```python
import pandas as pd
```

```python
import numpy as np
```

```python
df=pd.read_csv('https://raw.githubusercontent.com/YBI-Foundation/Dataset/main/Financial%20
```

```python
df.head()
```

	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6
0	01-01-2010	0	McIlroy's men catch cold from Gudjonsson	Obituary: Brian Walsh	Workplace blues leave employers in the red	Classical review: Rattle	Dance review: Merce Cunningham	Genetic tests to be used in setting premiums
1	02-01-2010	0	Warning from history points to crash	Investors flee to dollar haven	Banks and tobacco in favour	Review: Llama Farmers	War jitters lead to sell-off	Your not-so-secret history
2	03-01-2010	0	Comment: Why Israel's peaceniks feel betrayed	Court deals blow to seizure of drug assets	An ideal target for spooks	World steps between two sides intent on war	What the region's papers say	Comment: Fear and rage in Palestine
3	04-01-2010	1	£750,000-a-goal Weah aims parting shot	Newcastle pay for Fletcher years	Brown sent to the stands for Scotland qualifier	Tourists wary of breaking new ground	Canary Wharf climbs into the FTSE 100	Review: Bill Bailey
4	05-01-2010	1	Leeds arrive in Turkey to the silence of the fans	One woman's vision offers loan lifeline	Working Lives: How world leaders worked	Working Lives: Tricks of the trade	Working Lives: six-hour days, long lunches and...	Pop review: We Love UK

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4101 entries, 0 to 4100
Data columns (total 27 columns):
```

```
 #   Column      Non-Null Count   Dtype
---  ------      --------------   -----
 0   Date        4101 non-null    object
 1   Label       4101 non-null    int64
 2   News 1      4101 non-null    object
 3   News 2      4101 non-null    object
 4   News 3      4101 non-null    object
 5   News 4      4101 non-null    object
 6   News 5      4101 non-null    object
 7   News 6      4101 non-null    object
 8   News 7      4101 non-null    object
 9   News 8      4101 non-null    object
 10  News 9      4101 non-null    object
 11  News 10     4101 non-null    object
 12  News 11     4101 non-null    object
 13  News 12     4101 non-null    object
 14  News 13     4101 non-null    object
 15  News 14     4101 non-null    object
 16  News 15     4101 non-null    object
 17  News 16     4101 non-null    object
 18  News 17     4101 non-null    object
 19  News 18     4101 non-null    object
 20  News 19     4101 non-null    object
 21  News 20     4101 non-null    object
 22  News 21     4101 non-null    object
 23  News 22     4101 non-null    object
 24  News 23     4100 non-null    object
 25  News 24     4098 non-null    object
 26  News 25     4098 non-null    object
dtypes: int64(1), object(26)
memory usage: 865.2+ KB
```

```
df.shape
```

```
(4101, 27)
```

```
df.columns
```

```
Index(['Date', 'Label', 'News 1', 'News 2', 'News 3', 'News 4', 'News 5',
       'News 6', 'News 7', 'News 8', 'News 9', 'News 10', 'News 11', 'News 12',
       'News 13', 'News 14', 'News 15', 'News 16', 'News 17', 'News 18',
       'News 19', 'News 20', 'News 21', 'News 22', 'News 23', 'News 24',
       'News 25'],
      dtype='object')
```

```
#Get feature Selection
''.join(str(x) for x in df.iloc[1,2:27])
```

```
'Warning from history points to crashInvestors flee to dollar havenBanks and toba
cco in favourReview: Llama FarmersWar jitters lead to sell-offYour not-so-secret
historyReview: The Northern SinfoniaReview: HysteriaReview: The GuardsmanOpera: T
he Marriage of FigaroReview: The Turk in ItalyDeutsche spells out its plans for d
iversificationTraders' panic sends oil prices skywardTV sport chief leaves home o
```

```
df.index
```

```
RangeIndex(start=0, stop=4101, step=1)
```

```python
len(df.index)
```

```
4101
```

```python
news=[]
for row in range(0,len(df.index)):
    news.append(' '.join(str(x) for x in df.iloc[row,2:27]))
```

```python
type(news)
```

```
list
```

```python
news[0]
```

```
'McIlroy's men catch cold from Gudjonsson Obituary: Brian Walsh Workplace blues l
eave employers in the red Classical review: Rattle Dance review: Merce Cunningham
Genetic tests to be used in setting premiums Opera review: La Bohème Pop review:
Britney Spears Theatre review: The Circle Wales face a fraught night Under-21  ro
und-up Smith off to blot his copybook Finns taking the mickey Praise wasted as Br
```

```python
X=news
```

```python
type(X)
```

```
list
```

```python
#Get Feature Text Conversion to Bag of words
```

```python
from sklearn.feature_extraction.text import CountVectorizer
```

```python
cv=CountVectorizer(lowercase=True,ngram_range=(1,1))
```

```python
X=cv.fit_transform(X)
```

```python
X.shape
```

```
(4101, 48527)
```

```python
y=df['Label']
```

```python
y.shape
```

```
(4101,)
```

```python
#Get train test split
from sklearn.model_selection import train_test_split
```

```python
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_state=2529)
```

```python
from sklearn.ensemble import RandomForestClassifier
```

```python
rf=RandomForestClassifier(n_estimators=200)
```

```python
rf.fit(X_train,y_train)
```
```
    RandomForestClassifier(n_estimators=200)
```

```python
y_pred=rf.predict(X_test)
```

```python
from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
```

```python
confusion_matrix(y_test,y_pred)
```
```
    array([[123, 471],
           [117, 520]])
```

```python
print(classification_report(y_test,y_pred))
```
```
                  precision    recall  f1-score   support

               0       0.51      0.21      0.29       594
               1       0.52      0.82      0.64       637

        accuracy                           0.52      1231
       macro avg       0.52      0.51      0.47      1231
    weighted avg       0.52      0.52      0.47      1231
```

✓ 0s   completed at 12:33 AM   ● ✕