NAME OF THE PROJECT

Malignant Comment Classifier

Submitted by:

Nomaan Sayed

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the 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. 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.

  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. 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.

- ## Conceptual Background of the Domain Problem

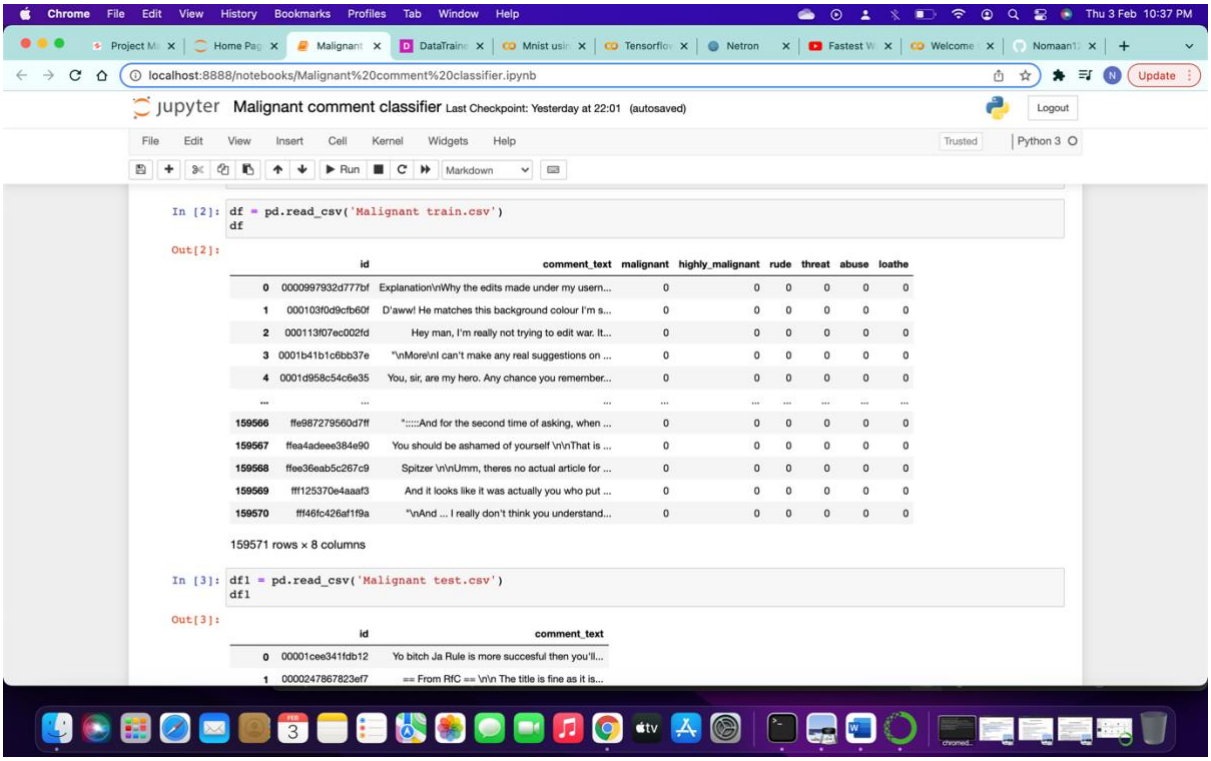
  Nowadays it is common that people shows their hatred towards a particular person on social media by commenting in abusive language and other things.

