INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 5)
Available online at: www.ijariit.com

A machine learning approach towards social media to tackle cyberbullying

Anjana J Mani
anjanajmani333@gmail.com
Mangalam College of Engineering, Ettumanoor, Kerala

Jinu P Sainudeen
jinu.sainudeen@mangalam.in
Mangalam College of Engineering, Ettumanoor, Kerala

ABSTRACT

The prevalence of social media is expanding step by step y. People of all age group are terribly interested in social networking. Social media connects people from different parts of the world. However, social media may have some side effects such as cyber bullying, which may have negative impacts on the life of people. Research shows that children and teenagers are the main victims of this cyber attack. Through the social media, people share their thoughts and emotions with their friends. There are large numbers of fraud accounts in social media. Cyber bullying is when someone, harass others on social media sites. Some people use it for cyber attack by making negative comments on others post. One way to tackle this problem is to detect those bullying messages and encrypt it. Machine learning techniques make automatic detection of cyber bullying messages. Weka is a powe full machine learning tool which can be used for this purpose. A combination of classification and lexical algorithms can detect whether a message is bullying or not. Cyber bullying is a major problem in society since social media has its presence in all fields of modern man's life.

Keywords— *Cyberbullying detection, Machine learning, Weka, Classification algorithms, Lexical analysis*

1. INTRODUCTION

Social media is the most popular innovation in the 21st century. A group of internet-based application which is built on the foundation of web 2.0[1]. Social media connects people from different parts of the world and they can share their opinion photos videos. Business persons use this as a medium for their marketing. Social media plays a major role in all fields whether it is business, politics, arts, politics or what not. Anywhere and everywhere there is an impact on social media. Social media has attained a great success in all fields and attracted people of different age groups. The disadvantage of social media is known as Cyber bullying which includes posting rumors, threats, sexual remarks, victim's personal information. Cyber bullying is harassing that happens over advanced gadgets like PDAs, PCs, and tablets. It can happen through SMS, Text, and applications, or online in web-based life, gatherings, or gaming where individuals can see, take part in, or share content and sending, posting, or sharing negative, unsafe, false, or mean substance about another person. This incorporates sharing individual or private data about another person causing shame or embarrassment. Posting negative comments regarding physical traits, religion, caste is a serious issue in society. A Study shows that cyber bullying victimization ranges from 10% to 40%as cited in [2]. About 43% of teenagers were bullied in USA [3]. This bullying has a negative impact on children.[4][5][6] and affect their education and personal life. We can prevent the side effects of cyber bullying by automatic detection of these bullying messages. There are numerous categories of cyber bully and different types of cyber bullying [7].

Namelessness and the absence of significant supervision in the electronic medium are two factors that have exacerbated this social threat. There are numerous internet based social networking sites, however, none of them gives a harassing free social condition. An effective way to solve this problem is to design a framework that will distinguish harassing words from the client contributions by contrasting it and the words put together by the Admin. The imperative part of digital tormenting is to check whether a given word is harassing or not rely upon the setting of sentence utilized. Machine learning is the best way to detect bullying messages [8][9]. The detection method can identify the presence of cyber bullying terms and classify cyber bullying activities in a social network such as Flaming, Harassment, Racism, and Terrorism [10].

2. RELATED WORKS

The rapid growth of social networking is supplementing the progression of cyber bullying activities. Most of the individuals involved in these activities belong to the younger generations, especially teenagers, who in the worst scenario are at more risk of suicidal attempts. One approach to detecting cyber bullying messages from social media through a weighting scheme of feature selection that presents a graph model to extract the cyber bullying network, which is used to identify the most active cyber bullying predators and victims through ranking algorithms [11].Detecting bullying messages can reduce the impact of bullying on victims. Machine learning is the best option for this detection. Another attempt is to try different things with a corpus of 4500 YouTube remarks, applying a scope of parallel and multiclass classifiers. Then locate that paired classifiers for singular names beat multiclass classifiers.

