Import Libraries

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

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

In [3]: ```python
df = pd.read_csv('Financial Market News.csv', encoding = "ISO-8859-1")
```

In [4]: ```python
df.head()
```

Out[4]:

	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6	New
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	Op review Bohé
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	Revi North Sinfc
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	Pov resentm fi Palesti
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	Revi Class
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	W m revi Ma Mc

5 rows × 27 columns

```
In [5]:  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 [6]:  df.shape
```

```
Out[6]:  (4101, 27)
```

```
In [7]:  df.columns
```

```
Out[7]:  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 Section

```
In [8]:  ' '.join(str(x) for x in df.iloc[1,2:27])
```

Out[8]: "Warning from history points to crash Investors flee to dollar haven Banks an
d tobacco in favour Review: Llama Farmers War jitters lead to sell-off Your n
ot-so-secret history Review: The Northern Sinfonia Review: Hysteria Review: T
he Guardsman Opera: The Marriage of Figaro Review: The Turk in Italy Deutsche
spells out its plans for diversification Traders' panic sends oil prices skyw
ard 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 o
f 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 contr
oversy"

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

Out[9]: RangeIndex(start=0, stop=4101, step=1)

```
In [10]:  len(df.index)
```

Out[10]: 4101

```
In [11]:  news = []
          for row in range(0,len(df.index)):
              news.append(' '.join(str(x) for x in df.iloc[row,2:27]))
```

```
In [12]:  type(news)
```

Out[12]: list

```
In [13]:  news[0]
```

Out[13]: "McIlroy's men catch cold from Gudjonsson Obituary: Brian Walsh Workplace blu
es leave employers in the red Classical review: Rattle Dance review: Merce Cu
nningham Genetic tests to be used in setting premiums Opera review: La Bohème
Pop review: Britney Spears Theatre review: The Circle Wales face a fraught ni
ght Under-21 round-up Smith off to blot his copybook Finns taking the mickey
Praise wasted as Brown studies injury options Ireland wary of minnows Finland
0 - 0 England Healy a marked man Happy birthday Harpers & Queen Win unlimited
access to the Raindance film festival Labour pledges £800m to bridge north-so
uth divide Wales: Lib-Lab pact firm despite resignation Donald Dewar Regenera
ting homes regenerates well-being in people Win £100 worth of underwear TV g
uide: Random views"

```
In [14]:  X = news
```

```
In [15]:  type(X)
```

Out[15]: list

Get Feature Text Conversion To Bag Of

```python
In [16]: from sklearn.feature_extraction.text import CountVectorizer
```

```python
In [17]: cv = CountVectorizer(lowercase = True, ngram_range=(1,1))
```

```python
In [18]: X = cv.fit_transform(X)
```

```python
In [19]: X.shape
```
Out[19]: (4101, 48527)

```python
In [20]: y = df['Label']
```

```python
In [21]: y.shape
```
Out[21]: (4101,)

Get Train Test Split

```python
In [22]: from sklearn.model_selection import train_test_split
```

```python
In [23]: X_train, X_test, y_train,y_test =train_test_split(X, y, train_size = 0.7, stra
```

```python
In [24]: from sklearn.ensemble import RandomForestClassifier
```

```python
In [25]: rf = RandomForestClassifier(n_estimators=200)
```

```python
In [26]: rf.fit(X_train, y_train)
```
Out[26]: RandomForestClassifier(n_estimators=200)

```python
In [27]: y_pred = rf.predict(X_test)
```

```python
In [28]: from sklearn.metrics import classification_report,confusion_matrix,accuracy_sc
```

```python
In [29]: confusion_matrix(y_test, y_pred)
```
Out[29]: array([[143, 438],
               [182, 468]], dtype=int64)

```
In [30]: print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.44      0.25      0.32       581
           1       0.52      0.72      0.60       650

    accuracy                           0.50      1231
   macro avg       0.48      0.48      0.46      1231
weighted avg       0.48      0.50      0.47      1231
```