iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** II **Month of publication:** February 2018

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

An Effective System for Automatic Detection & Prevention of Cybercrime in Micro Blog Systems

S.Kalaiselvi[1], J.Karthiyayini[2]
[1]*Dept of ISE, The Oxford College of Engineering, Bangalore, Karnataka, India,*
[2] *Dept of ISE, New Horizon College of Engineering, Bangalore, Karnataka ,india,*

Abstract: ***The Rapid growth of the social Networking sites is supplementing the progression of cyber crime activities. Cyber bullying is harassing or insulting a person by sending messages of hurting or threatening nature using electronic communication. Cyber bullying poses significant threat to physical and mental health of the victims Most of the individuals involved in these activities belong to the younger generations, especially teenagers. In particular, Cyber bullying may cause many serious and negative impacts on person's life and leads to teen suicide. Detection of cyber crimes and the provision of subsequent preventive measures are the main courses of action to combat cyber crimes. The proposed method is an effective method to detect and prevent cyber crime activities on social media. The detection method can identify the presence of cyber crime terms and classify cyber crime activities in social network such as Flaming, Harassment, Cyber-stalking, Denigration, impersonation and Terrorism, using machine learning algorithm. We propose a model to identify the most active predator and victim by extracting the text, user and network-based attributes and preventing the post of bully words by predator before access the social media.***

Keywords: Cyber bullying, Social Networks, Feature selection, SVM classifier, Victim, Cybercrime

I. INTRODUCTION

Due to advances of internet, Online Social Network (OSN) services or micro blog systems, such as Twitter, Face book are gaining in popularity as a source of communicating messages to other people. Messaging is widely used & useful in various purposes, example, business, education &socialization. More than millions of users have used websites as communication tools and as real-time, dynamic data sources. With the advent of web2.0, the social networks gained much popularity ever since the launch of the first social network sixDegrees.com in 1997[19].While web2.0 provides easy, interactive, anytime and anywhere access to the online communities, it also provides an opportunities for cybercrimes like cyber bullying, cyber aggression and Rumors. Cyber bullying involves the offensive information such as harassment, insult, and hate in the messages which are sent or post using OSN services for the purpose of intentionally hurting people emotionally, mentally, or physically [2].cyber bullying has also been extensively recognized as a serious problem, in which victims demonstrate a significantly high risk of suicidal ideation. One study conducted by national anti-bullying charity has shown that two out of three 13-22 years old who were surveyed have been victims of cyber bullying [3]. Cyber aggression as intentional harm delivered via electronic means to a person or a group of people who perceive such acts as offensive, deogatory, harmful or unwanted [18]. Rumors is also considered as a serious problem in OSN along with bullying and aggression. A rumor is a claim whose truthfulness is in doubt and has no clear source; even if its ideological origins and intents are clear [7]. Its tragic consequences have continuously reported among the youth. Recent research has shown that most teenagers experience bullying during their online activities including mobile usage, and also while involved in OSN.As highlighted by the National Crime Prevention Council, approximately 50% of the youth in America are victimized by cyber bullying[1].Since the number of cyber bullying ,cyber aggression and rumors experiences has recently been increasing[4],an intensive study of how effectively detect and prevent it from happening in real time manner is urgently required. As the problems mentioned, a number of studies are dedicated to explore various techniques to detect cyber bullying efficiently. Manual detection is considered the most accurate detection, but it is hard because it takes too much effort and resources. Automatic cyber bullying detection is therefore emphasized. Even though cyber bullying, cyber aggression and rumors has extensively been exploring, it remains a growing concern and the existing approaches are still inadequate especially with the large volume of data. Various kinds of OSN services can represent different forms or patterns of data. In addition, reduction in computation time becomes very crucial. The detection of cybercrime is therefore still challenging.

II. RELATED WORK

Chen et al. developed an approach for bullying detection that was equipped with a lexical syntactic feature, Although lexical features perform well in detecting offensive entities without considering the syntactical structure of the whole sentence, they fail to

distinguish sentence offensiveness which contain same words but in different orders [8]. Using data sets from MySpace, Dadvar et al. developed a gender based bullying detection approach that used the gender feature in enhancing the discriminative capacity of a classifier, not all the users provide complete information it leads to the imbalancing of the datasets it affects the efficiency of the model [5].Nalina and Sheela proposed an approach for detecting cyber bullying messages in Twitter by applying a feature selection weighting scheme and latent dirichlet allocation [13]. Chavan and Shylaja included pronouns, skip-gram, TF-IDF, and N-grams as additional features in improving the overall classification accuracy of their model, the TF-IDF is not semantic [9].

