International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Identification of Cyber Defamation in Social Network Using Machine Learning

[1]Minakshi Gururaj Sangolagi, [2]Khedkar Pooja Rajkumar, [3]Mohurle Ruchika Ashok, [4]Gakare Kirti Vikas, [5]Prof. Archana Priyadarshani

[1,2,3,4,5]Dhole Patil College of Engineering, Pune

Abstract —

It's important to note that cyberbullying may take place on any site that allows individuals to see, engage with, or share content, not only social media sites. Abusive or bullying behavior is easily recognizable since it lasts for an extended period of time and has an evil intent. Examples of cyberbullying include sending or sharing anything that is harmful, threatening, or offensive. Using the review paper provided below, we were able to do an effective and comprehensive analysis of the current cyberbullying identification techniques presented in this paper. Owing to these literature reviews, we were able to develop a method for identifying cyberbullying that combines the usage of Term Weight, Shannon Information Gain, and Fuzzy Classification. The methodology given above will be detailed effectively in this study's next iteration which will have even more emphasis with analysis of the approach with in-depth experimentation.

Keywords: *Cyberbullying detection, Term Weight, Shannon Information Gain, Fuzzy Classification.*

INTRODUCTION

Web, MBs, forums, and SM use is widespread among people of every age these days. With the proliferation of various online communication technologies, bullies have begun to transition their activities from the actual world to the virtual one. Whether perpetrated by a person or a group, bullying is a pervasive problem across the globe that leaves its victims defenseless. As its usage grows in all parts of society, online communication has become more important to the everyday lives of a large number of individuals. Due to the widespread nature of online social media, cyberbullying potentially happen to anyone, anywhere at any time, and in any place; moreover, the relative anonymity afforded by the internet makes it more difficult to stop similar perceived accusations than traditional bullying. Cyberbullying continues to be a topic of study, despite the fact that it is increasingly difficult to pinpoint the structural inequalities, encouragement, and deadly commitments that are often associated with old fashioned bullying due to its online context and cover of privacy and confidentiality. One definition of cyberbullying is the repeated use of electronic communication technologies to harass or intimidate a target. It’s really disturbing that cyberbullying is so commonplace. Bullying in its digital information occurs on various digital devices, such as computers, laptops, and tablets, and is known as cyberbullying. Cyberbullying may happen everywhere people can connect and converse resources, including in chat rooms, on message boards, in apps, on social media, and even during live streaming of video games. Some forms of cyberbullying include intimidation, accusations, harsh language, exclusion, confidentiality intrusions, intimidation, institutional discrimination, and defamation. When the victim has severe injuries, the act in question is considered criminal.

RELATED WORKS

This topic is discussed in Krithika V [4]. The scope of this study contained wide range of approaches that overcame a number of challenges associated with identifying cyberbullying through imagery and text. On the other hand, there has been an increase in the percentage of occurrences of bullying as well as allegations of it. This is as a result of the insufficient level of protection provided by a variety of websites. Teens are heavy users of social media websites such as Facebook, Instagram, and WhatsApp; nevertheless, no behavior identification schemes have really been built for these platforms. Teenagers may very well be the victims of cyberbullying for a variety of reasons, including but not limited to the dissemination of objectionable information online or the formation of connections with unknown individuals. This is a contributing factor in people attempting suicide, experiencing anxiety, and having other psychological issues. Supplemental studies in this area will be required to offer additional improved techniques and processes that really can clearly identify bullying posts and comments and immediately notify the poster even before to the posts being constructed, in addition to detecting bully texts in the process of being authored. This could be necessary in order to make these improvements.

Vivek K. Singh [5] justifies and motivates the usage of network characteristics, including network importance, as a proactive characteristic to test computational impartiality. This use is supported by the fact that network centrality is a system feature. An known method for identifying instances of cyberbullying was evaluated, and the results showed that the performance of the algorithm varied rather considerably depending on the individual's network centrality in relation to the potentially abusive text. It has been demonstrated that making recourse to a post-processing strategy known as adjusted odds greatly reduces the magnitude of this output gap. The results, notwithstanding the fact that they are premature, provide a significant addition to the existing body of expertise on the detection of cyberbullying along with the stability of interconnected technologies.

