iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38701

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Cyber Bullying Detection for Twitter Using ML Classification Algorithms

Muskan Patidar[1], Mahak Lathi[2], Manali Jain[3], Monika Dhakad[4], Prof. Yamini Barge[5]
[1, 2, 3, 4]*Student,* [5]*Assistant Professor, Department of Computer Science Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India.*

Abstract: ***Social networking platforms have given us incalculable opportunities than ever before, and its benefits are undeniable. Despite benefits, people may be humiliated, insulted, bullied, and harassed by anonymous users, strangers, or peers. Cyberbullying refers to the use of technology to humiliate and slander other people. It takes form of hate messages sent through social media and emails. With the exponential increase of social media users, cyberbullying has been emerged as a form of bullying through electronic messages. We have tried to propose a possible solution for the above problem, our project aims to detect cyberbullying in tweets using ML Classification algorithms like Naïve Bayes, KNN, Decision Tree, Random Forest, Support Vector etc. and also we will apply the NLTK (Natural language toolkit) which consist of bigram, trigram, n-gram and unigram on Naïve Bayes to check its accuracy. Finally, we will compare the results of proposed and baseline features with other machine learning algorithms. Findings of the comparison indicate the significance of the proposed features in cyberbullying detection.***

Keywords: Cyber bullying, Machine Learning Algorithms, Twitter, Natural Language Toolkit

I. INTRODUCTION

Cyberbullying can be defined as an aggressive or intentionally carried out harassment by group or individual through digital means repeatedly against a sufferer who is unable to defend themselves. This type of bullying includes threats, abusive or sexual remarks, rumours and hate speech. Cyberbullying is an ethical issue found on internet and the percentage of the victims is also alarming.

A. Cyberbullying on Social Media Sites

The major contributors to cyberbullying are social networking sites. The dynamic nature of these sites helps in the growth of online aggressive behaviour. The anonymous feature of user profiles increases the complexity to identify the bully. Social media is popular due to its connectivity in the form of networks. But this can be harmful when rumours or bullying posts are spread into the network which cannot be easily controlled. Twitter and Facebook can be taken as examples which are popular among various social media sites. According to Facebook users have more than 150 billion connections which gives the idea about how bullying content can be spread within the network in a fraction of time. To manually identify these bullying messages over this huge network is difficult. There should be an automated system where such kinds of things can be detected automatically thereby taking appropriate action. The victims mainly consist of women and teenagers. Intense effect on mental and physical health of the victims in such kind of activities higher' s the risk of depression leading to suicidal cases. Therefore to control cyberbullying there is need of automatic detection or monitoring systems.

B. Detection Models for Cyberbullying

Many works in this field have shown that machine algorithms can be used for predicting and detecting cyberbullying actions. As tremendous data is generated every second algorithms can be trained efficiently. Machine learning classifiers help to classify the content of the texts into non - bullying and bullying classes. After classification of the content the bullying ones can be stopped. This project aims to detect cyberbullying in tweets using ML Classification Algorithms. A training and predicting pipeline is implemented to contrast performance of various popular classification algorithms and determine the best suited model. Extracted features will be applied with Naïve Bayes, KNN, Decision Tree, Random Forest, Support Vector Machine algorithm etc. and also we will apply the NLTK(Natural language toolkit) which consist of bigram, trigram, n-gram and unigram on Naïve Bayes to check its accuracy. We will compare the algorithm and draw the results, results will indicate that our proposed framework provides a feasible solution to detect cyberbullying behavior and its severity in online social networks.

The motivation behind this project is to prevent teenagers from getting depressed or committing suicide due to cyber bullying activities and also to decrease the harassing incidents in cyberspace.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

II. PROBLEM FORMULATION

The social media network gives us to great communication platform opportunities they also increase the vulnerability of young people to threatening situations online. Cyberbullying on an social media network is a global phenomenon because of its huge volumes of active users. The trend shows that the cyber bullying on social network is growing rapidly every day. Recent studies report that cyberbullying constitutes a growing problem among youngsters. Successful prevention depends on the adequate detection of potentially harmful messages and the information overload on the Web requires intelligent systems to identify potential risks automatically. So, In this project we focus on to make a model on automatic cyberbullying detection in twitter by modelling posts written by bullies on twitter. We have tried to propose a possible solution for the above problem, our project aims to detect cyberbullying in tweets using ML algorithms like Naïve Bayes, KNN, Decision Tree, Random Forest, Support Vector etc. and also we will apply the NLTK(Natural language toolkit) which consist of bigram, trigram, n-gram and unigram on Naïve Bayes to check its accuracy. We will train the model using different algorithms and compare the results and check whether the model is able to provide the desired results or not that is whether it is able to predict the text is bullied or not.

III.LITERATURE REVIEW

Title	Problem	Solution	Result
An Effective Approach for Cyberbullying Detection and avoidance	The biggest problem regarding cyberbullying is that the age group of the offenders range from as young as eight to the legal adult age of eighteen and beyond. Once happen this activity then victims are often left permanently then difficult to find them.	In this paper focused on the issues of robust system and objectives are 1) Automatic detection and avoidance of cyberbully attack in internet. 2) Effective age authentication for website browsing and categorizing the links based on age	Represented a novel method on the current scenario of cyber-bullying and various methods available for the detection and prevention of cyber harassment. Our concept depends upon the text analysis, the data which is uploaded or text written by any user is first analyzed.
Using Machine Learning to Detect Cyberbullying	Teens and young adults, are finding new ways to bully one another over the Internet. in a study conducted by Symantec reported that, to their knowledge, their child has been involved in a cyberbullying incident	Used machine learning algorithm to detect cyberbullying. For training the data downloaded from website. The data was labeled using a web service. the labeled data, in conjunction with machine learning techniques provided by the Weka tool kit, to train a computer to recognize bullying content	Used a language-based method of detecting cyberbullying. By recording the percentage of curse and insult words within a post.
Cyberbullying Detection System on Twitter	Increased cyberbullying attacks on the social network services. To prevent these activities a system was proposed.	In this system, the users can identify the cyberbullying related tweets based on the keywords and populate it in a news feed form. By doing this, it allows users to determine the identities of the cyberbullies and the victims from the cyberbullying tweets.	With the advent of this cyberbullying detection and solution system in Twitter, it will help the authorities to monitor, regulate or at least decrease the harassing incidents in cyberspace

