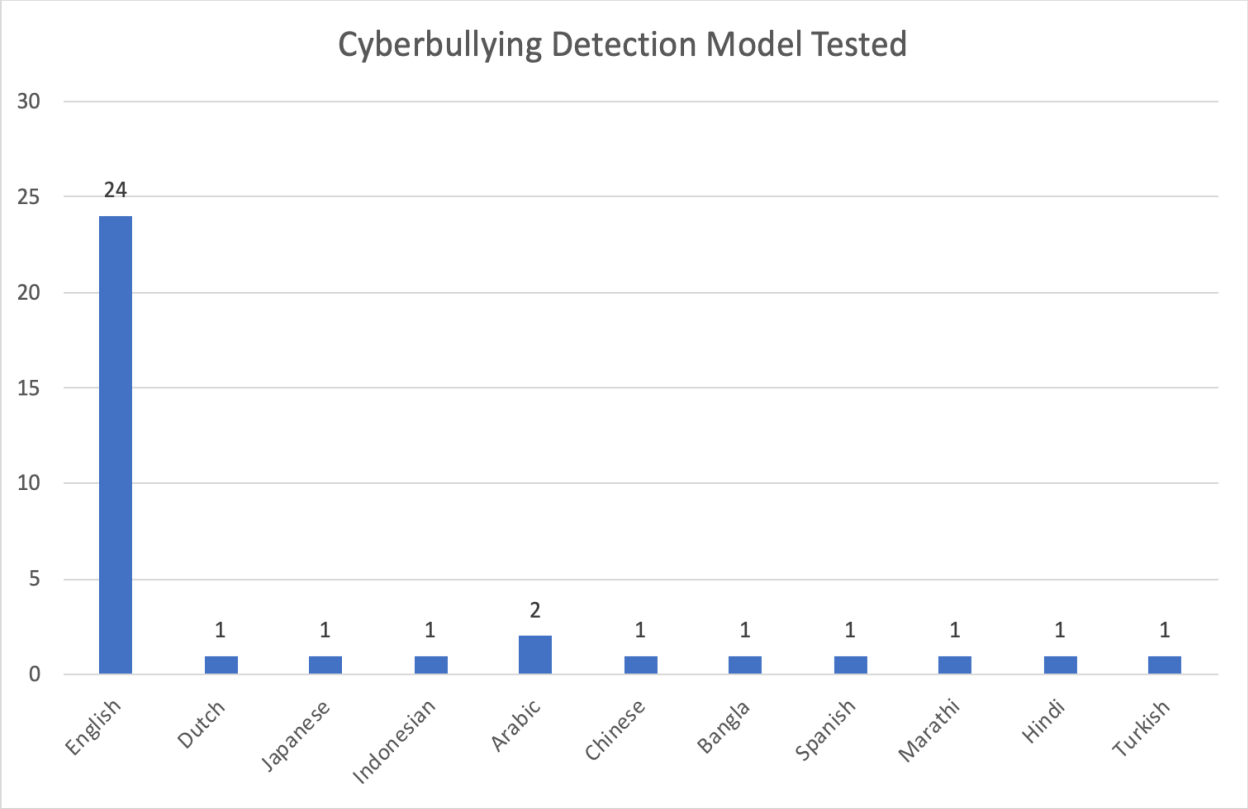

Figure 2. Proposed Cyberbullying Detection Models Tested in Different International Languages