There are a large number of related studies on rumor detection. Most works focused on detecting rumors by shallow features of messages, including content, blog features. This method obtains significant improvement, compared with the state-of-the-art approaches. Automatic rumor identification in micro blog systems is a relatively new field. There have so far been only a few works to address this problem, and most of these works primarily focus on using micro blogs' inherent features In [15], Castillo et al. extracted 68 features from posts of twitter and categorized them into four types: 1) content-based features, which consider characteristics of the tweet content, such as the length of a message and number of positive/negative sentiment words in a message;2) user-based features, which consider traits of Twitter users, such as registration age, number of followers, and number of followees; 3) topic-based features, which are aggregates computed from message-based features and user-based features, such as the fraction of tweets that contain URLs, the fraction of tweets with hash tags, and the fraction of sentiment positive and negative in a set; and 4) propagation-based features, which consider features related to the propagation tree of a post, such as the depth of the retweet tree, or the number of initial tweets of a topic. After the research of Castillo et al., research efforts have been focused on exploiting new features for rumor detection. Qazvinian et al. [16] extracted attributes related to contents of tweets, features about the network, and specific memes of Twitter to build different Bayes classifiers to detect the rumors spreading on the twitter. Yang et al. [14] proposed two new features: 1) client-based feature and 2) location-based feature and trained a support vector machine classifier to identify the misinformation and disinformation of Sina Weibo. In [12], Sun et al. first proposed multimedia-based features for event rumors identification. Cai et al. [11] proposed text features from retweets and comments to construct rumor classifier. Wang et al. [17] proposed graph-based features and applied them in spam bots detection. Zhang et al. [6] mined the deep information of microblog contents and extracted implicit features, such as popularity, sentiment or viewpoint of message contents, and user historical information, to detect rumors in micro blogs. In [10], Wu et al. studied message propagation patterns of Sina Weibo and used them as high-order features to construct a graph-kernel-based SVM classifier for rumor identification.

III. PROPOSED SYSTEM

The proposed work highlights the areas of importance that are necessary to create a single application for automating the detection and prevention of cyber-bullying ,cyber aggression, and rumors. The previous works have concentrated only on the textual features of the social networking sites & separate application for each. The proposed work combines the efficiency of both the textual features as well as the social networking features for the detection and prevention of cyber-bullying attack as a single application.

The proposed approach to detect bully, aggressive and rumor behavior on micro blogs as summarized in Figure 1.

