

# MALIGNANT COMMENTS CLASSIFIER PROJECT

SUBMITTED BY  
PROMILA

# INTRODUCTION

## BusinessProblemFraming

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.

## ConceptualBackgroundoftheDomainProblem

As now a days everyone is using social media and aware About the scenario that which type of comments people do it may be abusive, rude or anything else . Moreover to handle the data it needs to do pre processing of text data.

But to understand the problem domain it is not a out of box topic now a days.

## Review of Literature

The research starts with collecting the comments from different social media platforms in order to restrict and stop the increasing hatred through social media platform. There has been a remarkable increase in the cases of cyberbullying and trolls on various social media platforms. Many celebrities and influencers 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 inoffensive, but “u are an idiot” is clearly offensive.

## Motivation for the Problem Undertaken

Motivation behind this problem works in two ways first is to learn a kind of new problem statement and how to work on that also how to handle this size of dataset and which technique to use for that. And secondly by this model we can identify that how cyber bullying and hatred can be stop or restrict.

# Analytical Problem Framing

## Mathematical/Analytical Modeling of the Problem

The data consist of 1.5 lakh rows (approx.) and 8 columns. The large number of rows is a problem in terms of model building and data cleaning .

Also it is multi label classification problem it is important to evaluate that which model to use.

## Data Sources and their formats

The data is collected from the different social media platform to analysis the type of comments and hate in that environments.

And based on the data created six type of comments that are Malignant, Highly malignant , Abuse , Rude , Loathe , threat.

I have first read the data and converted it into the data frame.

```
# Loading data
```

```
data = pd.read_csv("train.csv", dtype=str)
```

In [3]:

```
punct=string.punctuation
```

In [4]:

```
data
```

Out [4]:

	id	comment_text	malignant	highly_malignant	rude	threat	abuse	loathe
0	0000997932d777bf	Explanation\nWhy the edits made under my username Hardcore Metallica Fan were reverted? They wer...	0	0	0	0	0	0
1	000103f0d9cfb60f	D'aww! He matches this background colour I'm seemingly	0	0	0	0	0	0

		stuck with. Thanks. (talk) 21:51, Januar...							
2	000113f07ec02fd	Hey man, I'm really not trying to edit war. It's just that this guy is constantly removing relev...	0	0	0	0	0	0	0
3	0001b41b1c6bb37e	"\nMore\nI can't make any real suggestions on improvement - I wondered if the section statistics...	0	0	0	0	0	0	0
4	0001d958c54c6e35	You, sir, are my hero. Any chance you remember what page that's on?	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...
159566	ffe987279560d7ff	"::::And for the second time of asking, when your view completely contradicts the coverage in r...	0	0	0	0	0	0	0
159567	ffea4adeeee384e90	You should be ashamed of yourself \n\nThat is a horrible thing you put on my talk page. 128.61....	0	0	0	0	0	0	0
159568	ffee36eab5c267c9	Spitzer \n\nUmm, theres no actual article for prostitution ring. - Crunch Captain.	0	0	0	0	0	0	0
159569	fff125370e4aaaf3	And it looks like it was actually you who put on the speedy to have the first version deleted no...	0	0	0	0	0	0	0
159570	fff46fc426af1f9a	"\nAnd ... I really don't think you understand. I came here and my idea was bad right away. Wh...	0	0	0	0	0	0	0

159571 rows × 8 columns

# DataPreprocessingDone

Data pre-processing is most important step to prepare the data for model building as in this problem statement we have only one independent column that is a text columns so following steps I have performed before model building:

## 1. Finding Null values :

There is no any null values in the data set

```
id          0
comment_text  0
malignant    0
highly_malignant  0
rude         0
threat       0
abuse        0
loathe       0
comment_text_new  0
dtype: int64
```

## 2. Removing stop words, punctuations, special characters and urls from the comments :

```
#removing punctuation , special character and urls
```

```
data['comment_textnew'] = data['comment_textnew'].apply(nfx.remove_punctuations)
```

```
data['comment_textnew'] = data['comment_textnew'].apply(nfx.remove_special_characters)
```

```
data['comment_textnew'] = data['comment_textnew'].apply(nfx.remove_phone_numbers)
```

```
data['comment_textnew'] = data['comment_textnew'].apply(nfx.remove_urls)
```

## DataInputs-Logic-OutputRelationships

Based on the comments the comments are classified though there are some common words like fuck , ass etc. which is most common in each kind of category but so far analyzed based on the intensity of the words used it is classifying the

comment.

## HardwareandSoftwareRequirementsandToolsUsed

Tool:Jupyter NoteBook 6.1.4:

Web-based interactive computing notebook Environment.

Software Requirement: The client environment may be Windows, macOS, or Linux.

Hardware Requirement:CPU: 2 x 64-bit, 2.8 GHz, 8.00 GT/s CPUs or better.

Memory: minimum RAM size of 32 GB, or 16 GB RAM with 1600 MHz

DDR3 installed, for a typical installation with 50 regular users.

