Assignment: 2

Roll No: 170001016

Name : B Venkata Pavan

1) TheDecisionTree:

ID3 Algorithm (Entropy and Information Gain)

```
variance <= 0.32016500000000003
    skewness <= 5.86535
        curtosis <= 3.0642 : 1
        curtosis > 3.0642
            skewness <= -1.81995
                variance <= -0.651195 : 1
                variance > -0.651195
                    curtosis <= 7.531300000000001 : 1
                    curtosis > 7.531300000000001 : 0
            skewness > -1.81995
                variance <= -2.185
                    variance <= -2.69165 : 0
                    variance > -2.69165 : 1
                variance > -2.185 : 0
    skewness > 5.86535
        variance <= -3.4448999999999996 : 1
        variance > -3.4448999999999996 : 0
 variance > 0.32016500000000003
    variance <= 1.7907000000000002
        curtosis <= -2.2721999999999998
            skewness <= 6.41995 : 1
            skewness > 6.41995 : 0
        curtosis > -2.2721999999999998
            entropy <= 0.09222649999999999
                variance <= 0.450925
                    variance <= 0.400865 : 0
                    variance > 0.400865 : 1
                variance > 0.450925 : 0
            entropy > 0.09222649999999999
                curtosis <= 0.866835 : 1
                curtosis > 0.866835 : 0
    variance > 1.7907000000000002
        curtosis <= -4.95885 : 1
        curtosis > -4.95885
            variance <= 2.03655
                variance <= 2.02515 : 0
                variance > 2.02515 : 1
            variance > 2.03655 : 0
```

CART Algorithm (Gini Index)

```
variance <= 0.32016500000000003
    skewness <= 5.86535
        curtosis <= 3.0642 : 1
        curtosis > 3.0642
            skewness <= -1.81995
                variance <= -0.651195 : 1
                variance > -0.651195
                    curtosis <= 7.531300000000001 : 1
                    curtosis > 7.531300000000001 : 0
            skewness > -1.81995
                variance <= -2.185
                    variance <= -2.69165 : 0
                    variance > -2.69165 : 1
                variance > -2.185 : 0
    skewness > 5.86535
        variance <= -3.4448999999999996 : 1
        variance > -3.4448999999999996 : 0
variance > 0.32016500000000003
    variance <= 1.7907000000000002
        curtosis <= -2.2721999999999998
            skewness <= 6.41995 : 1
            skewness > 6.41995 : 0
        curtosis > -2.2721999999999998
            entropy <= 0.09222649999999999
```

```
                    variance <= 0.450925
                        variance <= 0.400865 : 0
                        variance > 0.400865 : 1
                    variance > 0.450925 : 0
                entropy > 0.09222649999999999
                    curtosis <= 0.866835 : 1
                    curtosis > 0.866835 : 0
        variance > 1.7907000000000002
            curtosis <= -4.95885 : 1
            curtosis > -4.95885
                variance <= 2.03655
                    variance <= 2.02515 : 0
                    variance > 2.02515 : 1
                variance > 2.03655 : 0
```

Using Scikit-learn :

ID3 Algorithm :

The binary tree structure has 27 nodes and has the following tree structure:

```
node=0 test node: go to node 1 if variance <= 0.3201649934053421 else to node 16.
    node=1 test node: go to node 2 if skewness <= 5.865350008010864 else to node 11.
        node=2 test node: go to node 3 if curtosis <= 3.0642000436782837 else to node 4.
            node=3 leaf node.
            node=4 test node: go to node 5 if skewness <= -1.8199500441551208 else to node 8.
                node=5 test node: go to node 6 if variance <= -0.6511950194835663 else to node 7.
                    node=6 leaf node.
                    node=7 leaf node.
                node=8 test node: go to node 9 if curtosis <= 4.400849938392639 else to node 10.
                    node=9 leaf node.
                    node=10 leaf node.
        node=11 test node: go to node 12 if variance <= -3.4449000358581543 else to node 15.
            node=12 test node: go to node 13 if curtosis <= 2.111400008201599 else to node 14.
                node=13 leaf node.
                node=14 leaf node.
            node=15 leaf node.
    node=16 test node: go to node 17 if variance <= 2.2353999614715576 else to node 26.
        node=17 test node: go to node 18 if curtosis <= -2.2721999883651733 else to node 21.
            node=18 test node: go to node 19 if skewness <= 6.419950008392334 else to node 20.
                node=19 leaf node.
                node=20 leaf node.
            node=21 test node: go to node 22 if skewness <= 1.3636500239372253 else to node 25.
```

```
                node=22 test node: go to node 23 if curtosis <= 0.4701150059700012 else to node 24.
                    node=23 leaf node.
                    node=24 leaf node.
                node=25 leaf node.
        node=26 leaf node.
```

CART Algorithm :