In the present world, a significant issue that has to be addressed is cyberbullying of youngsters, as stated by Farhan Bashir [6]. It has been stated that cyberbullying has been a factor in the attempted and successful suicide of children and adolescents. Earlier researches generally, for the most part, ignored undergrads while focusing heavily on elementary, middle, and high school students. This study concentrated its research on concerns that are unique to young adults in attempt to better understand and assess cyberbullying. The practitioners are of the opinion that subjecting the TPB and SCT assumptions on the practice of cyberbullying towards the assessment could be beneficial use to researchers, academicians, parliamentarians, the Malaysian government, and experts by exposing essential facts. The purpose of this study was to evaluate the intellectual and situational factors that lead graduate students in Malaysian private and public institutions to participate in cyberbullying.

A cyberbullying detection method that draws across several fields of study was developed by Rohit Pawar [7] with the intention of detecting instances of cyberbullying in text messages and tweets, as well as in publication articles written in a number of Indian dialects. The findings of their experiments indicate that the Logistics Regression technique demonstrates greater efficiency on some of these datasets in comparison to any other approach. In particular, the production of content that has been consolidated would be of assistance in improving the functioning of the service. The results of this research show that these technologies are successful in more than two languages and multiple domains, which means that it is feasible to use them to detect instances of cyberbullying in a wide range of Indian languages too. The researchers have not looked at the usage of translators, contrasting the efficacy of different procedures, or pursuing other avenues of investigation such as natural language processing (NLP).

Farhan Bashir Shaikh [8] contributed an important participation by analyzing the factors alluring university college students to participate in cyberbullying behavior and offering an in-depth perception of the modules resulting to cyberbullying actions. This is in comparison to the traditional technique of focusing attention among one or two variables, which is intended to cover fewer ground. The cyberbullying factors that have been found may indeed be utilized as a guide to streamline the algorithm that is employed to predict the behavior of cyberbullying. Cyberbullying is a big issue in today's culture, especially among young people. As a direct consequence of being bullied online, several young people have apparently tried as well as managed to take their own lives. Previous studies have, for the most part, ignored university students in favor of putting substantial focus on elementary, middle, and high school students. Their study concentrated on identifying features that are unique to university students in attempt to provide a better understanding and analysis of the issue of cyberbullying.

An exclusive cyberbullying theoretical model has reportedly served as the basis for a recommended exclusion cyberbullying lexicon, as stated by Ong Chee Hang [9]. The dictionary includes many terminology related to cyberbullying, along with their meanings and categorizations. The Enchanted Learning negative vocabulary keyword collection was used in the selection process for these terms. The categorization was determined using the selected exclusion criteria for cyberbullying, while the meaning of the phrases was gleaned from the online English vocabulary database FrameNet. It would be highly beneficial for users of social networking platforms to comprehend which phrases were considered to be employed in cyberbullying if they use the proposed vocabulary as a dictionary. Users may use the vocabulary as a dictionary in the same way. It is possible that this might be another technique to reduce the amount of cyberbullying that occurs on social media platforms. In addition to this, it may be of use in identifying instances of cyberbullying that occur on social media platforms.

Using the Twitter API, Djedjiga Mouheb [10] proposes a system that may immediately detect instances of cyberbullying that occur inside Arabian tweets. The technique that has been presented screens posts based on the offensive keywords that are included within them and classifies them according to the degree to which the offensive words are used. Because of this, the system is able to carry out the appropriate procedures that are in line with the parameters that have been previously defined. The strategy that has been recommended is something that parents might use to monitor what their children are doing online and protect them from the dangers of cyberbullying. The Middle East's ongoing campaign versus cyberbullying has reached a new plateau because to the success of this program. The capacity of the system to identify instances of cyberbullying based on the content of the posts will be improved with the use of machines learning algorithms.

LSHWE, which stands for similarity-based embeddings, is a technique that was proposed by Zehua Zhao [11] in order to solve the problem of cyberbullying recognition jobs including keywords that have been purposely obscured. The first and second stages of the LSHWE. The first step is to build a co-occurrence vector C, a frequent lists of words R, a closest neighbor list NL obtained by local binary pattern hashing, and a k nearest matrix N for a training string. After that, the string will be used to train a machine learning model. Second, an LSH-based auto-encoder is used to train the word embeddings to be what they should be. Two aspects of the embedding strategy that have been proposed are designed in such a way that LSHWE can successfully reflect unique words. Because LSHWE is a method to word embedding that is based on universal similarities, the representations of unusual words that are learnt via the use of LSHWE should be as equivalent as feasible to the representations of the basic phrases to which they correspond. LSHWE is another algorithm that has shown to be rather successful.

