



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: VI Month of publication: June 2022

DOI: <https://doi.org/10.22214/ijraset.2022.44157>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Cyber Bullying Text Detection Using Machine Learning

A. Varsha Reddy¹, Gugulothu Kalpana², N. Satish Kumar³, Dr. Sundaragiri Dheeraj⁴

^{1, 2, 3} Students of Department of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Hyderabad

⁴ Associate Professor, Department of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Hyderabad

Abstract: With the upward thrust of the Internet, using social media has exploded, and it's emerged because the most powerful networking platform of the ordinal century. However, exaggerated social networking oftentimes has negative consequences for society, causative to some unwanted phenomena at the side of online abuse, harassment, cyberbullying, cybercrime, and trolling. Cyberbullying causes severe mental and physical distress in several people, particularly ladies and children, and might even cause suicide damaging social impact of online harassment attracts attention. Several incidences of online harassment, equivalent to sharing personal chats, spreading rumours, and creating sexual remarks, have recently occurred everywhere on the planet. As a result, specialists are paying nearer interest to detect bullying the big texts or messages on social media. By combining natural language processing and machine learning the aim of this observation is to create and construct a powerful method for detecting online abusive and bullying texts. The accuracy stage of six different machine learning techniques is evaluated the usage by of extraordinary features, particularly the count vectorizer.

Keywords: Cyber bullying, Machine learning, social media, Dataset

I. INTRODUCTION

Although cyber bullying differs from conventional harassment, it's far nonetheless distressing. The results and dangers are similar to before, if now no longer extra sizeable and widespread. Cyberbullying manifests itself in a number of paperwork at the perfect time. It does now no longer suggest hacking into a person's profile or posing as a person else.

It additionally consists of making unsightly statements approximately a person or spreading rumours to criticize them. Bullying at the net or on social media Bullying is described as any hobby or degree this is used to control, annoy, or stigmatize any other stigmatized heinous act this is obscenely dangerous and might have a sizeable and poor effect on anyone. They ordinarily take location thru on-line media, public gatherings, and different means.

In this regard, we propose a machine-learning-based cyberbullying detection model that can determine if a communication is related to cyberbullying or not. In the proposed cyberbullying detection model, we studied many machine learning methods, including Ada Boost, Support Vector Machine, Logistic Regression, Random Forest, XGBoost, and sgd. We run tests on datasets derived from Twitter and Facebook remarks and posts. We run a separate characteristic vector-like Count Vectorizer in which Logistic regression outperforms all the different devices getting to know strategies provided in this paper. Machine learning is know-primarily based totally cyber-bullying detection has been the problem of diverse studies. To discover the sentiment and contextual factors of a sentence, a supervised machine learning knows method primarily based totally on a count vectorizer was proposed. The accuracy of this approach is 80.01percent.

II. BULLYING DETECTION MODEL

A. Proposed System

The proposed system consists of the following fundamental segments shown in Fig.1in

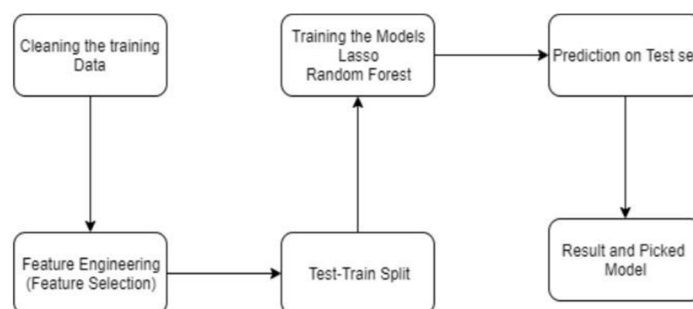


Figure.1:- Architecture diagram

The first component is primarily responsible for cleansing the data set; it can be thought of as a pre-processing phase. Because the info nonheritable might contain unstructured content, it's needed for us to wash or trim the data so as to realize bigger accuracy. Feature extraction is the next step within the model. during this phase, the pre-processed matter data is remodelled into an appropriate format in order to create it helpful for machine learning techniques. Classification is the final step in this process. The training and testing datasets will be examined using a variety of machine learning techniques or algorithms. All of these algorithms' performances will be compared to one another.

B. Implementation

1) *Training and Testing Data Set:* Importing libraries like pandas, NumPy, seaborn, matplotlib, and nltk is done. In this module, we'll concentrate on the dataset we've acquired, first removing all rows with null entries. At that time, we'll remove any features that could jeopardize the accuracy of our algorithm. We'll split the dataset into two pieces here: training and testing. We will use 70% of the dataset to train the model, and the remaining 30% will be used to assess the training model's precision. The information is manually classified as a bully (sexual, threat, troll, or religious) or not-bully. The dataset also includes three more columns that identify the remark category, the gender on which the comment is made, and the total number of reactions for each comment.

2) Data Pre-Processing

	label	untokenized_text	preprocessed_text
0	Non-offensive	As a woman you t complain about cleaning up yo...	woman t complain cleaning house ...
40	Non-offensive	said no pussy inside my doghouse	said pussy inside doghouse
63	Non-offensive	Addicted t hi woof woof hot lad	addicted t hi woof woof hot ...
66	Non-offensive	t woof woof and hot soles	t woof woof hot sol
67	Non-offensive	eat a do these One	eat one

Fig 2: -The above picture shows the table with tokenized text and pre-processed text.

The following steps include data pre-processing

- Data Normalization:* Data normalization is the organization of knowledge to seem similar across all records and fields. It will increase the cohesion of entry varieties resulting in cleansing, lead generation, segmentation, and better quality data.
- Data Standardization:* Data standardization is processing progress that converts the structure of various datasets into one common format of data. It deals with the transformation of datasets once the info is collected from different sources and before it's loaded into target systems.
- Data Analysis:* Data Analysis is the process of consistently applying applied math and/or logical techniques to explain and illustrate, condense and recap, and assess knowledge
- Data Visualization:* Data visualization is the observation of translating info into a visible context, such as a map or graph, to create data easier for the human brain to grasp and pull insights from. The main goal of information visual image is to make it easier to spot patterns, trends, and outliers in giant data sets

C. Feature Extraction

Machine learning algorithms cannot work directly with plain text. CountVectorizer symbolizes that it breaks sentences into words and does a very rudimentary pre-processing of the text. Removes all punctuation marks and lowercases all text. A vocabulary of recognized words is built up, which will eventually be utilized to encode unseen material within the actual world, there are totally different gratuitous characters or text in the postings or text. Numbers and grammar, for example, don't have any relation to the identification of bullying. we tend to should clean and prepare the comments for the detection part before applying the machine learning techniques. Numerous process tasks are performed during this step, together with the removal of any extraneous characters similar to stop-words, words, and tokenization. Following the pre-processing techniques like information visualization and data analysis. we prepare the 2 necessary options shown as follows

- 1) *CountVectorizer*: CountVectorizer tokenizes (that is, divides sentences into words) and does very rudimentary pre-processing on the text removes all punctuation marks and, lowercases all of the text. A vocabulary of recognized words is built up, which will eventually be utilized to encode unseen material.

D. Classifiers

The proposed model's final stage is classification, in which the retrieved characteristics are fed into an algorithm to train and take a look at the classifier and thereby evaluate whether or not it will with success notice cyberbullying. We'll use Random Forest, Ada boost, Support Vector Machine (SVM), Logistic Regression (LR), XGB, and SGD, among other machine learning methods and algorithms. By using these classifiers we find the accuracy of each of the classifiers.

- 1) *Random Forest*: A random forest is a meta estimator that employs averaging to increase predicted accuracy and control overfitting by fitting a number of decision tree classifiers on various sub-samples of the dataset. are an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set
- 2) *AdaBoost*: One of the ensembles boosting classifiers is Ada-boost or Adaptive Boosting. It combines several classifiers to enhance classifier accuracy. It's a technique that uses an associate repetitive ensemble approach. The AdaBoost classifier creates a powerful classifier by combining several low-performing classifiers, resulting in a high-accuracy classifier. Adaboost's core principle is to establish the weights of classifiers and train the data sample in each iteration so that reliable predictions of uncommon observations may be made.
- 3) *Linear Support Vector Classification*: A Linear SVC (Support Vector Classifier) is designed to fit the data you provide and provide a "best fit" hyperplane that divides or categorizes your data. Following that, you may input some features to your classifier to check what the "predicted" class is after t
- 4) *Logistic Regression*: Logistic regression is a classification algorithm that uses supervised learning to estimate the probability of a target variable. The nature of the target or dependent variable is dichotomous, implying that there are only two classes.
- 5) *XGBoost*: Extreme Gradient Boosting (XGBoost) is a distributed gradient-boosted decision tree (GBDT) machine learning toolkit that is scalable. It is the top machine learning package for regression, classification, and ranking tasks, and it includes parallel tree boosting
- 6) *SGD*: The Stochastic Gradient Descent (SGD) classifier is just a conventional SGD learning method that supports various loss functions and classification penalties. To implement SGD classification, Scikit-learn provides the SGDClassifier module. Because the update to the coefficients is conducted for each performing instance rather than at the end of examples, it has been successfully used to large-scale datasets.

III. RESULTS

	Algorithms	Accuracy
0	ada boost	0.763880
1	SVM	0.789219
2	Logistic Regression	0.800149
3	Random forest	0.776922
4	Xgboost	0.755186
5	sgd	0.797541

Fig: -3 The accuracy ratings of the classifiers.

Distribution of Tweets in the Dataset

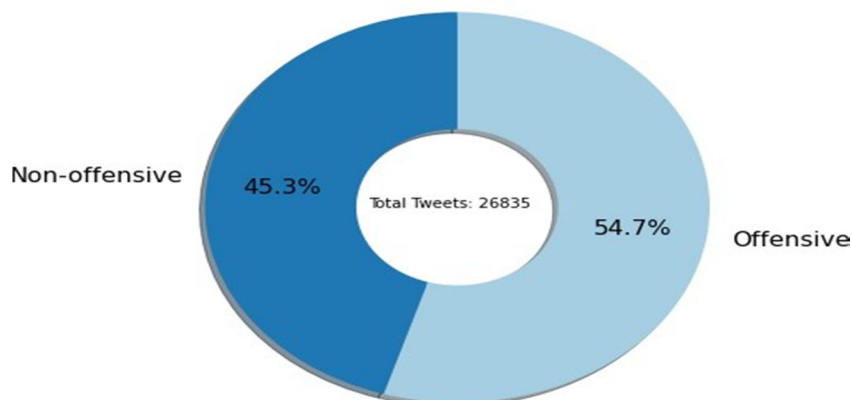


Fig: -4 Distribution of tweets in the data set

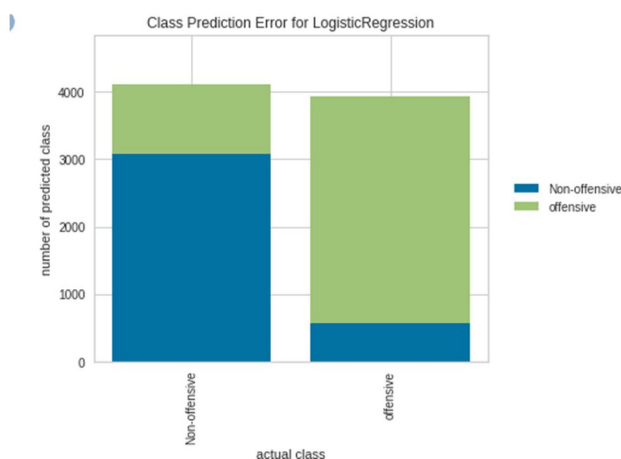


Fig: -5 Class prediction error for Logistic Regression

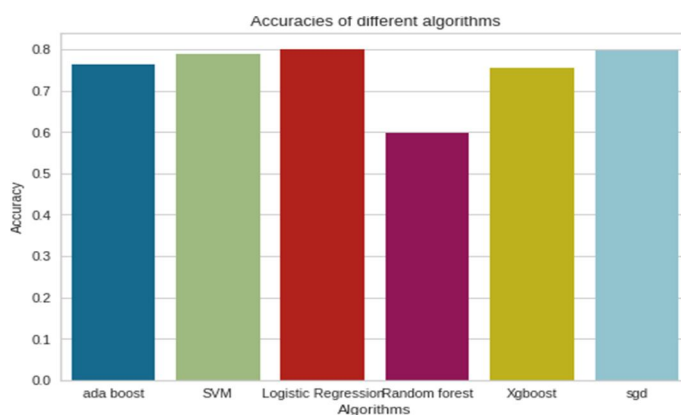


Fig:-6 Accuracies of different algorithms

IV. CONCLUSION

Using text classification techniques, we have a tendency to try to find cyberbullying text. We used a range of classification algorithms similar to SVM, logistical Regression, Xgboost, SGD, Random forest, and Ada boost classifiers, additional machine learning models may be applied in the future to cut back cyber bullying against individuals on social media



REFERENCES

- [1] soapboxie.com/social-issues/-Negative-Effects-Social-Networking
- [2] [dictionary.cambridge.org/dictionary/English trolling](http://dictionary.cambridge.org/dictionary/English/trolling)
- [3] www.merriam-webster.com/dictionary/text
- [4] coursehero.com/file/69808921/milestone-3-sophia-conflict-resolutiondocx
- [5] positiveaction.net/blog/social-bullying
- [6] merriam-webster.com/dictionary/paper
- [7] homes.cs.washington.edu/~lazowska/qsp/Images/Chap_14.pdf
- [8] techrepublic.com/article/unstructured-data-the-smart-persons-guide/
- [9] [dictionary/classification](#)
- [10] <https://www.chegg.com/homework-help/questions-and-answers/consider-comparing-performance-two-classification-algorithms-b1-based-k-fold-cross-validate-q89971862> 11:



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)