Fake News Detection

Importing Libraries

```python
In [1]:
import pandas as pd
import numpy as np
import seaborn as sns
import nltk as nlp
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string
```

```python
In [2]:
fake=r"C:\Users\Lenovo\Downloads\archive (5)\Fake.csv"
true=r"C:\Users\Lenovo\Downloads\archive (5)\True.csv"
df_fake = pd.read_csv(fake)
df_true = pd.read_csv(true)
```

```python
In [3]:
df_fake.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

```python
In [4]:
df_true.head(5)
```

Out[4]:

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

Inserting a column "class" as target feature

In [5]:
```python
df_fake["class"] = 0
df_true["class"] = 1
```

In [6]:
```python
df_fake.shape, df_true.shape
```

Out[6]: `((23481, 5), (21417, 5))`

In [7]:
```python
print(df_fake["class"])
```

```
0        0
1        0
2        0
3        0
4        0
        ..
23476    0
23477    0
23478    0
23479    0
23480    0
Name: class, Length: 23481, dtype: int64
```

In [8]:
```python
print(df_true["class"])
```

```
0        1
1        1
2        1
3        1
4        1
        ..
21412    1
21413    1
21414    1
21415    1
21416    1
Name: class, Length: 21417, dtype: int64
```

Merging True and Fake Dataframes

```
In [9]: df_merge = pd.concat([df_fake[0:1000], df_true[0:1000]], axis =0 )
        df_merge.head(10)
```

Out[9]:

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0

```
In [10]: df_merge["class"].unique()
```

```
Out[10]: array([0, 1], dtype=int64)
```

```
In [11]: df_merge.columns
```

```
Out[11]: Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')
```

```
In [12]: len(df_merge)
```

```
Out[12]: 2000
```

Removing columns which are not required

```
In [13]: df = df_merge.drop(["subject","date"], axis = 1)
```

```
In [14]: df.isnull().sum()
```

```
Out[14]:   title    0
           text     0
           class    0
           dtype: int64
```

Random Shuffling the dataframe

In [15]:
```python
df = df.sample(frac = 1) ## This basically reshuffles all the data
```

In [16]:
```python
df.head()
```

Out[16]:

	title	text	class
46	Senators seek to stop expansion of airport fac...	WASHINGTON (Reuters) - Two U.S. senators on Th...	1
649	Attorney General Sessions says now recalls mee...	WASHINGTON (Reuters) - U.S. Attorney General J...	1
754	Buy American? Even Trump's MAGA Gear Is Sold ...	When Donald Trump kicked off Made in America ...	0
281	Trump's eldest son questioned in Congress abou...	WASHINGTON (Reuters) - President Donald Trump'...	1
417	Trump weighs recognizing Jerusalem as Israel's...	WASHINGTON (Reuters) - President Donald Trump ...	1

In [17]:
```python
df.reset_index(inplace = True) ## Inpalce True means resetting the order in the exi
##df.drop(["index"], axis = 1, inplace = True)
```

In [18]:
```python
df.columns
```

Out[18]:
```
Index(['index', 'title', 'text', 'class'], dtype='object')
```

In [19]:
```python
df['full_text'] = df['title'] + ' ' + df['text']
```

In [20]: ```python
df.head()
```

Out[20]:

| | index | title | text | class | full_text |
|---|---|---|---|---|---|
| **0** | 46 | Senators seek to stop expansion of airport fac... | WASHINGTON (Reuters) - Two U.S. senators on Th... | 1 | Senators seek to stop expansion of airport fac... |
| **1** | 649 | Attorney General Sessions says now recalls mee... | WASHINGTON (Reuters) - U.S. Attorney General J... | 1 | Attorney General Sessions says now recalls mee... |
| **2** | 754 | Buy American? Even Trump's MAGA Gear Is Sold ... | When Donald Trump kicked off Made in America ... | 0 | Buy American? Even Trump's MAGA Gear Is Sold ... |
| **3** | 281 | Trump's eldest son questioned in Congress abou... | WASHINGTON (Reuters) - President Donald Trump'... | 1 | Trump's eldest son questioned in Congress abou... |
| **4** | 417 | Trump weighs recognizing Jerusalem as Israel's... | WASHINGTON (Reuters) - President Donald Trump ... | 1 | Trump weighs recognizing Jerusalem as Israel's... |

