**Project Proposal**

1. **Title of the Project:-**

**Sentiment analysis of IMDB movie reviews**

1. **Brief on the project:**

We are using dataset having 50K movie reviews for natural language processing or Text analytics. The task is to perform sentiment analysis on this movie reviews.

Sentiments analysis, a data mining technique that is used to measure and understand people's opinion and inclination through NLP. Computational linguistics and text analysis are used to analyze information from the web, social media, and other such sources

1. **Deliverables of the project**

We are using **Multinomial Naive Bayes, Logistic regression to be used for** Sentiment analysis of IMDB movie reviews .we are using Count Vectorizers vs TFID Vectorizers to convert the strings in to vectors

**First step is to clean the dataset**

Which includes

-Clean html tags,

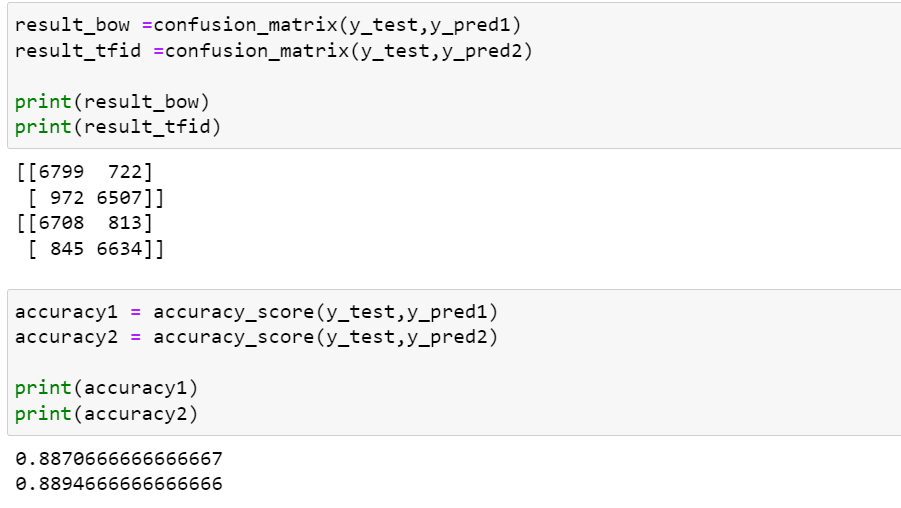
-Convert to lower,

-Remove special characters, r

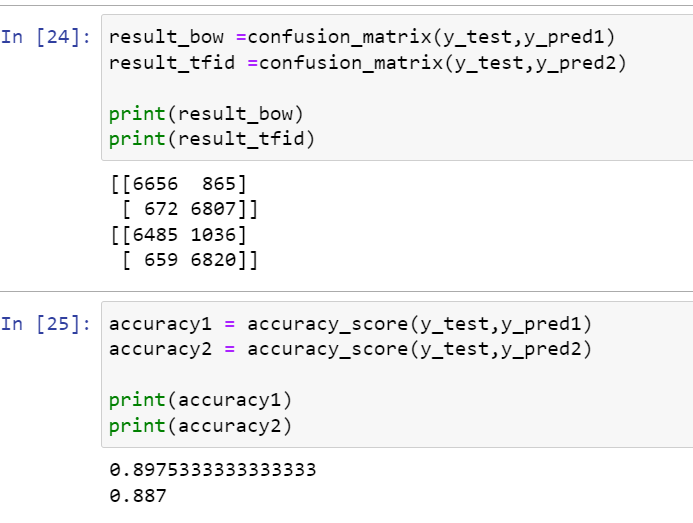
-Remove stop words.

- Count Vectorizers vs TFID Vectorizers to convert the strings in to vectors

Using **Multinomial Naive Bayes, confusion matrix & accuracy of model is as follows**



Using **Logical regression, confusion matrix & accuracy of model is as follows**



1. **Resources**
   * **Data set source:**

IMDB dataset having 50K movie reviews for natural language processing or Text analytics.  
 This is a dataset for binary sentiment classification.

The link for the resource is as

<https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

1. **Software:**

we have used python Software, in Jupiter Notebook that is open-source web-based interactive computing environment.

1. **Individual Details:**

Amandeep Singh

E-mail:- [Amandeep801@gmail.com](mailto:Amandeep801@gmail.com)

M.No:-9810043778