An Report
On

Machine Learning : Decision Tree
Assignment - 2

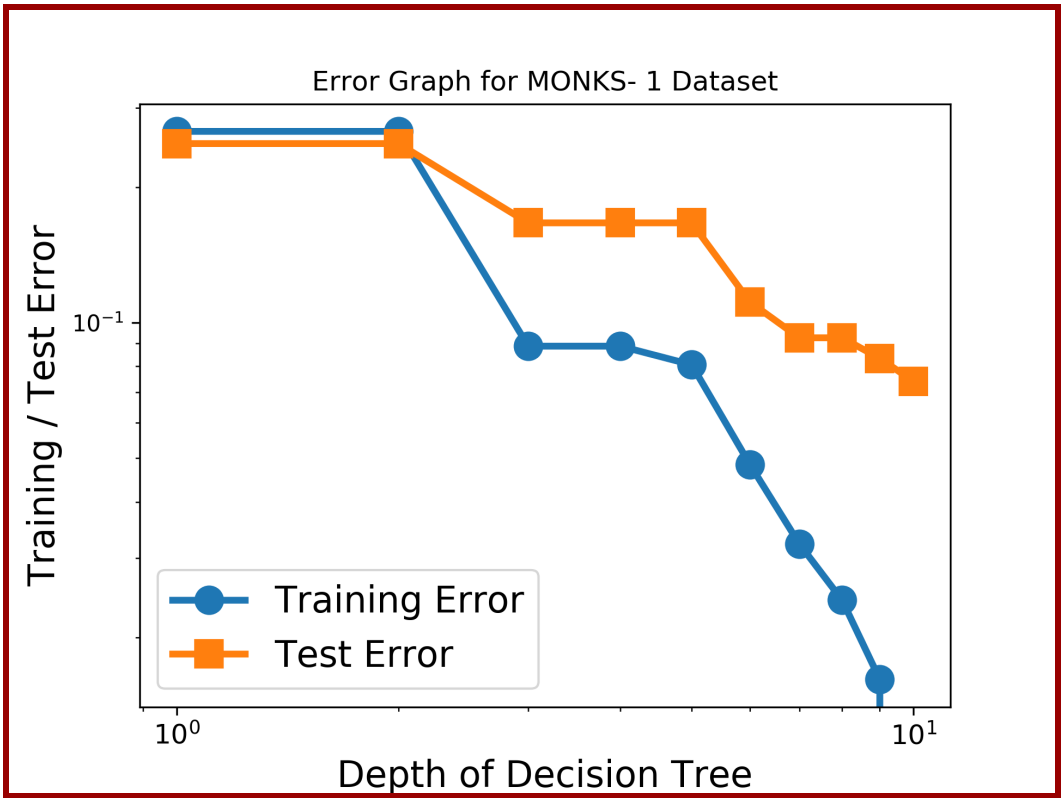
By

Soumyadeep Choudhury
Net ID :- sxc180056
UTD ID :- 2021439916

DECISION TREE RESULTS

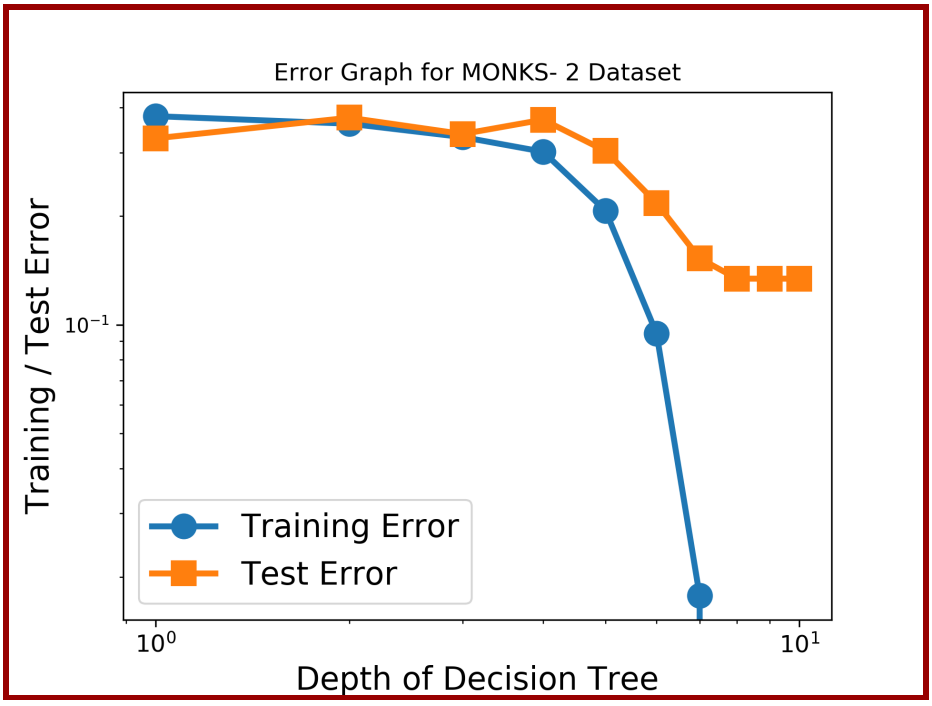
1. ***Monks - 1 Dataset : Implementing id3***

   - Average Training Error : **0.0911**
   - Average Testing Error : **0.1454**
   - Graph Comparison - Testing vs Training :



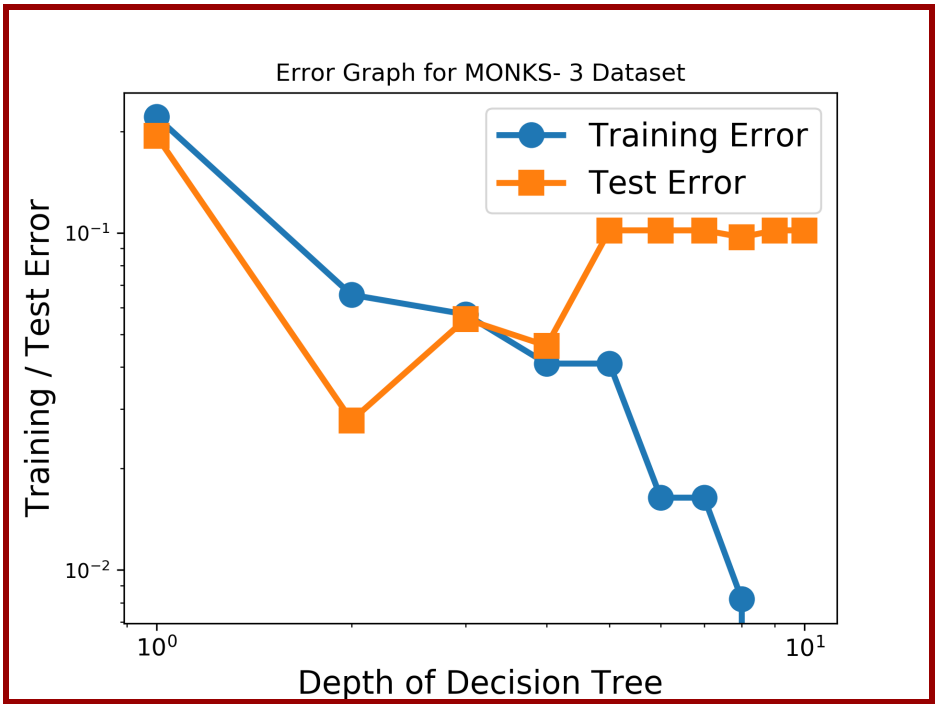
2. ***Monks - 2 Dataset : Implementing id3***

   - Average Training Error : **0.1692**
   - Average Testing Error : **0.2488**
   - Graph Comparison - Testing vs Training :



3. ***Monks - 3 Dataset : Implementing id3***

- Average Training Error : **0.0467**
- Average Testing Error : **0.0931**
- Graph Comparison - Testing vs Training :



4. ***Comparison Between Scikit-Learn and id3 (Monks : Dataset - 1)***

➔ **For Depth = 1 :**

A. Id3

- **Decision Tree :**

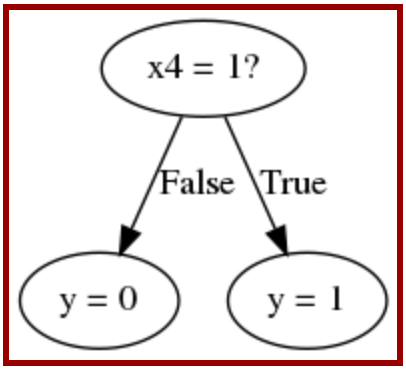
  TREE
  +-- [SPLIT: x4 = 1 False]
  | +-- [LABEL = 0]
  +-- [SPLIT: x4 = 1 True]
  | +-- [LABEL = 1]

- **Confusion Matrix :**

  MONKS-1 DATASET : Learning using id3 : Confusion matrix for depth -> 1

	Predicted Positives	Predicted Negatives
True Positives	216	0
True Negatives	108	108

- **Tree Graph / Image :**



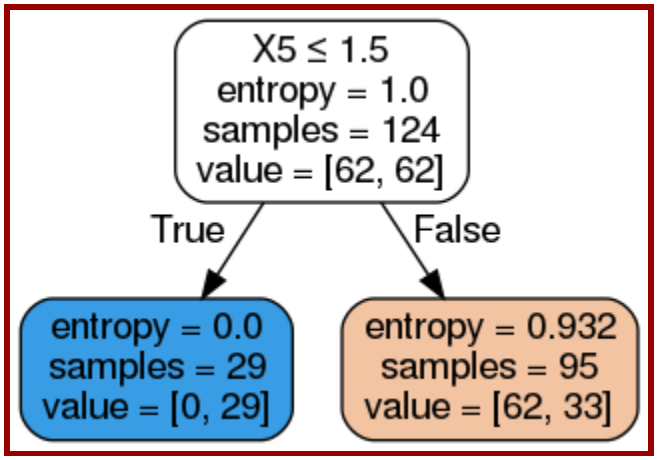
B. Scikit-Learn

- **Confusion Matrix :**

  MONKS-1 DATASET : Learning using scikit-learn : Confusion matrix for depth -> 1

	Predicted Positives	Predicted Negatives
True Positives	216	0
True Negatives	108	108

- **Tree Graph / Image :**



##

➔ **For Depth = 3 :**

A. Id3

- **Decision Tree :**

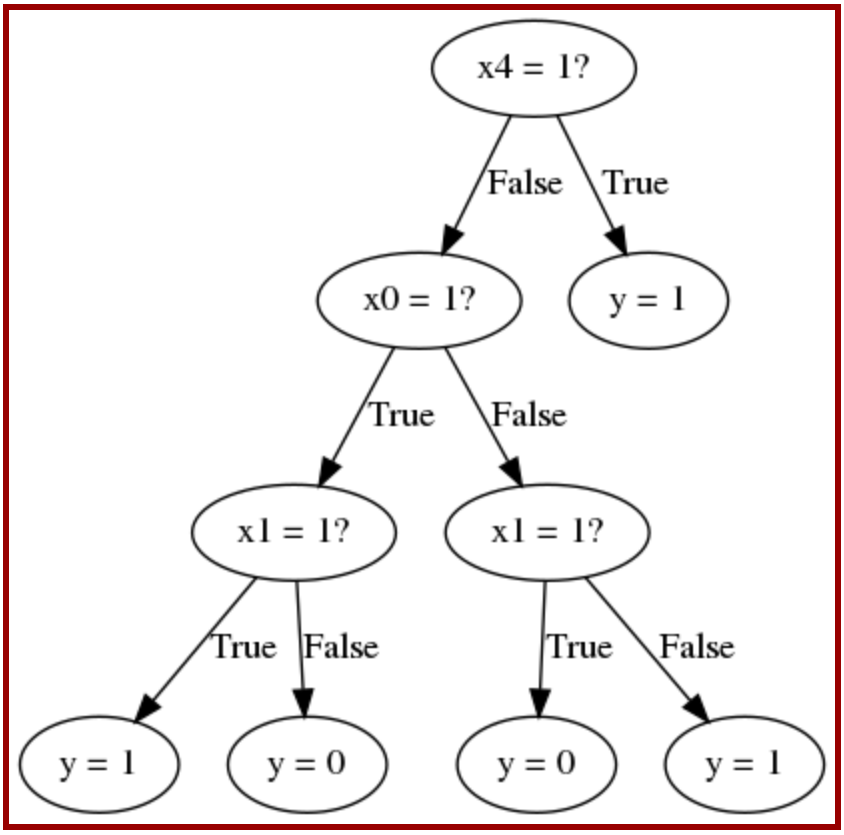
```
TREE
+-- [SPLIT: x4 = 1 False]
|       +-- [SPLIT: x0 = 1 True]
|       |       +-- [SPLIT: x1 = 1 True]
|       |       |       +-- [LABEL = 1]
|       |       +-- [SPLIT: x1 = 1 False]
|       |       |       +-- [LABEL = 0]
|       +-- [SPLIT: x0 = 1 False]
|       |       +-- [SPLIT: x1 = 1 True]
|       |       |       +-- [LABEL = 0]
|       |       +-- [SPLIT: x1 = 1 False]
|       |       |       +-- [LABEL = 1]
+-- [SPLIT: x4 = 1 True]
|       +-- [LABEL = 1]
```

- **Confusion Matrix :**

MONKS-1 DATASET : Learning using id3 : Confusion matrix for depth -> 3

	Predicted Positives	Predicted Negatives
True Positives	144	72
True Negatives	0	216

- **Tree Graph / Image :**

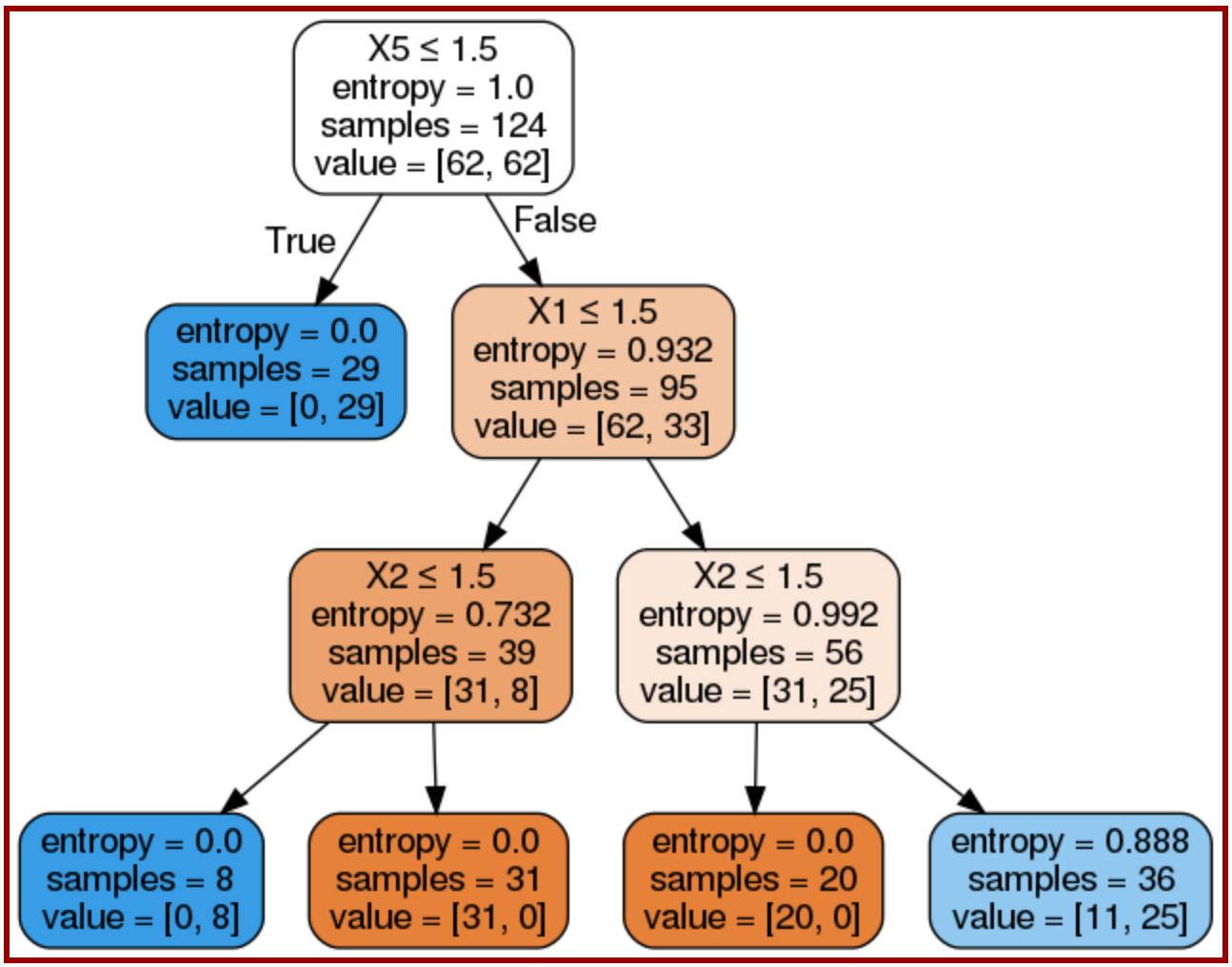


B. Scikit-Learn

- **Confusion Matrix :**
  MONKS-1 DATASET : Learning using scikit-learn : Confusion matrix for depth -> 3

	Predicted Positives	Predicted Negatives
True Positives	144	72
True Negatives	0	216

