Before log:

(Udacity) yaser@yaser-Inspiron-7520:~/Udacity/P5/P5packup./scripts$ ./poi\_id\_1.py

/usr/local/lib/python2.7/dist-packages/sklearn/cross\_validation.py:44: DeprecationWarning: This module was deprecated in version 0.18 in favor of the model\_selection module into which all the refactored classes and functions are moved. Also note that the interface of the new CV iterators are different from that of this module. This module will be removed in 0.20.

"This module will be removed in 0.20.", DeprecationWarning)

Dataset Exploration:

Data points: 146

# of POI: 18

# of not POI: 128

16 best features: ['salary', 'to\_messages', 'deferral\_payments', 'total\_payments', 'loan\_advances', 'bonus', 'director\_fees', 'total\_stock\_value', 'shared\_receipt\_with\_poi', 'from\_poi\_to\_this\_person', 'exercised\_stock\_options', 'other', 'from\_this\_person\_to\_poi', 'deferred\_income', 'expenses', 'restricted\_stock']

['poi', 'salary', 'to\_messages', 'deferral\_payments', 'total\_payments', 'loan\_advances', 'bonus', 'director\_fees', 'total\_stock\_value', 'shared\_receipt\_with\_poi', 'from\_poi\_to\_this\_person', 'exercised\_stock\_options', 'other', 'from\_this\_person\_to\_poi', 'deferred\_income', 'expenses', 'restricted\_stock', 'long\_term\_incentive']

Accuracy with LR: 0.767442

recall with LR: 0.400000

precision with LR: 0.222222

the F1 score with LR is: 0.285714

=========================================

Accuracy with NB: 0.906977

recall with NB: 0.600000

precision with NB: 0.600000

the F1 score with NB is: 0.600000

============= look above ============================

Accuracy of scalled features NB: 0.441860

recall of scalled features NB: 1.000000

precision of scalled features NB: 0.172414

the F1 score of scalled features NB: 0.294118

=================BernoulliNB========================

Accuracy of scalled features BernoulliNB: 0.790698

recall of scalled features BernoulliNB: 0.800000

precision of scalled features BernoulliNB: 0.333333

the F1 score of scalled featuresBernoulliNB: 0.470588

=========================================

Accuracy with DTree: 0.837209

recall with DTree: 0.000000

precision with DTree: 0.000000

./poi\_id\_1.py:336: RuntimeWarning: invalid value encountered in double\_scalars

F1 = 2 \* (precision \* recall) / (precision + recall)

the F1 score with DTree is: nan

=========================================

Accuracy of scalled featuresand tunned with DTree: 0.883721

recall of scalled features and tunned with DTree: 0.000000

/usr/local/lib/python2.7/dist-packages/sklearn/metrics/classification.py:1113: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 due to no predicted samples.

'precision', 'predicted', average, warn\_for)

precision of scalled features and tunned with DTree: 0.000000

./poi\_id\_1.py:368: RuntimeWarning: invalid value encountered in double\_scalars

F1 = 2 \* (precision \* recall) / (precision + recall)

the F1 score of scalled features and tunned with DTree is: nan

=========================================

Accuracy with MLPClassifier: 0.651163

recall with MLPClassifier: 0.400000

precision with MLPClassifier: 0.142857

the F1 score with MLPClassifier is: 0.210526

=========================================

/usr/local/lib/python2.7/dist-packages/sklearn/neural\_network/multilayer\_perceptron.py:563: ConvergenceWarning: Stochastic Optimizer: Maximum iterations reached and the optimization hasn't converged yet.

% (), ConvergenceWarning)

Accuracy with MLPClassifier scaled features: 0.883721

recall with MLPClassifier scaled features: 0.400000

precision with MLPClassifier scaled features: 0.500000

the F1 score with MLPClassifier scaled features is 0.444444

=========================================

Accuracy with tunned MLPClassifier: 0.883721

recall with tunned MLPClassifier: 0.000000

precision with tunned MLPClassifier: 0.000000

./poi\_id\_1.py:433: RuntimeWarning: invalid value encountered in double\_scalars

F1 = 2 \* (precision \* recall) / (precision + recall)

the F1 score with tunned MLPClassifier is: nan

=========================================

================= LR ========================

LogisticRegression(C=1.0, class\_weight=None, dual=False, fit\_intercept=True,

intercept\_scaling=1, max\_iter=100, multi\_class='ovr', n\_jobs=1,

penalty='l2', random\_state=None, solver='liblinear', tol=0.0001,

verbose=0, warm\_start=False)

Accuracy: 0.78420 Precision: 0.18003 Recall: 0.17400 F1: 0.17696 F2: 0.17517

Total predictions: 15000 True positives: 348 False positives: 1585 False negatives: 1652 True negatives: 11415

================== NB =======================

GridSearchCV(cv=None, error\_score='raise', estimator=GaussianNB(priors=None),

fit\_params={}, iid=True, n\_jobs=1,

param\_grid={'priors': [None, [0.1, 0.9], [0.2, 0.8]]},

pre\_dispatch='2\*n\_jobs', refit=True, return\_train\_score=True,

scoring=None, verbose=0)

Accuracy: 0.73827 Precision: 0.22626 Recall: 0.39800 F1: 0.28851 F2: 0.34555

Total predictions: 15000 True positives: 796 False positives: 2722 False negatives: 1204 True negatives: 10278

================ NB Scaled Features================

GaussianNB(priors=None)

Accuracy: 0.73900 Precision: 0.22604 Recall: 0.39500 F1: 0.28753 F2: 0.34363

Total predictions: 15000 True positives: 790 False positives: 2705 False negatives: 1210 True negatives: 10295

=============== BernoulliNB ================

GridSearchCV(cv=None, error\_score='raise',

estimator=BernoulliNB(alpha=1.0, binarize=0.0, class\_prior=None, fit\_prior=True),

fit\_params={}, iid=True, n\_jobs=1,

param\_grid={'binarize': [0.0, 0.1, 0.5, 0.05, 0.01, 0.001], 'alpha': [0.05, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], 'fit\_prior': [True, False]},

pre\_dispatch='2\*n\_jobs', refit=True, return\_train\_score=True,

scoring=None, verbose=0)

Accuracy: 0.71840 Precision: 0.29121 Recall: 0.77550 F1: 0.42342 F2: 0.58195

Total predictions: 15000 True positives: 1551 False positives: 3775 False negatives: 449 True negatives: 9225

================ DTree ========================

DecisionTreeClassifier(class\_weight=None, criterion='gini', max\_depth=None,

max\_features=None, max\_leaf\_nodes=None,

min\_impurity\_split=1e-07, min\_samples\_leaf=1,

min\_samples\_split=2, min\_weight\_fraction\_leaf=0.0,

presort=False, random\_state=None, splitter='best')

Accuracy: 0.79733 Precision: 0.22632 Recall: 0.21500 F1: 0.22051 F2: 0.21717

Total predictions: 15000 True positives: 430 False positives: 1470 False negatives: 1570 True negatives: 11530

==================== NN ============================

MLPClassifier(activation='relu', alpha=0.0001, batch\_size='auto', beta\_1=0.9,

beta\_2=0.999, early\_stopping=False, epsilon=1e-08,

hidden\_layer\_sizes=(100,), learning\_rate='constant',

learning\_rate\_init=0.001, max\_iter=200, momentum=0.9,

nesterovs\_momentum=True, power\_t=0.5, random\_state=None,

shuffle=True, solver='adam', tol=0.0001, validation\_fraction=0.1,

verbose=False, warm\_start=False)

Accuracy: 0.65047 Precision: 0.12750 Recall: 0.27750 F1: 0.17472 F2: 0.22464

Total predictions: 15000 True positives: 555 False positives: 3798 False negatives: 1445 True negatives: 9202

