In [1]:
```
import pandas as pd
```

In [2]:
```
import numpy as np
```

In [5]:
```
df=pd.read_csv("https://raw.githubusercontent.com/YBI-Foundation/Dataset/main/Financial%2
```

In [6]:
```
df.head()
```

Out[6]:

	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6	News 7	News 8
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	Opera review: La Bohème	Pop review: Britney Spears
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	Review: The Northern Sinfonia	Review: Hysteria
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	Poverty and resentment fuels Palestinian fury	Republican feud fear as dissident is killed
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	Review: Classical	Review: New Contemporaries 2000
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	World music review: Marisa Monte	Art review: Hollingsworth/Heyer

5 rows × 27 columns

In [7]:
```
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
```

In [8]:
```
df.shape
```

Out[8]:
```
(4101, 27)
```

In [9]:
```
df.columns
```

Out[9]:
```
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')
```

In [12]:
```
' '.join(str(x) for x in df.iloc[1,2:27])
```

Out[12]:
```
"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 Review: The Northern Sinfonia Review: Hysteria Review: The Guardsman Opera: The Marriage of Figaro Review: The Turk in Italy Deutsche spells out its plans for diversification Traders' panic sends oil prices skyward TV sport chief leaves home over romance Leader: Hi-tech twitch Why Wenger will stick to his Gunners Out of luck England hit rock bottom Wilkinson out of his depth Kinsella sparks Irish power play Brown banished as Scots rebound Battling Wales cling to lifeline Ehiogu close to sealing Boro move Man-to-man marking Match stats French referee at centre of storm is no stranger to controversy"
```

In [13]:
```
df.index
```

Out[13]:
```
RangeIndex(start=0, stop=4101, step=1)
```

In [14]:
```
len(df.index)
```

Out[14]:
```
4101
```

In [16]:
```
news=[]
for row in range(0,len(df.index)):
    news.append(' '.join(str(x) for x in df.iloc[row,2:27]))
```

In [17]:
```
type(news)
```

Out[17]:
```
list
```

In [18]:
```
news[0]
```

Out[18]:
```
"McIlroy's men catch cold from Gudjonsson Obituary: Brian Walsh Workplace blues leave empl
oyers 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 rev
iew: The Circle Wales face a fraught night Under-21  round-up Smith off to blot his copybo
ok Finns taking the mickey Praise wasted as Brown studies injury options Ireland wary of m
innows Finland 0 - 0 England Healy a marked man Happy birthday Harpers & Queen Win unlimit
ed access to the Raindance film festival Labour pledges £800m to bridge north-south divide
Wales: Lib-Lab pact firm despite resignation Donald Dewar Regenerating homes  regenerates
well-being in people Win £100 worth of underwear TV guide: Random views"
```

In [19]:
```
x=news
```

In [20]:
```
type(x)
```

Out[20]:
```
list
```

In [21]:
```
from  sklearn.feature_extraction.text  import  CountVectorizer
```

In [22]:
```
cv=CountVectorizer(lowercase=True,ngram_range=(1,1))
```

In [23]:
```
x=cv.fit_transform(x)
```

In [24]:
```
x.shape
```

Out[24]:
```
(4101, 48527)
```

In [25]:
```
y=df['Label']
y.shape
```

Out[25]:
```
(4101,)
```

In [26]:
```
from sklearn.model_selection import train_test_split
```

In [27]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,stratify=y,random_state=
```

In [29]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [30]:
```
rf=RandomForestClassifier(n_estimators=200)
```

In [32]:
```
rf.fit(x_train,y_train)
```

Out[32]:
```
RandomForestClassifier(n_estimators=200)
```

In [33]:
```
y_pred=rf.predict(x_test)
```

In [34]:
```
from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
```

In [35]:
```
confusion_matrix(y_test,y_pred)
```

Out[35]:
```
array([[146, 435],
       [166, 484]], dtype=int64)
```

In [36]:
```
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.47      0.25      0.33       581
           1       0.53      0.74      0.62       650

    accuracy                           0.51      1231
   macro avg       0.50      0.50      0.47      1231
weighted avg       0.50      0.51      0.48      1231
```

In [ ]:
```

```