IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101272

Twitter-Cyberbullying Detection using Machine Learning

Namrata Khade[1], Snehal Sarkate[2], Palak Kombade[3], Vaishnavi Alone[4], Vaishnavi Parate[5]

Assistant Professor Dept. of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India[1]

UG Students, Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India[2,3,4,5]

Abstract: Our paper provides detection of Cyberbullying using machine learning. In this project, we aim to build a system that tackles Cyberbullying by identifying the mean-spirited comments and also categorizing the comments as bullied one or not. The goal of this project is to show the implementation of software that will detect bullied tweets. A machine learning model is proposed to detect and prevent bullying on Twitter. Social media is a platform where many people are getting bullied. To identify word similarities in the tweets made by bullies and make use of machine learning and detect social media bullying actions. As social networking sites are increasing, cyberbullying is increasing day by day in today's world. Cyberbullying is a crime in which a perpetrator targets a person with online harassment and hate. As to detect the cyberbullying a GUI (Graphical User Interface) is created to detect where tweets are used to detect the cyberbullying. Cyberbullying includes insulting, humiliating and making fun of people on social media that can cause mental breakdowns for the victims, it can affect one physically as well to the extent that can also lead to suicidal attempts. We are using classifiers-Naive Bayes, SVM (Support Vector Machine), Random Forest, Decision Tree and Sklearn. As for the classification phase, machine learning will be used. Two classifiers i.e., SVM and Naïve Bayes are used for training and testing the Twitter bullying content. Both Naive Bayes and SVM (Support Vector Machine) were able to detect the true positives with 71.25% and 52.70% accuracy respectively. But Naive Bayes outperforms SVM of similar work on the same dataset.

Keywords: Cyberbullying detection · Machine Learning · Twitter· Tweets · Online harassment.

I. INTRODUCTION

Cyber bullying is a kind of bullying that occurs over digital devices that includes phones, laptop's computers, tablets etc. Social Media is the use of virtual platform for connecting, interacting, sharing of contents and opinion around the globe. Since the development of social sites platforms, its usage by teens and adults across the globe has seen great upsurge.

- Twitter serves as a platform to share both online and offline bullying experiences.
- Most tweets from the victim's perspective and commonly to report or self-disclosed in many ways.
- Bullying-related tweets were considerably longer than the average tweet.
- High profile bullying incidents bring about an increase in bullying-related tweets.
- The project included tactics expanded keywords with cyberbullying keywords and mean words.

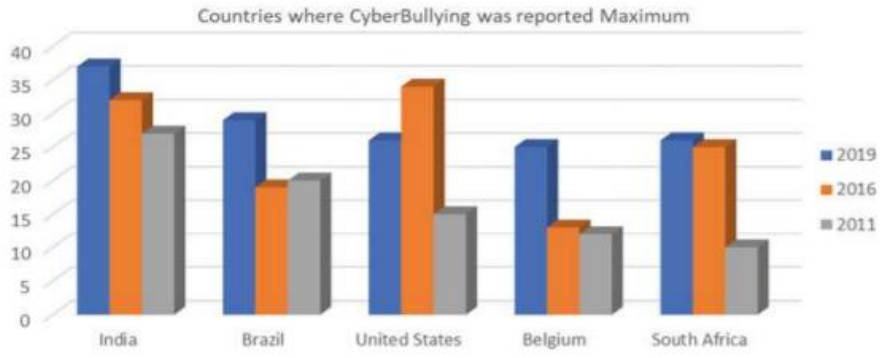


Fig 1: Countries where Cyberbullying was reported Maximum.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101272

II. PROPOSED SYSTEM:

- As Cyberbullying can be easily committed, it is considered a dangerous, life-threatening and fast spreading aggressive behavior. In this study, a content analysis is performed to predict aggressive behavior.
- In this project, we have considered two phases to predict the cyberbullying on the Twitter account by detecting the tweets as bullied or non-bullied one. To check this the first phase is to determine the data collection step as the prominent one to check whether the tweets are bullied.
- The second phase is to check the established paper accuracy model given and to modify them in the paper as the resource where the model can be a smart to predict itself as machine learning help itself to detect the predicting model.

