International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 1343-1348

Prevention and Suppression of Cyberbullying Using Machine Learning

Mrs. K.RAJESWARI 1st, MUSHRUF BASHA M 2nd, PRAVEEN S3rd , RANJITH S R 4th,SANDEAP V 5th

1st Assistant Professor, 2nd, 3rd, 4th, 5th UG Scholar(B.E), Department of Computer Science and Engineering, Mahendra Engineering College, Mahendhirapuri.

.

Abstract

In this project, cyberbullying is a widespread social behavior that can have many psychological, behavioral, and health adverse effects on victims of cyberbullying. Studies show that cyberbullying is occurring all over the world. Understanding the variables and processes that predict cyberbullying is important for interventions aimed at reducing online antisocial behavior. Cyberbullying is a major online problem and affects young people and adults alike. It led to misfortunes such as suicide and depression. There is an increasing need to regulate content on social media platforms. In the next study, we will use two different forms of cyberbullying, hate speech tweets from Twitter, and personal attack-based comments from the Wikipedia forums to net with text data based on natural language processing and machine learning learning. Create a model based on the detection of bullying. Three feature extraction methods and four classifiers are being considered to outline the best approach.

Keywords: *Cyber bulling Model, Predicting Cyber Crime, Suppression of Cyberbullying using Machine Learning.*

Date of Submission: 06-06-2022 Date of acceptance: 21-06-2022

I. INTRODUCTION

Cyberbullying 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.Given the consequences of cyberbullying on victims, it is urgently needed to find a proper actions to detect and hence to prevent it. One of the successful approaches that learns from data and generates a model that automatically classifiesproper actions is machine learning. Machine learning can be helpful to detect language patterns of the bullies and hence can generate a model to detect cyberbullying actions. Thus, the main contribution of this paper is to propose a supervised machine learning approach for detecting and preventing cyberbullying. The proposed approach is evaluated on a cyberbullying dataset from Kaggle which was collected and labeled by the authors Kelly Reynolds et al. in their paper. The performance of SVM and Neural Network classifiers are compared on both TFIDF and sentiment analysis feature extraction methods. Furthermore, experiments were made on different n-gram language model. 2-gram, 3-gram and 4-gram has been taken into consideration during the evaluation of the model produced by the classifiers. Finally, we evaluate our proposed approach with previous related work who used the same data.

The use of social media has grown exponentially over time with the growth of the Internet and has become the most influential networking platform in the 21st century. However, the enhancement of social connectivity often creates negative impacts on society that contribute to a couple of bad phenomena such as online abuse, harassment cyberbullying, cybercrime and online trolling. Cyberbullying frequently leads to serious mental and physical distress, particularly for women and children, and even sometimes force them to attempt suicide classifier which will accurately detect bully tweets. Pre- processing of data has two steps: Collection ofdata and Cleaning of data. The very first and basic step is collection of data that is done in two ways. The twitter API was accessed and tweets were extracted, rest of the tweets were obtained from Kaggle dataset. The dataset was divided into trainingand testing data. The tweets of the training data were labelled by the values 0 and 1.The bully tweets were represented by value 1 and the non-bully tweets were represented by value 0. The test data was not labelled.

II. LITERATURE SURVEY

1. Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study **AUTHOR:** Maral Dadvar and Kai Eckert 2019 Cyberbullying is a disturbing online misbehaviour with troubling consequences. It appears in different forms, and in most of the social networks, it isin textual format. Automatic detection of such incidents requires intelligent systems.Most of the existing studies have approached this problem with conventionalmachine learning models and the majority of the developed models in these studiesare adaptable to a single social network at a time. In recent studies, deep learning based models have found their way in the detection of cyberbullying incidents, claiming that they can overcome the limitations of the conventional models, and improve the detection performance. In this paper, we investigate the findings of a recent literature in this regard. We successfully reproduced the findings of this literature and validated their findings using the same datasets, namely Wikipedia, Twitter.

2. Cyberbullying Detection using Pre-Trained BERT ModelAUTHOR: Jaideep Yadav; Devesh Kumar; Dheeraj Chauhan 2020

Cyberbullying is spread across various social media platforms. It is a wrong deed in which the victim is harassed by receiving the derogatory / provocative / sensitive images or text messages by the bully. Detection of such message/post in such large platforms is very difficult and may sometimes lead to false detection. Recently, deep neural network based models have shown significant improvement over traditional models in detecting cyberbullying. Also, new and more complex deep learning architectures are being developed which are proving to be useful in various NLP tasks. Google researchers has recently developed a language learning model called BERT, which is capable of generating contextual embeddings and is also able to produce task specific embeddings for classification. A new approach isproposed to cyberbullying detection in social media platforms by using the novel pre-trained BERT model with a single linear neural network layer on top as a classifier, which improves over the existing results. The model is trained and evaluated on two social media datasets of which one dataset is small size and the second dataset is relatively larger size.

3. Cyberbullying severity detection: A machine learning approach AUTHOR: Bandeh Ali Talpur ,Declan O'Sullivan 2020 With widespread usage of online social networks and its popularity, social networking platforms have given us incalculable opportunities than ever before, andits benefits are undeniable. Despite benefits, people may be humiliated, insulted, bullied, and harassed by anonymous users, strangers, or peers. In this study, we haveproposed a cyberbullying detection framework to generate features from Twitter content by leveraging a pointwise mutual information technique. Based on these features, we developed a supervised machine learning solution for cyberbullying detection and multi-class categorization of its severity in Twitter. In the study we applied Embedding, Sentiment, and Lexicon features along with PMI-semantic orientation. Extracted features were applied with Naïve Bayes, KNN, Decision Tree,Random Forest, and Support Vector Machine algorithms. Results from experimentswith our proposed framework in a multi-class setting are promising both with respectto Kappa, classifier accuracy and f-measure metrics, as well as in a binary setting.

III. EXISTING METHOD

In this project the state of the art of different aspects related to our approach. On the one hand, bootstrapping technique and semantic similarity metrics are analyzed since both are the pillars of our approach. On the other hand, an exhaustivereview of emotion lexicons and corpora is carried out with the aim of obtaining conclusions and determining the pending issues. We adding and updating some processes to give the output as high accuracy. Lot of research have been done to find possible solutions to detect Cyberbullying on social networking sites. The method used was to select profiles for study, acquire information of tweets, select features to be used from profiles and using ML to find the author of tweets. 1900 tweets were used belonging to 19 different profiles. It had an accuracy of 68% for identifying author. Later it was used in a Case Study in a school in Spain where out of some suspected students for Cyberbullying the real owner of a profile had to be found and the method worked in the case. The following method still has some shortcomings. For example a case where trolling account doesn't have a real account to fool such systems or experts who can change writing styles and behaviors so that no patterns are found. For changing writing styles more efficient algorithms will be needed.

IV PROPOSED SYSTEM

Identifies the features of a cybercrime incident and their potential elements and Provides a two-level offence classification system based on specific criteria. The proposed schema can be extended with a list of recommended actions, corresponding measures and effective policies that counteract the offence type and subsequently the particular incident. In proposed system sentiment similarity analysis has been implemented using machine learning UN supervised and supervised algorithm. There are 4 machine learning algorithms has been used in order to get the based model based on accuracy score. In this method modeled the sentiment classificationproblems as learning sentiment specific word embedded issue and designed three neural network to effectively incorporate the super vision from text data with sentiment labels. Proposed a collaborative detection method where there are multipledetection nodes connected to each other where each nodes uses either different or same algorithm or data and results were combined to produce results. Suggested a B-LSTM technique based on concentration. Banerjee et al. used KNN with new embeddings to get a precision of 93%.

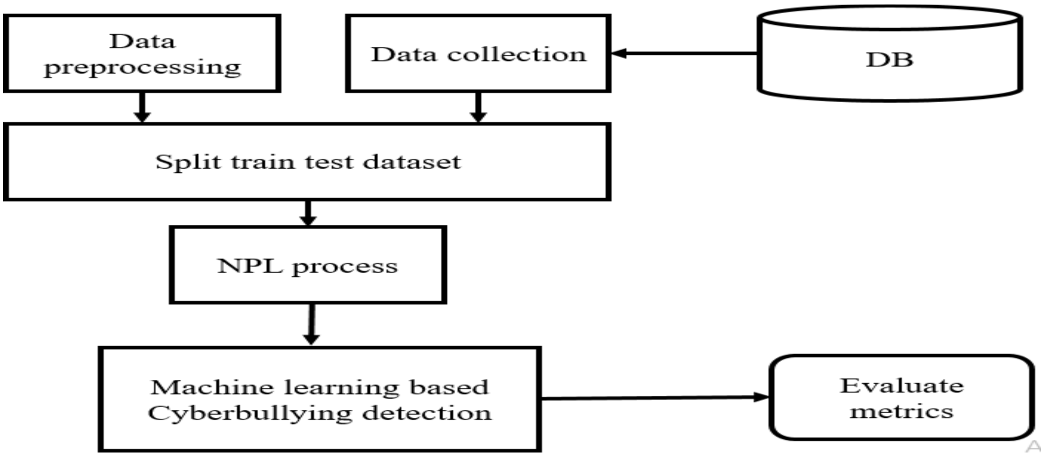


Fig 4.1 System Architecture

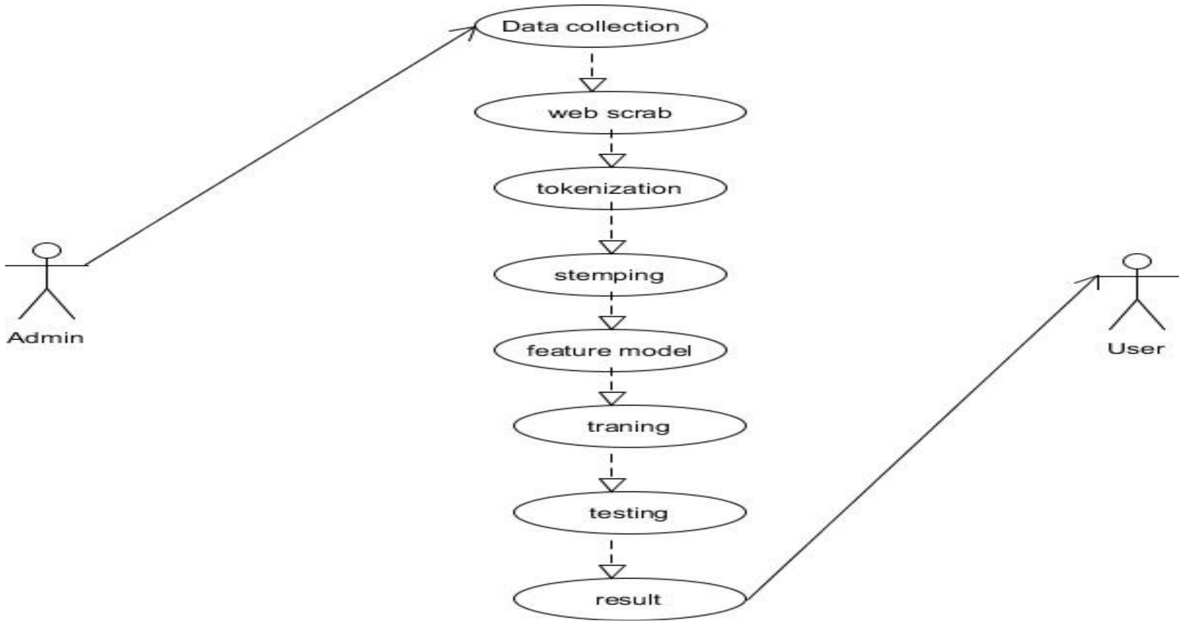
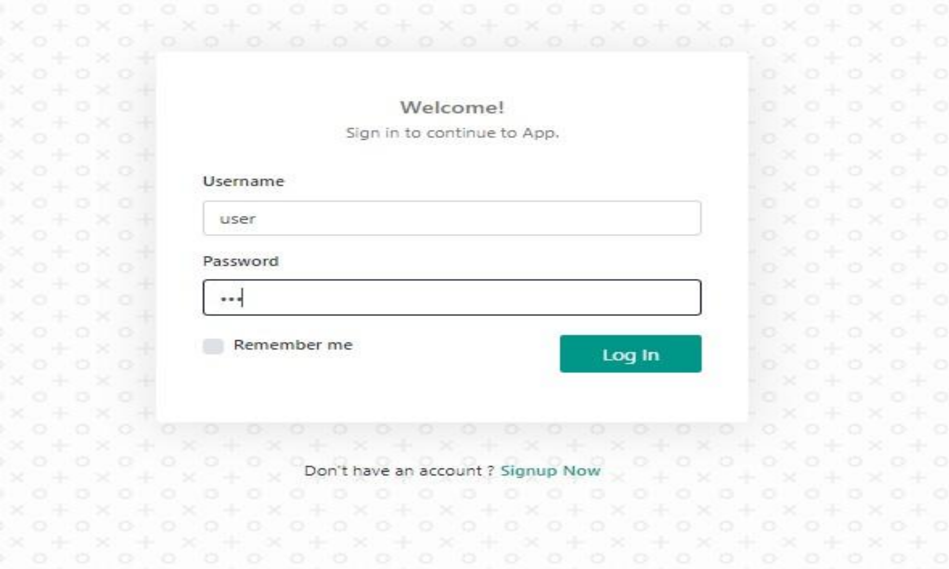


