International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

IDENTIFYING CYBER-BULLYING USING MACHINE LEARNING

A1: BABLI KUMARI

MCA Student, Dayananda Sagar College of Engineering
CO-AUTHOR: Dr. Suma S
Associate Professor, Department of MCA, Dayananda Sagar College of Engineering

---

Abstract–As Social media is a growing medium nowadays, which provides an easy access to an individual or a group thus sometimes results in misbehaving, harassment, bullying, threatening, use of vulgar words thus disturbs and hurt the sentiments of people using electronic means which is known as Cyber-Bullying or online bullying.
There are various strategies to torment other people which includes engaging themselves in warning-wars, actively participating in comment war or text attacks, posting mean, harsh or insulting remarks on the comment section, postingrumors, threats or embarrassing information which lead to exposing of privacy and hurts the sentiment of person.
It also includes the impersonation of another person, which causes deep effects in the person's life the impersonator could change the target's online profile to inappropriate things, create a duplicate profile of a person, set up a social media account to post as a victim and extort money, sometimes stealing the victim's password and other useful information.
So in this research paper we are going to detect the cyber-bullying content the comments, the remarks and take the appropriate actions against the person/people who does bully others.

Key Words:cyber-bullying, misbehaving, threatening, vulgarity.

1. INTRODUCTION

Cyber-Bullying is an act of intimidation or harassment of people through online medium whichis done on/by digital devices likes cell phones, laptops, desktop etc, the online medium includes the social media applications like facebook, Instagram or through text messaging or through email or online gaming communities. It has affected the society enormously by causing emotional, psychological and physical distress**.** Cyber-Bullying makes the victim feel threatened, it is traumatic for people because it can reach to their houses, where they generally feel the safest, Victims feels intense fear that they are inferior, Brain washing techniques makes them traumatized,The continous and pervasive cycle of cyber-bullying often results in victim suffering from depression, anxiety and inferior complexity, It leads to intense introvert behavior and sometimes result into suicide.

2. Body of Paper

A. Consequences of Cyber-bullying

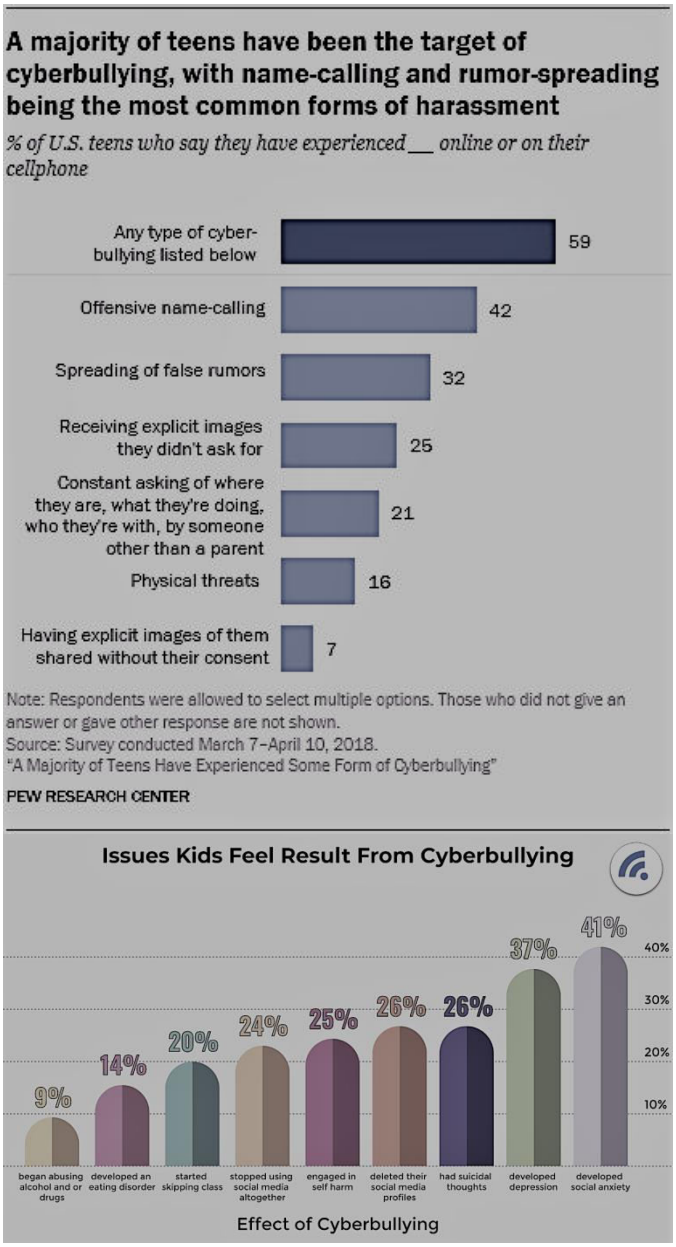
Consequences are way severe for the victims who have encountered such harassment and brutality. Invading the privacy of victim, ill-treating the victim often leads to psychological and physical distress. Often the victim's peers cut contacts or starts avoiding him/her in order to avoid the wrath of the bullies which causes the mental breakdown of the victim.
Victims end up feeling so inferior with low self-esteem, he/she goes far from self-actualization, they lose self-control as they don't believe in themselves anymore which results in poor health of the victim as anything put under constant pressure for a longer period of time couldn't function anymore and sometime leads to death.
Anxiety and fear arounds the victim all the time as he/she thinks herself/himself incapable to fight or correct what someone has done to them, also it makes them feel vulnerable as in their minds they feel unprotected and in danger all the times
Depression: As Cyber-bullying leads to the social dysfunction victim often loose the friends and family which leads them with a constant disability to make friends anymore, it drains victim's social life completely and sets the victim to depression especially at a young age.

