International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

DETECTING CYBER BULLYING ON SOCIAL MEDIA IN THE BIG DATA ERA USING MACHINE LEARNING TECHNIQUES

N.SIVAGANGA [1], B.RANJITH[2]

[1] PG SCHOLAR,DEPT OF CSE,ST.MARY'S GROUP OF INSTITUTION, GUNTUR, AP, INDIA.

[2]ASST. PROFESSOR [M.TECH], DEPARTMENT OF CSE , ST.MARY'S GROUP OF INSTITUTION, GUNTUR, AP, INDIA.

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 paper. The motivations for the construction of prediction models to fight aggressive behavior in SM are also outlined. We comprehensively review cyberbullying prediction models and identify the main issues related to the construction of cyberbullying prediction models in SM. This paper provides insights on the overall process for cyberbullying 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 and then using various machine learning algorithms for prediction of cyberbullying behaviors. Finally, the issues and challenges have been highlighted as well, which present new research directions for researchers to explore.

1.INTRODUCTION

Machine or deep learning algorithms help researchers understand big data [1]. Abundant information on humans and their societies can be obtained in this big data era, but this acquisition was previously impossible [2]. One of the main sources of human-related data is social media (SM). By applying machine learning algorithms to SM data, we can exploit historical data to predict the future of a wide range of applications. Machine learning algorithms provide an opportunity to effectively predict and detect negative forms of human behavior, such as cyberbullying [3]. Big data analysis can uncover hidden knowledge through deep learning from raw data [1]. Big data analytics has improved several applications, and forecasting the future has even become possible through the combination of big data and machine learning algorithms [4].

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 _elds. 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.

2.LITERATUREREVIEW

Predicting human behavior: The next Frontiers by V. Subrahmanian and S. Kumar
Machine learning has provided researchers with new tools for understanding human behavior. In this article, we briefly describe some successes in predicting behaviors
Homophily in the digital world: A LiveJournal case study by H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas
Are two users more likely to be friends if they share common interests? Are two users more likely to share common interests if they're friends? The authors study the phenomenon of homophily in the digital world by answering these central questions. Unlike the physical world, the digital world doesn't impose any geographic or organizational constraints on friendships. So, although online friends might share common interests, a priori there's no reason to believe that two users with common interests are more likely to be friends. Using data from LiveJournal, the authors show that the answer to both questions is yes.
Cybercrime detection
in online communications: The experimental case of cyberbullying detection in the Twitter network by M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana
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, Cyberbullying has grown into a major problem with the growth of online communication and social media. Cyberbullying has been recognized recently as a serious national health issue among online social network users and developing an efficient detection model holds tremendous practical significance. In this paper, we have proposed set of unique features derived from Twitter; network, activity, user, and tweet content, based on these feature, we developed a supervised machine learning solution for detecting cyberbullying in the Twitter. An evaluation demonstrates that our developed detection model based on our proposed features, achieved results with an area under the receiver-operating characteristic curve of 0.943 and an f-measure of 0.936. These results indicate that the proposed model based on these features provides a feasible solution to detecting Cyberbullying in online communication environments. Finally, we compare result obtained using our proposed features with the result obtained from two baseline features. The comparison outcomes show the significance of the proposed features.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

3.EXISTING SYSTEM

State-of-the-art research has developed features to improve the performance of cyberbullying 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 cyberbullying 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 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 cyberbullying prediction based on profane words as a feature. Similarly, a lexicon of profane words was constructed to indicate bullying, and these words were used as features for input to machine learning algorithms. 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 to machine learning in a previous work.The study concluded that the percentage of ``bad" words in a text is indicative of cyberbullying. 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 for a machine learning algorithm.

The System is not much affective due to Semi supervised machine learning techniques.

The system doesn't have sentiment classification for cyberbullying.

4.PROPOSED SYSTEM

The proposed system is constructing cyberbullying 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, cyberbullying 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 cyberbullying 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 cyberbullying prediction is to have a set of features that are extracted and engineered.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

The system is more effective due to LOGISTIC REGRESSION CLASSIFICATION and UNSUPERVISED MACHINE LEARNING.

An effective cyberbullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labeled text instance and also 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.

5. SYSTEM ARCHITECTURE:

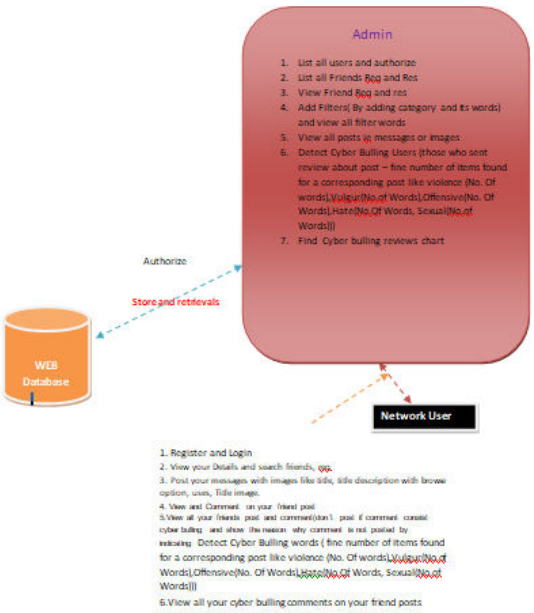

Fig 4.1 architecture Diagram

6.IMPLEMENTATION

Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as view and authorize users, view all friends request and responses, Add and View Filters, View all posts, Detect Cyber Bullying Users, Find Cyber Bullying Reviews Chart.

Viewing and Authorizing Users

In this module, the admin views all users details and authorize them for login permission. User Details such as User Name, Address, Email Id, Mobile Number.

Viewing all Friends Request and Response

In this module, the admin can see all the friends' requests and response history. Details such as Requested User Name and Image, and Requested to User Name and Image, status and date.

Add and View Filters

In this module, the admin can add filters (like Violence, Vulgar, Offensive, Hate, and Sexual) as Categories with the words those related to corresponding filters.

View all posts

In this module, the admin can see all the posts added by the users with post details like post name, description and post image.

Detect Cyber Bullying Users

In this module, the admin can see all the Cyber Bullying Users (The users who had posted a comment on posts using cyber bullying words which are all listed by the admin to detect and filter). In this, the results shown as, Number of items found for a corresponding post like Violence (no. of words belongs to Violence Filter used in comments by the users), Vulgar (no. of words belongs to Vulgar Filter used in comments by the users), Offensive (no. of words belongs to Offensive Filter used in comments by the users), Hate (no. of words belongs to Hate Filter used in comments by the users), Sexual (no. of words belongs to Sexual Filter used in comments by the users).

Find Cyber Bullying Reviews Chart

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

In this module, the admin can see all the posts with number of cyber bullying comments posted by users for particular post.

User

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like viewing their profile details, searching for friends and sending friend requests, Posting Your Messages as Posts by giving details, View and Comment on Friend Posts, viewing all friends posts and comment, view all your cyber bullying comments on your friend posts.

Viewing Profile Details, Search and Request Friends

In this module, the user can see their own profile details, such as their address, email, mobile number, profile Image.

The user can search for friends and can send friend requests or can accept friend requests.

Add Posts

In this, the user can add their own posts by giving post details such as, post title, description, uses, and image of post.

View and Comment on Your Friends Post

In this, the user can see his entire friend's post details (post title, description, uses, creator and image of post) and can comment on posts.

View all Friends Posts and Comment (Cyber bullying Related)

In this, the user can see his all friend's post details (post title, description, uses, creator and image of post) and can comment on posts.

Don't Post If the comment consists of Cyber bullying words and Shows the reason why comment is not posted by indicating Detected Cyber Bullying Words like Numbers of Cyber Bullying words Related to Filter Violence found in comment, Numbers of Cyber Bullying words Related to Filter Vulgar found in comment, Numbers of Cyber Bullying words Related to Offensive found in comment, Numbers of Cyber Bullying words Related to Hate found in comment, Numbers of Cyber Bullying words Related to Sexual found in comment, View all Your Cyber bullying comments on your friend posts

The user can see all his posted cyber bullying comments on their friend created posts.

.

7. SCREEN SHOTS

8.CONCLUSION

This study reviewed existing literature to detect aggressive behavior on SM websites by using machine learning approaches. We specifically reviewed four aspects of detecting cyberbullying messages by using

machine learning approaches, namely, data collection, feature engineering, construction of cyberbullying detection model, and evaluation of constructed cyberbullying detection models. Several types of discriminative features that were used to detect cyberbullying in online social networking sites were also summarized. In addition, the most effective supervised machine learning classifiers for classifying cyberbullying messages in online social networking sites were identified. One of the main contributions of current paper is the definition of evaluation metrics to successfully identify the significant parameter so the various machine learning algorithms can be evaluated against each other. Most importantly we summarized and identified the important factors for detecting cyberbullying through machine learning techniques specially supervised learning. For this purpose, we have used accuracy, precision recall and f-measure which gives us the area under the curve function for modeling the behaviors in cyberbullying. Finally, the main issues and open research challenges were described and discussed.

BIBILOGRAPHY

[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 cyberbullying 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. 4.

[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.

[10] 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.

[11] 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.

[12] 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.
[13] S. Yardi et al., ``Detecting spam in a Twitter network," First Monday, Jan. 2009. [Online]. Available: https://_rstmonday.org/ article/view/2793/2431
[14] 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.
[15] 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.
[16] M. J. Magro, ``A review of social media use in e-government," Administ. Sci., vol. 2, no. 2, pp. 148_161, 2012.
[17] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, ``Improving cyberbullying detection with user context," in Advances in Information Retrieval. Berlin, Germany: Springer, 2013, pp. 693_696.
[18] 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.
[19] V. S. Chavan and S. S. Shylaja, ``Machine learning approach for detection of cyber-aggressive comments by peers on social media network," in Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI), Aug. 2015, pp. 2354_2358.
[20] W. Dong, S. S. Liao, Y. Xu, and X. Feng, ``Leading effect of social media for _nancial fraud disclosure: A text mining based analytics," in Proc. AMCIS, San Diego, CA, USA, 2016.
[21] M. S. Rahman, T.-K. Huang, H. V. Madhyastha, and M. Faloutsos, ``FRAppE: Detecting malicious Facebook applications," in Proc. 8th Int. Conf. Emerg. Netw. Exp. Technol., 2012, pp. 313_324.
[22] S. Abu-Nimeh, T. Chen, and O. Alzubi, ``Malicious and spam posts in online social networks," Computer, vol. 44, no. 9, pp. 23_28, Sep. 2011.
[23] B. Doerr, M. Fouz, and T. Friedrich, ``Why rumors spread so quickly in social networks," Commun. ACM, vol. 55, no. 6, pp. 70_75, Jun. 2012.
[24] J. W. Patchin and S. Hinduja, Words Wound: Delete Cyberbul- lying and Make Kindness Go Viral. Golden Valley, MN, USA: Free Spirit Publishing, 2013.
[25] J. Cheng, C. Danescu-Niculescu-Mizil, and J. Leskovec, ``Antisocial behavior in online discussion communities," in Proc. 9th Int. AAAI Conf. Web Social Media, Apr. 2015.