III. RESEARCH METHOD

In this project, we collected the data from Kaggle. Initially it starts with a fetching of tweets from Twitter database or we can enter the tweets as an input using the keyboard too. Tweets are fetched and then pre-processed using python code and machine learning such that mean words, harassment words, noisy and irrelevant data are removed and then processed, words are then tokenized. The reason for selecting this dataset is that it is well-suited for our study as it contains the topics of cyber bullying that we are interested in.

As for the working project module a GUI interface is made to predict the bullying of the tweets in the Twitter account. To predict that we are using machine learning to detect the bullying in which the tweet contains mean words, harassment words which lead to victims go under depression. Victim go through various emotionally, physical and mentally behavior which is hard to predict once it gets hits to the victim. To cure it takes more to get the victim out of it.

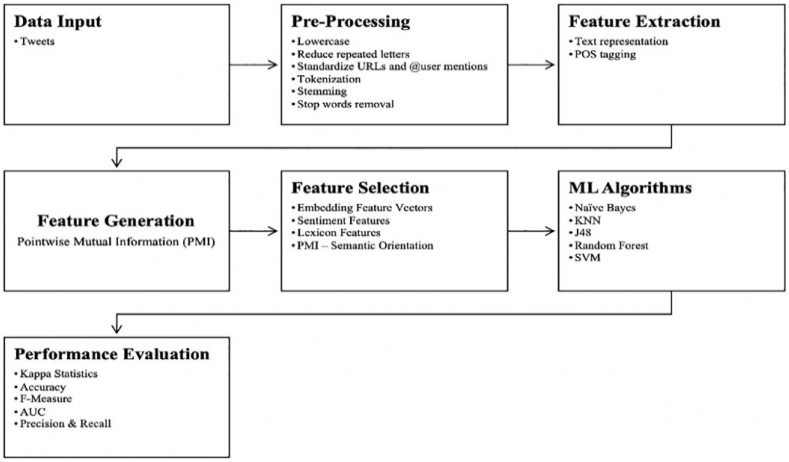


Fig 4: Proposed Approach

IV. FINDINGS AND ANALYSIS

There are various types of cyber bullying like sex, race, gender, religion, disability and many more. Here we provide a system where one gets to know if the tweet is bullying or not. We used machine learning model to classify if the tweet is bullying or not bullying. Our project has different modules such as :

1.Data Collection.
2. Data Preprocessing.
3. Training and Testing Dataset.
4.GUI Application.

1. Data Collection :

The Dataset is collected from Kaggle, an online website where datasets are available to use int the projects. This dataset

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101272

contains in general 1066 conversations messages collected from clean project dataset. The reasons for selecting this dataset are that it is well-suited for our study as it contains the topics of cyber bullying that we are interested in.

2. Data Preprocessing :

Data preprocessing is a technique by which raw data is converted into the useful information. At this stage the process of tokenization was applied. In this method data is processed to give a specific required result. The data is processed using where mean words, harassing words are recognized and distinguished using machine learning.