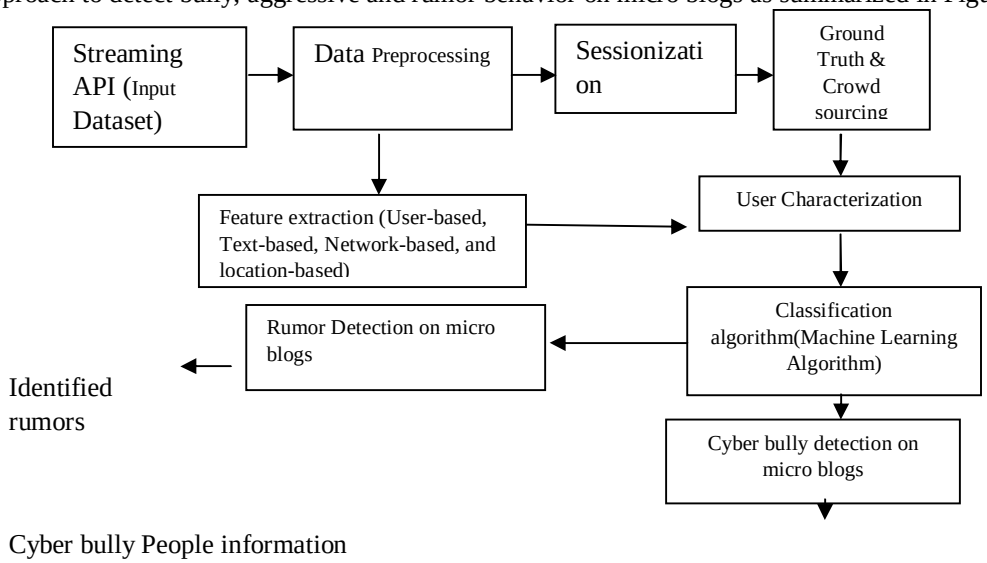


Figure 1.Overview of proposed methodology

IV. METHODOLOGY

1) *Data Collection:* Our first step is to collect data from micro blog such as twitter and, naturally, there are a few possible ways to do so. Here, we rely on Twitter's Streaming API, which provides free access to 1% of all tweets. The API returns each tweet in a JSON format, with the content of the tweet, some metadata (e.g., creation time, whether it is a reply or a retweet, etc.), as well as information about the poster (e.g., username, followers, friends, number of total posted tweets).
2) *Preprocessing:* Next, we remove stop words, URLs, and punctuation marks from the tweet text and perform normalization – i.e., we eliminate repeated letters and repetitive characters; e.g., the word "yessss" is converted to "yes". This step also involves the removal of spam content, which can be done using a few different techniques relying on tweeting behavior (e.g., many hash tags per tweet) and/or network features (e.g., spam accounts forming micro-clusters)
3) *Sessionization:* Since analyzing single tweets does not provide enough context to discern if a user is behaving in an aggressive or bullying way, we group tweets from the same user, based on time clusters, into sessions. This allows us to analyze contents of sessions rather than single tweets.
4) *Ground Truth:* We build ground truth (needed for machine learning classification, explained next) using human annotators. For this we use a crowd sourced approach, by recruiting Workers who are provided with a set of tweets from a user, and are asked to classify them according to predefined labels. If such an annotated dataset is already available, this step can be omitted.
5) *Feature Extraction:* We extract features from both tweets and user profiles. More specifically, features can be user-, text-, or network-based; e.g., the number of followers, tweets, hash tags, etc. The selection of appropriate features is obviously a very important step to speed up and improve learning quality.
6) *Classification:* The final step is to perform classification using the (extracted) features and the ground truth. Naturally, different machine learning techniques can be used for this task, including probabilistic classifiers (e.g., Naïve Bayes), decision trees (e.g., J48), ensembles (e.g., Random Forests), or neural networks.

V. CONCLUSION

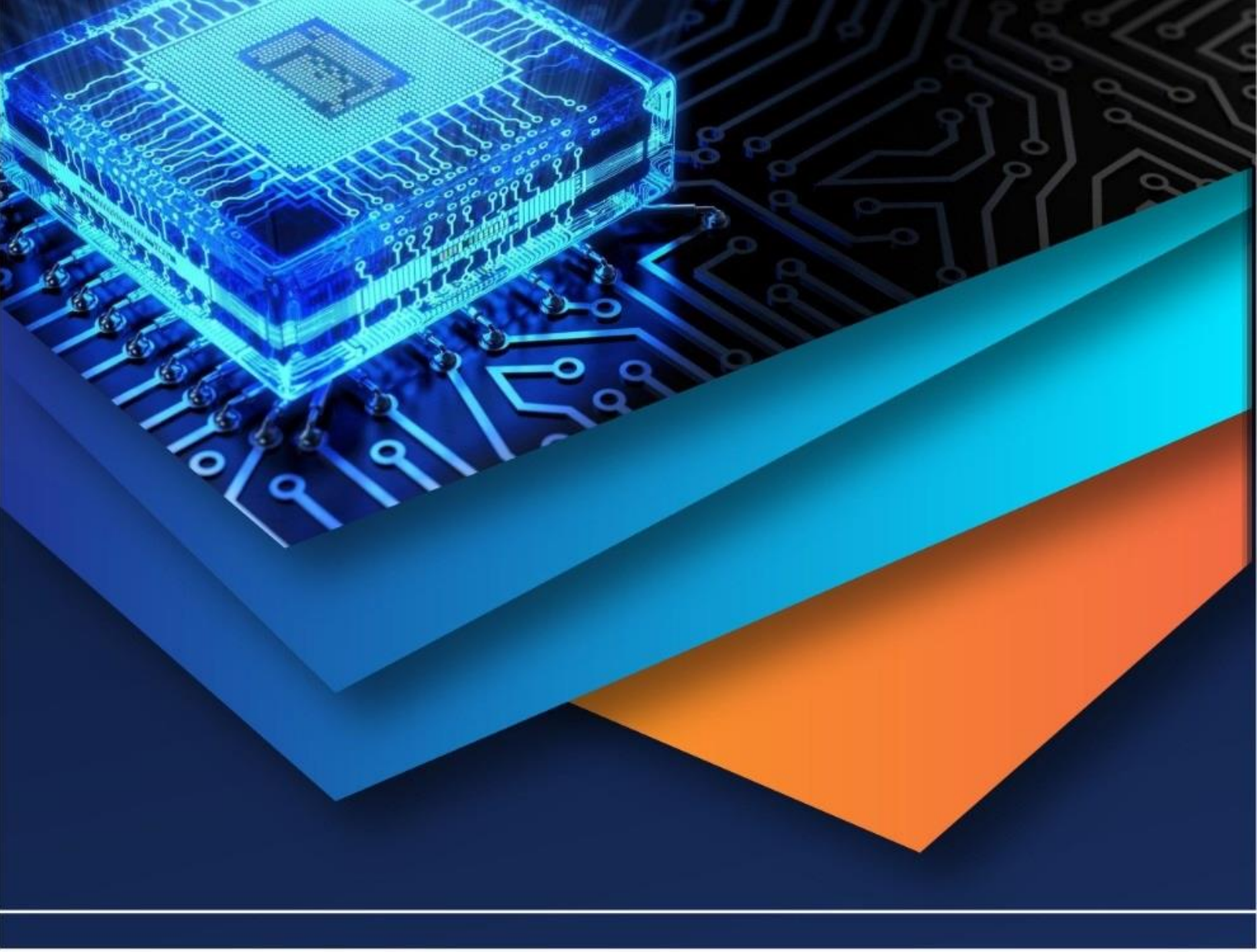
Online Social Network (OSN) services or microblog systems have become a new platform for information sharing, but they can also easily be utilized for bullying, aggressive and rumors spreading.

