

# Process\_data\_From\_Blockchain

May 31, 2020

```
In [130]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.impute import SimpleImputer
from sklearn.naive_bayes import GaussianNB, MultinomialNB
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score,
precision_score, roc_curve, auc, matthews_corrcoef
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression

In [97]: file_read = open("new_processed_from_blockchain.txt", "r")
input_data = file_read.readlines()
input_data[0]

Out[97]: '1:"0.28003056,-0.017351949,-0.10524618,-0.99700524,-0.9919519,-0.99088219,-0.9970652'

In [98]: for i in range(len(input_data)):
input_data[i] = (input_data[i].split(":")[1].split('\"')[1]).split(',')

In [99]: input_data = np.array(input_data)

In [100]: # input_data[0].split('\"')

In [191]: total_labels = input_data[:, -1].astype(np.float).astype(np.int)
total_data = input_data[:, :-1].astype(np.float)

In [192]: total_labels

Out[192]: array([6, 6, 1, 1, 2, 3, 6, 2, 1, 2, 5, 6, 3, 4, 1, 4, 1, 5, 4, 3, 4, 5,
2, 4, 5, 6, 2, 3, 3, 5])

In [193]: # total_labels = total_labels.tolist()
for i in range(len(total_labels)):
    if (total_labels[i] <= 3):
        total_labels[i] = 1
# total_labels = np.array(total_labels)
```

```
In [194]: total_labels
```

```
Out[194]: array([6, 6, 1, 1, 1, 1, 6, 1, 1, 1, 5, 6, 1, 4, 1, 4, 1, 5, 4, 1, 4, 5,
                1, 4, 5, 6, 1, 1, 1, 5])
```

```
In [195]: total_data
```

```
Out[195]: array([[ 0.28003056, -0.01735195, -0.10524618, -0.99700524, -0.9919519 ,
                  -0.99088219, -0.99706523, -0.99250103, -0.98972236, -0.94251059,
                  -0.57131514, -0.81966779,  0.85190119,  0.69118304,  0.83878502,
                  -0.99549419, -0.99998193, -0.99995672, -0.99982522, -0.9967439 ],
                 [-0.87239538,  0.1546078 ,  0.33075342, -0.0606394 , -0.30688449,
                  0.06857798, -0.10426969, -0.24459195,  0.14741974, -0.83507417,
                  0.03975068,  0.27674843, -0.70218484,  0.64224635,  0.79984485,
                  0.07234345, -0.29159684, -0.85311126, -0.20509594, -0.14242466],
                 [ 0.16830274, -0.01861949, -0.05469571, -0.24842517,  0.03499402,
                  -0.2684973 , -0.28172069,  0.02219311, -0.26545061, -0.13257279,
                  -0.05660474, -0.30108423,  0.13631085, -0.05969166,  0.38167955,
                  -0.09723705, -0.71388931, -0.79291376, -0.75309301, -0.31708652],
                 [ 0.31634897, -0.05139721, -0.06965648, -0.39990667, -0.31169523,
                  -0.12686989, -0.40480217, -0.31988822, -0.14773967, -0.32592299,
                  -0.2003643 , -0.16528211,  0.28370327,  0.34850224,  0.30215134,
                  -0.23771401, -0.81830467, -0.90541253, -0.6545277 , -0.46969237],
                 [ 0.33088458,  0.01141713, -0.30251751, -0.40604465, -0.09974635,
                  -0.08728796, -0.45214983, -0.11063233, -0.06548856, -0.21867298,
                  -0.12050775, -0.30541594,  0.40625586,  0.30897589,  0.36634209,
                  -0.15533843, -0.82170061, -0.84140606, -0.56989156, -0.57756604],
                 [ 0.17779799, -0.02676299, -0.18215308,  0.32048714,  0.28453352,
                  -0.31291746,  0.30487476,  0.21110028, -0.3220003 ,  0.41936807,
                  -0.08452928, -0.39336619, -0.21417567, -0.49040882,  0.3335263 ,
                  0.26750955, -0.12655083, -0.68204874, -0.77725978,  0.24786087],
                 [ 0.26677822, -0.01632738, -0.10877042, -0.96629565, -0.97322554,
                  -0.97910925, -0.96753538, -0.97081245, -0.97697002, -0.91965534,
                  -0.56400354, -0.81491318,  0.82046004,  0.68357936,  0.83583921,
                  -0.96923108, -0.99926698, -0.99975992, -0.99951377, -0.96816338],
                 [ 0.39052715, -0.01174237, -0.06712539, -0.30829041,  0.0525101 ,
                  -0.36619913, -0.35536285,  0.07263518, -0.34299755, -0.04739859,
                  -0.13799851, -0.41993469,  0.32897512,  0.16041695,  0.50446545,
                  -0.15985424, -0.75688219, -0.78586524, -0.81409151, -0.44119999],
                 [ 0.27157798,  0.03186125, -0.10773083, -0.44425848, -0.13977488,
                  -0.5011188 , -0.49952403, -0.13132148, -0.49633305, -0.37467025,
                  -0.14140635, -0.41304786,  0.32835154,  0.27219615,  0.66748005,
                  -0.32774411, -0.84424406, -0.85206465, -0.88481957, -0.59826388],
                 [ 0.35006469, -0.04310866, -0.26331135, -0.41036292, -0.17539352,
                  -0.16918479, -0.47170618, -0.14990026, -0.17110444, -0.19003834,
                  -0.08891896, -0.21378688,  0.41839241,  0.36581982,  0.43562447,
                  -0.21005446, -0.82377118, -0.86668993, -0.65318373, -0.62637306],
                 [ 0.27814006, -0.01625879, -0.10940422, -0.99812277, -0.97146729,
```