Fig 4.2 Activity Diagram for Receiver

V KEY RESULTS



Welcome!

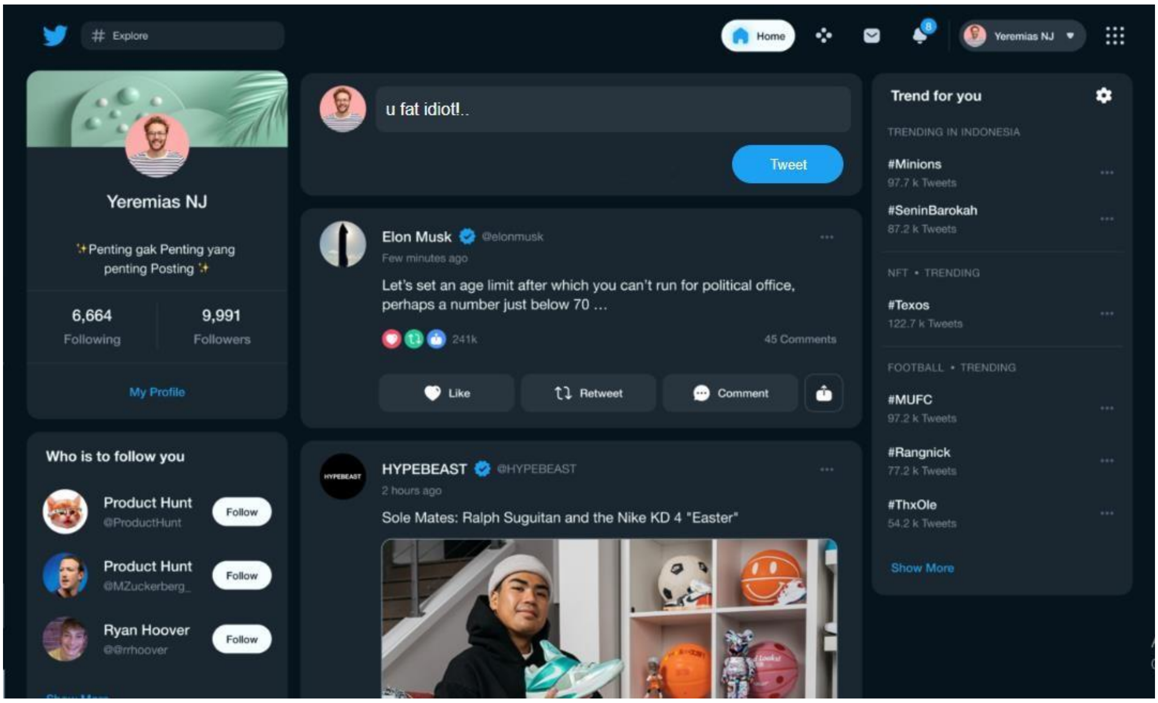
Sign in to continue to App.

