Detecting and Blocking Bullying Posts in Online Social Networks

Gopikrishnan M, Sd.Nafeesa Rosedar, Y.Sai Lokeswari, Y.Snehalatha

Department of Computer Science and Engineering

Prathyusha Engineering College

ABSTRACT

In recent times there are more rise in the users of social media since one of the most harmful consequences of social media causes the rise of cyber bullying, which tends to be more sinister than traditional bullying, given the online records typically live on the Internet for quite a long time moreover are hard to control. To tackle this issue, in this research paper, we are going to study the problem of how to minimize the bullying content while preserving the user experience. On analyzing this, we will present an effective method for detecting bullying posts in online social networks. We analyze tweets posted by the user to determine their relation to cyber bullying. The tweets that will be posted by the user, which contains bullying content will be detected and then these posts will be blocked while posting i.e., the users cannot be able to post hounding posts.

1. INTRODUCTION

Social networking sites are great tools for connecting with people. However, social networking has become widespread; people are finding illegal and corrupt ways to use these communities. The problems involved in social networking like privacy, online bullying, misuse, and trolling and many others. With the rise of so called trends of sharing or posting data or pictures on certain social networking sites and commenting on them have adding the risk of cyber defamation. Detecting and blocking bullying posts in online social networks is a way to create secured web chat applications in social networks.

As the application is able to analyze the posts posted by the user whether they are bullying in nature or not and if they contain bullying content such posts will be blocked.

2.LITERATURE REVIEW

Aparna Sankaran Srinath , Hannah Johnson, Gaby G. Dagher , and Min Long, "Bullynet : Unmasking Cyberbullies on Social Networks" in proceedings of IEEE, 2021.

Although the digital revolution and the rise of social media enabled great advances in communication platforms and social interactions, a wider proliferation of harmful behavior known as bullying has also emerged. This article presents a novel framework of BullyNet to identify bully users from the Twitter social network. We performed extensive research on mining SNs for better understanding of the relationships between users in social media, to build an SN based on bullying tendencies. We observed that by constructing conversations based on the context as well as content, we could effectively identify the emotions and the behavior behind bullying. In our experimental study, the evaluation of our proposed centrality measures to detect bullies from SN, we achieved around 80% accuracy with 81% precision in identifying bullies for various cases. There are still several open questions deserving further investigation. First, our approach focuses on extracting emotions and behavior from texts and emojis in tweets. However, it would be interesting to investigate images and videos, given that many users use them to bully others.

Second, it does not distinguish between bully and aggressive users. Devising new algorithms or techniques to distinguish bullies from aggressors would prove critical in better identification of cyberbullies. Another topic of interest would be to study the relationship between conversation graph dynamics and geographic location and how these dynamics are affected by the geographic dispersion of the users? Does proximity increase the bullying behavior?

L. Cheng, R. Guo, Y. Silva, D. Hall, and H. Liu, "Hierarchical attention networks for cyberbullying detection on the instagram social network," in Proceedings of the 2019 SIAM International Conference on Data Mining. SIAM, 2019.

In this paper, we proposed the Hierarchical Attention Networks for Cyberbullying Detection (HANCD) framework, which progressively builds a social media session by first aggregating words into comment vectors and then into session vectors. The proposed framework uses context to discover the relative importance of specific words and comments, rather than simply filter for words, devoid of context. To model the critical temporal information in a social media session, we jointly optimize the cyberbullying detection and time interval prediction tasks. By manipulating the weights of these two tasks, HANCD can capture their commonalities and differences to improve the performance of cyberbullying detection. Because comments are posted at discrete points in time, future work can be directed to time series analysis, which models a sequence of discrete temporal data in order to extract meaningful statistics and identify important trends. Another vital direction for future research may be time series forecasting, which could be used to predict future cyberbullying instances from previously observed cases. Efforts to more accurately detect cyberbullying on social media remain a critical step toward building safer, more inclusive social interaction spaces.

John Hani , Mohamed Nashaat , Mostafa Ahmed, Zeyad Emad, Eslam Amer, " Social Media Cyberbullying Detection using Machine Learning", in proceedings of IEEE, 2017.

