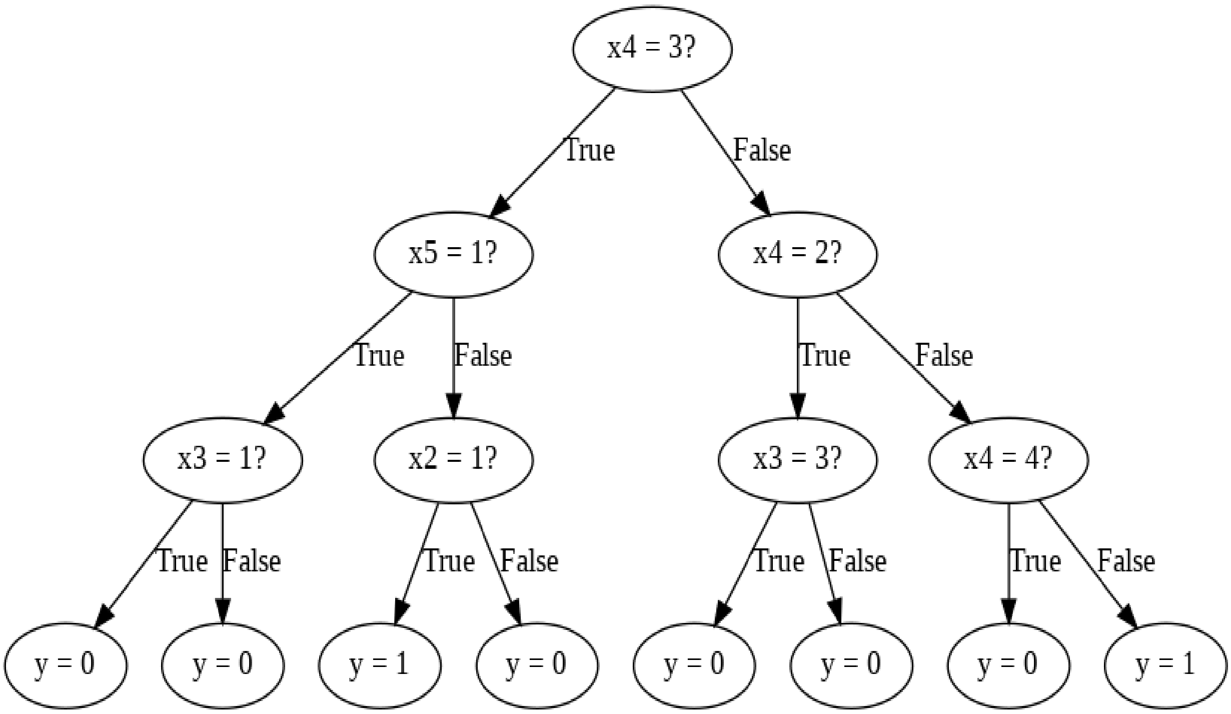
PROJECT REPORT

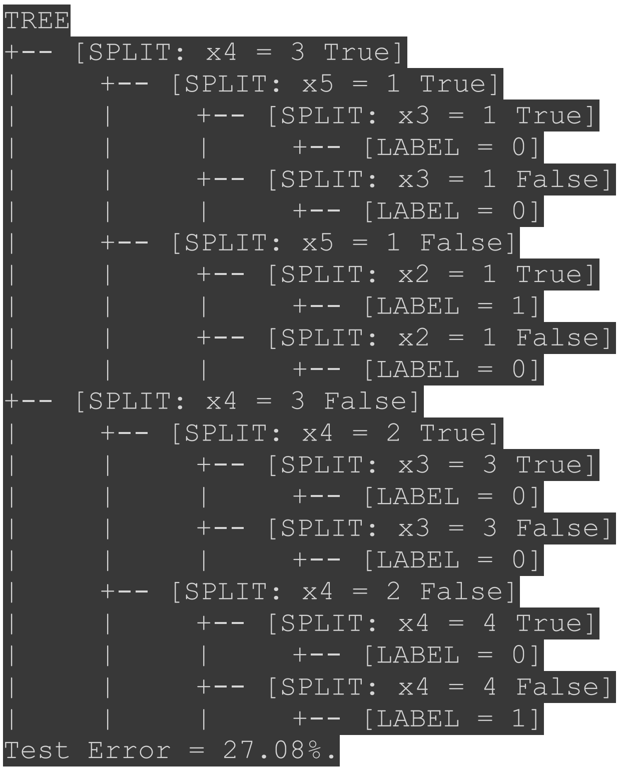
Fixed Depth Decision Tree

Bhattacharya, Abhirup

Part A:

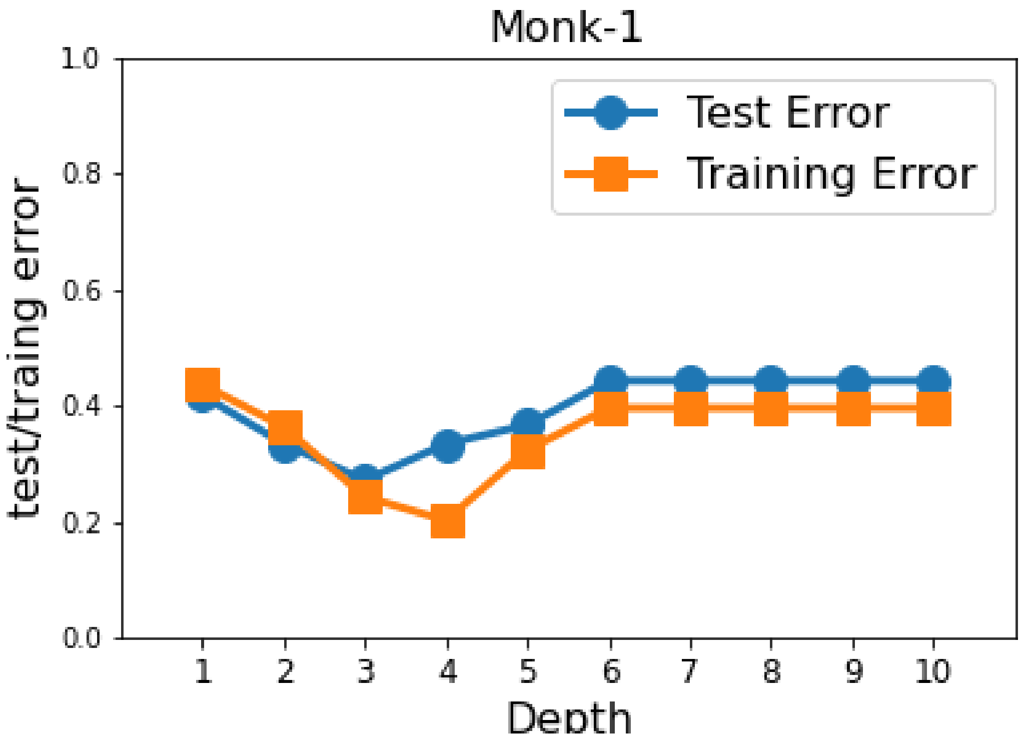


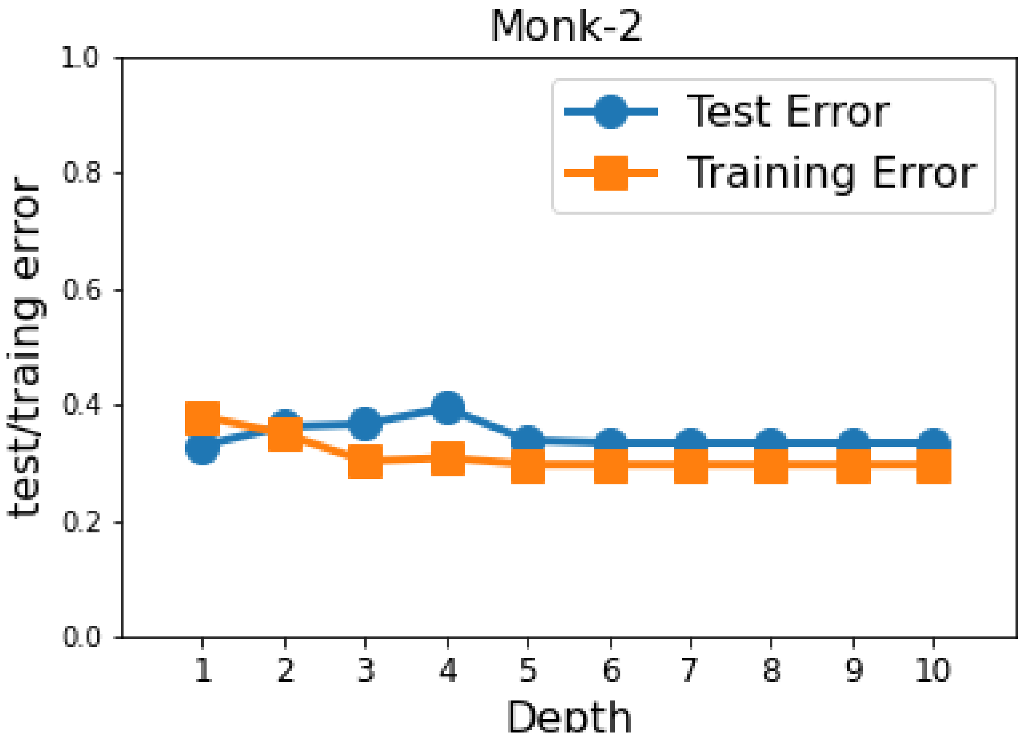
Code Output:

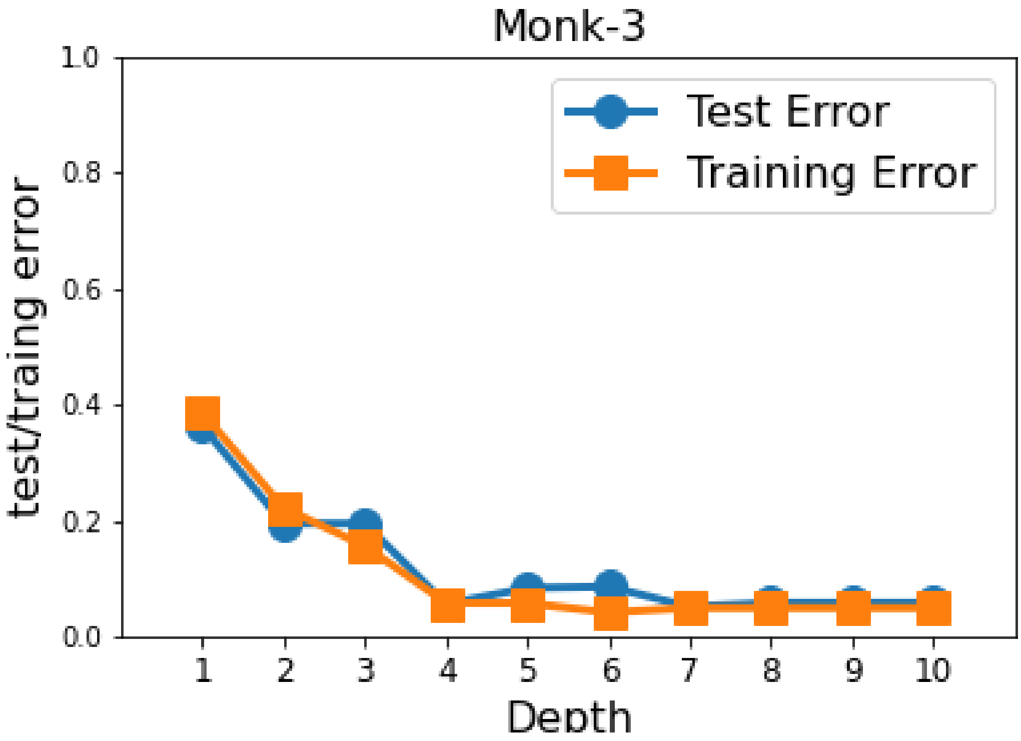


```
TREE
+-- [SPLIT: x4 = 3 True]
|      +-- [SPLIT: x5 = 1 True]
|      |      +-- [SPLIT: x3 = 1 True]
|      |      |      +-- [LABEL = 0]
|      |      +-- [SPLIT: x3 = 1 False]
|      |      |      +-- [LABEL = 0]
|      +-- [SPLIT: x5 = 1 False]
|      |      +-- [SPLIT: x2 = 1 True]
|      |      |      +-- [LABEL = 1]
|      |      +-- [SPLIT: x2 = 1 False]
|      |      |      +-- [LABEL = 0]
+-- [SPLIT: x4 = 3 False]
|      +-- [SPLIT: x4 = 2 True]
|      |      +-- [SPLIT: x3 = 3 True]
|      |      |      +-- [LABEL = 0]
|      |      +-- [SPLIT: x3 = 3 False]
|      |      |      +-- [LABEL = 0]
|      +-- [SPLIT: x4 = 2 False]
|      |      +-- [SPLIT: x4 = 4 True]
|      |      |      +-- [LABEL = 0]
|      |      +-- [SPLIT: x4 = 4 False]
|      |      |      +-- [LABEL = 1]
Test Error = 27.08%.
```

Part B:


Monk-1
Test Error
Training Error
test/traing error
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
10
Depth


Monk-2
Test Error
Training Error
test/traing error
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
10
Depth



Part C:

Depth - 1

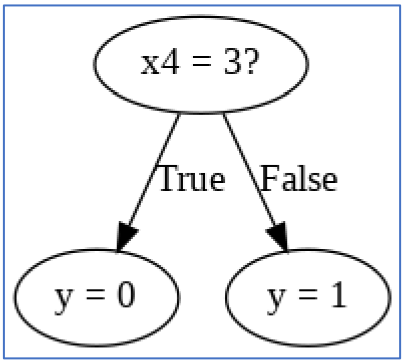
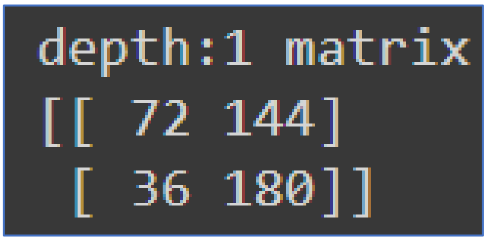