As shown by Syarifah Qonitatulhaq [12], it is essential to provide videos with explanations for the purpose of assisting children in gaining an understanding of cyberbullying. In addition, with the assistance of a respondent who provided the data shown in the schematic diagram, the individual in question fabricated an engaging and relevant story. An explanatory video that has some kind of interactive component will be more suited to the audience for whom it is designed. This video explanation has the potential to be expanded upon by being divided into a number of shorter videos, each of which would explain a different idea. In addition to developing expertise in animation, one needs also get familiar with the components and transformations used.

Following Saloni Mahesh Kargutkar [13], the technology breakthrough not only enhanced life expectancy but also made it easier for criminals to do harmful acts in an atmosphere that was conducive to their activities. Offenses committed online have now reached an extremely lethal level due to the fact that victims are always being searched for and have no opportunity of evading capture. Cyberbullying is one of the most serious acts that can be committed online, and several studies have demonstrated how badly it affects those who are bullied. This research presents a novel approach to the problem of identifying whether or not remarks made on Twitter about cyberbullying should be categorized as such. The approach makes it possible to get accurate results by making use of an efficient method of CNN application that is based on Keras. The approach that has been suggested may be used by the administration as well as any other university, supervisors, corporations, law enforcers, or regulating authorities.

Jason Wang [14] developed an effective and fine-grained categorization technique to combat cyberbullying with the goal of identifying malicious tweets prior to their progression into cyberbullying. The researchers begin by demonstrating that Dynamic Query Expansion effectively minimizes the imbalance issue. As a result, they recommend using it in more social media data analysis including natural language processing situations. In addition, researchers make exhaustive observations to build fine-grained cyberbullying recognition and exhibit equivalent improvements to prior binary cyberbullying identification studies. They do this by using a range of embedding approaches and classifier design configurations. In the end, these data suggest that the efficacy of the assumed Graph Convolutional Network design SOSNet, which harnesses the natural linguistic linkages between tweets, is already on par with or even better than that of traditional filters in this particular field of study.

Yuzana Win [15] developed a method for the identification of bully words in Myanmar writing by making use of a support vector machine classification algorithm. This method was published. In order to determine what constitutes bullying, the method used various processes. By employing a classification method that was based on a Support Vector Machine, the authors investigated whether or not the bullying-related terms that were obtained by the algorithm improved the accuracy of the classification. The recommended method was assessed by classifying comments and posts from Myanmar into one of two categories: positive or negative. The majority of the data set was arbitrarily split between training and validation, while the remaining data was used for testing. Throughout the course of the inquiry, the researchers use a tool called the TF-IDF vectorizer to convert concepts such as sexuality, violence, and profanity, as well as psychological, physical, and neutral keywords, into feature vectors.

PROJECT SCOPE

The purpose of this SRS document is to provide a detailed overview of our software product “Unmasking Cyberbullies on Social Networks”, its parameters and goals. This document describes the project’s target audience and its user interface, hardware and software requirements. It defines how our client; Unmasking Cyberbullies on Social Networks team and audience see the product and its functionality. The proposed methodology provides a good way to cyberbullying detection techniques that have been outlined to achieving our methodology for cyberbullying detection that utilizes Term Weight, Shannon Information Gain and Fuzzy Classification.

The major scope of this project is as follows: Easy to deploy the system. Easy user interface. Need moderate amount of data. Improved accuracy in Cyber Bullying Detection.

SYSTEM ARCHITECTURE

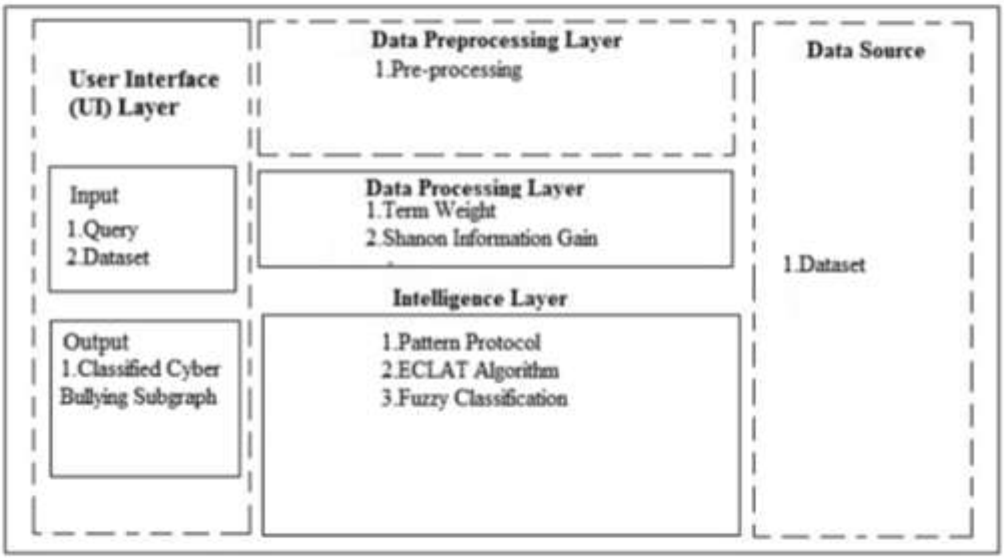

Fig.1 System Architecture

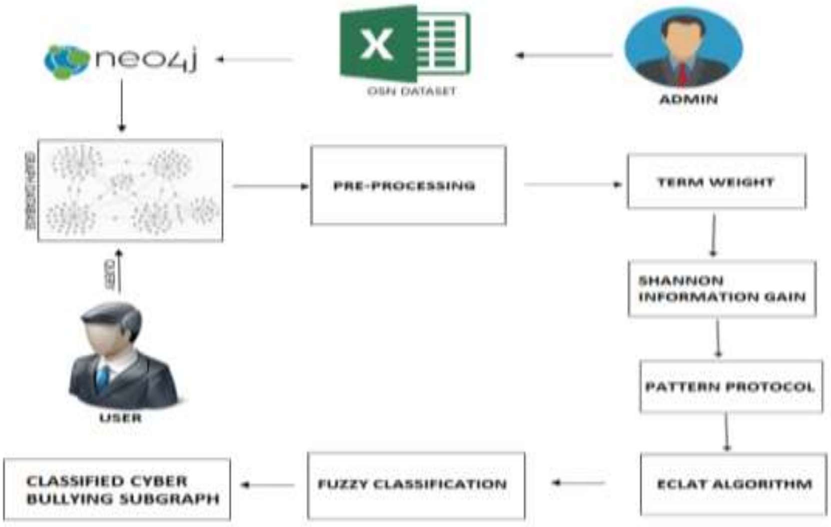

Fig.2 System Design

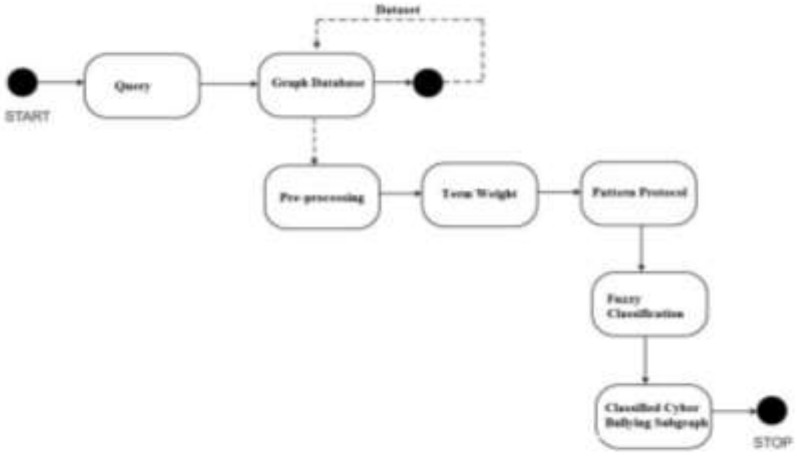

Fig.3 State Transition Diagram

CONCLUSION AND FUTURE SCOPE

Cyberbullying is a major issue because of the rise of the Internet and other forms of electronic communication. It's been defined as repeatedly harming or embarrassing someone using a mobile device, such as a smartphone, chat app, or media service. In this case, recognizing cyberbullying is essential for preserving social health. The task is challenging considering the importance placed on individuality and the connections that may be drawn between different factors, such as age, gender, and beliefs. The vocabulary and level of linguistic knowledge of users on different social media platforms varies greatly. The existing methods for detecting cyberbullying have been presented in this research piece, and this survey paper was employed to achieve an effective and complete study of those methods. Our strategy for detecting cyberbullying employs Term Weight, Shannon Information Gain, and Fuzzy Classification, and we owe a great deal to the articles discussed here. Future iterations of this study will have detailed demonstrations of the method given here.

REFERENCES

[1] A. Aggarwal, K. Maurya and A. Chaudhary, "Comparative Study for Predicting the Severity of Cyberbullying Across Multiple Social Media Platforms," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 871-877, doi: 10.1109/ICICCS48265.2020.9121046.

[2] A. Pradhan, V. M. Yatam and P. Bera, "Self-Attention for Cyberbullying Detection," 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), 2020, pp. 1-6, doi: 10.1109/CyberSA49311.2020.9139711.

[3] V. Banerjee, J. Telavane, P. Gaikwad and P. Vartak, "Detection of Cyberbullying Using Deep Neural Network," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 604-607, doi: 10.1109/ICACCS.2019.8728378.

[4] V. Krithika and V. Priya, "A Detailed Survey On Cyberbullying in Social Networks," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-10, doi: 10.1109/ic-ETITE47903.2020.031.