Username: user

Password: •••

☐ Remember me

Fig 5.1 Login Page



u fat idiot!..

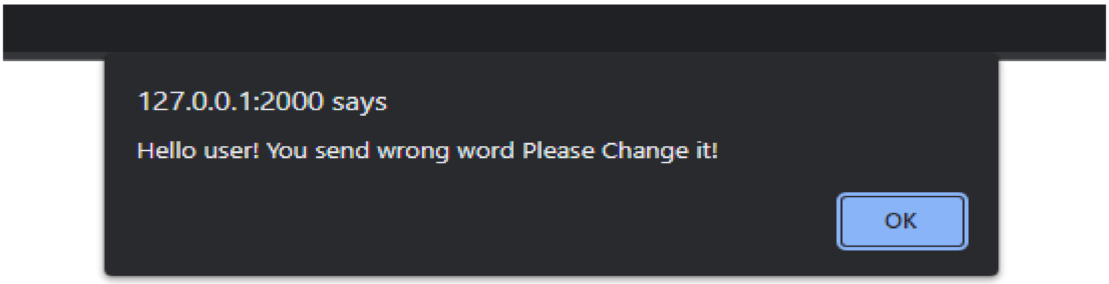
Elon Musk @elonmusk · Few minutes ago

Let's set an age limit after which you can't run for political office, perhaps a number just below 70 ...

HYPEBEAST @HYPEBEAST · 2 hours ago

Sole Mates: Ralph Suguitan and the Nike KD 4 "Easter"

Fig 5.2 Twitter Page User



127.0.0.1:2000 says

Hello user! You send wrong word Please Change it!

