PREDICTING CYBER BULLYING ON SOCIAL MEDIA

Dr. P M Prasuna [1] N. Sravya [2], K. Sadhvija [3], MD. Roshni [4], K. Dinesh [5]

[1] Professor, Department of CSE, Ramachandra College of Engineering, Eluru, A.P

[2,3,4,5] UG Students, Department of CSE, Ramachandra College of Engineering, Eluru, A.P

ABSTRACT

Prior to the innovation of information communication technologies (ICT), social interactions evolved within small cultural boundaries such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications. These social technologies have created a revolution in user-generated information, online human networks, and rich human behavior-related data. However, the misuse of social technologies such as social media (SM) platforms, has introduced a new form of aggression and violence that occurs exclusively online. A new means of demonstrating aggressive behavior in SM websites are highlighted in this project.

The motivations for the construction of prediction models to fight aggressive behavior in SM are also outlined. We comprehensively review cyber bullying prediction models and identify the main issues related to the construction of cyber bullying prediction models in SM. This project provides insights on the overall process for cyber bullying detection and most importantly overviews the methodology. Though data collection and feature engineering process has been elaborated, yet most of the emphasis is on feature selection algorithms for prediction of cyber bullying behaviors.

1. INTRODUCTION

Cyber bullying is bullying with the use of digital technologies. It can take place on social media, messaging platforms, gaming platforms and mobile phones. It is repeated behavior, aimed at scaring, angering or shaming those who are targeted. Examples include : Spreading lies about or posting embarrassing photos or videos of someone on social media, sending hurtful, abusive or threatening messages, images or videos via messaging platforms, impersonating someone and sending mean messages to others on their behalf or through fake accounts. Face-to-face bullying and cyber bullying can often happen alongside each other. But cyber bullying leaves a digital footprint – a record that can prove useful and provide evidence to help stop the abuse.

An insightful analysis of data on human behavior and interaction to detect and restrain aggressive behavior involves multifaceted angles and aspects and the merging of theorems and techniques from multidisciplinary and interdisciplinary. The accessibility of large-scale data produces new research questions, novel computational methods, interdisciplinary approaches, and outstanding opportunities to discover several vital inquiries quantitatively. However, using traditional methods (statistical methods) in this context is challenging in terms of scale and accuracy.

These methods are commonly based on organized data on human behavior and small-scale human networks (traditional social networks). Applying these methods to large online social networks (OSNs) in terms of scale and extent causes several issues. On the one hand, the explosive growth of OSNs

enhances and disseminates aggressive forms of behavior by providing platforms and networks to commit and propagate such behavior.

On the other hand, OSNs offer important data for exploring human behavior and interaction at a large scale, and these data can be used by researchers to develop effective methods of detecting and restraining misbehavior and/or aggressive behavior. OSNs provide criminals with tools to perform aggressive actions and networks to commit misconduct. Therefore, methods that address both aspects (content and network) should be optimized to detect and restrain aggressive behavior in complex systems. The purpose of the design phase is to arrange an answer of the matter such as by the necessity document.

This part is that the opening moves in moving the matter domain to the answer domain. The design phase satisfies the requirements of the system. The design of a system is probably the foremost crucial issue warm heartedness the standard of the software package. It's a serious impacton the later part, notably testing and maintenance. Detailed Design, the inner logic of every of the modules laid out in system design is determined. Throughout this part, the small print of the info of a module square measure sometimes laid out in a high-level style description language that is freelance of the target language within which the software package can eventually be enforced. In system design the main target is on distinguishing the modules, whereas throughout careful style the main target is on planning the logic for every of the modules.

2. LITERATURE SURVEY

Predicting human behavior: The next frontiers
Machine learning has provided researchers with new tools for understanding human behavior. In this article, we briefly describe some successes in predicting behaviors and describe the challenges over the next few years.
Subrahmanian, V. and S. Kumar, Predicting human behavior: The next frontiers. Science, 2017. 355(6324): p. 489-489.

Cybercrime detection in online communications
The popularity of online social networks has created massive social communication among their users and this leads to a huge amount of user-generated communication data. In recent years, cyber bullying has grown into a major problem with the growth of online communication and social media.
Al-garadi, M.A., K.D. Varathan, and S.D. Ravana, Cybercrime detection in online communications: The experimental case of cyber bullying detection in the Twitter network. Computers in Human Behavior, 2016. 63: p. 433-443.

