Sentiment analysis

Dataset:

Collected dataset are in bangle.

Total data : 27879

Training sample: 19500

Testing sample: 8379

Preprocessing:

Converted each data by Tfidf Vectorizer.

Random seed = 2

Scoring = 'accuracy'

```
kfold = model_selection.KFold(n_splits=10, random_state=seed)
```

Classifier used :

```
MultinomialNB
LogisticRegression
KNeighborsClassifier(n_neighbors=3)
```

Result:

Based on validation/Training dataset :

MultinomialNB [Accuracy : 40.11536208724385 % Deviation : 0.8981016722931661 %]

LogisticRegression [Accuracy : 40.36060470811732 % Deviation : 0.9061758926962192 %]

KNeighborsClassifier [Accuracy : 30.241871802794655 % Deviation : 1.967265622325506 %]

Now using MultinomialNB again for Matrix :

Accuracy : 40.11536208724385 %

Confusion Matrix:

```
[[  0   0   0   0   0   0   0 350   0   0   0   0   0   0
   0   0   0   0]
[  0   0   0   0   0   0   0 331   0   0   0   0   0   0
   0   1   0   0]
[  0   0   0   0   0   0   0 403   0   0   0   0   0   0
   0   0  11   0]
[  0   0   0   0   0   0   0 416   0   0   0   0   0   0
   0   0   0   1]
[  0   0   0   0   0   0   2 179   1   0   0   0   0   0
   0   0   0   0]
[  0   0   0   0   0   0   0 291   0   0   0   0   0   0
   0   0  14   0]
[  0   0   0   0   0   0   5 798   0   0   0   0   0   0
   0   0   3   0]
[  0   1   0   0   0   0   9 5504   1   0   0   0   0   0
   0   4  17   0]
[  0   0   0   0   0   0   2 631   2   0   0   0   0   0
   0   0   1   0]
[  0   0   0   0   0   0   0 228   0   0   0   0   0   0
   0   0   0   0]
[  0   0   0   0   0   0   1 312   0   0   0   0   0   0
   0   0   6   0]
[  0   0   0   0   0   0   0 359   0   0   0   0   0   0
   0   0   1   0]
[  0   0   0   0   0   0   1 752   0   0   0   0   0   0
```

```
   0   0  40   0]
[  0   0   0   0   0   0   1 248   0   0   0   0   0   0
   0   0   1   1]
[  0   0   0   0   0   0   1 183   0   0   0   0   0   0
   0   0   0   0]
[  0   0   0   0   0   0   0 656   0   0   0   0   0   0
   0  12  22   0]
[  0   0   0   0   0   0   0 1463   0   0   0   0   0   0
   0   0  34   0]
[  0   0   0   0   0   0   1 562   0   0   0   0   0   0
   0   1   2   0]]
 'precision', 'predicted', average, warn_for)
```

Report:

	precision	recall	f1-score	support
Angry (রাগান্বিত)	0.00	0.00	0.00	350
Bad (খারাপ)	0.00	0.00	0.00	332
Blush(গোলাপী আভা)	0.00	0.00	0.00	414
Consciousness (চেতনাবাদ)	0.00	0.00	0.00	417
Evil (জঘন্য)	0.00	0.00	0.00	182
Fail (ব্যর্থ)	0.00	0.00	0.00	305
HaHa(হা হা)	0.22	0.01	0.01	806
Like (ভাল)	0.40	0.99	0.57	5536
Love(ভালবাসা)	0.50	0.00	0.01	636
Protestant (প্রতিবাদমূলক)	0.00	0.00	0.00	228
Provocative (উস্কানিমুলক)	0.00	0.00	0.00	319
Rocking (আন্দোলিত হওয়া)	0.00	0.00	0.00	360
Sad (দু: খিত)	0.00	0.00	0.00	793

Shocking (অতিশয় বেদনাদায়ক)	0.00	0.00	0.00	251
Skeptical (সন্দেহপ্রবণ)	0.00	0.00	0.00	184
Skip (বোঝতে পারছি না)	0.67	0.02	0.03	690
Smiley (স্মাইলি)	0.22	0.02	0.04	1497
WOW(কি দারুন)	0.00	0.00	0.00	566
avg / total	0.25	0.40	0.24	13866

Given Data : আমি ভালো আছি

Prediction on given data : ['Like (ভাল)']

Fake News

Dataset:

Dataset is is English.

Total data: 82878

Training sample: 41439

Testing sample: 41439

Preprocessing:

Converted each data by Tfidf Vectorizer.

Random seed = 2

Scoring = ‘accuracy’

```
kfold = model_selection.KFold(n_splits=10, random_state=seed)
```

Classifier used :

```
MultinomialNB
LogisticRegression
KNeighborsClassifier(n_neighbors=3)
```

Result:

Based on validation/Training dataset :

MNB [Accuracy : 76.88675478177535 % Deviation : 2.4051092352520644 %]

LR [Accuracy : 90.52719322764844 % Deviation : 1.7858748260819919 %]

KNN [Accuracy : 71.32961306552728 % Deviation : 2.020683560805938 %]

Accuracy : 78.40909090909091%

Confusion Matrix:

[[903 673]

[11 1581]]

Report:

	precision	recall	f1-score	support
FAKE	0.99	0.57	0.73	1576
REAL	0.70	0.99	0.82	1592
avg / total	0.84	0.78	0.77	3168

Given Data : There are two possible explanations.

Prediction on given data : ['FAKE']

Apartment price prediction

Dataset:

Collected dataset is in numeric

Total data : 21614

Training sample: 10807

Testing sample: 10807

Preprocessing:

Dropped
'id','price','date','waterfront','lat','long','yr_renovated'

From dataset for better prediction.

Classifier used :

LinearDiscriminantAnalysis
KNeighborsClassifier
LinearRegression
GradientBoostingRegressor

Result:

Based on validation/Training dataset :

Accuracy:

LinearDiscriminantAnalysis : 0.8050337744054779 %

KNeighborsClassifier : 0.3423706856666975 %

LinearRegression : 63.62146648132394 %

GradientBoostingRegressor : 79.11162320580092 %

Root Mean Square in Linear Regression : 223666.15605586465

Given Data : 4,2,1200,5500,1,0,5,5,1200,0,1970,98178,1300,5000

Prediction on given data : [57170.33275595] Taka