Fig 5.3 Post Blocked

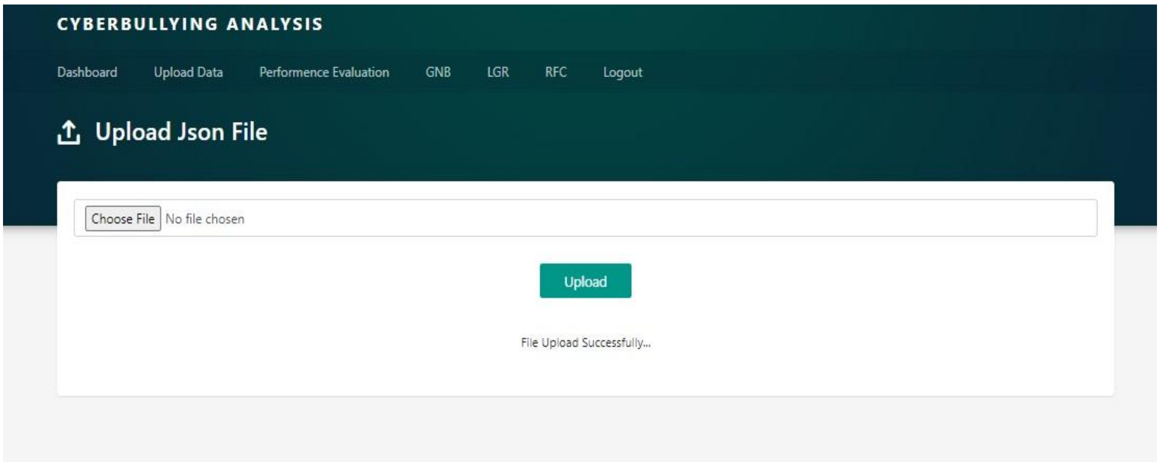


Fig 5.4 File Upload

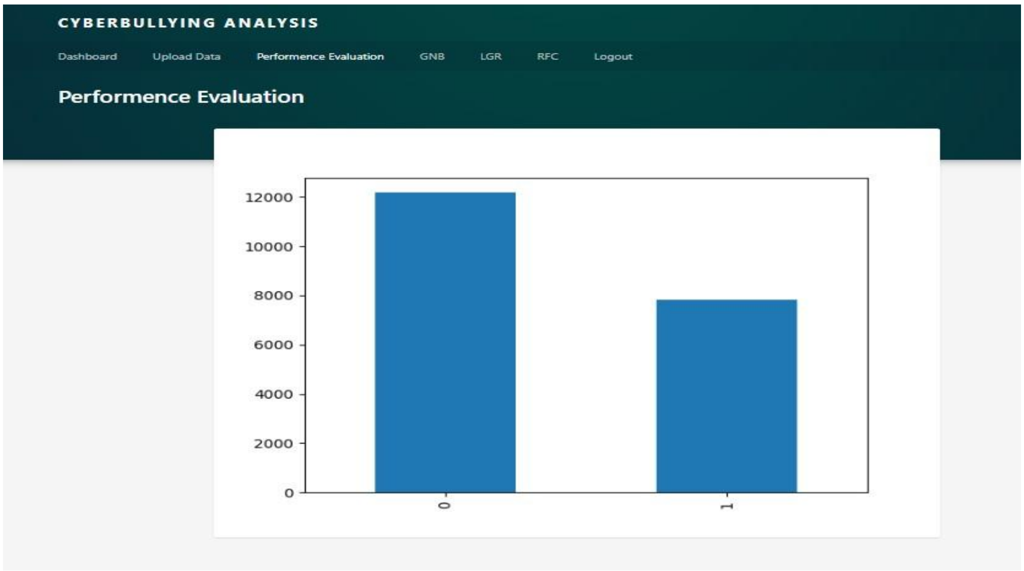


Fig 5.5 Performance Evaluation

VI FUTURE ENHANCEMENTS

In this project to hurt people and bring them harm. Cyberbullying is not a lightmatter. It needs to be taken seriously as it does have a lot of dangerous effects on thevictim. Moreover, it disturbs the peace of mind of a person. Many people are knownto experience depression after they are cyberbullied.

VII CONCLUSION

Cyber bullying across internet is dangerous and leads to mishappenings like suicides, depression etc. and therefore there is a need to control its spread. Therefore cyber bullying detection is vital on social media platforms. With viability of more data and better classified user information for various other forms of cyber-attacks Cyberbullying detection can be used on social media websites to ban users trying totake part in such activity In this paper we proposed an architecture for detection of cyber bullying to combat the situation. We discussed the architecture for two types of data: Hate speech Data on Twitter and Personal attacks on Wikipedia. For Hate speech Natural Language Processing techniques proved effective with accuracies ofover 90 percent using basic Machine learning algorithms because tweets containing Hate speech consisted of profanity which made it easily detectable.

REFERENCES

[1]. M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee,

[2]. H. Ting, W. S. Liou, D. Liberona, S. L. Wang, and G. M. T. Bermudez, “Towards the detection of cyberbullying based on social network mining techniques,” in Proceedings of 4th International Conference on Behavioral, Economic, and SocioCultural Computing, BESC 2017, 2017, vol. 2018-January, doi: 10.1109/BESC.2017.8256403.

[3]. P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, “Supervised machine learning for the detectionof troll profiles in twitter social network: Application to a real case of cyberbullying,” 2014, doi: 10.1007/978-3-319-01854-6_43.

[4]. A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of cyberbullying behavior in Twitter data," 2015, doi: 10.1109/EIT.2015.7293405.

[5]. R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," 2016, doi: 10.1145/2833312.2849567.

[6]. V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, "Detection of Cyberbullying Using Deep Neural Network," 2019, doi: 10.1109/ICACCS.2019.8728378.

[7]. K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," 2011, doi: 10.1109/ICMLA.2011.152. [7] J. Yadav, D. Kumar, and D. Chauhan, "Cyberbullying Detection using Pre-Trained BERT Model," 2020, doi:10.1109/ICESC48915.2020.9155700.

[8]. M. Dadvar and K. Eckert, "Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study," arXiv. 2018.

[9]. S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms," arXiv. 2018.

[10]. Y. N. Silva, C. Rich, and D. Hall, "BullyBlocker: Towards the identification ofcyberbullying in social networking sites," 2016, doi: 10.1109/ASONAM.2016.7752420.