ANALYSIS & REMOVING OF NEGATIVE

COMMENTS ON WOMEN

Divya Kandukuri1, A Tejasree Reddy2, Dodle Pranay Kumar3, K.Venkat Reddy41,2,3UG Scholar, Department of Information Technology, 4Assistant Professor, Department of Information Technology 1,2,3,4 Vignana

Bharathi Institute of Technology, Ghatkesar, India

1divyakandukuri86@gmail.com, 2tejasreeanugu10@gmail.com, 3pranayraos9@gmail.com

**ABSTRACT** Support, Decision Tree, and Random Forest in the proposed cyberbullying detection.

The use of social media has grown model. We conduct experiments

with two exponentially over time with the growth of the Internet and has become the most influential networking platform in the 21st century. However, the enhancement of social connectivity often creates negative impacts on society that contribute to a couple of bad phenomena such as online abuse, harassment cyberbullying, cybercrime and online trolling.

Cyberbullying frequently leads to serious mental and physical distress, particularly for women and children, and even sometimes force them to attempt suicide. Online harassment attracts attention due to its strong negative social impact. Many incidents have recently occurred worldwide due to online harassment, such as sharing private chats, rumours, and sexual remarks. Therefore, the identification of bullying text or message on social media has gained a growing amount of attention among researchers. The purpose of this research is to design and develop an effective technique to detect online abusive and bullying messages by merging natural language processing and machine learning. Two distinct freatures, namely Bag-of - Words (BoW) and term frequency-inverse text frequency

# OBJECTIVE

In this context, we suggest a cyberbullying detection model based on machine learning that can detect whether a text relates to cyberbullying or not. We have investigated several machine learning algorithms, including Naive Bayes, Vector Machines for datasets from twitter and Facebook’s comments and posts. For performance analysis, we use two different feature vectors BoW and TF-IDF. The results indicate that TF-IDF feature provides better accuracy than BoW where Randomforest provides better performance than any other machine learning algorithms used.

# INTRODUCTION

Social media is a platform that allows people to post anything like photos, videos, documents extensively and interact with society [1]. People connect with social media using their computers or smartphones. The most popular social media includes Facebook1, Twitter2, Instagram3, TikTok4 and so on. Nowadays, social media is involved in different sectors like education [2], business [3], and also for the noble cause [4]. Social media is also enhancing the world’s economy through creating many new job opportunities [5]. Although social media has a lot of benefits, it also has some drawbacks. Using this media, malevolent users conduct unethical and fraudulent acts to hurt others feelings and damage their reputation. Recently, cyberbullying has been one of the major social media issues. Cyberbullying or cyber-harassment refers to an electronic method of bullying or harassment. Cyberbullyingand cyberharassment are also known online bullying. As the digital realm has grown and technology has progressed, cyberbullying has become relatively common, particularly amongst adolescents.

# LITERATURE SURVEY

**Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets**

 **Authors:** Deepak Kumar1, Shivani

Aggarwal2

 **Published online:** September 2022

There are certain types of harassment and Violence that are very aggressive including staring and passing comments and these unacceptable practices are usually seen as a normal part of the urban life. There have been several studies that have been conducted in cities across India and women report similar type of sexual harassment and passing off comments by other unknown people. The study that was conducted across most popular Metropolitan cities of India including Delhi, Mumbai and Pune, it was shown that 60 % of the women feel unsafe while going out to work or while travelling in public transport. Women have the right to the city which means that they can go freely whenever they want whether it be too an Educational Institute, or any other place women want to go. But women feel that they are unsafe in places like malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point Safety or lack of concrete consequences in the life of women is the main reason of harassment of girls. There are instances when the harassment of girls was done by their neighbours while they were on the way to school or there was a lack of safety that created a sense of fear in the minds of small girls who throughout their lifetime suffer due to that one instance that happened in their lives where they were forced to do something unacceptable or was sexually harassed by one of their own neighbours or any other unknown person. Safest cities approach women safety from a perspective of women rights to the affect the city without fear of violence or sexual harassment. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need of protection of women and also recognizes that women and girls also have a right same as men have to be safe in the City. Throughout the research paper we have discussed about various machine learning algorithms that can help us to organize and analyze the huge amount of Twitter data obtained including millions of tweets and text messages shared every day. These machine learning algorithms are very effective and useful when it comes to analyzing of large amount of data including the SPC algorithm and linear algebraic Factor Model approaches.

# PROBLEM STATEMENT

We cannot be thrown in public, communities and opinion. The sense of the analysis of the most important are also useful in a wider range of important social instruments after a decision declaring that it allows them to call us a glimpse into the same subject:. And the ability to extract data from widely adopted in practice given social organizations worldwide.

# EXISTING SYSTEM

Approximately 50% of the teenagers America experience cyberbullying. This bullying has a physical and mental impact on the victim. The victims choose selfdestructive acts like suicide because the trauma of cyberbullying which is hard to be endured. Thus, the identification and prevention of cyberbullying is important to protect teenagers. The existing machine learning application of cyberbullying detection uses decision tree algorithms, but this algorithm is not effectively classifying the cyberbullying messages.

# PROPOSED SYSTEM

In this section, we describe the cyberbullying detection framework which consists of two major parts as shown in 1. The first part is called NLP (Natural Language Processing) and the second part is named as ML (Machine learning). In the first phase, datasets containing bullying texts, messages or post are collected and prepared for the machine learning algorithms using natural language processing. The processed datasets are then used to train the machine learning algorithms for detecting any harassing or bullying message on social media including Facebook and Twitter.

**DATA GATHERING:** The dataset

represented here is a collection of tweets which was collected using Twitter API. The number of data entries exceeded 1000 tweets which belong to different time periods. The following images depict the datasets indicating Text Labels.

**DATA PREPROCESSING:** Adaptation of

the raw data according to our need is

important before implementing the regression model. Since, raw data is most of the time inconsistent or incomplete or lacking in certain behavior or lacking in attributes or may contain noises. So, we need to remove all these abnormalities and convert the dataset into something which can be used by the machine learning algorithms. So, we processed data obtained from online sources to obtain useful data metrics, related to profanity in the output, on a daily basis which can be used to train our models. The comment data which we downloaded was in xlsx format. So, we had to convert the xlsx file format to csv format which is usual format used to train machine learning models. Further sometimes data contain various inconsistencies such as noisy data which model cannot interpret and valuem dominances of a variable over another which can cause model’s inconsistency to predict accurately.

**TRAINING PHASE:** For training the model, first we import a specific algorithm class/module and create an instance of it. Then using that instance, we fit the model to the training data. Then we validate it by testing its accuracy score and fine tune its parameters till we get required results.

**TESTING PHASE:** For testing the model, we compare its predicted values after the training phase with test data. Then input some different value for prediction and check whether it predicts it right. If it didn’t predict right then, fine tune the algorithmic parameters and fit the model again.It is concluded that the missing data values will result in inconsistency. For best results, and it is given pretreated to improve the efficiency of the algorithm. Outliers should be deleted, and a variable conversion must be performed.

# DESIGN ARCHITECTURE

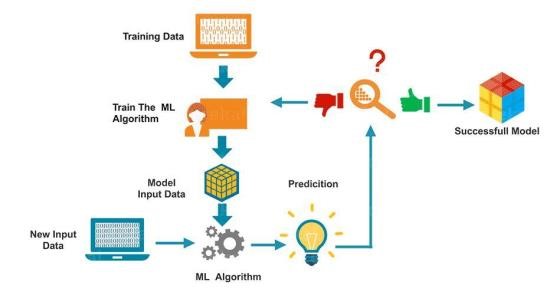


fig.System architecture for machine learning model.

**WORKING PROCESS:**

 Download and install the package to the machine that were useful for learning anaconda thou huge

Python.

 Land understand the game data, and the data structure of the summary statistics and for society.

 Machine learning models to choose the best to be sure that the truth will remain the same.