In this paper, we proposed an approach to detect cyberbullying using machine learning techniques. We evaluated our model on two classifiers SVM and Neural Network and we used TFIDF and sentiment analysis algorithms for features extraction. The classifications were evaluated on different n-gram language models. We achieved 92.8% accuracy using Neural Network with 3-grams and 90.3% accuracy using SVM with 4 grams while using both TFIDF and sentiment analysis together. We found that our Neural Network performed better than the SVM classifier as it also achieves average f-score 91.9% while the SVM achieves average f-score 89.8%. Furthermore, we compared our work with another related work that used the same dataset, finding that our Neural Network outperformed their classifiers in terms of accuracy and f-score. By achieving this accuracy, our work is definitely going to improve cyber bullying detection to help people to use social media safely. However, detecting cyber bullying pattern is limited by the size of training data. Thus, a larger cyber bullying data is needed to improve the performance. Hence, deep learning techniques will be suitable in the larger data as they are proven to outperform machine learning approaches over larger size data.

V. K. Singh, Q. Huang, and P. K. Atrey, "Cyberbullying detection using probabilistic socio-textual information fusion," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2016, pp. 884–887.

This paper advances the state of the art in cyberbullying detection beyond individual features to propose a novel method for fusing heterogeneous social and textual features for improved cyberbullying detection. The proposed method leverages the differences in the contributions of heterogeneous data features toward the classification goal and the associations between different features to generate a better classification performance. The obtained results were compared to a recent approach, which used similar dataset and features, and the proposed method resulted in significant improvements in the classification results. With the growth trends in multimodal data, better social network characterization, and textual feature analysis, the proposed fusion approach could provide the backbone for integrating such features for enhanced cyberbullying detection in different settings, thus paving the way for safer online social networks.

L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, "Xbully: Cyberbullying detection within a multi-modal context," in Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 2019.

With the growing popularity of social media platforms and rapid increases in social media use among teens, cyberbullying has become more prevalent and begun to raise serious societal concerns. The majority of previous efforts for detecting cyberbullying are based on text analysis. Although they mark an important step forward in combating cyberbullying, these works fail to consider the multimodal nature of social media data (e.g., texts, images, likes/shares, etc.). Our proposed model is based on the belief that multi-modal information can provide valuable insights for characterizing and detecting cyberbullying behaviours, which can complement and ultimately extend previous work. In this paper, we study the novel problem of cyberbullying detection within a multi-modal context. To address the challenges tied to multi-modal social media information, we propose an innovative cyberbullying detection framework, XBully, based on network representation learning. XBully first identifies representative mode hotspots to handle diverse feature types and then jointly maps both attributed and nominal nodes in a heterogeneous network into the same latent space by exploiting the cross-modal correlations and structural dependencies. Extensive experimental results on real-world datasets corroborate the effectiveness of the proposed framework. Future work directed towards building a deeper understanding of different modalities in characterizing cyberbullying behaviours will not only improve cyberbullying detection, but may also shed light on behaviours that are unique to users with different roles (e.g., victims, bullies) within cyberbullying interactions. Furthermore, we believe that the most efficient path forward entails interdisciplinary collaboration among researchers in computer science and psychology to address this major social issue.

3. EXISTING SYSTEM

It is hard to accurately interpret user's intentions and meanings in social media based merely on their messages (e.g., posts, tweets, comments), which are typically short, use expletive languages, or may include multimedia contents such as pictures and videos. For example, Twitter limits its users' messages to 140 characters, which could be a mix of text, words, emojis, and gifs. As a result, it is hard to determine the opinion expressed by a message promptly. In the existing methods they are only detecting the bullying content but not blocking it. In some existing methods they are blocking the users.

DISADVANTAGES OF EXISTING SYSTEM

1. Blocking too many users or social links will degrade user experiences and may arise complaints for the right violation.

2. In the existing methods they are only detecting the bullying content but not blocking it.

4. PROPOSED SYSTEM

To embark upon the issue, in this project, we are studying the problem of how to minimize the bullying content while preserving the user experience. In this view, we will present an effective method for detecting bullying posts in online social networks. We analyze tweets posted by the user to determine their relation to cyber-bullying. The tweets that will be posted by the user, which contains bullying content will be detected and then these posts will be blocked while posting i.e., the users cannot be able to post bullying posts.

ADVANTAGES OF PROPOSED SYSTEM

1. We take the first look into minimizing the bullying content while preserving the user experience.
2. In this we are not only detecting the bullying posts but also blocking the posts.
3. In this, there will be no mislead in the prediction of bullying content because only Admin can train the system.

5. MODULES

User registration and Login

User can be able to register in the application by giving all the required details, followed by the total data of the user is stored into database and admin is able to see the user credentials. After the user registration is completed, user can login the application by giving their username and password. User home page contains their profile details and user can able to find friends and can send friend requests.