==================== NN Scalled F =================

MLPClassifier(activation='relu', alpha=0.0001, batch\_size='auto', beta\_1=0.9,

beta\_2=0.999, early\_stopping=False, epsilon=1e-08,

hidden\_layer\_sizes=(100,), learning\_rate='constant',

learning\_rate\_init=0.001, max\_iter=200, momentum=0.9,

nesterovs\_momentum=True, power\_t=0.5, random\_state=None,

shuffle=True, solver='adam', tol=0.0001, validation\_fraction=0.1,

verbose=False, warm\_start=False)

Accuracy: 0.65700 Precision: 0.12763 Recall: 0.26950 F1: 0.17323 F2: 0.22049

Total predictions: 15000 True positives: 539 False positives: 3684 False negatives: 1461 True negatives: 9316

==================== NN Tunned =====================

GridSearchCV(cv=None, error\_score='raise',

estimator=MLPClassifier(activation='relu', alpha=0.0001, batch\_size='auto', beta\_1=0.9,

beta\_2=0.999, early\_stopping=False, epsilon=1e-08,

hidden\_layer\_sizes=(100,), learning\_rate='constant',

learning\_rate\_init=0.001, max\_iter=200, momentum=0.9,

nesterovs\_momentum=True, power\_t=0.5, random\_state=None,

shuffle=True, solver='adam', tol=0.0001, validation\_fraction=0.1,

verbose=False, warm\_start=False),

fit\_params={}, iid=True, n\_jobs=1,

param\_grid={'activation': ['logistic', 'relu'], 'max\_iter': [150, 200], 'hidden\_layer\_sizes': [10, 20]},

pre\_dispatch='2\*n\_jobs', refit=True, return\_train\_score=True,

scoring=None, verbose=0)

Accuracy: 0.84680 Precision: 0.13835 Recall: 0.02850 F1: 0.04726 F2: 0.03388

Total predictions: 15000 True positives: 57 False positives: 355 False negatives: 1943 True negatives: 12645

=============== DT- tunning ======================

GridSearchCV(cv=None, error\_score='raise',

estimator=DecisionTreeClassifier(class\_weight=None, criterion='gini', max\_depth=None,

max\_features=None, max\_leaf\_nodes=None,

min\_impurity\_split=1e-07, min\_samples\_leaf=1,

min\_samples\_split=2, min\_weight\_fraction\_leaf=0.0,

presort=False, random\_state=None, splitter='best'),

fit\_params={}, iid=True, n\_jobs=1,

param\_grid={'min\_samples\_split': [2, 3], 'max\_features': ['sqrt', 'log2'], 'criterion': ('gini', 'entropy'), 'max\_depth': [2, 3], 'min\_samples\_leaf': [2, 3, 5, 10]},

pre\_dispatch='2\*n\_jobs', refit=True, return\_train\_score=True,

scoring=None, verbose=0)

Accuracy: 0.84973 Precision: 0.30462 Recall: 0.09900 F1: 0.14943 F2: 0.11445

Total predictions: 15000 True positives: 198 False positives: 452 False negatives: 1802 True negatives: 12548

After Log

(Udacity) yaser@yaser-Inspiron-7520:~/Udacity/P5/P5\_YasserAlnakhli/scripts$ ./poi\_id\_2.py

/usr/local/lib/python2.7/dist-packages/sklearn/cross\_validation.py:44: DeprecationWarning: This module was deprecated in version 0.18 in favor of the model\_selection module into which all the refactored classes and functions are moved. Also note that the interface of the new CV iterators are different from that of this module. This module will be removed in 0.20.

"This module will be removed in 0.20.", DeprecationWarning)

Dataset Exploration:

Data points: 146

# of POI: 18

# of not POI: 128

/usr/local/lib/python2.7/dist-packages/sklearn/feature\_selection/univariate\_selection.py:113: UserWarning: Features [19 20 21 22 23] are constant.

UserWarning)

/usr/local/lib/python2.7/dist-packages/sklearn/feature\_selection/univariate\_selection.py:114: RuntimeWarning: invalid value encountered in divide

f = msb / msw

16 best features: ['salary', 'total\_stock\_value\_log', 'total\_payments', 'salary\_log', 'bonus', 'bonus\_log', 'total\_stock\_value', 'shared\_receipt\_with\_poi', 'exercised\_stock\_options', 'exercised\_stock\_options\_log', 'total\_payments\_log', 'deferred\_income', 'expenses', 'restricted\_stock', 'long\_term\_incentive', 'loan\_advances']

['poi', 'salary', 'total\_stock\_value\_log', 'total\_payments', 'salary\_log', 'bonus', 'bonus\_log', 'total\_stock\_value', 'shared\_receipt\_with\_poi', 'exercised\_stock\_options', 'exercised\_stock\_options\_log', 'total\_payments\_log', 'deferred\_income', 'expenses', 'restricted\_stock', 'long\_term\_incentive', 'loan\_advances']

Accuracy with LR: 0.790698

recall with LR: 0.400000

precision with LR: 0.250000

the F1 score with LR is: 0.307692

=========================================

Accuracy with NB: 0.860465

recall with NB: 0.400000

precision with NB: 0.400000

the F1 score with NB is: 0.400000

============= look above ============================

Accuracy of scalled features NB: 0.627907

recall of scalled features NB: 1.000000

precision of scalled features NB: 0.238095

the F1 score of scalled features NB: 0.384615

=================BernoulliNB========================

Accuracy of scalled features BernoulliNB: 0.651163

recall of scalled features BernoulliNB: 0.400000

precision of scalled features BernoulliNB: 0.142857

the F1 score of scalled featuresBernoulliNB: 0.210526

=========================================

Accuracy with DTree: 0.813953

recall with DTree: 0.000000

precision with DTree: 0.000000

./poi\_id\_2.py:336: RuntimeWarning: invalid value encountered in double\_scalars

F1 = 2 \* (precision \* recall) / (precision + recall)

the F1 score with DTree is: nan

=========================================

Accuracy of scalled featuresand tunned with DTree: 0.883721

recall of scalled features and tunned with DTree: 0.400000

precision of scalled features and tunned with DTree: 0.500000

the F1 score of scalled features and tunned with DTree is: 0.444444

=========================================

Accuracy with MLPClassifier: 0.511628

recall with MLPClassifier: 0.400000

precision with MLPClassifier: 0.100000

the F1 score with MLPClassifier is: 0.160000

=========================================

/usr/local/lib/python2.7/dist-packages/sklearn/neural\_network/multilayer\_perceptron.py:563: ConvergenceWarning: Stochastic Optimizer: Maximum iterations reached and the optimization hasn't converged yet.

% (), ConvergenceWarning)

Accuracy with MLPClassifier scaled features: 0.883721

recall with MLPClassifier scaled features: 0.400000

precision with MLPClassifier scaled features: 0.500000

the F1 score with MLPClassifier scaled features is 0.444444

=========================================

Accuracy with tunned MLPClassifier: 0.883721

recall with tunned MLPClassifier: 0.000000

/usr/local/lib/python2.7/dist-packages/sklearn/metrics/classification.py:1113: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 due to no predicted samples.

'precision', 'predicted', average, warn\_for)

precision with tunned MLPClassifier: 0.000000

./poi\_id\_2.py:431: RuntimeWarning: invalid value encountered in double\_scalars

F1 = 2 \* (precision \* recall) / (precision + recall)

the F1 score with tunned MLPClassifier is: nan

=========================================

================= LR ========================

LogisticRegression(C=1.0, class\_weight=None, dual=False, fit\_intercept=True,

intercept\_scaling=1, max\_iter=100, multi\_class='ovr', n\_jobs=1,

penalty='l2', random\_state=None, solver='liblinear', tol=0.0001,

verbose=0, warm\_start=False)

Accuracy: 0.78420 Precision: 0.18003 Recall: 0.17400 F1: 0.17696 F2: 0.17517

Total predictions: 15000 True positives: 348 False positives: 1585 False negatives: 1652 True negatives: 11415

