

```
In [2]: #Make necessary imports:
```

```
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 [ ]: #Now, Let's read the data into a DataFrame, and get the shape of the data and the first 5 records.
```

```
In [5]: # Read the data
df = pd.read_csv(r"C:\Users\KIIT\Downloads\news.csv")
```

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

```
(6335, 4)
   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

      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
```

```
In [ ]: #And get the Labels from the DataFrame.
```

```
In [6]: #DataFlair - Get the Labels
labels=df.label
labels.head()
```

```
Out[6]: 0    FAKE
        1    FAKE
        2    REAL
        3    FAKE
        4    REAL
        Name: label, dtype: object
```

```
In [ ]: #Split the dataset into training and testing sets.
```

```
In [8]: #DataFlair - Split the dataset
x_train,x_test,y_train,y_test=train_test_split(df['text'], labels, test_size=0.2, random_state=7)
```

```
In [9]: #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 [ ]: #Next, we'll initialize a PassiveAggressiveClassifier. This is. We'll fit this on tfidf_train and
```

```
In [11]: #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 [ ]: #We got an accuracy of 92.5% with this model. Finally, let's print out a confusion matrix to gain
```

```
In [12]: #DataFlair - Build confusion matrix
confusion_matrix(y_test,y_pred, labels=['FAKE','REAL'])
```

```
Out[12]: array([[586,  52],
               [ 43, 586]], dtype=int64)
```

```
In [ ]: #So with this model, we have 586 true positives, 586 true negatives, 43 false positives, and 52 ;
```