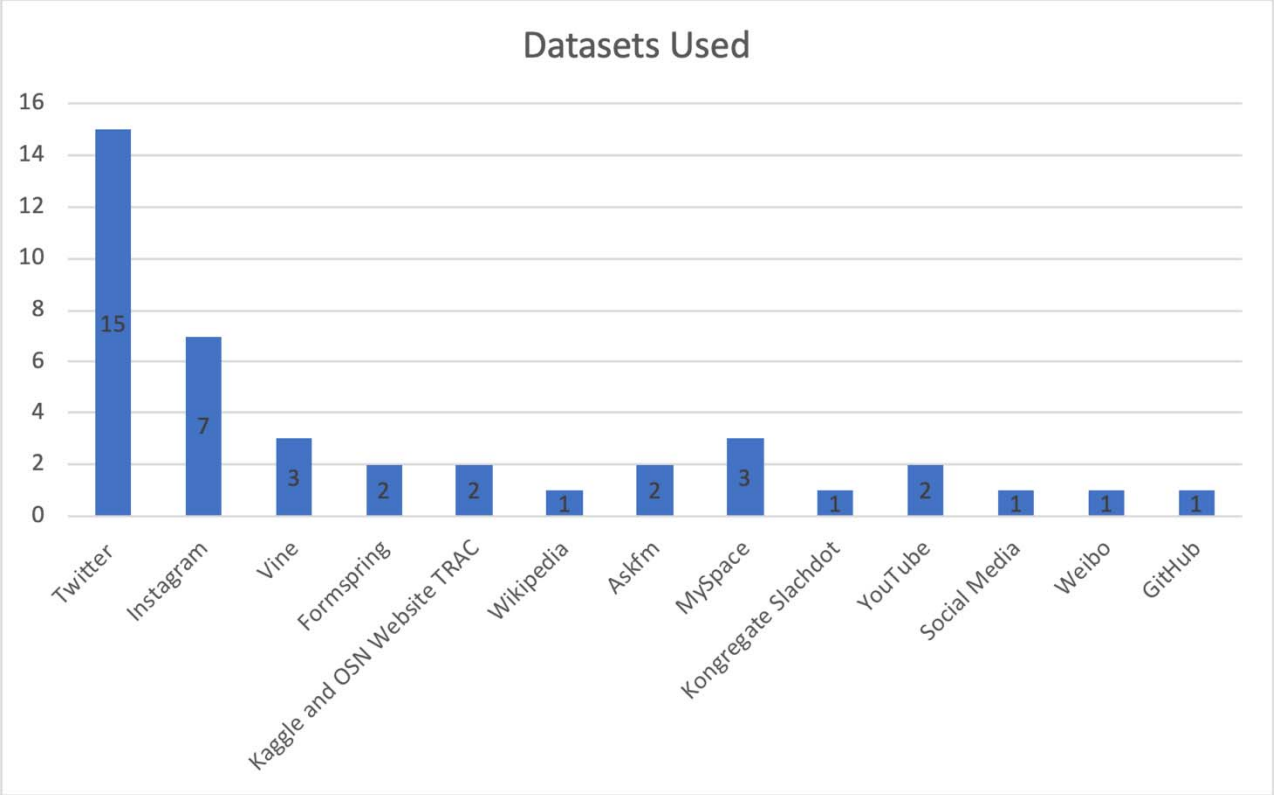

Figure 3. Various Datasets Used in Cyberbullying Detection.

8. CONLUSION

With rapid growth in ICT and internet, more and more humans connect each other in same town or rest of the world. Nonetheless, any new modern technology come with the potential for misuse. Unfortunately, these methods lead to harassment, misbehavior or online criminal act like cyberbullying. Moreover, it is very easier for online users to expand their human network particularly through OSNs.

Conversely, if people misuse or abuse social media in order to engage in cyberbullying, they may be classified as barbaric fellow of human being. Mostly of researcher has been worked on detecting bullying phrases, acronyms and keywords within the datasets, and utilizing text classification techniques in Natural Language Processing (NLP) and machine leaning methods.

In the future research addressing the role of modern technologies particularly mobile phones and peer-to-peer tools and devices should be considered for further research studies.

Also, research on cyberbullying may be able to apply deep learning since it can work adequately within text classification as research study conduct by [50] for spam detection. In the further research studies regards to cyberbullying may co-operated with other discipline fields such as sociologist and psychologist to increase the detection of cyberbullying.

9. ACKNOWLEDGMENTS

The authors would like to acknowledge Universiti Teknologi Malaysia (UTM) to support this paper is under the GUP grants and under the Grant ID of Q.K130000.2656.16J51

REFRENCES:

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

[2] Robin M. Kowalski, Susan P. Limber, Patricia W. Agatston, *Cyberbullying: Bullying in the Digital Age, 2nd Edition*. 2012.

