Project Name:-Malignant-Comments-Classifier

SUBMITTED BY :-

VARSHA V. SHINDE

INDEX

- ➢ Problem Statements
- ➢ HEATMAP ISNULL
- ➢ Correlation HEATMAP
- ➢ Understanding EDA
- ➢ Exploratory data analysis
- ➢ Malignant plot
- ➢ Highly-Malignant
- ➢ Approach
- ➢ MODEL BUILDING
- ➢ Training, Validation & Test Metrics
- ➢ Model Dashboard
- ➢ Conclusion

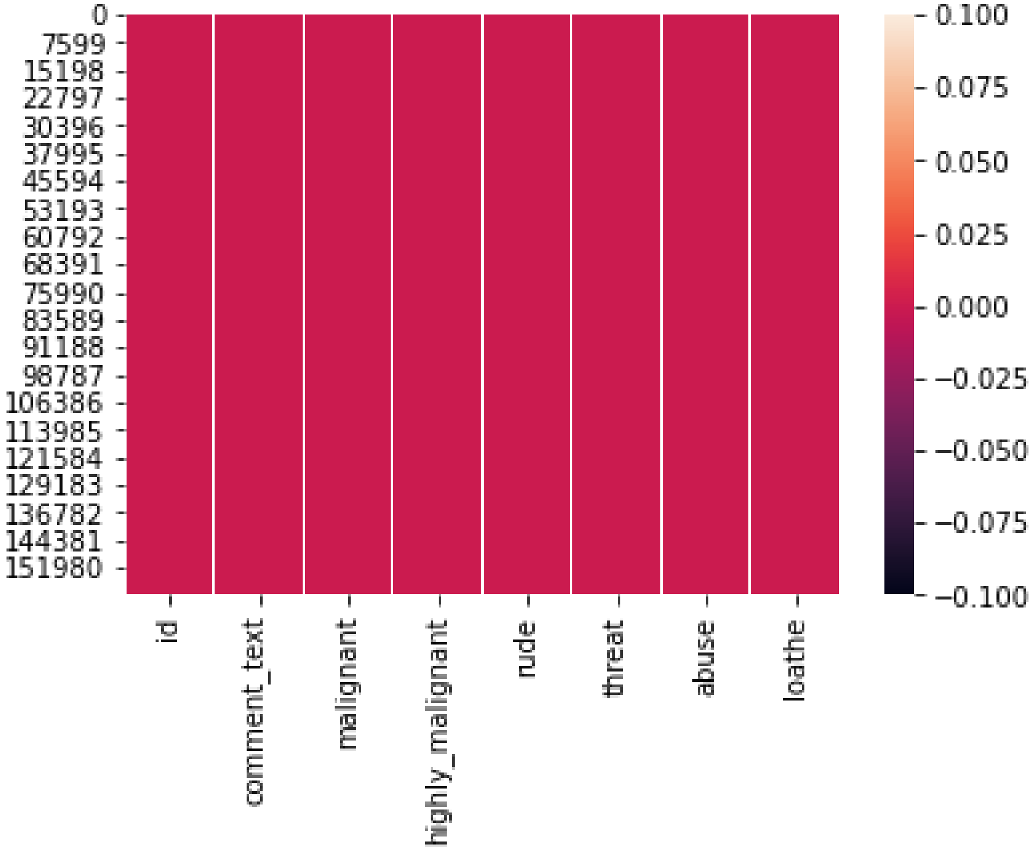
PROBLEM STATEMENTS

- 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 problem across multiple platforms, there is a lack of models for online hate detection.
- Online hate, described as abusive language, aggression, cyber bullying, 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 cyber bullying 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.

