

Malignant Comments Classifier Project

Submitted by:

Gowtham k

**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

This problem helps to detect the toxic comments and rectify it from depression, mental illness, self-hatred and suicidal throughts, etx…

* Review of Literature

First of all we are importing the dictionaries that we are needed for this project and then importing the csv file into the jupiter note book and displaying maximum columns in it,then checking the shape of the data base given and Next checking the datatype given in the data base by using df.info command and made a count for each toxic comments split up and it is listed as per the count

.

* Motivation for the Problem Undertaken

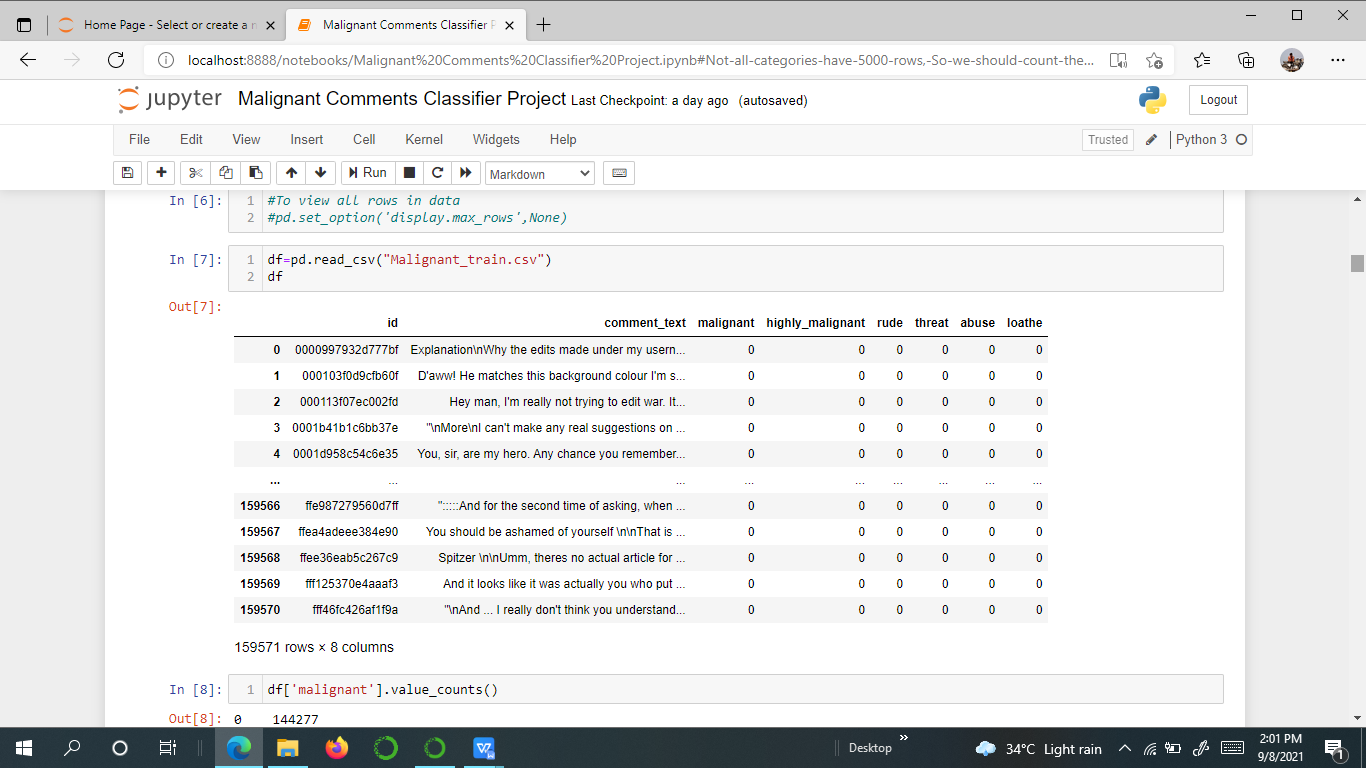
There has been a remarkable increase in the cases of cyberbullying and trolls on various social media platforms. Many celebrities and influences are facing backlashes from people and have to come across hateful and offensive comments. This can take a toll on anyone and affect them mentally leading to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

