Detection of Cyberbullying on Social Media by Using Machine Learning

G TEJASRI[1], B. MURALI[2]
[1]PG Student, CSE, Quba College of Engineering & Technology
[2]Associate professor, CSE, Quba College of Engineering & Technology

Abstract*—Cyberbullying is a major problem encountered on internet that affects teenagers and also adults.It has lead to mishappenings like suicide and depression.Regulation of content on Social media platorms has become a growing need. The following study uses data from two different forms of cyberbullying, hate speech tweets from Twittter and comments based on personal attacks from Wikipedia forums to build a model based on detection of Cyberbullying in text data using Natural Language Processing and Machine learning. Three methods for Feature extraction and four classifiers are studied to outline the best approach. For Tweet data the model provides accuracies above 90% and for Wikipedia data it gives accuracies above 80%.. Researches on Cyberbullying Incidents show that 11.4% of 720 young peoples surveyed in the NCT DELHI were victims of cyberbullying in a 2018 survey by Child Right and You, an NGO in India, and almost half of them did not even mention it to their teachers, parents or guardians. From this project we hope to reduce the amount of cyberattacks across the globe. This paper mainly works on algorithms proposed by famous researchers and educators.

***Index Terms*— Cyberbullying, Internet Safety, Suicide Prevention, Depression Prevention, Social Media Regulation, Hate Speech Detection**

I. INTRODUCTION

Now more than ever technology has become an integral part of our life. With the evolution of the internet. Social media is trending these days. But as all the other things misusers will pop out sometimes late sometime early but there will be for sure.Now Cyberbullying is common these days. Sites for social networking are excellent tools for communication within individuals. Use of social networking has become widespread over the years, though, in general people find immoral and unethical ways of negative stuff. We see this happening between teens or sometimes between young adults. One of the negative stuffs they do is bullying each other over the internet.In online environment we cannot easily said that whether someone is saying something just for fun or there may be other intention of him. Often,with just a joke,"or don't take it so seriously," they'll laugh it off.Cyberbullying is the use of technology to harass, threaten, embarrass, or target another person. Often this internet fight results into real life threats for some individual. Some people have turned to suicide. It is necessary to stop such activities at the beginning. Any actions could be taken to avoid this for example if an individual's tweet/post is found offensive then maybe his/her account can be terminated or suspended for a particular period.

what is cyberbullying?? Cyberbullying is harassment, threatening, embarrassing or targeting someone for the purpose of having fun or even by well-planned means Researches on Cyberbullying Incidents show that 11.4% of 720 young peoples surveyed in the NCT DELHI were victims of cyberbullying in a 2018 survey by Child Right and You, an NGO in India, and almost half of them did not even mention it to their teachers, parents or guardians. 22.8% aged 13-18 who used the internet for around 3 hours a day were vulnerable to Cyberbullying while 28% of people who use internet more than 4 hours a day were victims. There are so many other reports suggested us that the impact of Cyberbullying is affecting badly the peoples and children between age of 13 to 20 face so many difficulties in terms of health, mental fitness and their decision making capability in any work. Researchers suggest that every country should have to take this matter seriously and try to find solution. In 2016 an incident called Blue Whale Challenge led to lots of child suicides in Russia and other countries . It was a game that spread over different social networks and it was a relationship between an administrator and a participant. For fifty days certain tasks are given to participants . Initially they are easy like waking up at 4:30 AM or watching a horror movie . But later they

escalated to self harm which let to suicides. The administrators were found later to be children between ages 12-14.

II.LITERATURE SURVEY

An approach using keyword matching, opinion mining and social network analysis and got a precision of 0.79 and recall of 0.71 from datasets from four websites.

AUTHOR: Hsien

DESCRIPTION: The approach utilizes a multifaceted strategy incorporating keyword matching, opinion mining, and social network analysis to glean insights from datasets sourced from four distinct websites. Keyword matching serves as the initial filter, identifying relevant words or phrases indicative of pertinent information within the datasets. Opinion mining, or sentiment analysis, delves deeper, scrutinizing text to discern the prevailing sentiments or attitudes expressed towards specific subjects.

A hypothesis that a troll (one who cyberbullies) on a social networking sites.

AUTHOR: Patxi Gal´an-Garc´ıa et al DESCRIPTION: Trolls are prevalent on social networking sites and contribute significantly to cyberbullying behavior. Specifically, it is hypothesized that individuals exhibiting troll-like behavior, characterized by deliberately inflammatory or offensive comments, frequent engagement in online arguments, and a lack of empathy towards others, play a central role in perpetuating cyberbullying dynamics on social media platforms. The anonymity and distance afforded by online interactions may embolden trolls to engage in harmful behavior that they might not exhibit in face-to-face interactions.

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM: Mangaonkar et al. [3] proposed a collaborative detection method where there are multiple detection nodes connected to each other where each nodes uses either different or same algorithm and data and results were combined to produce results. P. Zhou et al.[4] suggested a B-LSTM technique based on concentration. Banerjee et al.[5]. used KNN with new embeddings to get an precision of 70%. DISADVANTAGES OF EXISTING SYSTEM:

- Accuracy is Low.
- Only using Data Mining algorithms in existing methods.

3.2 PROPOSED SYSTEM: For banks has become very difficult for detecting the fraud in bank payments. Machine learning plays a vital role for detecting the financial fraud in the transactions. For predicting these transactions in the proposed The proposed device is developing cyberbullying prediction fashions is to use a textual content classification strategy that entails the building of desktop gaining knowledge of classifiers from labeled textual content instances. Another potential is to use a lexicon-based mannequin that includes computing orientation for a record from the semantic orientation of phrases or phrases in the document. Generally, the lexicon in lexicon-based fashions can be developed manually or mechanically by way of the use of seed phrases to make bigger the listing of words. However, cyberbullying prediction the use of the lexicon-based method is uncommon in literature.

3.3.SYSTEM REQUIREMENTS

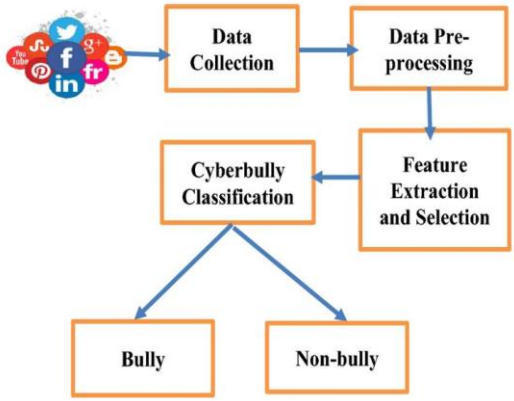
3.3.1.HARDWARE REQUIREMENTS(minimum):

- System : Pentium IV 2.4 GHz
- Hard Disk : 40 GB
- Ram : 512 Mb. 3.3.2.

SOFTWARE REQUIREMENTS:

- Operating System: Windows
- Coding Language: Python 3.7

IV. SYSTEM ARCHITECTURE

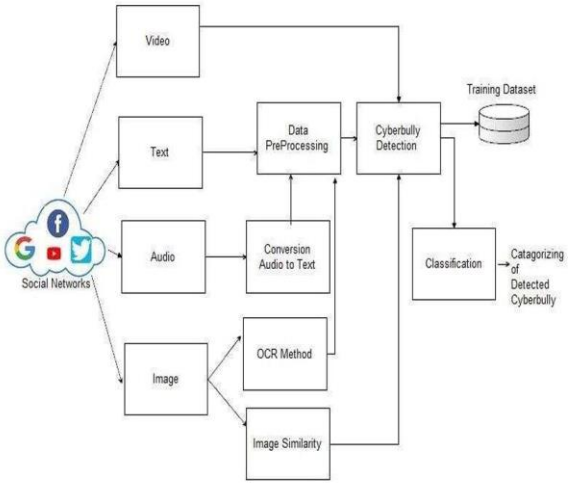


V. SYSTEM DESIGN

DATA FLOW DIAGRAM:

The data flow for detecting cyberbullying on social media using machine learning involves several

interconnected modules that process and manipulate data at different stages of the detection pipeline. Initially, data is sourced from social media platforms such as Twitter, Facebook, or Instagram, where users post content. This raw data is then collected using various techniques, including APIs provided by the platforms or web scraping methods. Upon collection, the data undergoes preprocessing, where it is cleaned of noise, tokenized into individual words or tokens, and standardized through processes like removing stopwords and normalization. Subsequently, the preprocessed data is labeled to identify instances of cyberbullying behavior, either manually or through crowdsourcing platforms, marking each instance as cyberbullying or non-cyberbullying.