- **Tree Graph / Image :**



##

➔ **For Depth = 5 :**

A. Id3

- **Decision Tree :**

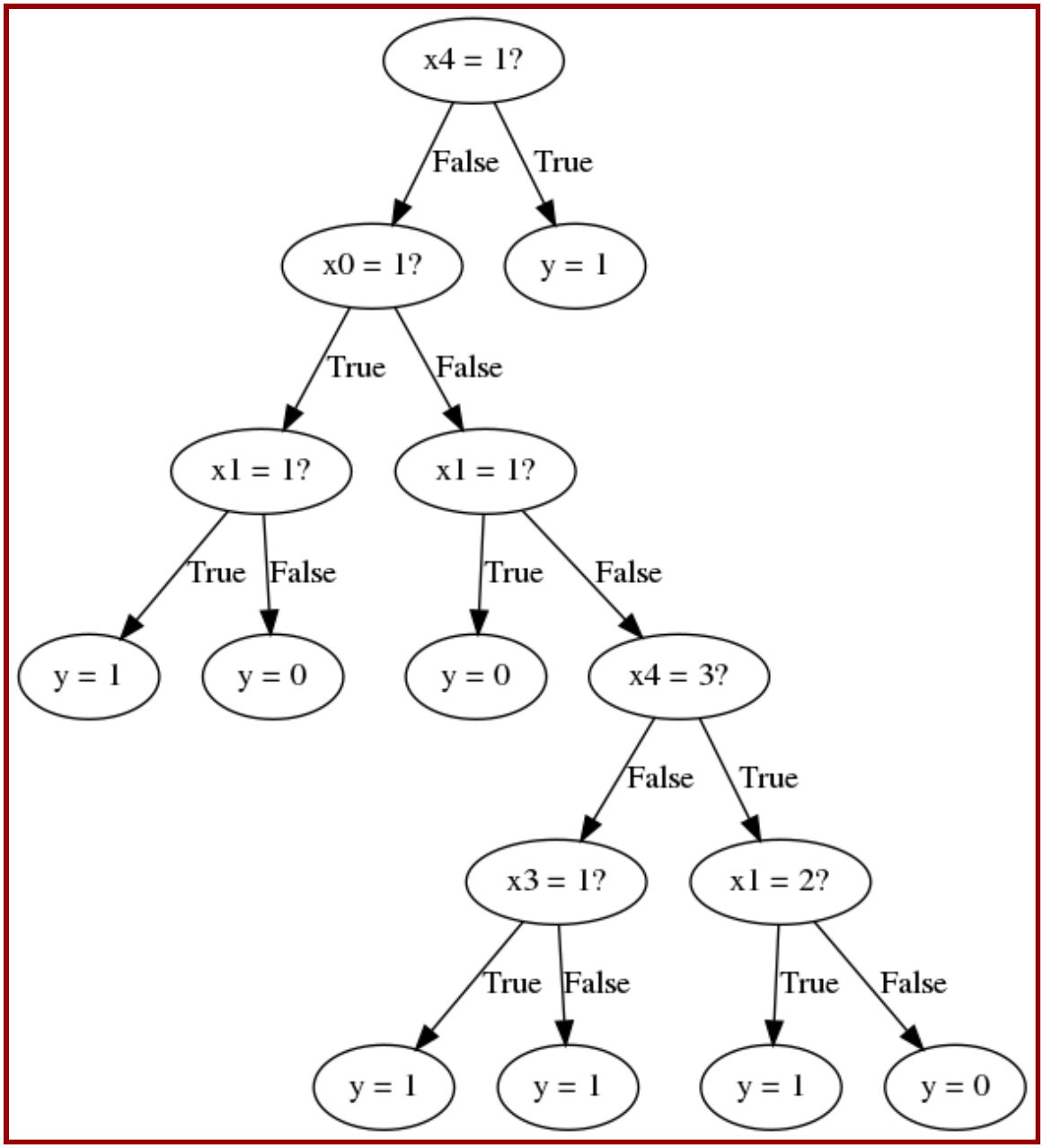
TREE
+-- [SPLIT: x4 = 1 False]
| +-- [SPLIT: x0 = 1 True]
| | +-- [SPLIT: x1 = 1 True]
| | | +-- [LABEL = 1]
| | +-- [SPLIT: x1 = 1 False]
| | | +-- [LABEL = 0]
| +-- [SPLIT: x0 = 1 False]
| | +-- [SPLIT: x1 = 1 True]

```
|       |       |       +-- [LABEL = 0]
|       |       +-- [SPLIT: x1 = 1 False]
|       |       |       +-- [SPLIT: x4 = 3 False]
|       |       |       |       +-- [SPLIT: x3 = 1 True]
|       |       |       |       |       +-- [LABEL = 1]
|       |       |       |       +-- [SPLIT: x3 = 1 False]
|       |       |       |       |       +-- [LABEL = 1]
|       |       |       +-- [SPLIT: x4 = 3 True]
|       |       |       |       +-- [SPLIT: x1 = 2 True]
|       |       |       |       |       +-- [LABEL = 1]
|       |       |       |       +-- [SPLIT: x1 = 2 False]
|       |       |       |       |       +-- [LABEL = 0]
+-- [SPLIT: x4 = 1 True]
|       +-- [LABEL = 1]
```

- **Confusion Matrix :**
  MONKS-1 DATASET : Learning using id3 : Confusion matrix for depth -> 5

	Predicted Positives	Predicted Negatives
True Positives	156	60
True Negatives	12	204