User Creates a Post

In the user home page users can be able to create a tweet by giving the tweet title, content for the tweet and by adding the image of the tweet which they want to post. User can view all the tweets they have created. User can be able to view the tweets posted by their friends and can re-tweet the posts created by their friends.

Admin Adds Bullying Words

Admin have all users credentials and all the users information is stored in the database by using MYSQL. Admin can be able to view all the tweets and re-tweets posted by the users which are stored in the database. Admin can add bullying words to the system based on which the tweets are analyzed later.

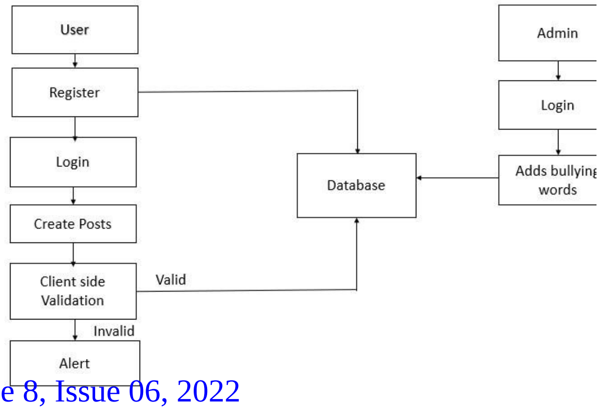
Detection of Bullying Posts

While the user creating a post the system will be able to detect the posts which contain bullying words. This detection can be done based on the content the user creates while posting the post. Here client side validation will be done.

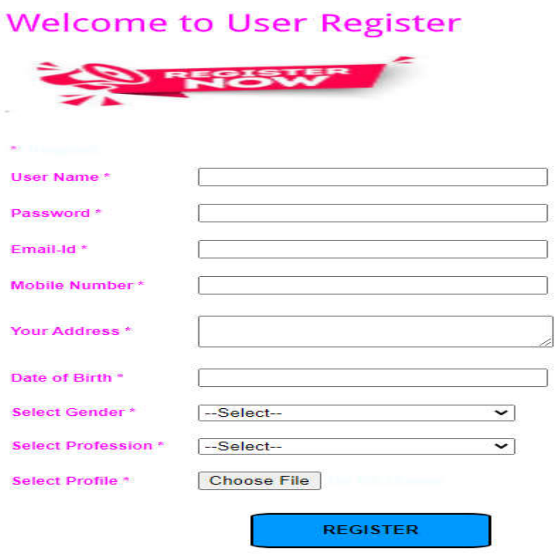
Blocking of Bullying Posts

After detecting the posts that are in bullying in nature will be blocked and the user will be getting an alert that the post is invalid.

6. SYSTEM ARCHITECTURE



7. OUTPUT


Welcome to User Register

REGISTER NOW

User Name *

Password *

Email-Id *

Mobile Number *

Your Address *

Date of Birth *

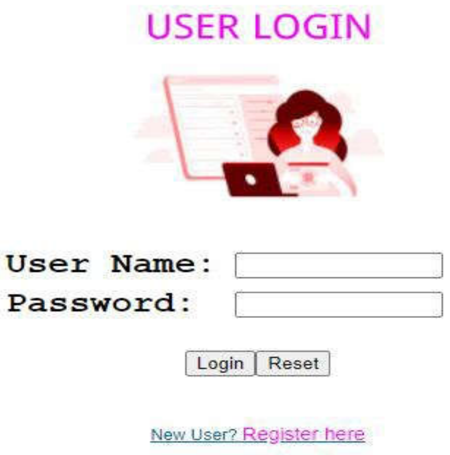
Select Gender * --Select--

Select Profession * --Select--

Select Profile * Choose File

REGISTER

Figure 1. User Registration Page


USER LOGIN

User Name:

Password:

