

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

In [2]: import numpy as np

In [3]: df = pd.read_csv('https://raw.githubusercontent.com/Lorddhaval/Dataset/main/Financial%20Market%20News.csv', encoding = "ISO-8859-1")

In [4]: 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	News 25
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 Raindance film fes...	Labour pledges £800m to bridge north-south divide	Wales: Lib-Lab pact firm despite resignation	Donald Dewar	Regenerating homes regenerates well-being in ...	Win £100 worth of underwear	TV guide: Random views
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	French referee at centre of storm is no strang...
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	Maverick who sparked the new intifada
3	04-01-2010	1	£750,000-a-goal Weah aims paring 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	Weah on his way as City march on
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	England given olive-branch reception

5 rows × 27 columns

```
In [5]: 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 [6]: df.shape

(4101, 27)

Out[6]:

In [7]: 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')

In [8]: ' '.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 O pera: 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 c entre of storm is no stranger to controversy"

In [9]: df.index

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

Out[9]:

In [10]: len(df.index)

4101

Out[10]:

In [11]: news = []
for row in range(0,len(df.index)):
    news.append(' '.join(str(x) for x in df.iloc[row,2:27]))

In [12]: type(news)

list

Out[12]:

In [13]: news[0]

"McIlroy's men catch cold from GudjonssonObituary: Brian WalshWorkplace blues leave employers in the redClassical review: RattleDance review: Merce CunninghamGenetic tests to be used in setting premiumsOpera review: La BohèmePop review: Britney SpearsTheatre review: The CircleWales face a fraught nightUnder-21 round-upSmith off to blot his copybookFinns taking the mickeyPraise wasted as Brown studies injury optionsIreland wary of minnowsFinland 0 - 0 EnglandHealy a marked manHappy bir thday Harpers & QueenWin unlimited access to the Raindance film festivalLabour pledges £800m to bridge north-south dividewales: Lib-Lab pact firm despite resignationDonald DewarRegenerating homes regenerates well-being in peopleWin £100 worth o f underwearTV guide: Random views"

In [14]: X = news
type(X)

list

Out[14]:

In [15]: from sklearn.feature_extraction.text import CountVectorizer

In [16]: cv = CountVectorizer(lowercase = True, ngram_range=(1,1))

In [17]: X = cv.fit_transform(X)

In [18]: X.shape

(4101, 108682)

Out[18]:

In [19]: y = df['Label']

In [20]: y.shape

(4101,)

Out[20]:

In [21]: from sklearn.model_selection import train_test_split

In [22]: X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.3, stratify = y, random_state=22529)

In [23]: from sklearn.ensemble import RandomForestClassifier

In [24]: rf = RandomForestClassifier(n_estimators=200)

In [25]: rf.fit(X_train, y_train)

Out[25]:
RandomForestClassifier
RandomForestClassifier(n_estimators=200)

In [26]: y_pred = rf.predict(X_test)

In [27]: from sklearn.metrics import classification_report,confusion_matrix,accuracy_score

In [28]: confusion_matrix(y_test, y_pred)

array([[ 91, 490],
       [115, 535]], dtype=int64)

Out[28]:

In [29]: print(classification_report(y_test, y_pred))

              precision    recall  f1-score   support

    0               0.44       0.16       0.23         581
    1               0.52       0.82       0.64         650

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

In [ ]:
```