These discoveries demonstrate that the recognition of literary digital harassing can be handled by building singular point delicate classifiers [12].

Cyber bullying has demonstrated weighty to youth Internet clients and past techniques depended intensely on the utilization of physically created word references. This task portrays starter comes about for a framework that utilizations Latent Semantic Indexing (LSI) for the identification of digital harassing in a named gathering of posts from Form spring me. After pre-handling to represent varieties in spelling and utilization of emojis, a pursuit framework was produced. This framework fundamentally beats the pattern with an exceptionally basic inquiry and isn't subject to a word reference of tormenting terms [13].

Another method developed makes use of substance depiction shown in perspective of a variety of SDA: underestimated stacked Denoising auto encoders (mSDA), which gets straight as opposed to nonlinear projection to stimulate getting ready and limits interminable uproar scattering to take in more solid depictions. Then utilized semantic information to develop mSDA and make Semantic-enhanced Marginalized Stacked Denoising Autoencoders (smSDA).

The semantic information involves bothering words. A modified extraction of tormenting words in perspective of word embeddings is proposed with the goal that the included human work can be reduced. In the midst of planning of smSDA, we attempt to imitate harassing features from another run of the mill words by finding the sit out of gear structure, i.e. association, among tormenting and common words. The nature behind this musing is that some irritating messages don't contain tormenting words [14].

3. EXISTING SYSTEM

There are numerous social networking sites however none of them gives a tormenting free social condition. Social networking services offer friends space where they can maintain their relationships, chat with each other and share information. It gives an opportunity to make new friends via mutual friends. On the first use of the system, users are required to submit a profile containing personal information such as their name, date of birth, and a photo. The personal information is made available to other users of the system and is used to identify friends on the network and to add them to a list of contacts. In most systems, users cannot only view their friends but also second-degree friends (friends of their friends). Some networks follow an “invitation only” approach. Hence, every person in the system is automatically connected to at least one other person [15]. Users post their photos videos and thoughts through social media and their friends can comment on these post and some people post bullying messages as comments. This would severely influence the victims. Cyber bullying can be especially destructive as usually an open type of embarrassment and numerous others can perceive what is composed or posted. When something is distributed on the web, it is troublesome if not difficult to evacuate all hints of it. Automatic discovery of tormenting message is the best answer to handle this issue.

4. PROPOSED SYSTEM

There are many social media websites but none of them provides a bullying-free social environment. A system that can automatically detect bullying messages is set using the Machine Learning instrument Weka. The system will recognize tormenting words from the customer commitments by differentiating it and the words set up together by the Admin and past customers. The basic piece of computerized irritating is to check whether a given word is tormenting or not depend upon the setting of sentence used. A sentence can make sure or negative depends upon the earnestness and hugeness of the words used as a piece of it. We can't expect that a sentence with tormenting word is continually negative thusly we require an instrument to take in the sentence and to choose if it's a positive or negative one. Weka can do this with its awesome plan and gathering counts.

Weka is an accumulation of machine learning calculations for information mining errands. The calculations can either be connected specifically to a dataset or called from your own particular java code. Weka contains apparatuses for information pre-handling, grouping, relapse, bunching, affiliation principles, and perception. It is additionally appropriate for growing new machine learning plans. Weka makes an extensive number of characterization calculations accessible. The expansive number of machine learning calculations accessible is one of the advantages of utilizing the Weka stage to work through your machine learning problems.5 top characterization calculations in Weka Logistic Regression Naive Bayes, Decision Tree, k-Nearest Neighbor Support Vector Machines.

