

Malignant Comment Classifier

Submitted by:

Amey Prabhune

**ACKNOWLEDGMENT**

I take this an opportunity to thank all those who have directly inspired and helped us towards successful completion of this project report. We express our sincere thanks to SME MR. Shubham Yadav for his guidance.

**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 a problem across multiple platforms, there is a lack of models for online hate detection.

* Conceptual Background of the Domain Problem

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.

* 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
  + For the visualization I have used numpy, sklearn(sikit learn), pandas, matplotlib.
  + For measuring our model accuracy I have used accuracy score, confusion matrix and classification report.
  + For pre-processing I have used min-max scaler, power transform.
  + For model selection I have used train\_test\_split, and cross validation.
* Data Sources and their formats

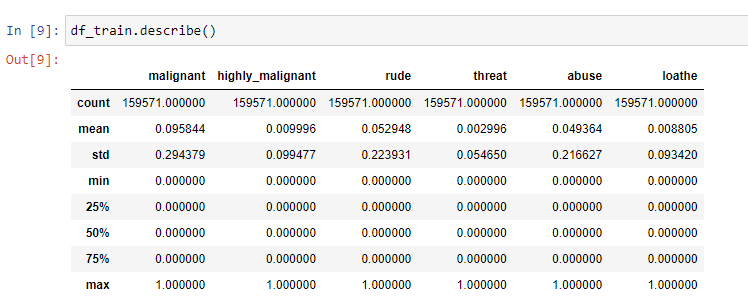
The data set contains the training set, which has approximately 1,59,000 samples and the test set which contains nearly 1,53,000 samples. All the data samples contain 8 fields which includes ‘Id’, ‘Comments’, ‘Malignant’, ‘Highly malignant’, ‘Rude’, ‘Threat’, ‘Abuse’ and ‘Loathe’.

The label can be either 0 or 1, where 0 denotes a NO while 1 denotes a YES. There are various comments which have multiple labels. The first attribute is a unique ID associated with each comment.

The data set includes:

* 1. **Malignant:** It is the Label column, which includes values 0 and 1, denoting if the comment is malignant or not.
  2. **Highly Malignant:** It denotes comments that are highly malignant and hurtful.
  3. **Rude:** It denotes comments that are very rude and offensive.
  4. **Threat:** It contains indication of the comments that are giving any threat to someone.
  5. **Abuse:** It is for comments that are abusive in nature.
  6. **Loathe:** It describes the comments which are hateful and loathing in nature.
  7. **ID:** It includes unique Ids associated with each comment text given.
  8. **text:** This column contains the comments extracted from various social media platforms.
* Data Inputs- Logic- Output Relationships

First I checked null values. In our data set there are no null values. In our dataset we have multiple target variable and we have separate data set for training and testing. We need to apply NLP of our comment column for cleaning the text. So we can find predict malignant comment.



* Hardware and Software Requirements and Tools Used

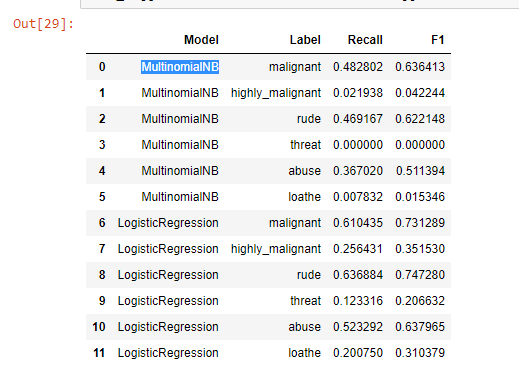
1. Software Requirement :
   1. Excel
   2. OS – windows , Linux
   3. Jupyter Notebook
   4. Internet browser
2. Hardware Requirement:
   1. RAM: 4 GB or more than.
   2. ROM: 50 GM or more than.
   3. Internet connection.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

In target variable we have only two values 0 and 1 so it is a classification problem. For model building we will use classification model like. Random Forest, SVC, AD Boost etc. in dataset we have multiple target variable and two feature columns means independent and dependent variable. First I drop column which are not useful for the model building.

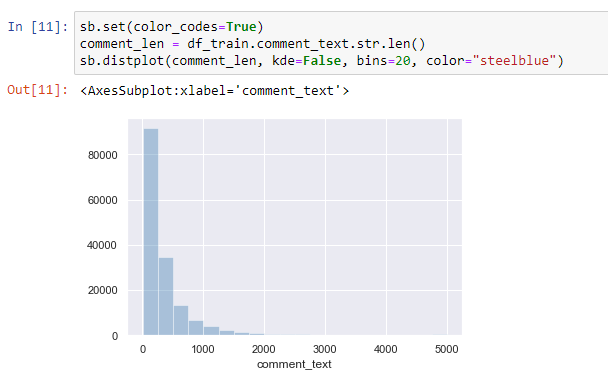
* Testing of Identified Approaches (Algorithms)
  1. Logistic Regression
  2. MultinomialNB
  3. Support Vector Classifier
  4. Gaussian NB
* Run and Evaluate selected models



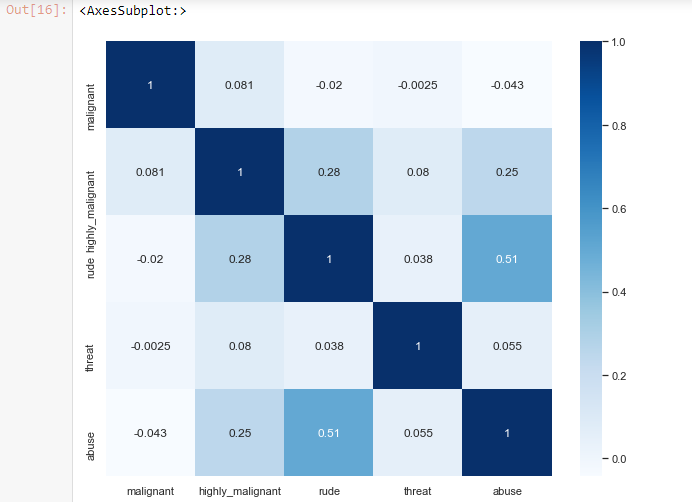
* Key Metrics for success in solving problem under consideration

In dataset, we can see the correlation with the help of box plot means there are some big values in our dataset. There are no null values in dataset. Most of the columns are in numeric type. Dataset is imbalance.

* Visualizations









**CONCLUSION**

* Key Findings and Conclusions of the Study
  1. We have 2 values in our target variable.
  2. We have a multiple target variable.
  3. There are no null values.
  4. We have to perform NLP on our dataset.

* Learning Outcomes of the Study in respect of Data Science

With the help of visualization we can easily understand our data.

Visuals and diagrams makes it easier for us to identify strongly correlated parameters. Visualization can improve speed of decision making. Clean data can give us more accurate result. If data is noisy then our model won’t work as we expect. In this model random forest algorithm gives us best accuracy score.

* Limitations of this work and Scope for Future Work
  1. Data: Lack of Good Data.
  2. Time: building a machine learning model is time consuming.
  3. Performance: Performance cannot guaranteed.