

```
In [38]: import pandas as pd
```

```
In [39]: d=pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Financial%20Mar
```

```
In [40]: d.head()
```

Out[40]:

	Date	Label	News 1	News 2	News 3	News 4	News 5	News 6	News 7
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
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
	03-		Comment: Why Israel's	Court deals blow to	An ideal	World steps between	What the	Comment: Fear and	Poverty and resentment

```
In [41]: d.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 [42]: d.shape
```

```
Out[42]: (4101, 27)
```

```
In [43]: d.columns
```

```
Out[43]: 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

```
In [44]: ' '.join(str(x) for x in d.iloc[1,2:27])
```

```
Out[44]: "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 [45]: d.index
```

```
Out[45]: RangeIndex(start=0, stop=4101, step=1)
```

```
In [46]: len(d.index)
```

```
Out[46]: 4101
```

```
In [47]: news=[]
         for r in range(0,len(d.index)):
             news.append(' '.join(str(x) for x in d.iloc[r,2:27]))
```

```
In [48]: type(news)
```

```
Out[48]: list
```

```
In [49]: news[0]
```

```
Out[49]: "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 Engl and Healy a marked man Happy birthday Harpers & Queen Win unlimited access to the Raindance 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 [50]: x=news
```

```
In [51]: type(x)
```

```
Out[51]: list
```

## Get Feature Text Conversion to Bag of

# Words

```
In [52]: from sklearn.feature_extraction.text import CountVectorizer
```

```
In [53]: cv=CountVectorizer(lowercase=True,ngram_range=(1,1))
```

```
In [54]: x=cv.fit_transform(x)
```

```
In [55]: x.shape
```

```
Out[55]: (4101, 48527)
```

```
In [56]: y=d['Label']
```

```
In [57]: y.shape
```

```
Out[57]: (4101,)
```

## Train Test Split

```
In [58]: from sklearn.model_selection import train_test_split
```

```
In [60]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,stratify=y,random_state=42)
```

```
In [61]: from sklearn.ensemble import RandomForestClassifier
```

```
In [62]: l=RandomForestClassifier()
```

```
In [63]: l.fit(x_train,y_train)
```

```
Out[63]: 

▼ RandomForestClassifier



RandomForestClassifier()


```

```
In [64]: y_pred=l.predict(x_test)
```

```
In [66]: from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
```

```
In [68]: print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.48	0.31	0.38	581
1	0.53	0.70	0.60	650
accuracy			0.52	1231
macro avg	0.51	0.50	0.49	1231
weighted avg	0.51	0.52	0.50	1231

```
In [69]: confusion_matrix(y_test,y_pred)
```

```
Out[69]: array([[180, 401],  
               [195, 455]], dtype=int64)
```

```
In [ ]:
```