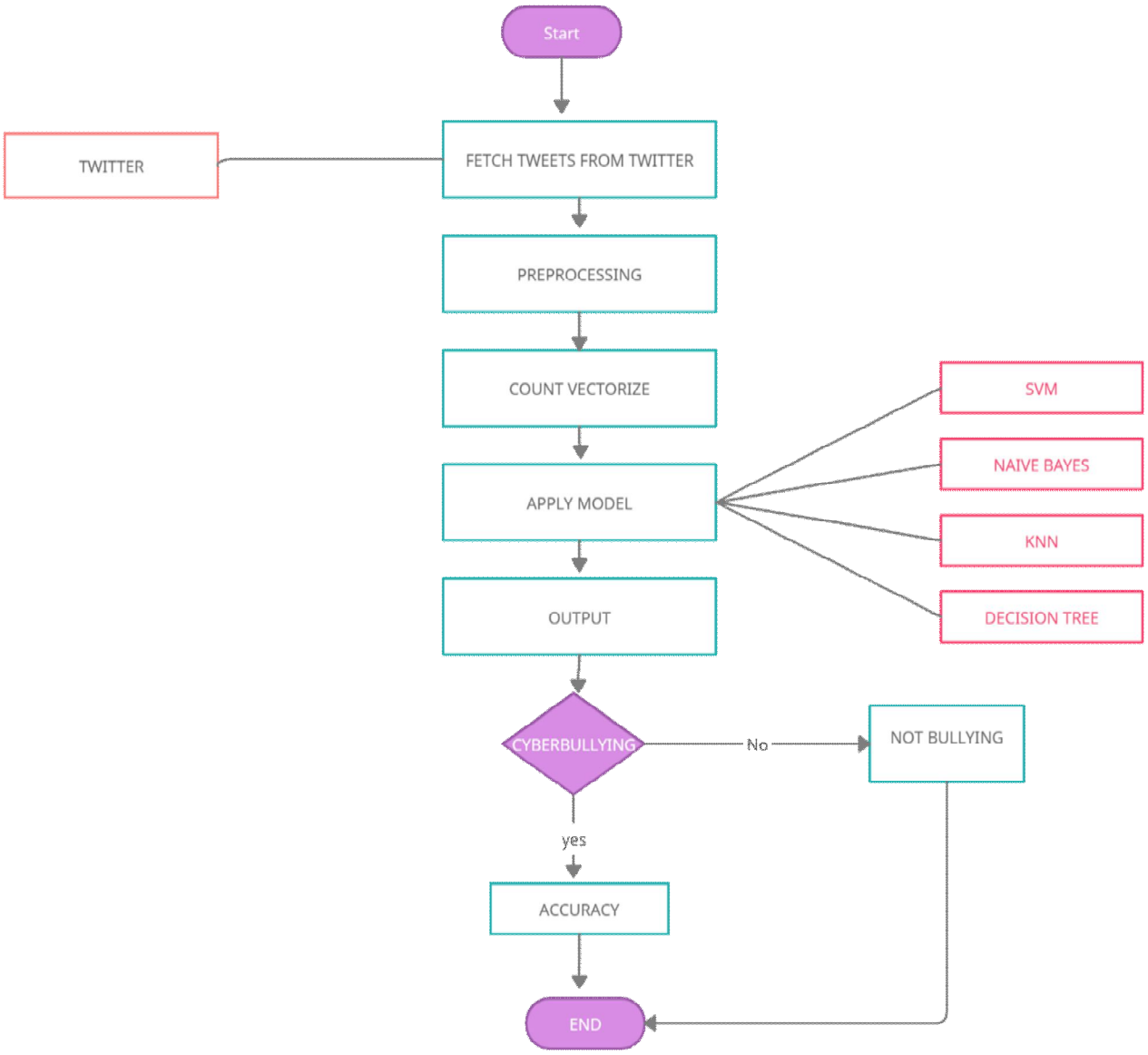
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

IV. METHODOLOGY

This project will be developed using python, ML and web technology.

A. First we will search and find the dataset from Kaggle or Github and download it to train the model.
B. After downloading we will pre-process the data, clean the data then with the help of naïve bayes, SVM (Support vector machine), KNN and Random Forest we will train the dataset and generate models separately and also we will apply the NLTK(Natural language toolkit) which consist of bigram, trigram, n-gram and unigram on Naïve Bayes to check its accuracy.
C. We will fetch the real time tweets from twitter and then we apply the generated model to these fetched tweets and check if the text is cyberbullying or not, also we will compare different algorithms and check the accuracy of various models and select the best model.
D. For the frontend purpose we will create a platform where there will be different logins for user and admin and the user will be able to post the tweet and the admin will be able to view the tweets and classify whether the tweet is bullied or not.
E. These all-purpose we are using python as backend, for training the model we are using ML algorithms.

The flow chart depicting the flow of our system is-



The dataset we collected for training the model is-

Pre-processing was carried out by first changing all the uppercase words to lowercase. The next step was to remove all the punctuation marks and emojis as they are not required for our purpose. The next important step is to remove the stop words. They are the common words which are not useful in detection like “a”, “on” , “all” These words do not carry important meaning and are usually removed from texts. A text file containing bad words was created. If the tweet has the words present in the bad words then it is labelled as “TRUE” else labelled as “FALSE”. This final dataset contains the tweets and classification which is now used for algorithm implementation.

6	.Am I the only one who thinks justin bieber is fugly af now?	Non-Bullying
7	We carry on? We as in fugly lookin unwanted people?	Non-Bullying
8	Don?t know what?s worse, the fact he?s hoying milk in before the water or his fat fugly face at the end of that vid.	Non-Bullying
9	Enjoy your New Smyrna Beach..full of seaweed, shallow fugly green water and sharks. Nothing comp.	Non-Bullying
10	Yeah honestly that?s fugly.	Non-Bullying
11	No one wants to see those fugly 3T bikes?	Non-Bullying
12	This luxury bldg they?re putting up across from .thepinhook is so fugly. I would still be mad even if it were aesthetically pleasing but ew.	Non-Bullying
13	Everyone else was in the line up except for her and they got to debut together while she gad to wear a fugly purple.	Non-Bullying
14	Fell for that again didnt ya XD #PhoneCallTrick.	Non-Bullying
15	fugly Am alright starting to get a headache myself.	Non-Bullying
16	Fugly?. Whom?. You two? Henny.....	Non-Bullying
17	o wow conq varus actually looks good unlike fucking horrifically fugly conq karma.	Non-Bullying
18	I'm fed up with this migraine.","How's you	Non-Bullying
19	King Show off.	Non-Bullying
20	if anyone references talking about aliens when explaining aquarius they are fraudulent.', 'do not trust them, they are a fugly.	Non-Bullying
21	She's fugly.	Non-Bullying
22	.I am fugly crying!', 'I can?t handle this!', 'This was so beautiful!', 'Euphoria is such a bop too.', '????.	Non-Bullying
23	i literally only know one person who uses the word fugly.', 'lol fuck off anna..	Non-Bullying
24	.Fugly ass.	Non-Bullying
25	.today i am fatish and fugly.	Non-Bullying
26	The ?fugly? friend.', 'Funny &; ugly.	Non-Bullying

Our proposed framework provides a feasible solution to detect cyberbullying behaviour and its severity in online social networks like twitter. After completion of the project we will get a system which would successfully detect the offensive text or comments or the cyberbullied statements on Twitter. The system will help the society especially the teenagers and adults for preventing them from further bullying.

V. RESULT DISCUSSIONS

The system is successfully able to predict whether the content is bullied or not. We have achieved the following output-