Cyber violence
The existing literature on the relationship between social media and violence, including prevalence rates, typologies, and the overlap between cyber and in person violence. This review explores the individual-level correlates and risk factors associated with cyber violence. Peterson, J.K. and J. Densley, Is Social Media a Gang? Toward a Selection, Facilitation, or Enhancement Explanation of Cyber Violence. Aggression and Violent Behavior, 2016.

Improving cyber bullying detection with user context
The negative consequences of cyber bullying are becoming more alarming every day and technical solutions that allow for taking appropriate action by means of automated detection are still very limited.

Up until now, studies on cyber bullying detection have focused on individual comments only, disregarding context such as users' characteristics and profile information. In this paper we show that taking user context into account improves the detection of cyber bullying.

Dadvar, M., et al., Improving cyber bullying detection with user context, in Advances in Information Retrieval. 2013, Springer. p.693-696.

Statistical detection of online drifting twitter spam

Spam has become a critical problem in online social networks. This project focuses on Twitter spam detection. Recent research works focus on applying machine learning techniques for Twitter spam detection, which make use of the statistical features of tweets.

Liu, S., J. Zhang, and Y. Xiang. Statistical Detection of Online Drifting Twitter Spam: Invited Paper. in Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security. 2016. ACM.

Cyber bullying detection: a step toward a safer internet yard

As a result of the invention of social networks friendships, relationships and social communications have all gone to a new level with new definitions. One may have hundreds of friends without even seeing their faces. Meanwhile, alongside this transition there is increasing evidence that online social applications have been used by children and adolescents for bullying.

Dadvar, M. and F. De Jong. Cyber bullying detection: a step toward a safer Internet yard. in Proceedings of the 21st international conference companion on World Wide Web. 2012.

Automatic monitoring of cyber bullying on social networking sites

The automatic monitoring of cyber bullying on social networking sites has potential for signaling harmful messages, preventing these messages from remaining online and providing timely responses. Although technological advancements are made to optimize automatic cyber bullying detection systems, little is known about its desirability and requirements.

Van Royen, K., et al., Automatic monitoring of cyber bullying on social networking sites: From technological feasibility to desirability. Telematics and Informatics, 2015. 32(1): p.

Detection of cyber bullying incidents on the Instagram social network Cyber bullying is a growing problem affecting more than half of all American teens. The main goal of this paper is to investigate fundamentally new approaches to understand and automatically detect incidents of cyber bullying over images in Instagram, a media-based

mobile social network. To this end, they have collected a sample Instagram data set consisting of images and their associated comments, and designed a labeling study for cyber bullying as well as image content using human labelers at the crowd-sourced Crowd flower Web site.

Hosseinmardi, H., et al., Detection of cyber bullying incidents on the Instagram social network. arXiv preprint arXiv:1503.03909,2015.

3.   EXISTING SYSTEM

State-of-the-art research has developed features to improve the performance of cyber bullying prediction. For example, a lexical syntactic feature has been proposed to deal with the prediction of offensive language; this method is better than traditional learning-based approaches in terms of precision. Dadvar *et al.* examined gender information from profile information and developed a gender-based approach for cyber bullying prediction by using datasets from Myspace as a basis. The gender feature was selected to improve the discrimination capability of a classifier. Age and gender

International Journal of Engineering Science and Advanced Technology (IJESAT)

	Open Access Research Article
IJESAT (Enriching the Research)	Volume: 23 Issue: 07
	July, 2023

were included as features in other studies, but these features are limited to the information provided by users in their online profiles.

- Several studies focused on cyber bullying prediction based on profane words as a feature. Similarly, a lexicon of profane words was constructed to indicate bullying. Using profane words as features demonstrates a significant improvement in model performance. For example, the number of ``bad'' words and the density of ``bad'' words were proposed as features for input. The study concluded that the percentage of ``bad'' words in a text is indicative of cyber bullying. Another research expanded a list of pre-defined profane words and allocated different weights to create bullying features. These features were concatenated with bag-of-words and latent semantic features and used as a feature input.

DISADVANTAGES

- Mostly children are facing these cyber bullying in their social media networks, but they don't know a particular way to stop this and to whom to share with. Because of this sometimes children are taking harm full decisions to themselves.
- This may even increase their mental stress.

4. PROPOSED SYSTEM

The proposed system is constructing cyber bullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labeled text instances. Another means is to use a lexicon-based model that involves computing orientation for a document from the semantic orientation of words or phrases in the document. Generally, the lexicon in lexicon-based models can be constructed manually or automatically by using seed words to expand the list of words. However, cyber bullying prediction using the lexicon-based approach is rare in literature.

The primary reason is that the texts on SM websites are written in an unstructured manner, thus making it difficult for the lexicon-based approach to detect cyber bullying based only on lexicons. However, lexicons are used to extract features, which are often utilized as inputs to machine learning algorithms. For example, lexicon based approaches, such as using a profane-based dictionary to detect the number of profane words in a post, are adopted as profane features to machine learning models. The key to effective cyber bullying prediction is to have a set of features that are extracted and engineered.

ADVANTAGES

- Because of this cyber bullying they can't share these abusive type of words to their parents. By using our project this problem will be solved with in the two persons without knowing to the third eye.
- Through this website we can easily identify and prevent those abusive content in order to build a good environment with in the users.
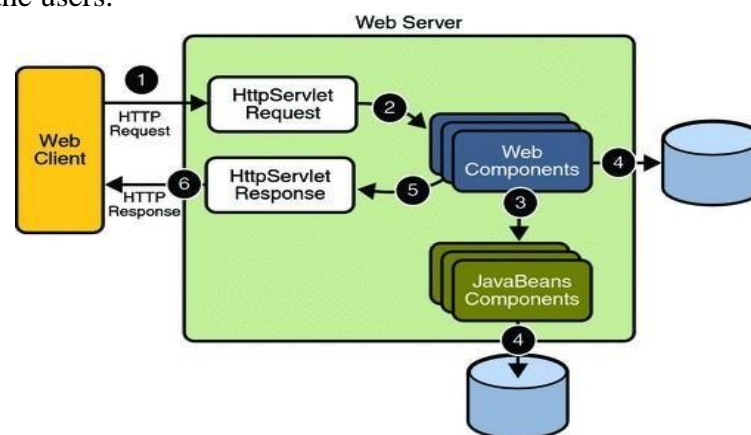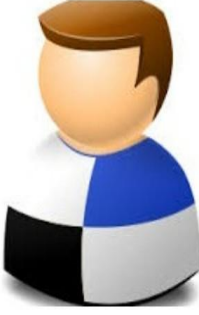

Fig1.System Architecture

5. RESULTS

Screenshot1: shows the existed user will login to account and the new user willregister to account.

Screenshot2 : The user can see their own profile details, such as their address, etc.

Screenshot3 : User can search for friends and can send friend request or can accept request.

Screenshot4: The user can see all his posted cyber bullying comments on theirfriend created posts.

Screenshot5 : The user can see his friend's post details and can comment on

International Journal of Engineering Science and Advanced Technology (IJESAT)

	Open Access Research Article
IJESAT (Enriching the Research)	Volume: 23 Issue: 07
	July, 2023

posts.

Screenshot6: The admin has to login by using valid user name and password.

Screenshot7 : The admin views all users details and authorize them for login permission.

Screenshot8 : The admin can add filters as categories with the words those related to corresponding filters.
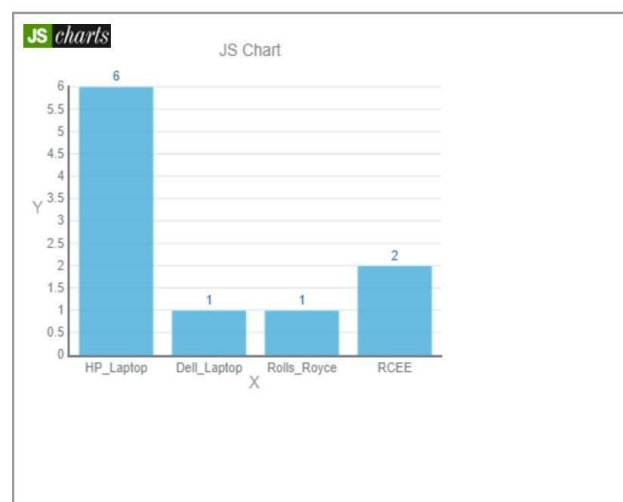
Screenshot9: The admin can see all the posts with number of cyber bullyingcomments posted by

theusers.

