financial Market News-sentiment Analysis

```
import pandas as pd
```

Double-click (or enter) to edit

```
import numpy as np
```

```
rket
```

```
df=pd.read_csv(r'https://raw.githubusercontent.com/YBI-Foundation/Dataset/main/Financial%20Market%20News.csv',encoding="IS
```

```
df.head()
```

	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6	News 7	News 8	..
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	
1	02-01-2010	0	Warning from history	Investors flee to	Banks and	Review: Home	War jitters lead to sell	Your not-so-secret	Review: The	Review: Hysteria	

✓ 0s   completed at 5:44 PM   ● ✕

		crash	haven	favour				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	Republican feud fear as dissident is killed
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
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

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 Feture Selection

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

> 'Warning from history points to crashInvestors flee to dollar havenBanks and tobacco in favourReview: Llama Farmers
> War jitters lead to sell-offYour not-so-secret historyReview: The Northern SinfoniaReview: HysteriaReview: The Guar
> dsmanOpera: 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 ba

```
df.index
```

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

```
len(df.indexm)
```

> 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 r
> eview: RattleDance review: Merce CunninghamGenetic tests to be used in setting premiumsOpera review: La BohèmePop r
> eview: Britney SpearsTheatre review: The CircleWales face a fraught nightUnder-21 round-upSmith off to blot his co
> pybookFinns taking the mickeyPraise wasted as Brown studies injury optionsIreland wary of minnowsFinland 0 - 0 Engl

andHealy a marked manHappy birthday Harpers & QueenWin unlimited access to the Raindance film festivalLabour pledges £800m to bridge north-south divideWales: Lib-Lab pact firm despite resignationDonald DewarRegenerating homes, neg...

```
x=news
```

```
type(x)
```

```
list
```

Get feature Text Conversion to bag of word

```
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(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([[103, 478],
            [131, 519]])

print(classification_report(y_test,y_pred))

                   precision    recall  f1-score   support

                0       0.44      0.18      0.25       581
                1       0.52      0.80      0.63       650

         accuracy                           0.51      1231
        macro avg       0.48      0.49      0.44      1231
     weighted avg       0.48      0.51      0.45      1231
```