Anay Gupta
CSE 494: AI for Cyber Security
Shakarian - Friday 1 pm
March 15th, 2019

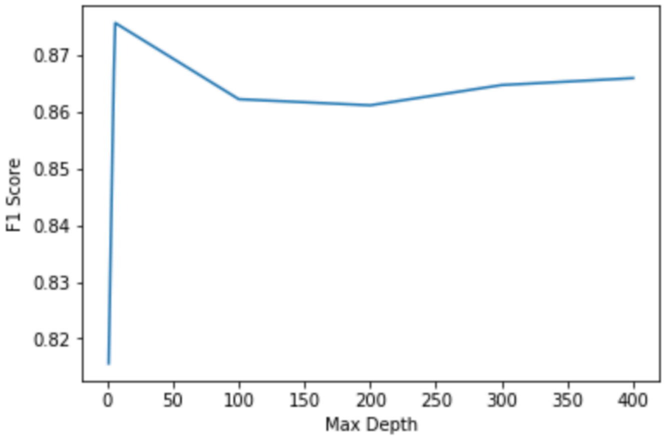
CSE 494 Homework 2

1. Decision Tree Classifier

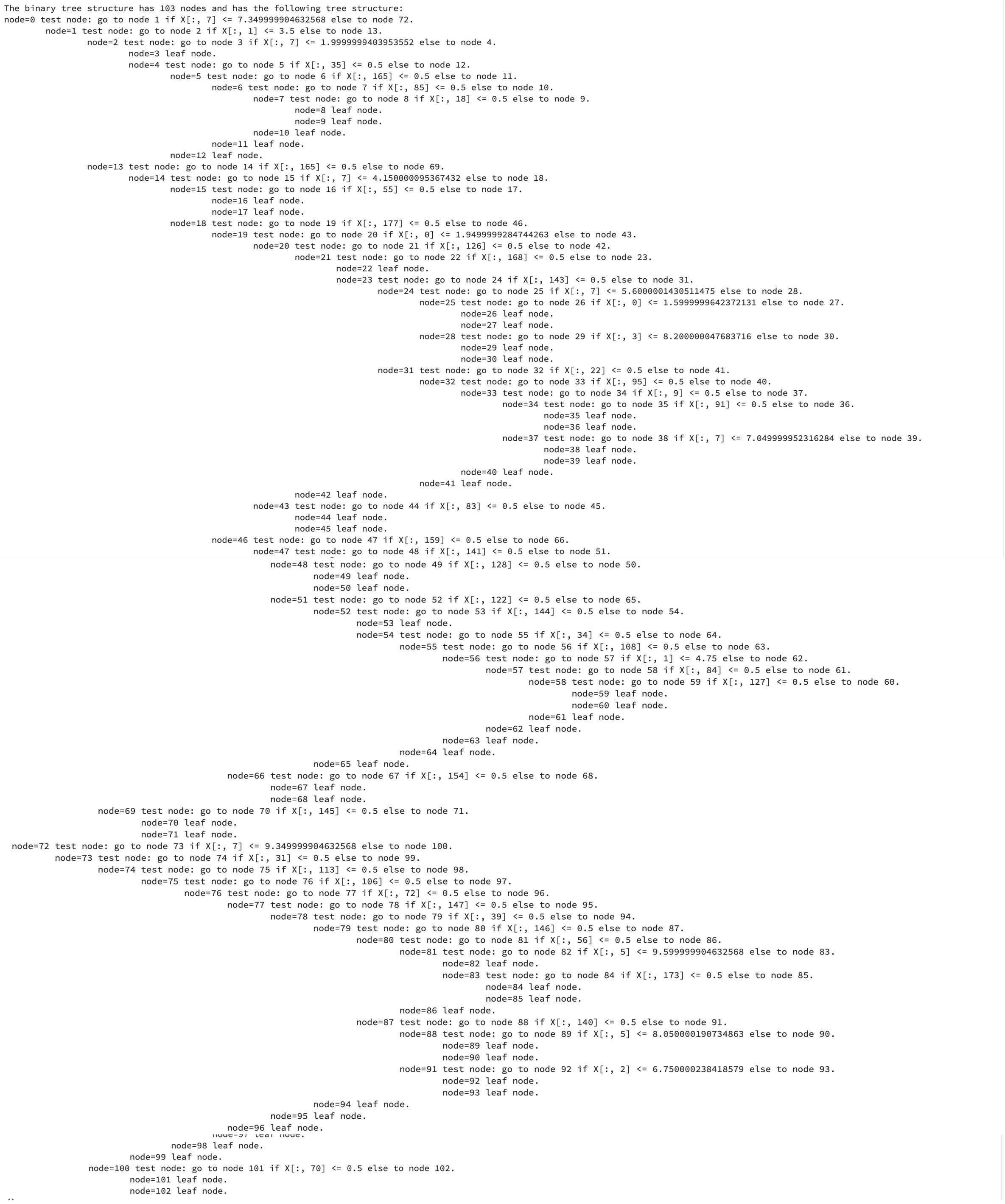
a. Classification Metrics using the best attributes for predicting whether a CVE would be exploited (10 fold Cross-Validation):

Precision: 0.8503457675753229
Recall: 0.8897444717444717
F1: 0.8643339276968938

b. The results are in the following format: (Precision, Recall, F1)
Max_depth = 1 -> **(0.8078005073861508, 0.8377330057330058, 0.8155863139185904)**
Max_depth = 5 -> **(0.8665461090881259, 0.8803390663390663, 0.8672669362686862)**
Max_depth = 6 -> **(0.8703820354052553, 0.8923390663390665, 0.875680102863954)**
Max_depth = 100 -> **(0.8458079524492724, 0.8897444717444717, 0.8622212516405557)**
Max_depth = 200 -> **(0.8484408397505344, 0.8857444717444718, 0.8613930628561046)**
Max_depth = 300 -> **(0.8368962219033955, 0.8937444717444718, 0.8590948489723298)**
Max_depth = 400 -> **(0.8551950611389021, 0.8857444717444718, 0.8654471169101587)**



c. Tree Properties/Rules (Implemented from Lab 3?)



```
The binary tree structure has 103 nodes and has the following tree structure:
node=0 test node: go to node 1 if X[:, 7] <= 7.349999904632568 else to node 72.
        node=1 test node: go to node 2 if X[:, 1] <= 3.5 else to node 13.
                node=2 test node: go to node 3 if X[:, 7] <= 1.9999999403953552 else to node 4.
                        node=3 leaf node.
                        node=4 test node: go to node 5 if X[:, 35] <= 0.5 else to node 12.
                                node=5 test node: go to node 6 if X[:, 165] <= 0.5 else to node 11.
                                        node=6 test node: go to node 7 if X[:, 85] <= 0.5 else to node 10.
                                                node=7 test node: go to node 8 if X[:, 18] <= 0.5 else to node 9.
                                                        node=8 leaf node.
                                                        node=9 leaf node.
                                                node=10 leaf node.
                                        node=11 leaf node.
                                node=12 leaf node.
                node=13 test node: go to node 14 if X[:, 165] <= 0.5 else to node 69.
                        node=14 test node: go to node 15 if X[:, 7] <= 4.150000095367432 else to node 18.
                                node=15 test node: go to node 16 if X[:, 55] <= 0.5 else to node 17.
                                        node=16 leaf node.
                                        node=17 leaf node.
                                node=18 test node: go to node 19 if X[:, 177] <= 0.5 else to node 46.
                                        node=19 test node: go to node 20 if X[:, 0] <= 1.9499999284744263 else to node 43.
                                                node=20 test node: go to node 21 if X[:, 126] <= 0.5 else to node 42.
                                                        node=21 test node: go to node 22 if X[:, 168] <= 0.5 else to node 23.
                                                                node=22 leaf node.
                                                                node=23 test node: go to node 24 if X[:, 143] <= 0.5 else to node 31.
                                                                        node=24 test node: go to node 25 if X[:, 7] <= 5.6000001430511475 else to node 28.
                                                                                node=25 test node: go to node 26 if X[:, 0] <= 1.5999999642372131 else to node 27.
                                                                                        node=26 leaf node.
                                                                                        node=27 leaf node.
                                                                                node=28 test node: go to node 29 if X[:, 3] <= 8.200000047683716 else to node 30.
                                                                                        node=29 leaf node.
                                                                                        node=30 leaf node.
                                                                        node=31 test node: go to node 32 if X[:, 22] <= 0.5 else to node 41.
                                                                                node=32 test node: go to node 33 if X[:, 95] <= 0.5 else to node 40.
                                                                                        node=33 test node: go to node 34 if X[:, 9] <= 0.5 else to node 37.
                                                                                                node=34 test node: go to node 35 if X[:, 91] <= 0.5 else to node 36.
                                                                                                        node=35 leaf node.
                                                                                                        node=36 leaf node.
                                                                                                node=37 test node: go to node 38 if X[:, 7] <= 7.049999952316284 else to node 39.
                                                                                                        node=38 leaf node.
                                                                                                        node=39 leaf node.
                                                                                        node=40 leaf node.
                                                                                node=41 leaf node.
                                                        node=42 leaf node.
                                                node=43 test node: go to node 44 if X[:, 83] <= 0.5 else to node 45.
                                                        node=44 leaf node.
                                                        node=45 leaf node.
                                        node=46 test node: go to node 47 if X[:, 159] <= 0.5 else to node 66.
                                                node=47 test node: go to node 48 if X[:, 141] <= 0.5 else to node 51.
                                                        node=48 test node: go to node 49 if X[:, 128] <= 0.5 else to node 50.
                                                                node=49 leaf node.
                                                                node=50 leaf node.
                                                        node=51 test node: go to node 52 if X[:, 122] <= 0.5 else to node 65.
                                                                node=52 test node: go to node 53 if X[:, 144] <= 0.5 else to node 54.
                                                                        node=53 leaf node.
                                                                        node=54 test node: go to node 55 if X[:, 34] <= 0.5 else to node 64.
                                                                                node=55 test node: go to node 56 if X[:, 108] <= 0.5 else to node 63.
                                                                                        node=56 test node: go to node 57 if X[:, 1] <= 4.75 else to node 62.
                                                                                                node=57 test node: go to node 58 if X[:, 84] <= 0.5 else to node 61.
                                                                                                        node=58 test node: go to node 59 if X[:, 127] <= 0.5 else to node 60.
                                                                                                                node=59 leaf node.
                                                                                                                node=60 leaf node.
                                                                                                        node=61 leaf node.
                                                                                                node=62 leaf node.
                                                                                        node=63 leaf node.
                                                                                node=64 leaf node.
                                                                node=65 leaf node.
                                                node=66 test node: go to node 67 if X[:, 154] <= 0.5 else to node 68.
                                                        node=67 leaf node.
                                                        node=68 leaf node.
                        node=69 test node: go to node 70 if X[:, 145] <= 0.5 else to node 71.
                                node=70 leaf node.
                                node=71 leaf node.
        node=72 test node: go to node 73 if X[:, 7] <= 9.349999904632568 else to node 100.
                node=73 test node: go to node 74 if X[:, 31] <= 0.5 else to node 99.
                        node=74 test node: go to node 75 if X[:, 113] <= 0.5 else to node 98.
                                node=75 test node: go to node 76 if X[:, 106] <= 0.5 else to node 97.
                                        node=76 test node: go to node 77 if X[:, 72] <= 0.5 else to node 96.
                                                node=77 test node: go to node 78 if X[:, 147] <= 0.5 else to node 95.
                                                        node=78 test node: go to node 79 if X[:, 39] <= 0.5 else to node 94.
                                                                node=79 test node: go to node 80 if X[:, 146] <= 0.5 else to node 87.
                                                                        node=80 test node: go to node 81 if X[:, 56] <= 0.5 else to node 86.
                                                                                node=81 test node: go to node 82 if X[:, 5] <= 9.599999904632568 else to node 83.
                                                                                        node=82 leaf node.
                                                                                        node=83 test node: go to node 84 if X[:, 173] <= 0.5 else to node 85.
                                                                                                node=84 leaf node.
                                                                                                node=85 leaf node.
                                                                                node=86 leaf node.
                                                                        node=87 test node: go to node 88 if X[:, 140] <= 0.5 else to node 91.
                                                                                node=88 test node: go to node 89 if X[:, 5] <= 8.050000190734863 else to node 90.
                                                                                        node=89 leaf node.
                                                                                        node=90 leaf node.
                                                                                node=91 test node: go to node 92 if X[:, 2] <= 6.750000238418579 else to node 93.
                                                                                        node=92 leaf node.
                                                                                        node=93 leaf node.
                                                                node=94 leaf node.
                                                        node=95 leaf node.
                                                node=96 leaf node.
                                        node=97 leaf node.
                                node=98 leaf node.
                        node=99 leaf node.
                node=100 test node: go to node 101 if X[:, 70] <= 0.5 else to node 102.
                        node=101 leaf node.
                        node=102 leaf node.
```