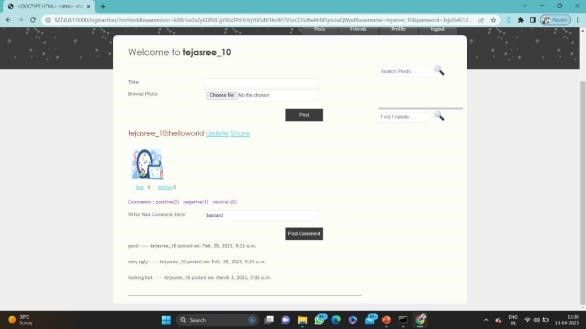
Python is an interpreted language popular and powerful. Unlike A Python is a comprehensive and language development platform and can be used for research and development of production systems. There are also numerous modules to choose from, the laboratory libraries, offering that they were doing the work of a number of the road between

them. This seems to be able to see.

The best way to start using that, too, Python, and it is the doctrine of the machine to finish the project.

* This will need to install Python's prey, interpreted start (at least).
* Does not make you feel you are in a small sail project.
* This will give you confidence, perhaps to pursue your own little projects.

The best way to get new business or a new tool to browse an automatic learning platform will cover the key project from beginning to end, and in stages. The five princes of the Philistines, and the burden of sum up given to the information, evaluate algorithms that predict.

Here is an overview of what we are going to cover:

1. Installing the Python anaconda platform.
2. Loading the dataset.
3. Summarizing the dataset.
4. Visualizing the dataset.
5. Evaluating some algorithms.
6. Making some predictions.

**MODULES:**

* Twitter data for Sentiment

Analysis(module-01)

* Data validation and Preprocessing technique (Module-02)
* Performance measurements of

logistic regression and dec tree

* Performance measurements of
* Support vector classify Random forest (Module-04)
* GUI based sentiment analysis (Module-5)

**RESULTS:**

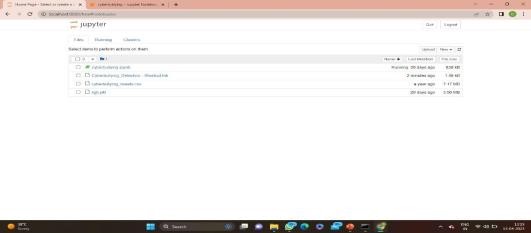


Fig: Open the anaconda navigator.

Then launch to jupyter notebook.

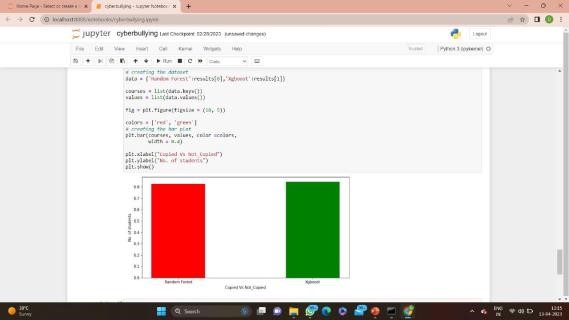


Fig: Open the corresponding folder in jupyter notebook. Run the file then as a result a bar graph will be shown.

Fig: A twitter platform is shown where comments can be posted.

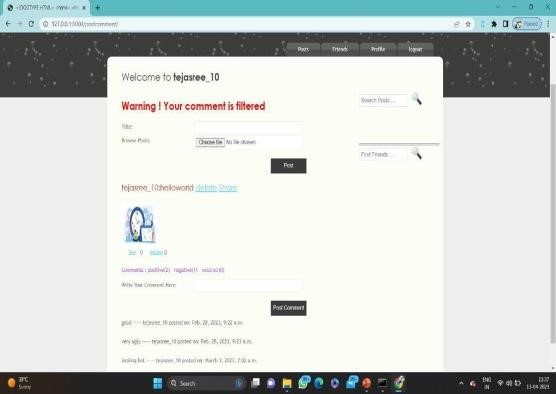


Fig: Here all negative Comments will be Removed.

**CONCLUSION:**

In particular, cyber bullying has become more common and has begun to raise significant social issues with the rising prevalence of social media sites and increased social media use by teenagers. There needs to design automatic cyberbullying detection method to avoid bad consequences of cyber harassment. Considering the significance of cyberbullying detection, in this study, we investigated the automated identification of posts on social media related cyberbullying by considering two features BoW and TFIDF. Four machine learning algorithms are used to identify bullying text and SVM for both BoW and TF-IDF. In future we are planning to design a framework for automatic detection and classification of cyberbullying from Bengali texts using deep learning algorithms.