[5] V. K. Singh and C. Hofenbitzer, "Fairness across Network Positions in Cyberbullying Detection Algorithms," 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2019, pp. 557-559, doi: 10.1145/3341161.3342949.

[6] F. B. Shaikh, M. Rehman, A. Amin, A. Shamim and M. A. Hashmani, "Cyberbullying Behaviour: A Study of Undergraduate University Students," in IEEE Access, vol. 9, pp. 92715-92734, 2021, doi: 10.1109/ACCESS.2021.3086679.

[7] R. Pawar and R. R. Raje, "Multilingual Cyberbullying Detection System," 2019 IEEE International Conference on Electro Information Technology (EIT), 2019, pp. 040-044, doi: 10.1109/EIT.2019.8833846.

[8] F. Bashir Shaikh, M. Rehman and A. Amin, "Cyberbullying: A Systematic Literature Review to Identify the Factors Impelling University Students Towards Cyberbullying," in IEEE Access, vol. 8, pp. 148031-148051, 2020, doi: 10.1109/ACCESS.2020.3015669.

[9] O. C. Hang and H. M. Dahlan, "Cyberbullying Lexicon for Social Media," 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS), 2019, pp. 1-6, doi: 10.1109/ICRIIS48246.2019.9073679.

[10] D. Mouheb, M. H. Abushamleh, M. H. Abushamleh, Z. A. Aghbari and I. Kamel, "Real-Time Detection of Cyberbullying in Arabic Twitter Streams," 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2019, pp. 1-5, doi: 10.1109/NTMS.2019.8763808.

[11] Z. Zhao, M. Gao, F. Luo, Y. Zhang and Q. Xiong, "LSHWE: Improving Similarity-Based Word Embedding with Locality Sensitive Hashing for Cyberbullying Detection," 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9207640.

[12] S. Qonitatulhaq, N. Astin and W. Sarinastiti, "Creative Media for Cyberbullying Education," 2019 International Electronics Symposium (IES), 2019, pp. 622-627, doi: 10.1109/ELECSYM.2019.8901646.

[13] S. M. Kargutkar and V. Chitre, "A Study of Cyberbullying Detection Using Machine Learning Techniques," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 734-739, doi: 10.1109/ICCMC48092.2020.ICCMC-000137

[14] J. Wang, K. Fu and C. -T. Lu, "SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1699-1708, doi: 10.1109/BigData50022.2020.9378065.

[15] Y. Win, "Classification using Support Vector Machine to Detect Cyberbullying in Social Media for Myanmar Language," 2019 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), 2019, pp. 122-125, doi: 10.1109/ICCE-Asia46551.2019.8942212.