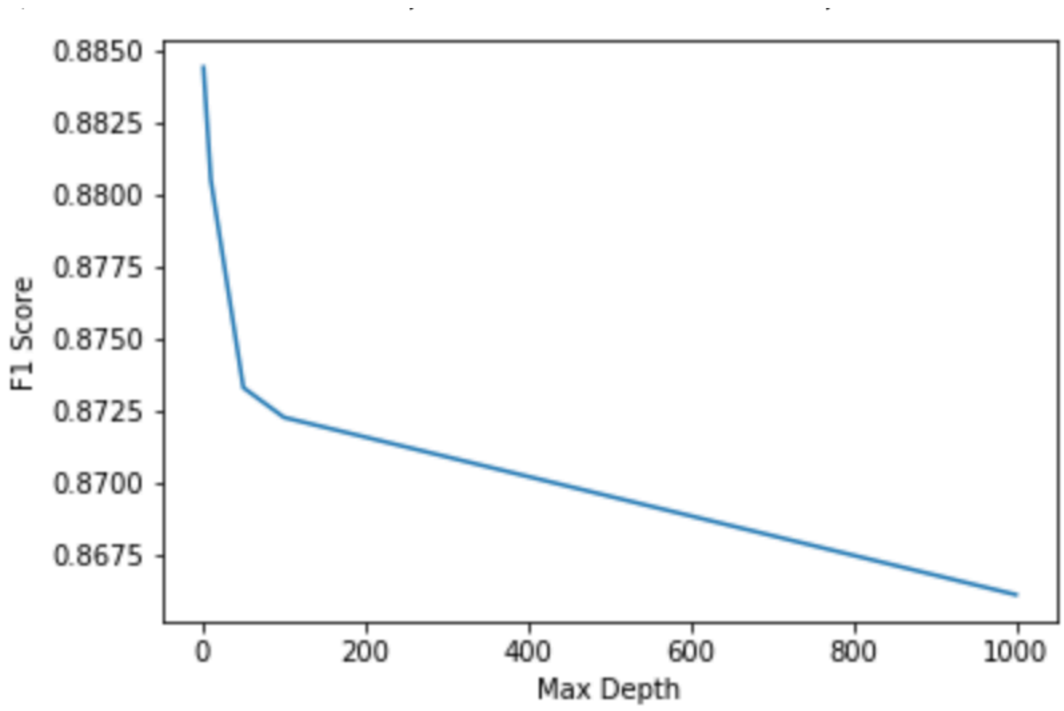
2. Logistic Regression Classifier

a. Classification metrics using default hyper parameters and 10-fold Cross Validation:

Precision: 0.8744640304282247
Recall: 0.8989336609336609
F1: 0.8844305345903992

b. The plotted points are in the following format: (C, F1 Score)

(1000, 0.8660935516008672)
(100, 0.8722585871804949)
(50, 0.8732916082352702)
(10, 0.8805247620840486)
(1, 0.8844305345903992)



c. Upon comparing the overall results of the Decision Tree classifier and the Logistic Regression classifier, it is apparent that the Logistic Regression classifier outperforms the Decision Tree classifier as it plateaus off at a higher F1 score. The Logistic Regression classifier's F1 score is 88.4% while the Decision Tree classifier's F1 Score is 86.4% with both using 10-fold cross-validation and default hyper-parameters. Similarly, the Logistic Regression classifier's precision and recall scores are 87.4% and 89.9% while the Decision Tree classifier's precision and recall scores are 85% and 88.9% respectively. The categorical features such as attackVector, attackComplexity, privelegesRequired, etc. work better with Logistic Regression rather than Decision Trees because they do not have to be converted to multiple binary features. Instead, Logistic Regression classifier can use these types of categorical features as is.

Bonus Question:

Random Forest Classifier Results with 10-fold cross validation, 100 n_estimators, max_depth = None, and criterion = ‘entropy’:

Precision: 0.6829268292682927

Recall: 0.7777777777777778

F1-Score: 0.7272727272727273

Comparing to Decision Tree Classifier:

The results are in the following format: (Precision, Recall, F1)

Max_depth = 1 -> **(0.9230769230769231, 0.3333333333333333, 0.489795918367347)**

Max_depth = 5 -> **(0.7058823529411765, 0.6666666666666666, 0.6857142857142857)**

Max_depth = 6 -> **(0.7142857142857143, 0.6944444444444444, 0.7042253521126761)**

Max_depth = 100 -> **(0.7105263157894737, 0.75, 0.7297297297297298)**

Max_depth = 200 -> **(0.6923076923076923, 0.75, 0.7199999999999999)**

Max_depth = 300 -> **(0.7, 0.7777777777777778, 0.7368421052631577)**

Max_depth = 400 -> **(0.6829268292682927, 0.7777777777777778, 0.7272727272727273)**