DMG Assignment 1

STEPS TO RUN THE CODE

1. Download all the .py and .ipynb files.
2. Run the Data_Preprocessing.ipynb file to save the preprocessed data.
3. Run the automate.py file to get results for Q1(accuracy,precision,recall,auc-roc), the visualization results for Q2 and the results for the statistical tests performed in Q3.
4. Run the automate2.py file to get the results for k-fold cross-validation of single attribute split. K-fold-cross validation for multiple attributes can be done similarly.

ASSUMPTIONS FOR VISUALIZATIONS

1. Did not create a graphic representation for every tree.
2. Stored the information regarding the splits at each node.
3. Tree is presented as level order traversal

Ques 1)

a. For single attribute split:

```
Single Attribute Accuracy for dataset fetal_health is: 0.8943661971830986
Single Attribute Recall for dataset fetal_health is: 0.8943661971830986
Single Attribute Precision for dataset fetal_health is: 0.899548463628636
Single Attribute F1-Score for dataset fetal_health is: 0.8965688705495971

Single Attribute Accuracy for dataset banking_dataset is: 0.9021607186210245
Single Attribute Recall for dataset banking_dataset is: 0.9021607186210245
Single Attribute Precision for dataset banking_dataset is: 0.8978487749725944
Single Attribute F1-Score for dataset banking_dataset is: 0.8997595299016761

Single Attribute Accuracy for dataset cervical_cancer is: 0.9534883720930233
Single Attribute Recall for dataset cervical_cancer is: 0.9534883720930233
Single Attribute Precision for dataset cervical_cancer is: 0.9503588860178007
Single Attribute F1-Score for dataset cervical_cancer is: 0.9515186532375612
```

b. For multiple attribute split:

```
Multi Attribute Accuracy for dataset fetal_health is: 0.9272300469483568
Multi Attribute Recall for dataset fetal_health is: 0.9272300469483568
Multi Attribute Precision for dataset fetal_health is: 0.9246593066790796
Multi Attribute F1-Score for dataset fetal_health is: 0.9246760902616017

Multi Attribute Accuracy for dataset banking_dataset is: 0.9047098810390871
Multi Attribute Recall for dataset banking_dataset is: 0.9047098810390871
Multi Attribute Precision for dataset banking_dataset is: 0.8945751315570651
Multi Attribute F1-Score for dataset banking_dataset is: 0.8976631563563743

Multi Attribute Accuracy for dataset cervical_cancer is: 0.9767441860465116
Multi Attribute Recall for dataset cervical_cancer is: 0.9767441860465116
Multi Attribute Precision for dataset cervical_cancer is: 0.9757967269595177
Multi Attribute F1-Score for dataset cervical_cancer is: 0.9757593266187806
```

Ques 2)

Visualization of single attribute decision tree for Fetal_Health Dataset

```
the tree is splitted according to the following features
Level  0
Nodes are :
18 ( Threshold : 107.0 )
Level  1
Nodes are :
1 ( Threshold : 110.0 )  9 ( Threshold : 0.5 )
Level  2
Nodes are :
19 ( Threshold : 97.0 )  16 ( Threshold : 1.0 )  10 ( Threshold : 68.0 )  10 ( Threshold : 6.0 )
Level  3
Nodes are :
5 ( Threshold : 0.012 )  15 ( Threshold : 9.0 )  8 ( Threshold : 79.0 )  9 ( Threshold : 0.4 )  16 (
Threshold : 3.0 )  14 ( Threshold : 185.0 )
Level  4
Nodes are :
12 ( Threshold : 68.0 )  8 ( Threshold : 58.0 )  20 ( Threshold : 36.0 )  8 ( Threshold : 59.0 )  11 (
Threshold : 6.3 )  14 ( Threshold : 153.0 )  9 ( Threshold : 0.5 )  7 ( Threshold : 0.001 )  7 ( Threshold
: 0.0 )  19 ( Threshold : 140.0 )  17 ( Threshold : 125.0 )
Level  5
Nodes are :
```

```
2 ( Threshold : 0.0 )  1 ( Threshold : 137.0 )  10 ( Threshold : 6.0 )  0 ( Threshold : 1330.0 )  3 (
Threshold : 0.0 )  19 ( Threshold : 137.0 )  7 ( Threshold : 0.0 )  8 ( Threshold : 35.0 )  8 ( Threshold :
44.0 )  21 ( Threshold : 0.0 )  17 ( Threshold : 154.0 )
Level  6
Nodes are :
18 ( Threshold : 106.0 )  11 ( Threshold : 9.5 )  21 ( Threshold : -1.0 )  21 ( Threshold : -1.0 )  13 (
Threshold : 133.0 )  10 ( Threshold : 31.0 )  0 ( Threshold : 53.0 )  6 ( Threshold : 0.0 )  14 (
Threshold : 181.0 )  13 ( Threshold : 51.0 )  11 ( Threshold : 6.7 )  2 ( Threshold : 0.0 )  1 ( Threshold
: 150.0 )
Level  7
Nodes are :
16 ( Threshold : 0.0 )  20 ( Threshold : 6.0 )  10 ( Threshold : 3.0 )  11 ( Threshold : 4.6 )  4 (
Threshold : 0.0 )  0 ( Threshold : 78.0 )  16 ( Threshold : 2.0 )  18 ( Threshold : 116.0 )  13 (
Threshold : 57.0 )  12 ( Threshold : 111.0 )  6 ( Threshold : 0.0 )  11 ( Threshold : 9.4 )  11 (
Threshold : 10.4 )  18 ( Threshold : 147.0 )  11 ( Threshold : 7.1 )
Level  8
Nodes are :
21 ( Threshold : 0.0 )  11 ( Threshold : 6.3 )  9 ( Threshold : 0.4 )  5 ( Threshold : 0.0 )  9 ( Threshold
: 0.3 )  8 ( Threshold : 70.0 )  11 ( Threshold : 6.5 )  0 ( Threshold : 1755.0 )  19 ( Threshold : 119.0 )
7 ( Threshold : 0.001 )  5 ( Threshold : 0.0 )  2 ( Threshold : 0.0 )  3 ( Threshold : 0.0 )  10 (
Threshold : 15.0 )  15 ( Threshold : 5.0 )  6 ( Threshold : 0.0 )  15 ( Threshold : 4.0 )
Level  9
Nodes are :
3 ( Threshold : 0.0 )  1 ( Threshold : 148.0 )  5 ( Threshold : 0.0 )  3 ( Threshold : 0.0 )  12 (
Threshold : 26.0 )  8 ( Threshold : 69.0 )  6 ( Threshold : 0.0 )  3 ( Threshold : 0.0 )  5 ( Threshold :
0.0 )  11 ( Threshold : 14.2 )  21 ( Threshold : 1.0 )  11 ( Threshold : 10.3 )  8 ( Threshold : 15.0 )  9 (
Threshold : 0.6 )  3 ( Threshold : 0.0 )  5 ( Threshold : 0.0 )  18 ( Threshold : 152.0 )  20 ( Threshold :
3.0 )  13 ( Threshold : 81.0 )
```

Visualization for single attribute split for Banking Dataset

```
the tree is splitted according to the following features
Level  0
Nodes are :
11 ( Threshold : 446.0 )
Level  1
Nodes are :
13 ( Threshold : 27.0 )  11 ( Threshold : 644.0 )
Level  2
Nodes are :
11 ( Threshold : 162.0 )  1 ( Threshold : 60.0 )  19 ( Threshold : 1.286 )  16 ( Threshold : -1.1 )
Level  3
Nodes are :
12 ( Threshold : 3.0 )  19 ( Threshold : 1.032 )  10 ( Threshold : 1.0 )  11 ( Threshold : 122.0 )  9 (
Threshold : 5.0 )  3 ( Threshold : 1.0 )  10 ( Threshold : 2.0 )  11 ( Threshold : 870.0 )
```

```
Level  4
Nodes are :
18 ( Threshold : -42.0 )  18 ( Threshold : -49.5 )  4 ( Threshold : 3.0 )  11 ( Threshold : 178.0 )  15 (
Threshold : 0.0 )  0 ( Threshold : 28962.0 )  11 ( Threshold : 238.0 )  19 ( Threshold : 0.851 )  2 (
Threshold : 1.0 )  13 ( Threshold : 10.0 )  0 ( Threshold : 35509.0 )  5 ( Threshold : 0.0 )  5 (
Threshold : 0.0 )  2 ( Threshold : 8.0 )  9 ( Threshold : 4.0 )
Level  5
Nodes are :
16 ( Threshold : -1.1 )  19 ( Threshold : 0.715 )  1 ( Threshold : 65.0 )  11 ( Threshold : 251.0 )  1 (
Threshold : 60.0 )  19 ( Threshold : 1.726 )  16 ( Threshold : -0.1 )  17 ( Threshold : 92.843 )  4 (
Threshold : 2.0 )  19 ( Threshold : 1.799 )  9 ( Threshold : 7.0 )  20 ( Threshold : 5008.7 )  8 (
Threshold : 0.0 )  8 ( Threshold : 0.0 )  10 ( Threshold : 3.0 )  5 ( Threshold : 0.0 )  17 ( Threshold :
92.893 )  12 ( Threshold : 6.0 )  13 ( Threshold : 0.0 )  5 ( Threshold : 0.0 )  10 ( Threshold : 1.0 )  5 (
Threshold : 0.0 )  17 ( Threshold : 93.369 )  11 ( Threshold : 857.0 )  8 ( Threshold : 0.0 )  4 (
Threshold : 6.0 )  2 ( Threshold : 5.0 )  2 ( Threshold : 9.0 )  15 ( Threshold : 1.0 )  10 ( Threshold :
2.0 )
Level  6
Nodes are :
15 ( Threshold : 0.0 )  10 ( Threshold : 1.0 )  16 ( Threshold : -2.9 )  4 ( Threshold : 5.0 )  15 (
Threshold : 0.0 )  11 ( Threshold : 257.0 )  9 ( Threshold : 5.0 )  15 ( Threshold : 0.0 )  5 ( Threshold :
0.0 )  0 ( Threshold : 34559.0 )  8 ( Threshold : 0.0 )  1 ( Threshold : 22.0 )  12 ( Threshold : 1.0 )  4 (
Threshold : 5.0 )  14 ( Threshold : 0.0 )  6 ( Threshold : 1.0 )  10 ( Threshold : 2.0 )  1 ( Threshold :
83.0 )  5 ( Threshold : 0.0 )  20 ( Threshold : 5023.5 )  3 ( Threshold : 2.0 )  11 ( Threshold : 261.0 )
7 ( Threshold : 0.0 )  9 ( Threshold : 1.0 )  5 ( Threshold : 1.0 )  3 ( Threshold : 1.0 )  2 ( Threshold :
0.0 )  0 ( Threshold : 40601.0 )  0 ( Threshold : 7887.0 )  16 ( Threshold : -1.8 )  2 ( Threshold : 5.0 )
16 ( Threshold : 1.1 )  18 ( Threshold : -50.0 )  15 ( Threshold : 1.0 )  16 ( Threshold : -1.8 )  3 (
Threshold : 2.0 )  14 ( Threshold : 2.0 )  5 ( Threshold : 0.0 )  8 ( Threshold : 0.0 )  20 ( Threshold :
5076.2 )  11 ( Threshold : 647.0 )  7 ( Threshold : 1.0 )  2 ( Threshold : 7.0 )  3 ( Threshold : 2.0 )  11
( Threshold : 857.0 )  13 ( Threshold : 999.0 )  4 ( Threshold : 0.0 )  15 ( Threshold : 1.0 )  7 (
Threshold : 1.0 )  12 ( Threshold : 4.0 )
Level  7
Nodes are :
17 ( Threshold : 94.601 )  14 ( Threshold : 2.0 )  8 ( Threshold : 0.0 )  16 ( Threshold : -3.4 )  9 (
Threshold : 1.0 )  14 ( Threshold : 2.0 )  14 ( Threshold : 2.0 )  0 ( Threshold : 8106.0 )  8 ( Threshold
: 0.0 )  18 ( Threshold : -37.5 )  11 ( Threshold : 233.0 )  15 ( Threshold : 0.0 )  12 ( Threshold : 4.0 )
8 ( Threshold : 0.0 )  1 ( Threshold : 42.0 )  9 ( Threshold : 4.0 )  4 ( Threshold : 5.0 )  11 ( Threshold
: 359.0 )  4 ( Threshold : 5.0 )  1 ( Threshold : 27.0 )  1 ( Threshold : 55.0 )  13 ( Threshold : 999.0 )
13 ( Threshold : 999.0 )  13 ( Threshold : 999.0 )  0 ( Threshold : 328.0 )  1 ( Threshold : 24.0 )  2 (
Threshold : 2.0 )  15 ( Threshold : 0.0 )  2 ( Threshold : 6.0 )  2 ( Threshold : 6.0 )  1 ( Threshold :
88.0 )  13 ( Threshold : 999.0 )  16 ( Threshold : -1.8 )  3 ( Threshold : 0.0 )  14 ( Threshold : 0.0 )  0 (
Threshold : 14971.0 )  1 ( Threshold : 74.0 )  13 ( Threshold : 14.0 )  13 ( Threshold : 21.0 )  17 (
Threshold : 92.201 )  3 ( Threshold : 0.0 )  17 ( Threshold : 92.963 )  15 ( Threshold : 0.0 )  11 (
Threshold : 447.0 )  7 ( Threshold : 0.0 )  6 ( Threshold : 0.0 )  15 ( Threshold : 1.0 )  3 ( Threshold :
0.0 )  16 ( Threshold : -0.1 )  9 ( Threshold : 7.0 )  16 ( Threshold : -0.1 )  20 ( Threshold : 5228.1 )
18 ( Threshold : -41.8 )  0 ( Threshold : 23851.0 )  17 ( Threshold : 92.893 )  1 ( Threshold : 29.0 )
18 ( Threshold : -42.7 )  20 ( Threshold : 5195.8 )  20 ( Threshold : 5008.7 )  18 ( Threshold : -39.8 )
```

```
9 ( Threshold : 6.0 )  3 ( Threshold : 1.0 )  10 ( Threshold : 3.0 )  13 ( Threshold : 12.0 )  7 (
Threshold : 2.0 )  1 ( Threshold : 46.0 )  12 ( Threshold : 10.0 )  4 ( Threshold : 6.0 )  2 ( Threshold :
9.0 )  9 ( Threshold : 1.0 )  1 ( Threshold : 32.0 )  12 ( Threshold : 11.0 )  3 ( Threshold : 1.0 )  11 (
Threshold : 886.0 )  3 ( Threshold : 0.0 )
Level  8
Nodes are :
18 ( Threshold : -47.1 )  10 ( Threshold : 1.0 )  5 ( Threshold : 0.0 )  5 ( Threshold : 0.0 )  17 (
Threshold : 93.749 )  6 ( Threshold : 1.0 )  4 ( Threshold : 0.0 )  0 ( Threshold : 29549.0 )  16 (
Threshold : -1.7 )  8 ( Threshold : 0.0 )  6 ( Threshold : 0.0 )  13 ( Threshold : 6.0 )  8 ( Threshold :
0.0 )  5 ( Threshold : 0.0 )  17 ( Threshold : 94.199 )  0 ( Threshold : 203.0 )  3 ( Threshold : 0.0 )  15
( Threshold : 0.0 )  9 ( Threshold : 6.0 )  7 ( Threshold : 0.0 )  2 ( Threshold : 9.0 )  19 ( Threshold :
4.076 )  11 ( Threshold : 77.0 )  20 ( Threshold : 5076.2 )  0 ( Threshold : 34570.0 )  2 ( Threshold :
7.0 )  14 ( Threshold : 1.0 )  5 ( Threshold : 0.0 )  20 ( Threshold : 5023.5 )  1 ( Threshold : 21.0 )  19
( Threshold : 0.663 )  6 ( Threshold : 1.0 )  6 ( Threshold : 1.0 )  0 ( Threshold : 2255.0 )  17 (
Threshold : 92.431 )  20 ( Threshold : 5076.2 )  5 ( Threshold : 0.0 )  13 ( Threshold : 999.0 )  2 (
Threshold : 0.0 )  12 ( Threshold : 1.0 )  17 ( Threshold : 93.798 )  4 ( Threshold : 2.0 )  4 ( Threshold
: 6.0 )  6 ( Threshold : 0.0 )  13 ( Threshold : 999.0 )  20 ( Threshold : 5076.2 )  19 ( Threshold : 1.4 )
19 ( Threshold : 1.354 )  13 ( Threshold : 999.0 )  6 ( Threshold : 0.0 )  4 ( Threshold : 0.0 )  13 (
Threshold : 999.0 )  18 ( Threshold : -39.8 )  14 ( Threshold : 2.0 )  12 ( Threshold : 2.0 )  9 (
Threshold : 3.0 )  17 ( Threshold : 93.749 )  8 ( Threshold : 0.0 )  12 ( Threshold : 2.0 )  10 (
Threshold : 1.0 )  14 ( Threshold : 0.0 )  19 ( Threshold : 0.908 )  5 ( Threshold : 1.0 )  14 ( Threshold
: 1.0 )  2 ( Threshold : 0.0 )  19 ( Threshold : 0.73 )  8 ( Threshold : 0.0 )  2 ( Threshold : 1.0 )  16 (
Threshold : -1.8 )  20 ( Threshold : 5017.5 )  16 ( Threshold : -1.8 )  11 ( Threshold : 592.0 )  13 (
Threshold : 11.0 )  0 ( Threshold : 6491.0 )  19 ( Threshold : 1.372 )  15 ( Threshold : 0.0 )  14 (
Threshold : 0.0 )  6 ( Threshold : 0.0 )  2 ( Threshold : 9.0 )  20 ( Threshold : 5191.0 )  17 ( Threshold
: 93.075 )  2 ( Threshold : 10.0 )  19 ( Threshold : 1.365 )  14 ( Threshold : 0.0 )  9 ( Threshold : 0.0 )
1 ( Threshold : 25.0 )  4 ( Threshold : 2.0 )  2 ( Threshold : 1.0 )  11 ( Threshold : 753.0 )  1 (
Threshold : 41.0 )  19 ( Threshold : 0.898 )  17 ( Threshold : 92.893 )  15 ( Threshold : 0.0 )  6 (
Threshold : 0.0 )  10 ( Threshold : 3.0 )  16 ( Threshold : -3.4 )  2 ( Threshold : 5.0 )  13 ( Threshold :
10.0 )  4 ( Threshold : 2.0 )  19 ( Threshold : 1.291 )  3 ( Threshold : 1.0 )  2 ( Threshold : 1.0 )  11 (
Threshold : 835.0 )  12 ( Threshold : 7.0 )  16 ( Threshold : 1.1 )  20 ( Threshold : 5195.8 )  16 (
Threshold : -0.1 )  8 ( Threshold : 0.0 )  15 ( Threshold : 1.0 )  14 ( Threshold : 0.0 )  14 ( Threshold :
0.0 )  7 ( Threshold : 2.0 )  11 ( Threshold : 1435.0 )  5 ( Threshold : 0.0 )
Level  9
Nodes are :
9 ( Threshold : 6.0 )  1 ( Threshold : 48.0 )  2 ( Threshold : 5.0 )  5 ( Threshold : 0.0 )  4 ( Threshold :
0.0 )  16 ( Threshold : -3.4 )  14 ( Threshold : 1.0 )  9 ( Threshold : 3.0 )  2 ( Threshold : 4.0 )  4 (
Threshold : 5.0 )  16 ( Threshold : -1.1 )  20 ( Threshold : 4963.6 )  6 ( Threshold : 2.0 )  12 (
Threshold : 3.0 )  16 ( Threshold : -2.9 )  15 ( Threshold : 2.0 )  16 ( Threshold : -1.7 )  15 ( Threshold
: 2.0 )  16 ( Threshold : -3.0 )  12 ( Threshold : 1.0 )  17 ( Threshold : 92.469 )  4 ( Threshold : 3.0 )  9
( Threshold : 7.0 )  12 ( Threshold : 1.0 )  20 ( Threshold : 5076.2 )  7 ( Threshold : 0.0 )  19 (
Threshold : 1.266 )  12 ( Threshold : 3.0 )  16 ( Threshold : -1.8 )  1 ( Threshold : 53.0 )  1 (
Threshold : 25.0 )  4 ( Threshold : 3.0 )  4 ( Threshold : 2.0 )  2 ( Threshold : 10.0 )  9 ( Threshold :
1.0 )  12 ( Threshold : 3.0 )  16 ( Threshold : -3.4 )  19 ( Threshold : 0.773 )  1 ( Threshold : 36.0 )  15
( Threshold : 0.0 )  9 ( Threshold : 3.0 )  0 ( Threshold : 40060.0 )  14 ( Threshold : 0.0 )  13 (
Threshold : 999.0 )  6 ( Threshold : 2.0 )  10 ( Threshold : 2.0 )  13 ( Threshold : 999.0 )  2 (
Threshold : 5.0 )  17 ( Threshold : 92.649 )  6 ( Threshold : 0.0 )  4 ( Threshold : 4.0 )  17 ( Threshold
: 92.843 )  13 ( Threshold : 999.0 )  0 ( Threshold : 335.0 )  0 ( Threshold : 4984.0 )  14 ( Threshold :
0.0 )  10 ( Threshold : 3.0 )  3 ( Threshold : 1.0 )  10 ( Threshold : 2.0 )  4 ( Threshold : 1.0 )  6 (
Threshold : 2.0 )  18 ( Threshold : -31.4 )  15 ( Threshold : 0.0 )  5 ( Threshold : 0.0 )  5 ( Threshold :
0.0 )  14 ( Threshold : 1.0 )  3 ( Threshold : 0.0 )  1 ( Threshold : 67.0 )  16 ( Threshold : -2.9 )  16 (
Threshold : -1.8 )  0 ( Threshold : 12121.0 )  0 ( Threshold : 15968.0 )  18 ( Threshold : -49.5 )  1 (
Threshold : 66.0 )  15 ( Threshold : 0.0 )  11 ( Threshold : 501.0 )  8 ( Threshold : 0.0 )  12 (
Threshold : 1.0 )  11 ( Threshold : 578.0 )  12 ( Threshold : 2.0 )  18 ( Threshold : -40.8 )  16 (
Threshold : -2.9 )  8 ( Threshold : 0.0 )  13 ( Threshold : 3.0 )  9 ( Threshold : 6.0 )  7 ( Threshold :
0.0 )  17 ( Threshold : 93.876 )  18 ( Threshold : -49.5 )  1 ( Threshold : 59.0 )  0 ( Threshold : 7405.0
)  1 ( Threshold : 32.0 )  13 ( Threshold : 2.0 )  19 ( Threshold : 1.354 )  5 ( Threshold : 0.0 )  4 (
Threshold : 1.0 )  18 ( Threshold : -36.4 )  14 ( Threshold : 0.0 )  1 ( Threshold : 25.0 )  12 (
Threshold : 3.0 )  17 ( Threshold : 93.075 )  1 ( Threshold : 47.0 )  9 ( Threshold : 3.0 )  2 ( Threshold
: 10.0 )  17 ( Threshold : 93.2 )  11 ( Threshold : 611.0 )  4 ( Threshold : 3.0 )  3 ( Threshold : 2.0 )  14
( Threshold : 0.0 )  17 ( Threshold : 93.444 )  13 ( Threshold : 9.0 )  6 ( Threshold : 0.0 )  7 (
Threshold : 0.0 )  10 ( Threshold : 1.0 )  2 ( Threshold : 0.0 )  7 ( Threshold : 1.0 )  17 ( Threshold :
93.749 )  6 ( Threshold : 1.0 )  9 ( Threshold : 0.0 )  4 ( Threshold : 3.0 )  9 ( Threshold : 6.0 )  2 (
Threshold : 1.0 )  18 ( Threshold : -46.2 )  3 ( Threshold : 1.0 )  2 ( Threshold : 5.0 )  7 ( Threshold :
0.0 )  1 ( Threshold : 31.0 )  10 ( Threshold : 3.0 )  16 ( Threshold : -1.8 )  0 ( Threshold : 28293.0 )  5
( Threshold : 0.0 )  17 ( Threshold : 93.2 )  8 ( Threshold : 0.0 )  3 ( Threshold : 1.0 )  9 ( Threshold :
3.0 )  12 ( Threshold : 1.0 )  17 ( Threshold : 93.918 )  20 ( Threshold : 5228.1 )  6 ( Threshold : 0.0 )
15 ( Threshold : 1.0 )  1 ( Threshold : 58.0 )  5 ( Threshold : 0.0 )  18 ( Threshold : -40.4 )  15 (
Threshold : 0.0 )  12 ( Threshold : 3.0 )  17 ( Threshold : 93.2 )  12 ( Threshold : 3.0 )  17 ( Threshold
: 93.2 )  12 ( Threshold : 3.0 )
```

Visualization for mutli attribute split for Fetal Health Dataset

```
the tree is splitted according to the following features
Level  0
Nodes are :
[10, 18] ( Threshold : [6.0, 107.0] )
Level  1
Nodes are :
[6, 8] ( Threshold : [0.0, 24.0] )  [10, 8] ( Threshold : [6.0, 59.0] )
Level  2
Nodes are :
[0, 11] ( Threshold : [913.0, 0.0] )  [9, 14] ( Threshold : [3.3, 211.0] )  [13, 10] ( Threshold : [134.0,
7.0] )  [9, 17] ( Threshold : [0.5, 139.0] )
Level  3
Nodes are :
[19, 0] ( Threshold : [91.0, 1177.0] )  [10, 13] ( Threshold : [0.0, 50.0] )  [19, 11] ( Threshold : [128.0,
3.2] )  [18, 15] ( Threshold : [147.0, 7.0] )  [9, 11] ( Threshold : [0.5, 4.8] )  [10, 11] ( Threshold : [6.0,
2.7] )
Level  4
Nodes are :
```

```
[16, 7] ( Threshold : [0.0, 0.001] )  [1, 17] ( Threshold : [143.0, 99.0] )  [1, 19] ( Threshold : [137.0,
140.0] )  [20, 19] ( Threshold : [32.0, 147.0] )  [14, 8] ( Threshold : [136.0, 72.0] )  [8, 9] ( Threshold :
[79.0, 0.5] )  [1, 10] ( Threshold : [143.0, 6.0] )
Level  5
Nodes are :
[2, 21] ( Threshold : [0.0, 0.0] )  [2, 7] ( Threshold : [0.001, 0.002] )  [9, 12] ( Threshold : [1.2, 91.0] )
[19, 9] ( Threshold : [152.0, 1.9] )  [18, 8] ( Threshold : [147.0, 47.0] )  [2, 10] ( Threshold : [0.0, 32.0]
)  [12, 19] ( Threshold : [65.0, 157.0] )  [10, 20] ( Threshold : [67.0, 44.0] )  [20, 14] ( Threshold : [0.0,
139.0] )  [8, 1] ( Threshold : [79.0, 114.0] )  [16, 1] ( Threshold : [0.0, 147.0] )  [8, 10] ( Threshold :
[78.0, 68.0] )
Level  6
Nodes are :
[11, 19] ( Threshold : [0.7, 123.0] )  [15, 18] ( Threshold : [6.0, 115.0] )  [6, 18] ( Threshold : [0.0,
110.0] )  [17, 15] ( Threshold : [146.0, 13.0] )  [0, 6] ( Threshold : [38.0, 0.0] )  [13, 14] ( Threshold :
[77.0, 184.0] )  [15, 20] ( Threshold : [9.0, 32.0] )  [1, 15] ( Threshold : [138.0, 8.0] )  [18, 12] (
Threshold : [129.0, 10.0] )  [5, 11] ( Threshold : [0.003, 0.0] )  [6, 16] ( Threshold : [0.0, 0.0] )  [11, 8] (
Threshold : [2.5, 79.0] )  [18, 12] ( Threshold : [125.0, 25.0] )  [13, 21] ( Threshold : [134.0, 0.0] )  [13,
11] ( Threshold : [138.0, 10.2] )  [5, 16] ( Threshold : [0.0, 0.0] )
Level  7
Nodes are :
[20, 7] ( Threshold : [15.0, 0.0] )  [6, 18] ( Threshold : [0.0, 108.0] )  [0, 13] ( Threshold : [1098.0,
54.0] )  [7, 12] ( Threshold : [0.0, 8.0] )  [17, 18] ( Threshold : [146.0, 159.0] )  [2, 1] ( Threshold : [0.0,
123.0] )  [5, 20] ( Threshold : [0.003, 71.0] )  [0, 20] ( Threshold : [84.0, 23.0] )  [16, 15] ( Threshold :
[0.0, 1.0] )  [18, 8] ( Threshold : [124.0, 79.0] )  [1, 17] ( Threshold : [131.0, 123.0] )  [18, 19] (
Threshold : [147.0, 153.0] )  [7, 4] ( Threshold : [0.0, 0.003] )  [15, 12] ( Threshold : [2.0, 19.0] )  [10,
11] ( Threshold : [35.0, 5.0] )
Level  8
Nodes are :
[9, 13] ( Threshold : [0.9, 134.0] )  [4, 12] ( Threshold : [0.007, 123.0] )  [4, 7] ( Threshold : [0.001,
0.0] )  [5, 17] ( Threshold : [0.003, 129.0] )  [19, 14] ( Threshold : [161.0, 168.0] )  [16, 17] ( Threshold
: [1.0, 150.0] )  [10, 9] ( Threshold : [2.0, 1.6] )  [15, 10] ( Threshold : [4.0, 35.0] )  [19, 17] ( Threshold
: [123.0, 122.0] )  [16, 8] ( Threshold : [0.0, 61.0] )  [14, 13] ( Threshold : [159.0, 120.0] )  [16, 1] (
Threshold : [0.0, 137.0] )  [1, 8] ( Threshold : [143.0, 79.0] )  [21, 1] ( Threshold : [0.0, 142.0] )
Level  9
Nodes are :
[8, 14] ( Threshold : [29.0, 168.0] )  [10, 5] ( Threshold : [0.0, 0.0] )  [9, 6] ( Threshold : [1.2, 0.0] )  [6,
9] ( Threshold : [0.0, 1.7] )  [17, 0] ( Threshold : [146.0, 487.0] )  [21, 9] ( Threshold : [-1.0, 2.3] )  [7,
13] ( Threshold : [0.002, 54.0] )  [21, 14] ( Threshold : [0.0, 154.0] )  [15, 17] ( Threshold : [5.0, 162.0]
)  [14, 12] ( Threshold : [154.0, 43.0] )  [13, 18] ( Threshold : [120.0, 153.0] )  [7, 5] ( Threshold :
[0.001, 0.001] )  [16, 15] ( Threshold : [0.0, 2.0] )  [19, 13] ( Threshold : [123.0, 84.0] )  [21, 20] (
Threshold : [1.0, 3.0] )  [1, 14] ( Threshold : [136.0, 155.0] )  [8, 2] ( Threshold : [79.0, 0.0] )  [13, 19] (
Threshold : [72.0, 138.0] )  [18, 1] ( Threshold : [145.0, 140.0] )
```

Visualization for Single Attribute Split for Cervical Cancer Dataset

```
the tree is splitted according to the following features
Level  0
Nodes are :
31 ( Threshold : 0.0 )
Level  1
Nodes are :
25 ( Threshold : 0.0 )  6 ( Threshold : 7.0 )
Level  2
Nodes are :
28 ( Threshold : 0.0 )  12 ( Threshold : 0.0 )
Level  3
Nodes are :
16 ( Threshold : 0.0 )  12 ( Threshold : 0.0 )  31 ( Threshold : 1.0 )  12 ( Threshold : 1.0 )
Level  4
Nodes are :
23 ( Threshold : 0.0 )  32 ( Threshold : 0.0 )  19 ( Threshold : 0.0 )  8 ( Threshold : 0.0 )
Level  5
Nodes are :
26 ( Threshold : 0.0 )  29 ( Threshold : 0.0 )  33 ( Threshold : 0.0 )  8 ( Threshold : 0.0 )  3 (
Threshold : 15.0 )
Level  6
Nodes are :
27 ( Threshold : 0.0 )  3 ( Threshold : 17.0 )  10 ( Threshold : 0.0 )  12 ( Threshold : 0.0 )  24 (
Threshold : 0.0 )  6 ( Threshold : 0.0 )
Level  7
Nodes are :
17 ( Threshold : 0.0 )  24 ( Threshold : 0.0 )  2 ( Threshold : 2.0 )  14 ( Threshold : 0.0 )  33 (
Threshold : 0.0 )  1 ( Threshold : 17.0 )  0 ( Threshold : 152.0 )
Level  8
Nodes are :
31 ( Threshold : 0.0 )  28 ( Threshold : 0.0 )  22 ( Threshold : 0.0 )  17 ( Threshold : 0.0 )  28 (
Threshold : 0.0 )  30 ( Threshold : 1.0 )  25 ( Threshold : 0.0 )  11 ( Threshold : 0.0 )  22 ( Threshold :
0.0 )
Level  9
Nodes are :
31 ( Threshold : 0.0 )  26 ( Threshold : 0.0 )  20 ( Threshold : 0.0 )  19 ( Threshold : 0.0 )  11 (
Threshold : 0.0 )  28 ( Threshold : 1.0 )  7 ( Threshold : 0.0 )  31 ( Threshold : 1.0 )  7 ( Threshold :
0.0 )
```

Visualization for Mulit Attribute Split for Cervical Cancer Dataset

```
the tree is splitted according to the following features
Level  0
Nodes are :
```

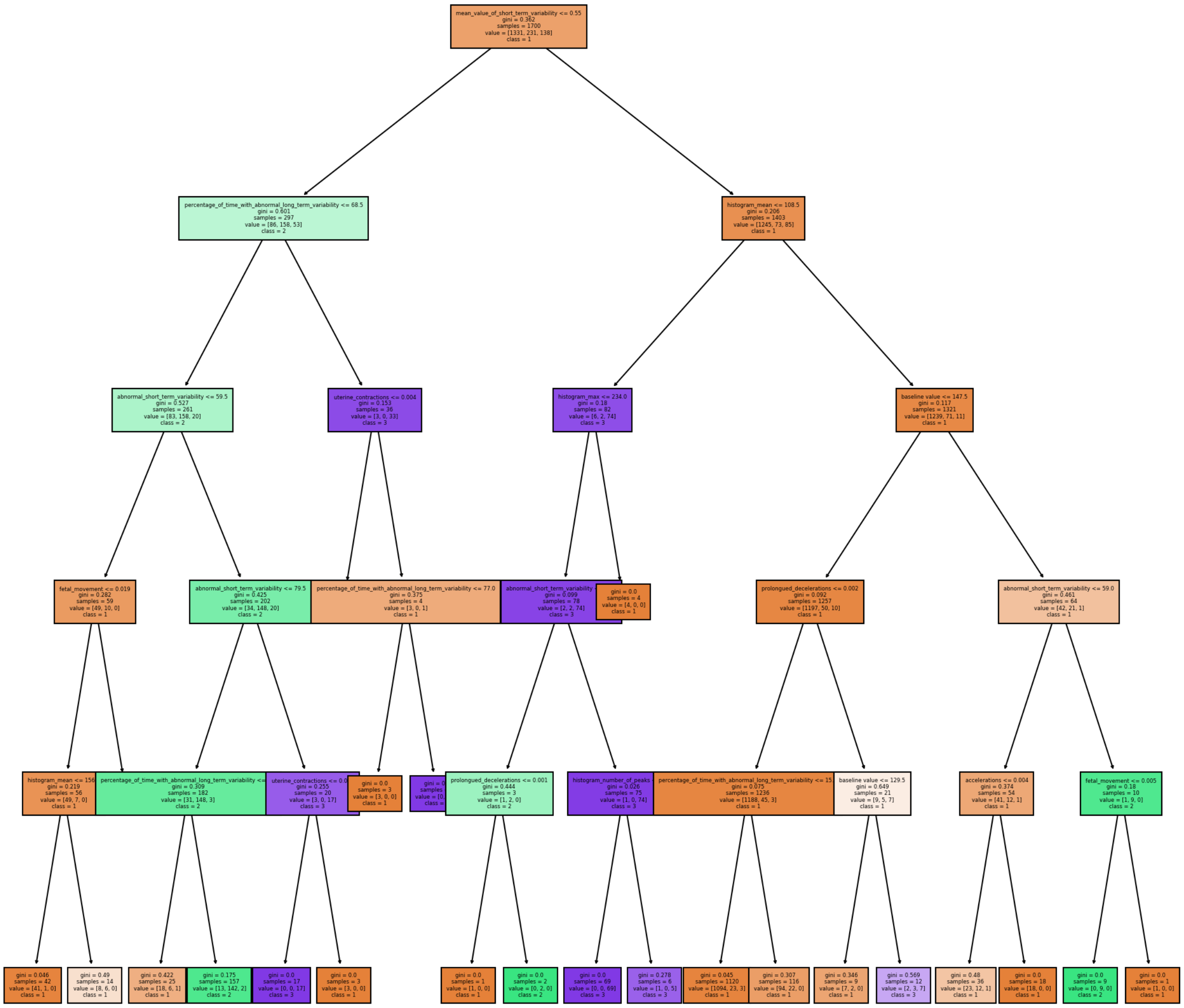
```
[17, 31] ( Threshold : [0.0, 0.0] )
Level  1
Nodes are :
[30, 32] ( Threshold : [0.0, 0.0] )  [17, 11] ( Threshold : [0.0, 0.0] )
Level  2
Nodes are :
[21, 12] ( Threshold : [0.0, 0.0] )  [29, 6] ( Threshold : [0.0, 14.0] )  [1, 4] ( Threshold : [26.0, 3.0] )
Level  3
Nodes are :
[31, 3] ( Threshold : [0.0, 10.0] )  [27, 3] ( Threshold : [0.0, 18.0] )  [10, 4] ( Threshold : [0.0, 5.0] )
[20, 33] ( Threshold : [0.0, 0.0] )
Level  4
Nodes are :
[32, 20] ( Threshold : [0.0, 0.0] )  [0, 4] ( Threshold : [325.0, 1.0] )  [0, 3] ( Threshold : [103.0, 14.0] )
[17, 23] ( Threshold : [0.0, 0.0] )  [1, 2] ( Threshold : [17.0, 3.0] )
Level  5
Nodes are :
[3, 15] ( Threshold : [11.0, 0.0] )  [27, 33] ( Threshold : [0.0, 0.0] )  [33, 10] ( Threshold : [0.0, 0.0] )
[25, 7] ( Threshold : [0.0, 0.0] )  [33, 16] ( Threshold : [0.0, 0.0] )  [15, 24] ( Threshold : [0.0, 0.0] )
Level  6
Nodes are :
[25, 28] ( Threshold : [0.0, 0.0] )  [10, 12] ( Threshold : [0.0, 1.0] )  [20, 26] ( Threshold : [0.0, 0.0] )
[27, 4] ( Threshold : [0.0, 3.0] )  [6, 14] ( Threshold : [0.0, 0.0] )  [9, 12] ( Threshold : [0.0, 1.0] )  [4,
26] ( Threshold : [1.0, 0.0] )
Level  7
Nodes are :
[27, 17] ( Threshold : [0.0, 0.0] )  [33, 29] ( Threshold : [0.0, 0.0] )  [8, 32] ( Threshold : [0.0, 0.0] )
[21, 0] ( Threshold : [0.0, 44.0] )  [19, 15] ( Threshold : [0.0, 0.0] )  [10, 18] ( Threshold : [1.0, 0.0] )
[31, 32] ( Threshold : [1.0, 1.0] )  [33, 2] ( Threshold : [0.0, 2.0] )  [2, 9] ( Threshold : [1.0, 1.0] )
Level  8
Nodes are :
[12, 33] ( Threshold : [0.0, 0.0] )  [14, 12] ( Threshold : [0.0, 0.0] )  [12, 9] ( Threshold : [1.0, 0.0] )  [0,
18] ( Threshold : [64.0, 0.0] )  [8, 5] ( Threshold : [0.0, 0.0] )  [13, 11] ( Threshold : [0.0, 1.0] )  [10, 26]
( Threshold : [0.0, 0.0] )  [22, 9] ( Threshold : [0.0, 0.0] )  [27, 18] ( Threshold : [0.0, 0.0] )
Level  9
Nodes are :
[23, 17] ( Threshold : [0.0, 0.0] )  [14, 15] ( Threshold : [0.0, 0.0] )  [14, 30] ( Threshold : [0.0, 1.0] )
[24, 3] ( Threshold : [0.0, 19.0] )  [8, 10] ( Threshold : [0.0, 0.0] )  [17, 2] ( Threshold : [0.0, 3.0] )  [25,
28] ( Threshold : [0.0, 0.0] )  [20, 32] ( Threshold : [0.0, 1.0] )  [33, 16] ( Threshold : [0.0, 0.0] )  [19,
22] ( Threshold : [0.0, 0.0] )
```

Visualization for multi attribute split for Banking Dataset

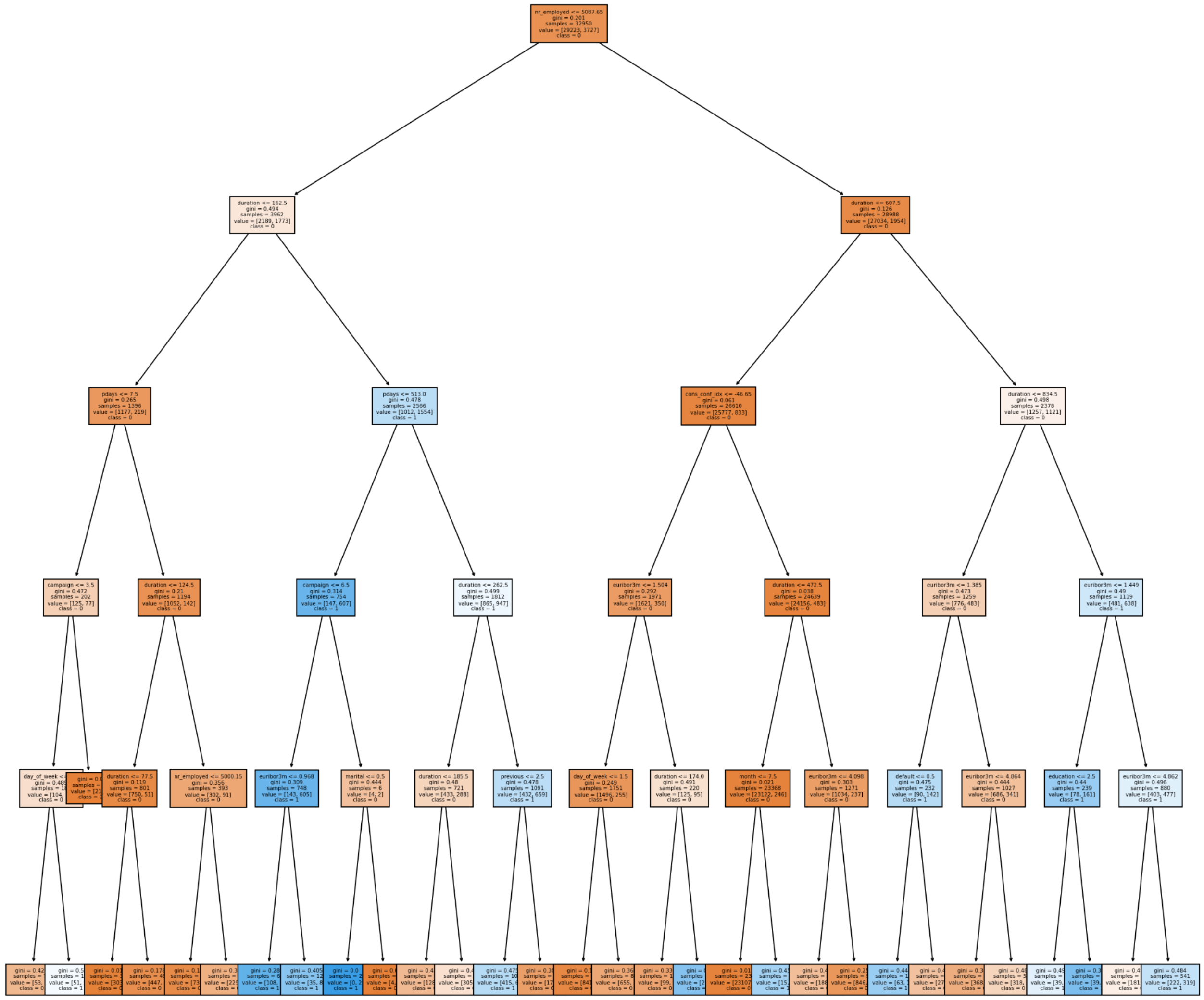
```
the tree is splitted according to the following features
Level  0
Nodes are :
```

```
[21, 31] ( Threshold : [0.0, 0.0] )
Level  1
Nodes are :
[32, 30] ( Threshold : [0.0, 0.0] )  [7, 25] ( Threshold : [0.0, 0.0] )
Level  2
Nodes are :
[33, 32] ( Threshold : [0.0, 0.0] )  [28, 27] ( Threshold : [0.0, 0.0] )  [5, 13] ( Threshold : [0.0, 2.0] )
Level  3
Nodes are :
[1, 14] ( Threshold : [20.0, 0.0] )  [21, 6] ( Threshold : [0.0, 0.0] )  [18, 13] ( Threshold : [0.0, 0.0] )  [6,
19] ( Threshold : [0.0, 0.0] )  [13, 8] ( Threshold : [0.0, 0.0] )
Level  4
Nodes are :
[2, 16] ( Threshold : [1.0, 0.0] )  [26, 3] ( Threshold : [0.0, 18.0] )  [26, 9] ( Threshold : [0.0, 0.0] )  [33,
7] ( Threshold : [0.0, 0.0] )  [26, 31] ( Threshold : [0.0, 0.0] )  [33, 6] ( Threshold : [0.0, 0.0] )  [25, 13]
( Threshold : [0.0, 0.0] )
Level  5
Nodes are :
[17, 24] ( Threshold : [0.0, 0.0] )  [13, 4] ( Threshold : [0.0, 3.0] )  [20, 22] ( Threshold : [0.0, 0.0] )
[21, 9] ( Threshold : [0.0, 1.0] )  [17, 6] ( Threshold : [0.0, 0.0] )  [12, 1] ( Threshold : [0.0, 29.0] )  [24,
18] ( Threshold : [0.0, 0.0] )  [21, 17] ( Threshold : [0.0, 0.0] )
Level  6
Nodes are :
[0, 33] ( Threshold : [162.0, 0.0] )  [6, 3] ( Threshold : [0.0, 14.0] )  [24, 3] ( Threshold : [0.0, 15.0] )
[24, 3] ( Threshold : [0.0, 15.0] )  [25, 29] ( Threshold : [0.0, 0.0] )  [30, 31] ( Threshold : [1.0, 0.0] )
[9, 20] ( Threshold : [0.0, 0.0] )  [25, 12] ( Threshold : [0.0, 0.0] )  [20, 3] ( Threshold : [0.0, 11.0] )
Level  7
Nodes are :
[16, 25] ( Threshold : [0.0, 0.0] )  [25, 1] ( Threshold : [0.0, 15.0] )  [33, 18] ( Threshold : [0.0, 0.0] )
[28, 2] ( Threshold : [0.0, 1.0] )  [30, 20] ( Threshold : [0.0, 0.0] )  [28, 13] ( Threshold : [0.0, 0.0] )
[25, 31] ( Threshold : [0.0, 0.0] )  [16, 10] ( Threshold : [0.0, 0.0] )  [27, 13] ( Threshold : [0.0, 0.0] )
[5, 6] ( Threshold : [0.0, 7.0] )
Level  8
Nodes are :
[0, 20] ( Threshold : [162.0, 0.0] )  [8, 12] ( Threshold : [0.0, 0.0] )  [22, 17] ( Threshold : [0.0, 0.0] )
[16, 10] ( Threshold : [0.0, 0.0] )  [12, 2] ( Threshold : [0.0, 2.0] )  [0, 16] ( Threshold : [109.0, 0.0] )
[28, 9] ( Threshold : [0.0, 0.0] )  [13, 17] ( Threshold : [0.0, 0.0] )  [16, 9] ( Threshold : [0.0, 0.08] )
Level  9
Nodes are :
[5, 1] ( Threshold : [0.0, 28.0] )  [17, 8] ( Threshold : [0.0, 0.0] )  [10, 32] ( Threshold : [0.0, 1.0] )  [7,
5] ( Threshold : [0.0, 0.0] )  [18, 25] ( Threshold : [0.0, 0.0] )  [11, 6] ( Threshold : [0.0, 0.0] )  [2, 0] (
Threshold : [2.0, 152.0] )
```

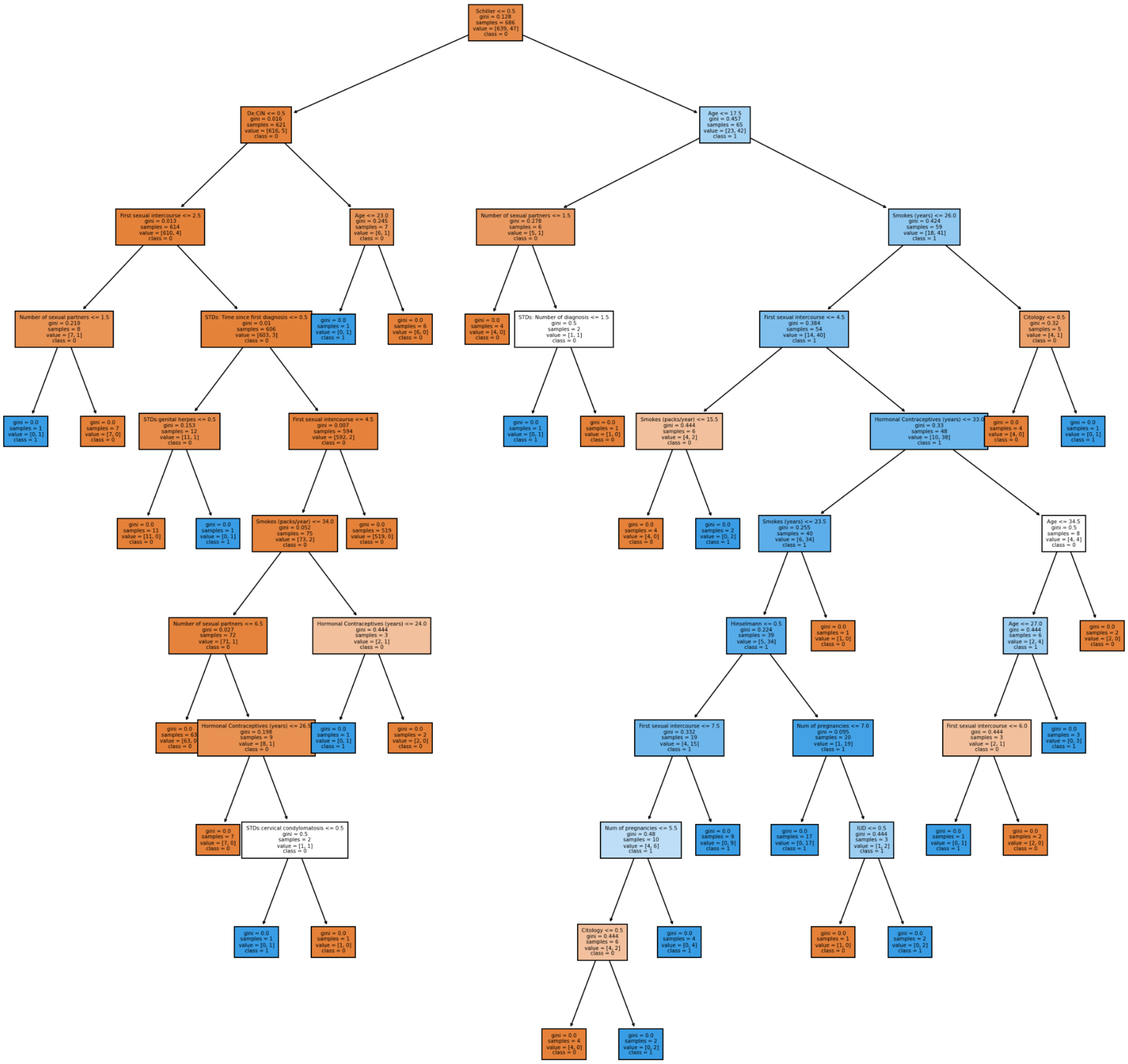
Visualization of normal decision tree for Fetal_Health Dataset



Visualization of normal decision tree for Banking Dataset


nr_employed <= 5087.65 gini = 0.201 samples = 32950 value = [29223, 3727] class = 0
duration <= 162.5 gini = 0.494 samples = 3962 value = [2189, 1773] class = 0
duration <= 607.5 gini = 0.126 samples = 28988 value = [27034, 1954] class = 0
pdays <= 7.5 gini = 0.265 samples = 1396 value = [1177, 219] class = 0
pdays <= 513.0 gini = 0.478 samples = 2566 value = [1012, 1554] class = 1
cons_conf_idx <= -46.65 gini = 0.061 samples = 26610 value = [25777, 833] class = 0
duration <= 834.5 gini = 0.498 samples = 2378 value = [1257, 1121] class = 0
campaign <= 3.5 gini = 0.472 samples = 202 value = [125, 77] class = 0
duration <= 124.5 gini = 0.21 samples = 1194 value = [1052, 142] class = 0
campaign <= 6.5 gini = 0.314 samples = 754 value = [147, 607] class = 1
duration <= 262.5 gini = 0.499 samples = 1812 value = [865, 947] class = 1
euribor3m <= 1.504 gini = 0.292 samples = 1971 value = [1621, 350] class = 0
duration <= 472.5 gini = 0.038 samples = 24639 value = [24156, 483] class = 0
euribor3m <= 1.385 gini = 0.473 samples = 1259 value = [776, 483] class = 0
euribor3m <= 1.449 gini = 0.49 samples = 1119 value = [481, 638] class = 1
day_of_week <= gini = 0.489 samples = 18 value = [104, class = 0
gini = 0.0 samples = value = [21 class = 0
duration <= 77.5 gini = 0.119 samples = 801 value = [750, 51] class = 0
nr_employed <= 5000.15 gini = 0.356 samples = 393 value = [302, 91] class = 0
euribor3m <= 0.968 gini = 0.309 samples = 748 value = [143, 605] class = 1
marital <= 0.5 gini = 0.444 samples = 6 value = [4, 2] class = 0
duration <= 185.5 gini = 0.48 samples = 721 value = [433, 288] class = 0
previous <= 2.5 gini = 0.478 samples = 1091 value = [432, 659] class = 1
day_of_week <= 1.5 gini = 0.249 samples = 1751 value = [1496, 255] class = 0
duration <= 174.0 gini = 0.491 samples = 220 value = [125, 95] class = 0
month <= 7.5 gini = 0.021 samples = 23368 value = [23122, 246] class = 0
euribor3m <= 4.098 gini = 0.303 samples = 1271 value = [1034, 237] class = 0
default <= 0.5 gini = 0.475 samples = 232 value = [90, 142] class = 1
euribor3m <= 4.864 gini = 0.444 samples = 1027 value = [686, 341] class = 0
education <= 2.5 gini = 0.44 samples = 239 value = [78, 161] class = 1
euribor3m <= 4.862 gini = 0.496 samples = 880 value = [403, 477] class = 1
gini = 0.42 samples = value = [53, class = 0
gini = 0.5 samples = 1 value = [51, class = 1
gini = 0.01 samples = 3 value = [303 class = 0
gini = 0.178 samples = 49 value = [447, class = 0
gini = 0.1 samples = value = [73 class = 0
gini = 0.3 samples = value = [229 class =
gini = 0.28 samples = 6 value = [108, class = 1
gini = 0.405 samples = 12 value = [35, 8 class = 1
gini = 0.0 samples = 2 value = [0, 2 class = 1
gini = 0.0 samples = value = [4, class = 0
gini = 0.4 samples = value = [128 class =
gini = 0.4 samples = value = [305 class =
gini = 0.475 samples = 10 value = [415, 6 class = 1
gini = 0.30 samples = value = [17 class = 0
gini = 0.1 samples = value = [841 class =
gini = 0.36 samples = 8 value = [655, class = 0
gini = 0.33 samples = 1 value = [99, class = 0
gini = 0 samples value = [2 class =
gini = 0.01 samples = 23 value = [23107 class = 0
gini = 0.45 samples = value = [15, class = 1
gini = 0.4 samples = value = [188 class =
gini = 0.25 samples = 9 value = [846, class = 0
gini = 0.44 samples = 1 value = [63, class = 1
gini = 0.4 samples = value = [27 class =
gini = 0.3 samples = value = [368 class =
gini = 0.48 samples = 5 value = [318, class = 0
gini = 0.49 samples = value = [39, class = 1
gini = 0.3 samples = value = [39, class =
gini = 0.4 samples = value = [181 class =
gini = 0.484 samples = 541 value = [222, 319] class = 1

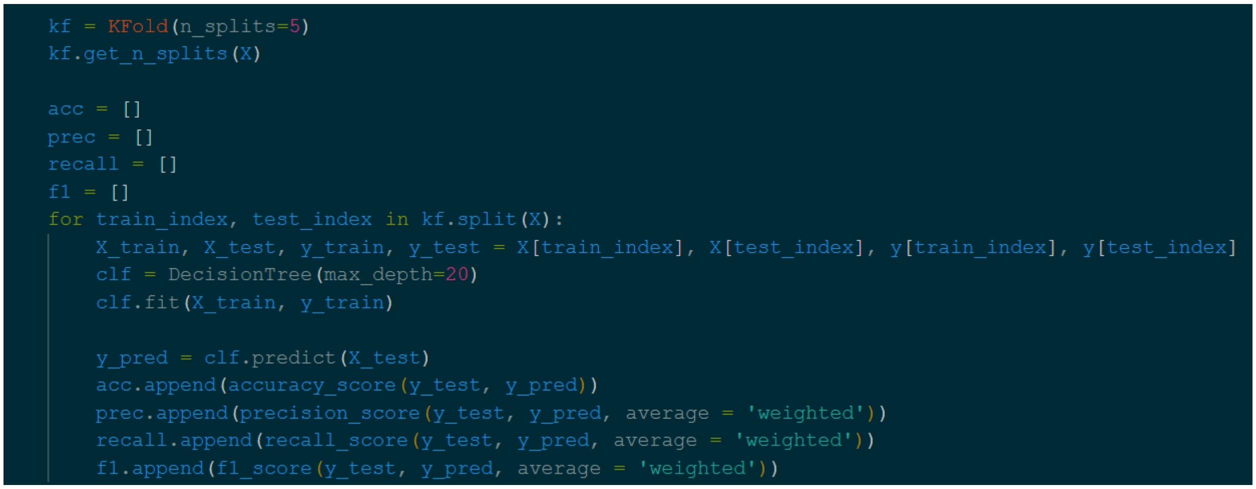
Visualization of normal decision tree for Cervical cancer Dataset


