

MALIGNANT-COOMENTS-CLASSIFIER

Submitted by:

MAHESH AGASHIMANI

**ACKNOWLEDGMENT**

I take great pleasure to thank and acknowledgement the allowance by Data Trained Education and permission by Flip Robo. I extend whole hearted thanks to them I worked and learned a lot and sharing me the knowledge and experience.

Data Trained Education and Flip Robo provided training is the very important to completion of project.

**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

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

This dataset contains information about Malignant comments classifier listed on train dataset. This data can be used for a lot of purposes such as Malignant prediction to exemplify the use of Random Forest regression in Machine Learning.

* Data Sources and their formats

The columns in the given dataset is as follows:..

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’.

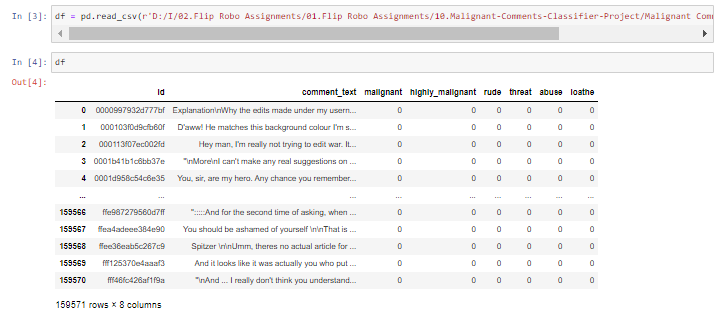
The data set includes:

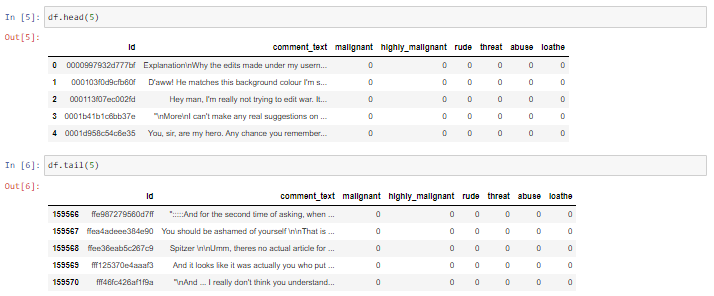
* **Malignant:** It is the Label column, which includes values 0 and 1, denoting if the comment is malignant or not.
* **Highly Malignant:** It denotes comments that are highly malignant and hurtful.
* **Rude:** It denotes comments that are very rude and offensive.
* **Threat:** It contains indication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

Importing the Libraries:

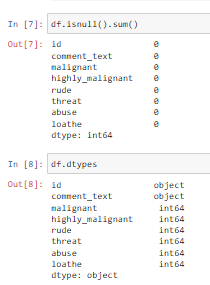


. Loading the Dataset:





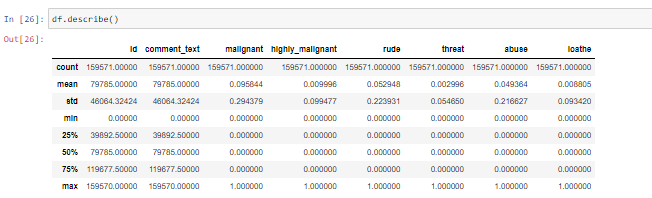
After that we are checking if Null values are present in datasets.



In checked above data sets, there is no Null values present.

* Data Inputs- Logic- Output Relationships

Using describe method its showing the value of columns like count, mean, std, min, 25%, 50%, 75% max value.



* State the set of assumptions (if any) related to the problem under consideration

As given in datasets my assumption is predicting malignantand this problem is Classification Problem.

* Hardware and Software Requirements and Tools Used

The needed time to train the model depends on the capability of the used system during the experiment. Some libraries use GPU resources over the CPU to take a shorter time to train a model.

|  |  |
| --- | --- |
| Operating System | Windows 10 |
| Processor | Core i7 |
| RAM | 16GB |
| Graphics card | 1080 TI OC |

Also we are using Jupiter notebook for running the code

.

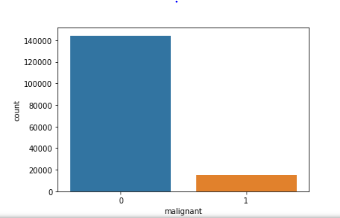
**Model/s Development and Evaluation**

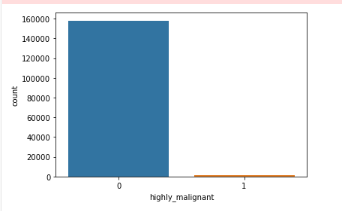
* Identification of possible problem-solving approaches (methods)

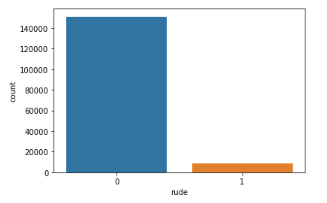
After describing the data, we are checking the EDA part for Categorical & Numerical data.

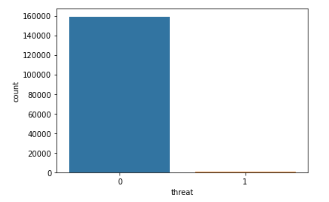
EDA part for Categorical data using Count Plot:.

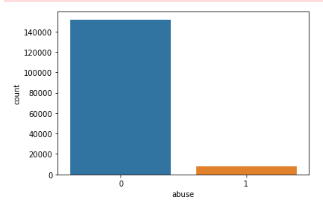


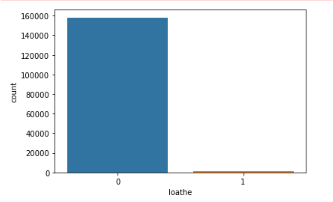




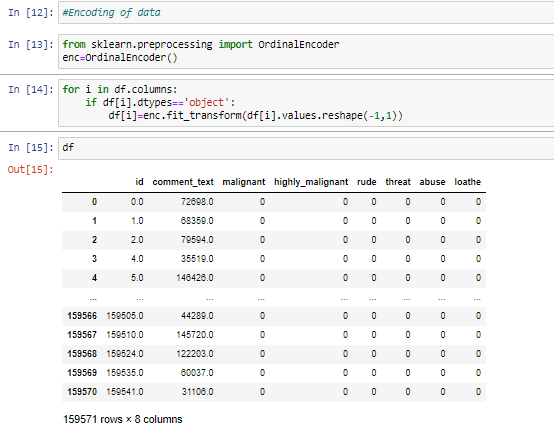








After this we are going to checking dtypes & encoding the data in categorical to numerical data. Machine learning technique is processing only numerical data.



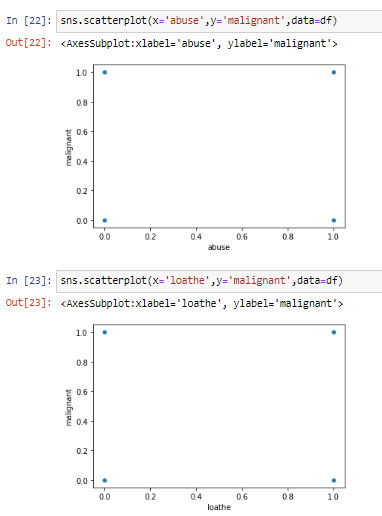
As above image we encoded categorical data into numerical data.

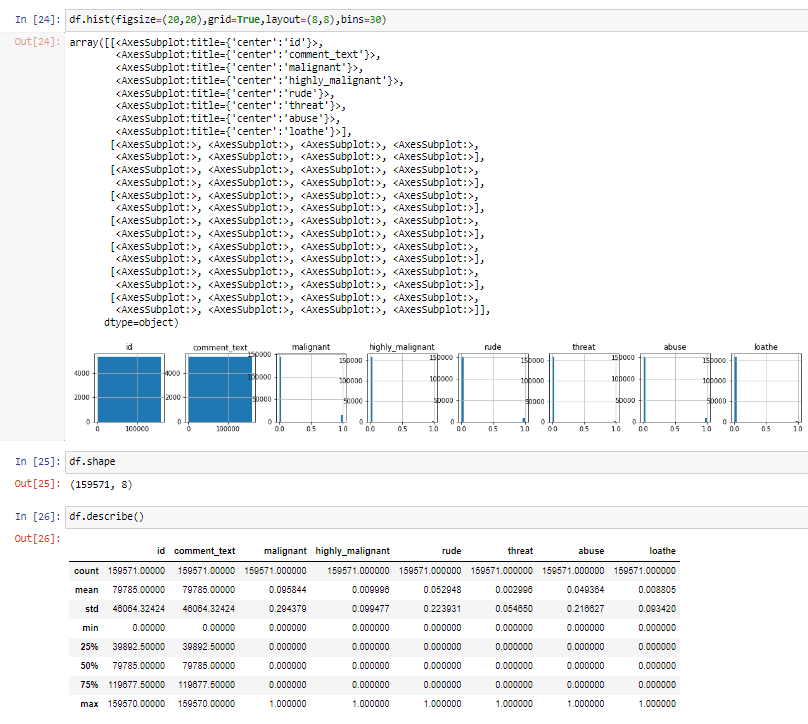
Visualizing data using Scatterplot: -

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs we will understand the data.

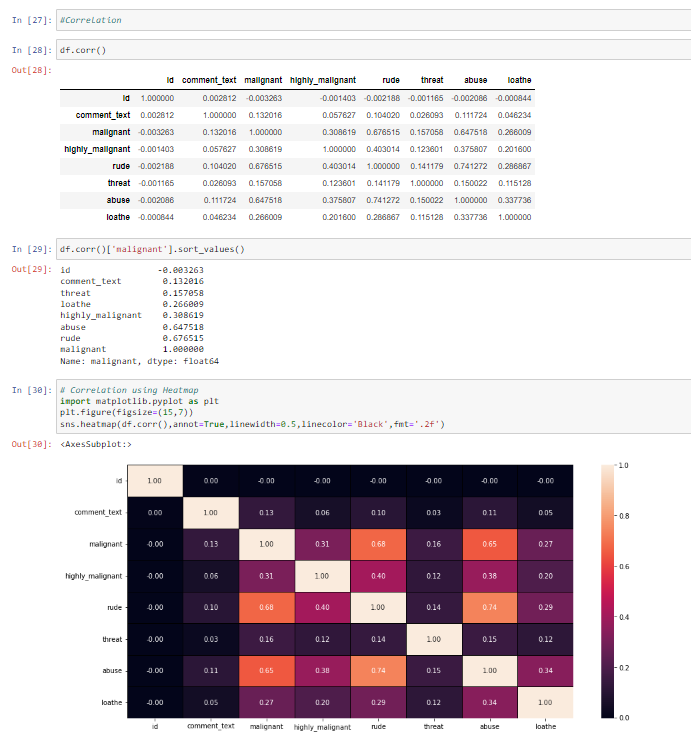






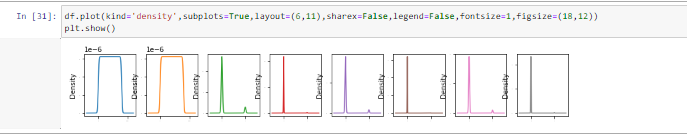


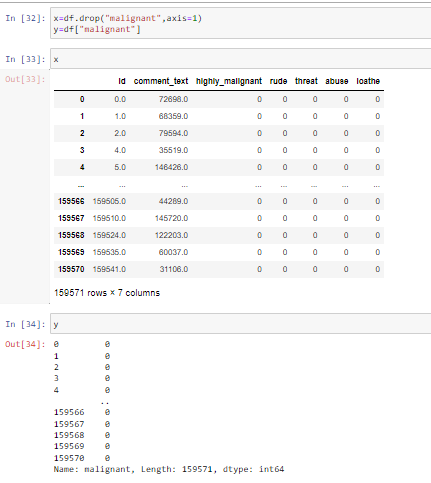
After this we are going to check Correlation of the data.



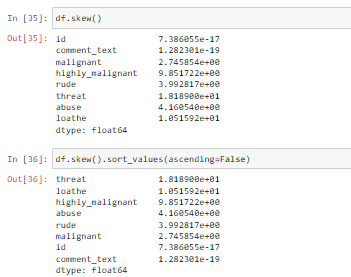
Observation:

* comment\_text and malignant features are highly correlated with each other.
* highly\_malignant and malignant features are highly correlated with each other.
* threat and malignant columns features are highly correlated with each other.
* loathe and malignant is also highly correlated with Price column.
* abuse and malignant is highly correlated with each other.
* rude and malignant is highly correlated with each other.

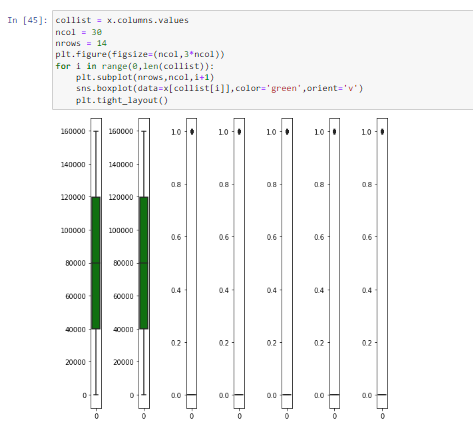


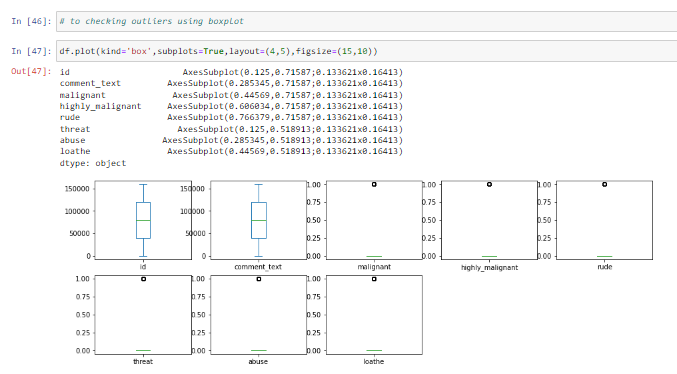


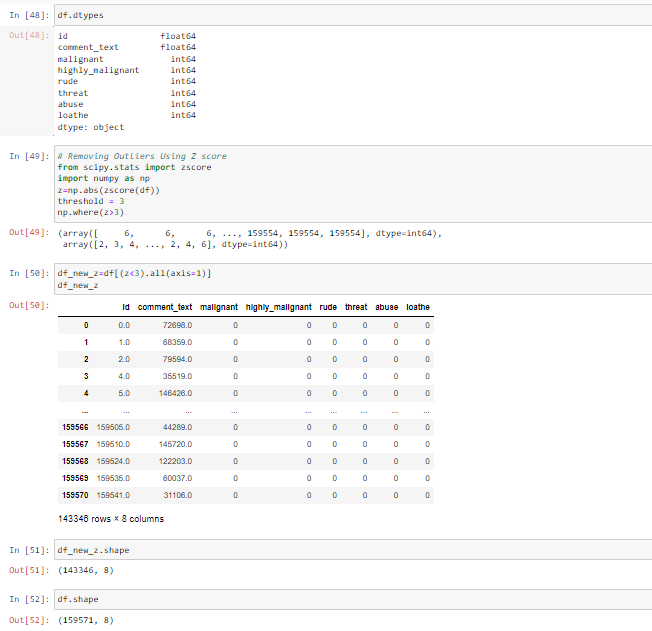
Checking skewness & outliers for Categorical data.

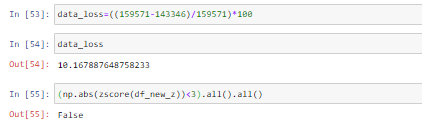








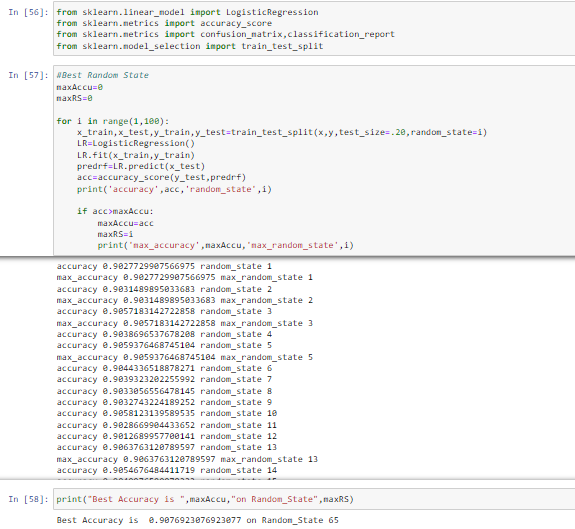


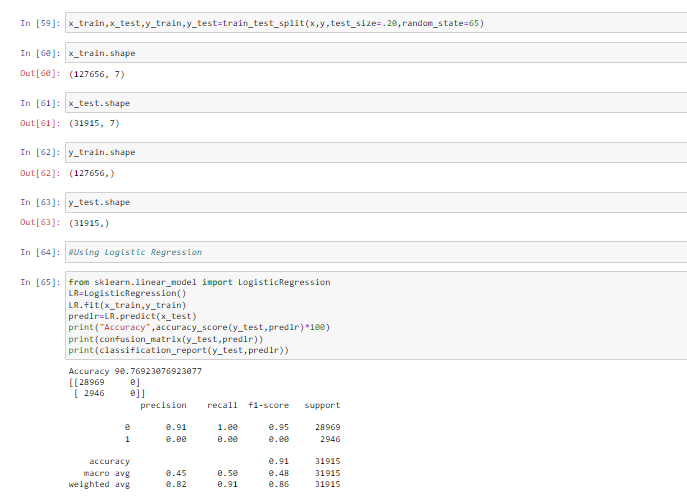


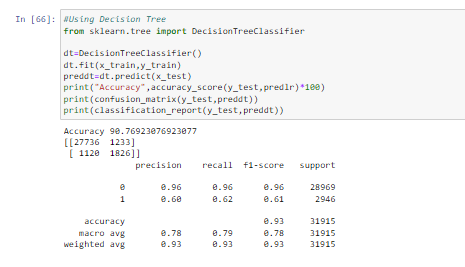
* Testing of Identified Approaches (Algorithms)

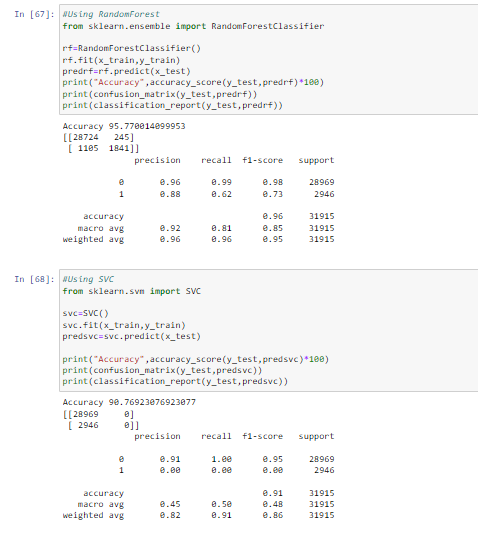
For Training & Testing data using RandomForestClassifier model.

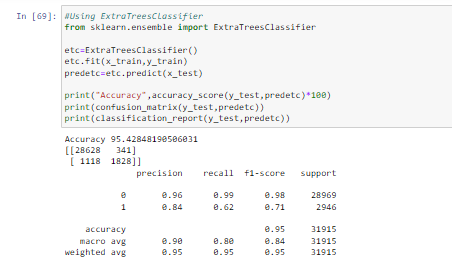
Importing necessary libraries.

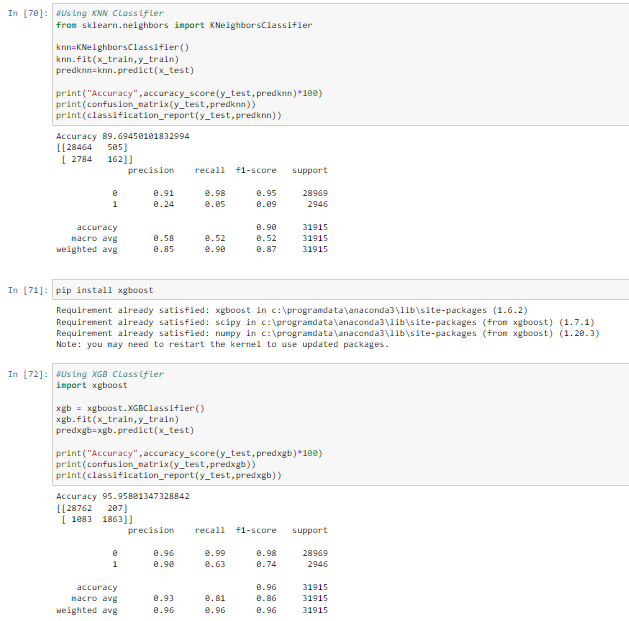


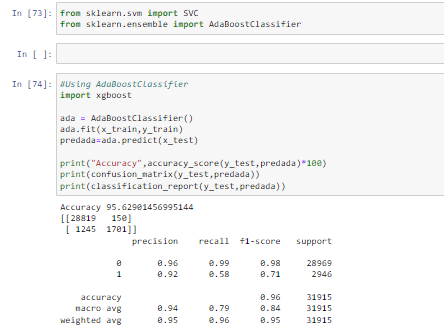












* Run and Evaluate selected models

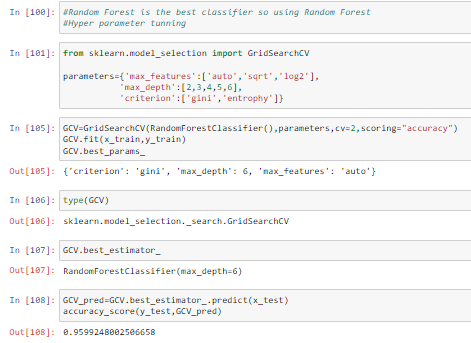




As above checked algorithm Random Forest Classifier accuracy is good.

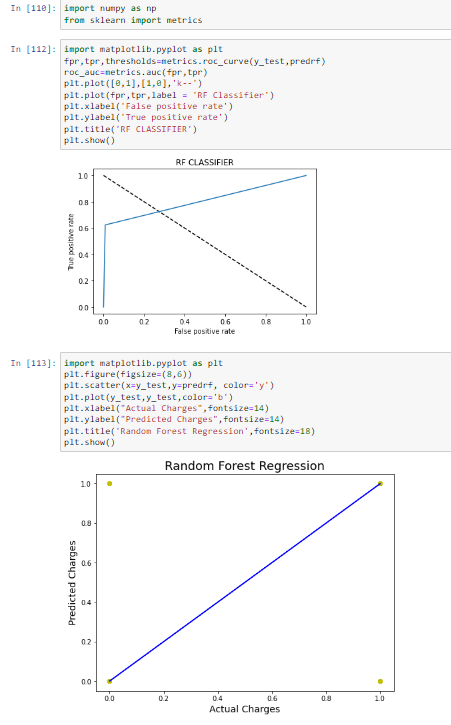
We have achieved approximately 95% accuracy using RandomForest Classifier.

Hyperparameter tuning of data-



* Visualizations

AUC ROC Plot:

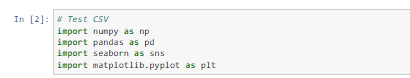


* Interpretation of the Results

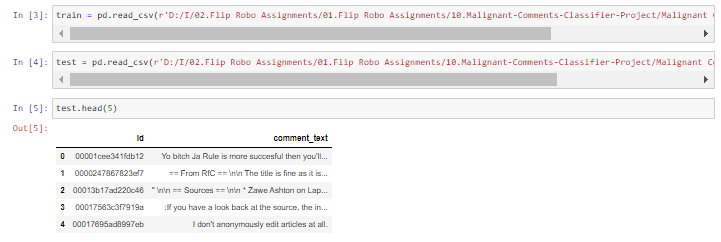
After Building a model RandomForestClassifier method is good for predict the output.

* Model Building for Test Data(test.csv):-

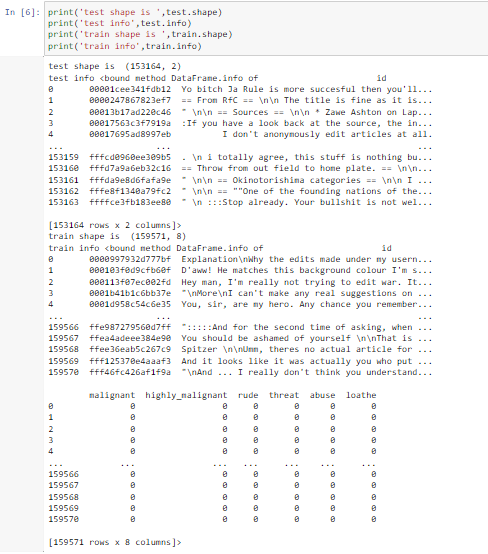
Importing the Libraries:

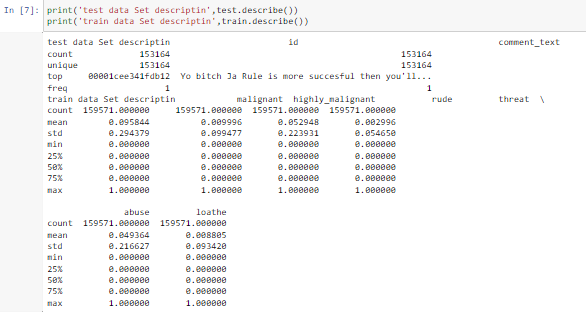


Loading the Dataset:

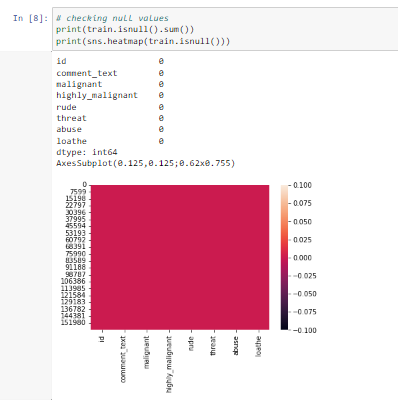


After that we are checking shape of values & describing the data:





After that we are checking if Null values are present in datasets.

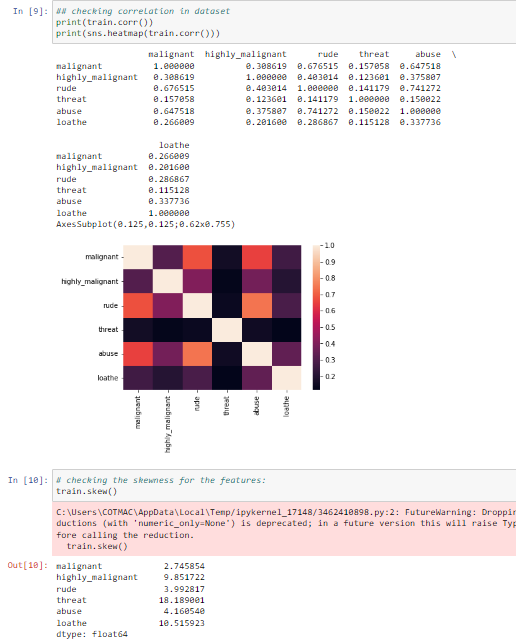


In checked above data sets, there is no Null values present.

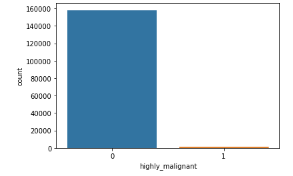
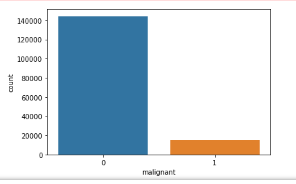
**Model/s Development and Evaluation**

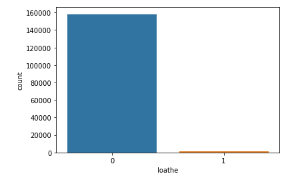
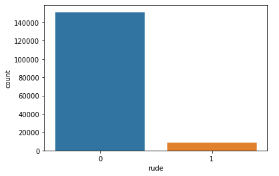
* Identification of possible problem-solving approaches (methods)

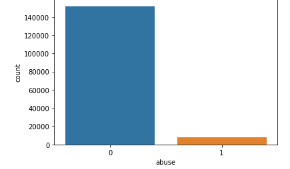
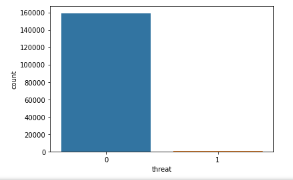
Checking correlation & skewness of the data:



After that Checking EDA Part:



Methodology

1. Data Preprocessing

The following steps were taken to process the data:

→ A string without all punctuations to be prepared:

1. The string library contains punctuation characters. This is imported and all numbers are appended to this string. Our comment\_text field contains strings such as won't, didn't, etc. which contain apostrophe character('). To prevent these words from being converted to wont or didn't, the character ' represented as \' in escape sequence notation is replaced by empty character in the punctuation string.

→ Updating the list of stop words:

1. Stop words are those words that are frequently used in both written and verbal communication and thereby do not have either a positive or negative impact on our statement like “is, this, us, etc.”.

2. Single letter words if existing or created due to any preprocessing step do not convey any useful meaning and so they can be directly removed. Hence letters from b to z, will be added to the list of stop words imported directly.

→ Stemming and Lemmatizing:

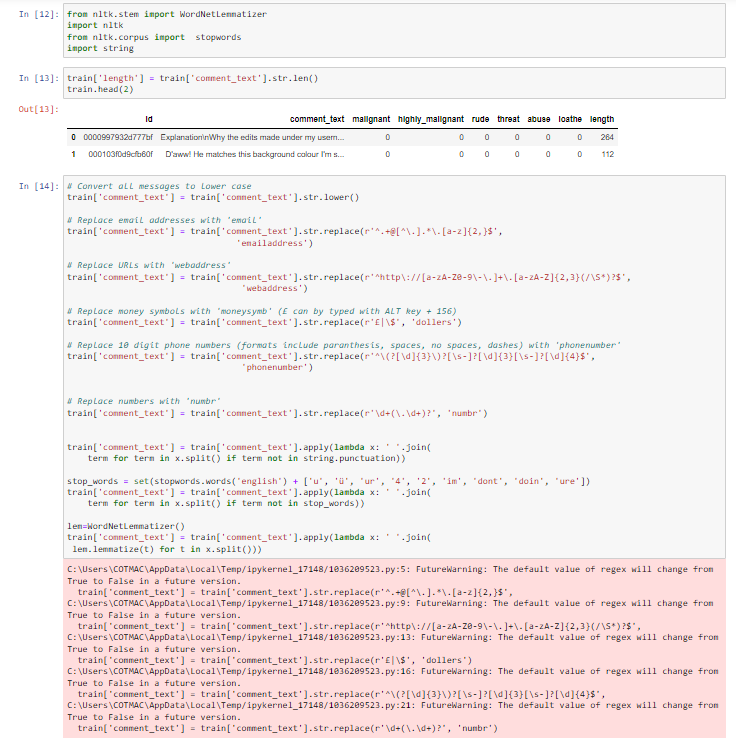
1. The process of converting inflected/derived words to their word stem or the root form is called stemming. Many similar origin words are converted to the same word e.g. words like "stems", "stemmer", "stemming", "stemmed" as based on "stem".

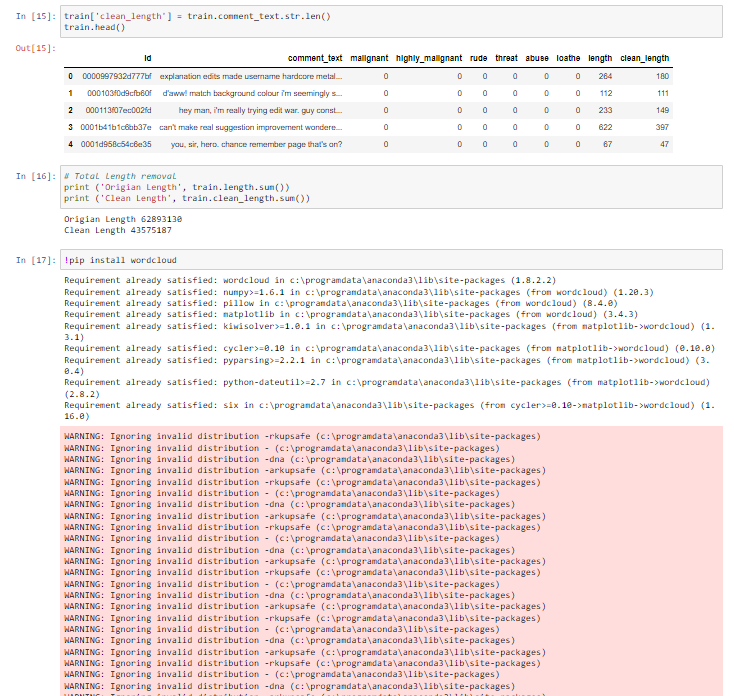
2. Lemmatizing is the process of grouping together the inflected forms of a word so they can be analyzed as a single item. This is quite similar to stemming in its working but differs since it depends on correctly identifying the intended part of speech and meaning of a word in a sentence, as well as within the larger context surrounding that sentence, such as neighboring sentences or even an entire document.

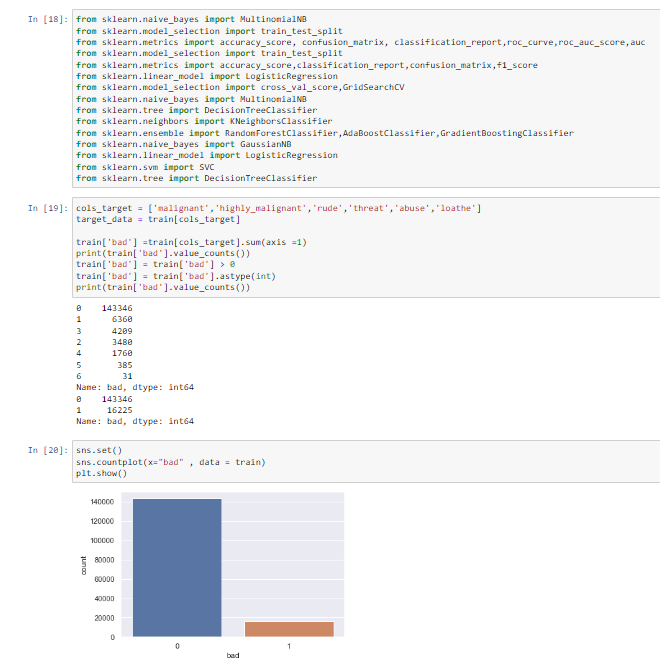
3. The wordnet library in nltk will be used for this purpose. Stemmer and Lemmatizer are also imported from nltk.

We have performed the following preprocessing on the data:

* Removed punctuations
* Removed the stop words
* Stemming and lemmatization
* Applied counter vectorizer





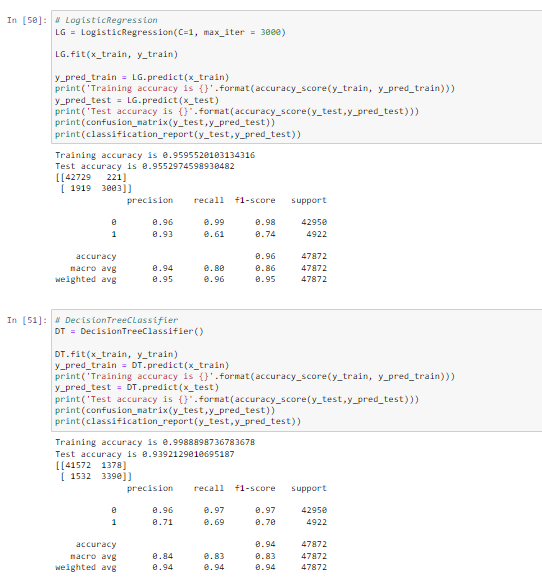


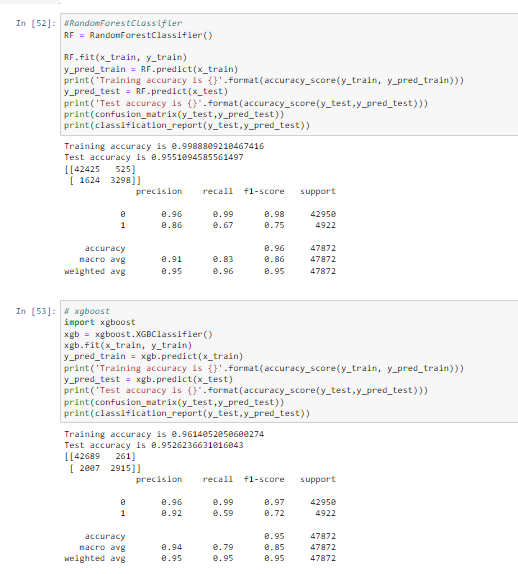


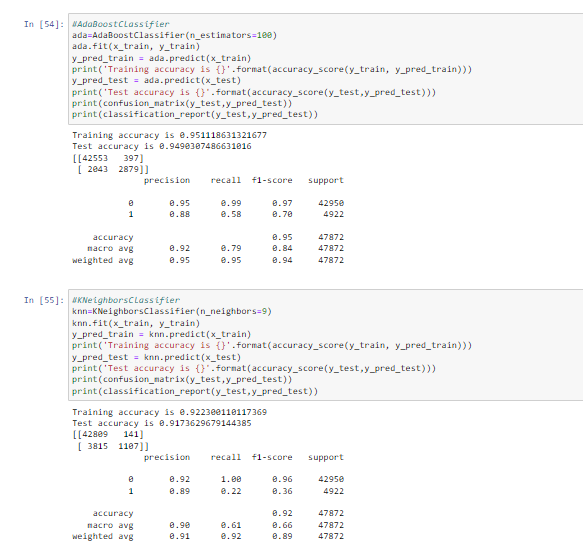
→ Splitting dataset into Training and Testing:

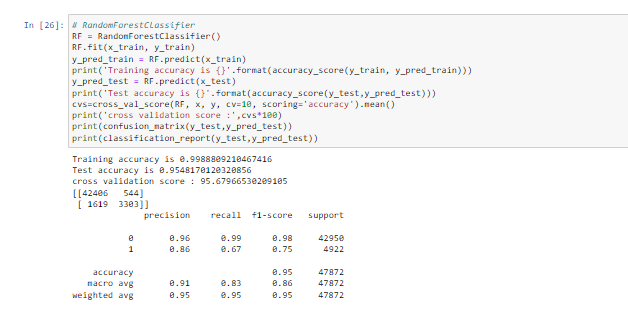
1. Since the system was going out of memory using train\_test\_split, I had jumbled all the indexes in the beginning itself.

2. The shuffle function defined here performs the task of assigning first 2/3rd values to train and remaining 1/3rd values to the test set.





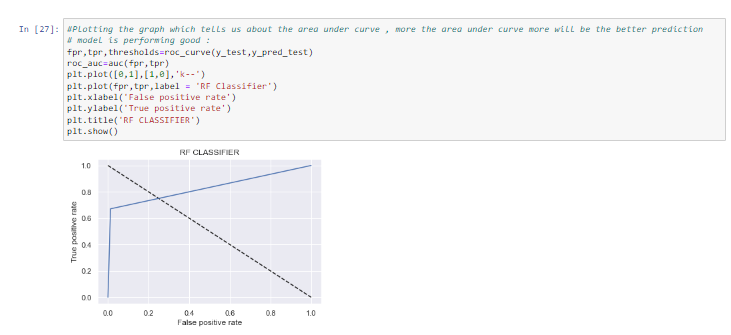




As above checked algorithm Random Forest Classifier accuracy is good.

* Visualizations

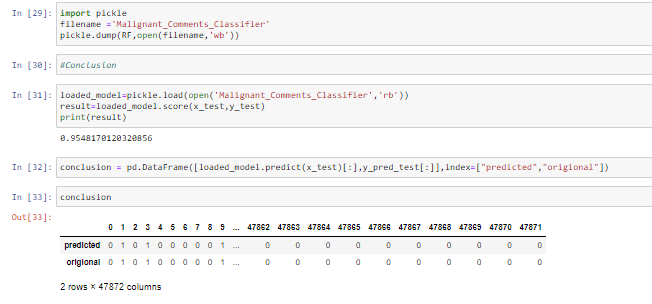
AUC ROC Plot:



**CONCLUSION**

* Key Findings and Conclusions of the Study

Using different models for Malignant Comments Classifier prediction. Seven different types of Machine Learning methods including DecisionTreeClassifier, RandomForestClassifier, SVC, ExtraTreesClassifier , KNeighborsClassifier, XGBClassifier , AdaBoostClassifier are compared and analyzed for optimal solutions. Even though all of those methods achieved desirable results, different models have their own pros and cons. The Random Forest method has the lowest error on the training set but is prone to be overfitting. So using RandomForestClassifier method draw the AUC ROC Curve.



* Learning Outcomes of the Study in respect of Data Science

As we checked in dataset not present null values.

Visualizing the categorical data using countplot.

Converted Categorical data into numerical data using Encoder method.

After all this split the data into train & test split.

* Limitations of this work and Scope for Future Work

The used pre-processing methods do help in the prediction accuracy. However, experimenting with different combination of pre-processing methods to achieve better prediction accuracy.

Make available features in combining the features and predict it may be improved performance.