FLIP ROBO

NAME OF THE PROJECT

Malignant-Comments-Classifier

Submitted by:

Chandan Kumar

FLIPROBO SME:

Ms. Khushboo Garg

ACKNOWLEDGMENT

I would like to express my special gratitude to “Flip Robo” team, who has given me this opportunity to deal with a beautiful dataset and it has helped me to improve my analyzation skills. And I want to express my huge gratitude to Ms. Khushboo Garg (SME Flip Robo), she is the person who has helped me to get out of all the difficulties I faced while doing the project.

A huge thanks to “Data trained” who are the reason behind my Internship at Fliprobo. Last but not least my parents who have been my backbone in every step of my life.

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.

- 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 inoffensive, but “u are an idiot” is clearly offensive.

  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

  Describe your objective behind to make this project, this domain and what is the motivation behind.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  The Problem is of Classification.

  The Data consists of 8 features in the dataset.

  There are more than 200K samples in the dataset.

  All the Data is of Int datatype.

  Use various algorithms to build up the model.

Almost all of the features are highly skewed, skewness is to be addresses.

There are some samples with non-negotiable values to the respective features that must be addressed.

The data is unscaled.

There are no missing values in the dataset.

The target classes are highly imbalance; imbalance must be addressed.

- Data Sources and their formats
- The data was provided by the client to “FlipRobo Technologies”. The data is in the form of a comma separated file (CSV). The data i.e. the features and the target are in the single file.

```
mcc=pd.read_csv('Documents/train.csv')
```

```
mcc.shape

(159571, 8)
```

```
mcc.columns

Index(['id', 'comment_text', 'malignant', 'highly_malignant', 'rude', 'threat',
       'abuse', 'loathe'],
      dtype='object')
```

The different datatypes of these features are as shown in above figure. Out of all features only two features with object datatypes and rest are int64.

- Data Inputs- Logic- Output Relationships

Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

Data is inputted in the form of a Pandas data frame to the model. The model is evaluated on a validation set using 3 Time Repeated, 5 fold Stratified cross validation set with AUC_ROC as the scoring parameter, And the model which preforms the best on the validation set is used for prediction of the classes and their respective probability per record on the test set.

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

There are no such formal assumptions as we are Random-Forest to be the best model, producing better results than any other algorithms. Random Forests Algorithms are non-parametric and can thus handle skewed and multi-modal data as well as categorical data that are ordinal or non-ordinal.

- Hardware and Software Requirements and Tools Used

Hardware Required:

- A computer with a processor i3 or above.
- More than 4 GiB of Ram.
- GPU preferred.
- Around 100 Mib of Storage Space.

Software Required:

- Python 3.6 or above
- Jupyter Notebook.
- Excel

Tools/Libraries Used:

1. Computing Tools:
   - Numpy
   - Pandas
   - Scipy
   - Sk-learn
2. Visualizing Tools:
   - Matplotlib
   - Seaborn
3. Saving Tools:
   - Joblib

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Describe the approaches you followed, both statistical and analytical, for solving of this problem.

- Testing of Identified Approaches (Algorithms)

  Listing down all the algorithms used for the training and testing.

  The Algorithms used for testing, training and Validating the models are as follows:

- ➢ Logistic Regression
- ➢ SVC (with an rbf kernel)
- ➢ Decision Tree
- ➢ K Nearest Neighbour
- ➢ Naïve Bayes
- ➢ Random Forest
- ➢ Gradient Boosting

- Run and Evaluate selected models

```
-----------------------------------------------------------------------------------------
urrent Model in Progress: Logistic Regression
=========================================================================================
raining:  BinaryRelevance(classifier=LogisticRegression(), require_dense=[True, True]
esting:

       Hamming Loss  : 0.022066084314470186
       Accuracy Score: 0.9123433345993164
             precision    recall  f1-score   support

          0       0.92      0.51      0.66      1281
          1       0.61      0.18      0.28       150
          2       0.95      0.54      0.69       724
          3       0.00      0.00      0.00        44
          4       0.81      0.45      0.58       650
          5       0.88      0.13      0.22       109

  micro avg       0.89      0.47      0.61      2958
  macro avg       0.70      0.30      0.40      2958
eighted avg       0.87      0.47      0.60      2958
samples avg       0.05      0.04      0.04      2958
```

```
=====================================================================================
urrent Model in Progress: Random Forest Classifier
=====================================================================================
raining:  BinaryRelevance(classifier=RandomForestClassifier(), require_dense=[True, True]
esting:

       Hamming Loss  : 0.022066084314470186
       Accuracy Score: 0.906114698063046
             precision    recall  f1-score   support

          0       0.81      0.60      0.69      1281
          1       0.57      0.05      0.10       150
          2       0.87      0.69      0.77       724
          3       0.00      0.00      0.00        44
          4       0.69      0.53      0.60       650
          5       0.81      0.16      0.26       109

  micro avg       0.79      0.55      0.65      2958
  macro avg       0.62      0.34      0.40      2958
eighted avg       0.77      0.55      0.63      2958
samples avg       0.05      0.05      0.05      2958
ompleted in [959.1025076999999 sec.]
```

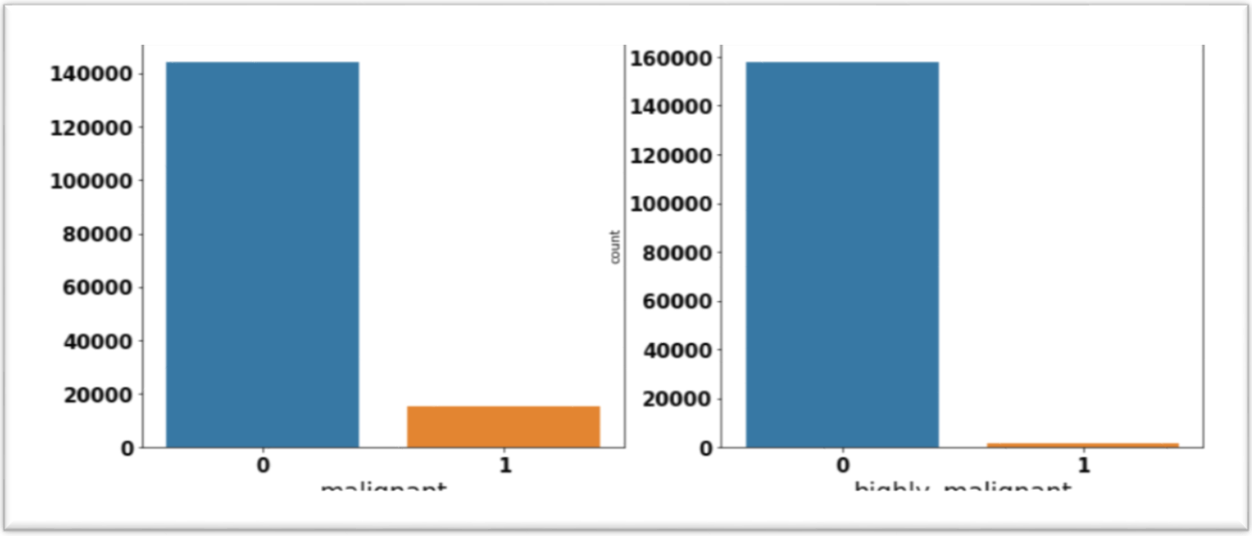
```
urrent Model in Progress: Support Vector Classifier
=====================================================================================
raining:  BinaryRelevance(classifier=LinearSVC(max_iter=3000), require_dense=[True, True]
esting:

       Hamming Loss  : 0.020952019242942144
       Accuracy Score: 0.9115077857956704
             precision    recall  f1-score   support

          0       0.85      0.60      0.71      1281
          1       0.49      0.16      0.24       150
          2       0.90      0.65      0.75       724
          3       0.50      0.18      0.27        44
          4       0.75      0.53      0.62       650
          5       0.80      0.32      0.46       109

  micro avg       0.82      0.56      0.67      2958
  macro avg       0.71      0.41      0.51      2958
eighted avg       0.81      0.56      0.66      2958
samples avg       0.06      0.05      0.05      2958
ompleted in [3.748981000001853 sec.]
```

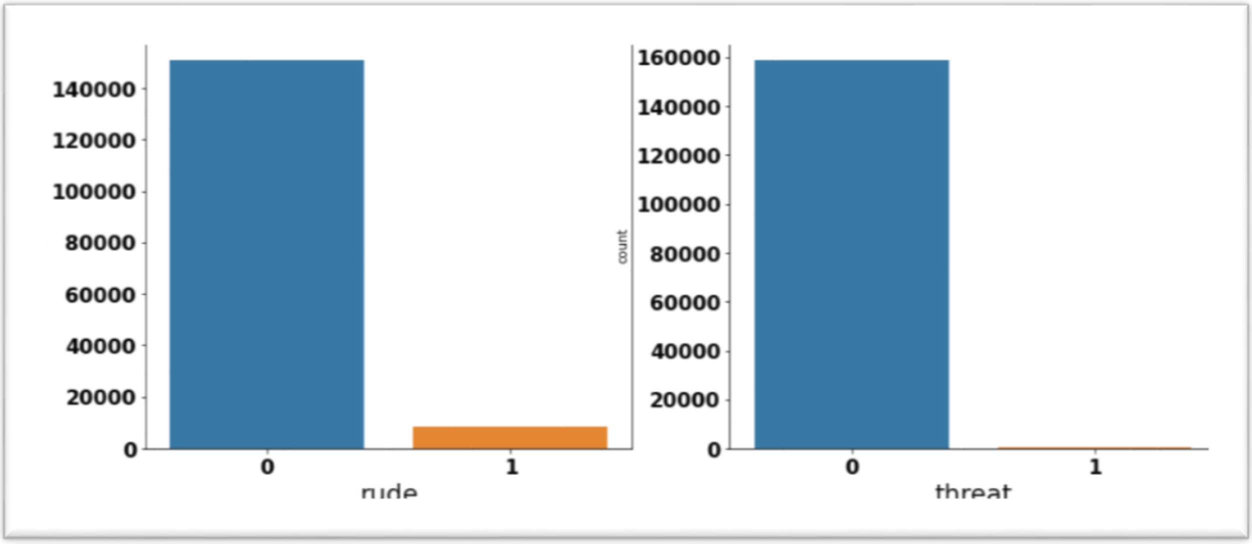
```
urrent Model in Progress: Ada Boost Classifier
================================================================================
raining:  BinaryRelevance(classifier=AdaBoostClassifier(), require_dense=[True, True]
esting:

      Hamming Loss  : 0.023446005823521965
      Accuracy Score: 0.9057349031522978
           precision    recall  f1-score   support

         0       0.82      0.53      0.65      1281
         1       0.51      0.23      0.32       150
         2       0.90      0.60      0.72       724
         3       0.53      0.18      0.27        44
         4       0.71      0.44      0.54       650
         5       0.60      0.28      0.39       109

  micro avg       0.80      0.50      0.61      2958
  macro avg       0.68      0.38      0.48      2958
eighted avg       0.79      0.50      0.61      2958
samples avg       0.05      0.04      0.04      2958
```

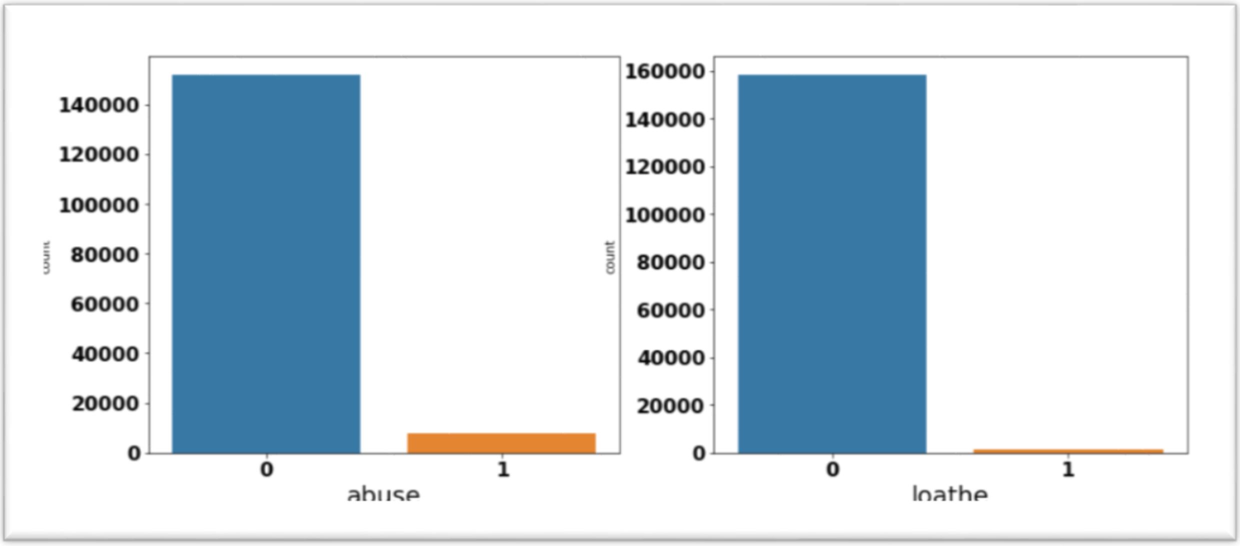
- Key Metrics for success in solving problem under consideration

  The key-metric under considerations is AUC_ROC although the model was finalized on basis of other metrics as Matthew's Correlation Coefficient (MCC) as well as F1-score.

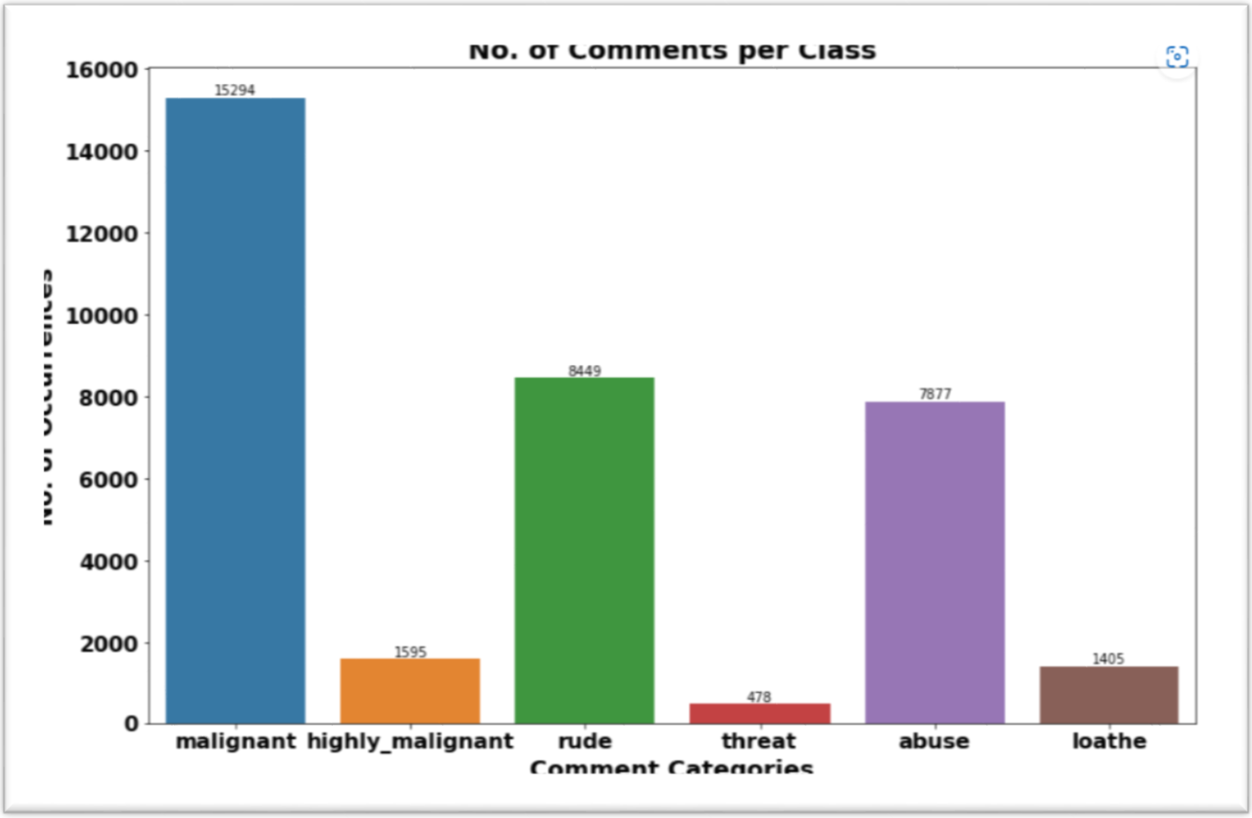
- Visualizations

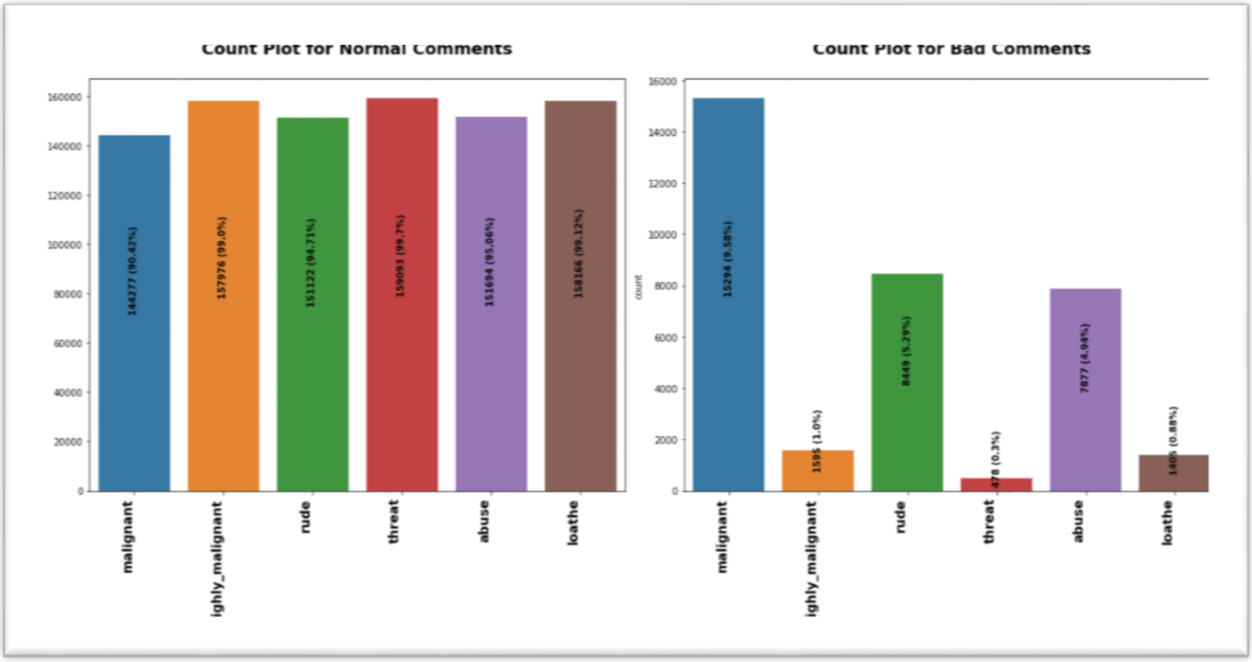



140000
120000
100000
80000
60000
40000
20000
0
0
1
rude
160000
140000
120000
100000
80000
60000
40000
20000
0
count
0
1
threat

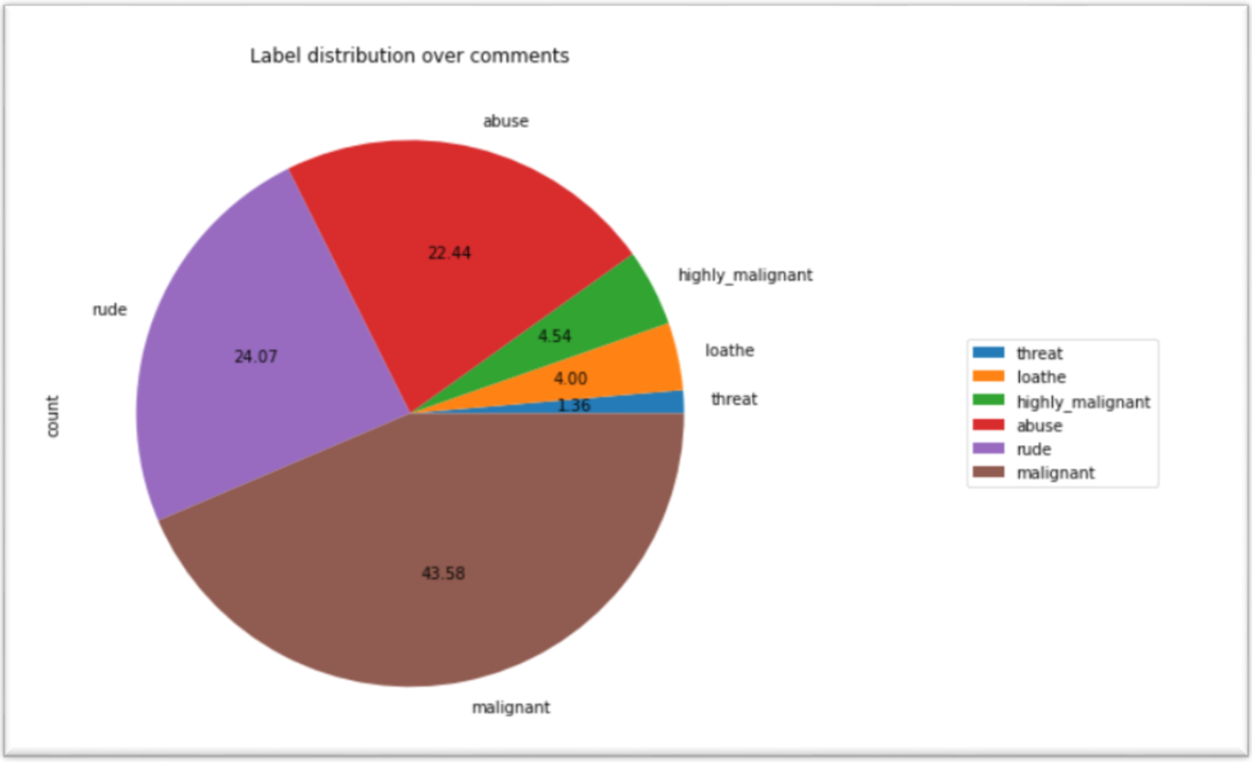

140000
120000
100000
80000
60000
40000
20000
0
0
1
abuse
160000
140000
120000
100000
80000
60000
40000
20000
0
count
0
1
loathe

- Out of total Negative comments the maximum negative comments come with Malignant in nature followed by rude categories.
- Around 90% comments are Good/Neutral in nature while rest 10% comments are Negative in nature.
- Very few comments come with threatening nature.

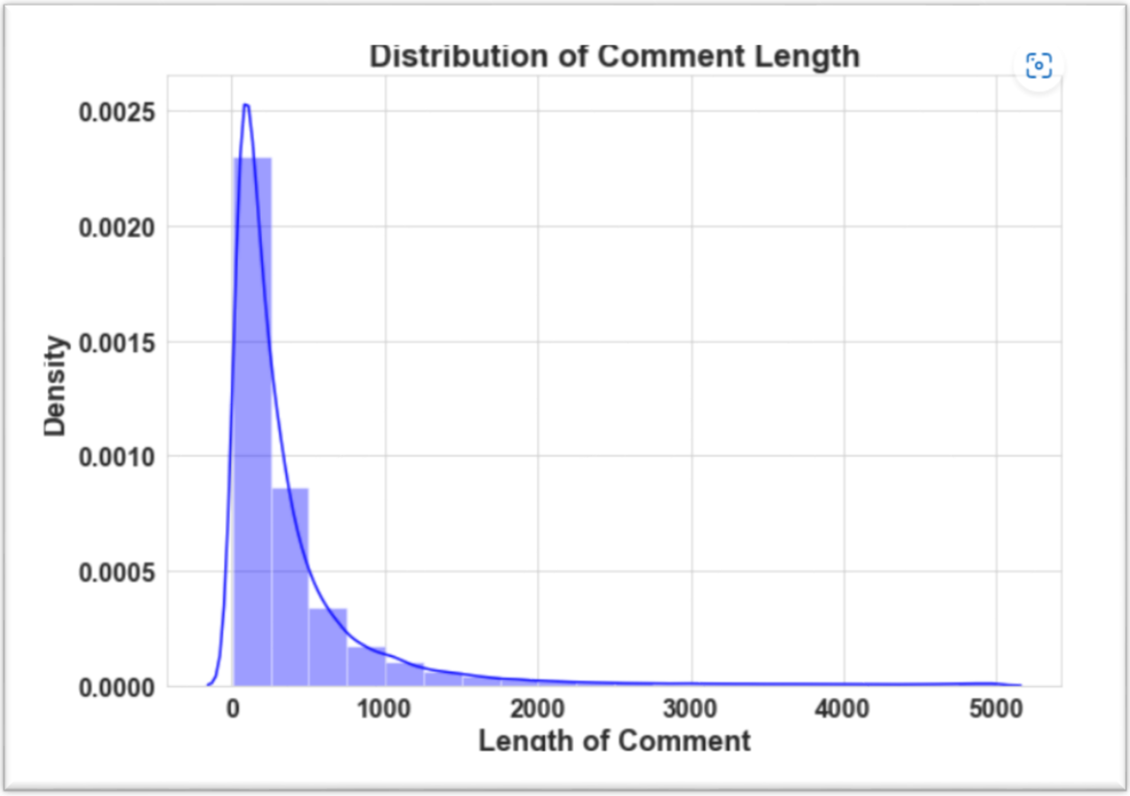




Out of total negative comments around 43.58% are malignant in nature followed by 24.07% are rude comments



- Above is a plot showing the comment length frequency. As noticed, most of the comments are short with only a few comments longer than 1000 words.
- Majority of the comments are of length 500, where maximum length is 5000 and minimum length is 5. Median length being 250.



- Interpretation of the Results

```
from sklearn.metrics import roc_curve, auc, roc_auc_score, multilabel_confusion_matrix
print("Confusion matrix:\n\n", multilabel_confusion_matrix(y_test, fmod_pred))

Confusion matrix:

 [[[10720    73]
  [  507   668]]

 [[11833     0]
  [  135     0]]

 [[11268    42]
  [  231   427]]

 [[11930     0]
  [   38     0]]

 [[11274    98]
  [  289   307]]

 [[11869     3]
  [   79    17]]]
```

This is the classifications report on the test set. Since we have high imbalance in our target classes we used AUC_ROC to evaluate the model.

CONCLUSION

- Key Findings and Conclusions of the Study
- Linear Support Vector Classifier performs better with Accuracy Score: 91.15077857956704 % and Hamming Loss: 2.0952019242942144 % than the other classification models.
- Final Model (Hyper parameter Tuning) is giving us Accuracy score of 91.26% which is slightly improved compare to earlier Accuracy score of 91.15%.
- SVM classifier is fastest algorithm compare to others.

Thank You