In [21]: ```python
columns_to_drop=['index', 'title', 'text']
df_new=df.drop(columns=columns_to_drop, inplace=False)
```

In [22]: ```python
df_new.head()
```

Out[22]:

| | class | full_text |
|---|---|---|
| **0** | 1 | Senators seek to stop expansion of airport fac... |
| **1** | 1 | Attorney General Sessions says now recalls mee... |
| **2** | 0 | Buy American? Even Trump's MAGA Gear Is Sold ... |
| **3** | 1 | Trump's eldest son questioned in Congress abou... |
| **4** | 1 | Trump weighs recognizing Jerusalem as Israel's... |

In [23]: ```python
print(len(df_new))
```

```
2000
```

Preprocessing

In [24]: ```python
nlp.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\Lenovo\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[24]: True

In [25]: ```python
print(stopwords.words('english'))
print(f"No of stop words are -> {len(stopwords.words('english'))}")
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you
've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'h
is', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'its
elf', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'who
m', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were',
'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing'
, 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'o
f', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'd
uring', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out
', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there
', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'mos
t', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than
', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "shoul
d've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'could
n', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "ha
sn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mu
stn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn
't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]
No of stop words are -> 179
```

Stemming

In [26]:
```python
# Stemming process
def stemming_process(y):
    first_step = re.sub(r"^[A-Za-z]", ' ', y).lower()
    second_step = first_step.split()
    porter_stemmer = PorterStemmer()
    result = []
    for w in second_step:
        if w not in stopwords.words('english'):
            result.append(porter_stemmer.stem(w))
    return ' '.join(result)
```

In [27]:
```python
df_new['full_text'] = df_new['full_text'].apply(stemming_process)
```

In [28]:
```python
df_new.head()
```

Out[28]:

	class	full_text
0	1	enat seek stop expans airport facial scan wash…
1	1	ttorney gener session say recal meet flag russ…
2	0	buy american? even trump' maga gear sold via c…
3	1	rump' eldest son question congress russia wash…
4	1	rump weigh recogn jerusalem israel' capital: o…

```python
In [30]: def wordopt(text):
             text = text.lower()
             text = re.sub('\[.*?\]', '', text)
             text = re.sub("\\W"," ",text)
             text = re.sub('https?://\S+|www\.\S+', '', text)
             text = re.sub('<.*?>+', '', text)
             text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
             text = re.sub('\n', '', text)
             text = re.sub('\w*\d\w*', '', text)
             return text
```

```python
In [31]: df_new["full_text"] = df["full_text"].apply(wordopt)
```

```python
In [32]: df_new = df_new.sample(frac = 1)
```

Defining dependent and independent variables and vectorization

```python
In [33]: x = df_new["full_text"]
         y = df_new["class"]
```

```python
In [34]: vectorizer = TfidfVectorizer()
         vectorizer.fit(x)
         X = vectorizer.transform(x)
```

Splitting Training and Testing

```python
In [35]: x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.30)
```