The binary tree structure has 35 nodes and has the following tree structure:

```
node=0 test node: go to node 1 if variance <= 0.3201649934053421 else to node 18.
    node=1 test node: go to node 2 if skewness <= 7.7639501094818115 else to node 15.
        node=2 test node: go to node 3 if variance <= -0.45856499671936035 else to node 10.
            node=3 test node: go to node 4 if curtosis <= 6.218649864196777 else to node 7.
                node=4 test node: go to node 5 if skewness <= 7.224900007247925 else to node 6.
                    node=5 leaf node.
                    node=6 leaf node.
                node=7 test node: go to node 8 if skewness <= -4.674500226974487 else to node 9.
                    node=8 leaf node.
                    node=9 leaf node.
            node=10 test node: go to node 11 if curtosis <= 2.624650001525879 else to node 14.
                node=11 test node: go to node 12 if skewness <= 5.453549861907959 else to node 13.
                    node=12 leaf node.
                    node=13 leaf node.
                node=14 leaf node.
        node=15 test node: go to node 16 if variance <= -4.726000070571899 else to node 17.
            node=16 leaf node.
            node=17 leaf node.
    node=18 test node: go to node 19 if curtosis <= -4.455850124359131 else to node 22.
        node=19 test node: go to node 20 if variance <= 3.2221500873565674 else to node 21.
            node=20 leaf node.
            node=21 leaf node.
        node=22 test node: go to node 23 if variance <= 1.5651999711990356 else to node 30.
            node=23 test node: go to node 24 if curtosis <= -2.2721999883651733 else to node 27.
    node=24 test node: go to node 25 if skewness <= 5.657400012016296 else to node 26.
                    node=25 leaf node.
                    node=26 leaf node.
                node=27 test node: go to node 28 if skewness <= 1.367799997329712 else to node 29.
                    node=28 leaf node.
                    node=29 leaf node.
            node=30 test node: go to node 31 if variance <= 2.038800001144409 else to node 34.
                node=31 test node: go to node 32 if curtosis <= -2.6483500003814697 else to node 33.
                    node=32 leaf node.
                    node=33 leaf node.
                node=34 leaf node.
```

2) The value of information gain and gini index of the root

node Using My Model :

ID3 Algorithm :

Root Node is : variance
Entropy of root Node is : 0.993
Information gain of root node is : 0.409

CART Algorithm :

Root Node is : variance
Gini Index of root node is : 0.494
Gini index of diversity of root node is : 0.238

Using Scikit-learn:

ID3 Algorithm :

Root Node is : variance
Entropy of root Node is : 0.990
Information gain of root node is : 0.430

CART Algorithm :

Root Node is : variance
Gini Index of root node is : 0.493
Gini index of diversity of root node is : 0.263

3) The labels generated on the test data and accuracy on the test datausing

a) My Model

Labels generated using ID3 Algorithm :

```
[1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1,
1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0,
0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1,
0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0,
1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0,
1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0,
0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1,
0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0,
0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1,
0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1]
```

Accuracy using ID3 Algorithm = 97.08%

Labels generated using CART Algorithm :

```
  [0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0,
1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0,
1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0,
0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1,
0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0,
0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1,
1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1,
0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1,
1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0]
```

Accuracy using CART Algorithm = 98.18%

b) Scikit Learn

Labels generated using ID3 Algorithm :

```
        [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 1. 0. 1. 1. 1. 1. 1. 0. 0.
1. 0.1. 0. 0. 1. 0. 0. 1. 0. 0. 1. 1. 0. 1. 1. 1. 0. 0. 1. 1. 0. 1. 1. 1. 0.
 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 1. 1. 0.
 1. 0. 0. 0. 0. 1. 1. 1. 1. 0. 1. 0. 0. 0. 0. 1. 1. 0. 0. 0. 1. 1. 0. 1.
 1. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 1. 1. 1. 1. 1. 0. 1. 1. 1. 0. 1. 1. 0.
 1. 0. 1. 0. 1. 0. 1. 1. 0. 1. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0.
 1. 1. 1. 1. 1. 0. 1. 1. 1. 0. 1. 0. 1. 0. 0. 0. 1. 1. 1. 1. 1. 0. 1. 0.
 0. 0. 0. 0. 0. 1. 0. 0. 1. 1. 0. 0. 0. 0. 1. 0. 1. 0. 1. 1. 0. 0. 1. 0.
 0. 1. 1. 1. 1. 0. 0. 1. 1. 1. 0. 0. 1. 1. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 1. 1. 1. 1. 0. 1. 0. 0. 1. 1. 1. 1. 0. 0. 0. 1. 1. 1. 1. 0. 0.
 0. 1. 0. 1. 1. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 1.
 1. 0. 1. 0. 1. 1. 1. 1. 0. 0. 0.]
```

Accuracy using ID3 Algorithm = 98.91%

Labels generated using CART Algorithm :

```
            [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 1. 0. 1. 1. 1. 1. 1. 0. 0.
1. 0.1. 0. 0. 1. 0. 0. 1. 0. 0. 1. 1. 0. 1. 0. 1. 0. 0. 1. 1. 0. 1. 1. 1. 0.
 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 1. 1. 0.
 1. 0. 0. 0. 0. 1. 1. 1. 1. 0. 1. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 1. 0. 1.
 1. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 1. 1. 1. 1. 1. 0. 1. 1. 1. 0. 1. 1. 0.
 1. 0. 0. 0. 1. 0. 1. 1. 0. 1. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0.
 1. 1. 1. 1. 1. 0. 1. 1. 1. 0. 1. 0. 1. 0. 0. 0. 1. 1. 1. 1. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 1. 0. 0. 1. 1. 0. 0. 0. 0. 1. 0. 1. 0. 1. 1. 0. 0. 1. 0.
 0. 1. 1. 1. 0. 0. 0. 1. 1. 1. 0. 0. 1. 1. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 1. 1. 1. 1. 0. 1. 0. 0. 1. 1. 1. 1. 0. 0. 0. 1. 1. 1. 1. 0. 0.
 0. 1. 0. 1. 1. 1. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 1. 1. 0. 0. 0. 1.
 1. 0. 1. 0. 1. 0. 1. 0. 0. 0. 0.]
```

Accuracy using CART Algorithm = 96.78%