-0.99005629, -0.99828223, -0.97215303, -0.98970551, -0.94243196,  
 -0.55756832, -0.8189303, 0.8513212, 0.68145969, 0.8458757,  
 -0.99087986, -0.99999063, -0.99973451, -0.99981839, -0.99806278],  
 [ 0.23690107, -0.0176074, -0.10244162, -0.98223308, -0.98003856,  
 -0.98137767, -0.98317476, -0.97905033, -0.98167108, -0.9389744,  
 -0.56965256, -0.8105017, 0.82211804, 0.68766325, 0.83859504,  
 -0.97528702, -0.99943376, -0.99984729, -0.99953531, -0.98346987],  
 [ 0.2269563, -0.01114583, -0.08540512, -0.19657033, -0.14708527,  
 -0.34940388, -0.29236418, -0.1726459, -0.3723118, 0.26206694,  
 -0.03280512, -0.15607812, 0.29188068, 0.14871582, 0.48856775,  
 -0.18953474, -0.67551838, -0.85872777, -0.80622075, -0.54941886],  
 [ 0.27788779, -0.01818021, -0.10692821, -0.9962534, -0.99398783,  
 -0.98712891, -0.99624597, -0.99420912, -0.98535622, -0.93848665,  
 -0.5738312, -0.82180683, 0.85123639, 0.69144565, 0.84460746,  
 -0.99436486, -0.99997735, -0.9999681, -0.99974569, -0.99577729],  
 [ 0.21843518, -0.00653951, -0.10683856, -0.32894525, -0.15976382,  
 -0.25149535, -0.35450148, -0.16083562, -0.2753807, -0.17739764,  
 -0.03046609, -0.20202377, 0.29398469, 0.15396412, 0.33169834,  
 -0.20181959, -0.77281104, -0.86270892, -0.7461697, -0.40186394],  
 [ 0.3080432, -0.02732928, -0.14022197, -0.97452606, -0.96197777,  
 -0.92594823, -0.97317968, -0.95731377, -0.91642835, -0.91312502,  
 -0.57142025, -0.80388257, 0.84445397, 0.67589896, 0.811405,  
 -0.94462838, -0.99938108, -0.99939254, -0.9950483, -0.97105791],  
 [ 0.24847489, -0.00513955, -0.0995467, -0.41659387, -0.26658815,  
 -0.17906707, -0.47807354, -0.26999367, -0.1808865, -0.1048889,  
 -0.13599542, -0.31855517, 0.36164604, 0.33787997, 0.28052445,  
 -0.27622, -0.82831449, -0.89493514, -0.6957178, -0.59388094],  
 [ 0.28342531, -0.01620339, -0.09712822, -0.99704778, -0.97285921,  
 -0.98785117, -0.99747136, -0.97404253, -0.9880223, -0.93860892,  
 -0.56637718, -0.8032079, 0.85021987, 0.67443567, 0.84721261,  
 -0.98804712, -0.99997657, -0.99975437, -0.99958259, -0.997753 ],  
 [ 0.28023702, -0.01895768, -0.10726972, -0.99707669, -0.99384898,  
 -0.98966128, -0.99747272, -0.99361463, -0.98933414, -0.9403408,  
 -0.57731838, -0.81791684, 0.85129886, 0.68836548, 0.84470474,  
 -0.99578777, -0.99998224, -0.99996324, -0.99980934, -0.99745782],  
 [ 0.22571228, -0.01572238, -0.10553341, 0.0950629, 0.34114544,  
 -0.20826985, 0.03981343, 0.32453292, -0.33420084, 0.44858912,  
 0.02126267, -0.34561956, 0.16057504, -0.09958053, 0.12298263,  
 0.14273764, -0.39934067, -0.65381145, -0.71666458, -0.14159751],  