================== NB =======================

GridSearchCV(cv=None, error\_score='raise', estimator=GaussianNB(priors=None),

fit\_params={}, iid=True, n\_jobs=1,

param\_grid={'priors': [None, [0.1, 0.9], [0.2, 0.8]]},

pre\_dispatch='2\*n\_jobs', refit=True, return\_train\_score=True,

scoring=None, verbose=0)

Accuracy: 0.73827 Precision: 0.22626 Recall: 0.39800 F1: 0.28851 F2: 0.34555

Total predictions: 15000 True positives: 796 False positives: 2722 False negatives: 1204 True negatives: 10278

================ NB Scaled Features================

GaussianNB(priors=None)

Accuracy: 0.73900 Precision: 0.22604 Recall: 0.39500 F1: 0.28753 F2: 0.34363

Total predictions: 15000 True positives: 790 False positives: 2705 False negatives: 1210 True negatives: 10295

=============== BernoulliNB ================

GridSearchCV(cv=None, error\_score='raise',

estimator=BernoulliNB(alpha=1.0, binarize=0.0, class\_prior=None, fit\_prior=True),

fit\_params={}, iid=True, n\_jobs=1,

param\_grid={'binarize': [0.0, 0.1, 0.5, 0.05, 0.01, 0.001], 'alpha': [0.05, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], 'fit\_prior': [True, False]},

pre\_dispatch='2\*n\_jobs', refit=True, return\_train\_score=True,

scoring=None, verbose=0)

Accuracy: 0.72127 Precision: 0.29358 Recall: 0.77550 F1: 0.42592 F2: 0.58383

Total predictions: 15000 True positives: 1551 False positives: 3732 False negatives: 449 True negatives: 9268

================ DTree ========================

DecisionTreeClassifier(class\_weight=None, criterion='gini', max\_depth=None,

max\_features=None, max\_leaf\_nodes=None,

min\_impurity\_split=1e-07, min\_samples\_leaf=1,

min\_samples\_split=2, min\_weight\_fraction\_leaf=0.0,

presort=False, random\_state=None, splitter='best')

Accuracy: 0.79473 Precision: 0.22489 Recall: 0.22050 F1: 0.22267 F2: 0.22136

Total predictions: 15000 True positives: 441 False positives: 1520 False negatives: 1559 True negatives: 11480

==================== NN ============================

MLPClassifier(activation='relu', alpha=0.0001, batch\_size='auto', beta\_1=0.9,

beta\_2=0.999, early\_stopping=False, epsilon=1e-08,

hidden\_layer\_sizes=(100,), learning\_rate='constant',

learning\_rate\_init=0.001, max\_iter=200, momentum=0.9,

nesterovs\_momentum=True, power\_t=0.5, random\_state=None,

shuffle=True, solver='adam', tol=0.0001, validation\_fraction=0.1,

verbose=False, warm\_start=False)

Accuracy: 0.66027 Precision: 0.12627 Recall: 0.26150 F1: 0.17030 F2: 0.21537

Total predictions: 15000 True positives: 523 False positives: 3619 False negatives: 1477 True negatives: 9381

==================== NN Scalled F =================

MLPClassifier(activation='relu', alpha=0.0001, batch\_size='auto', beta\_1=0.9,

beta\_2=0.999, early\_stopping=False, epsilon=1e-08,

hidden\_layer\_sizes=(100,), learning\_rate='constant',

learning\_rate\_init=0.001, max\_iter=200, momentum=0.9,

nesterovs\_momentum=True, power\_t=0.5, random\_state=None,

shuffle=True, solver='adam', tol=0.0001, validation\_fraction=0.1,

verbose=False, warm\_start=False)

Accuracy: 0.66280 Precision: 0.12451 Recall: 0.25350 F1: 0.16700 F2: 0.20999

Total predictions: 15000 True positives: 507 False positives: 3565 False negatives: 1493 True negatives: 9435

(Udacity) yaser@yaser-Inspiron-7520:~/Udacity/P5/P5\_YasserAlnakhli/scripts$