

In [37]:

```
#fake email detection using the machine learning classification algorithm
#term frequency-the number of times word appears in a documents is called the term frequency
#a higher value means term appear more than other frequency
#idf-inverse document frequency-word that occur many times in a document but also occur many times
in many others
#passivealgorithmclassifier is a online learning algorithm its update the correct the loss causing
very little change in terms
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
```

In [38]:

```
#read the data
df=pd.read_csv('news.csv')
print(df)
#get the shape and head
df.shape
df.head()
```

```
   Unnamed: 0  title \
0      8476  You Can Smell Hillary's Fear
1     10294  Watch The Exact Moment Paul Ryan Committed Pol...
2      3608  Kerry to go to Paris in gesture of sympathy
3     10142  Bernie supporters on Twitter erupt in anger ag...
4       875  The Battle of New York: Why This Primary Matters
...      ...
6330     4490  State Department says it can't find emails fro...
6331     8062  The 'P' in PBS Should Stand for 'Plutocratic' ...
6332     8622  Anti-Trump Protesters Are Tools of the Oligarc...
6333     4021  In Ethiopia, Obama seeks progress on peace, se...
6334     4330  Jeb Bush Is Suddenly Attacking Trump. Here's W...
```

```
   text label
0  Daniel Greenfield, a Shillman Journalism Fello...  FAKE
1  Google Pinterest Digg Linkedin Reddit Stumbleu...  FAKE
2  U.S. Secretary of State John F. Kerry said Mon...  REAL
3  - Kaydee King (@KaydeeKing) November 9, 2016 T...  FAKE
4  It's primary day in New York and front-runners...  REAL
...      ...
6330  The State Department told the Republican Natio...  REAL
6331  The 'P' in PBS Should Stand for 'Plutocratic' ...  FAKE
6332  Anti-Trump Protesters Are Tools of the Oligar...  FAKE
6333  ADDIS ABABA, Ethiopia -President Obama convene...  REAL
6334  Jeb Bush Is Suddenly Attacking Trump. Here's W...  REAL
```

[6335 rows x 4 columns]

Out[38]:

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 [39]:

```
#label the of the frame
```

```
labels=df.label
labels.head()
print(labels)
```

```
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 [40]:

```
#split the dataset into the traning and testing
x_train,x_test,y_train,y_test=train_test_split(df['text'],labels,test_size=0.2,random_state=7)
```

In [41]:

```
#DataFlair - Initialize a TfidfVectorizer
tfidf_vectorizer=TfidfVectorizer(stop_words='english', max_df=0.7)
#DataFlair - Fit and transform train set, transform test set
tfidf_train=tfidf_vectorizer.fit_transform(x_train)
tfidf_test=tfidf_vectorizer.transform(x_test)
```

In [42]:

```
#DataFlair - Initialize a PassiveAggressiveClassifier
pac=PassiveAggressiveClassifier(max_iter=50)
pac.fit(tfidf_train,y_train)
#DataFlair - 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 [43]:

```
#DataFlair - Build confusion matrix
confusion_matrix(y_test,y_pred, labels=['FAKE','REAL'])
```

Out[43]:

```
array([[588,  50],
       [ 45, 584]], dtype=int64)
```