Schiller <= 0.5 gini = 0.128 samples = 686 value = [639, 47] class = 0
Dx:CIN <= 0.5 gini = 0.016 samples = 621 value = [616, 5] class = 0
Age <= 17.5 gini = 0.457 samples = 65 value = [23, 42] class = 1
First sexual intercourse <= 2.5 gini = 0.013 samples = 614 value = [610, 4] class = 0
Age <= 23.0 gini = 0.245 samples = 7 value = [6, 1] class = 0
Number of sexual partners <= 1.5 gini = 0.278 samples = 6 value = [5, 1] class = 0
Smokes (years) <= 26.0 gini = 0.424 samples = 59 value = [18, 41] class = 1
Number of sexual partners <= 1.5 gini = 0.219 samples = 8 value = [7, 1] class = 0
STDs: Time since first diagnosis <= 0.5 gini = 0.01 samples = 606 value = [603, 3] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 6 value = [6, 0] class = 0
gini = 0.0 samples = 4 value = [4, 0] class = 0
STDs: Number of diagnosis <= 1.5 gini = 0.5 samples = 2 value = [1, 1] class = 0
First sexual intercourse <= 4.5 gini = 0.384 samples = 54 value = [14, 40] class = 1
Citology <= 0.5 gini = 0.32 samples = 5 value = [4, 1] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 7 value = [7, 0] class = 0
STDs:genital herpes <= 0.5 gini = 0.153 samples = 12 value = [11, 1] class = 0
First sexual intercourse <= 4.5 gini = 0.007 samples = 594 value = [592, 2] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
Smokes (packs/year) <= 15.5 gini = 0.444 samples = 6 value = [4, 2] class = 0
Hormonal Contraceptives (years) <= 33.0 gini = 0.33 samples = 48 value = [10, 38] class = 1
gini = 0.0 samples = 4 value = [4, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 11 value = [11, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
Smokes (packs/year) <= 34.0 gini = 0.052 samples = 75 value = [73, 2] class = 0
gini = 0.0 samples = 519 value = [519, 0] class = 0
gini = 0.0 samples = 4 value = [4, 0] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
Smokes (years) <= 23.5 gini = 0.255 samples = 40 value = [6, 34] class = 1
Age <= 34.5 gini = 0.5 samples = 8 value = [4, 4] class = 0
Number of sexual partners <= 6.5 gini = 0.027 samples = 72 value = [71, 1] class = 0
Hormonal Contraceptives (years) <= 24.0 gini = 0.444 samples = 3 value = [2, 1] class = 0
Hinselmann <= 0.5 gini = 0.224 samples = 39 value = [5, 34] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
Age <= 27.0 gini = 0.444 samples = 6 value = [2, 4] class = 1
gini = 0.0 samples = 2 value = [2, 0] class = 0
gini = 0.0 samples = 63 value = [63, 0 class = 0
Hormonal Contraceptives (years) <= 26.5 gini = 0.198 samples = 9 value = [8, 1] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 2 value = [2, 0] class = 0
First sexual intercourse <= 7.5 gini = 0.332 samples = 19 value = [4, 15] class = 1
Num of pregnancies <= 7.0 gini = 0.095 samples = 20 value = [1, 19] class = 1
First sexual intercourse <= 6.0 gini = 0.444 samples = 3 value = [2, 1] class = 0
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.0 samples = 7 value = [7, 0] class = 0
STDs:cervical condylomatosis <= 0.5 gini = 0.5 samples = 2 value = [1, 1] class = 0
Num of pregnancies <= 5.5 gini = 0.48 samples = 10 value = [4, 6] class = 1
gini = 0.0 samples = 9 value = [0, 9] class = 1
gini = 0.0 samples = 17 value = [0, 17] class = 1
IUD <= 0.5 gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 2 value = [2, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
Citology <= 0.5 gini = 0.444 samples = 6 value = [4, 2] class = 0
gini = 0.0 samples = 4 value = [0, 4] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.0 samples = 4 value = [4, 0] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1

We see that the results are pretty similar, to each other however, the depths of both trees are varying. From our visualization, we can see that the tree created through logistic regression is more dense and has better splitting thresholds. Also, between one attribute and 2 attributes, we can clearly see that 2 attribute splitting generates a denser tree overall.

Ques 3)

K-Fold Cross Validation
Code for K-Fold Cross Validation

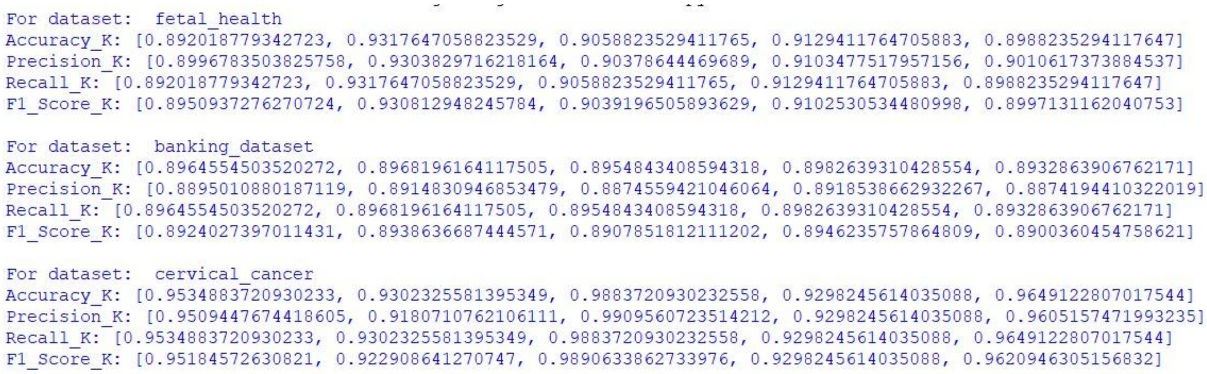


```
kf = KFold(n_splits=5)
kf.get_n_splits(X)

acc = []
prec = []
recall = []
f1 = []
for train_index, test_index in kf.split(X):
    X_train, X_test, y_train, y_test = X[train_index], X[test_index], y[train_index], y[test_index]
    clf = DecisionTree(max_depth=20)
    clf.fit(X_train, y_train)

    y_pred = clf.predict(X_test)
    acc.append(accuracy_score(y_test, y_pred))
    prec.append(precision_score(y_test, y_pred, average = 'weighted'))
    recall.append(recall_score(y_test, y_pred, average = 'weighted'))
    f1.append(f1_score(y_test, y_pred, average = 'weighted'))
```

Results



```
For dataset:  fetal_health
Accuracy_K: [0.892018779342723, 0.9317647058823529, 0.9058823529411765, 0.9129411764705883, 0.8988235294117647]
Precision_K: [0.8996783503825758, 0.9303829716218164, 0.90378644469689, 0.9103477517957156, 0.9010617373884537]
Recall_K: [0.892018779342723, 0.9317647058823529, 0.9058823529411765, 0.9129411764705883, 0.8988235294117647]
F1_Score_K: [0.8950937276270724, 0.930812948245784, 0.9039196505893629, 0.9102530534480998, 0.8997131162040753]

For dataset:  banking_dataset
Accuracy_K: [0.8964554503520272, 0.8968196164117505, 0.8954843408594318, 0.8982639310428554, 0.8932863906762171]
Precision_K: [0.8895010880187119, 0.8914830946853479, 0.8874559421046064, 0.8918538662932267, 0.8874194410322019]
Recall_K: [0.8964554503520272, 0.8968196164117505, 0.8954843408594318, 0.8982639310428554, 0.8932863906762171]
F1_Score_K: [0.8924027397011431, 0.8938636687444571, 0.8907851812111202, 0.8946235757864809, 0.8900360454758621]

For dataset:  cervical_cancer
Accuracy_K: [0.9534883720930233, 0.9302325581395349, 0.9883720930232558, 0.9298245614035088, 0.9649122807017544]
Precision_K: [0.9509447674418605, 0.9180710762106111, 0.9909560723514212, 0.9298245614035088, 0.9605157471993235]
Recall_K: [0.9534883720930233, 0.9302325581395349, 0.9883720930232558, 0.9298245614035088, 0.9649122807017544]
F1_Score_K: [0.95184572630821, 0.922908641270747, 0.9890633862733976, 0.9298245614035088, 0.9620946305156832]
```

Hypothesis Testing :

1. **Student t-test**
2. **Oneway ANOVA test**
3. **Kruskal-Wallis H test**

For Fetal Health Dataset

t_ttest_ind: 35.815812899261174 **p_ttest_ind:** 3.7202295686941064e-10

t_f_oneway: 128.27724536348759 **p_f_oneway:** 2.9565642521973021e-15

t_kruskal: 13.12403312585693   **p_kruskal:** 1.788883827645674e-21

For Cervical Cancer Dataset
t_ttest_ind: -0.6460747096989726   **p_ttest_ind:** 2.6422449161585092e-18
t_f_oneway: 4.1741253051261196   **p_f_oneway:** 4.518664249618514e-16
t_kruskal: 41.812271434343534   **p_kruskal:** 1.8752412627101588e-14

For Banking Dataset

t_ttest_ind: 1.4818724459095483   **p_ttest_ind:** 0.053874485150239152
t_f_oneway: 2.1959459459459465   **p_f_oneway:** 0.0545481513502957
t_kruskal: 1.7347239029613857   **p_kruskal:** 0.051878087340702118

Analysis:

1. Most researchers use alpha = 0.05 as truth value so we are also choosing alpha as 0.05. So, if p_value is less than alpha than we can reject null hypothesis and the difference between accuracies of model is justified otherwise model is not performing well according to statistics.
2. In Fetal and Cervical dataset p_value is less than 0.05 so we can conclusively reject ht null hypothesis and accept our model. And the result we are getting matches the real expectations. And the accuracy is coming out to be better in pair of attribute case which is expected too as we are increasing number of features.
3. But in case of Banking dataset p_value is greater than 0.05 so we cannot reject the null hypothesis in this case and we can see that there is not much improvement in the result of accuracy and f1_score is also less for pair of attribute vs single attribute which is not intuitive too.