Libraries:Pandas: For reading CSV file, Converting dataset into a data frame,handling date datatype, and more.

Seaborn and matplotlib: For EDA and Visualization.

# Model/s Development and Evaluation

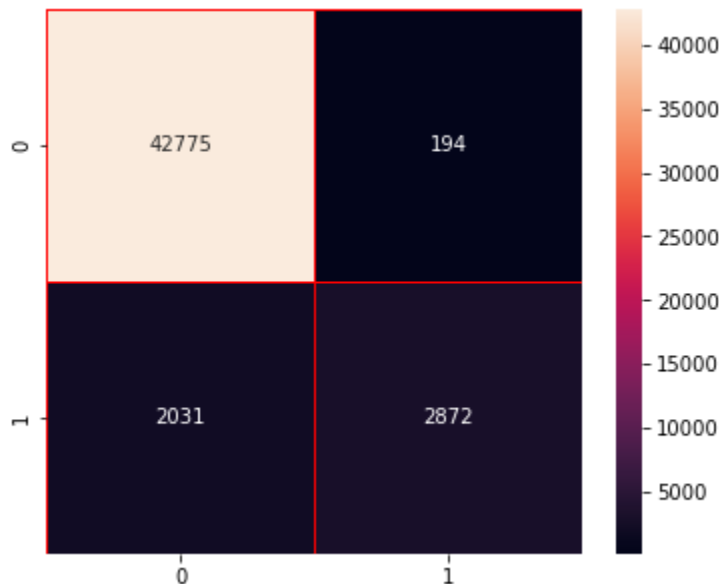
## Identification of possible problem-solving approaches (methods)

Before building of the model data cleaning is important and in this problem I have removed the stop words and Punctuations , special characters and urls from the comment text . Also this steps are very time consuming Because of large data set. Because of large data set I have first divided the data into two parts then perform model building.

I have used OneVsRest classifier and different ensemble techniques for this problem.

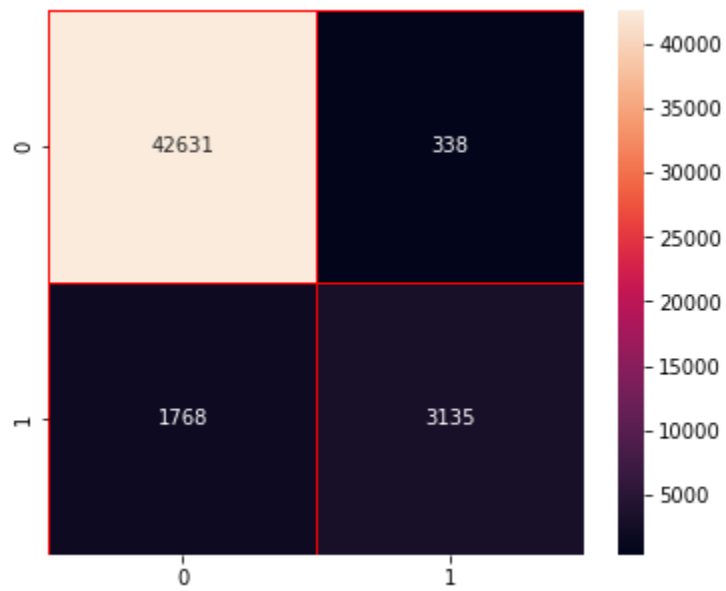
Also Binary relevance and Power set, though the model which is giving Best accuracy is onevsrest classifier with svc.

