## **Detailed Steps**

**Imports**

Imports common data science and NLP libraries. Also imports tokenizers and stemmers from NLTK.

**Loading data**

Loads the IMDB dataset from a CSV file into a Pandas DataFrame.

**Exploratory Data Analysis**

Checks basic stats about the data like shape, columns, data types, duplicates etc.

**Text preprocessing**

Does initial text cleaning and preprocessing:

* Lowercases the review text
* Removes HTML tags using regex
* Tokenizes into words using WordPunctTokenizer

**Text cleansing**

Further cleans tokens:

* Removes stopwords
* Lemmatizes words using WordNetLemmatizer

**Train-test split**

Splits data into train and test sets for modelling.

**Vectorization**

Vectorizes text data using TF-IDF vectorizer from Sklearn. This converts text into numeric vectors.

**Model training**

Trains 3 models:

* Multinomial Naive Bayes
* Linear SVM
* Logistic Regression

Makes predictions on test set for each model.

**Evaluation**

Multinomial Naive Bayes

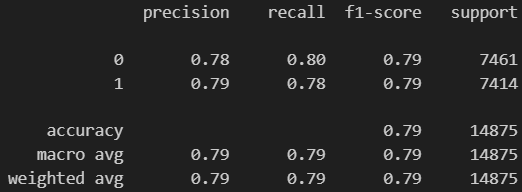
Accuracy Score: **0.787**

Confusion Matrix:

[[1263 65]

[ 213 1919]]

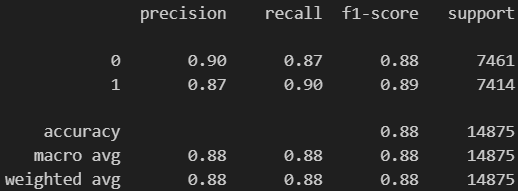
Classification Report:

****

**SVM (Support Vector Machine)**

Accuracy Score: **0.884**

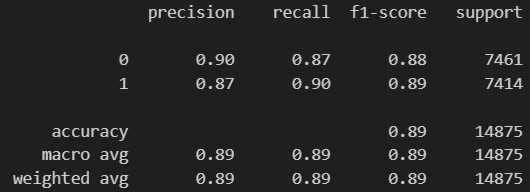
Classification Report:

****

**Logistic Regression**

Accuracy Score: **0.886**

Classification Report:

****

**The Logistic Regression model has the highest accuracy, followed by SVM. Both models achieve over 88% accuracy in classifying positive/negative sentiment movie reviews.Naive Bayes has slightly lower performance compared to the other two.**