International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 05, May-2021

Cyber Bullying Detection using SGD Classifier

Prof. D. H. Patil [1], Gautami Kharul [2], Pranjali Gaikwad [3], Vaishali Khawse [4]
Information Technology
JSPM's Rajarshi Shahu College of Engineering, PUNE

***Abstract*— This paper describes a system for automatic detection of cyberbullying issues and challenges. The system is built to detect an activity on social media which can harm people such as hate speech, abusive language. People spread hatred toward a person in social networking. It does not affect only for health but also in many different aspects. Social media may have some side effects such as cyberbullying, which may have negative impacts on the life of people, especially children and teenagers. Automatic detection of such incidents requires intelligent systems. The subject discussed in this paper begins with a cyberbullying introduction: concept, categories and roles. Then, available data sources, features and classification methods used are analyzed in the discussion of cyberbullying identification. The famous methods used to classify bullying keywords inside the corpus are Natural Language Processing (NLP) and machine learning algorithms. The objective of this paper is to develop a social media network with some functionalities and demonstrate how cyberbullying can be prevented on social media.**

Keywords*— *Cyber bullying detection, social media, pre-processing, classifier algorithm, feature extraction, NLP, SGD classifier etc.

I. INTRODUCTION

Cyberbullying includes a person doing threatening acts against another person, stalking, etc. Cyberbullying means a groups or individuals of individuals who take advantage of social media to harass other individuals. The use of electronic media or communication channel to bully a person, typically by sending messages of a threatening nature is known as cyberbullying. Cyberbullying and victims are the key positions involved with cyberbullying cases. There are different explanations why it occurs.

As the online communication is increasing day by day, cyberbullying is a bigger problem than traditional bullying. Hence to minimize and stop the cyberbullying the detection of is very important and it has really a practical significance. It is possible to detect bullying features used by various bullies and their victims and develop an effective model to detect cyberbullying content using the various machine learning technique to ensure healthy social environment. It happens by making use of information and communication technologies such as cellular phone, text messaging, E-mail, chat sites, social media platforms etc. in various social networking devices or sites intended to hurt or bully people. These activity poses a real threat to a number of young adults, teenagers and individuals who are the active online users.

Cyberbullying can be identified by repeated behavior and an intent to harm. Automatic detection and prevention of these incidents can substantially help to tackle this problem. Cyberbullying can harm the online reputations of everyone involved – not just the person being bullied, but those doing the bullying or participating in it. With the rapid growth of social media, users especially teenagers are spending significant amount of time online and beside all the benefits that it might bring them, their online presence also make them vulnerable to threats and social misbehaviors such as cyberbullying.

It includes-

- Sending improper text messages.
- Posting statements online that are unacceptable.
- Making negative comments.
- Blackmailing with certain demands.
- Stalking and use of intimidation.
- Threats of violence or death.
- Hate-related communications or actions.

II. LITERATURE SURVEY

[1] A "Deeper" Look at Detecting Cyberbullying in Social Networks
Author: Ricardo Ribeiro, Luisa Coheur

As cyberbullying becomes more common in social media, automatically identifying it and taking proactive steps to address it becomes critical. A thorough examination of the current state-of-the-art in cyberbullying detection revealed that, despite their increasing reputation in other text-based classification tasks, deep learning techniques have been rarely used to address this issue.

[2] "Detecting Suspicious Texts using Machine Learning Techniques"
Author: Omar Sharif1, Mohammed Moshiul Hoque ∗, A. S. M. Kayes2

The performance of the proposed system is compared with the human baseline and existing

ML techniques. The SGD classifier 'tf-idf' with the combination of unigram and bigram

features are used to achieve the highest accuracy

[3] "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter"
Author: Amgad Muneer 1,* and Suliman Mohamed Fati 2

In this study, various classifiers have been used to classify whether the tweet is cyberbullying or non-cyberbullying. The classifier models constructed are LR, Light LGBM, SGD, RF, AdaBoost, naïve Bayes, and SVM. These classifiers have been discussed, and the evaluation of their performance is carried out.