It is of great importance to develop an automatic tool to identify the cybercrime activities on the social networks or microblog systems. The proposed work deals with developing a single application to detect & prevent the cybercrime activities by using the evolutionary feature selection on the machine learning algorithm.

REFERENCES

[1] NCPC.org.Cyberbullying.Available :http://www.ncpc.org/cyberbullying
[2] Cyber bullying Research Center, "What is cyber bullying?" 2016. [Online]
[3] Rui Zhao, Anna Zhou,Kezhi Mao,"Automatic Detection of Cyber bullying on Social Networks based on Bulling features",ICDCN'16 Article No.43,January 2016,ACM.
[4] Cyber bullying Research Center, "summary of our Cyber bullying Research (2004-2016)", 2016 [online].
[5] Mohammed Ali Al-garadi, Kasturi DewiVarathan, Sri Devi Ravana. Cybercrime detection in online communications: The experimental case of cyber bullying detection in the Twitter network, Computers in Human Behavior 63 (2016) 433-443, Elsevier
[6] Sardar Hamidian and Mona Diab,"Rumor Detection and Classification for Twitter Data,IARIA(2015),71-77,SOTICS 2015:The fifth international conference on social Media Technologies, Communication, and Informatics,ISBN:978-1-61208-443-5
[7] Nalini, K., & Sheela, L. J. (2015). Classification of Tweets using text classifier to detect cyber bullying. In Emerging ICT for bridging the future-Proceedings of the 49th Annual convention of the Computer Society of India CSI (Vol. 2, pp. 637-645). Springer
[8] Chavan, V. S., & Shylaja, S. (2015). Machine learning approach for detection of cyber aggressive comments by peers on social media network. In Advances in computing, communications and informatics (ICACCI), 2015 International Conference on (pp.2354-2358). IEEE.
[9] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on Sina Weibo by propagation structures," in Proc. IEEE Int. Conf. Data Eng. (ICDE),2015, pp. 651–662
[10] G. Cai, H. Wu, and R. Lv, "Rumors detection in Chinese via crowd responses," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Min.(ASONAM'14), 2014, pp. 912–917
[11] S. Sun, H. Liu, J. He, and X. Du, "Detecting event rumors on Sina Weibo automatically," in Web Technologies and Applications, New York, NY,USA: Springer, 2013, pp. 120–131
[12] Chen, Ying, Yilu Zhou, Sencun Zhu, and Heng Xu. "Detecting offensive language in social media to protect adolescent online safety." In Privacy, Security, Risk and Trust (PASSAT), 2012International Conference on and 2012 International Conference on Social Computing (SocialCom), pp. 71-80. IEEE, 2012
[13] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on SinaWeibo," in Proc. ACM SIGKDDWorkshop Min. Data Semant., 2012,p. 13
[14] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in Proc. 20th Int. Conf. World Wide Web, 2011, pp. 675–684
[15] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in Proc. Conf. Empirical Methods Nat. Lang. Process., 2011, pp. 1589–1599

[16] Wang, S. Xie, B. Liu, and P. S. Yu, "Review graph based online store review spammer detection," in Proc. IEEE 11th Int.
[17] D.W.Grigg.Cyber-aggression: Definition and concept of Cyber bullying .Australian Journal of Guidance and counseling, 20(2), 2010.
[18] D.Boyd and N.B.Ellison," Social Networks Sites: Definition, History and Scholarship", Computer-Mediated commun.vol.no:13, 2007.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY