

OBJECTIVE: To develop a predictive model that uses financial market news headlines to classify and predict market movements.


Import Library

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

Data Import

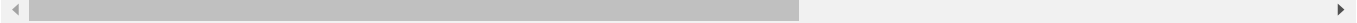
```
df = pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Financial%20Market%20News.csv',encoding ="ISO-8859-1")
```

```
df.head()
```




	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6	News 7	News 8	...	News 16	News 17
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	...	Finland 0 - 0 England	Healy a market mar
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	...	Why Wenger will stick to his Gunners	Out of luck Englanc hit rock bottom
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	...	FTSE goes upwardly mobile	At this price? BF Amocc
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	...	More cash on way for counties	Cairns carries Kiwis tc victory
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	...	Duisenberg in double trouble	Pru tc cul pensior charges

5 rows × 27 columns



```
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')

```

Feature Selection

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

```

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 Marriag
e 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 botto
n Wilkinsoe out of his death Kissella speaks Irish never play Beave banished as Caste reborn Battling Wales cling to lifeline This

```

```
df.index
```

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

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

```

)
4101

```

```

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

```

```
type(news)
```

```
list
```

```
news[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 Thea
tre review: The Circle Wales face a fraught night 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 & Qu
an His unlimited access to the Bridgese film festival Labour pledges 6800m to bridge north-south divide Wales: Lili Leh post-fin

```

```
X= news
```

```
type(X)
```

```
list
```

Text Conversion

```

from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(lowercase=True, ngram_range=(1,2))

```

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

```
X.shape
```

```
(4101, 668654)
```

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

```
y.shape
```

```
(4101,)
```

Training Model

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,random_state =2529,stratify=y,test_size=0.4)
```

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

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

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

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

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

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

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

```
array([[ 74, 700],
       [ 85, 782]])
```

Evaluation

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

```
precision    recall  f1-score   support

0           0.47      0.10      0.16         774
1           0.53      0.90      0.67         867

 accuracy          0.52         1641
 macro avg          0.50      0.50      0.41         1641
 weighted avg          0.50      0.52      0.43         1641
```

```
accuracy_score(y_test,y_pred)
```

```
0.5216331505179769
```

Start coding or [generate](#) with AI.

