Import Libraries

```python
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score,classification_report
```

Import Dataset

In [2]:
```python
true_data = pd.read_csv('True.csv')
fake_data = pd.read_csv('Fake.csv')
```

In [3]:
```python
fake_data.head()
```

Out[3]:

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

In [4]:
```python
#add a column target label
true_data['Target']=['True']*len(true_data)
fake_data['Target']=['Fake']*len(fake_data)
```

```
In [5]:  # data=true_data.append(fake_data).sample(frac=1).reset_index().drop(columns=[
         # data.head()

         data = pd.concat([true_data, fake_data], ignore_index=True)
         data = data.sample(frac=1).reset_index(drop=True)
         data.head()
```

Out[5]:

	title	text	subject	date	Target
0	What Ted Cruz Said In Church Sounds Like Shar...	Presidential candidate Senator Ted Cruz (R-TX)...	News	February 15, 2016	Fake
1	THEY KNEW! CLOSE CLINTON FRIEND Warned Against...	In June of 2016, the disgraced sexual predator...	left-news	Oct 18, 2017	Fake
2	Former RNC Chair Michael Steele Refuses To Vo...	Donald Trump has dragged the Republican Party ...	News	October 21, 2016	Fake
3	Cambodia's main opposition party dissolved by ...	PHNOM PENH (Reuters) - Cambodia s highest cour...	worldnews	November 16, 2017	True
4	CNN INTERVIEW TURNS INTO SCREAMING MATCH When ...	(From CNN) New video captured hours before Mic...	politics	Mar 13, 2017	Fake

```
In [6]:  #create a dummy column for target column
         data['label']=pd.get_dummies(data.Target)['Fake']
         data.head()
```

Out[6]:

	title	text	subject	date	Target	label
0	What Ted Cruz Said In Church Sounds Like Shar...	Presidential candidate Senator Ted Cruz (R-TX)...	News	February 15, 2016	Fake	1
1	THEY KNEW! CLOSE CLINTON FRIEND Warned Against...	In June of 2016, the disgraced sexual predator...	left-news	Oct 18, 2017	Fake	1
2	Former RNC Chair Michael Steele Refuses To Vo...	Donald Trump has dragged the Republican Party ...	News	October 21, 2016	Fake	1
3	Cambodia's main opposition party dissolved by ...	PHNOM PENH (Reuters) - Cambodia s highest cour...	worldnews	November 16, 2017	True	0
4	CNN INTERVIEW TURNS INTO SCREAMING MATCH When ...	(From CNN) New video captured hours before Mic...	politics	Mar 13, 2017	Fake	1

```
In [7]:  def cleaning_data(row):
             row=row.lower()
             row=re.sub('[^a-zA-Z]',' ', row)
             token=row.split()
             news=[ps.lemmatize(word) for word in token if not word in stopwords]
             cleaned_news=' '.join(news)
             return cleaned_news
```

Split the dataset into training and test data

```
In [8]: xtrain,xtest,ytrain,ytest=train_test_split(data['text'],data['label'],test_siz
```

```
In [9]: xtrain
```

```
Out[9]: 38388    DC Antifa Leader Moved to Turkey with Man She ...
        23136    Consortium News Exclusive: The mainstream medi...
        24221    Leave it to Obama to politicize our money some...
        8837     NEW YORK (Reuters) - A federal appeals court o...
        15902    LAHORE, Pakistan (Reuters) - A Christian man h...
                                   ...
        30403    COLOMBO (Reuters) - European lawmakers said on...
        21243    SEOUL (Reuters) - North Korea's leader Kim Jon...
        42613    Megyn Kelly may have departed Fox News for mor...
        43567    BERLIN (Reuters) - Largely unperturbed by Ange...
        2732     BARCELONA (Reuters) - Catalonia s regional par...
        Name: text, Length: 35918, dtype: object
```

```
In [10]: xtest
```

```
Out[10]: 35305    TOKYO (Reuters) - Japanese Prime Minister Shin...
         29180    (Reuters) - U.S. President Donald Trump named ...
         29805    PARIS (Reuters) - France is close to finalizin...
         38237    21st Century Wire says Our weekly documentary ...
         5099     MOSCOW (Reuters) - Politicians from North and ...
                                    ...
         37781    WASHINGTON (Reuters) - Several environmental g...
         2275      Snitches get stitches  is Kremlin s message t...
         19324    If there s anyone who has clearly had enough o...
         1178     THE PURGE CONTINUES THE NYC MAYOR has just ope...
         34472    DANANG, Vietnam (Reuters) - Vietnam released w...
         Name: text, Length: 8980, dtype: object
```

```
In [11]: ytrain
```

```
Out[11]: 38388    1
         23136    1
         24221    1
         8837     0
         15902    0
                 ..
         30403    0
         21243    0
         42613    1
         43567    0
         2732     0
         Name: label, Length: 35918, dtype: uint8
```

```
In [12]: ytest
```

```
Out[12]: 35305    0
         29180    0
         29805    0
         38237    1
         5099     0
                 ..
         37781    0
         2275     1
         19324    1
         1178     1
         34472    0
         Name: label, Length: 8980, dtype: uint8
```