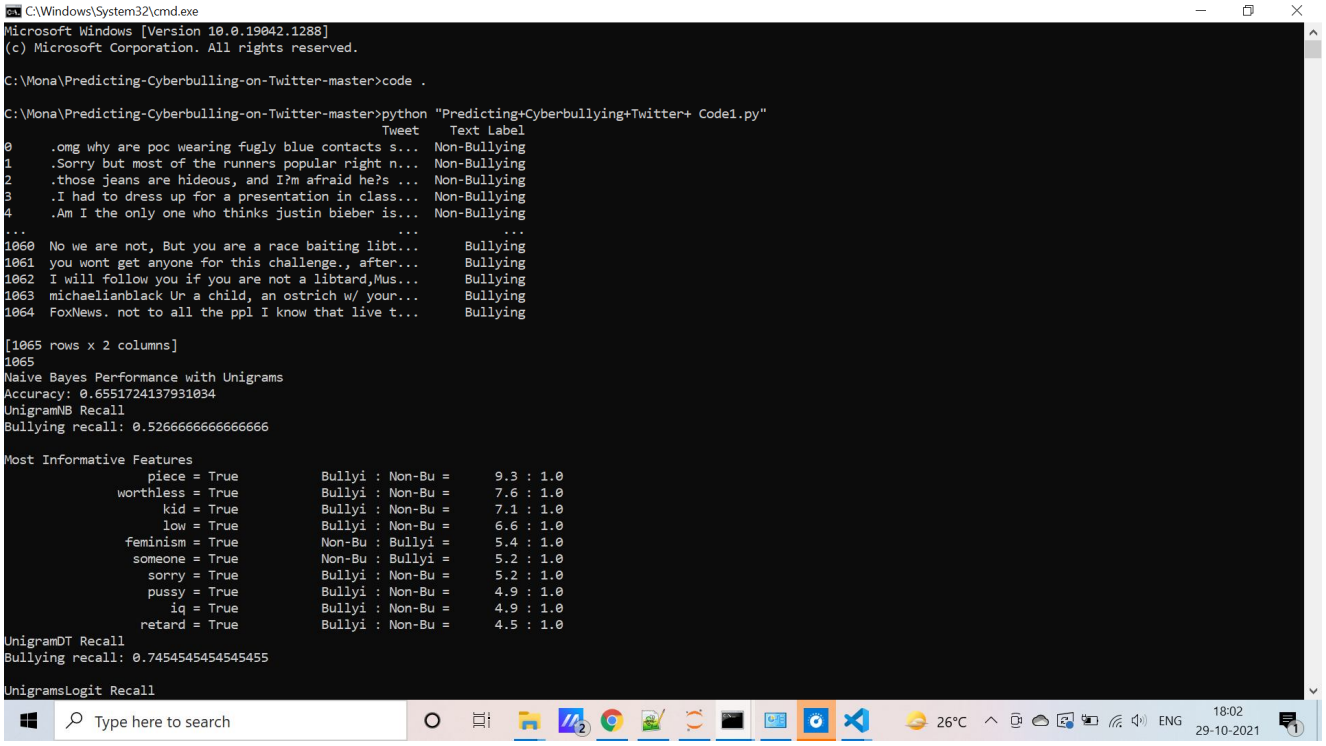


Fig – Depicts the statement identified as bullied or non-bullied

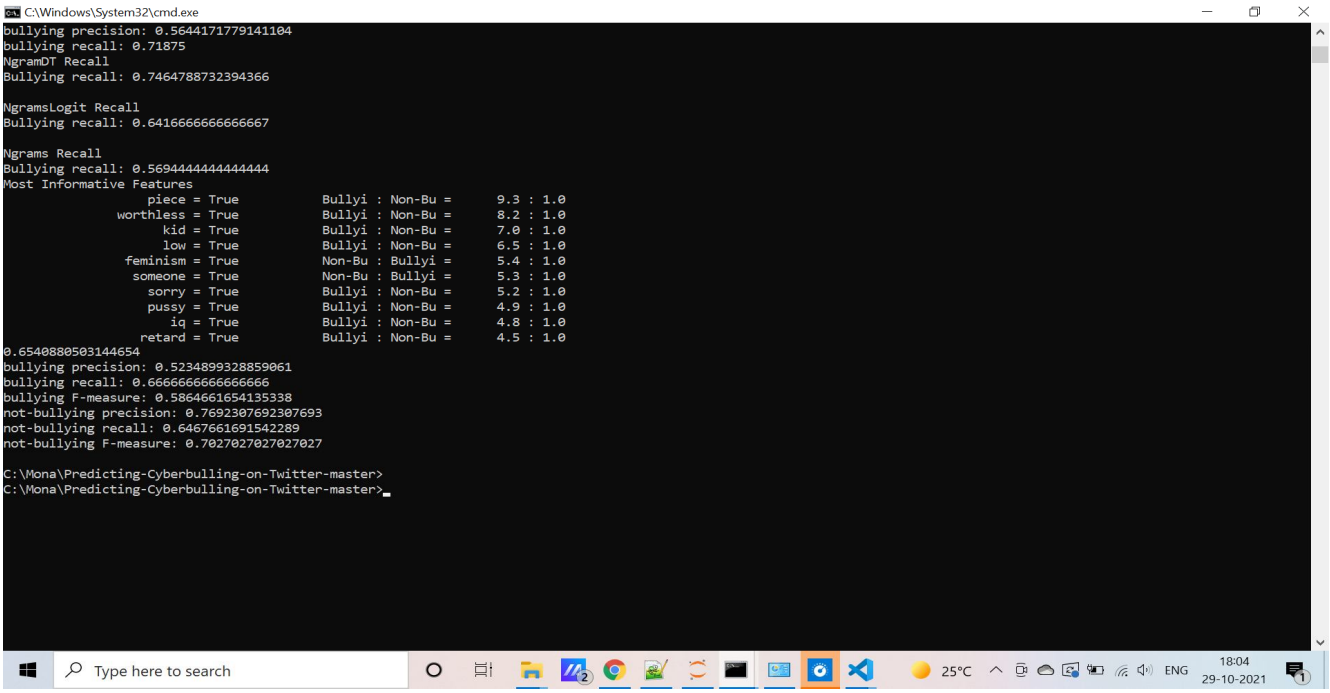


Fig- Depicts the precision and Accuracy

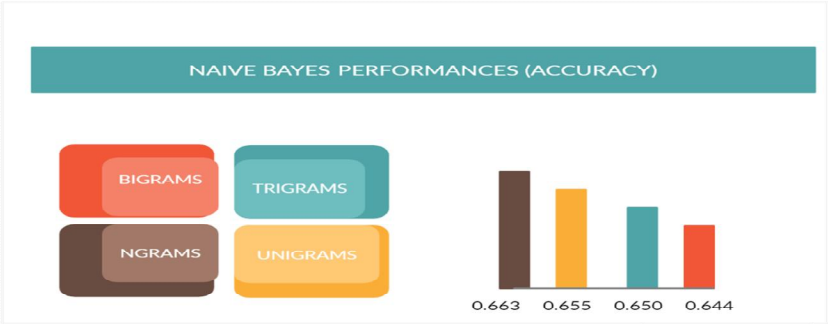
Table 1- Shows the Naïve Bayes Algorithm Accuracy with Unigram, Bigram, Trigram, N-gram

Model	Accuracy
Naïve Bayes Unigram	0.6551
Naïve Bayes Bigram	0.6446
Naïve Bayes Trigram	0.6509
Naïve Bayes N-Gram	0.6634

Table 2- Shows Bullying and Non-Bullying Precision and Recall

Text	Precision	Recall
Bullying	0.5234	0.6666
Non-Bullying	0.7692	0.6467

Graph- Shows the Naïve Bayes Algorithm Accuracy with Unigram, Bigram, Trigram, N-gram



