MALIGNANT COMMENTS CLASSIFIER PROJECT

Submitted by:

ANJANA.P

ACKNOWLEDGMENT

I would like to express my appreciation to team Flyprobo for giving such a realistic data for analysis, with a full-length description of the project. My mentor Mr.Harsh Ayush has helped me in many stages of this project where I was stuck with problems. I use this opportunity to thank him for helping me at the right time without any delay.

I also thank DataTrained academy team for their wonderful classes and also their live support team who have been there at any time to help.

Also, this project made me search for a lot of data's in several webpages and sites, that helped me to rectify my doubts and, I was able to study more about data analysis.

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.

Conceptual Background of the Domain Problem

This project is to build a model which can be used to predict whether the comments are malignant,abusive,threatening,abusive etc.Thus we can classify the comments as offensive or not.This may help to control and restrict from spreading hatred and cyberbullying.

Points to Remember:

There are no null values

Output is multi classified as malignant,highly malignant,rude,threat,abuse,loathe etc

Train and test data are separate and contains above 1,50,000 datas in each

There are comments with multiple labels

Review of Literature

First of all both the datas,train and test are saved in a csv file. Then its shape, datatypes, column value counts are all checked to get an outline of the data collected. There are no null values in this dataset.Unwanted column like id is dropped from both train and test data as it provide no information for our analysis.All comment types category are integer datatype and comment texts are in objective datatype.159571 comments are there in train dataset and from this only 35098 comments are offensive type.others are normal messages.Data comments are preprocessed .All punctuations,stop words etc are removed and they are converted to lower cases.Then train and test datas are combined and train test split is done and model is performed.

Motivation for the Problem Undertaken

The main objective behind doing this project is to make an understanding of the types of comments appear in social medias nowadays.It is a realtime problem which everyone faces today.These abusive comments may bring lot of problems like hatefulness,depression etc in the users.This analysis may help to build a model which will classify the comments in different categories, and the user can decide whether to read the offensive comments or can neglect them.

Data Sources and their formats

FlipRobo technologies have provided this dataset for detailed analysis which was collected from various online social medias and sites. The data collected was in an excel sheet with a very detailed description of each columns with it. The data is converted into csv file and loaded in Jupiter notebook first. There are 8 columns and 159571 rows in train dataset and 2 columns and153163 rows in test dataset. The data was in integer and object datatypes. There are multiple output columns in train dataset.They are malignant,highly malignant,abuse,loathe,threat etc.We have to build a model which will predit the type of comments in test data.

ANALYTICAL PROBLEM FRAMING

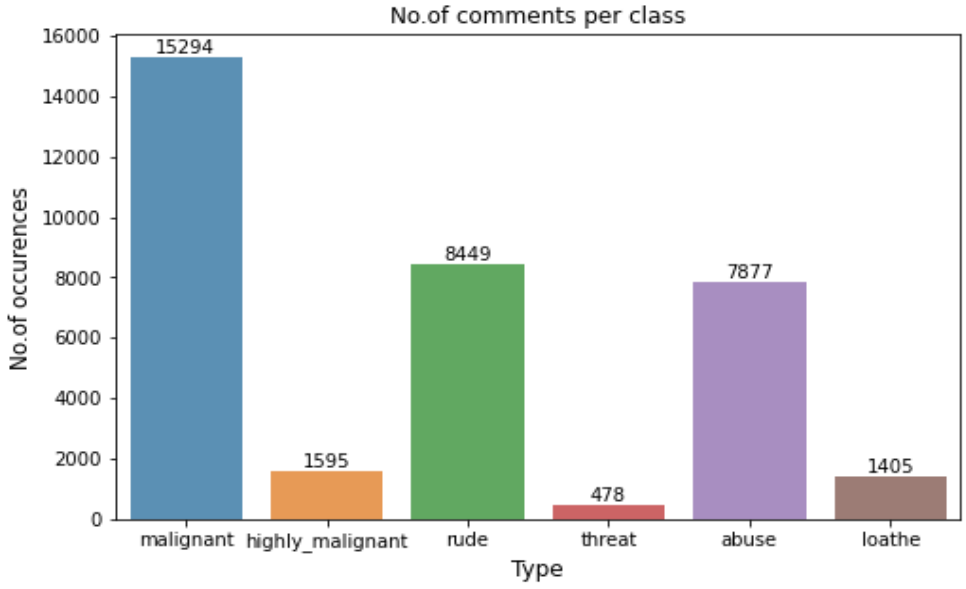
Data Preprocessing Done

First the important libraries for preprocessing and also csv file for analysis is imported. Shape and datatypes of columns are checked. There are 159571 rows and 8 columns in our train dataset and 153163 rows and 2 columns in test dataset.. comment text is object datatype in both train and test dataset and comment types are integer.Comment types are represented in 0's and 1's.If it shows 1 under a particular category,the comment belongs to that category.. Unnecessary column id is dropped as they provide no necessary information for our analysis. Null values are checked and should be cleared if there is any.

We can see that there is only 2 unique values in comment types,that means only 0's and 1's are present in these columns.Describe functions provide information with minimum,maximum,mean,std.deviation,25th percentile,50th percentile 75th percentile of each column. Mean is higher for malignant_comments and lower for threat comments.Also for all the columns minimum is 0 and maximum is 1.