Vectorization

Vectorization is a process in machine learning that converts text or other categorical data into numerical vectors.

TF-IDF stands for term frequency-inverse document frequency. It is a statistical measure that is used to quantify the importance of a word in a document. TF-IDF is calculated by multiplying the term frequency by the inverse document frequency.

```
In [13]: vectorizer = TfidfVectorizer(max_features=500, lowercase=False, ngram_range=(1
```

```
In [14]: vec_xtrain=vectorizer.fit_transform(xtrain).toarray()
         vec_xtest=vectorizer.fit_transform(xtest).toarray()
```

```
In [15]: vec_xtrain
```

```
Out[15]: array([[0.        , 0.        , 0.        , ..., 0.        , 0.2970844 ,
                 0.        ],
                [0.        , 0.        , 0.        , ..., 0.        , 0.        ,
                 0.        ],
                [0.        , 0.06643891, 0.        , ..., 0.        , 0.04445308,
                 0.        ],
                ...,
                [0.        , 0.        , 0.        , ..., 0.        , 0.16900815,
                 0.        ],
                [0.        , 0.        , 0.        , ..., 0.02577194, 0.        ,
                 0.        ],
                [0.        , 0.20230865, 0.        , ..., 0.        , 0.        ,
                 0.        ]])
```

```
In [16]: vec_xtest
```

```
Out[16]: array([[0.        , 0.        , 0.        , ..., 0.        , 0.        ,
          0.        ],
         [0.        , 0.        , 0.        , ..., 0.04813364, 0.        ,
          0.        ],
         [0.        , 0.09176602, 0.09438542, ..., 0.06845066, 0.        ,
          0.        ],
         ...,
         [0.        , 0.        , 0.        , ..., 0.        , 0.04259164,
          0.06187009],
         [0.        , 0.        , 0.        , ..., 0.        , 0.03242363,
          0.        ],
         [0.        , 0.        , 0.        , ..., 0.        , 0.        ,
          0.        ]])
```

Model :- MultinomialNB

```
In [17]: clf=MultinomialNB()
         clf.fit(vec_xtrain,ytrain)
         y_pred=clf.predict(vec_xtest)
```

Classification Report

```
In [18]: print(classification_report(ytest, y_pred))
```

```
               precision    recall  f1-score   support

           0       0.91      0.83      0.87      4328
           1       0.85      0.92      0.89      4652

    accuracy                           0.88      8980
   macro avg       0.88      0.87      0.88      8980
weighted avg       0.88      0.88      0.88      8980
```

```
In [19]: y_pred_train= clf.predict(vec_xtrain)
         print(classification_report(ytrain, y_pred_train))
```

```
               precision    recall  f1-score   support

           0       0.92      0.95      0.94     17089
           1       0.95      0.93      0.94     18829

    accuracy                           0.94     35918
   macro avg       0.94      0.94      0.94     35918
weighted avg       0.94      0.94      0.94     35918
```

Accuracy score

```
In [20]: accuracy_score(ytrain,y_pred_train)
```

Out[20]: 0.9394175622250682

```
In [21]: accuracy_score(ytest,y_pred)
```

Out[21]: 0.8766146993318485

```
In [ ]:
```