

MALIGNANT COMMENTS CLASSIFICATION

Submitted by:

Lokesh Bisen

**ACKNOWLEDGMENT**

Online forums and social media platforms have provided individuals with the means to put forward their thoughts and freely express their opinion on various issues and incidents. In some cases, these online comments contain explicit language which may hurt the readers. Comments containing explicit language can be classified into myriad categories such as Toxic, Severe Toxic, Obscene, Threat, Insult, and Identity Hate. The threat of abuse and harassment means that many people stop expressing themselves and give up on seeking different opinions.

**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.

* Conceptual Background of the Domain Problem

To protect users from being exposed to offensive language on online forums or social media sites, companies have started flagging comments and blocking users who are found guilty of using unpleasant language. Several Machine Learning models have been developed and deployed to filter out the unruly language and protect internet users from becoming victims of online harassment and cyberbulling.

* Review of Literature

While surfing for some good problem statement on Kaggle, I landed at the Toxic Comment Classification Challenge which was running live then.

The background for the problem originates from the multitude of online forums, where-in people participate actively and make comments. As the comments some times may be abusive, insulting or even hate-based, it becomes the responsibility of the hosting organizations to ensure that these conversations are not of negative type. The task was thus to build a model which could make prediction to classify the comments into various categories.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

I had a dataset of 159571 samples of comments along with their labels. I observed that every 1 in 10 samples was toxic, every 1 in 50 samples was obscene and insulting, but the occurrences of sample being severe-toxic, threat and identity hate was extremely rare.

* Data Preprocessing Done

Data exploration and preprocessing is the first step in data analysis and typically involves summarizing the main characteristics of a data set, including its size, accuracy, initial patterns in the data and other attributes. It is commonly

conducted by data analysts using visual analytics tools, but it can also be done in more advanced statistical software, Python. Before it can conduct analysis on data collected by multiple data sources and stored in data warehouses, an organization must know how many cases are in a data set, what variables are included, how many missing values there are and what general hypotheses the data is likely to support. An initial exploration of the data set can help answer these questions by familiarizing analysts with the data with which they are working.

In the Data preprocessing we have checked null values using pandas library and in order to get better insights from data set used visualization technique using matplotlib and seaborne library. In addition text data might have some numeric and symbols which we have to replace it.

* Hardware and Software Requirements and Tools Used

In order to build a good model it is necessary to have good computational power hardware and software’s. we have used python and its respective libraries. Our data was huge hence we I have used google colab pro as it offers good computational power.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

After exploratory data analysis and preprocessing the main steps comes up that is model building and evaluation is the technique where we make assumptions that which model giving the best accuracy. We have performed some punctuation and stopwords removal. And we have used Lemmatization also, Lemmatization is the process of grouping together the different inflected forms of a word so they can be analyzed as a single item.

**CONCLUSION**

To conclude, In this project we have performed various techniques like Data analysis which has proven way for any organizations and enterprises to gain the information they need to make better decisions, serve their customers, and increase productivity and revenue. The benefits of data analysis are almost too numerous to count, and some of the most rewarding benefits include getting the right information for business, getting more value out of it. And it creates more effective marketing campaigns, gaining a better understanding of customers, and so on. In addition, Stemming is the process of converting inflected/derived words to their word stem or the root form. Basically, a large number of similar origin words are converted to the same word. E.g. words like “stems”, “stemmer”, “stemming”, “stemmed” are based on “stem”. This helps in achieving the training process with a better accuracy.

Lemmatising is the process of grouping together the inflected forms of a word so they can be analyzed as a single item. This is quite similar to stemming in its working but not exactly same. Lemmatising depends on correctly identifying the intended part of speech and meaning of a word in a sentence, as well as within the larger context surrounding that sentence, such as neighboring sentences or even an entire document. I used the word-net library in nltk for this purpose. Stemmer and Lemmatizer were imported from nltk.

Tfidf Vectorizer is used for converting a string of words into a matrix of words. Column headers have the words themselves and the cell values signify the frequency of occurrence of the word.Finally I reached the core part of the project, where I could start building the classifier. I had use 4 major algorithm . Where, our random forest model giving the good accuracy hence selected it and saved it using joblib library.