- **Tree Graph / Image :**



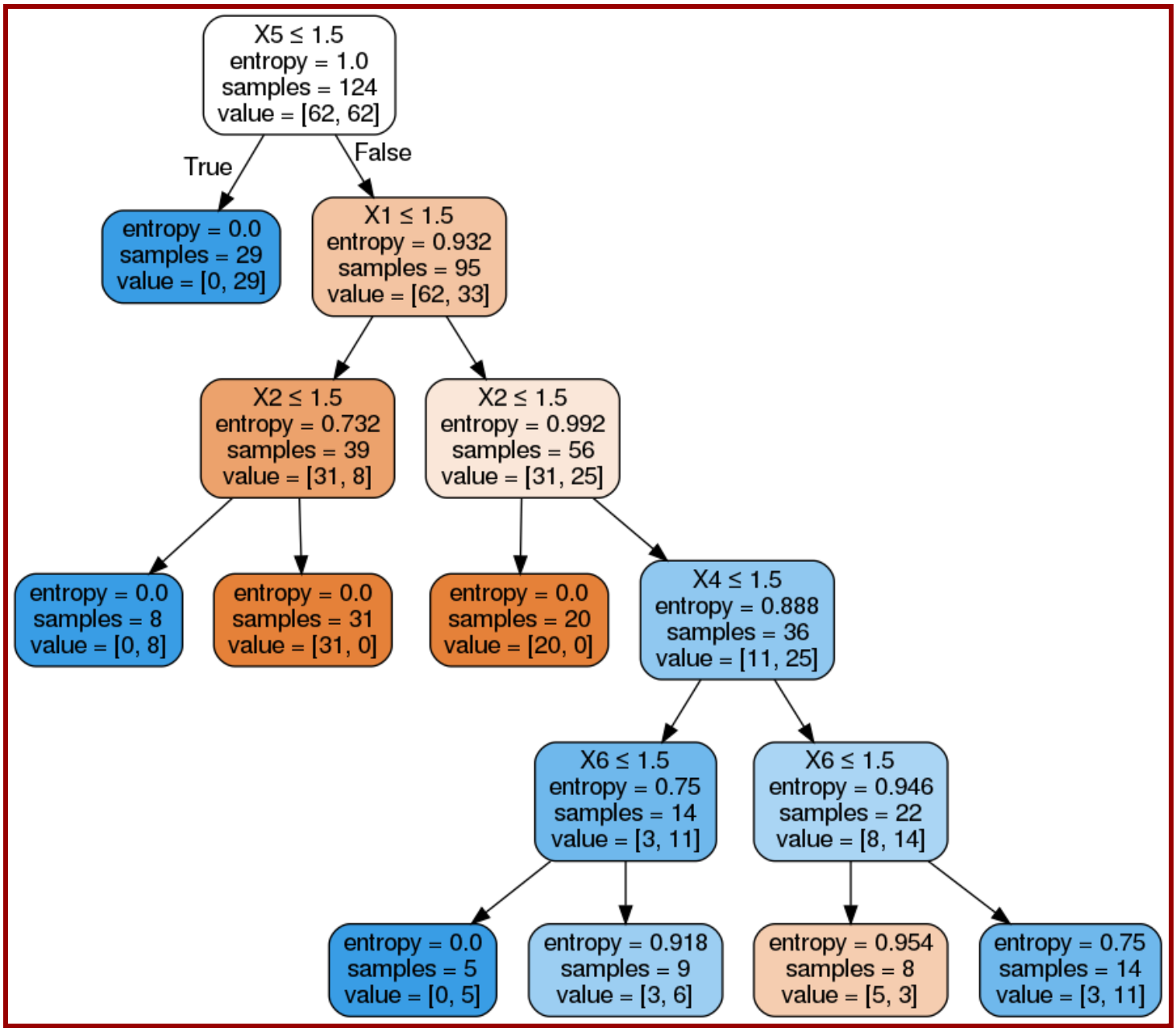
B. Scikit-Learn

- **Confusion Matrix :**

  MONKS-1 DATASET : Learning using scikit-learn : Confusion matrix for depth -> 5

	Predicted Positives	Predicted Negatives
True Positives	168	48
True Negatives	24	192

- **Tree Graph / Image :**



5. ***Results of UCI dataset : Scikit-Learn vs id3***

➔ **For Depth = 1 :**

C. Id3

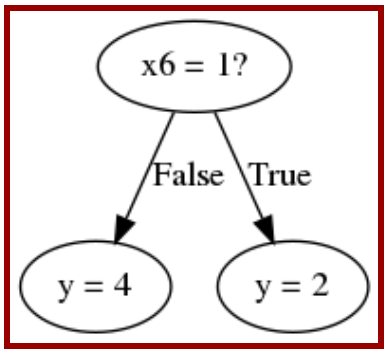
- **Decision Tree :**

  TREE

+-- [SPLIT: x6 = 1 False]
| +-- [LABEL = 4]
+-- [SPLIT: x6 = 1 True]
| +-- [LABEL = 2]

- **Confusion Matrix :**
  Winsconsin breast cancer : id3 : Depth -> 1 : Confusion Matrix
  [[115 28]
  [7 60]]

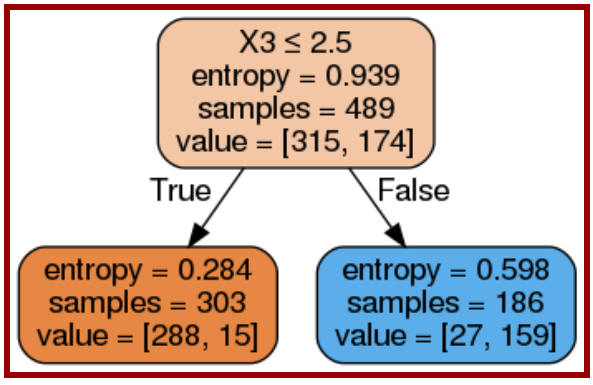
- **Tree Graph / Image :**



D. Scikit-Learn

- **Confusion Matrix :**
  Winsconsin breast cancer : scikit-learn : Depth -> 1: Confusion Matrix
  [[129 14]
  [2 65]]

- **Tree Graph / Image :**



##

➔ **For Depth = 3 :**

C. Id3

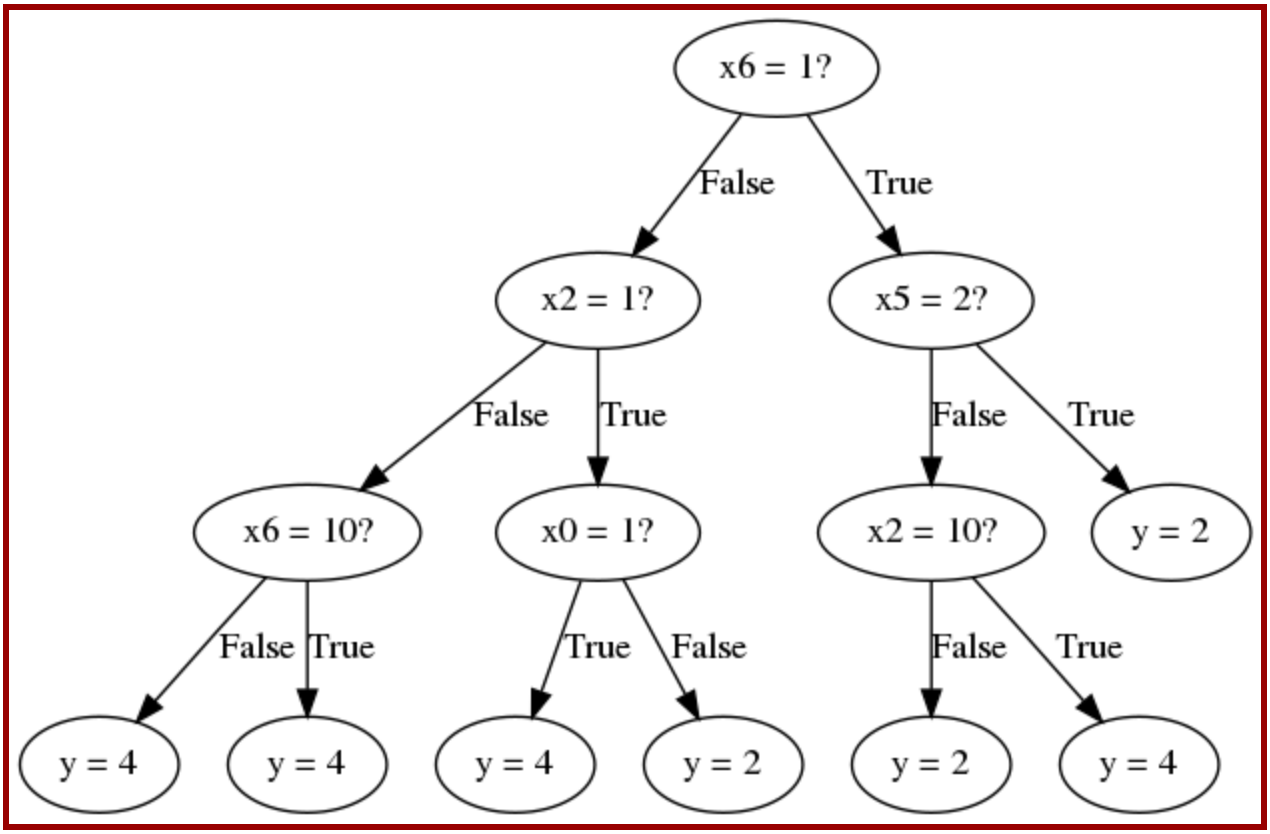
- **Decision Tree :**

```
TREE
+-- [SPLIT: x6 = 1 False]
|       +-- [SPLIT: x2 = 1 False]
|       |       +-- [SPLIT: x6 = 10 False]
|       |       |       +-- [LABEL = 4]
|       |       +-- [SPLIT: x6 = 10 True]
|       |       |       +-- [LABEL = 4]
|       +-- [SPLIT: x2 = 1 True]
|       |       +-- [SPLIT: x0 = 1 True]
|       |       |       +-- [LABEL = 4]
|       |       +-- [SPLIT: x0 = 1 False]
|       |       |       +-- [LABEL = 2]
+-- [SPLIT: x6 = 1 True]
|       +-- [SPLIT: x5 = 2 False]
|       |       +-- [SPLIT: x2 = 10 False]
|       |       |       +-- [LABEL = 2]
|       |       +-- [SPLIT: x2 = 10 True]
|       |       |       +-- [LABEL = 4]
|       +-- [SPLIT: x5 = 2 True]
|       |       +-- [LABEL = 2]
```

- **Confusion Matrix :**

```
Winsconsin breast cancer : id3 : Depth -> 3 : Confusion Matrix
[[135   8]
 [  6  61]]
```

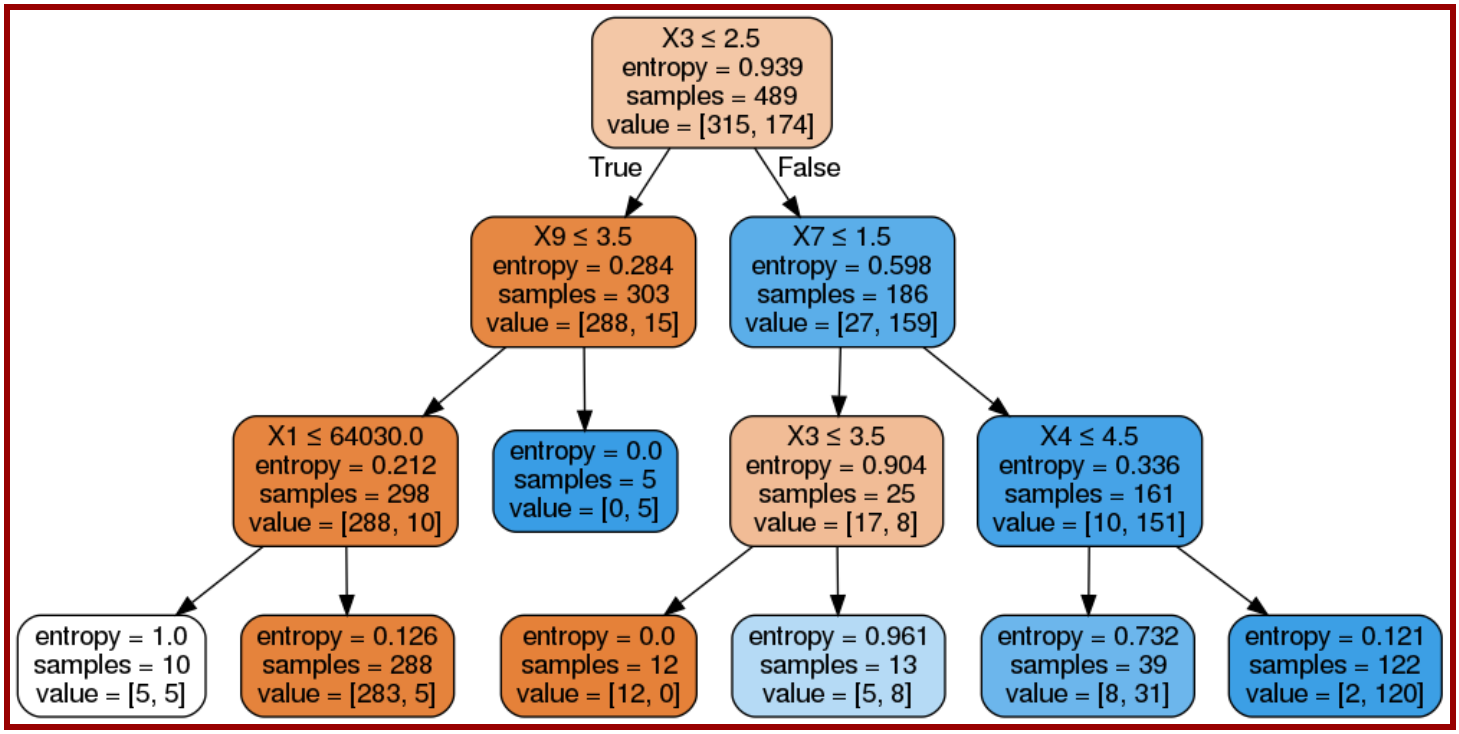
- **Tree Graph / Image :**



D. Scikit-Learn

- **Confusion Matrix :**
  Winsconsin breast cancer : scikit-learn : Depth -> 3: Confusion Matrix
  [[134 9]
  [0 67]]