6. CONCLUSION

This study reviewed existing literature to detect aggressive behavior on SM websites. We specifically reviewed four aspects of detecting cyber bullying messages namely, data collection, feature engineering, construction of cyber bullying detection model, and evaluation of constructed cyber bullying detection models. Several types of discriminative features that were used to detect cyber bullying in online social networking sites were also summarized. Most importantly we summarized and identified the important factors for detecting cyber bullying by using our website. For this purpose, we have used tokenization and bag of words which gives us the area under the curve function for modeling the behaviors in cyber bullying. Finally, the main issues and open research challenges were described and discussed.

REFERENCES

[1] V. Subrahmanian and S. Kumar, ``Predicting human behavior: The next frontiers,''
Science, vol. 355, no. 6324, p. 489, 2017.

[2] H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, ``Homophily in the digital world: A LiveJournal case study,'' IEEE Internet Comput., vol. 14, no. 2, pp. 15_23, Mar./Apr. 2010.

[3] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, ``Cybercrime detection in online communications: The experimental case of cyber bullying detection in the Twitter network,'' Comput. Hum. Behav., vol. 63, pp. 433_443, Oct. 2016.

[4] L. Phillips, C. Dowling, K. Shaffer, N. Hodas, and S. Volkova, ``Using social media to predict the future: A systematic literature review,'' 2017,arXiv:1706.06134. [Online]. Available: https://arxiv.org/abs/1706.06134

[5] H. Quan, J. Wu, and Y. Shi, ``Online social networks & social network services: A technical survey,'' in Pervasive Communication Handbook.Boca Raton, FL, USA: CRC Press, 2011, p.

[6] J. K. Peterson and J. Densley, ``Is social media a gang? Toward a selection,facilitation, or enhancement explanation of cyber violence,'' Aggression Violent Behav., 2016.

[7] BBC. (2012).Huge Rise in Social Media. [Online]. Available:http://www.bbc.com/news/uk-20851797

[8] P. A.Watters and N. Phair, ``Detecting illicit drugs on social media using automated social media intelligence analysis (ASMIA),'' in Cyberspace Safety and Security. Berlin, Germany: Springer, 2012, pp. 66_76.

[9] M. Fire, R. Goldschmidt, and Y. Elovici, ``Online social networks:Threats and solutions,'' IEEE Commun. Surveys Tuts., vol. 16, no. 4,pp. 2019_2036, 4th Quart., 2014. N. M. Shekokar and K. B. Kansara, ``Security against sybil attack in social network,'' in Proc. Int. Conf. Inf. Commun. Embedded Syst. (ICICES),2016, pp. 1_5.

[10] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, ``Detecting and tracking political abuse in social media,'' in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 297_304.

[11] A. Aggarwal, A. Rajadesingan, and P. Kumaraguru, ``PhishAri: Automatic realtime phishing detection on Twitter,'' in Proc. eCrime Res.Summit (eCrime), Oct. 2012, pp. 1_12.

[12] S. Yardi et al., ``Detecting spam in a Twitter network,'' First Monday, Jan. 2009. [Online]. Available: https:// rstmonday.org/article/view/2793/2431

[13] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, ``Analyzing spammers' social networks for fun and pro_t: A case study of cyber criminal ecosystem on twitter,'' in Proc. 21st Int. Conf. World Wide Web,2012, pp. 71_80.

[14] G. R. S.Weir, F. Toolan, and D. Smeed, ``The threats of social networking:Old wine in new bottles?'' Inf. Secur. Tech. Rep., vol. 16, no. 2, pp. 38_43,2011.

[15] M. J. Magro, ``A review of social media use in e-government,'' Administ.Sci., vol. 2, no. 2, pp. 148_161, 2012.

[16] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, ``Improving cyber bullying detection with user context,'' in Advances in Information Retrieval. Berlin, Germany: Springer, 2013, pp. 693_696.

[17] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, ``Detecting offensive language in social media to protect adolescent online safety,'' in Proc. Int. Conf. Privacy, Secur., Risk Trust (PASSAT), Sep. 2012, pp. 71_80.

WWW.IJESAT.COM