## Logistic Regression

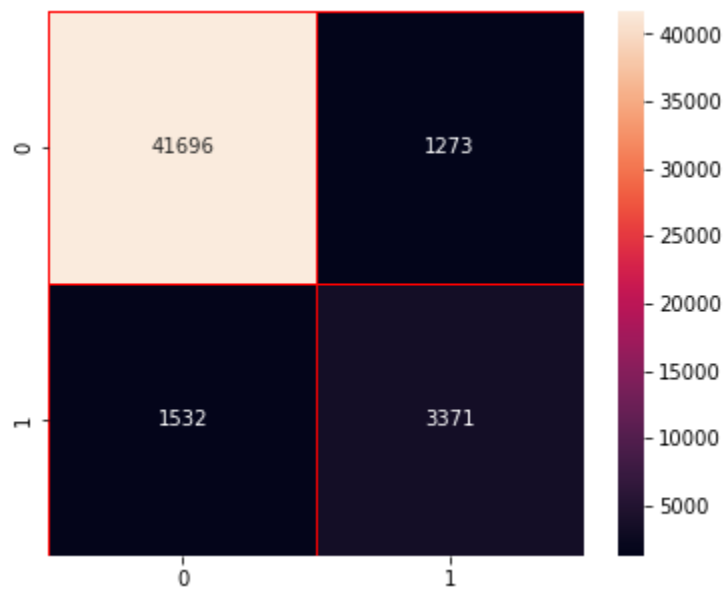




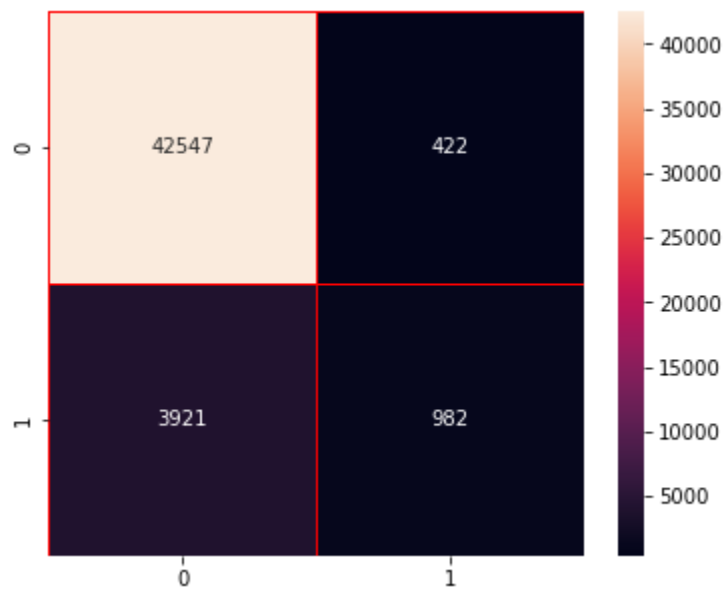
## RandomForestClassifier



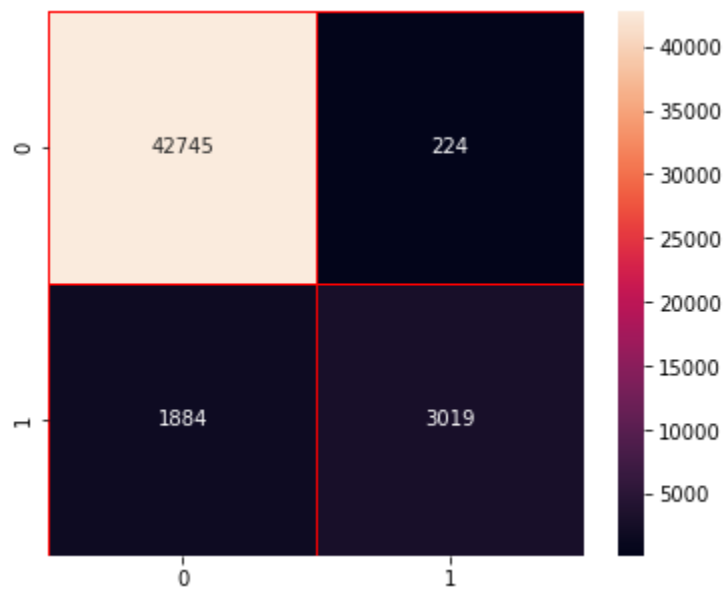
## DecisionTreeClassifier



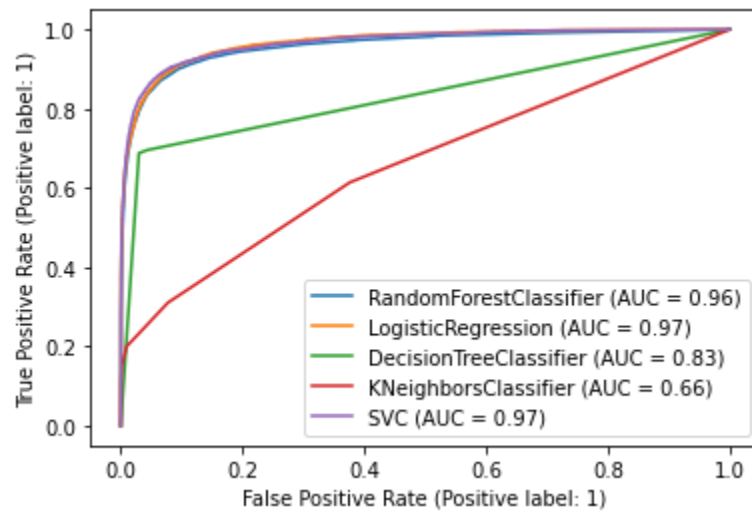
## KNeighborsClassifier



## SVC



# Plot roc curve



# CONCLUSION

## 1.Key Findings and Conclusions of the Study

In the given problem statement the sampling of the data

Is more important also the size of the dataset is a challenge because of it models takes time to build and give the predictions.

More the labels are unbalanced less value for f1 score.

To get more accurate values need to use sampling before modelling

## 2.Learning Outcomes of the Study in respect of Data Science

This problem gives knowledge of different multi label and multi class classification and also how to use different types of metrics and how to treat large size dataset.

Especially in case of word to vector phase size of dataset Matters a lot.

## 3.Limitations of this work and Scope for Future Work

As it is a multi-class classification problem we can explore more Algorithms and hyper parametric tuning to get more accurate results. Also I have not used any hyper parametric tuning because of the large dataset my system did not support though I tried couple of times.