[3] T. Mahlangu, C. Tu, and P. Owolawi, "A review of automated detection methods for cyberbullying," in *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, 2018, pp. 1–5.

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

[4] P. K. Smith, J. Mahdavi, M. Carvalho, S. Fisher, S. Russell, and N. Tippett, "Cyberbullying: Its nature and impact in secondary school pupils," *Journal of child psychology and psychiatry*, vol. 49, no. 4, pp. 376–385, 2008.

[5] S. Hinduja and J. W. Patchin, "Cyberbullying: Neither an epidemic nor a rarity," *European Journal of Developmental Psychology*, vol. 9, no. 5, pp. 539–543, 2012.

[6] J. W. Patchin and S. Hinduja, "Bullies move beyond the schoolyard: A preliminary look at cyberbullying," *Youth violence and juvenile justice*, vol. 4, no. 2, pp. 148–169, 2006.

[7] R. S. Griffin and A. M. Gross, "Childhood bullying: Current empirical findings and future directions for research," *Aggression and violent behavior*, vol. 9, no. 4, pp. 379–400, 2004.

[8] D. Maher, "Cyberbullying: An ethnographic case study of one Australian upper primary school class," *Youth Studies Australia*, vol. 27, no. 4, p. 50, 2008.

[9] N. E. Willard, *Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress*. Research press, 2007.

[10] A. Pascucci, V. Masucci, and J. Monti, "Computational Stylometry and Machine Learning for Gender and Age Detection in Cyberbullying Texts," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2019, pp. 1–6.

[11] C. L. Nixon, "Current perspectives: the impact of cyberbullying on adolescent health," *Adolescent health, medicine and therapeutics*, vol. 5, p. 143, 2014.

[12] T. Safaria, "Prevalence and Impact of Cyberbullying in a Sample of Indonesian Junior High School Students.," *Turkish Online Journal of Educational Technology-TOJET*, vol. 15, no. 1, pp. 82–91, 2016.

[13] D. Goebert, I. Else, C. Matsu, J. Chung-Do, and J. Y. Chang, "The impact of cyberbullying on substance use and mental health in a multiethnic sample," *Maternal and child health journal*, vol. 15, no. 8, pp. 1282–1286, 2011.

[14] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, "Detection of harassment on web 2.0," *Proceedings of the Content Analysis in the WEB*, vol. 2, pp. 1–7, 2009.

[15] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of Textual Cyberbullying.," *The Social Mobile Web*, vol. 11, no. 02, pp. 11–17, 2011.

[16] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," in *Machine learning and applications and workshops (ICMLA), 2011 10th International Conference on*, 2011, vol. 2, pp. 241–244.

[17] S. A. Özel, E. Saraç, 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.

[18] 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*, 2014, pp. 3–6.

[19] C. Van Hee *et al.*, "Automatic detection of cyberbullying in social media text," *PloS one*, vol. 13, no. 10, p. e0203794, 2018.

[20] P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying," *Logic Journal of the IGPL*, vol. 24, no. 1, pp. 42–53, 2016.

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

[22] M. Yao, "Robust Detection of Cyberbullying in Social Media," in *Companion Proceedings of The 2019 World Wide Web Conference*, 2019, pp. 61–66.

[23] C. Chelmis and M. Yao, "Minority Report: Cyberbullying Prediction on Instagram," in *Proceedings of the 10th ACM Conference on Web Science*, 2019, pp. 37–45.

[24] L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, "Xbully: Cyberbullying detection within a multi-modal context," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 2019, pp. 339–347.

[25] L. Cheng, R. Guo, and H. Liu, "Robust cyberbullying detection with causal interpretation," in *Companion Proceedings of The 2019 World Wide Web Conference*, 2019, pp. 169–175.

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

[26] D.-S. Zois, A. Kapodistria, M. Yao, and C. Chelmis, "Optimal online cyberbullying detection," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2017–2021.

[27] N. Tahmasbi and E. Rastegari, "A Socio-Contextual Approach in Automated Detection of Public Cyberbullying on Twitter," *ACM Transactions on Social Computing*, vol. 1, no. 4, pp. 1–22, 2018.

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

[29] M. Yao, C. Chelmis, and D.-S. Zois, "Cyberbullying detection on instagram with optimal online feature selection," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018, pp. 401–408.

[30] R. I. Rafiq, H. Hosseinmardi, R. Han, Q. Lv, and S. Mishra, "Scalable and timely detection of cyberbullying in online social networks," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 2018, pp. 1738–1747.

[31] S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms," in *European Conference on Information Retrieval*, 2018, pp. 141–153.

[32] E. Raisi and B. Huang, "Cyberbullying detection with weakly supervised machine learning," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 409–416.

[33] H. Dani, J. Li, and H. Liu, "Sentiment informed cyberbullying detection in social media," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2017, pp. 52–67.

[34] R. Zhao and K. Mao, "Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder," *IEEE Transactions on Affective Computing*, vol. 8, no. 3, pp. 328–339, 2016.

[35] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Detection of cyberbullying incidents on the instagram social network," *arXiv preprint arXiv:1503.03909*, 2015.

[36] R. I. Rafiq, H. Hosseinmardi, R. Han, Q. Lv, S. Mishra, and S. A. Mattson, "Careful what you share in six seconds: Detecting cyberbullying instances in Vine," in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2015, pp. 617–622.

[37] V. Nahar, S. Al-Maskari, X. Li, and C. Pang, "Semi-supervised learning for cyberbullying detection in social networks," in *Australasian Database Conference*, 2014, pp. 160–171.

[38] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Improving cyberbullying detection with user context," in *European Conference on Information Retrieval*, 2013, pp. 693–696.

[39] E. Raisi and B. Huang, "Cyberbullying identification using participant-vocabulary consistency," *arXiv preprint arXiv:1606.08084*, 2016.

[40] H. A. Schwartz *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, p. e73791, 2013.

[41] B. Haidar, M. Chamoun, and A. Serhrouchni, "Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content," in *2017 1st Cyber Security in Networking Conference (CSNet)*, 2017, pp. 1–8.

[42] R. Pawar and R. R. Raje, "Multilingual Cyberbullying Detection System," in *2019 IEEE International Conference on Electro Information Technology (EIT)*, 2019, pp. 040–044.

[43] G. A. León-Paredes *et al.*, "Presumptive Detection of Cyberbullying on Twitter through Natural Language Processing and Machine Learning in the Spanish Language," in *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2019, pp. 1–7.

[44] S. Akhter, "Social media bullying detection using machine learning on Bangla text," in *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, 2018, pp. 385–388.

[45] N. Lu, G. Wu, Z. Zhang, Y. Zheng, Y. Ren, and K.-K. R. Choo, "Cyberbullying detection in social media text based on character-level convolutional neural network with shortcuts," *Concurrency and Computation: Practice and Experience*, p. e5627, 2020.

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

[46] D. Mouheb, M. H. Abushamleh, M. H. Abushamleh, Z. Al Aghbari, and I. Kamel, "Real-Time Detection of Cyberbullying in Arabic Twitter Streams," in *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 2019, pp. 1–5.

[47] H. Nurrahmi and D. Nurjanah, "Indonesian Twitter Cyberbullying Detection using Text Classification and User Credibility," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 543–548.

[48] R. R. Dalvi, S. B. Chavan, and A. Halbe, "Detecting A Twitter Cyberbullying Using Machine Learning," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 297–301.

[49] J. Zhang, T. Otomo, L. Li, and S. Nakajima, "Cyberbullying Detection on Twitter using Multiple Textual Features," in *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, 2019, pp. 1–6.

[50] G. Jain and B. Agarwal, "An overview of RNN and CNN techniques for spam detection in social media," *Int J Adv Res Comput Sci Softw Eng*, vol. 6, no. 10, pp. 126–132, 2016.