3. Training and testing Dataset :

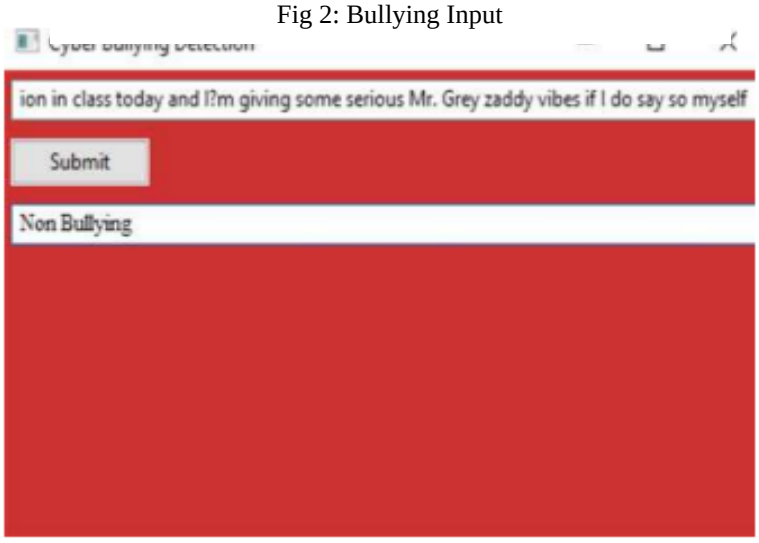
After this process the training and testing process was carried out. Training is done on the dataset where after preprocessing of the data is used for training and testing the model and is also used as the final dataset which will be used for predicting the bullying tweets.

4.GUI (Graphical User Interface) Application :

This is the last module and the working module which is the frontend used by the used to detect the bullying tweets. By pasting the tweet in the search bar and then predicting, gives the result of bullying tweet or non-bullying tweet.

Various machine learning classifier is used which gives the highest accuracy and will be used in the model to predict the bullying status. In this project we are using classifiers such as Naïve Bayes, Support Vector Machine, Decision Tree and Sklearn. In which Naïve Bayes gives the highest accuracy followed by Support Vector Machine.

Fig 2: Bullying Input



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101272

V. RESULT

Above figure shows the end user display of the proposed approach where the system efficiently distinguishes the text into its bullying category.

VI. FUTURE SCOPE

As this project predict bullying by pasting the tweet in the search bar and the detecting bullied or not and as for the future scope it can also be developed by adding features which will detect in the real time. By commenting tweet in the Twitter account, it will detect the tweet as the harassment one and will not let the attacker to tweet it. By adding new features in the Twitter application, it can be further modified to use as a natural social media app. As in this project it is applied in the Twitter, so it can be further be applied in various social media app.

VII. CONCLUSION

The use of internet and social media has clear advantages for societies, but their frequent use may also have significant adverse consequences. This involves unwanted sexual exposure, cybercrime and cyberbullying. We developed a model for detecting cyberbullying behavior and its severity in Twitter. The developed model is a feature-based model that uses features from tweets contents to develop a machine learning classifier for classifying the tweets as cyberbullying or non-cyberbullying.

VIII. REFERENCES

1. https://doi.org/10.1016/j.ipm.2009.03.002
2. https://doi.org/10.1371/journal.pone.017161
3. Fire M, Goldschmidt R, Elovici Y. Online Social Networks: Threats and Solutions. IEEE Commun Surv Tutor. 2014; 16: 2019–2036
4. https://doi.org/10.1109/JBHII.2019.2902303
5. https://doi.org/10.1109/TKDE.2005.50
6. https://doi.org/10.2307/2529310
7. Abu-Nimeh S, Chen T, Alzubi O. Malicious and Spam Posts in Online Social Networks. Computer. 2011; 44: 23–28
8. https://doi.org/10.2307/3315487
9. https://doi.org/10.2147/AHMT.S36456
10. https://doi.org/10.1080/09627250408553239
11. https://doi.org/10.1109/MC.2011.222
12. https://doi.org/10..1145/2184319.2184338

IX. BIOGRAPHY

Mrs. Namrata S. Khade is working as a Asst. Professor at Priyadarshini College of Engineering. She is having 12 years of experience in the field of teaching to engineering students. She completed her Engineering in 2007 and Master in Engineering in 2013.She is a member of IEEE, ISTE and CSI. She is having more than 30 research published in International Journals and Conferences. Her interests include distributed parallel computation, System Programming, Computer Graphics and Wireless Sensor Network.