## TASK 6

#### **Data Science Example**

# Sentiment Analysis for Movie Reviews Using Python

Name: Nishtha Singh (19BCE0037)

Email: nishtha.singh2019@vitstudent.ac.in

**College:** VIT VELLORE

#### **About The Project**

**Sentiment analysis** is the process of analyzing and categorizing opinions conveyed in a piece of text as positive, negative, or neutral. In this code, sentiment analysis is performed on a dataset consisting of movie reviews and its executed using Python in jupyter notebook.

The dataset used for this task consists of 20 reviews, with the sentiment labels 'positive' and 'negative'.

The algorithms used in this code are Naive Bayes algorithm and MultinomialNB algorithm.

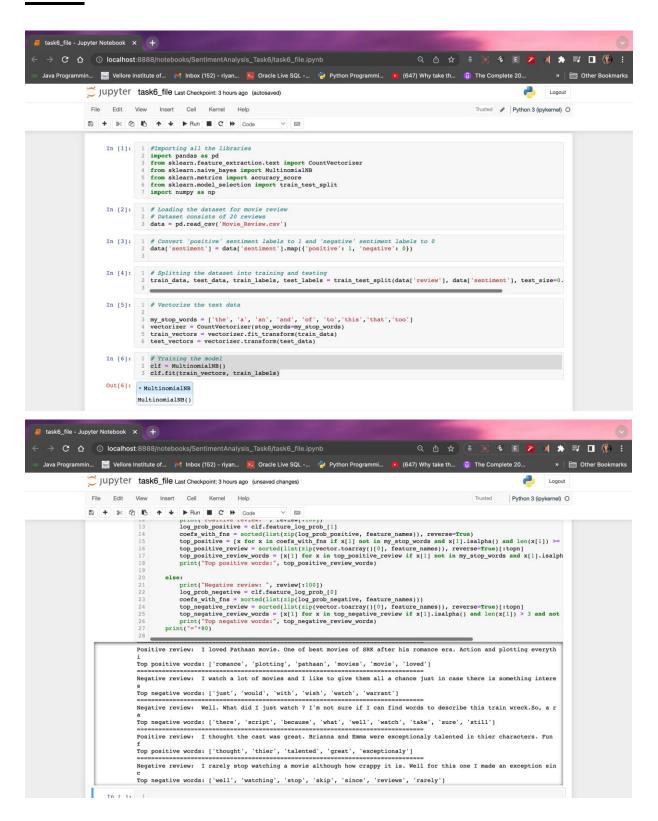
Naive Bayes algorithm is a probabilistic algorithm that uses Bayes' theorem to classify text.

The MultinomialNB algorithm is then used to train the model on the training data. It's designed for text classification tasks and is commonly used in sentiment analysis.

#### **Steps**

- 1. The first part of the code imports the necessary libraries and loads the dataset. The 'positive' sentiment labels are changed to 1 and the 'negative' labels to 0.
- 2. The dataset is then split into training and testing sets, with 80% of the data used for training.
- 3. The text data is then vectorized using the CountVectorizer class from scikit-learn. Then this class is used to convert text data into a matrix of word counts. Stop words are removed from the text data using the 'stop\_words' parameter, which is set to a list of commonly occurring words such as 'the', 'a', 'an', etc.
- 4. The MultinomialNB class is then used to train the model on the training set. The model is then used to predict the sentiment of each review in the dataset.
- 5. For each review, the first 100 words of the review are printed, along with the top positive or negative words depending on the predicted sentiment.
- 6. The top words are obtained by sorting the features according to their log probabilities and selecting the top features. The 'reverse' parameter is used to sort the features in descending order. Stop words are excluded from the top words list.
- 7. At the end, we get the sentiment for every movie review in terms of positive and negative.

#### Code



### **Results**

#### Positive Review with top positive words

Negative words with top negative words