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Not all categories have 5000 rows, So we should count them first and make them balanced to get accurate result

* Data Sources and their formats

Train data: 

* Data Preprocessing Done

Removing the punctuation marks and alphanumeric formats it the comments

* Data Inputs- Logic- Output Relationships

The input data will not affect the output because we are using training data different and testing data is different

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

Made a count for the toxic comment split up and balancing is done for the data to get the accurate probability output

* Hardware and Software Requirements and Tools Used

Hard ware used is HP Pavilion software used to build the model is Jupiter note book.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

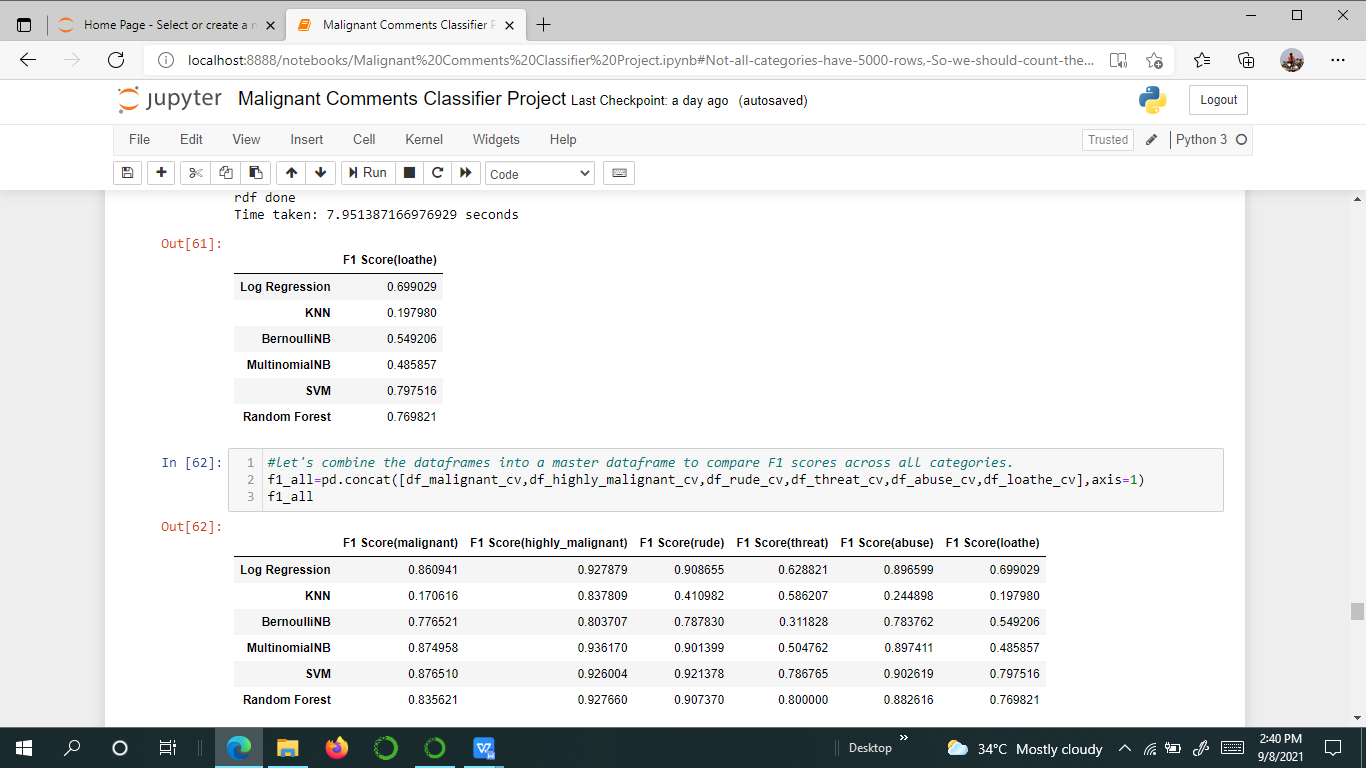
Describe the approaches you followed, both statistical and analytical, for solving of this problem.

Initially cleaned the data by removing unwanted symbols and numbers in it then further balancing the data and applying respective algorithms in it

* Testing of Identified Approaches (Algorithms)

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

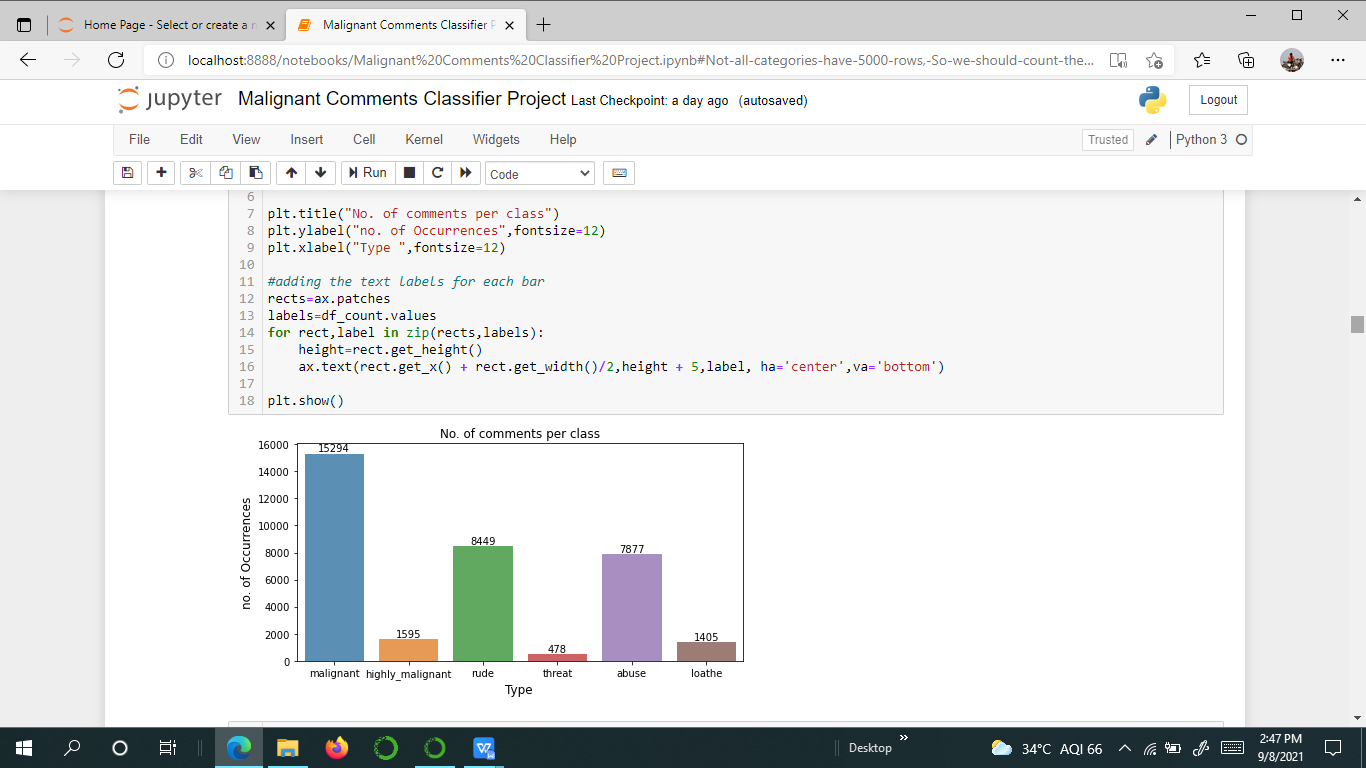
* LogisticRegression
* KNeighborsClassifier
* MultinomialNB
* BernoulliNB
* LinearSVC
* Random forest
* Run and Evaluate selected models

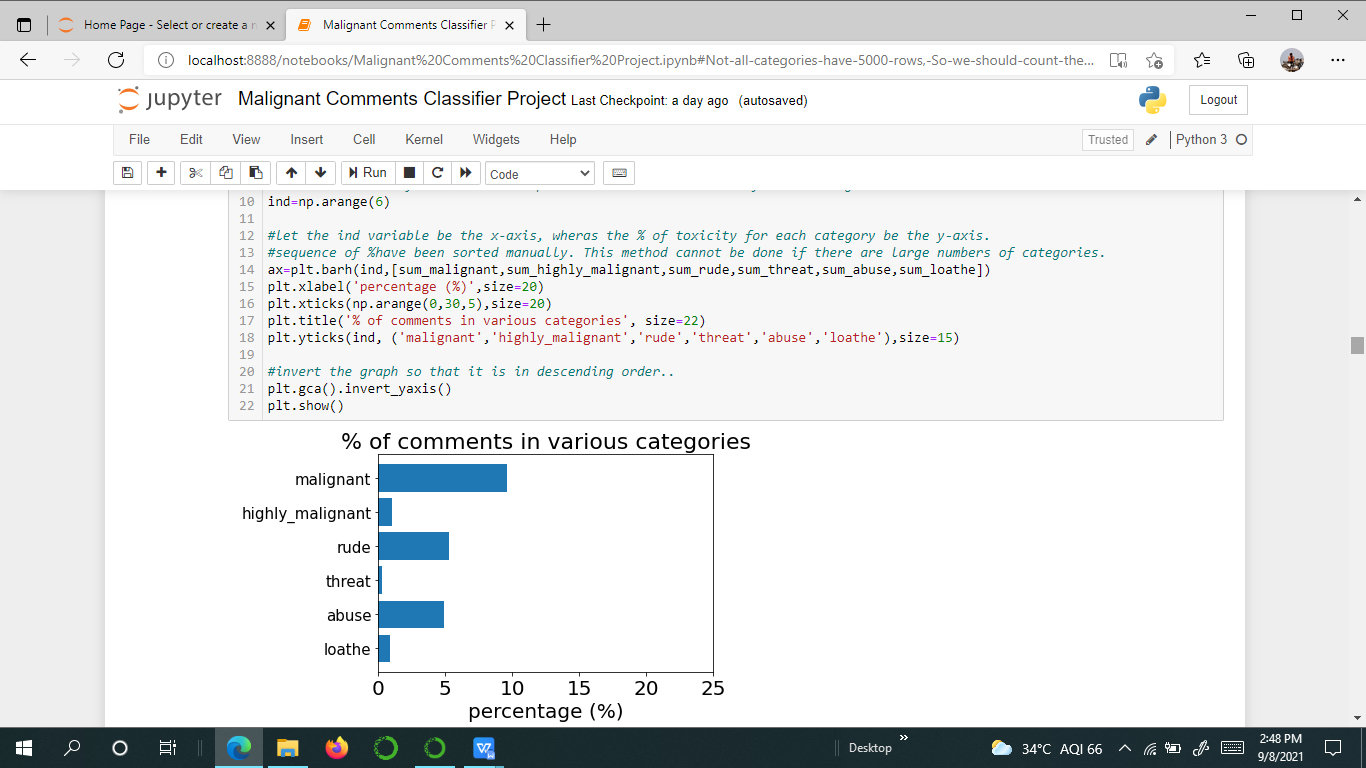


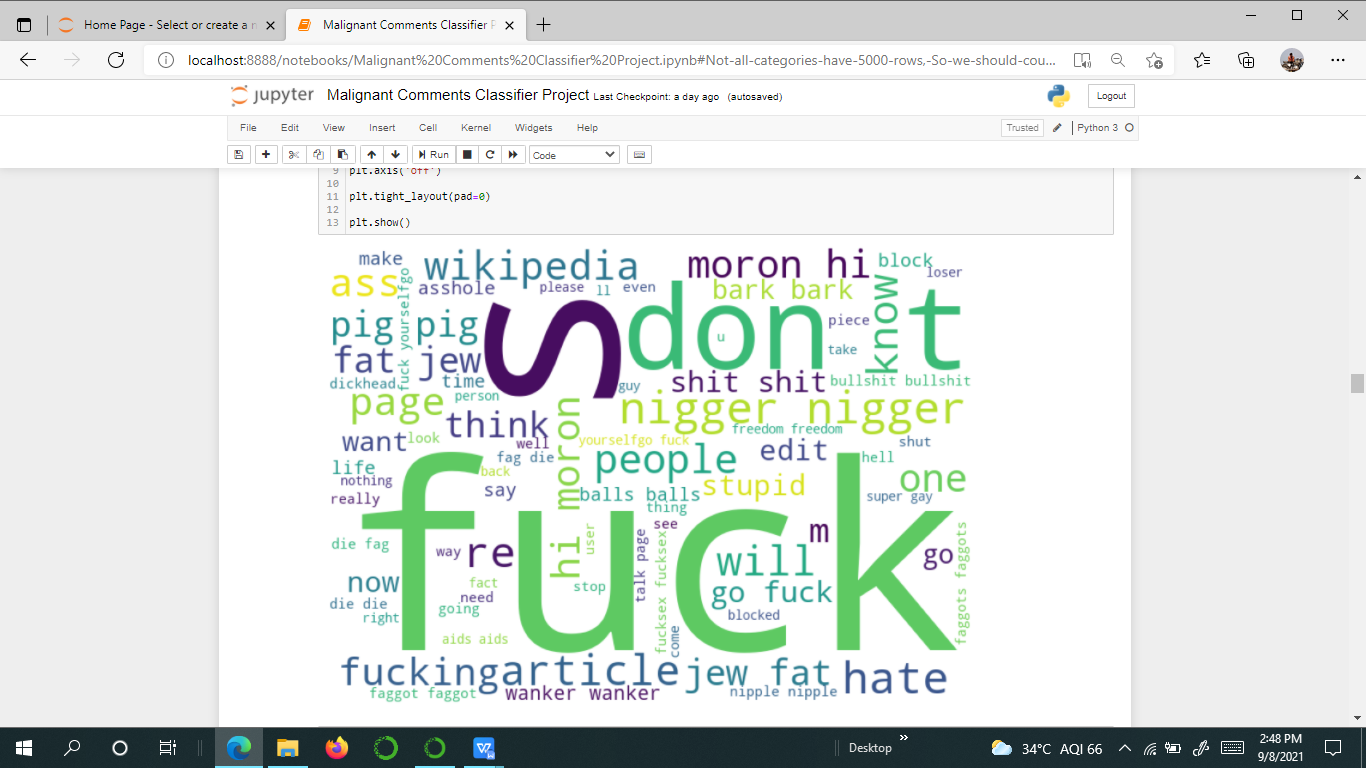
* Key Metrics for success in solving problem under consideration

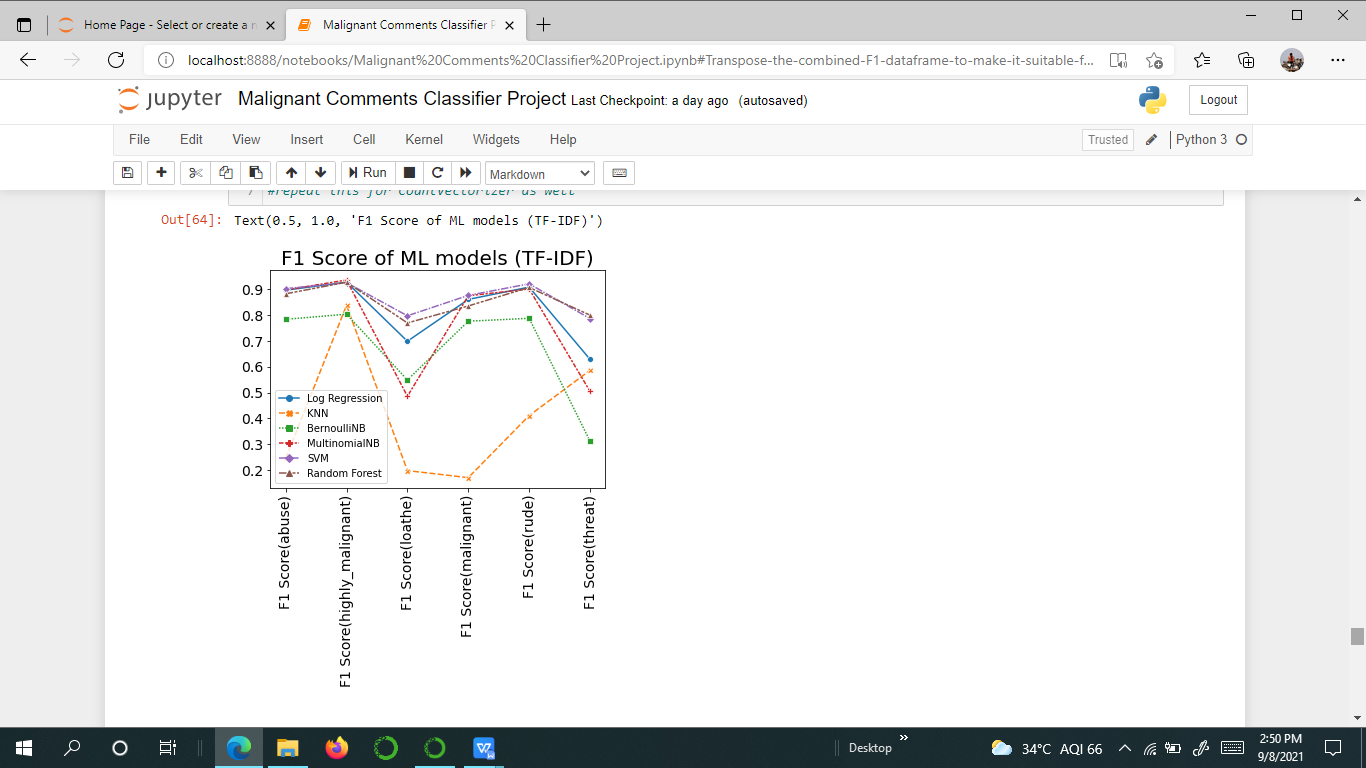
Used six kind of algorithms and getting the six prediction value and by comparing these then choosing the best models

* Visualizations









* Interpretation of the Results

Transpose the combined F1 dataframe to make it suitable for presentation on a graph

**CONCLUSION**

* Key Findings and Conclusions of the Study

Describe the key findings, inferences, observations from the whole problem.

By comparing these algorithms we choose Random Forest Classifier as our final Model because the accuracy is high and less the difference between the accuracy and cross validation, Finally we get an accuracy of 92% for this model, hence it is a very good model.

* Learning Outcomes of the Study in respect of Data Science

Actually handling data is an easier task but holding the correct cleaned data and further we use it in correct form is very much difficult and I faced many problems when I building the model, and displayed the best results

* Limitations of this work and Scope for Future Work

This models can be used in social media platform, It will avoid any violance in the real world activities and people who are using social media platforms they will be getting unwanted comments so this will be the solution to solve, Thank you