Financial Market News - Sentiment Analysis

In [4]:

```python
import pandas as pd
```

In [5]:

```python
import numpy as np
```

In [9]:

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

In [11]:

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

Out[11]:

	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6	News 7
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
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
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
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
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

5 rows × 27 columns
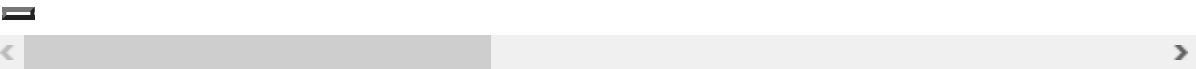
In [13]:

```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
```

In [16]:

```python
df.shape
```

Out[16]:

```
(4101, 27)
```

In [18]:

```python
df.columns
```

Out[18]:

```
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 1
2',
       '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 [20]:

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

Out[20]:

"Warning from history points to crashInvestors flee to dollar havenBanks and tobacco in favourReview: Llama FarmersWar jitters lead to sell-offYour not-so-secret historyReview: The Northern SinfoniaReview: HysteriaReview: The GuardsmanOpera: The Marriage of FigaroReview: The Turk in ItalyDeutsche spells out its plans for diversificationTraders' panic sends oil prices skywardTV sport chief leaves home over romanceLeader: Hi-tech twitchWhy Wenger will stick to his GunnersOut of luck England hit rock bottomWilkinson out of his depthKinsella sparks Irish power playBrown banished as Scots reboundBattling Wales cling to lifelineEhiogu close to sealing Boro moveMan-to-man markingMatch statsFrench referee at centre of storm is no stranger to controversy"

In [22]:

```python
df.index
```

Out[22]:

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

In [24]:

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

Out[24]:

4101

In [27]:

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

In [29]:

```python
type(news)
```

Out[29]:

list

In [31]:

```python
news[0]
```

Out[31]:

"McIlroy's men catch cold from GudjonssonObituary: Brian WalshWorkplace blue
s leave employers in the redClassical review: RattleDance review: Merce Cunn
inghamGenetic tests to be used in setting premiumsOpera review: La BohèmePop
review: Britney SpearsTheatre review: The CircleWales face a fraught nightUn
der-21 round-upSmith off to blot his copybookFinns taking the mickeyPraise
wasted as Brown studies injury optionsIreland wary of minnowsFinland 0 - 0 E
nglandHealy a marked manHappy birthday Harpers & QueenWin unlimited access t
o the Raindance film festivalLabour pledges £800m to bridge north-south divi
deWales: Lib-Lab pact firm despite resignationDonald DewarRegenerating homes
regenerates well-being in peopleWin £100 worth of underwearTV guide: Random
views"

In [32]:

```python
X=news
```

In [34]:

```python
type(X)
```

Out[34]:

list

In [35]:

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

In [37]:

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

In [38]:

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

In [40]:

```python
X.shape
```

Out[40]:

(4101, 108682)

In [41]:

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

In [43]:

```python
y.shape
```

Out[43]:

```
(4101,)
```

In [45]:

```python
from sklearn.model_selection import train_test_split
```

In [47]:

```python
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, stratify=y, rand
```

In [49]:

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

In [52]:

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

In [53]:

```python
rf.fit(X_train, y_train)
```

Out[53]:

```
RandomForestClassifier(n_estimators=200)
```

In [54]:

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

In [55]:

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

In [56]:

```python
confusion_matrix(y_test, y_pred)
```

Out[56]:

```
array([[114, 467],
       [132, 518]])
```

In [59]:

```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.46      0.20      0.28       581
           1       0.53      0.80      0.63       650

    accuracy                           0.51      1231
   macro avg       0.49      0.50      0.45      1231
weighted avg       0.50      0.51      0.46      1231
```