 [ 0.2783322, -0.01700362, -0.1116865, -0.99823176, -0.99171166,  
 -0.98992232, -0.99804568, -0.99267792, -0.99041508, -0.94383966,  
 -0.56737382, -0.81085575, 0.85256286, 0.69278233, 0.84016142,  
 -0.99632581, -0.99999125, -0.9999551, -0.99980021, -0.99738684],  
 [ 0.2769146, -0.01327666, -0.10974388, -0.99836013, -0.98293077,  
 -0.99323506, -0.99852534, -0.98100811, -0.99212414, -0.94458281,  
 -0.5646813, -0.82564197, 0.85237876, 0.69282801, 0.84359968,  
 -0.9938327, -0.99999205, -0.99985226, -0.99988653, -0.99839916],  
 [ 0.2250419, -0.0912645, -0.36960062, -0.12719559, -0.32877187,

```

0.02111189, -0.17809992, -0.33300086, 0.02991902, 0.15733914,
-0.18829286, -0.41491428, 0.18458693, 0.31276594, 0.12596649,
-0.02762942, -0.61744203, -0.90209583, -0.4323667, -0.28539135],
[ 0.26689552, -0.00942394, -0.08519007, -0.97325473, -0.87953542,
-0.90937518, -0.97698255, -0.87821831, -0.90875622, -0.92057503,
-0.49785502, -0.74440613, 0.80085229, 0.66374319, 0.81466137,
-0.93040058, -0.99950652, -0.9966779, -0.99433697, -0.98443193],
[ 0.27701495, -0.01443921, -0.10368149, -0.99783369, -0.99048982,
-0.98884984, -0.99797355, -0.98916684, -0.98786494, -0.94264451,
-0.57066592, -0.81599108, 0.85165766, 0.69474976, 0.84735711,
-0.99360041, -0.9999888, -0.99993318, -0.99975997, -0.99747743],
[ 0.38002077, -0.04051057, -0.13901039, -0.85876042, -0.88803272,
-0.88369419, -0.88267837, -0.89126254, -0.89223201, -0.83507417,
-0.50568749, -0.68429669, 0.62575124, 0.64224635, 0.79984485,
-0.84814356, -0.98737114, -0.99620818, -0.99105624, -0.91259281],
[ 0.18155133, 0.0159864, -0.16614154, -0.21471492, -0.00897888,
0.12841003, -0.28110618, -0.06410186, 0.1434152, -0.12144027,
0.20957708, -0.12595501, 0.049721, 0.26855908, 0.02675893,
-0.01214942, -0.68856867, -0.80800152, -0.42653781, -0.42188751],
[ 0.43906334, -0.02897953, -0.14925098, 0.20752477, -0.23589028,
-0.25317118, 0.15228917, -0.2692441, -0.25908643, 0.41766641,
0.0843052, -0.1520279, -0.14313645, 0.24236629, 0.1301467,
0.04398838, -0.26557718, -0.88610781, -0.7448121, -0.02486511],
[ 0.31205089, -0.00119141, -0.11730152, 0.03385787, -0.19032387,
-0.26311693, -0.08131291, -0.21765187, -0.28692833, 0.58430698,
0.03018861, -0.39367574, 0.21427424, 0.25330471, 0.24620623,
-0.0705795, -0.46460106, -0.87213056, -0.75370778, -0.32169548],
[ 0.27803836, -0.01901653, -0.10659311, -0.99796138, -0.98432594,
-0.98378654, -0.9982562, -0.98409429, -0.98370204, -0.9385799,
-0.56556963, -0.80831823, 0.85263625, 0.6875945, 0.8456663,
-0.99221488, -0.99998964, -0.9998868, -0.99965324, -0.99810472]]])

```

```

In [243]: from sklearn.model_selection import train_test_split
          X_train, X_test, y_train, y_test = train_test_split(total_data, total_labels, test_s

```

```

In [244]: for i in range(len(y_train)):
          if(y_train[i] == 1):
              y_train[i] = 4
          elif(y_train[i] == 4):
              y_train[i] = 1
          # elif(y_train[i] == 5):
          #     y_train[i] = 4

```

```

In [245]: nbclf = GaussianNB()
          rfclf = RandomForestClassifier(n_estimators=100)
          svmclf = svm.SVC(kernel='linear') #Linear Kernel

```

```

In [246]: def classify_and_report(classifier,X_train, X_test, y_train, y_test):
          classifier.fit(X_train,y_train)

```

```

y_predicted = classifier.predict(X_test)
#print(confusion_matrix(Y_test, Y_predicted))
# print("Naive Bayes Classifier: \n")
print("Multiclass classification: ")
print('Accuracy:', accuracy_score(y_test, y_predicted))
#print('F1 score:', f1_score(Y_test, Y_predicted,average='macro'))
#print('F1 score:', f1_score(Y_test, Y_predicted,average='weighted'))
print('F1 score:', f1_score(y_test, y_predicted,average='macro'))
print('Recall:', recall_score(y_test, y_predicted,average='macro'))
print('Precision:', precision_score(y_test, y_predicted,average='macro'))
print("Matthews Correlation Coefficient: ",matthews_corrcoef(y_test, y_predicted))
print('Classification report:', classification_report(y_test, y_predicted))

```

```

In [247]: print ("Naive Bayes")
          classify_and_report(nbcclf,X_train, X_test, y_train, y_test)

```

Naive Bayes

Multiclass classification:

Accuracy: 0.1

F1 score: 0.16666666666666666

Recall: 0.125

Precision: 0.25

Matthews Correlation Coefficient: -0.2858966759567453

Classification report:                    precision      recall    f1-score      support

1	0.00	0.00	0.00	6
4	0.00	0.00	0.00	1
5	0.00	0.00	0.00	1
6	1.00	0.50	0.67	2

accuracy			0.10	10
macro avg	0.25	0.12	0.17	10
weighted avg	0.20	0.10	0.13	10

```

/home/ashish/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1437: Und
'precision', 'predicted', average, warn_for)
/home/ashish/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1437: Und
'precision', 'predicted', average, warn_for)
/home/ashish/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1437: Und
'precision', 'predicted', average, warn_for)

```

```

In [248]: print ("Random Forest")
          classify_and_report(rfcclf,X_train, X_test, y_train, y_test)

```

Random Forest

Multiclass classification:

```

Accuracy: 0.2
F1 score: 0.375
Recall: 0.375
Precision: 0.375
Matthews Correlation Coefficient: 0.05172413793103448
Classification report:

```

		precision	recall	f1-score	support
	1	0.00	0.00	0.00	6
	4	0.00	0.00	0.00	1
	5	1.00	1.00	1.00	1
	6	0.50	0.50	0.50	2
	accuracy			0.20	10
	macro avg	0.38	0.38	0.38	10
	weighted avg	0.20	0.20	0.20	10

```

In [249]: print ("SVM")
          classify_and_report(svmclf,X_train, X_test, y_train, y_test)

```

```

SVM
Multiclass classification:
Accuracy: 0.1
F1 score: 0.125
Recall: 0.25
Precision: 0.08333333333333333
Matthews Correlation Coefficient: -0.08934271123648291
Classification report:

```

		precision	recall	f1-score	support
	1	0.00	0.00	0.00	6
	4	0.00	0.00	0.00	1
	5	0.33	1.00	0.50	1
	6	0.00	0.00	0.00	2
	accuracy			0.10	10
	macro avg	0.08	0.25	0.12	10
	weighted avg	0.03	0.10	0.05	10

```

/home/ashish/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1437: Und
'precision', 'predicted', average, warn_for)
/home/ashish/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1437: Und
'precision', 'predicted', average, warn_for)
/home/ashish/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1437: Und
'precision', 'predicted', average, warn_for)

```