Figure 1 : Decision Tree for Depth 1



```
depth:1 matrix
[[ 72 144]
 [ 36 180]]
```

Figure 2 : Confusion Matrix for Depth 1

Depth - 3

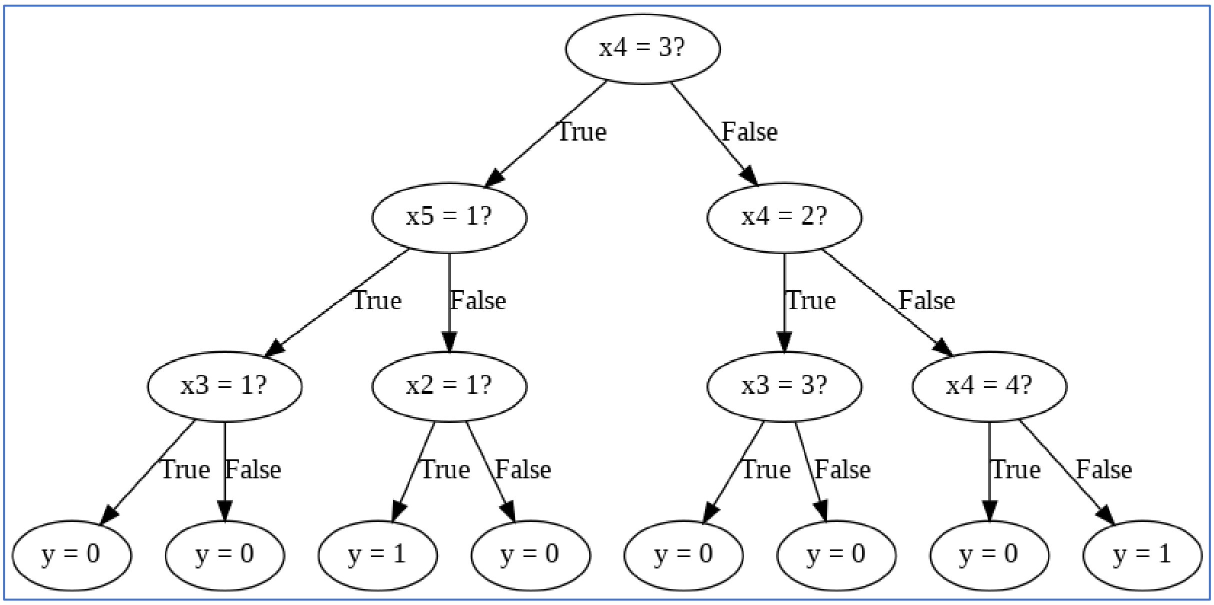
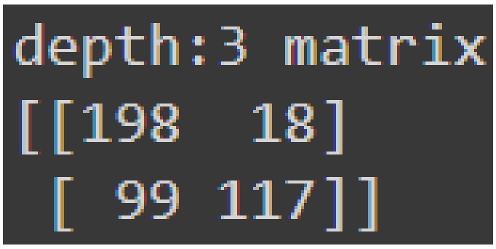


Figure 3 : Decision Tree for Depth - 3



```
depth:3 matrix
[[198  18]
 [ 99 117]]
```

Figure 4 : Confusion Matrix for Depth 3

Depth -5

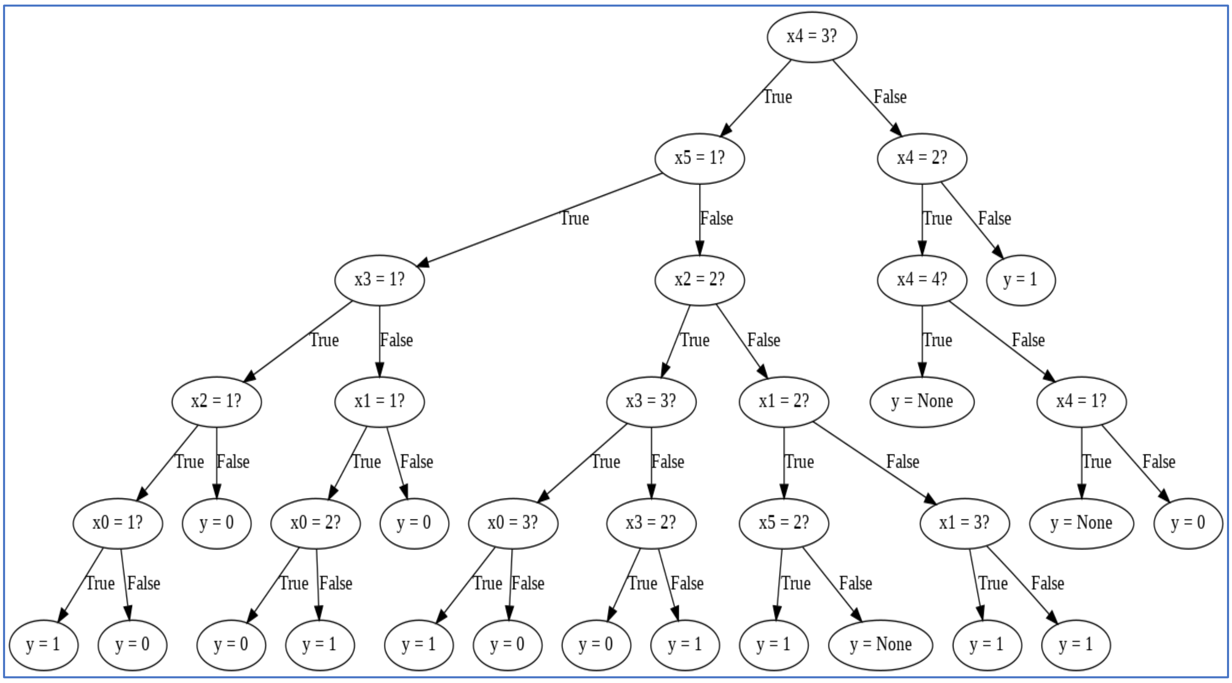
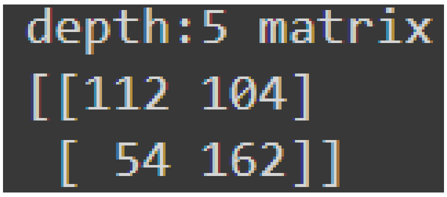


Figure 5 : Decision Tree for Depth – 5



```
depth:5 matrix
[[112 104]
 [ 54 162]]
```

Figure 6 : Confusion Matrix for Depth 5

Part D:

In this part, I have used the decision tree implementation from scikit learn package to learn the decision tree using criteria Entropy. Following are the decision trees and confusion matrix for depth 1,3 and 5

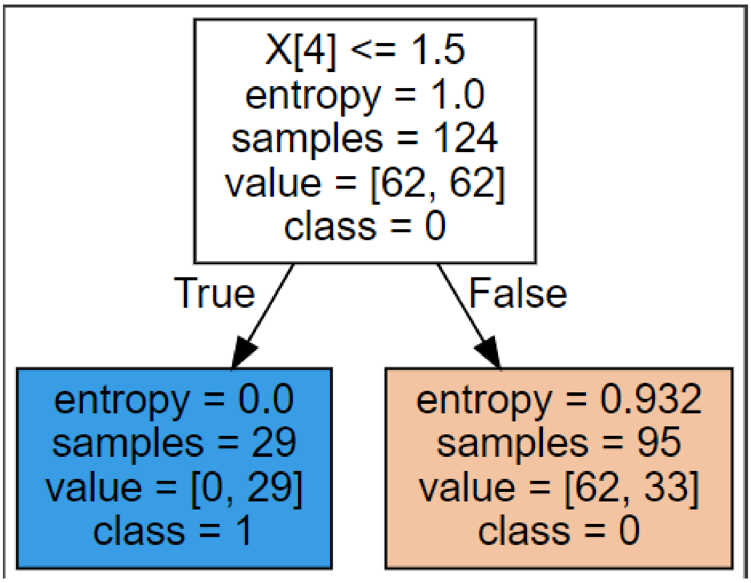
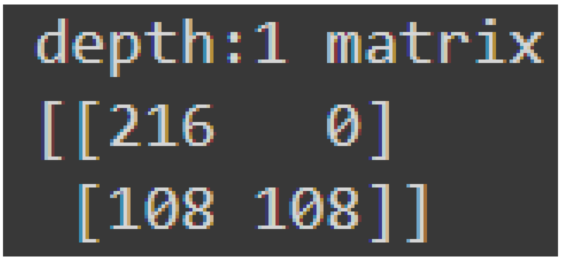


Figure 7 : Scikit Learn Depth 1 Decision Tree



```
depth:1 matrix
[[216   0]
 [108 108]]
```

Figure 8 : Scikit Learn Depth 1 Confusion Matrix

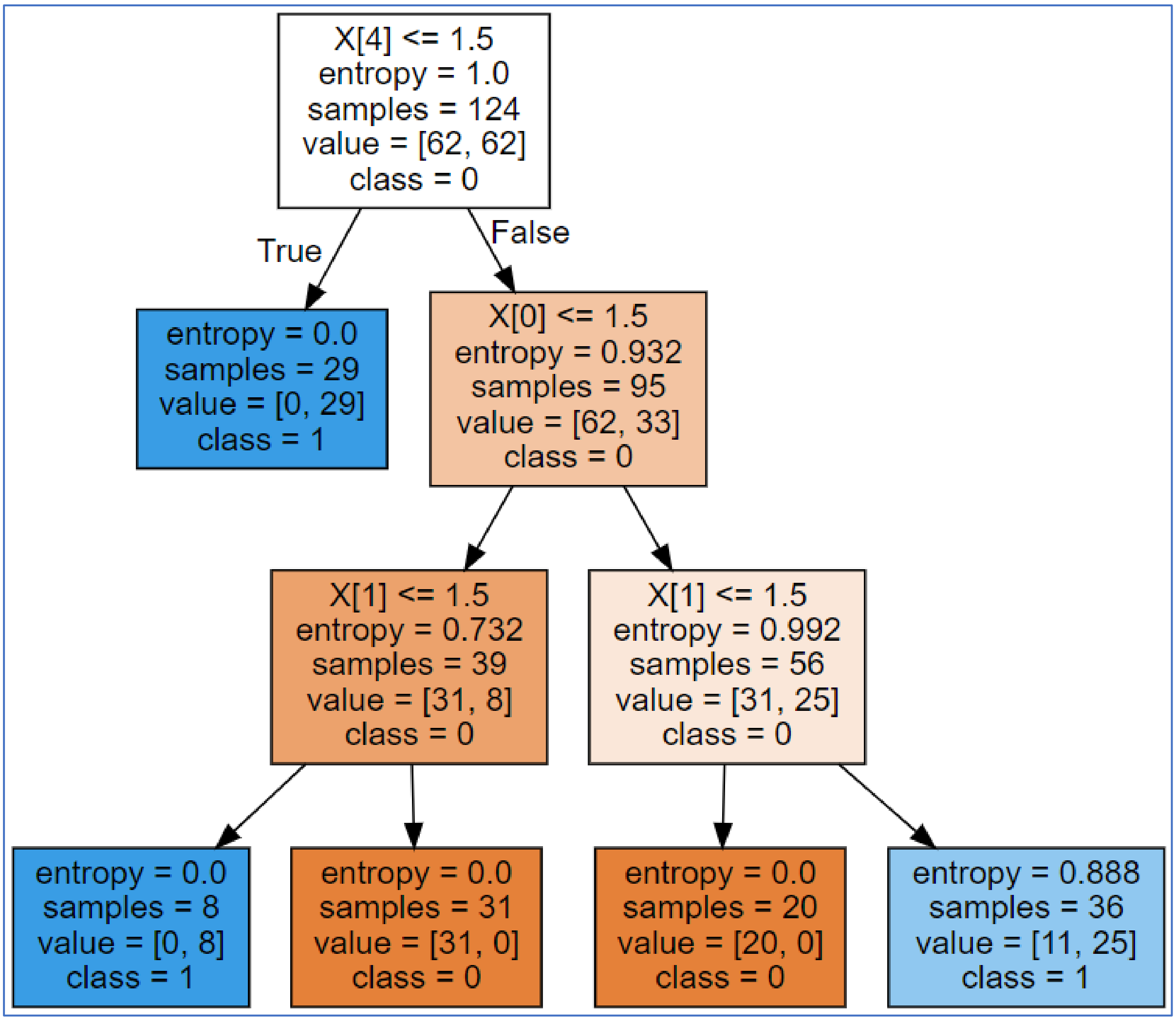
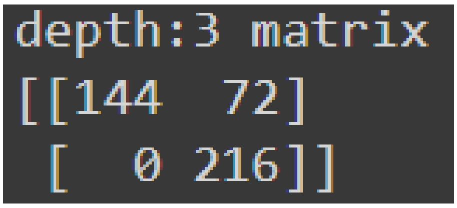
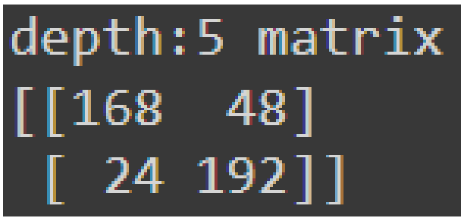


Figure 9 : Scikit Learn Depth 3 Decision Tree



```
depth:3 matrix
[[144  72]
 [  0 216]]
```