Logistic Regression

```python
In [36]: from sklearn.linear_model import LogisticRegression

         LR = LogisticRegression()
         LR.fit(x_train,y_train)
```

```
Out[36]: ▾ LogisticRegression

         LogisticRegression()
```

```python
In [37]: pred_lr=LR.predict(x_test)
```

```python
In [38]: print(pred_lr)
```

```
[1 0 1 1 1 1 1 0 0 0 1 1 1 0 1 1 1 0 1 0 1 0 0 0 0 0 1 1 0 1 1 0 0 1 1 1 1
 0 1 0 0 0 0 0 1 1 0 1 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 1 0 0 0 1 1 0
 1 1 0 0 0 0 0 0 0 1 1 0 0 0 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 0 1 0 1 0 1 1
 0 0 0 0 1 0 0 1 1 0 0 1 0 0 0 1 0 1 0 0 0 1 0 0 1 1 1 0 0 0 0 0 1 0 0 1 1
 0 0 0 0 0 0 1 1 1 0 1 0 0 1 1 1 0 1 0 1 1 1 0 0 1 1 1 0 1 0 0 1 1 1 1 0 0
 0 1 0 1 0 1 0 1 1 1 1 1 1 0 1 0 1 0 0 1 0 0 0 0 1 0 1 1 0 1 0 0 1 0 0 1 1
 1 0 1 0 1 1 1 1 0 0 0 0 0 0 1 0 0 1 0 0 0 1 1 1 1 0 1 1 0 1 0 0 1 0 0 0 0
 0 1 1 0 1 1 1 0 1 1 0 1 0 0 1 1 1 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 1 1 1 1
 0 1 0 1 0 0 1 0 1 1 1 1 1 0 1 0 1 1 0 1 1 0 1 1 0 1 0 1 1 1 1 0 1 0 0 1 1
 0 1 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 1 1 1 0 0
 0 1 0 0 0 1 0 0 1 1 1 0 1 0 0 1 0 1 0 0 0 1 1 1 1 1 0 1 0 0 1 1 1 1 1 0 1
 0 0 0 0 0 0 1 0 1 1 1 0 1 1 1 0 1 0 1 1 0 1 1 1 0 0 1 1 0 1 1 0 0 1 1 1 1
 1 0 1 1 1 1 0 1 0 0 0 1 1 0 0 0 0 1 0 1 1 0 0 0 1 1 0 1 1 0 1 1 0 0 1 1 0
 1 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0 1 1 1 1 1 1 1 0 1 0 1 0 1 1 1 1
 1 1 0 0 1 1 0 1 1 1 1 1 1 1 1 1 0 0 1 0 0 1 1 0 1 0 0 0 0 0 1 0 1 1 0 0
 0 1 0 0 1 1 0 0 1 0 1 0 1 1 1 0 0 0 1 0 0 0 0 1 1 0 1 1 1 0 0 0 0 0 1 0 1
 1 1 0 1 0 0 1 1]
```

In [39]: `LR.score(x_test, y_test)`

Out[39]: `0.9866666666666667`

In [40]: `print(classification_report(y_test, pred_lr))`

```
              precision    recall  f1-score   support

           0       0.98      0.99      0.99       298
           1       0.99      0.98      0.99       302

    accuracy                           0.99       600
   macro avg       0.99      0.99      0.99       600
weighted avg       0.99      0.99      0.99       600
```

Decision Tree Classification

In [41]: `from sklearn.tree import DecisionTreeClassifier`

In [42]: `DTC = DecisionTreeClassifier()`

In [43]: `DTC.fit(x_train, y_train)`

Out[43]:
```
▾ DecisionTreeClassifier

DecisionTreeClassifier()
```

In [44]: `pred_dtc = DTC.predict(x_test)`

In [45]:
```
print("Decision Tree Classifier:")
print("Accuracy:", accuracy_score(y_test, pred_dtc))
print(classification_report(y_test, pred_dtc))
```

```
Decision Tree Classifier:
Accuracy: 0.9966666666666667
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       298
           1       1.00      1.00      1.00       302

    accuracy                           1.00       600
   macro avg       1.00      1.00      1.00       600
weighted avg       1.00      1.00      1.00       600
```

Random forest classifier

In [46]:
```python
from sklearn.ensemble import RandomForestClassifier
```

In [47]:
```python
RFC = RandomForestClassifier()
RFC.fit(x_train, y_train)
pred_rfc = RFC.predict(x_test)
```

In [48]:
```python
print("Random Forest Classifier:")
print("Accuracy:", accuracy_score(y_test, pred_rfc))
print(classification_report(y_test, pred_rfc))
```

```
Random Forest Classifier:
Accuracy: 0.9983333333333333
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       298
           1       1.00      1.00      1.00       302

    accuracy                           1.00       600
   macro avg       1.00      1.00      1.00       600
weighted avg       1.00      1.00      1.00       600
```

Naive Bayes Classifier

In [49]:
```python
from sklearn.naive_bayes import MultinomialNB
```

In [50]:
```python
NB = MultinomialNB()
NB.fit(x_train, y_train)
pred_nb = NB.predict(x_test)
```

In [51]:
```python
print("Naive Bayes Classifier:")
print("Accuracy:", accuracy_score(y_test, pred_nb))
print(classification_report(y_test, pred_nb))
```

```
Naive Bayes Classifier:
Accuracy: 0.9666666666666667
              precision    recall  f1-score   support

           0       0.96      0.97      0.97       298
           1       0.97      0.96      0.97       302

    accuracy                           0.97       600
   macro avg       0.97      0.97      0.97       600
weighted avg       0.97      0.97      0.97       600
```