**FUTURE WORK:**

* To automate this processes by show the result in desktop application.
* To optimize the work to implement in Artificial
* Intelligence environment.

The task of sentiment analysis, especially in the domain of micro-bloging, is still in the developing stage and far from complete. So we propose a couple of ideas which we feel are worth exploring in the future and may result in further improved performance. Right now we have worked with only the very simplest unigram models; we can improve those models by adding extra information like closeness of the word with a negation word. We could specify a window prior to the word (a window could for example be of 2 or 3 words) under consideration and the effect of negation may be incorporated into the model if it lies within that window. The closer the negation word is to the unigram word whose prior polarity is to be calculated, the more it should affect the polarity. For example if the negation is right next to the word, it may simply reverse the polarity of that word and farther the negation is from the word the more minimized ifs effect should be.

**REFERENCES:**

1. Ong Kok Chien, Poo Kuan Hoong,and Chiung Ching Ho. A comparative study of hits vs pagerank algorithms for twitter user’s analysis. In 2014 International Conference on Computational Science and Technology (ICCST), pages 1–6.

IEEE, 2014.

1. Adam G Hart, William S Carpenter, Estelle Hlustik-Smith, Matt Reed, and Anne E Goodenough. Testing the potential of twitter mining methods for data acquisition: Evaluating novel opportunities for ecological research in multiple taxa. Methods in Ecology and Evolution, 9(11):2194–2205, 2018.
2. Ari Z Klein, Abeed Sarker, Haitao Cai, Davy Weissenbacher, and Graciela Gonzalez-Hernandez. Social media mining for birth defects research: a rule-based, bootstrapping approach to collecting data for rare health-related events on twitter. Journal of biomedical informatics, 87:68– 78, 2018.
3. Jeremy YL Yap, Chiung Ching Ho, and Choo-Yee Ting. Analytichierarchy process (ahp) for business site selection. In AIP Conference Proceedings, volume 2016, page 020151. AIP Publishing, 2018. 5. Hui-Jia Yee, Choo-Yee Ting, and Chiung Ching Ho. Optimal geospatial features for sales analytics. In AIP Conference Proceedings, volume 2016, page 020152. AIP Publishing, 2018.

6. Choo-Yee Ting, Chiung Ching Ho, Hui Jia Yee, and Wan Razali Matsah.Geospa- tial analytics in retail site selection and sales prediction. Big data, 6(1):42–52, 2018. 7. Amita Jain and Minni Jain. Location based twitter opinion mining using common-sense information. Global Journal of Enterprise Information System, 9(2), 2017.

1. Prabhsimran Singh, Ravinder Singh Sawhney, and Karanjeet Singh Kahlon. Sentiment analysis of demonetization of 500

&1000rupee banknotes by indian government. ICT Express, 4(3):124–129, 2018.

1. Hyo Jin Do and Ho-Jin Choi. Sentiment analysis of real-life situations using loca- tion, people and time as contextual features. In 2015 International Conference on Big Data and Smart Computing (BIGCOMP), pages 39–42. IEEE, 2015. 10. mXin Zheng, Jialong Han, and Aixin Sun. A survey of location prediction on twitter. IEEE Transactions on Knowledge and Data Engineering, 30(9):1652–1671, 2018.

11. mShiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng. Cnn for situations understanding based on sentiment analysis of twitter data. Procedia computer science,111:376–381, 2017. 12. Yuebing Zhang, Zhifei Zhang, Duoqian Miao, and Jiaqi Wang. Three-way en- hanced convolutionalneural networks for sentence-levelsentiment classification. Information Sciences, 477:55–64, 2019.

13.S. Agrawal and A. Awekar, “Deep learning for detecting cyberbullying across multiple social media platforms,” in European Conference on Information

Retrieval. Springer, 2018, pp. 141–153.

14.P. Badjatiya, S. Gupta, M. Gupta, and V.

Varma, “Deep learning for hate speech detection in tweets,” in Proceedings of the 26th International Conference on World Wide Web Companion, 2017, pp. 759–760.

15.M. A. Al-Ajlan and M. Ykhlef, “Deep learning algorithm for cyberbullying detection,” International Journal of Advanced Computer Science and

Applications, vol. 9, no. 9, 2018.