Figure 10 : Scikit Learn Depth 3 Confusion Matrix



```
depth:5 matrix
[[168  48]
 [ 24 192]]
```

Figure 11 : Scikit Learn Depth 5 DT

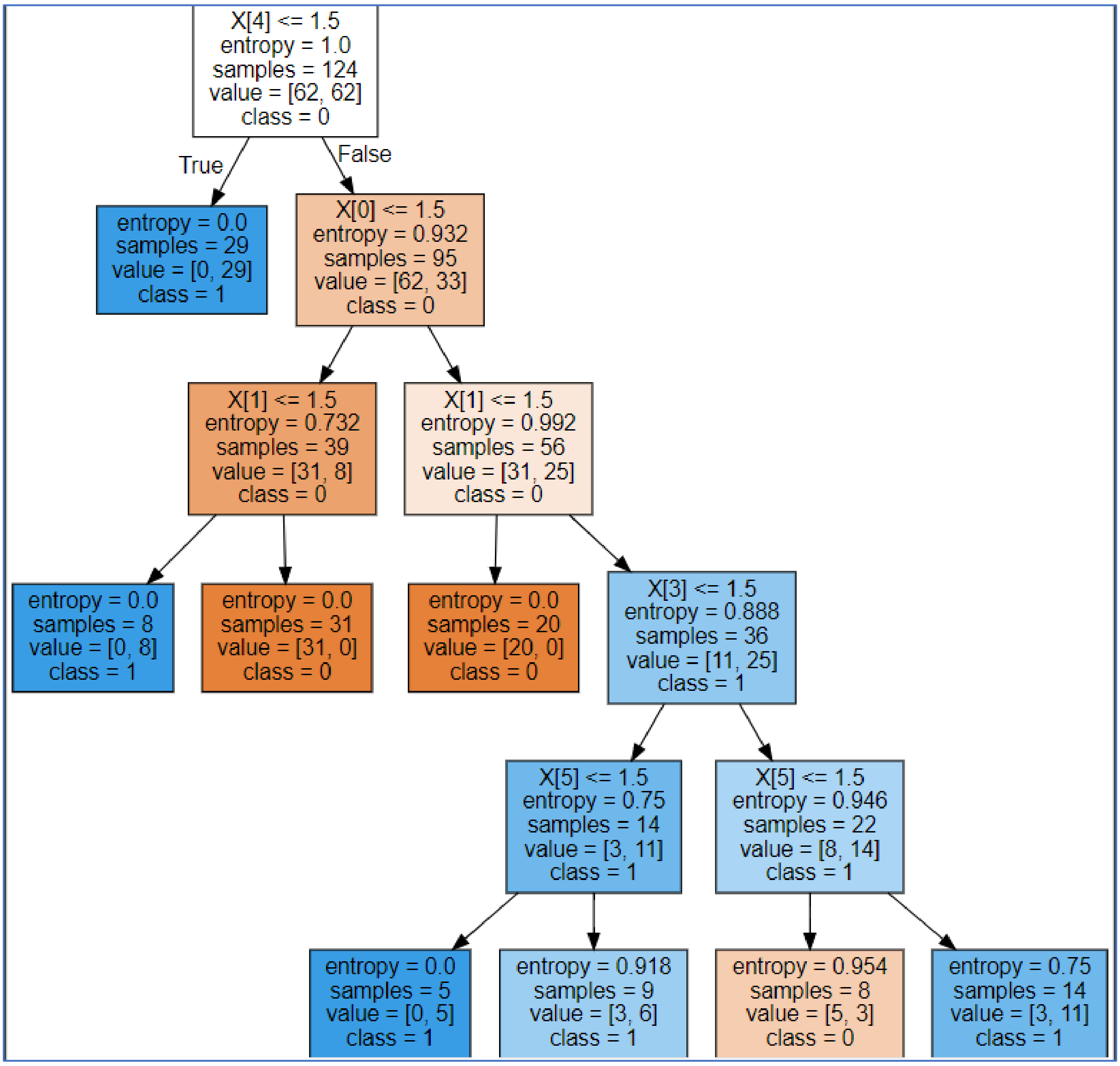


Figure 12 : Scikit Learn Depth-5 Decision Tree

Part E:

I have implemented part C and D for the iris dataset from the UCI machine learning repository. Following are the decision trees and corresponding confusion matrix.

Using the ID3 Decision Tree Algorithm Implemented in Part A

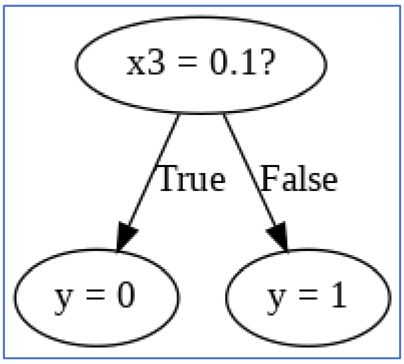
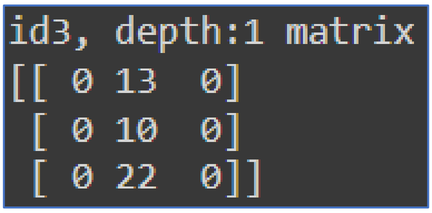


Figure 13 : Decision Tree for Iris Dataset - Depth 1



```
id3, depth:1 matrix
[[ 0 13  0]
 [ 0 10  0]
 [ 0 22  0]]
```

Figure 14 : Confusion Matrix for Iris Dataset - Depth 1

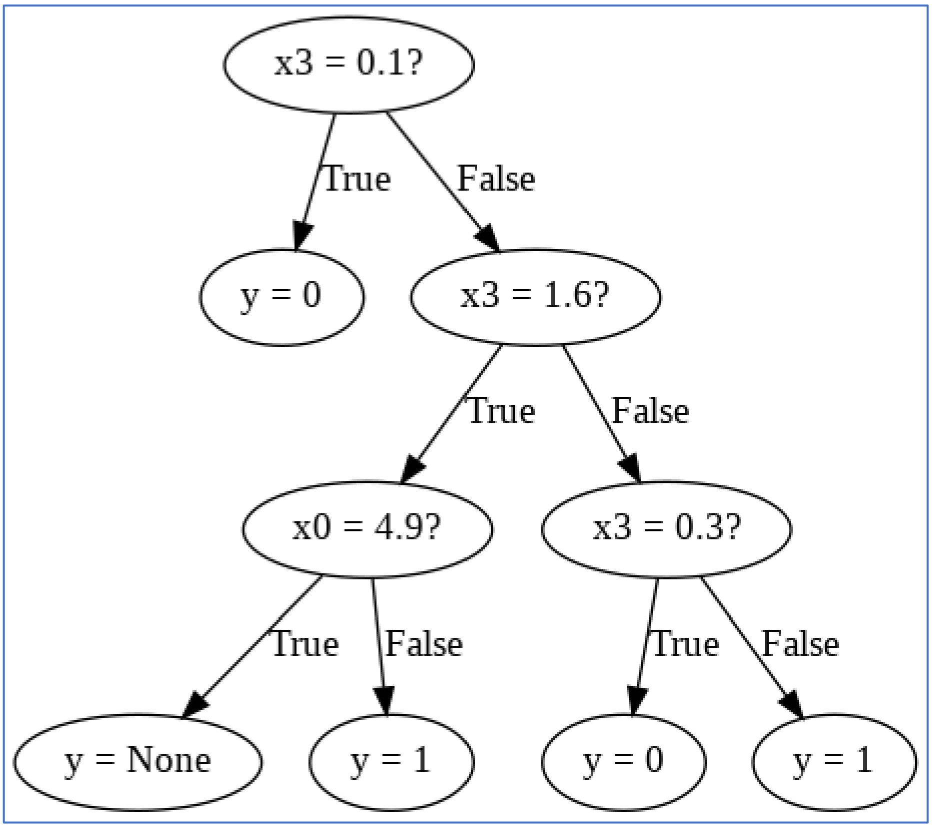
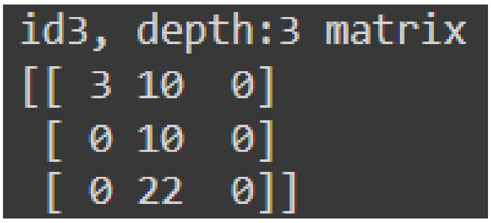


