C:\Users\Nelson\SG-DAT-NL\Project>python NL\_Project\_naive\_Bayes.py

NL\_Project\_naive\_Bayes.py:22: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy

dataset.dropna(inplace = True)

>>>BernoulliNB:

-----

Accuracy: 52.82%

Accuracy on training data: 52.86%

Accuracy on test data: 52.72%

-----

Accuracy score: 52.82%

Accuracy score on training data: 52.86%

Accuracy score on test data: 52.72%

-----

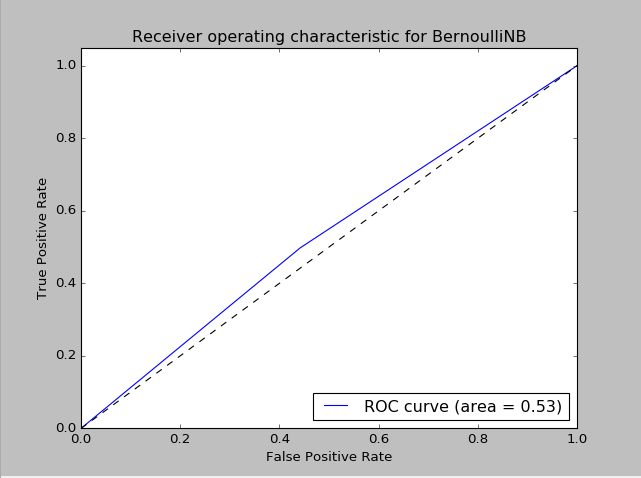
Confusion matrix (on entire data):

[[12056 9535]

[10640 10533]]

-----

ROC (on entire data)



>>>Logistic Regression with binary features:

-----

Accuracy: 53.14%

Accuracy on training data: 53.28%

Accuracy on test data: 52.71%

-----

Accuracy score: 53.14%

Accuracy score on training data: 53.28%

Accuracy score on test data: 52.71%

-----

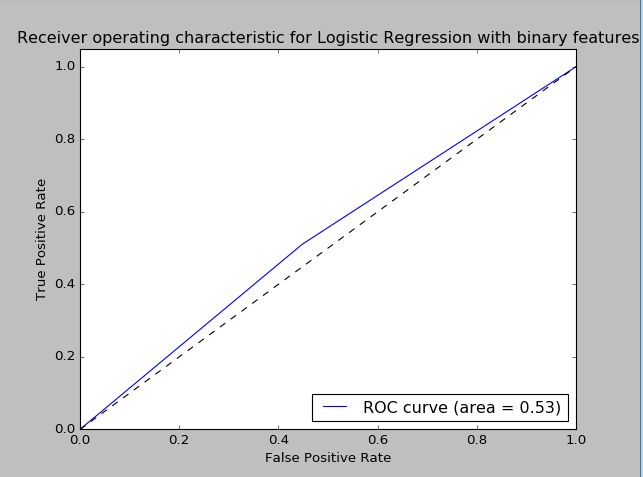
Confusion matrix (on entire data):

[[11918 9673]

[10366 10807]]

-----

ROC (on entire data)



NL\_Project\_naive\_Bayes.py:124: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy

dataset.dropna(inplace = True)

>>>GaussianNB:

-----

Accuracy: 53.29%

Accuracy on training data: 53.25%

Accuracy on test data: 53.41%

-----

Accuracy score: 53.29%

Accuracy score on training data: 53.25%

Accuracy score on test data: 53.41%

-----

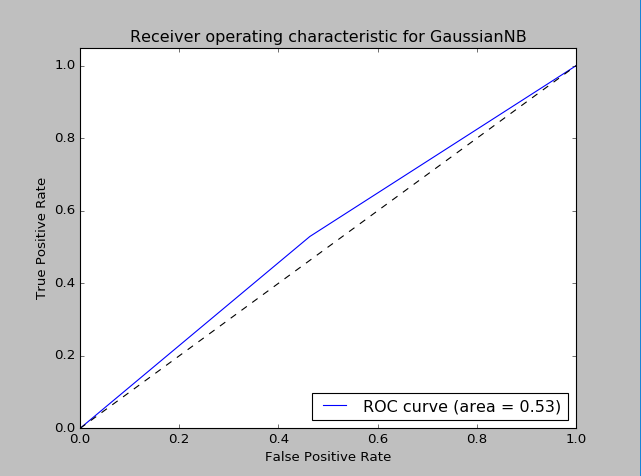
Confusion matrix (on entire data):

[[12165 10504]

[10451 11745]]

-----

ROC (on entire data)



>>>Logistic Regression with continuous features:

-----

Accuracy: 51.07%

Accuracy on training data: 50.69%

Accuracy on test data: 52.21%

-----

Accuracy score: 51.07%

Accuracy score on training data: 50.69%

Accuracy score on test data: 52.21%

-----

Confusion matrix (on entire data):

[[20885 1784]

[20169 2027]]

-----

ROC (on entire data)

