Spam filter

احمد هشام فتحي سكشن 1

مجدي ياسر محمود خليل سكشن 3

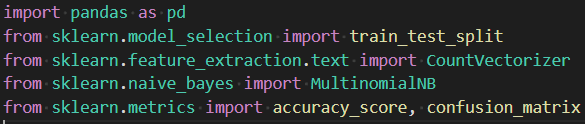
AI project

Spam filter considered to be in Fully observable, Episodic, Stochastic environment

This project indicates whether a given (input) email is a spam or not.

So first we imported these libraries:

* **Pandas**
* **train\_test\_split** from **sklearn.model\_selection**
* **CounterVectorizer** from **sklearn.** **feature\_extraction.text**
* **MultinomialNB** from **sklearn.naive\_bayes**
* **accuracy\_score & confusion\_matrix** from **sklearn.metrics**

****

Then we searched the **Kaggle** website and found our **Dataset** we are going to use:

[Here is the link to our Dataset](https://www.kaggle.com/datasets/karthickveerakumar/spam-filter?resource=download)

Or

[View it on excel here](Dataset/emails.csv)

After we got the **Dataset,** we need to read it in the compiler using

“dataset = pd.read\_csv() function”



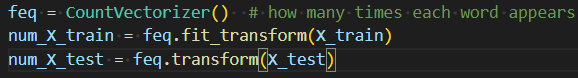
Now we can use the **Dataset** in the compiler using “print(dataset.columns)” we know the columns names(text, spam).

Now we split the **Dataset** using this command

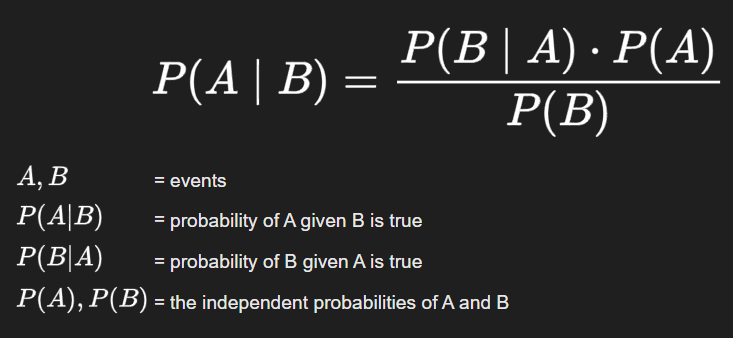
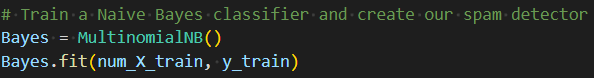
To see the content of each email and to see if it’s a spam or not.

Then we want to train our model using 80% of the data for training which leaves 20% of the data for testing using this command



Now to count the number of words in each email we used **CountVectorizer()** to see the frequency of repeated words to help us detect whether this email is a spam or not

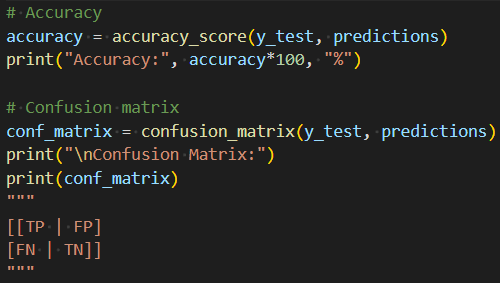
And then train a Naive Bayes classifier

****

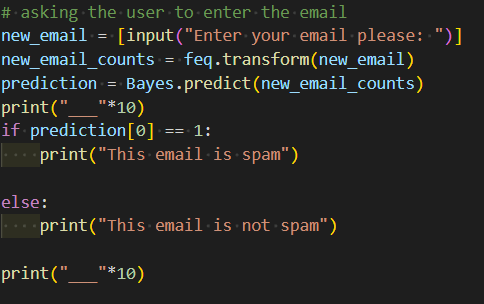
Now making the model predict



Now we want to see the **Accuracy & Confusion Matrix** which indicates how accurate the model is and indicates the **True Positive, False Positive, False Negative, True Negative.**



Now the last thing to do is to ask the user to enter the suspicious email and wait for the model to predict whether is it a spam or not



Here are some outputs:

