In [1]:
```
pip install numpy pandas sklearn
```

```
Requirement already satisfied: numpy in c:\python\python39\lib\site-package
s (1.19.3)
Requirement already satisfied: pandas in c:\python\python39\lib\site-packag
es (1.1.5)
Note: you may need to restart the kernel to use updated packages.Requiremen
t already satisfied: sklearn in c:\python\python39\lib\site-packages (0.0)
Requirement already satisfied: pytz>=2017.2 in c:\python\python39\lib\site-
packages (from pandas) (2020.4)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\python\python3
9\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: six>=1.5 in c:\python\python39\lib\site-pack
ages (from python-dateutil>=2.7.3->pandas) (1.15.0)
Requirement already satisfied: scikit-learn in c:\python\python39\lib\site-
packages (from sklearn) (0.24.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\python\python39\l
ib\site-packages (from scikit-learn->sklearn) (2.1.0)
Requirement already satisfied: joblib>=0.11 in c:\python\python39\lib\site-
packages (from scikit-learn->sklearn) (1.0.0)
Requirement already satisfied: scipy>=0.19.1 in c:\python\python39\lib\site
-packages (from scikit-learn->sklearn) (1.5.4)

WARNING: You are using pip version 20.3.3; however, version 21.1.1 is avail
able.
You should consider upgrading via the 'c:\python\python39\python.exe -m pip
install --upgrade pip' command.
```

In [2]:
```
import numpy as np
import pandas as pd
import itertools
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
```

```
c:\python\python39\lib\site-packages\numpy\_distributor_init.py:30: UserWar
ning: loaded more than 1 DLL from .libs:
c:\python\python39\lib\site-packages\numpy\.libs\libopenblas.NOIJJG62EMASZI
6NYURL6JBKM4EVBGM7.gfortran-win_amd64.dll
c:\python\python39\lib\site-packages\numpy\.libs\libopenblas.QVLO2T66WEPI7J
Z63PS3HMOHFEY472BC.gfortran-win_amd64.dll
  warnings.warn("loaded more than 1 DLL from .libs:"
```

In [4]:
```
#Read the data
df=pd.read_csv('news.csv')

#Get shape and head
df.shape
df.head()
```

```
c:\python\python39\lib\site-packages\IPython\core\interactiveshell.py:3146:
DtypeWarning: Columns (24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,4
1,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,6
6,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,9
1,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,1
12,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,
131,132,133,134,135,136,137,138,139,140) have mixed types.Specify dtype opt
ion on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[4]:

	Unnamed: 0	title	text	label	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Un

	Unnamed: 0	title	text	label	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Un
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE	NaN	NaN	NaN	NaN	
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE	NaN	NaN	NaN	NaN	
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL	NaN	NaN	NaN	NaN	
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE	NaN	NaN	NaN	NaN	
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL	NaN	NaN	NaN	NaN	

In [5]:
```
#DataFlair - Get the labels
labels=df.label
labels.head()
```

Out[5]:
```
0    FAKE
1    FAKE
2    REAL
3    FAKE
4    REAL
Name: label, dtype: object
```

In [6]:
```
#DataFlair - Split the dataset
x_train,x_test,y_train,y_test=train_test_split(df['text'], labels, test_siz
```

In [ ]: