#Financial Market Sentiment Analysis

Importing the libraries

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Importing the dataset

```
df =
pd.read_csv('https://raw.github.com/YBI-Foundation/Dataset/main/Financ
ial%20Market%20News.csv',encoding="ISO-8859-1")
```

Get Information of Dataframe

```
df.info()    #gives column name, count,  not null category, D-type(data
type)
```

```
<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.describe()   *#gives the linear relation of each column with another column*

```
            Label
count  4101.000000
mean      0.528164
std       0.499267
min       0.000000
25%       0.000000
50%       1.000000
75%       1.000000
max       1.000000
```

df.head()

```
         Date  Label                                              News
1  \
0  01-01-2010      0          McIlroy's men catch cold from
Gudjonsson
1  02-01-2010      0             Warning from history points to
crash
2  03-01-2010      0       Comment: Why Israel's peaceniks feel
betrayed
3  04-01-2010      1             £750,000-a-goal Weah aims parting
shot
4  05-01-2010      1  Leeds arrive in Turkey to the silence of the
fans


                                              News 2  \
0                      Obituary: Brian Walsh
1              Investors flee to dollar haven
2  Court deals blow to seizure of drug assets
3             Newcastle pay for Fletcher years
4     One woman's vision offers loan lifeline


                                              News 3  \
0      Workplace blues leave employers in the red
1                     Banks and tobacco in favour
2                     An ideal target for spooks
3  Brown sent to the stands for Scotland qualifier
4          Working Lives: How world leaders worked


                                              News 4  \
0                 Classical review: Rattle
1                     Review: Llama Farmers
```

```
2   World steps between two sides intent on war
3               Tourists wary of breaking new ground
4                  Working Lives: Tricks of the trade


                                             News 5  \
0                   Dance review: Merce Cunningham
1                        War jitters lead to sell-off
2                        What the region's papers say
3             Canary Wharf climbs into the FTSE 100
4   Working Lives: six-hour days, long lunches and...


                                             News 6  \
0   Genetic tests to be used in setting premiums
1                        Your not-so-secret history
2             Comment: Fear and rage in Palestine
3                              Review: Bill Bailey
4                        Pop review: We Love UK


                                             News 7  \
0                        Opera review: La Bohème
1                   Review: The Northern Sinfonia
2   Poverty and resentment fuels Palestinian fury
3                              Review: Classical
4             World music review: Marisa Monte


                                             News 8  ...  \
0                   Pop review: Britney Spears  ...
1                             Review: Hysteria  ...
2   Republican feud fear as dissident is killed  ...
3                   Review: New Contemporaries 2000  ...
4             Art review:  Hollingsworth/Heyer  ...

                             News 16
News 17  \
0                   Finland 0 - 0 England                    Healy a
marked man
1   Why Wenger will stick to his Gunners  Out of luck England hit rock
bottom
2                   FTSE goes upwardly mobile                    At this price?  BP
Amoco
3           More cash on way for counties     Cairns carries Kiwis to
victory
4           Duisenberg in double trouble          Pru to cut pension
charges


                                             News 18  \
0                   Happy birthday Harpers & Queen
1                        Wilkinson out of his depth
2                                        Go fish
3   Year of Blanchflower's flourish when Spurs sto...
```

```
4                                    Art review: Paul Graham


                                       News 19  \
0  Win unlimited access to the Raindance film fes...
1                        Kinsella sparks Irish power play
2              Bosnian Serb blows himself up to evade law
3              New direct approach brings only pay-per-blues
4                      Shearer shot sparks Boro humiliation

                                       News 20  \
0  Labour pledges £800m to bridge north-south divide
1                         Brown banished as Scots rebound
2                            Orange float delayed to 2001
3                            Third Division  round-up
4  Ridsdale's lingering fears as Leeds revisit Tu...

                                       News 21  \
0       Wales: Lib-Lab pact firm despite resignation
1                         Battling Wales cling to lifeline
2  Angry factory workers root out fear, favours a...
3                            Second Division  round-up
4              Champions League: Rangers v Galatasaray


                        News 22  \
0                      Donald Dewar
1   Ehiogu close to sealing Boro move
2  Smith defied advice on dome payout
3              First Division  round-up
4   Champions League: Lazio v Arsenal


                                       News 23  \
0  Regenerating homes  regenerates well-being in ...
1                            Man-to-man marking
2                    Xerox takes the axe to jobs
3              McLean ends his career with a punch
4                            Lazio 1 - 1 Arsenal


                        News 24  \
0   Win £100 worth of underwear
1                  Match stats
2  Comment: Refugees in Britain
3     Heskey grabs triple crown
4            England in Pakistan


                                       News 25
0                            TV guide: Random views
1  French referee at centre of storm is no strang...
2                  Maverick who sparked the new intifada
3                  Weah on his way as City march on
```

```
[5 rows x 27 columns]

df.isnull().sum() #(df.isna().sum() gives same result)
#gives the sum of all null values columns-wise

Date          0
Label         0
News  1       0
News  2       0
News  3       0
News  4       0
News  5       0
News  6       0
News  7       0
News  8       0
News  9       0
News  10      0
News  11      0
News  12      0
News  13      0
News  14      0
News  15      0
News  16      0
News  17      0
News  18      0
News  19      0
News  20      0
News  21      0
News  22      0
News  23      1
News  24      3
News  25      3
dtype: int64

df.nunique()    #gives total no. of unique entries

Date        4101
Label          2
News  1     4074
News  2     4082
News  3     4081
News  4     4087
News  5     4070
News  6     4075
News  7     4078
News  8     4080
News  9     4085
News  10    4077
News  11    4074
```

```
News 12    4080
News 13    4082
News 14    4081
News 15    4083
News 16    4083
News 17    4077
News 18    4082
News 19    4078
News 20    4078
News 21    4077
News 22    4078
News 23    4072
News 24    4078
News 25    4066
dtype: int64
```

```python
df.columns   #give column names in the dataframe
```

```
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')
```

```python
df.shape
```

```
(4101, 27)
```

```python
#Get Feature Selection
```

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

```
{"type":"string"}
```

```python
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

news[0]

{"type":"string"}

X=news

type(news)

list
```

Get Feature Text conversion to Bag of Words

```
from sklearn.feature_extraction.text import CountVectorizer

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

X=cv.fit_transform(X)

X.shape

(4101, 48527)

X

<4101x48527 sparse matrix of type '<class 'numpy.int64'>'
      with 872946 stored elements in Compressed Sparse Row format>

y=df['Label']

y.shape

(4101,)
```

Splitting the dataset into the Training set and Test set

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3,
                                        stratify=y, random_state = 1)

X_train.shape, X_test.shape, y_train.shape, y_test.shape

((2870, 48527), (1231, 48527), (2870,), (1231,))
```

Model

```
from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier()
rf.fit(X_train,y_train)

RandomForestClassifier()
```

Model Prediction

```
y_pred=rf.predict(X_test)

y_pred.shape

(1231,)
```

Get Model Evaluation

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

accuracy_score(y_test,y_pred)

0.5174654752233956

confusion_matrix(y_test,y_pred)

array([[168, 413],
       [181, 469]])

print(classification_report(y_test,y_pred))

              precision    recall  f1-score   support

           0       0.48      0.29      0.36       581
           1       0.53      0.72      0.61       650

    accuracy                           0.52      1231
   macro avg       0.51      0.51      0.49      1231
weighted avg       0.51      0.52      0.49      1231


rf.predict_proba(X_test)

array([[0.58, 0.42],
       [0.49, 0.51],
       [0.6 , 0.4 ],
       ...,
       [0.6 , 0.4 ],
       [0.5 , 0.5 ],
       [0.57, 0.43]])
```