Documentation on Twitter Sentiment Analysis

**Dataset** : https://www.kaggle.com/datasets/kazanova/sentiment140

**Overview :**   
This is the sentiment140 dataset. It contains 1,600,000 tweets extracted using the twitter api . The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment

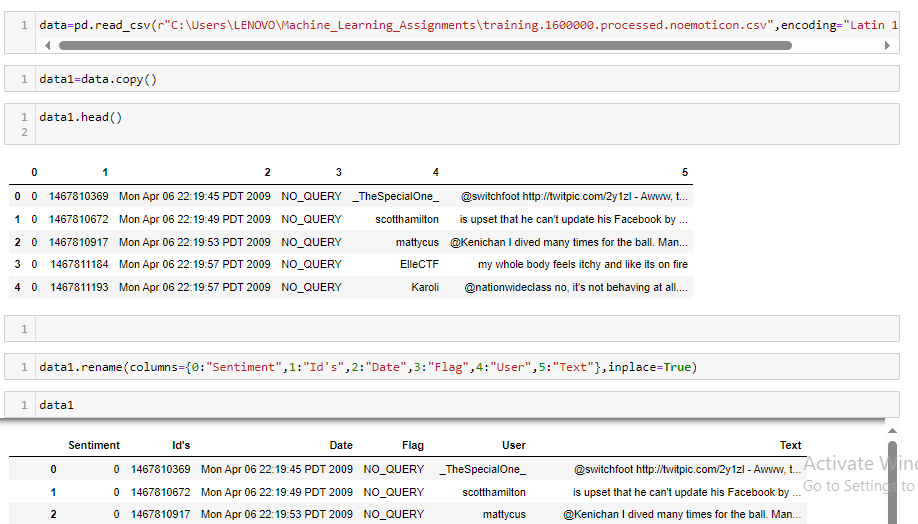
**Problem Statement :** Fit a model to predict the sentiment of the text using Naïve Bayes .

1. **Import Libraries**

**

**Data Collection**

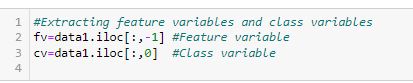
1. Importing the dataset and preview of the data :



Dropped unwanted columns



**Feature Engineering :**

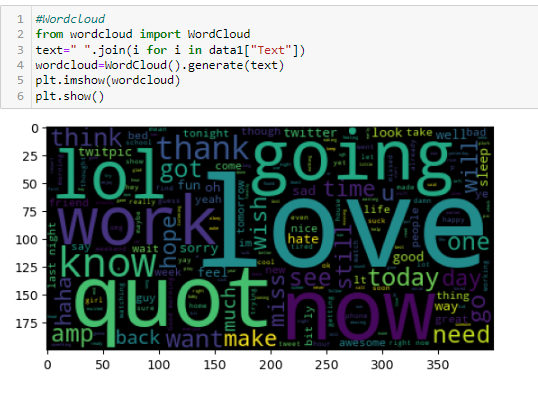
****

Splitting the data to train and test datasets

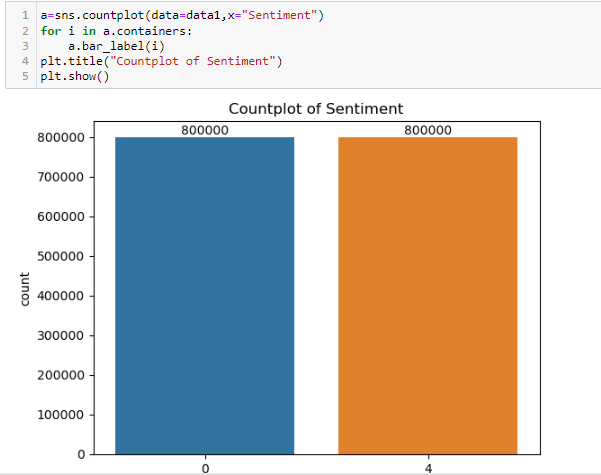


**Exploratory Data Analysis (Before Data cleaning ):**

* Most repeated words :