Digital tormenting location framework makes utilization of the Naive Bayes calculation. Lexical examination and characterization calculation cooperate in the framework for better execution Lexical investigation, lexing or tokenization is the way toward changing over a grouping of characters, (for example, in a PC program or page) into a succession of tokens (strings with an allotted and along these lines recognized importance). A lexical token or just token is a string with an allotted and along these lines recognized the significance. It is organized as a couple comprising of a token name and a discretionary token esteem. The token name is a class of lexical unit. There are mainly three modules for this system. Admin, user, and cyber authority. The user needs to register first. During registration, they get a username and a password. The user can log in to the system by submitting this username and password. Cyber bullying detection system provides all functionalities that are basically provided by all social networking sites. The user can send a friend request, accept the friend request and can upload photos and can post any events or their views regarding any political or social issues. The user can chat with their friends. They can find new friends. When the user performs these activities it may lead to cyber bullying either purposefully or unknowingly. The system will automatically detect the bullying words and encrypt it thus the user can see those bullying messages and comments in an encrypted format.

The admin is authenticated with the system using a username and password. Admin can keep track of the bullying users. Admin can analyze which all users are using bullying words, where it is used more whether in posts comments or chats. Admin has the capability to add a third party Cyber. Admin adds them by giving a secure username and password. Admin has the power to block the bullying user. Admin has the provision for adding and editing bullying words which are used by the system to perform classification and clustering using Weka tool. The cyber is authenticated with a username and password provided by the Admin later the cyber has the provision for changing this username and password. The cyber can obtain the bullying user details from the Admin then the cyber can take disciplinary based on the severity of the cyber bullying.

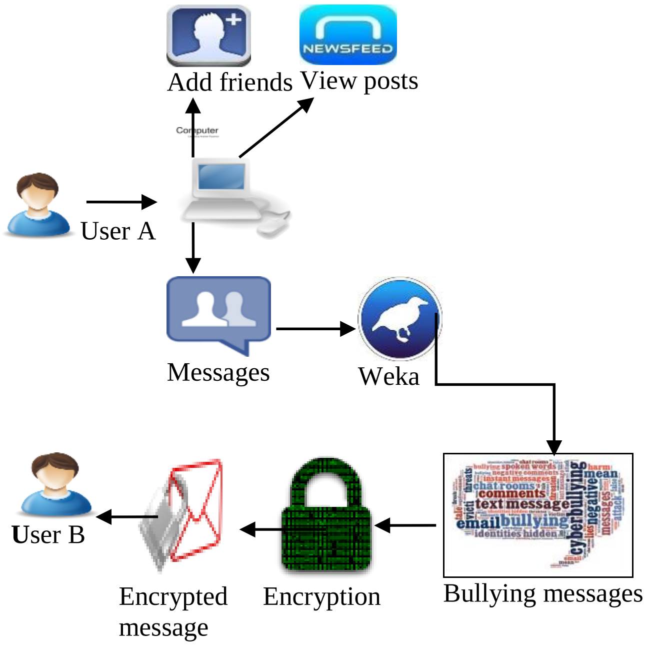

Fig. 1: Working of cyber bullying detection system

5. RESULT

An ever-increasing number of young people in online networks are presented to and hurt by digital harassing. Studies demonstrate that in Europe around 18% of the kids have been associated with digital harassing, prompting serious sorrows and even suicide endeavors. Digital tormenting is characterized as a forceful, deliberate act did by a gathering or individual, utilizing electronic types of contact more than once or after some time, against a casualty who can't without much of a stretch guard him-or herself.

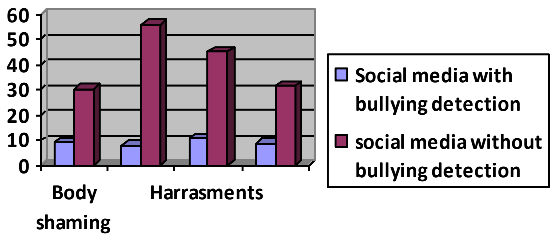

Fig. 2: Comparison of cyber crime rate with and without bullying detection

One way to deal with address the computerized tormenting issue is to thus distinguish and quickly report tormenting messages with the objective that fitting measures can be taken to neutralize advanced infringement. The consequence of the trials appears in the accompanying chart. From the chart plainly location enhances execution when we include all the more harassing particular highlights and that it enhances facilitate when setting data is included. One approach to address the digital tormenting issue is to consequently distinguish and speedily report harassing messages so legitimate measures can be taken to avoid digital wrongdoings. Exploratory outcomes demonstrate that digital wrongdoing rates are brought down when web based life with harassing location is utilized.