[4] "Perceived Distress Associated with Acts of Conventional and Cyber Bullying"
Author: Jaideep Yadav, Dheeraj Chauhan

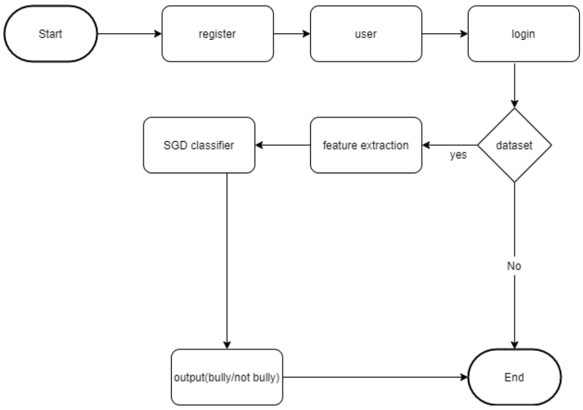
Cyberbullying has received considerable attention, and experts have made several assumptions about this phenomenon. In particular, experts have speculated that the potential harm from cyberbullying is greater than that from conventional bullying, but this assumption has not been confirmed empirically

III. ALGORITHM

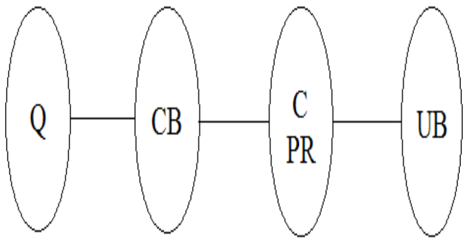
- SGD Classifier is a linear classifier (SVM, logistic regression) optimized by the SGD. These are two different concepts. While SGD is an optimization method, Logistic Regression or linear Support Vector Machine is a machine learning algorithm/model

- After the training the classifier, we'll check the model accuracy score. Now, we can predict the test data by using the trained model. After the prediction, we'll check the accuracy level by using the confusion matrix function

- Stochastic Gradient Descent (SGD) is a simple yet efficient optimization algorithm used to find the values of parameters/coefficients of functions that minimize a cost function. In other words, it is used for discriminative learning of linear classifiers under convex loss functions such as SVM and Logistic regression.
- Proposed system
  (Diagram)



IV. MATHEMATICAL MODELLING



Where,
Q = User entered input
CB = preprocess
C = feature selection
PR = preprocess request evaluation
UB = predict outcome

Set Theory

1) Let S be as system which input image
S = {In, P, Op, Φ}
2) Identify Input In as
In = {Q}

Where,
Q = User entered input (dataset)

3) Identify Process P as

P = {CB, C, PR}
Where,
CB = Pre-process
C = feature selection
PR = Pre-process request evaluation
4) Identify Output Op as
Op = {UB}
Where,
UB = Predict outcome
Φ=Failures and Success conditions.

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 cybercrime and cyberbullying. We developed a model for detecting cyberbullying activities and its severity in social media networks. The feature extraction is a critical process in machine learning to enhance the detection accuracy. One of the improvements is to incorporate and test different feature extractions to improve the detection rate of both classifiers LR and SGD. 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.

REFERENCES

[1] A “Deeper” Look at Detecting Cyberbullying in Social Networks Hugo Rosa INESC-ID Instituto Superior Tecnico, Universidade de Lisboa ´ Faculdade de Psicologia, Universidade de Lisboa hugo.rosa@l2f.inesc-id.pt Ricardo Ribeiro INESC-ID ISCTE-IUL, Instituto Universitario de Lisboa ´ ricardo.ribeiro@inesc-id.pt

[2] Bangla Text Document Categorization Using Stochastic Gradient Descent (SGD) Classifier Fasihul Kabir∗ , Sabbir Siddique† , Mohammed Rokibul Alam Kotwal‡ , Mohammad Nurul Huda§ Department of Computer Science and Engineering United International University Dhaka, Bangladesh

[3] 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE-2020), 07-08 February 2020, (IEEE Conference Record 48199 Text Classification on Twitter Data Dr. Priyanka Harjule, Astha Gurjar, Harshita Seth, Priya Thakur Department of Computer Science IIIT Kota

[4] “Detecting Suspicious Texts using Machine Learning Techniques” Author: Omar Sharif1 , Mohammed Moshiul Hoque ∗ , A. S. M. Kayes2

[5] “A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter” Author: Amgad Muneer 1,* and Suliman Mohamed Fati 2

[6] “Perceived Distress Associated with Acts of Conventional and Cyber Bullying” Author: Jaideep Yadav, Dheeraj Chauhan

[7] M. ElSherief, V. Kulkarni, D. Nguyen, W. Y. Wang, and E. Belding, “Hate lingo: A target-based linguistic analysis of hate speech in social media,” in 12th Intl. AAAI Conference on Web and Social Media, 2018.