B. Recent Studies

Recent studies have found that with the growing technology and the excessive use of social media. People are letting social media to enter into their private space, which include bringing their bullies home, which hasnot remain safe anymore.
Almost 70% of children aged 10-17 are the frequent victims of being bullied or harassed over the internet but due to circumstances 91% of these victims never ever reports the

incidents which gives more liberty to the offenders.

New Research center survey claims that 60% of teenagers have experienced abusive online behaviors.



C. Related Work

There have been many approaches which propose systems that could detect the cyber-bullying automatically with high accuracy. One of which is Naïve Bayes machine learning approach with accuracy rate 91%. Another one done by advancing and enhancing the Naïve Bayes classifier where the words are extracted and examined using pattern clustering which made it 96% accurate, but with a drawback that this process doesn't not work in parallel.

Another type of approaches uses deep learning and neutral network. One of the approaches is done on twitter where the twitter data is checked, its accuracy could not be calculated because of its imbalancy but it was able to achieve 56% precision and 96% accuracy.

D. Data Collection

A. Origin of Dataset

We have taken the social media application Reddit which is a platform where a user can ask question as well as answer in his/her interesting field and accordingly they can upvote or downvote answers which other users have written.

It is a great platform where an individual can write about their thoughts, share their experiences and connect with others too, it brings people together.

The reason behind choosing this platform is as its people oriented depends on the interaction, content is producing at a higher rate with all age group people interacting with one another through comments ,messaging and answers, hence the bullying rate is quite high.

B. Data Labeling

The labeling of data is done by analyzing the posts and the comments related to it on different sub reddits by using PRAW which is an abbreviation for Python wrapper; it is a python package that allows for simple access to reddit's API, it basically fetches the comments of various posts and thus classified using Naïve Bayes classifier.

E. Proposed Approach

The online versionIt contains three main steps preprocessing, featuring extraction and classification step. At preprocessing step the data is filtered by removing the noise and unnecessary text. Its steps are:

- **Tokenization:**Theinput text is taken in form of sentence or paragraph then converting it into a separate words of a list.
- **Lowering Text:**The tokenization words are then converted into lowercase letters for eg 'I AM GREAT' to 'i am great'.
- **Stop words and encoding cleaning:**Its theimportant part of the preprocessing where the text is

cleaned from stop words and encoding characters like \t or \n which are not of much significance.

- **Stemming:** This step takes the list of words to convert it back into its original form where no suffixes or prefixes are used.

The next step involves transformation of textual data into a format that is accepted by machine learning algorithm, in this step the textual data is extracted using TFIDF (Term frequency-Inverse Document Frequency).
The key feature of TFIDF is to get the weight of the words with respect to sentence or paragraph.

The last step involves the classification step where the extracted data is sent into a classification algorithm to test and train the classifier, thus used in the prediction phase, the classifier used is SVM (Support Vector Machine).
The evaluation of classifiers is done by using various evaluation metrices whose criteria involves Accuracy, precision, recall and F-score.

Algorithm

It involves a vocabulary of known texts and measure of presence of known words while extracting the document vectors we use Boolean values 0 for not present and 1 for present.

We have used SVM for the proposed model because of following reasons:

- It is highly effective in dimensional space
- It is memory efficient
- It is versatile as different kernel functions can be applied to it
- More effective when number of dimensions exceed than number of inputs.

3. CONCLUSIONS

In this research paper, the proposed approach is to identify cyber-bullying using the machine learning techniques, the proposed model is evaluated on two classifiers SVM and Naïve Bays classifier and for feature extraction the TFIDF has been used. We have also used SVM for the custom feature extraction that implies any user who is getting bullied or offended with any word can mention it and eliminate it. The main motive of our proposed approach is to help people to use social media safely and without any hesitation.

ACKNOWLEDGEMENT

This research was carried out based on my self-experience and self-studying of different incidentsof the cyber-bullying.

REFERENCES

1. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108–121
2. Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, Berlin Heidelberg New York (1997) 415–438
3. van Leeuwen, J. (ed.): Computer Science Today. Recent Trends and Developments. Lecture Notes in Computer Science, Vol. 1000. Springer-Verlag, Berlin Heidelberg New York (1995)
4. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin Heidelberg New York (1996)