Malignant Comments Classifier

Submitted by:

ABHISHEK PAI

ACKNOWLEDGMENT

The background information relating to the project was been provided by fliprobo as a part of the internship phase.

The data was collected from various websites to aid this project.

Related guidance was been provided by fliprobo for the completion of this project

INTRODUCTION

- Business Problem Framing

  The goal is to build a prototype of online hate and abuse comment classifier which can be used to classify hate and offensive comments so that they can be controlled and restricted from spreading hatred and cyberbullying.

- Conceptual Background of the Domain Problem

  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

- Review of Literature

  There is not much research performed as the Data and related information was provided by the source itself, which was been taken into consideration based on the information given by Flip Robo.

- Motivation for the Problem Undertaken

  The Project was assigned by flip Robo as part of the internship phase for better understanding the concept and getting the idea of the industry.

Analytical Problem Framing

- Data Sources and their formats

  After loading the data, the information of data was been checked and the samples was been observed.

- Data Pre-processing Done

  The entire data was in form of CSV and was a mixture of numbers, and objects. There are 6 output variables in the pattern of 1 and 0 each having a significant meaning. The output was based on the data which was provided by a source on the behavioural pattern of the entity. The object part was been converted and extracted to perform ML

- Hardware and Software Requirements and Tools Used

  The system with a 16 core processor was been used,
  The operating system was Windows 10,
  Anaconda 3 was been used for performing ML
  Libraries:
  import pandas as pd

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings # Ignores any warning
warnings.filterwarnings("ignore")

import sys
import re

import nltk
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords

import tensorflow as tf
```

Model/s Development and Evaluation

.

- Testing of Identified Approaches (Algorithms)

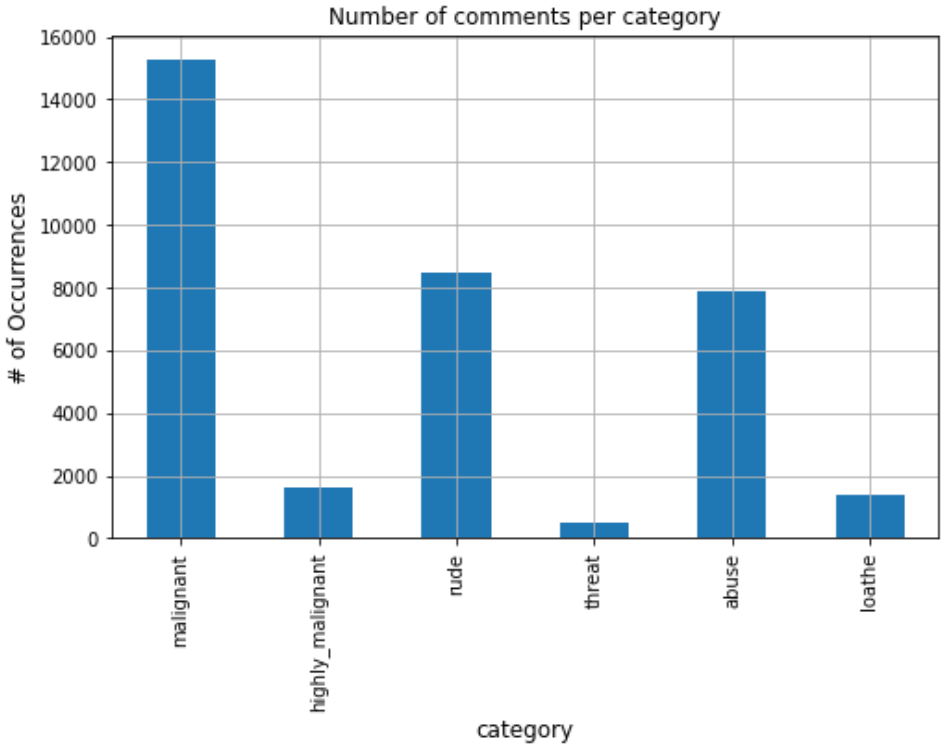
Keras with Bidirectional long-short term memory(bi-lstm) was used which gave 97% accuracy

.

Visualizations

Bag of words

Balanced dataset



	category	number_of_comments
0	malignant	15294
1	highly_malignant	1595
2	rude	8449
3	threat	478
4	abuse	7877
5	loathe	1405

- Interpretation of the Results

  Keras with Bidirectional long-short term memory(bi-lstm) gave the best output and was used which gave 97% accuracy

CONCLUSION

- Key Findings and Conclusions of the Study

  The model performed great and the output obtained had much reasonable accuracy

  This model can help understand real-world situations for the comments which are in the English language

  Learning Outcomes of the Study in respect of Data Science

  Adding more data can help to increase the accuracy.

- Limitations of this work and Scope for Future Work

There is a lot of scopes, more tweaks in a model can help to get better results.