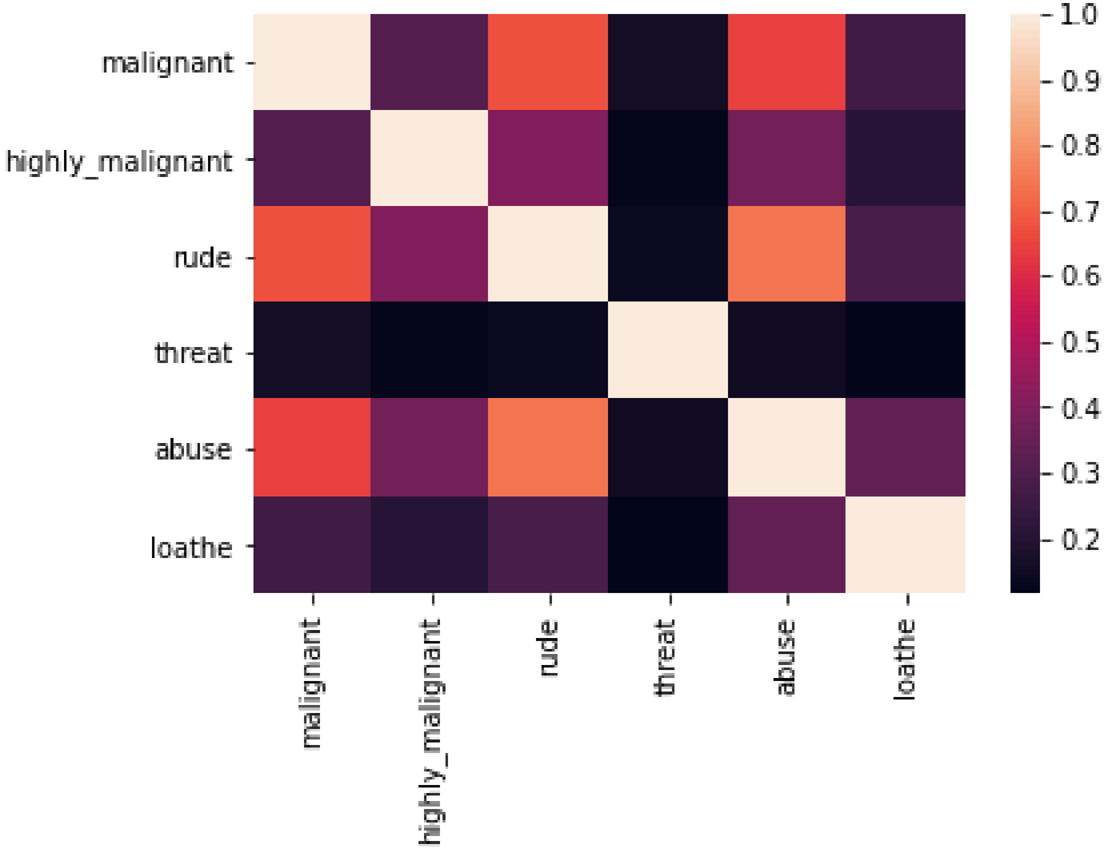
CONTINUE..

- 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.
- 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 cyber bullying.

HEATMAP ISNULL



CORRELATION HEATMAP



UNDERSTANDING EDA

- Data Exploration Among 159,000 comments in the main database, there were about 8% of the cases that were labelled as toxic, while the remaining 92% were labelled as nontoxic. Since normally fewer comments are toxic in every general social media related data set [, the distribution of classes of the data set used in this study is expected.
- Therefore, because the main goal here is to develop an algorithm that is more highly accurate than others using the same data set (or similarly structured data sets), this issue does not create a problem in comparing results to similar methods and algorithms.
- In order to assess the frequency of words size being repeated in two different categories, the overall frequency over number of words was plotted Interestingly, the toxic group were usually shorter than non-toxic and therefore it helped us to focus more on shorter sentences to improve algorithm performance.

EXPLORATORY DATA ANALYSIS

- Exploratory data analysis is a crucial step in the data analysis process. The main aim of EDA is to gain a better understanding of the given data and to analyze their key characteristics. This is achieved by using data visualization techniques.

➢ Steps for Data (Text) Cleaning:

❑ Removing Punctuations and other special/ non-ASCII characters.

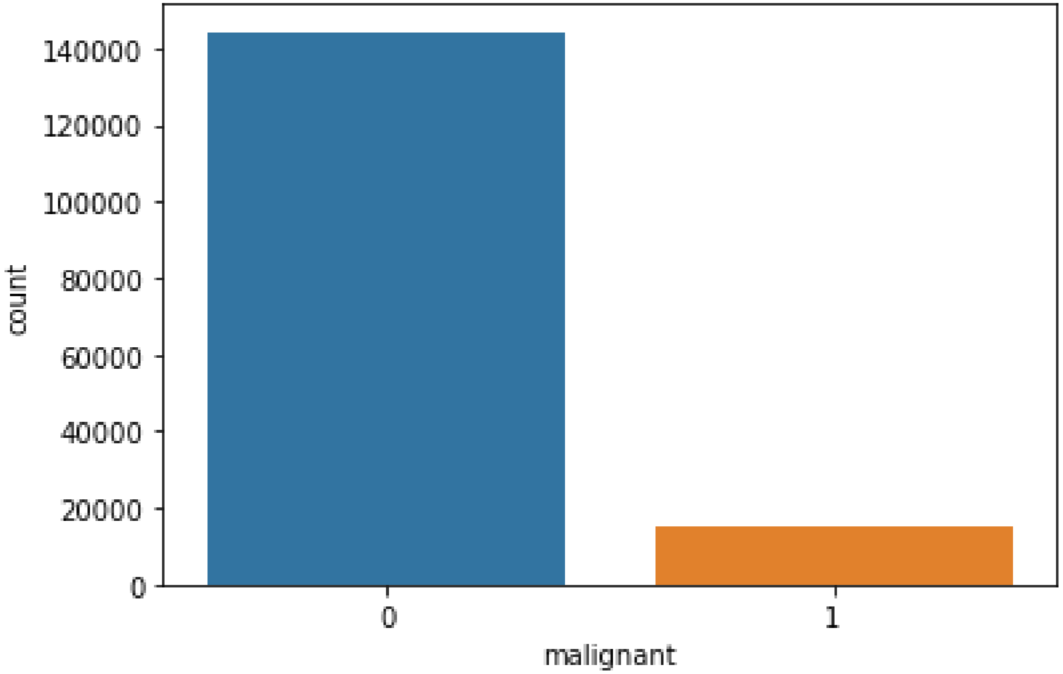
❑ Splitting the comments into individual words.

❑ Removing Stop Words.

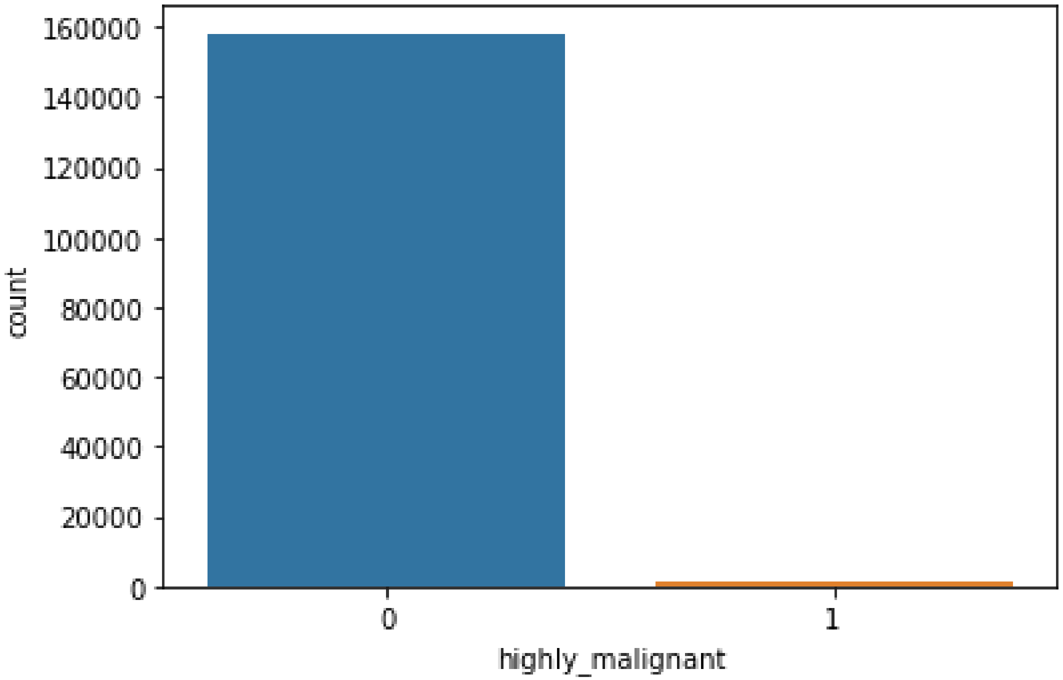
❑ Stemming and Lemmatising.

❑ Stemming and Lemmatising.

MALIGNANT PLOT



HIGHLY-MALIGNANT



APPROACH

- This project focus on studying the effects of three different kind of neural network models (MLP,LSTM, and CNN)at two levels of granularity-word level and character –level-on both binary and multi-label classification tasks.
- We details these approaches, along with our baseline ,below :-
- Tasks:-
- Binary Classification
- Multi-Label Classification
- Baelines:Support Vector Machine
- Neural Network Models
- Multilayered Perceptron
- Long short Term memory
- Convolutional Neural Network

MODEL BUILDING

- .Our basic pipeline consisted of count vectorizer or a tf-idf vectorizer and a classifier.
- we used OneVsRest Classifier model.
- we trained the model with Logistic Regression(LR),Random Forest(RF) and Gradient Boosting(GB) CLASSIFIERS.
- Among them LR gave good probabilities with default parameters.
- So, we then improved the I,R model by changing its parameter.

TRAINING,VALIDATION & TEST METRICS

- To know whether was generalizable or not, we divided the into train and validation sets in 80:20 ratio.
- we then trained various models on the training data, then we ran the models on validation data and we checked whether the model is generalizable or not.
- Also we trained data and we checked whether the model is generalizable or not.

MODEL DASHBOARD

1] SVM

2] Multilayer perceptron

- ❑ Binary Classification
- ❑ Multi-Label Classification

3] Linear Regression

4] Naive Bayes.

CONCLUSION

- Additionally, the followings are some suggested studies to be considered as future work in this area:
- We suggest a plan to improve the NLP classifiers: first by using other algorithms which such as Support Vector Clustering (SVC) and Convolution Neural Networks (CNN); secondly, extend the classifiers to the overall goal of Kaggle competition which is multi-label classifiers. In the current study, the problem simplified into two classes but it worth to pursue a main goal which is 7 classes of comments.
- We also suggest using SVM for text processing and text classification. It requires a grid search for hyper-parameter tuning to get the best results.
- Using Other DNN techniques (CNN)) because some recently published papers such as have shown that CNN proves to have a very high performance for various NLP tasks.