Data Inputs- Logic- Output Relationships

Data counts of each column gives an idea that ,from 159571 comments, only 35098 comments are offensive.All other messages are normal in nature. From these offensive comments malignant comments are more than others and threat comments are the least.There are 15294 malignant comments,8449 rude comments, 7877 abuse comments, 1595 highly malignant comments ,1405 loathe comments and 478 threat comments.These number of comments are plotted in bar plots.It is as shown below.



Then both the comment texts in train and test datasets are converted to lower cases and the data is cleaned by removing all the punctuations, urls, email id's, numbers ,stopwords etc..
Word cloud is used to display the most frequent words appeared in our datasets.

Hardware and Software Requirements and Tools Used

I used intel core i3 processor, 4GB RAM and 64 bit operating system as hardware and windows 10, MS excel, MS word and python 3 Jupyter notebook as software for the completion of this project. In jupyter notebook various libraries are also used. They include pandas, numpy, matplotlib , seaborn , imblearn and sklearn.

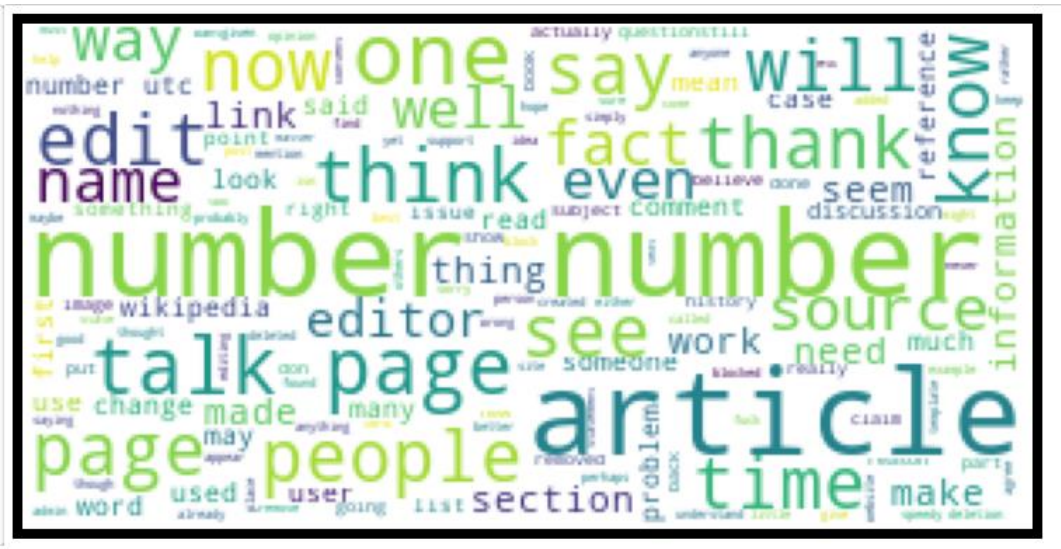
MODEL/S DEVELOPMENT AND EVALUATION

Identification of possible problem-solving approaches

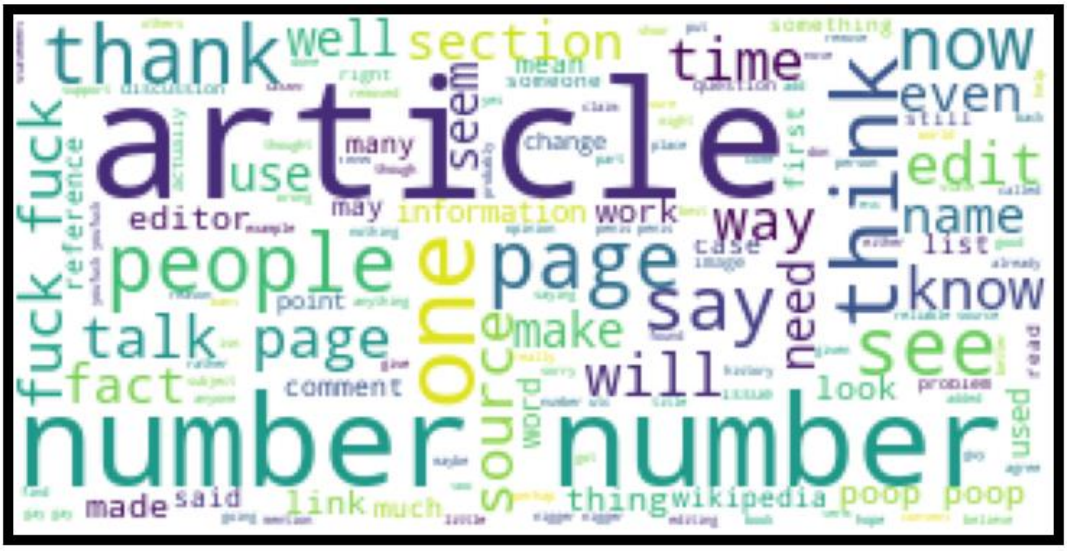
The major problems we dealt with this dataset. They are.

1. Too large dataset –.Both the train and test datasets are too large. So it was impossible to concatenate them to predict the model. The model shows error due to too large dataset. So algorithm was performed separately to avoid this problem
2. Multi classification output columns- There is multiple output columns in the dataset. So maximum function is used to print the maximum of all the output columns and a new column is inserted to print this new output.

WORDCLOUD



Word cloud of train dataset.Here it will display the most frequent words appearing in train dataset comments



Word cloud of test dataset.Here it will display the most frequent words appearing in test dataset comments.