International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Detection of Cyber Bullying on Social Media using Machine Learning

*[1]M Sravanthi, [2]G Niharika, [2]V Ramya

[1]Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

[2]Students, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

Article Info
Volume 9, Issue 5
Page Number : 626-629

Publication Issue
September-October-2022
Pages : 606-611

Article History
Accepted : 10 Oct 2022
Published : 30 Oct 2022

ABSTRACT

In the modern era, the usage of the internet has increased tremendously which in turn has led to the evolution of large amounts of data. Cyber world has its own pros and cons. One of the alarming situations in web 4.0 is cyber bullying, a type of cyber-crime. When bullying occurs online with the aid of technology it is known as cyber bullying. This research paper has surveyed the work done by 30 different researchers on cyber bullying, and elaborated on different methodologies adopted by them for the detection of bullying. Three types of features namely textual, behavioral and demographic features are extracted from the dataset as compared to earlier study over the same dataset where only textual features were considered. Textual features include certain bullying words that if exists within the text may lead to a true outcome for cyber bullying. Personality trait features are extracted for the user if it is involved once in bullying may bully in future too. While demographic features extracted from the dataset include age, gender and location. The system is evaluated through different performance measures for both classifiers used and the performance of the Support Vector Machine classifier is found better than the Bernoulli NB with overall 87.14 accuracy.

Keywords : Cyber Bullying, Machine Learning, SVM, NLP.

I. INTRODUCTION

Across the globe due to the tremendous increase in the availability of data services, addiction of social media among society has increased proportionally. Just like other countries, India has also witnessed a drastic rise in cyber bullying. In this era of web 4.0 where people live on digital and online platforms, it is very difficult to protect society from the alarming rise in cyber-crime. It has been surveyed that the major victims of cyberbullying are adolescents. Different cyber bullying attacks that are performed by attackers are: (1) Sending or posting hateful or abusive comments with an intention to harm the character of an individual (2) Posting an inappropriate image or video. (3) Creation of a false or improper website. (4) Issuing Online threats that cause a person to kill themselves or injure another person. (5) Triggering online religious, racial, ethnic or political hatred by posting hate comments. Because of the tremendous increase in the availability of data services around the world, social media addiction in society has increased proportionally. It is extremely difficult to protect society from the alarming rise in cyber-crime in this era of web 4.0, where people live on digital and online platforms. According to surveys, adolescents are the most common victims of cyberbullying. The following are examples of cyberbullying attacks carried out by attackers: (1) sending or posting hateful or abusive comments with the intent to harm an individual's character (2) Posting an offensive image or video. (3) Creating a false or inappropriate website. (4) Making online threats that cause someone to kill themselves or injure someone else. (5) Inciting religious, racial, ethnic, or political hatred online through the posting of hateful comments or videos.

II. RELATED WORK

Cyberbullying is a persistent problem in Saudi schools, exacerbated by the advancement of digital technology and its pervasive presence in almost every societal aspect. With such technologies, it is unsurprising that harassment has spread to the virtual world of teenagers, where it is rampant. The intensity and outcome measures of this phenomenon have alarmed interested parties, but researchers who examined the causes and motivating factors behind cyberspace bullying participation are few and far between. 2

The Theory of Planned Behavior, a well-known theory, was used to examine this issue (TPB). This study specifically looked at the effects of attitudes, normative beliefs, subjective norms, and perceived behavioral control/self-efficacy on cyberbullying intentions and expected societal outcomes. 1The study distributed 395 questionnaires to Saudi high school students from the ninth to the twelfth grades. The collected data were subjected to multiple linear regressions, with the results revealing that behavioral attitudes, social norms, perceived behavioral controls, social media use, a lack of parental controls, and a lack of regulations all had a direct effect on intentions to interact in cyberbullying. The findings also revealed that cyberbullying intentions had a direct impact on student academic performance.

This study adds to our understanding of students' intentions toward cyberbullying and the relationship between the Theory of Planned (TPB) variables and the predictive utility model. Finally, the findings of this study can be used to develop prevention and intervention strategies, which have many implications for theory, practice, and policy. Cybercrime refers to any crime committed using the internet as an access medium and using an electronic device such as a computer or a mobile phone. The main factors limiting previous research in cyberbullying detection have been a lack of datasets, predators' hidden identities, and victims' privacy. Taking these factors into account, an effective text mining approach based on machine learning algorithms is proposed for proactively detecting bullying text. The dataset gathered from myspace.com and PervertedJustice.com was being used to assess the system's performance. When compared to a previous study on the same dataset that only considered textual features, three types of features are extracted from the dataset: textual, behavioral, and demographic features. Age, gender, and location are among the demographic features extracted from the dataset. The system is evaluated using various performance measures for both classifiers used, and the performance of the Support Vector Machine classifier is found to be better than the Bernoulli NB with an overall accuracy of 87.51.

III.PROPOSED SYSTEM

Twitter dataset may easier to extracted compared to other mediums such as Facebook , Instagram, and YouTube. Even though statistic brain. Come aforementioned stated that cyber bullying occurred most in Facebook but only data from public profiles could be extracted easily such as Twitter that the data is publicly available. The main function was to extract social media public data using available API. Then next step is to data cleaning and Pre-processing. As the extracted data had multilingual unstructured content along with lot of emoji, it was required to clean the data for higher accuracy. Several supervised machine learning algorithms were compared to identify the best one.