VI. CONCLUSION

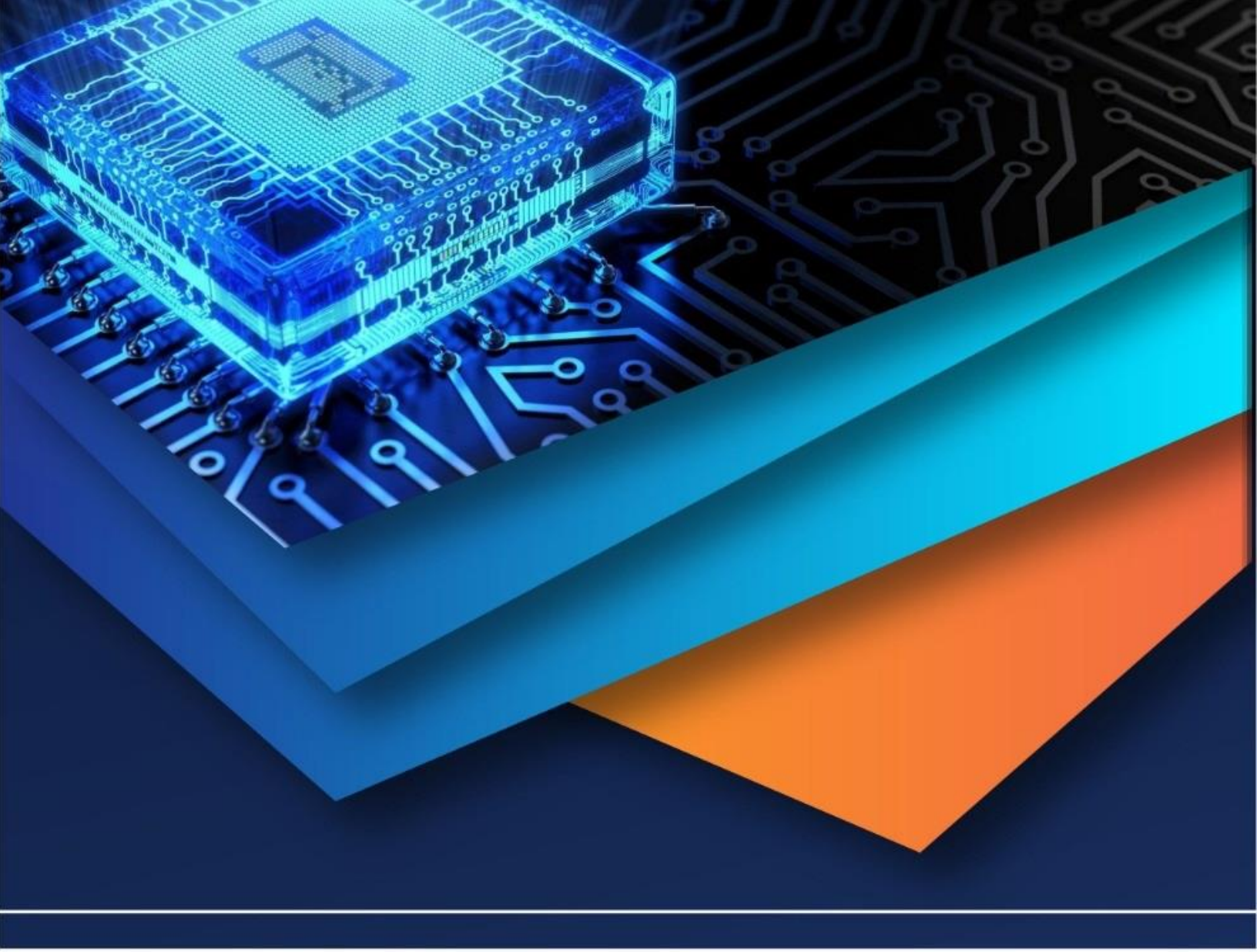
The study reviewed the existing literature for various machine learning algorithms and identified Naïve bayes N-gram gives the best accuracy and also the system is able to identify the bullied and non-bullied statements. The goal of this project is to the automatic detection of cyberbullying-related posts on Twitter. Automatic detection of signals of cyberbullying would enhance moderation and allow them to respond quickly when necessary. However, these posts could just as well indicate that cyberbullying is going on. The main aim of this project is that it presents a system to automatically detect signals of cyberbullying on social media handle Twitter, including different types of cyberbullying, covering posts from bullies, victims and bystanders.

VII. ACKNOWLEDGEMENT

We as the authors would like to extend a special thanks of vote to the reviewers of this paper for their valuable suggestions to improve this paper. The paper is supported by Acropolis Institute of Technology and Research, Indore (M.P.)

REFERENCES

[1] Poeter. (2011) Study: A Quarter of Parents Say Their Child Involved in Cyberbullying. pcmag.com. [Online].Available: http://www.pcmag.com/article2/0,2817,2388540,00.asp
[2] J. W. Patchin and S. Hinduja, “Bullies move Beyond the Schoolyard; a Preliminary Look at Cyberbullying,” Youth Violence and Juvenile Justice, vol. 4, no. 2, pp. 148–169,2006
[3] Anti Defamation League. (2011) Glossary of Cyberbullying Terms.adl.org.[Online].Available:http://www.adl.org/education/curriculum connections/cyberbullying /glossary.pdf
[4] N. E. Willard, Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress. Research Press, 2007.
[5] D. Maher, “Cyberbullying: an Ethnographic Case Study of one Australian Upper Primary School Class,” Youth Studies Australia, vol. 27, no. 4, pp. 50–57, 2008.
[6] https://www.sciencedirect.com/topics/computer-science/deep-neural-network
[7] An Effective Approach for Cyberbullying Detection and avoidance IEEE paper
[8] Approaches to Automated Detection of Cyberbullying: A Survey IEEE paper
[9] Cyberbullying Detection System on Twitter IEEE paper
[10] Methods for Detection of Cyberbullying: A Survey IEEE paper



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY