Stock Sentiment Analysis using News Headlines

```
import pandas as pd
```

```
df=pd.read_csv('Data.csv', encoding = "ISO-8859-1")
```

```
df.head()
```

	Date	Label	Top1	Top2	Top3	Top4	Top5	Top6
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...
1	2000-01-04	0	Scorecard	The best lake scene	Leader: German sleaze inquiry	Cheerio, boyo	The main recommendations	Has Cubie killed fees?
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
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...
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

5 rows × 27 columns

```
df.tail()
```

	Date	Label	Top1	Top2	Top3	Top4	Top5	To
4096	2016-06-27	0	Barclays and RBS shares suspended from trading...	Pope says Church should ask forgiveness from g...	Poland 'shocked' by xenophobic abuse of Poles ...	There will be no second referendum, cabinet ag...	Scotland welcome to join EU, Merkel ally says	Sterli dips bel Friday's 3 year l amid
4097	2016-06-28	1	2,500 Scientists To Australia: If You	The personal details of 112,000 French	S&P cuts United Kingdom sovereign	Huge helium deposit found in	CEO of the South African state broadcaster	Brexit c investc $2 trillic the wo

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

```
len(train)
```

```
3975
```

```
len(test)
```

```
378
```

```
# Removing punctuations
data=train.iloc[:,2:27]
data.replace("[^a-zA-Z]"," ",regex=True, inplace=True)

# Renaming column names for ease of access
list1= [i for i in range(25)]
new_Index=[str(i) for i in list1]
data.columns= new_Index
data.head(5)
```

	0	1	2	3	4	5	6	
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	Gun so wic an ta
1	Scorecard	The best lake scene	Leader German sleaze inquiry	Cheerio boyo	The main recommendations	Has Cubie killed fees	Has Cubie killed fees	Has C killed
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	Cull contir his C mono
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	Hus hu sham but blam
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	Whin formid min

5 rows × 25 columns

```
# Convertng headlines to lower case
for index in new_Index:
    data[index]=data[index].str.lower()
data.head(1)
```

	0	1	2	3	4	5	6	7	8	
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	sou lee

1 rows × 25 columns

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

```
'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 tal
e 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 critica
l 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 me
```

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

```
headlines[0]
```

```
'a  hindrance to operations   extracts from the leaked reports scorecard hughes  instant hit buoys blues jack gets his skates on a
t ice cold alex chaos as maracana builds up for united depleted leicester prevail as elliott spoils everton s party hungry spurs s
ense rich pickings gunners so wide of an easy target derby raise a glass to strupar s debut double southgate strikes  leeds pay th
e penalty hammers hand robson a youthful lesson saints party like it s      wear wolves have turned into lambs stump mike catches
testy gough 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 lan
dmine crusade put tories in a panic yeltsin s resignation caught opposition flat footed russian roulette sold out recovering a tit
le'
```

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.ensemble import RandomForestClassifier
```

```
## implement BAG OF WORDS
countvector=CountVectorizer(ngram_range=(2,2))
traindataset=countvector.fit_transform(headlines)
```

```
# implement RandomForest Classifier
randomclassifier=RandomForestClassifier(n_estimators=200,criterion='entropy')
randomclassifier.fit(traindataset,train['Label'])
```

```
RandomForestClassifier(criterion='entropy', n_estimators=200)
```

```
## Predict for the Test Dataset
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 = randomclassifier.predict(test_dataset)
```

```
## Import library to check accuracy
from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
```

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

```
[[138  48]
 [ 11 181]]
0.843915343915344
              precision    recall  f1-score   support
```

```
              0       0.93      0.74      0.82       186
              1       0.79      0.94      0.86       192

       accuracy                           0.84       378
      macro avg       0.86      0.84      0.84       378
   weighted avg       0.86      0.84      0.84       378
```

```
import seaborn as sns
import matplotlib.pyplot as plt

#Plot the confusion matrix.
sns.heatmap(matrix,
            annot=True,
            fmt='g')

plt.ylabel('Prediction',fontsize=13)
plt.xlabel('Actual',fontsize=13)
plt.title('Confusion Matrix',fontsize=17)
plt.show()
```