Decision trees

1. A)Graphs are:
   1. No of nodes : 13094
   2. Train: 88.9
   3. Valid: 79.9
   4. Test: 80.6

The train accuracies are highest means tree is over fitted.

B) Scikit-learn Decision Trees

With default values accuracies are :-

Train accuracy is: 0.8893333333333333

Valid accuracy is: 0.8007142857142857

Test accuracy is: 0.7973333333333333

|  |  |  |  |
| --- | --- | --- | --- |
| max\_depth | min\_samples\_leaf | min\_samples\_split | Valid accuracy |
| 2 | 1 | 2 | 80.45 |
| 4 | 1 | 2 | 81.2 |
| 8 | 1 | 2 | 82.1 |
| 10 | 4 | 2 | 825 |

Highest accuracy of valid data: 82.5

Parameter values are:-

Max\_depth = 10, mini\_samples\_leaf = 4, mini\_sample\_split= 2

At these parameters accuracy are:-

Train accuracy is : 0.8402592592592593

Valid accuracy is : 0.8255714285714286

Test accuracy is : 0.8236666666666667

Almost same to above part in pruning for valid data

C) Random Forest:-

Accuracies at default values are:-

Train accuracy is : 0.8843703703703704

Valid accuracy is : 0.8066666666666666

Test accuracy is : 0.8151428571428572

|  |  |  |  |
| --- | --- | --- | --- |
| n\_estimators | Max\_features | Bootstrap | Valid accuracy |
| 10 | Auto | True | 80.7 |
| 50 | Auto | True | 80.8 |
| 200 | 4 | True | 80.9 |
| 200 | 6 | True | 81.1 |
| 200 | 9,10 | True | 81.1 |

Train accuracy is : 0.8893333333333333

Valid accuracy is : 0.8116666666666666

Test accuracy is : 0.8162857142857143