Fake News Detection Using NLP

Phase 3 submission

Problem Statement:

Begin building the fake news detection model by loading and preprocessing the dataset. Load the fake news dataset and preprocess the textual data.

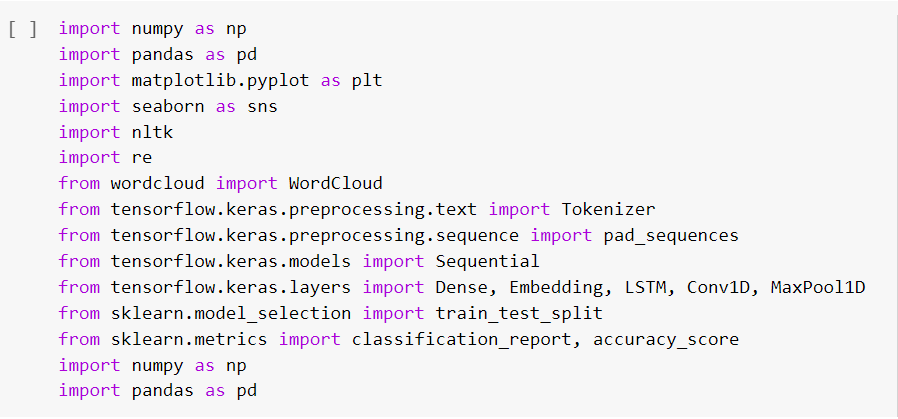
# Data Cleaning:



Data cleaning is a process of removing inconsistencies in the dataset and incorrect values .It also in involves handling missing values either by removing them or assigning them average values. It helps to improve the efficiency of the model.

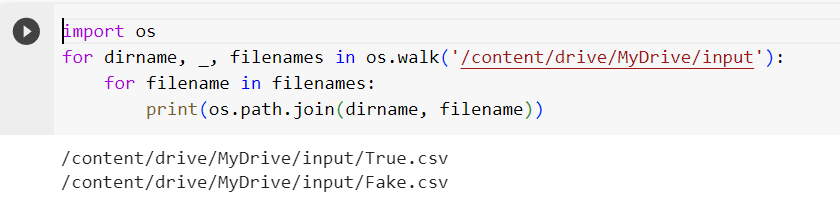
In the first step, we will only remove the unnecessary data points from the dataset which does not helps in improving the model performance.

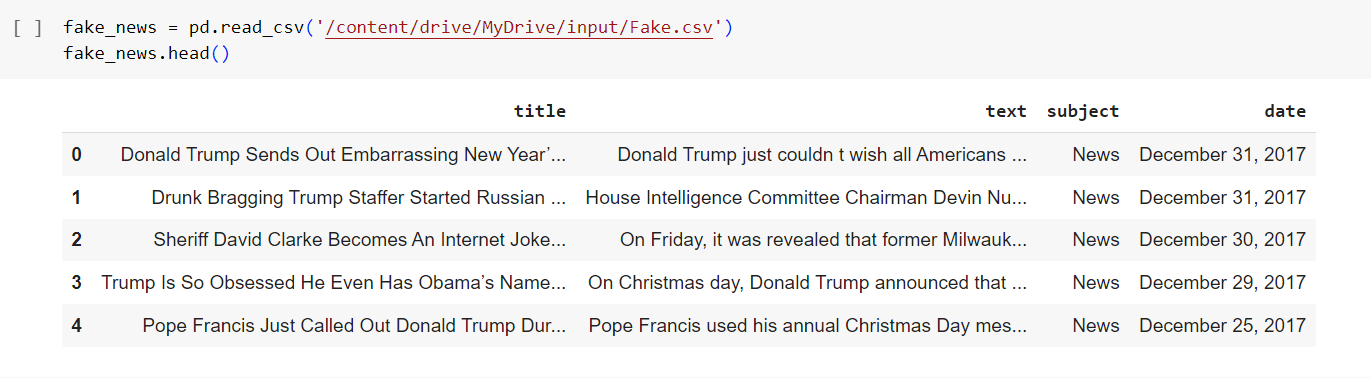
Initially we import the necessary packages for our data cleaning process and also in the future purposes,

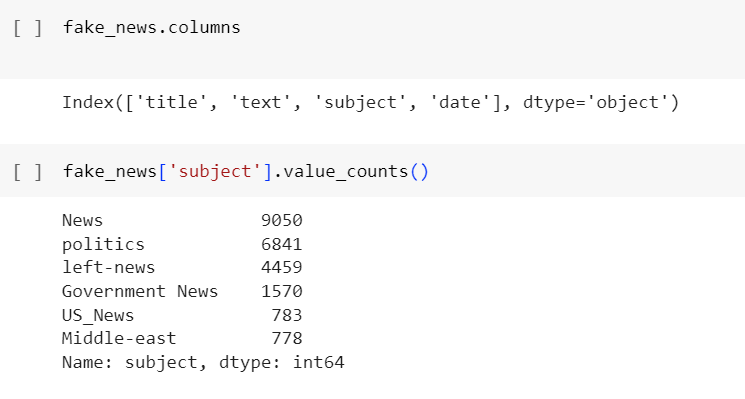


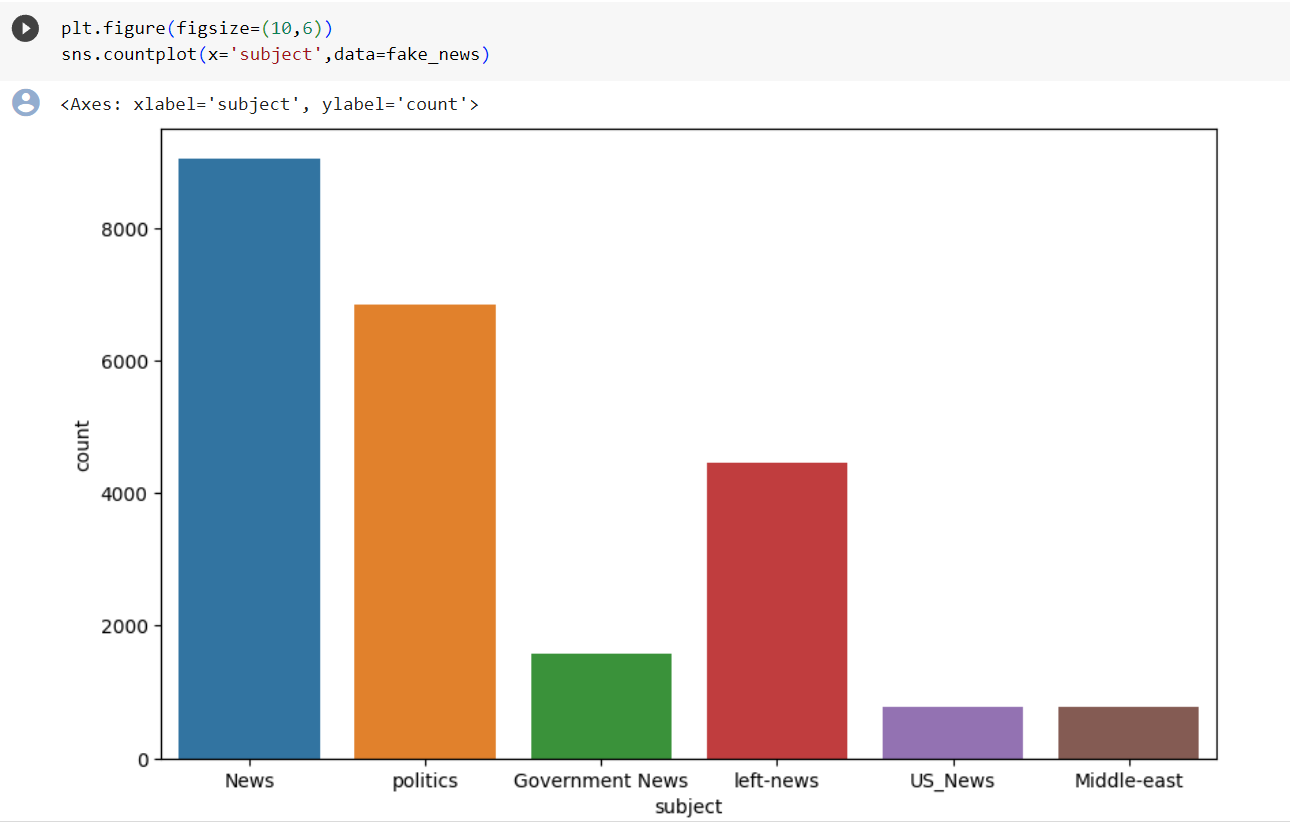
we use these packages in various stages of our cleaning process and also in the future in which we need to build models.

Here, we read the .csv files of true and fake news and then explore the count values of their subjects







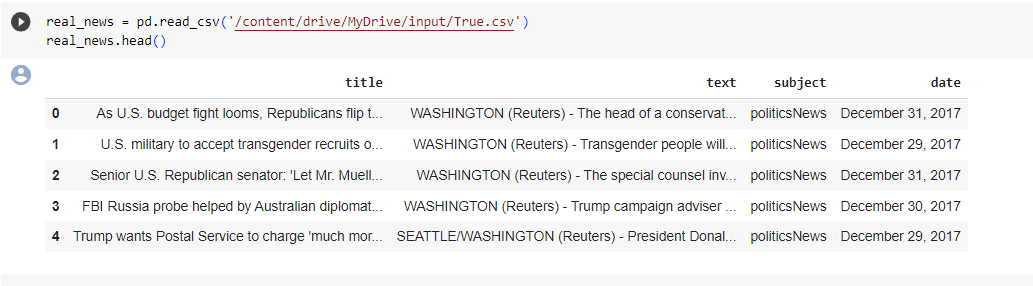


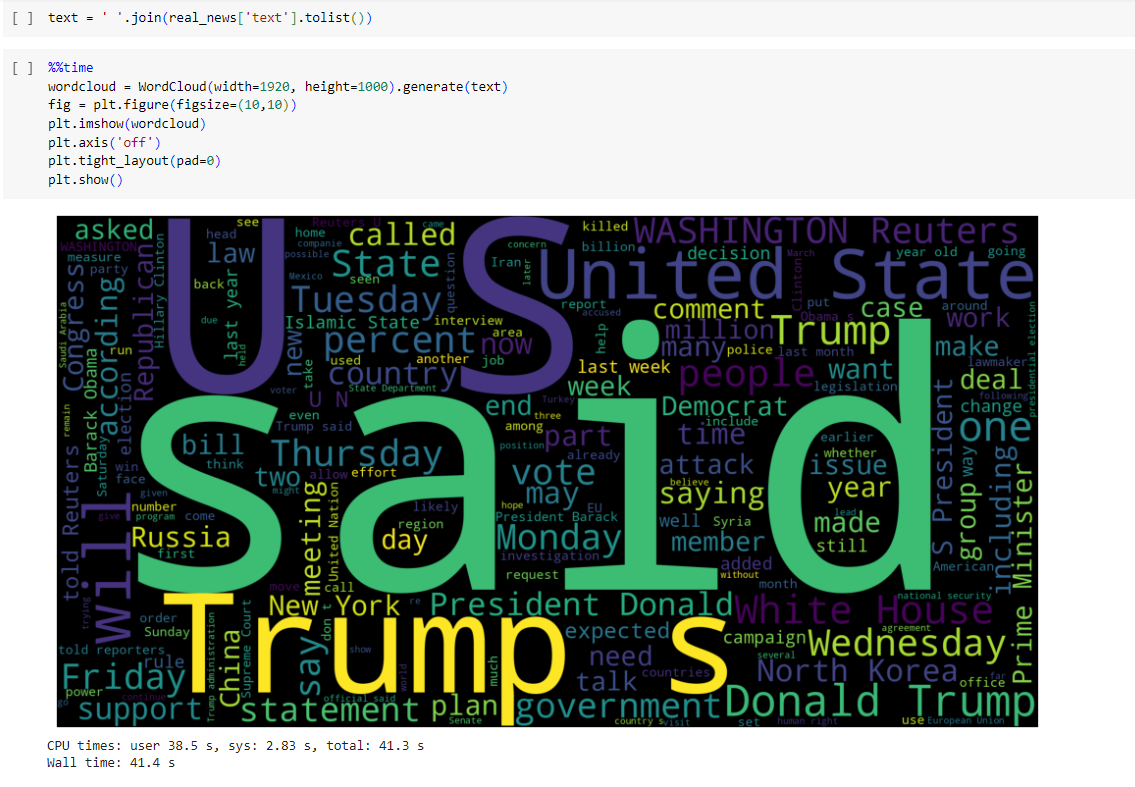
# Here , we have used wordcloud to see that which word has mostly used for the fake news. By seeing that we can make a conclusion that which topic(about a person, event or anything) is mostly contains fake news).We also do the same for true news.

# Word Cloud for Fake News:

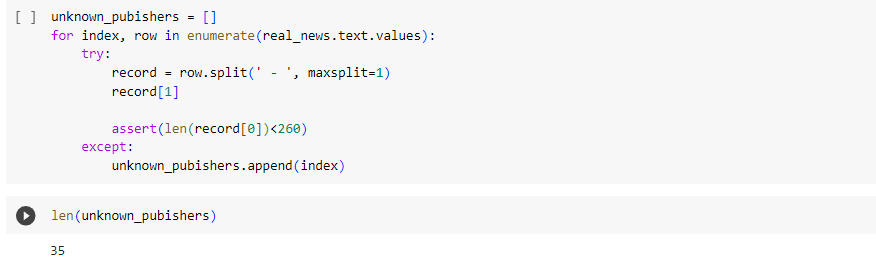


# Word cloud for True News:







Let’s create a list of news lists in real\_news.csv with unknown publishers by using the following code snippets

