**Machine Learning Assignment 2 :**

**Decision Trees , Bagged Trees and Random Forests**

1. **Decision Tree :**
   1. Bank Note Authentication

Accuracy for testing : 94.9275 %

Hyper parameters used:

Depth = 5

Minimum Size =10

Discretization fineness = 10

Tree Structure :

( in terms of the attribute for splitting and the value of split and the depth)

|  |  |  |
| --- | --- | --- |
| Attribute | Split Value | Depth |
| 0 | 2.02816666667 | 0 |
| 1 | 4.04336666667 | 1 |
| 3 | -0.0552 | 1 |
| 0 | -0.268 | 2 |
| 3 | -3.1119 | 2 |
| 0 | 3.835844 | 2 |
| 2 | 5.9421 | 3 |
| 0 | 1.2634 | 3 |
| 1 | 10.033 | 3 |
| 3 | -1.4188 | 3 |
| 3 | 0.515 | 3 |

Tree Leafs (terminal nodes) :

(in terms of the class at each node and number of datasets , in left to right fashion)

|  |  |  |
| --- | --- | --- |
| Class | Number of Dataset | Depth |
| 0.0 | 333 | 2 |
| 0.0 | 71 | 4 |
| 0.0 | 56 | 4 |
| 0.0 | 77 | 4 |
| 0.0 | 86 | 4 |
| 1.0 | 80 | 4 |
| 1.0 | 59 | 4 |
| 0.0 | 80 | 4 |
| 0.0 | 71 | 4 |
| 0.0 | 73 | 4 |
| 0.0 | 79 | 4 |
| 0.0 | 19 | 4 |

* 1. Sensorless Drive Diagnosis

Accuracy for testing : 87.454%

Hyper parameters used:

Depth = 6

Minimum Size =10

Discretization fineness = 10

Tree Structure :

( in terms of the attribute for splitting and the value of split and the depth)

Node : The values are in the order of SplitIndex , Value , Depth

Leafs : The values are in the order of Depth , Class , Number of Datasets

Nodes : index , value , depth 18 1.02812666667 0

Nodes : index , value , depth 15 0.000763697777778 1

Nodes : index , value , depth 11 -0.0129038888889 2

Nodes : index , value , depth 7 -0.000135888888889 3

Leafs :  4 10.0 339

Nodes : index , value , depth 11 -0.0368147777778 4

Leafs :  5 1.0 168

Nodes : index , value , depth 43 -1.49704444444 5

Leafs :  6 1.0 98

Leafs :  6 1.0 85

Nodes : index , value , depth 7 0.0297976666667 3

Nodes : index , value , depth 23 0.878918888889 4

Nodes : index , value , depth 23 0.863156666667 5

Leafs :  6 3.0 89

Leafs :  6 3.0 92

Nodes : index , value , depth 19 0.887474444444 5

Leafs :  6 3.0 97

Nodes : index , value , depth 27 0.00117344444444 6

Leafs :  7 5.0 57

Leafs :  7 3.0 54

Nodes : index , value , depth 18 0.861226666667 4

Leafs :  5 7.0 149

Nodes : index , value , depth 36 -0.613857777778 5

Leafs :  6 7.0 98

Leafs :  6 7.0 108

Nodes : index , value , depth 18 0.97805 2

Nodes : index , value , depth 18 0.944743333333 3

Nodes : index , value , depth 6 -0.000730555555556 4

Nodes : index , value , depth 44 -1.49721111111 5

Leafs :  6 4.0 93

Leafs :  6 4.0 95

Nodes : index , value , depth 9 0.0113775555556 5

Leafs :  6 3.0 76

Leafs :  6 5.0 88

Nodes : index , value , depth 23 0.961072222222 4

Nodes : index , value , depth 11 0.0179752222222 5

Nodes : index , value , depth 12 0.000769483333333 6

Leafs :  7 8.0 55

Leafs :  7 6.0 51

Leafs :  6 5.0 122

Nodes : index , value , depth 22 0.967363333333 5

Nodes : index , value , depth 23 0.963233333333 6

Leafs :  7 5.0 56

Leafs :  7 5.0 65

Nodes : index , value , depth 47 -1.49725555556 6

Leafs :  7 6.0 63

Leafs :  7 8.0 63

Nodes : index , value , depth 19 1.00581333333 3

Nodes : index , value , depth 21 0.994513333333 4

Nodes : index , value , depth 11 0.0385461111111 5

Leafs :  6 6.0 51

Leafs :  6 4.0 54

Nodes : index , value , depth 23 1.00047222222 5

Leafs :  6 4.0 66

Leafs :  6 4.0 63

Nodes : index , value , depth 18 1.0134 4

Nodes : index , value , depth 42 -1.49597777778 5

Nodes : index , value , depth 23 1.01086666667 6

Leafs :  7 6.0 47

Leafs :  7 6.0 71

Leafs :  6 11.0 89

Nodes : index , value , depth 19 1.01664444444 5

Leafs :  6 11.0 99

Leafs :  6 8.0 79

Nodes : index , value , depth 18 1.13062222222 1

Nodes : index , value , depth 23 1.06408888889 2

Nodes : index , value , depth 22 1.04373333333 3

Nodes : index , value , depth 20 1.0353 4

Nodes : index , value , depth 18 1.03286666667 5

Leafs :  6 8.0 65

Leafs :  6 8.0 53

Nodes : index , value , depth 20 1.0383 5

Leafs :  6 11.0 54

Leafs :  6 11.0 61

Nodes : index , value , depth 21 1.0505 4

Nodes : index , value , depth 0 -4.854e-06 5

Leafs :  6 11.0 79

Leafs :  6 9.0 69

Leafs :  5 9.0 98

Nodes : index , value , depth 18 1.0777 3

Nodes : index , value , depth 21 1.07576666667 4

Nodes : index , value , depth 18 1.0722 5

Leafs :  6 2.0 77

Leafs :  6 2.0 68

Nodes : index , value , depth 18 1.0761 5

Leafs :  6 2.0 65

Leafs :  6 2.0 72

Nodes : index , value , depth 18 1.08441111111 4

Nodes : index , value , depth 18 1.08142222222 5

Leafs :  6 2.0 90

Leafs :  6 2.0 79

Nodes : index , value , depth 39 -0.593153333333 5

Leafs :  6 9.0 58

Leafs :  6 9.0 58

Nodes : index , value , depth 18 1.36904444444 2

Nodes : index , value , depth 7 0.0121106666667 3

Nodes : index , value , depth 19 1.30207777778 4

Nodes : index , value , depth 23 1.29274444444 5

Leafs :  6 3.0 63

Leafs :  6 3.0 74

Nodes : index , value , depth 20 1.34563333333 5

Leafs :  6 4.0 66

Leafs :  6 10.0 75

Nodes : index , value , depth 19 1.31553333333 4

Nodes : index , value , depth 19 1.29661111111 5

Leafs :  6 7.0 92

Leafs :  6 1.0 71

Nodes : index , value , depth 18 1.3628 5

Leafs :  6 5.0 68

Leafs :  6 5.0 73

Nodes : index , value , depth 19 1.46243333333 3

Nodes : index , value , depth 18 1.43103333333 4

Nodes : index , value , depth 22 1.37883333333 5

Leafs :  6 8.0 88

Leafs :  6 2.0 74

Leafs :  5 6.0 95

Nodes : index , value , depth 9 -0.0287392444444 4

Leafs :  5 11.0 95

Leafs :  5 9.0 90

The size of the dataset being 58,000 with 50 features was very computationally intense and needed a long time , hence values were sampled from each class and these were used to train the data.

1. **Bagged Tree :**
   1. Bank Note Authentication

Hyper parameters used:

Depth = 5

Minimum Size =10

Discretization fineness = 10

K Value = 13

Accuracy = 89.50% (might vary due to the shuffle of training data)

* 1. Sensorless Drive Diagnosis

Hyper parameters used:

Depth = 5

Minimum Size =10

Discretization fineness = 10

K Value = 9

Accuracy = 73..23%

1. **Random Forests :**
   1. Bank Note Authentication

Hyper parameters used:

Depth = 5

Minimum Size =10

Discretization fineness = 10

K Value = 5

Choice beween top random = 5 values

Accuracy = 94.65% (varying between 70 – 98) due to random choice)