

Financial Market News - Sentiment Analysis

#This is a data(dummy) of Financial market Top 25 News for the day and task is to train and Predict Model for Overall sentiment analysis


```
#import Library
import pandas as pd
```

```
import numpy as np
```

```
#import dataset
```


```
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	Mclroy'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 to 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	Prut 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')

```

Get 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
m Wilkieson out of his death Kinella speaks Irish news play Beowulf banished as Cato behaved 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 GudjonssonObituary: Brian WalshWorkplace blues leave employers in the redClassical review: RattleDan
ce review: Merce CunninghamGenetic tests to be used in setting premiumsOpera review: La BohèmePop review: Britney SpearsTheatre rev
iew: The CircleWales face a fraught nightUnder-21 round-upSmith off to blot his copybookFinns taking the mickeyPraise wasted as Br
own studies injury optionsIreland wary of minnowsFinland 0 - 0 EnglandHealy a marked manHappy birthday Harpers & QueenWin unlimited
access to the Bridgerton film festivalLakers pledge $800m to bridge north-south divideLalor: Lib Lab next firm despite resignationD

```

```
x = news
```

```
type(x)
```

```
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, 108682)
```

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

```
y.shape
```

```
(4101,)
```

### Get Train Test Split

```
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 = 2529)
```

```
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, confusion_matrix, accuracy_score
```

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

```
array([[104, 477],
       [136, 514]])
```

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

```
precision    recall  f1-score   support

0           0.43      0.18      0.25         581
1           0.52      0.79      0.63         650

accuracy          0.50         1231
macro avg          0.48      0.48      0.44         1231
weighted avg          0.48      0.50      0.45         1231
```