Login Reset

New User? Register here

Figure 2. User Login Page

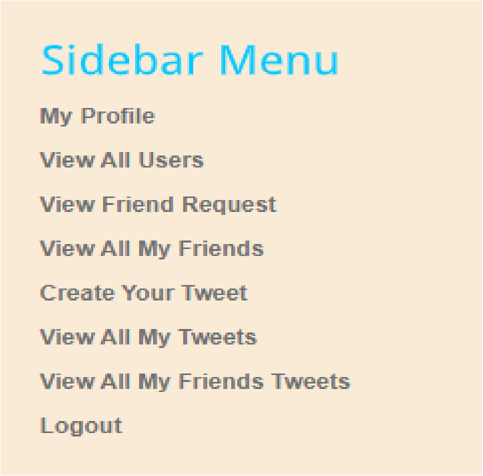


Figure 3. User Home Page

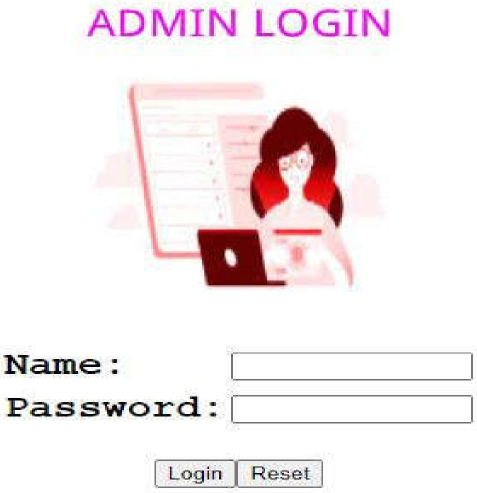
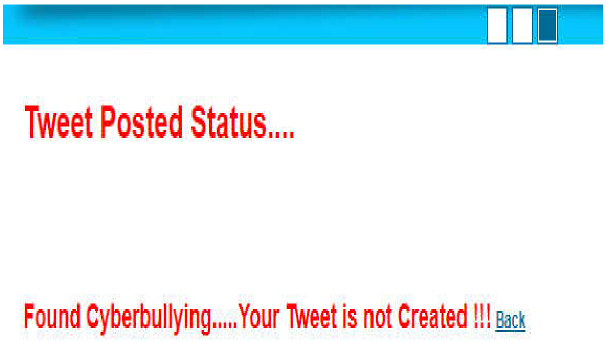


Figure 4. Admin Login Page



Tweet Posted Status....

Found Cyberbullying.....Your Tweet is not Created !!! Back

Figure 5. Bullying posts will be detected and blocked

8. CONCLUSION

Detecting and blocking bullying posts in online social networks is an effective way to create secure web-chat application in social networks. The application is able to categorize the information posted by the user whether it contains any bullying contents or not. If the post is bullying in nature, it will not be posted and the user gets an alert. The admin adds some of the cursing, swearing or expletive words to the system which are bullying in nature. These words are used by the system to predict the tweet posted by user as bullying or not and can be able to block the bullying posts. Evaluation shows that detecting and blocking bullying posts in online social networks achieves high information category detection accuracy.

9. REFERENCES

[1] Aparna Sankaran Srinath , Hannah Johnson, Gaby G. Dagher , and Min Long, “Bullynet : Unmasking Cyberbullies on Social Networks” in proceedings of IEEE, 2021.

[2] L. Cheng, R. Guo, Y. Silva, D. Hall, and H. Liu, “Hierarchical attention networks for cyberbullying detection on the instagram social network,” in Proceedings of the 2019 SIAM International Conference on Data Mining. SIAM, 2019.

[3] John Hani , Mohamed Nashaat , Mostafa Ahmed, Zeyad Emad, Eslam Amer, “ Social Media Cyberbullying Detection using Machine Learning”, in proceedings of IEEE, 2017.

[4] V. K. Singh, Q. Huang, and P. K. Atrey, “Cyberbullying detection using probabilistic socio- textual information fusion,” in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2016, pp. 884–887.

[5] L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, “Xbully: Cyberbullying detection within a multimodal context,” in Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 2019.

[6] R. Zhao, A. Zhou, and K. Mao, “Automatic detection of cyberbullying on

social networks based on bullying features," in Proc. 17th Int. Conf. Distrib. Comput. Netw., Jan. 2016, pp. 1–6.

[7] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Detection of cyberbullying incidents on the Instagram social network," CoRR, vol. 1503.03909, 2015.

[8] A. Squicciarini, S. Rajtmajer, Y. Liu, and C. Griffin, "Identification and characterization of cyberbullying dynamics in an online social network," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Aug. 2015, pp. 280–285.

[9] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Mean birds: Detecting aggression and bullying on Twitter," in Proc. ACM Web Sci. Conf., Jun. 2017, pp. 13–22.

[10] L. Cheng, J. Li, Y. Silva, D. Hall, and H. Liu, "PI-bully: Personalized cyberbullying detection with peer influence," in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 5829–5835.