Frequent use SVM by researches shows that SVM is popular among other classifiers in supervised learning approach. SVM is suitable for high-skew text classification such as to detect cyber bullying using content based features. Any circumstances such as missing data, type of feature and computer performance, SVM still perform other classifier. These features are generally obtained by statistical analysis of documents (tweets or sentences):

Bad words:

From literature is quite evident and intuitive that some "bad" words make a text a suitable candidate to be labeled as a possible cyber bullying sentence. As just done in other works, and in this work we have identified a list of insults and swear words (550 terms), collecting these terms from different online available sources.

Bad words density:

In this work we check also the density of "bad" words as a single feature. This features is equivalent to the number of bad words that appear in a sentence, for each severity level, divided by the words in the same sentence.

Badness of a sentence:

We also add a feature to our work in order to measure the overall "badness" of a text. This feature is computed by taking a weighted average of the "bad" words (weighted by a severity assigned).

Density of upper case letters:

This feature is based on Dadvar et al. [7] results. The presence of capital letters in a text message is selected as a feature, considering it as possible 'shouting' at someone behavior, as commonly treated in social networks netiquette. This feature is given by the ratio between the number of upper case letter and the length (number of chars) of the whole sentence.

Exclamations and questions marks:

Just like capital letters, also exclamation points and question marks can be considered as emotional comments. We just stated that cyber bullying is related to an extreme case of sentiment analysis and so it can be connected to the strong (usually bad) emotions. With this premise, we consider helpful to introduce the number of exclamation points and question marks as a feature in our work.

The preprocessing step is done in the following:

- Tokenization: In this part we take the text as sentences or whole paragraph and then output the entered text as separated words in list.
- Lowering text: This takes the list of words that got the out of tokenization and then lower all the letters Like: "THIS IS AWESOME" is going to be 'this is awesome'.
- Stop words and encoding cleaning: This is an essential part of the preprocessing where we clean the text from those stop words and encoding characters like '*' which do not provide a meaningful information to the classifiers.

IV. CONCLUSION

In this work tries to address the issue of cyber-bullying in Twitter platform using Machine Learning. Experiments were carried out with both supervised and unsupervised machine learning techniques. It was observed that identifying the right set of keywords is an essential step for getting better results during sentiment analysis Results indicate that our model achieves reasonable performance and could be usefully applied to build concrete monitoring applications to mitigate the heavy social problem of cyber bullying.

An experimental result indicates that the SVM based method achieves best accuracy and the performance improves if user specific data can be included. Due to highdimensional input space, few irrelevant features and linearly separable nature of text dataset, SVM performs better than other classification algorithm for text classification. In future, significance of individual features can be studied for further enhancement of the method.

V. REFERENCES

[1]. Rice, Eric, et al. “Cyber bullying perpetration and victimization among middle-school students.” American Journal of Public Health (ajph), pp. e66-e72, Washington, 2015.

[2]. Bangladesh Telecommunication RegulatoryCommission, http://www.btrc.gov.bd/content/internet-subscribers-Bangladeshjanuary-2018, [Last Accessed on 18 Mar 2018].

[3]. Mandal, Ashis Kumar,Rikta Sen. "Supervised learning methods for Bangla web document categorization." International Journal of Artificial Intelligence & Applications, IJAIA, Vol 5, pp. 5,10.5121/ijaia.2014.5508

[4]. Dani Harsh, Jundong Li, and Huan Liu, “Sentiment Informed Cyberbullying Detection in Social Media” Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Spinger, Cham, 2017

[5]. Dinakar, Karthik, Roi Reichart, and Henry Lieberman. “Modeling the detection of Textual Cyberbullying.” The Social Mobile Web 11.02(2011):11-17

[6]. K. Dinkar, R. Reichart and H. Liebernman, “Modeling the Detection of Textual Cyberbullying,” MIT. International Conference on Weblog nd Social Media. Barcelona, Spain, 2011.

[7]. M. Dadvar and F.de Jong. 2012.”Cyberbullying detection:astep toward a safer internet yard”. In Proceedings of the 21st International Conference on World Wide Web(WWW ’12 Companion). ACM, New York, NY, USA, 121-126

[8]. Sunil B. Mane, Yashwanth Sawant, Saif Kazi, Vaibhav Shinde,“Real Time Sentiment Analysis of Twitter Data Using Hadoop”, International Journal of computer Science and Information Technologies,(3098-3100),Vol.5(3),2014.

[9]. Riya Suchdev, Pallavi Kotkar,Rahul Ravindran, “twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach”, International Journal of Computer Applications(0975-8887),Volume 103 a No.4, October 2014.

[10]. J. Xu, K. Jun, X. Zhu, and A. Bellmore, “Learning fromBulling Traces in Social Media,"Proc.2012 Conf.North Am. Chapter Assoc. Comput. Linguist. Hun. Lang. Technol, pp. 656-666,2012

[11]. S. Hnduja and J. W. Patchin “Cyberbullying: Identification, Prevention, & Response,”Cyberbullying Res. Cent, no. October, pp. 1-9,2018

[12]. A. saravanaraj, J. I. sheebaassistant, S. Pradeep, and D. Dean, “Automatic Detection of Cyberbullying From Twitter.” IRACST-International J. Comput. Sci. Inf. Technol. Secur., vol. 6, no. 6,pp. 2249-9555,2016.

Cite this article as :

M Sravanthi, G Niharika, V Ramya, "Detection of Cyber Bullying on Social Media using Machine Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 5, pp. 626-629, September-October 2022.
Journal URL : https://ijsrst.com/IJSRST2182943