- **Tree Graph / Image :**



##

➔ **For Depth = 5 :**

C. Id3

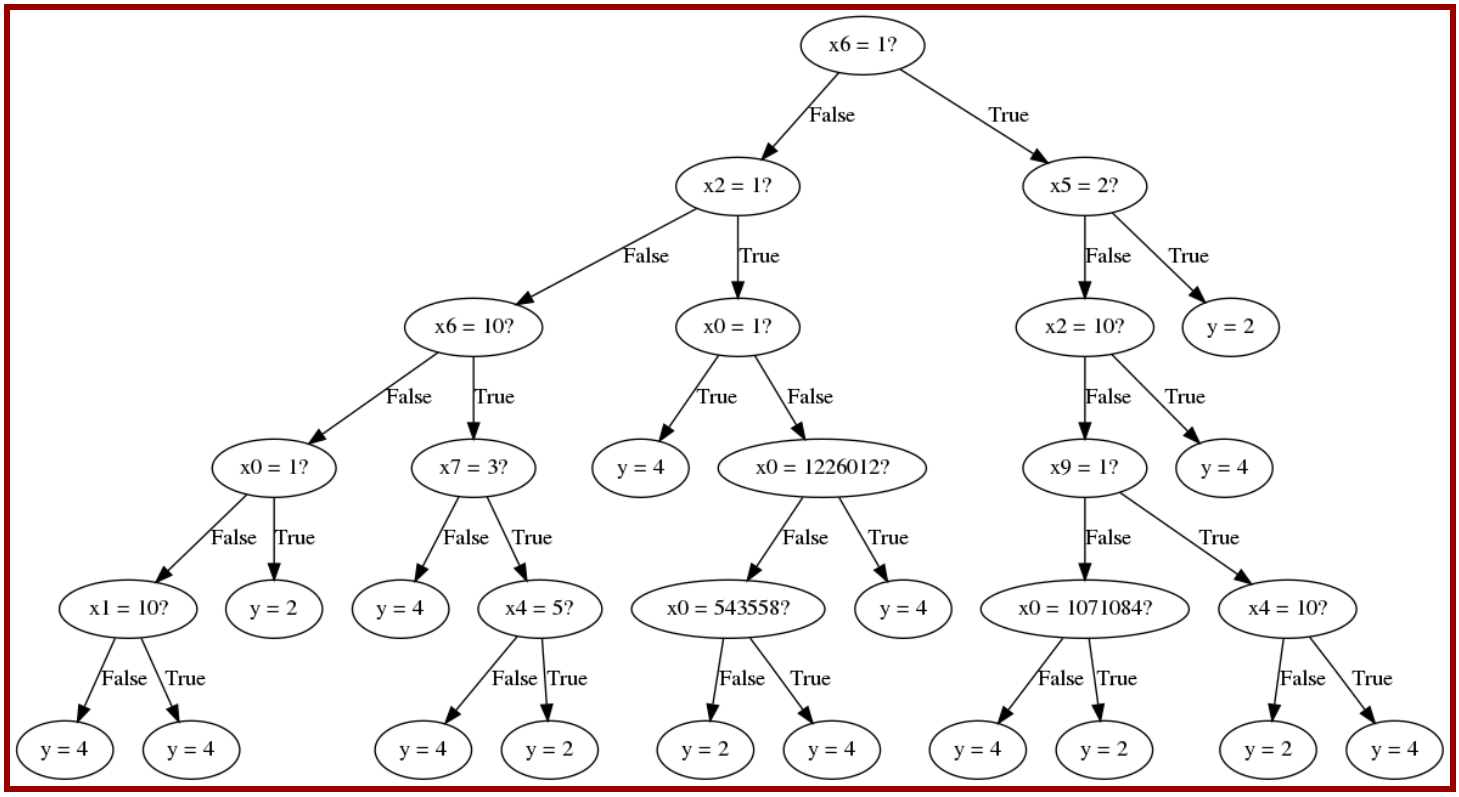
- **Decision Tree :**

```
TREE
+-- [SPLIT: x6 = 1 False]
|       +-- [SPLIT: x2 = 1 False]
|       |       +-- [SPLIT: x6 = 10 False]
|       |       |       +-- [SPLIT: x0 = 1 False]
|       |       |       |       +-- [SPLIT: x1 = 10 False]
|       |       |       |       |       +-- [LABEL = 4]
|       |       |       |       +-- [SPLIT: x1 = 10 True]
|       |       |       |       |       +-- [LABEL = 4]
|       |       |       +-- [SPLIT: x0 = 1 True]
|       |       |       |       +-- [LABEL = 2]
|       |       +-- [SPLIT: x6 = 10 True]
|       |       |       +-- [SPLIT: x7 = 3 False]
|       |       |       |       +-- [LABEL = 4]
|       |       |       +-- [SPLIT: x7 = 3 True]
```

```
|       |       |       |       +-- [SPLIT: x4 = 5 False]
|       |       |       |       |       +-- [LABEL = 4]
|       |       |       |       +-- [SPLIT: x4 = 5 True]
|       |       |       |       |       +-- [LABEL = 2]
|       |       +-- [SPLIT: x2 = 1 True]
|       |       |       +-- [SPLIT: x0 = 1 True]
|       |       |       |       +-- [LABEL = 4]
|       |       |       +-- [SPLIT: x0 = 1 False]
|       |       |       |       +-- [SPLIT: x0 = 1226012 False]
|       |       |       |       |       +-- [SPLIT: x0 = 543558 False]
|       |       |       |       |       |       +-- [LABEL = 2]
|       |       |       |       |       +-- [SPLIT: x0 = 543558 True]
|       |       |       |       |       |       +-- [LABEL = 4]
|       |       |       |       +-- [SPLIT: x0 = 1226012 True]
|       |       |       |       |       +-- [LABEL = 4]
+-- [SPLIT: x6 = 1 True]
|       +-- [SPLIT: x5 = 2 False]
|       |       +-- [SPLIT: x2 = 10 False]
|       |       |       +-- [SPLIT: x9 = 1 False]
|       |       |       |       +-- [SPLIT: x0 = 1071084 False]
|       |       |       |       |       +-- [LABEL = 4]
|       |       |       |       +-- [SPLIT: x0 = 1071084 True]
|       |       |       |       |       +-- [LABEL = 2]
|       |       |       +-- [SPLIT: x9 = 1 True]
|       |       |       |       +-- [SPLIT: x4 = 10 False]
|       |       |       |       |       +-- [LABEL = 2]
|       |       |       |       +-- [SPLIT: x4 = 10 True]
|       |       |       |       |       +-- [LABEL = 4]
|       |       +-- [SPLIT: x2 = 10 True]
|       |       |       +-- [LABEL = 4]
|       +-- [SPLIT: x5 = 2 True]
|       |       +-- [LABEL = 2]
```

- **Confusion Matrix :**
  Winsconsin breast cancer : id3 : Depth -> 5 : Confusion Matrix
  [[135 8]
  [4 63]]

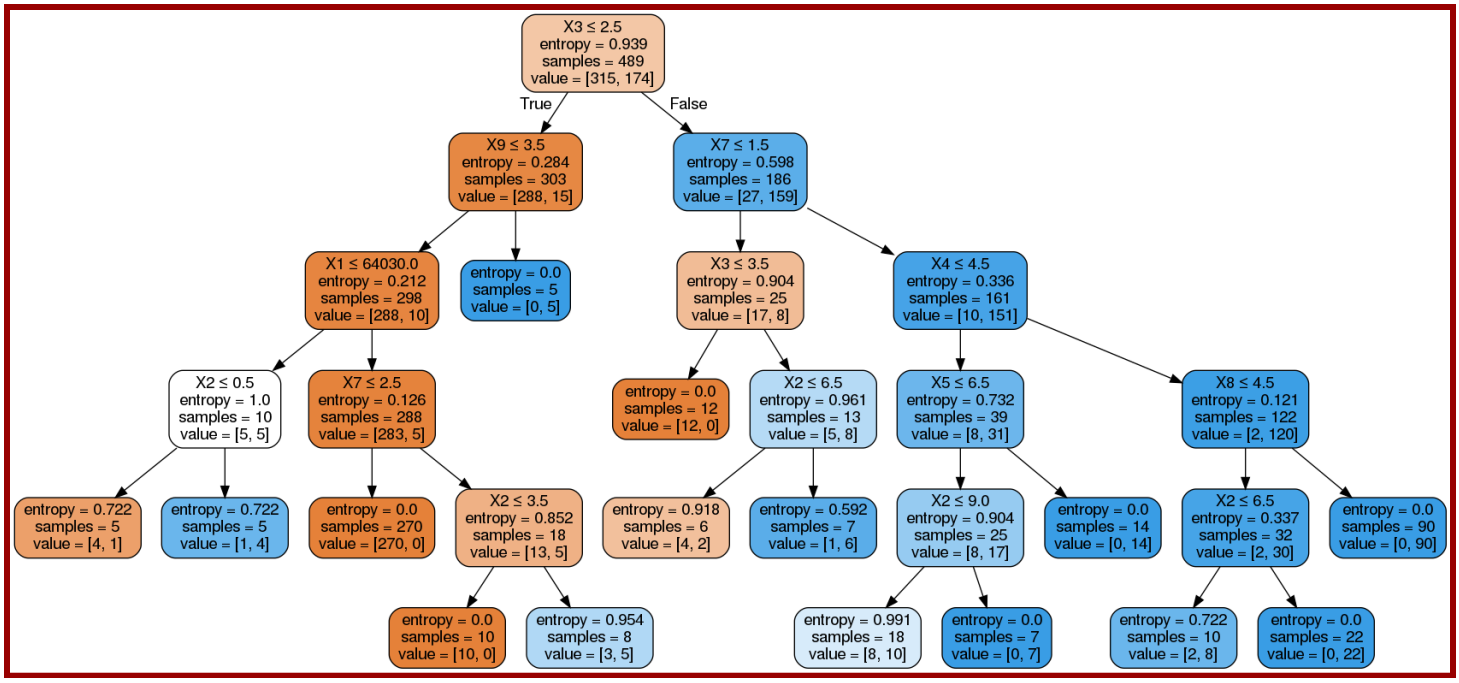
- **Tree Graph / Image :**



D. Scikit-Learn

- **Confusion Matrix :**
  Winsconsin breast cancer : scikit-learn : Depth -> 5: Confusion Matrix
  [[133 10]
  [3 64]]

- **Tree Graph / Image :**



6. ***Discussion Project***

ID3 -

- For comparing the information between the data points, it uses : “equal to” (=)
- The data points of an attribute having the same value are split to the left side of the tree
- The data points of an attribute having different value are split to the right side of the tree
- The testing error tends to saturate after depth = 5 : My choice : depth = 5

Scikit Learn -

- For comparing the information between the data points, it uses : “less than or equal to” (<=)
- The data points of an attribute having less or equal value are split to the left side of the tree
- The data points of an attribute having higher value are split to the right side of the tree
- More efficient computation on the information gain due to its E_mac approximation in entropy
- It has lower false positives than id3. Best depth (my choice) = 5

Conclusion - The final results and the generated results are highly similar with very low error which can be observed from the graph and confusion matrix. The results may vary due to better approximation, smoother gradient error computation and effective data preprocessing or handling error values through pruning or mean-bias approximation.

---------- **END** ----------