VI. SOFTWARE ENVIRONMENT

What is Machine Learning : Before we take a look at the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of building models of data. Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models tunable parameters that can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain. Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here. Machine learning is used in many different applications, from image and speech recognition to natural language processing, recommendation systems, fraud detection, portfolio optimization, automated task, and so on. Machine learning models are also used to power autonomous vehicles, drones, and robots, making them more intelligent and adaptable to changing environments.

VII. SYSTEM IMPLEMENTATION

Sample code:

```
import numpy as np

import pandas as pd

from flask import Flask, request, jsonify, render_template, redirect, flash, send_file

from sklearn.preprocessing import MinMaxScaler

from werkzeug.utils import secure_filename

import pickle

import numpy as np

import pandas as pd

from flask import Flask, request, jsonify, render_template, redirect, flash, send_file

from sklearn.preprocessing import MinMaxScaler

from werkzeug.utils import secure_filename

import pickle

import numpy as np

import pandas as pd

from sklearn.tree import DecisionTreeClassifier

from sklearn.svm import SVC

app = Flask(__name__) #Initialize the flask App

fraud = pickle.load(open('fraud.pkl','rb'))

@app.route('/')
```

```
@app.route('/first')

def first():

return render_template('first.html')
```

VIII SYSTEM TESTING

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results. Integration testing Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional Test Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items:

Valid Input : identified classes of application outputs must be exercised.

Input: : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of valid input must be accepted.

Systems/Procedures: interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

IX SCREENSHOTS

HOME PAGE:

Fig 9.1 : Home Page

PROCESSING LINK PAGE:

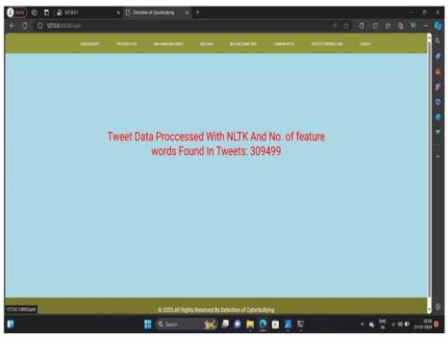


Fig 9.4: Processing Link Page

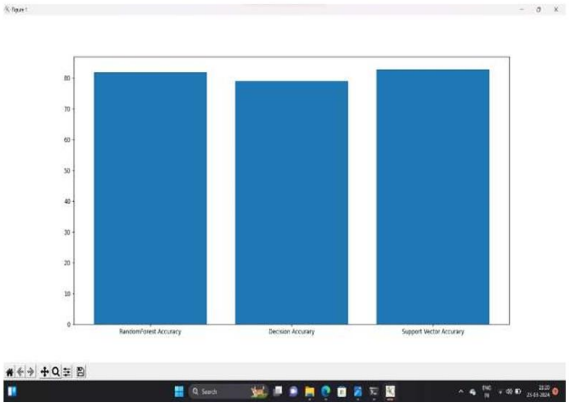


Fig 9.8 : Comparison Page

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 avaibility 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 to take 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 of over 90 percent using basic Machine learning algorithms because tweets containing Hate speech consisted of profanity which made it easily detectable. Due to this it gives better results with BoW and Tf-Idf models rather than Word2Vec models However, Personal attacks were difficult to detect through the same model because the comments generally did not use any common sentiment that could be learned however the three feature selection methods performed similarly.Word2Vec models that use context of features proved effective in both datasets giving similar results in comparatively less features when combined with Multi Layered Perceptrons. As seen by changing nature.

REFERENCES

[1] I. 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 Socio Cultural Computing, BESC 2017, 2017, vol. 2018-January, doi: 10.1109/BESC.2017.8256403.

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

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

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

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

[6] 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: