Fake News Detection

Import library

In [4]:

```
pip install pandas
```

```
     --------- ---------------------------- 3.4/14.8 MB 545.0 kB/s et
a 0:00:21
     --------- ---------------------------- 3.5/14.8 MB 549.4 kB/s et
a 0:00:21
     --------- ---------------------------- 3.5/14.8 MB 559.7 kB/s et
a 0:00:21
     --------- ---------------------------- 3.6/14.8 MB 563.6 kB/s et
a 0:00:20
     --------- ---------------------------- 3.6/14.8 MB 564.3 kB/s et
a 0:00:20
     --------- ---------------------------- 3.7/14.8 MB 568.1 kB/s et
a 0:00:20
     --------- ---------------------------- 3.7/14.8 MB 568.7 kB/s et
a 0:00:20
     --------- ---------------------------- 3.8/14.8 MB 572.5 kB/s et
a 0:00:20
     ---------- ---------------------------- 3.8/14.8 MB 577.7 kB/s et
a 0:00:20
     ---------- ---------------------------- 3.9/14.8 MB 581.3 kB/s et
a 0:00:19
```

In [5]:

```
import pandas as pd
import numpy as np
```

In [7]:

```
import pandas as pd
import numpy as np
import itertools
```

In [29]:

```
df = pd.read_csv("news.csv")
```

In [30]:

```
df.head()
```

Out[30]:

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

In [31]:

```
df.shape
```

Out[31]:

```
(6335, 4)
```

In [32]:

```
df.isnull().sum()
```

Out[32]:

```
Unnamed: 0    0
title         0
text          0
label         0
dtype: int64
```

In [33]:

```
labels = df.label
```

In [34]:

```
labels
```

Out[34]:

```
0       FAKE
1       FAKE
2       REAL
3       FAKE
4       REAL
        ...
6330    REAL
6331    FAKE
6332    FAKE
6333    REAL
6334    REAL
Name: label, Length: 6335, dtype: object
```

In [35]:

```
labels.head()
```

Out[35]:

```
0    FAKE
1    FAKE
2    REAL
3    FAKE
4    REAL
Name: label, dtype: object
```

In [18]:

```
pip install sklearn
```

```
Collecting sklearnNote: you may need to restart the kernel to use updated
packages.

  DEPRECATION: sklearn is being installed using the legacy 'setup.py inst
all' method, because it does not have a 'pyproject.toml' and the 'wheel'
package is not installed. pip 23.1 will enforce this behaviour change. A
possible replacement is to enable the '--use-pep517' option. Discussion c
an be found at https://github.com/pypa/pip/issues/8559 (https://github.co
m/pypa/pip/issues/8559)


  Downloading sklearn-0.0.post1.tar.gz (3.6 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Installing collected packages: sklearn
  Running setup.py install for sklearn: started
  Running setup.py install for sklearn: finished with status 'done'
Successfully installed sklearn-0.0.post1
```

In [20]:

```
pip install scikit-learn
```

```
0:00:08
     - -------------------------------------- 0.2/8.3 MB 1.1 MB/s eta
0:00:08
     - -------------------------------------- 0.3/8.3 MB 981.5 kB/s et
a 0:00:09
     - -------------------------------------- 0.4/8.3 MB 1.0 MB/s eta
0:00:08
     -- ------------------------------------- 0.5/8.3 MB 1.2 MB/s eta
0:00:07
     --- ------------------------------------ 0.6/8.3 MB 1.4 MB/s eta
0:00:06
     --- ------------------------------------ 0.7/8.3 MB 1.5 MB/s eta
0:00:06
     --- ------------------------------------ 0.8/8.3 MB 1.5 MB/s eta
0:00:05
     ---- ----------------------------------- 0.9/8.3 MB 1.5 MB/s eta
0:00:06
     ---- ----------------------------------- 1.0/8.3 MB 1.5 MB/s eta
0:00:05
     ---- ----------------------------------- 1.0/8.3 MB 1.5 MB/s eta
```

In [36]:

```
from sklearn.model_selection import train_test_split
```

In [56]:

```
x_train, x_test, y_train, y_test = train_test_split(df["text"], labels, test_size = 0.2,
```

In [57]:

```
x_train.head()
```

Out[57]:

```
4741    NAIROBI, Kenya – President Obama spoke out Sun...
2089    Killing Obama administration rules, dismantlin...
4074    Dean Obeidallah, a former attorney, is the hos...
5376      WashingtonsBlog \nCNN’s Jake Tapper hit the ...
6028    Some of the biggest issues facing America this...
Name: text, dtype: object
```

In [58]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
```

In [59]:

```
# initialise a Tfidvectorizer
vector = TfidfVectorizer(stop_words='english', max_df=0.7)
```

In [60]:

```
# fit and transform
tf_train = vector.fit_transform(x_train)
tf_test = vector.transform(x_test)
```

In [61]:

```
# initialise a PassiveAggressiveClassifier
pac = PassiveAggressiveClassifier(max_iter=50)
pac.fit(tf_train, y_train)
```

Out[61]:

```
PassiveAggressiveClassifier(max_iter=50)
```

In [62]:

```
# prediction on test dataset
from sklearn.metrics import accuracy_score, confusion_matrix
y_pred = pac.predict(tf_test)
```

In [63]:

```
score = accuracy_score(y_test, y_pred)
```

In [67]:

```
print(f"Accuracy : {round(score*100,2)}%")
```

```
Accuracy : 95.03%
```

In [65]:

```
# confusion metrics
confusion_matrix(y_test, y_pred, labels=['FAKE', 'REAL'])
```

Out[65]:

```
array([[623,  25],
       [ 38, 581]], dtype=int64)
```

In [66]:

```
# save model
import pickle
filename = 'finalized_model.pkl'
pickle.dump(pac, open(filename, 'wb'))
```

In [68]:

```
# save vectorizer
filename = 'vectorizer.pkl'
pickle.dump(vector, open(filename, 'wb'))
```

In [ ]:

```

```