

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

In [2]: import numpy as np

In [4]: df=pd.read_csv(r"https://github.com/ybifoundation/Dataset/raw/main/Financial%20Market%20News.csv" , encoding="ISO-8859-1")

In [5]: df.head()
```

	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6	News 7	News 8	...	News 16	News 17	News 18	News 19	News 20	News 21	News 22	News 23	News 24
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 marked man	Happy birthday Harpers & Queen	Win unlimited access to the Rairdance film fes...	Labour pledges £800m to bridge north-south divide	Wales: Lib-Lab pact despite resignation	Donald Dewar	Regenerating homes regenerates well-being in ...	Win £100 worth of underwear
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 England hit rock bottom	Wilkinson out of 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
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? BP Amoco	Go fish	Bosnian Serb blows himself up to evade law	Orange float delayed to 2001	Angry factory workers root out fear, favours a...	Smith defied advice on dome payout	Xerox takes the axe to jobs	Comment: Refugees in Britain
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	Year of Blanchflower's flourish when Spurs sto...	New direct approach brings only pay-per-blues	Third Division round-up	Second Division round-up	First Division round-up	McLean ends his career with a punch	Heskey grabs triple crown
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 to cut pension charges	Art review: Paul Graham	Shearer shot sparks Boro humiliation	Ridsdale's lingering fears as Leeds revisit Tu...	Champions League: Rangers v Galatasaray	Champions League: Lazio v Arsenal	Lazio 1 - 1 Arsenal	England in Pakistan

5 rows x 27 columns

```
In [6]: 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 [7]: df.shape

Out[7]: (4101, 27)

In [8]: df.columns

Out[8]: 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')

In [9]: ' '.join(str(x) for x in df.iloc[1,2:27])

Out[9]: "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 Marriage 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 bottom Wilkinson out of 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 controversy"

In [10]: df.index

Out[10]: RangeIndex(start=0, stop=4101, step=1)

In [11]: len(df.index)

Out[11]: 4101

In [13]: news = []
for row in range(0 , len(df.index)):
    news.append(' '.join(str(x) for x in df.iloc[row,2:27]))

In [14]: type(news)

Out[14]: list

In [15]: news[0]

Out[15]: "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 Theatre 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 & Queen Win unlimited access to the Rairdance film festival Labour pledges £800m to bridge north-south divide Wales: Lib-Lab pact firm despite resignation Donald Dewar Regenerating homes regenerates well-being in people Win £100 worth of underwear TV guide: Random views"

In [16]: X = news

In [17]: type(X)

Out[17]: list

In [18]: from sklearn.feature_extraction.text import CountVectorizer

In [19]: cv = CountVectorizer(lowercase = True , ngram_range=(1,1))

In [20]: X = cv.fit_transform(X)

In [21]: X.shape

Out[21]: (4101, 48527)

In [22]: y = df['Label']

In [23]: y.shape

Out[23]: (4101,)

In [24]: from sklearn.model_selection import train_test_split

In [25]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.3, stratify = y , random_state = 72529)

In [26]: from sklearn.ensemble import RandomForestClassifier

In [27]: rf = RandomForestClassifier(n_estimators = 200)

In [28]: rf.fit(X_train,y_train)

Out[28]: RandomForestClassifier
RandomForestClassifier(n_estimators=200)

In [29]: y_pred = rf.predict(X_test)

In [30]: from sklearn.metrics import confusion_matrix, classification_report,accuracy_score

In [31]: confusion_matrix(y_test, y_pred)

Out[31]: array([[149, 432],
      [163, 487]], dtype=int64)

In [32]: print(classification_report(y_test, y_pred))

              precision    recall  f1-score   support

         0       0.48      0.26      0.33         581
         1       0.53      0.75      0.62         650

 accuracy          0.50
 macro avg         0.50      0.50      0.48         1231
 weighted avg         0.51      0.52      0.49         1231

In [ ]:
```