

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

```
In [3]: df = pd.read_csv('Data.csv', encoding = "ISO-8859-1")
df.head()
```

	Date	Label	Top1	Top2	Top3	Top4	Top5	Top6	Top7	Top8	...	Top16	Top17	Top18	Top19	Top20	Top21	Top22	Top23
0	2000-01-03	0	A hindrance to operations: extracts from the...	Scorecard	Hughes' instant hit buoys Blues	Jack gets his skates on at ice-cold Alex	Chaos as Maracana builds up for United	Depleted Leicester prevail as Elliott spoils E...	Hungry Spurs sense rich pickings	Gunners so wide of an easy target	...	Flintoff injury piles on woe for England	Hunters threaten Jospin with new battle of the...	Kohl's successor drawn into scandal	The difference between men and women	Sara Denver, nurse turned solicitor	Diana's landmine crusade put Tones in a panic	Yeltsin's resignation caught opposition flat-f...	Ru ro
1	2000-01-04	0	Scorecard	The best lake scene	Leader: German sleaze inquiry	Cheerio, boyo	The main recommendations	Has Cubie killed fees?	Has Cubie killed fees?	Has Cubie killed fees?	...	On the critical list	The timing of their lives	Dear doctor	Irish court halts IRA man's extradition to Nor...	Burundi peace initiative fades after rebels re...	PE points the way forward to the ECB	Campaigners keep up pressure on Nazi war crime...	Ra
2	2000-01-05	0	Coventry caught on counter by Flo	United's rivals on the road to Rio	Thatcher issues defence before trial by video	Police help Smith lay down the law at Everton	Tale of Trautmann bears two more retellings	England on the rack	Pakistan retaliate with call for video of Walsh	Cullinan continues his Cape monopoly	...	South Melbourne (Australia)	Necaxa (Mexico)	Real Madrid (Spain)	Raja Casablanca (Morocco)	Corinthians (Brazil)	Tony's pet project	Al Nassr (Saudi Arabia)	Hc
3	2000-01-06	1	Pilgrim knows how to progress	Thatcher facing ban	McIlroy calls for Irish fighting spirit	Leicester bin stadium blueprint	United braced for Mexican wave	Auntie back in fashion, even if the dress look...	Shoaib appeal goes to the top	Hussain hurt by 'shambles' but lays blame on e...	...	Putin admits Yeltsin quit to give him a head s...	BBC worst hit as digital TV begins to bite	How much can you pay for...	Christmas glitches	Upending a table, Chopping a line and Scoring ...	Scientific evidence 'unreliable', defence claims	Fusco wins judicial review in extradition case	R t Ru ad
4	2000-01-07	1	Hitches and Horlocks	Beckham off but United survive	Breast cancer screening	Alan Parker	Guardian readers: are you all whingers?	Hollywood Beyond	Ashes and diamonds	Whingers - a formidable minority	...	Most everywhere: UDIs	Most wanted: Chloe lunettes	Return of the cane 'completely off the agenda'	From Sleepy Hollow to Greenland	Blunkett outlines vision for over 11s	Embattled Dobson attacks 'play now, pay later'...	Doom and the Dome	W i di

5 rows × 27 columns

```
In [4]: train = df[df['Date']<='20150101']
test = df[df['Date']>'20141231']
```

```
In [5]: data = train.iloc[:,2:27]
data.replace("[^a-zA-z]", " ", regex = True, inplace = True)
```

```
In [6]: data.head()
```

	Top1	Top2	Top3	Top4	Top5	Top6	Top7	Top8	Top9	Top10	...	Top16	Top17	Top18	Top19	Top20	Top21	Top22	Top23
0	A hindrance to operations extracts from the...	Scorecard	Hughes' instant hit buoys Blues	Jack gets his skates on at ice cold Alex	Chaos as Maracana builds up for United	Depleted Leicester prevail as Elliott spoils E...	Hungry Spurs sense rich pickings	Gunners so wide of an easy target	Derby raise a glass to Strupar's debut double	Southgate strikes Leeds pay the penalty	...	Flintoff injury piles on woe for England	Hunters threaten Jospin with new battle of the...	Kohl's successor drawn into scandal	The difference between men and women	Sara Denver nurse turned solicitor	Diana's landmine crusade put Tories in a panic	Yeltsin's resignation caught opposition flat t	
1	Scorecard	The best lake scene	Leader German sleaze inquiry	Cheerio boyo	The main recommendations	Has Cubie killed fees	Has Cubie killed fees	Has Cubie killed fees	Hopkins furious at Foster's lack of Hannibal...	Has Cubie killed fees	...	On the critical list	The timing of their lives	Dear doctor	Irish court halts IRA man's extradition to Nor...	Burundi peace initiative fades after rebels re...	PE points the way forward to the ECB	Campaigners keep up pressure on Nazi war crime	
2	Coventry caught on counter by Flo	United's rivals on the road to Rio	Thatcher issues defence before trial by video	Police help Smith lay down the law at Everton	Tale of Trautmann bears two more retellings	England on the rack	Pakistan retaliate with call for video of Walsh	Cullinan continues his Cape monopoly	McGrath puts India out of their misery	Blair Witch bandwagon rolls on	...	South Melbourne Australia	Necaxa Mexico	Real Madrid Spain	Raja Casablanca Morocco	Corinthians Brazil	Tony's pet project	Al Nas Saudi Arat	
3	Pilgrim knows how to progress	Thatcher facing ban	McIlroy calls for Irish fighting spirit	Leicester bin stadium blueprint	United braced for Mexican wave	Auntie back in fashion even if the dress look...	Shoaib appeal goes to the top	Hussain hurt by shambles but lays blame on e...	England's decade of disasters	Revenge is sweet for jubilant Cronje	...	Putin admits Yeltsin quit to give him a head s...	BBC worst hit as digital TV begins to bite	How much can you pay for	Christmas glitches	Upending a table Chopping a line and Scoring ...	Scientific evidence unreliable defence claims	Fusco wi judic review extraditi ca	
4	Hitches and Horlocks	Beckham off but United survive	Breast cancer screening	Alan Parker	Guardian readers are you all whingers	Hollywood Beyond	Ashes and diamonds	Whingers a formidable minority	Alan Parker part two	Thuggery Toxins and Ties	...	Most everywhere UDIs	Most wanted Chloe lunettes	Return of the cane completely off the agenda	From Sleepy Hollow to Greenland	Blunkett outlines vision for over s	Embattled Dobson attacks play now pay later ...	Doom at the Dor	

5 rows × 25 columns

```
In [7]: list1 = [i for i in range(25)]
new_Index = [str(i) for i in list1]
data.columns = new_Index
data.head()
```

	0	1	2	3	4	5	6	7	8	9	...	15	16	17	18	19	20	21	22
0	A hindrance to operations extracts from the...	Scorecard	Hughes' instant hit buoys Blues	Jack gets his skates on at ice cold Alex	Chaos as Maracana builds up for United	Depleted Leicester prevail as Elliott spoils E...	Hungry Spurs sense rich pickings	Gunners so wide of an easy target	Derby raise a glass to Strupar's debut double	Southgate strikes Leeds pay the penalty	...	Flintoff injury piles on woe for England	Hunters threaten Jospin with new battle of the...	Kohl's successor drawn into scandal	The difference between men and women	Sara Denver nurse turned solicitor	Diana's landmine crusade put Tories in a panic	Yeltsin's resignation caught opposition flat t	
1	Scorecard	The best lake scene	Leader German sleaze inquiry	Cheerio boyo	The main recommendations	Has Cubie killed fees	Has Cubie killed fees	Has Cubie killed fees	Hopkins furious at Foster's lack of Hannibal...	Has Cubie killed fees	...	On the critical list	The timing of their lives	Dear doctor	Irish court halts IRA man's extradition to Nor...	Burundi peace initiative fades after rebels re...	PE points the way forward to the ECB	Campaigners keep up pressure on Nazi war crime	
2	Coventry caught on counter by Flo	United's rivals on the road to Rio	Thatcher issues defence before trial by video	Police help Smith lay down the law at Everton	Tale of Trautmann bears two more retellings	England on the rack	Pakistan retaliate with call for video of Walsh	Cullinan continues his Cape monopoly	McGrath puts India out of their misery	Blair Witch bandwagon rolls on	...	South Melbourne Australia	Necaxa Mexico	Real Madrid Spain	Raja Casablanca Morocco	Corinthians Brazil	Tony's pet project	Al Nas Saudi Arat	
3	Pilgrim knows how to progress	Thatcher facing ban	McIlroy calls for Irish fighting spirit	Leicester bin stadium blueprint	United braced for Mexican wave	Auntie back in fashion even if the dress look...	Shoaib appeal goes to the top	Hussain hurt by shambles but lays blame on e...	England's decade of disasters	Revenge is sweet for jubilant Cronje	...	Putin admits Yeltsin quit to give him a head s...	BBC worst hit as digital TV begins to bite	How much can you pay for	Christmas glitches	Upending a table Chopping a line and Scoring ...	Scientific evidence unreliable defence claims	Fusco wi judic review extraditi ca	
4	Hitches and Horlocks	Beckham off but United survive	Breast cancer screening	Alan Parker	Guardian readers are you all whingers	Hollywood Beyond	Ashes and diamonds	Whingers a formidable minority	Alan Parker part two	Thuggery Toxins and Ties	...	Most everywhere UDIs	Most wanted Chloe lunettes	Return of the cane completely off the agenda	From Sleepy Hollow to Greenland	Blunkett outlines vision for over s	Embattled Dobson attacks play now pay later ...	Doom at the Dor	

5 rows × 25 columns

```
In [9]: for index in new_Index:
data[index] = data[index].str.lower()
data.head()
```

	0	1	2	3	4	5	6	7	8	9	...	15	16	17	18	19	20	21	22
0	a hindrance to operations extracts from the...	scorecard	hughes' instant hit buoys blues	jack gets his skates on at ice cold alex	chaos as maracana builds up for united	depleted leicester prevail as elliot spoils e...	hungry spurs sense rich pickings	gunners so wide of an easy target	derby raise a glass to strupar's debut double	southgate strikes Leeds pay the penalty	...	flintoff injury piles on woe for england	hunters threaten jospin with new battle of the...	kohl's successor drawn into scandal	the difference between men and women	sara denver nurse turned solicitor	diana's landmine crusade put tories in a panic	yeltsin's resignation caught opposition flat l...	
1	scorecard	the best lake scene	leader german sleaze inquiry	cheerio boyo	the main recommendations	has cubie killed fees	has cubie killed fees	has cubie killed fees	hopkins furious at foster's lack of hannibal...	has cubie killed fees	...	on the critical list	the timing of their lives	dear doctor	irish court halts ira man's extradition to nor...	burundi peace initiative fades after rebels re...	pe points the way forward to the ecb	campaigners keep up pressure on nazi war crime...	
2	coventry caught on counter by flo	united's rivals on the road to rio	thatcher issues defence before trial by video	police help smith lay down the law at everton	tale of trautmann bears two more retellings	england on the rack	pakistan retaliate with call for video of wals	cullinan continues his cape monopoly	mcgrath puts india out of their misery	blair witch bandwagon rolls on	...	south melbourne australia	necaxa mexico	real madrid spain	raja casablanca morocco	corinthians brazil	tony's pet project	al nassr saudi arabia	
3	pilgrim knows how to progress	thatcher facing ban	mcilroy calls for irish fighting spirit	leicester bin stadium blueprint	united braced for mexican wave	auntie back in fashion even if the dress look...	shoaib appeal goes to the top	hussain hurt by shambles but lays blame on e...	england's decade of disasters	revenge is sweet for jubilant cronje	...	putin admits yeltsin quit to give him a head s...	bbc worst hit as digital tv begins to bite	how much can you pay for	christmas glitches	upending a table chopping a line and scoring ...	scientific evidence unreliable defence claims	fusco wins judicial review in extradition case	
4	hitches and horlocks	beckham off but united survive	breast cancer screening	alan parker	guardian readers are you all whingers	hollywood beyond	ashes and diamonds	whingers a formidable minority	alan parker part two	thuggery toxins and ties	...	most everywhere udis	most wanted chloe lunettes	return of the cane completely off the agenda	from sleepy hollow to greenland	blunkett outlines vision for over s	embattled dobson attacks play now pay later ...	doom and the dome	

5 rows × 25 columns

```
In [11]: headline = []
for row in range(0,len(data.index)):
    headline.append(' '.join(str(x) for x in data.iloc[row, 0:25]))
```

```
In [12]: headline[0]
```

```
Out[12]: 'a hindrance to operations extracts from the leaked reports scorecard hughes' instant hit buoys blues jack gets his skates on at ice cold alex chaos as maracana builds up for united depleted leicester prevail as elliot spoils everton's party hungry spurs sense rich pickings gunners so wide of an easy target derby raise a glass to strupar's debut double southgate strikes Leeds pay the penalty hammers hand robson a youthful lesson saints party like it's wear wol ves have turned into lamb's stump mike catches testy toughie's taunt langer escapes to hit flintoff injury piles on woe for england hunters threaten jospin with new battle of the somme kohl's successor drawn into scandal the difference between men and women sara denver nurse turned solicitor diana's landmine crusade put tories in a panic yeltsin's resignation caught opposition flat footed russian roulette sold out recovering a title'
```

```
In [13]: ' '.join(str(x) for x in data.iloc[1,0:25])
```

```
Out[13]: 'scorecard the best lake scene leader german sleaze inquiry cheerio boyo the main recommendations has cubie killed fees has cubie killed fees has cubie killed fees hopkins furious at foster's lack of hannibal appetite has cubie killed fees a tale of two tails i say what i like and i like what i say elbows eyes and nipples task force to assess risk of asteroid collision how i found myself at last on the critical list the timing of their lives dear doctor irish court halts ira man's extradition to northern ireland burundi peace initiative fades after rebels reject mandela as mediator pe points the way forward to the ECB campaigners keep up pressure on nazi war crimes suspect jane ratcliffe yet more things you wouldn't know without the movies millennium bug fails to bite'
```

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

```
In [16]: countVector = CountVectorizer(ngram_range = (2,2))
train_dataset = countVector.fit_transform(headline)
```

```
In [18]: randomforest = RandomForestClassifier(n_estimators = 200, criterion = 'entropy')
randomforest.fit(train_dataset, train['Label'])
```

```
Out[18]: RandomForestClassifier(criterion='entropy', n_estimators=200)
```

```
In [22]: test_transform= []
for row in range(0,len(test.index)):
    test_transform.append(' '.join(str(x) for x in test.iloc[row,2:27]))
test_dataset = countVector.transform(test_transform)
predictions = randomforest.predict(test_dataset)
```

```
In [23]: from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
```

```
In [24]: matrix=confusion_matrix(test['Label'],predictions)
print(matrix)
score=accuracy_score(test['Label'],predictions)
print(score)
report=classification_report(test['Label'],predictions)
print(report)
```

[[137 49] [4 188]]					
0.8597883597883598	precision	recall	f1-score	support	
0	0.97	0.74	0.84	186	
1	0.79	0.98	0.88	192	
			0.86	378	
accuracy			0.86	378	
macro avg	0.88	0.86	0.86	378	
weighted avg	0.88	0.86	0.86	378	