6. CONCLUSION

Cyber bullying is the utilization of innovation to irritate, debilitate, humiliate, or focus on another person. The results for casualties under digital harassing may even be awful, for example, the event of self-harmful conduct or suicides. One approach to address the digital harassing issue is to naturally distinguish and immediately report tormenting messages with the goal that appropriate measures can be taken to counteract conceivable tragedies. Online networking is where a large portion of this digital tormenting happens. One approach to keep this is the programmed location of harassing words and scrambles those words. Machine learning systems can recognize the harassing words. In this manner, WEKA is utilized for this reason and it gives more precise outcomes to the checking of tormenting words. The digital tormenting identification framework in this manner gives a harassing free online life condition for its clients.

7. REFERENCES

[1] A. M. Kaplan and M. Haenlein, “Users of the world, unite! The challenges and opportunities of social media,” Business horizons, vol. 53, no. 1, pp. 59–68, 2010.

[2] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, “Bullying in the digital age: A critical review and meta-analysis of cyber bullying research among youth.” 2014.

[3] M. Ybarra, “Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression, “National summit on interpersonal violence and abuse across the lifespan: Forging a Shared Agenda, 2010.

[4] B. K. Biggs, J. M. Nelson, and M. L. Sampilo, “Peer relations inthe anxiety–depression link: Test of a mediation model,” Anxiety, Stress, & Coping, vol. 23, no. 4, pp. 431–447, 2010.

[5] S. R. Jimerson, S. M. Swearer, and D. L. Espelage, Handbook of bullying in schools: An international perspective. Routledge/Taylor & Francis Group, 2010.

[6] G. Gini and T. Pozzoli, “Association between bullying and psychosomaticproblems: A meta-analysis,” Pediatrics, vol. 123, no. 3, pp. 1059–1065, 2009.

[7] Demystifying and Deescalating Cyber Bullying ‘Barbara Trolley, Ph.D. CRCConnie Hanel, M.S.E.d & Linda Shields, M.S.E.d

[8] A. Kontostathis, L. Edwards, and A. Leatherman, “Text miningand cybercrime,” Text Mining: Applications and Theory. John Wiley& Sons, Ltd, Chichester, UK, 2010.

[9] J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, “Learning from bullying traces in social media,” in Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies. Association for Computational Linguistics, 2012, pp. 656–666.

[10] 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.

[11] An Effective Approach for Cyberbullying Detection Vinita Nahar1, Xue Li2, Chaoyi Pang3 1, 2School of Information Technology and Electrical Engineering, the University of Queensland, Brisbane, Queensland 4072, Australia.

[12] Modeling the Detection of Textual Cyberbullying. Karthik Dinakar, Roi Reichart Henry, Lieberman. MIT Media Lab, Computer Science & Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge, MA 02139 the USA.

[13] Cyber Bullying Detection Using Social and Textual Analysis Qianjia Huang Department of Applied Computer Science The University of Winnipeg Winnipeg, MB, Canadahuangq17@webmail.uwinnipeg.ca Vivek K. Singh The Media Lab Massachusetts Institute of Technology Cambridge, MA, USA singhv@mit.edu Pradeep K. Atrey

Department of Computer Science University at Albany-SUNY Albany, NY, USA patrey@albany.edu.

[14] Trends in technology-based sexual and nonsexual aggression over time and linkages to non-technology aggression Michele Ybarra MPH Ph.D. Center for Innovative Public Health Research.

[15] On Social Network Web Sites: Definition, Features, Architectures and Analysis Tools Vala Ali Rohani Department OfSoftware Engineering Faculty of Computer Science and Information Technology University Of Malaya Kuala Lumpur, Malaysia V.Rohani@perdana.um.edu.my Ow Siew Hock.