In [32]:

```
import os
os.chdir("D:\Data Science\Projects")
```

In [33]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
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
%matplotlib inline
```

In [34]:

```
os.getcwd()
```

Out[34]:

```
'D:\\Data Science\\Projects'
```

In [35]:

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

In [36]:

```
news_data.head(5)
```

Out[36]:

	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 [37]:

```
news_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6335 entries, 0 to 6334
Data columns (total 4 columns):
Unnamed: 0    6335 non-null int64
title         6335 non-null object
text          6335 non-null object
label         6335 non-null object
dtypes: int64(1), object(3)
memory usage: 198.1+ KB
```

In [38]:

```
news_data.shape
```

Out[38]:

```
(6335, 4)
```

In [39]:

```
news_data.isnull().sum()
```

Out[39]:

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

In [40]:

```
labels = news_data.label
labels.head()
```

Out[40]:

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

In [41]:

```
x_train,x_test,y_train,y_test = train_test_split(news_data['text'], labels, test_size=0.2,
```

In [42]:

```
# Initialize a TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer(stop_words='english', max_df=0.7)
```

In [43]:

```
# Fit and transform train set, transform test set

tfidf_train = tfidf_vectorizer.fit_transform(x_train)

tfidf_test = tfidf_vectorizer.transform(x_test)
```

In [44]:

```
# Initialize a PassiveAggressiveClassifier

pac=PassiveAggressiveClassifier(max_iter=50)

pac.fit(tfidf_train,y_train)

#Predict on the test set and calculate accuracy

y_pred=pac.predict(tfidf_test)

score=accuracy_score(y_test,y_pred)

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

```
Accuracy: 92.5%
```

In [45]:

```
# Build confusion matrix

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

Out[45]:

```
array([[589,  49],
       [ 46, 583]], dtype=int64)
```

In [ ]:

```

```