

**NAME OF THE PROJECT**

**MALIGNANT COMMENTS CLASSIFIER**

Submitted by:

Muktikanta Sahoo

**ACKNOWLEDGMENT**

I would like to express my deep and sincere gratitude to my Project supervisor, Miss. Sapna Verma, Project coordinator, Flip Robo Technologies for giving me the opportunity to do research and providing invaluable guidance throughout this project. His dynamism, vision, sincerity and motivation have deeply inspired me.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

**INTRODUCTION**

* **Business Problem Framing**

The proliferation of social media enables people to express their opinions widely online. However, at the same time, this has resulted in the emergence of conflict and hate, making online environments uninviting for users. Although researchers have found that hate is aproblem across multiple platforms, there is a lack of models for online hate detection.

* **Conceptual Background of the Domain Problem**

The proliferation of social media enables people to express their opinions widely online. However, at the same time, this has resulted in the emergence of conflict and hate, making online environments uninviting for users. Although researchers have found that hate is aproblem across multiple platforms, there is a lack of models for online hate detection.

Online hate, described as abusive language, aggression, cyberbullying, hatefulness and many others has been identified as a major threat on online social media platforms. Social media platforms are the most prominent grounds for such toxic behaviour.

There has been a remarkable increase in the cases of cyberbullying and trolls on various social media platforms. Many celebrities and influences are facing backlashes from people and have to come across hateful and offensive comments. This can take a toll on anyone and affect them mentally leading to depression, mental illness, self-hatred and suicidal thoughts.

* **Review of Literature**

Internet comments are bastions of hatred and vitriol. While online anonymity has provided a new outlet for aggression and hate speech, machine learning can be used to fight it. The problem we sought to solve was the tagging of internet comments that are aggressive towards other users. This means that insults to third parties such as celebrities will be tagged as unoffensive, but “u are an idiot” is clearly offensive.

* **Motivation for the Problem Undertaken**

Our goal is to build a prototype of online hate and abuse comment classifier which can used to classify hate and offensive comments so that it can be controlled and restricted from spreading hatred and cyberbullying.

Analytical Problem Framing

* **Mathematical/ Analytical Modeling of the Problem**

We have implemented the classification algorithm to determine the price of the houses. We divided the training set (of 100 examples) into 2 parts for training and test (of 80% and 20% examples respectively) to evaluate baseline performance to evaluate the regression statistics.

* **Data Sources and their formats**

The data set contains the training set, which has approximately 1,59,571 samples and the test set which contains nearly 1,53,164samples. All the data samples contain 8 fields which includes ‘Id’, ‘Comments’, ‘Malignant’, ‘Highly malignant’, ‘Rude’, ‘Threat’, ‘Abuse’ and ‘Loathe’.

* **Data Preprocessing Done**

Once you know what your data has and looks like you will have to transform it in order to make it suitable for algorithms to process and work more efficiently in order to give more accurate and precise results. This is essentially Data Pre-Processing which is the most important and the most time consuming stage of any ML project.

Real life data is not arranged and presented to you nicely and in a dataframe with no abnormalities. Data usually has a lot of so called abnormalities like missing values, a lot of features with incorrect format, features on different scales etc.

1. I have dropped some columns which has small impact or no impact on house pricing.
2. I have dropped all null values in this dataset.
3. Separate all categorical columns and integer columns

Hardware and Software Requirements and Tools Used

* Jupitar Notebook
* Scipy
* Matplotlib
* Seaborn

**Model/s Development and Evaluation**

* **Identification of possible problem-solving approaches (methods**)

Importing modules🡪Loading the dataset & processing the dataset----🡪Checking the corelation-🡪Visualization--🡪Model training-🡪Model selection🡪Hyperparameter testing-🡪Plotting the prediction

* **Testing of Identified Approaches (Algorithms)**

This is regression based problem so we use different classification related models to analyse and solve the problem. The algorithms used for the training and testing:

* LogisticRegression
* DecisionTreeClassifier
* RandomForestClassifier
* AdaBoostClassifier
* XGB Boost
* Kneighbors Classifier

Run and Evaluate selected models

In this process, we are going to build and train five different types of linear regression models which are the LogisticRegression,DecisionTreeClassifier,RandomForestClassifie,AdaBoostClassifier,GradientBoostingClassifier,ExtraTreesClassifier

. For all the models, we are going to use the pre-built algorithms provided by the scikit-learn package in python. And the process for all the models are the same, first, we define a variable to store the model algorithm, next, we fit the train set variables into the model, and finally make some predictions in the test set.

* Visualizations

We use different libraries for visualization like : seaborn

Matplotlib

* Interpretation of the Results

After going through a bunch of processes, we have successfully built and evaluated five different types of linear regression models in python also, choosing the best model for our given dataset. But, this won’t stop here. Each and every model we built have their own statistical and mathematical concepts.

.

**CONCLUSION**

* **Key Findings and Conclusions of the Study**

The finding of the study is that only few users over online use unparliamentary language. And most of these sentences have more stop words and are being quite long. As discussed before few motivated disrespectful crowds use these foul languages in the online forum to bully the people around and to stop them from doing these things that they are not supposed to do. Our study helps the online forums and social media to induce a ban to profanity or usage of profanity over these forums.

* **Learning Outcomes of the Study in respect of Data Science**

Through this project we were able to learn various Natural language processing techniques like lemmatization, stemming, removal of stop words. We were also able to learn to convert strings into vectors through hash vectorizer. In this project we applied different evaluation metrics like log loss, hamming loss besides accuracy.

* **Limitations of this work and Scope for Future Work**

Problems faced while working in this project:

* More computational power was required as it took more than 2 hours
* Imbalanced dataset and bad comment texts
* Good parameters could not be obtained using hyperparameter tuning as time was consumed more

Areas of improvement:

* Could be provided with a good dataset which does not take more time.
* Less time complexity
* Providing a proper balanced dataset with less errors.