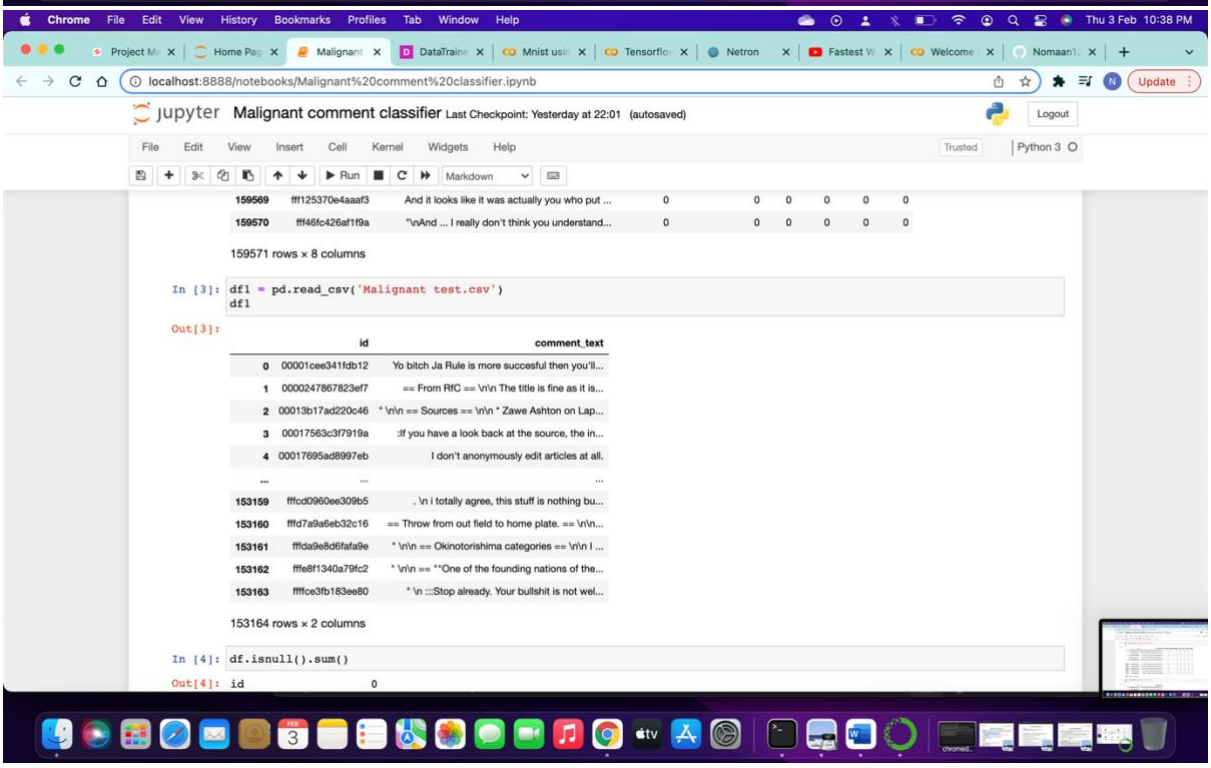
- ## Review of Literature

  It is common to abuse someone on social media, E-commerce site, twitter etc. We all have seen some or the other time , Machine learning helps us to get through this problem and detect this comments on real time.

Analytical Problem Framing

- Data Sources and their formats
  Training and Testing data is provides separately.





- Data Preprocessing Done

  1- Load the data

  2- Checking null values

  3- Encoding dataset

  4- Data Visualization using seaborn

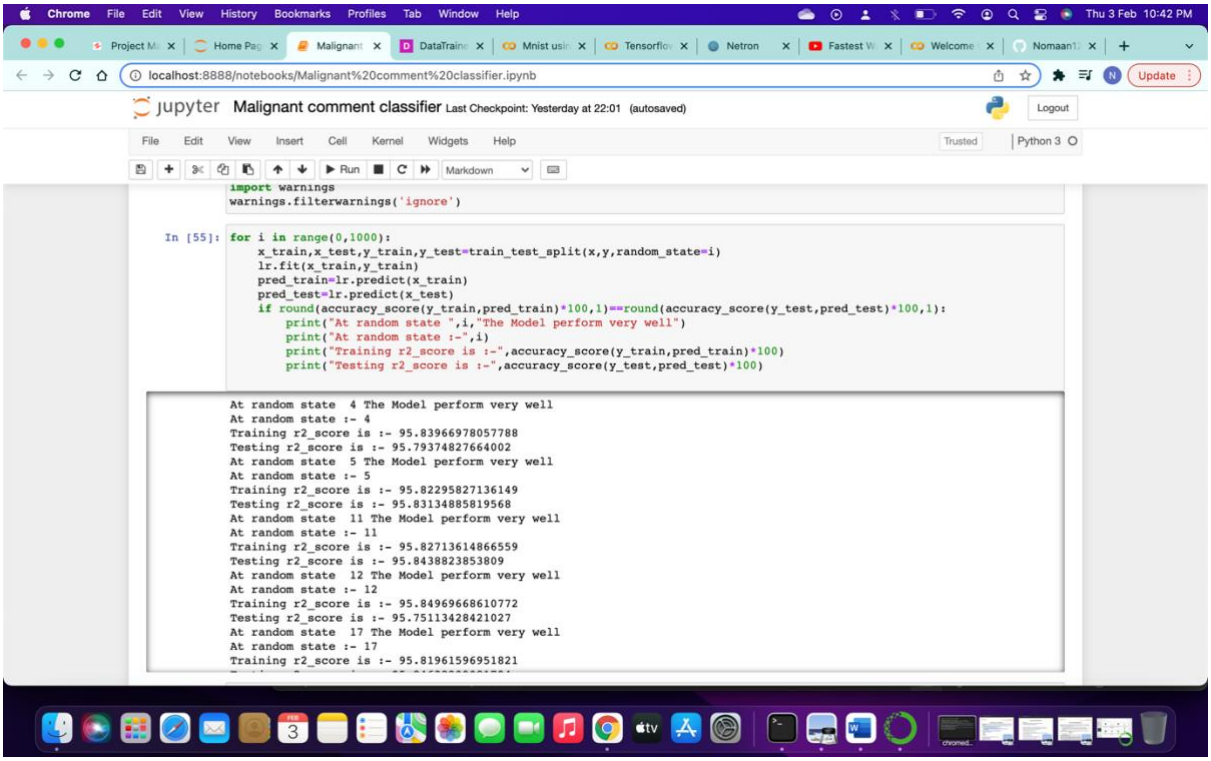
5- Describing dataset
6- Correlation of the dataset
7- Heatmap
8- Checking outliers and removing it
9- Transforming dataset
10- Scaling dataset for better understanding

- State the set of assumptions (if any) related to the problem under consideration
  No such assumptions are done.
- Hardware and Software Requirements and Tools Used

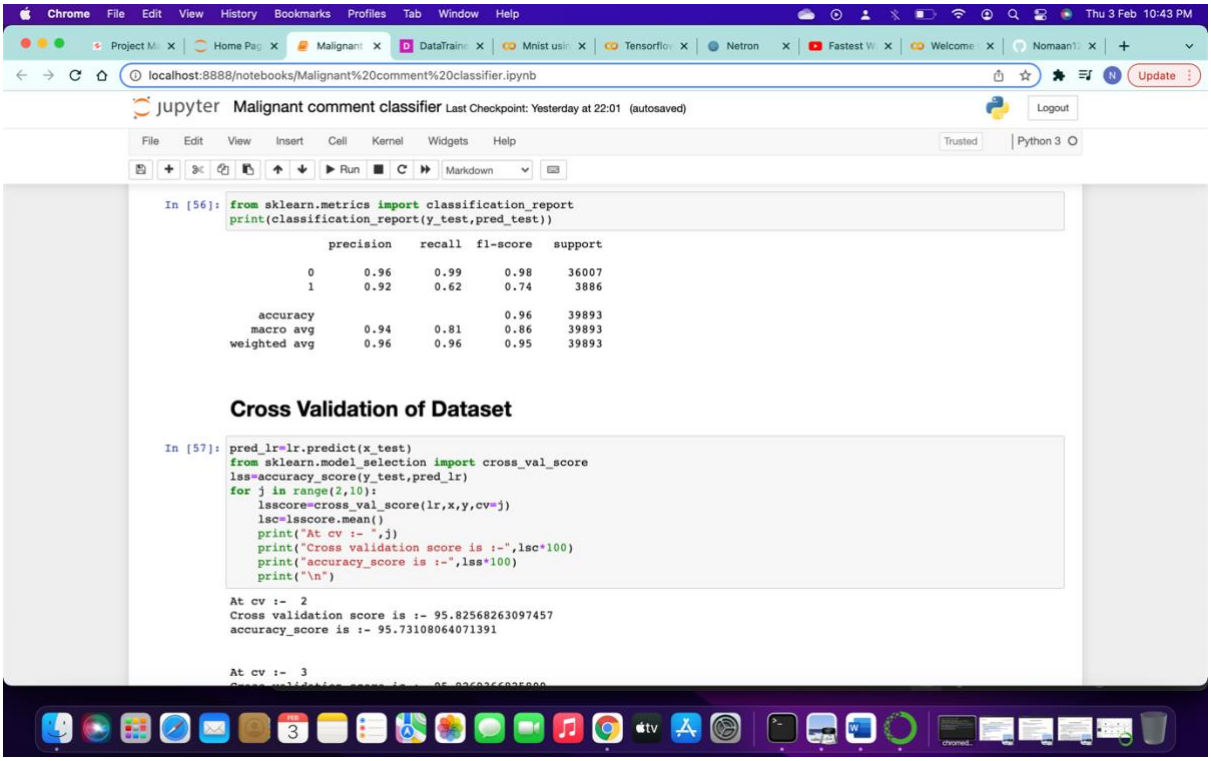
Hardware- core i5 9th gen with 8gb ram
Software-Jupyter Notebook by using Python

Model/s Development and Evaluation

- Testing of Identified Approaches (Algorithms)
  Logistic Regression is used
- Run and Evaluate selected models



- Key Metrics for success in solving problem under consideration



- Visualizations
- Interpretation of the Results

Chrome File Edit View History Bookmarks Profiles Tab Window Help Thu 3 Feb 10:44 PM

localhost:8888/notebooks/Malignant%20comment%20classifier.ipynb

jupyter **Malignant comment classifier** Last Checkpoint: Yesterday at 22:01 (autosaved)

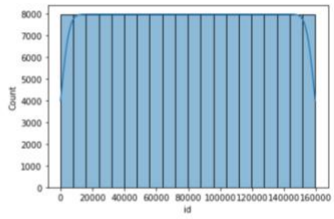
Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

In [15]:
```
sns.histplot(x='id',bins=20,data=df,kde=True)
```

Out[15]: <AxesSubplot:xlabel='id', ylabel='Count'>



In [16]:
```
sns.histplot(x='id',data=df1,bins=20,kde=True)
```

Out[16]: <AxesSubplot:xlabel='id', ylabel='Count'>

Chrome File Edit View History Bookmarks Profiles Tab Window Help Thu 3 Feb 10:44 PM

localhost:8888/notebooks/Malignant%20comment%20classifier.ipynb

jupyter **Malignant comment classifier** Last Checkpoint: Yesterday at 22:01 (autosaved)

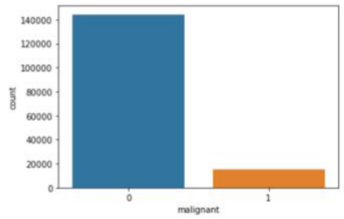
Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

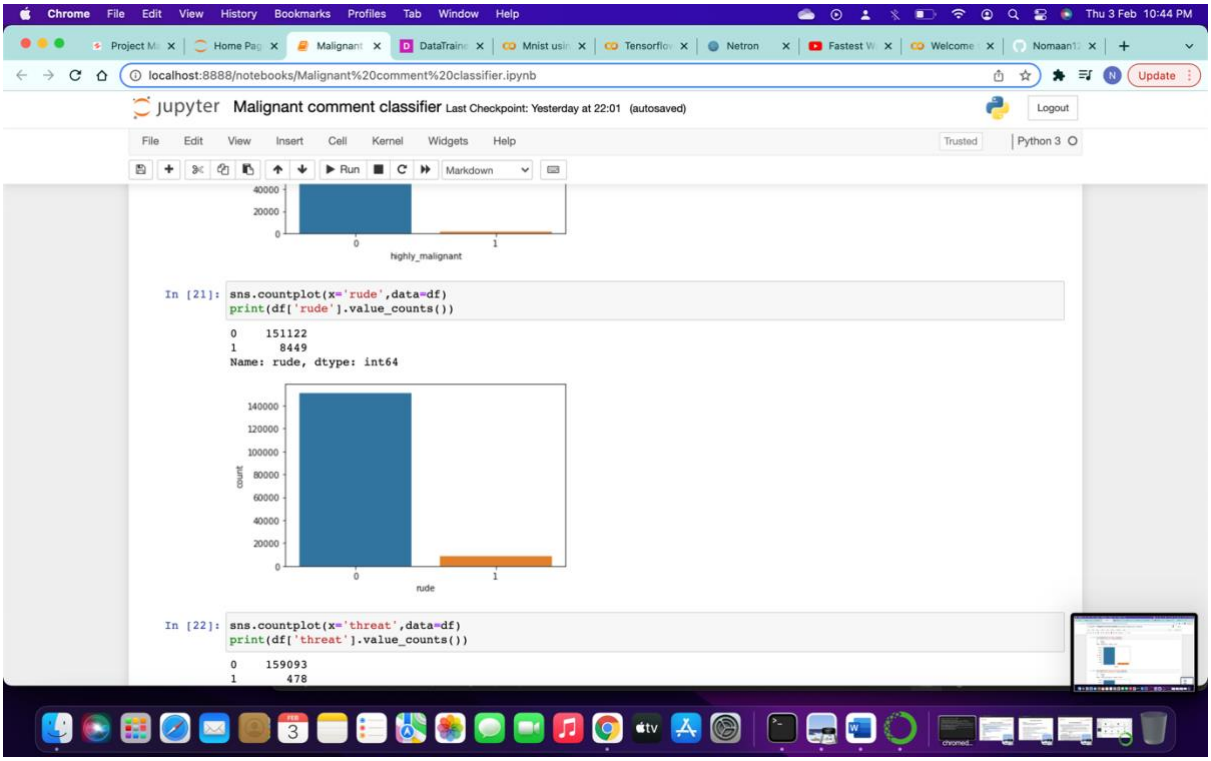
In [19]:
```
sns.countplot(x='malignant',data=df)
print(df['malignant'].value_counts())
```

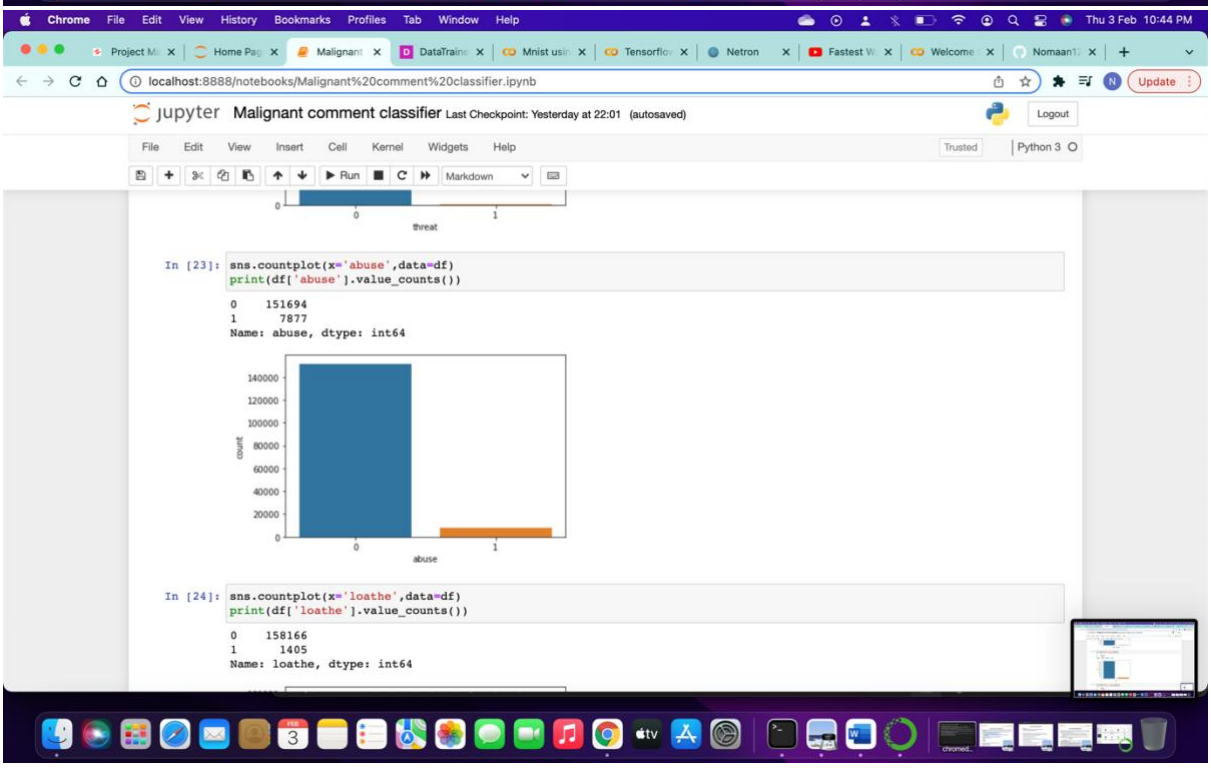
```
0    144277
1     15294
Name: malignant, dtype: int64
```



In [20]:
```
sns.countplot(x='highly_malignant',data=df)
print(df['highly_malignant'].value_counts())
```

```
0    157976
1      1595
Name: highly_malignant, dtype: int64
```


jupyter Malignant comment classifier Last Checkpoint: Yesterday at 22:01 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
Run Markdown
highly_malignant
In [21]:
sns.countplot(x='rude',data=df)
print(df['rude'].value_counts())
0 151122
1 8449
Name: rude, dtype: int64
count
rude
In [22]:
sns.countplot(x='threat',data=df)
print(df['threat'].value_counts())
0 159093
1 478


jupyter Malignant comment classifier Last Checkpoint: Yesterday at 22:01 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
Run Markdown
threat
In [23]:
sns.countplot(x='abuse',data=df)
print(df['abuse'].value_counts())
0 151694
1 7877
Name: abuse, dtype: int64
count
abuse
In [24]:
sns.countplot(x='loathe',data=df)
print(df['loathe'].value_counts())
0 158166
1 1405
Name: loathe, dtype: int64

CONCLUSION

- Learning Outcomes of the Study in respect of Data Science

  By using Data Science we can detect the real time malignant comments done which is very good. We can let our model thinks by using Datascience.
- Limitations of this work and Scope for Future Work

  We can only detect it once it is commented, we cannot stop any person to comment from scratch.