Figure 15 : Decision Tree for Iris Dataset - Depth 3



```
id3, depth:3 matrix
[[ 3 10  0]
 [ 0 10  0]
 [ 0 22  0]]
```

Figure 16 : Confusion Matrix for Iris Dataset - Depth 3

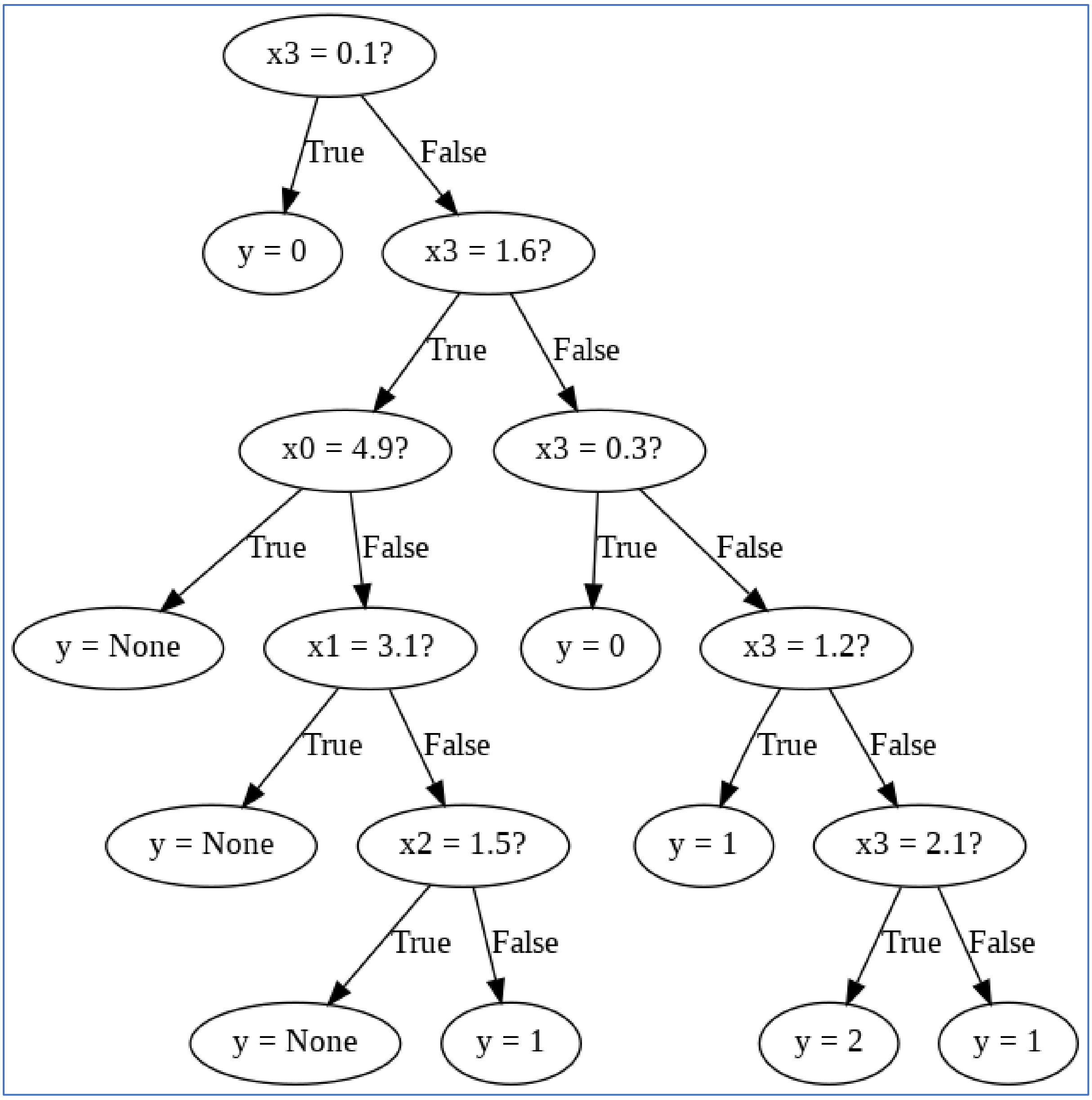
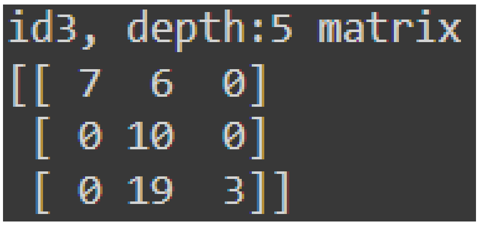


Figure 17 : Decision Tree for Iris Dataset - Depth 5



```
id3, depth:5 matrix
[[ 7  6  0]
 [ 0 10  0]
 [ 0 19  3]]
```

Figure 18 : Confusion Matrix for Iris Dataset - Depth 5

Using Scikit Learn’s ID3 Implementation:

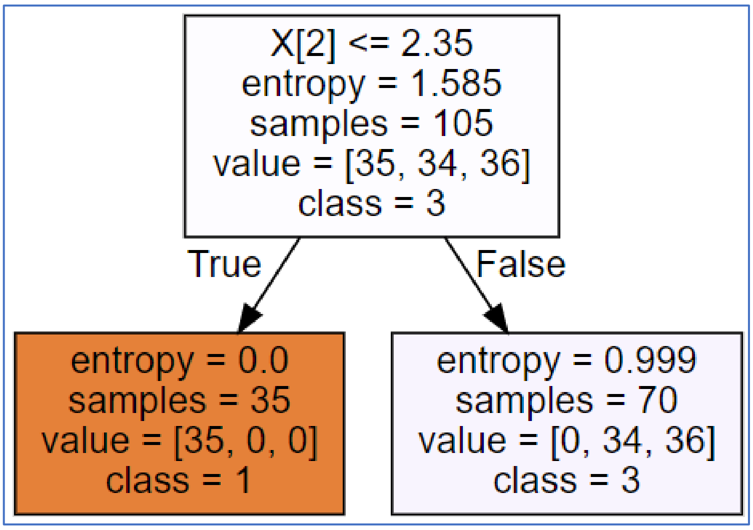
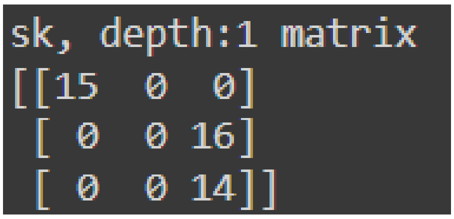


Figure 19 : Decision Tree (Depth 1) for Iris Dataset with Scikit Learn



```
sk, depth:1 matrix
[[15  0  0]
 [ 0  0 16]
 [ 0  0 14]]
```

Figure 20: Confusion Matrix (Depth 1) for Iris Dataset with Scikit Learn

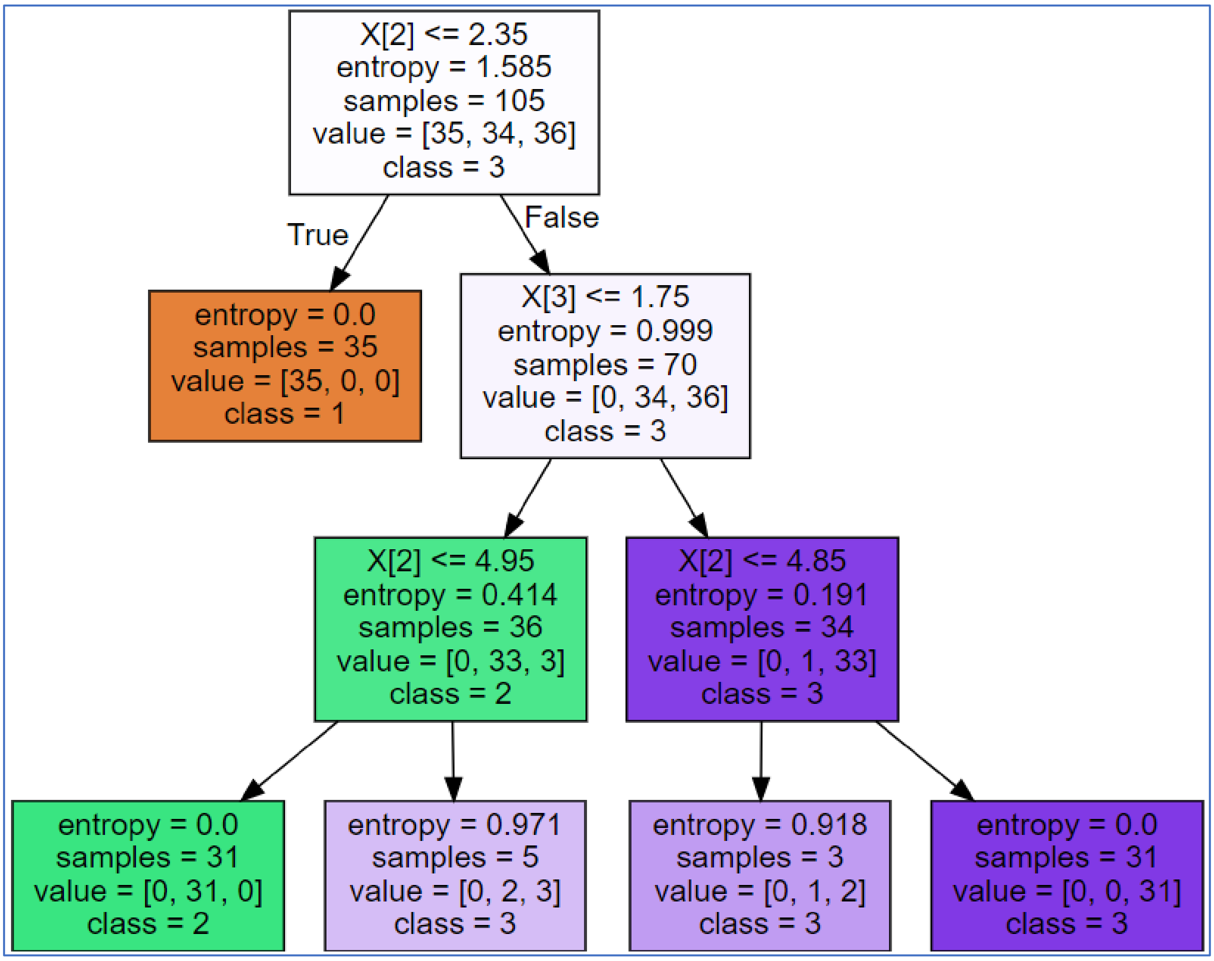
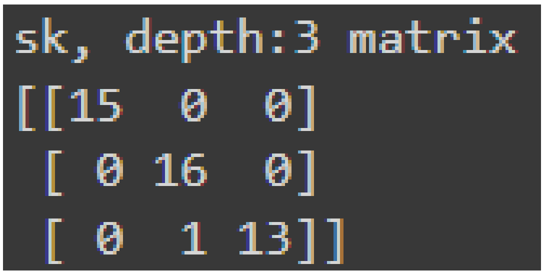


Figure 21 : Decision Tree (Depth 3) for Iris Dataset with Scikit Learn



```
sk, depth:3 matrix
[[15  0  0]
 [ 0 16  0]
 [ 0  1 13]]
```