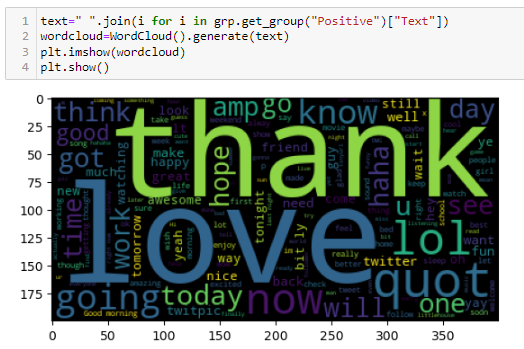
* Countplot of Class label:

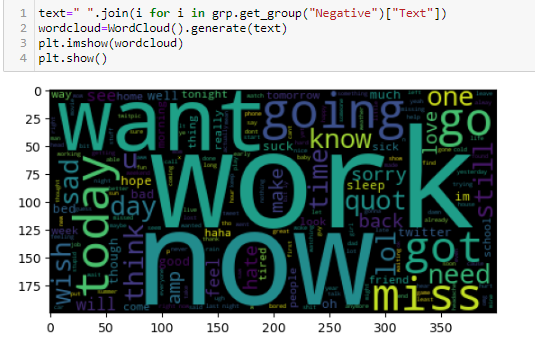


Based on the count plot of class variable, the dataset is balanced.

* Wordcloud for each sentiment

Positive :

  
Negative :



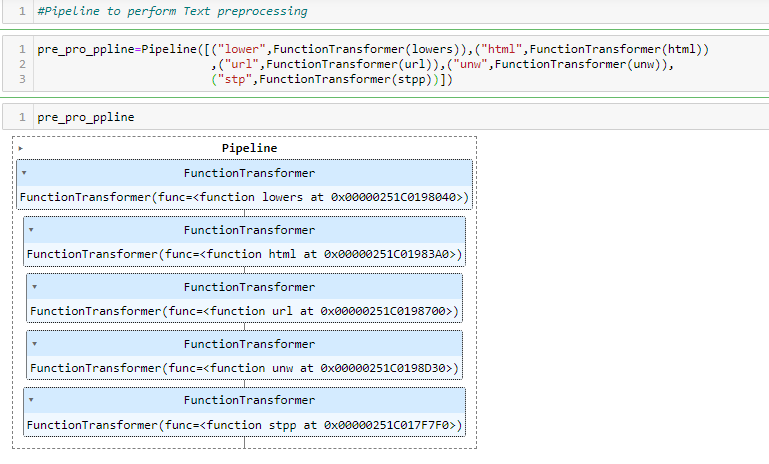
**Pipelines for Data Cleaning / Data manipulation :**

1. Data Cleaning :

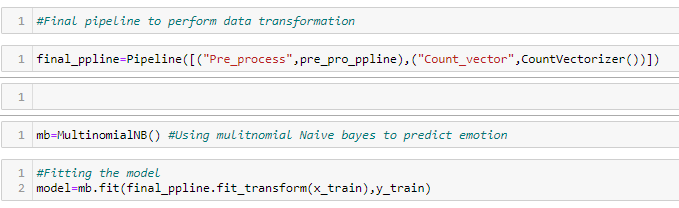
Defined different functions to perform data cleaning and transformation



Pipeline created is

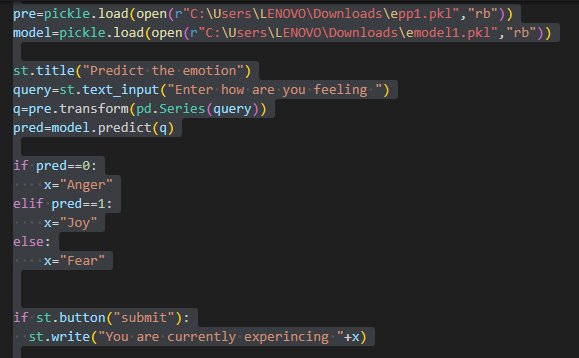


**Used MultinomialBayes to Predict Emotion :**

****

**Deployment :**

Dumped the model and loaded in VSCode to deploy the model



AND the result using streamlit is