Figure 22 : Confusion Matrix (Depth 3) for Iris Dataset with Scikit Learn

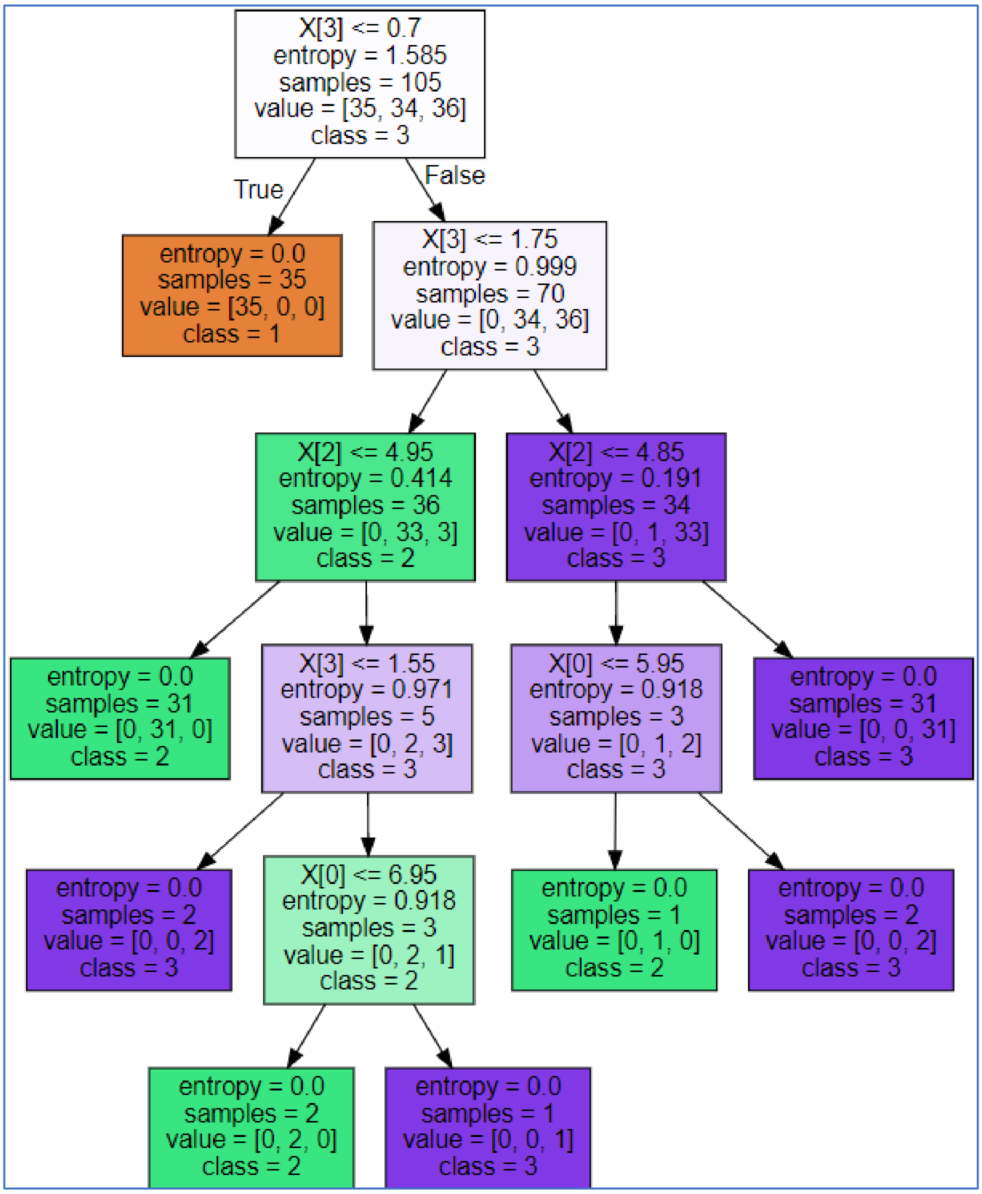
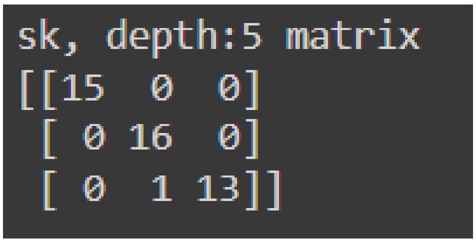


Figure 23 : Decision Tree (Depth 5) for Iris Dataset with Scikit Learn



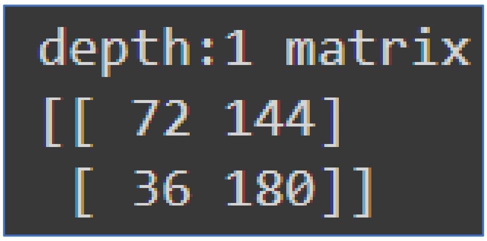
```
sk, depth:5 matrix
[[15  0  0]
 [ 0 16  0]
 [ 0  1 13]]
```

Figure 24 : Confusion Matrix (Depth 5) for Iris Dataset with Scikit Learn

Comparison of Confusion Matrices

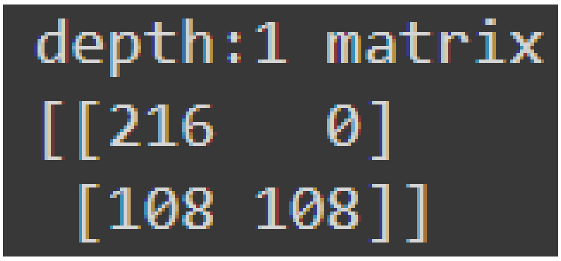
Depth 1:

ID3 Implementation:



```
depth:1 matrix
[[ 72 144]
 [ 36 180]]
```

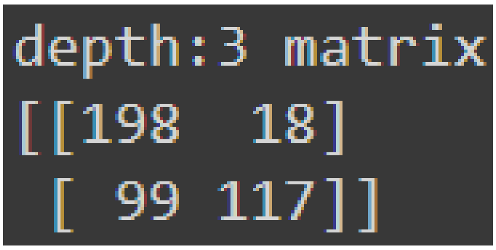
Scikit Learn:



```
depth:1 matrix
[[216   0]
 [108 108]]
```

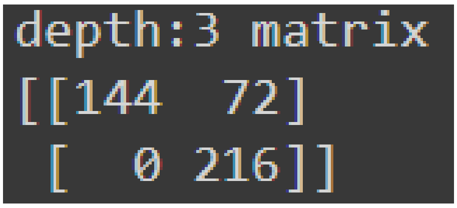
Depth 3:

ID3 Implementation



```
depth:3 matrix
[[198  18]
 [ 99 117]]
```

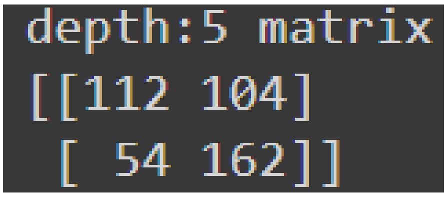
Scikit Learn:



```
depth:3 matrix
[[144  72]
 [  0 216]]
```

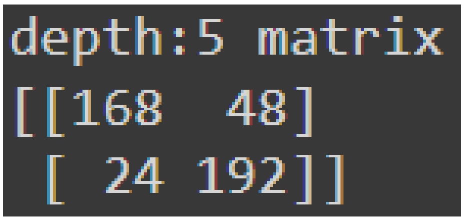
Depth 5:

ID3 Implementation



```
 depth:5 matrix
 [[112 104]
  [ 54 162]]
```

Scikit Learn



```
depth:5 matrix
[[168  48]
 [ 24 192]]
```

We can see that in both the implementations of the decision tree algorithm, the best confusion matrix is depicted in case the max_depth is set to 3. In both cases, the number of false positives and false negatives(non-diagonal elements) are high in case of max_depth 1 and 5. This can be due to underfitting in case of max_depth 1 and overfitting in case of max depth 5.

In general, we can see that the performance of the scikit learn algorithm is slightly better than the ID3 implementation of the decision tree.