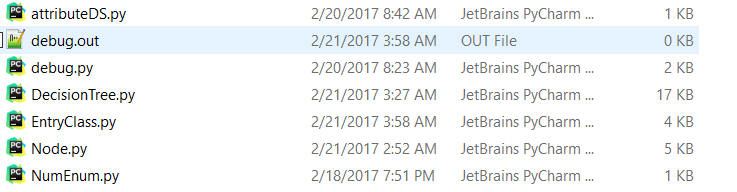
Name: Akhilesh Kumar Kagalvadi Chinnaswamy(axk167131)  
Machine Learning Assignment 1.  
  
Read Me file:  
----------------------------------------------------------------------

The program is coded in Python using a Pycharm IDE.  
unzip decision tree.zip  
open /DecisionTree you will find the coding files along with the datasets.

# The entry class for the file is : EntryClass.py

Also contains remaining files:  


Library dependencies include pandas:

**import** random  
**import** copy  
**import** pandas **as** pd  
**import** math

To run the file, use following command.  
\DecisionTree> python EntryClass.py 10 20 "./data\_sets1/test\_set.csv" "./data\_sets1/training\_set.csv" "./data\_sets1/validation\_set.csv" "yes"

\DecisionTree> python EntryClass.py 10 20 "./data\_sets1/test\_set.csv" "./data\_sets1/training\_set.csv" "./data\_sets1/validation\_set.csv" "no"

\DecisionTree> python EntryClass.py 10 20 "./data\_sets2/test\_set.csv" "./data\_sets2/training\_set.csv" "./data\_sets2/validation\_set.csv" "yes"

\DecisionTree> python EntryClass.py 10 20 "./data\_sets2/test\_set.csv" "./data\_sets2/training\_set.csv" "./data\_sets2/validation\_set.csv" "no"

Report :

Output for both stored in text file: (output.txt) and also embedded in this page.



Without tree print is shown below:

Sample output without tree:

|  |  |
| --- | --- |
| .\data\_sets1\test\_set.csv  .\data\_sets1\training\_set.csv  .\data\_sets1\validation\_set.csv  Fitting the DataSet with Entropy Method  Accuracy of Test data with Entropy Method is 0.7585  Accuracy of Validation data with Entropy Method is 0.759  PrunedTree with Entropy Method for Validation Data K = 10 l= 5  The Best Accuracy of Validation Data is 0.763  PrunedTree with validation data, the Test data results for Entropy method K = 10 l= 5  The Accuracy for test data in Pruned Tree is 0.7595  Fitting the Dataset with variance impurity  Accuracy of Test data with variance impurity is 0.752  Accuracy of Validation data with variance impurity is 0.758  PrunedTree with validation data for info Gain K = 10 l= 5  The Best Accuracy of validation data for Pruned Tree is 0.758  PrunedTree with validation data, the Test data results for variance impurity K = 10 l= 5  The Accuracy of test data for Pruned Tree is 0.752 | .\data\_sets2\test\_set.csv  .\data\_sets2\training\_set.csv  .\data\_sets2\validation\_set.csv  Fitting the DataSet with Entropy Method  Accuracy of Test data with Entropy Method is 0.7233333333333334  Accuracy of Validation data with Entropy Method is 0.7733333333333333  PrunedTree with Entropy Method for Validation Data K = 10 l= 5  The Best Accuracy of Validation Data is 0.7766666666666666  PrunedTree with validation data, the Test data results for Entropy method K = 10 l= 5  The Accuracy for test data in Pruned Tree is 0.73  Fitting the Dataset with variance impurity  Accuracy of Test data with variance impurity is 0.725  Accuracy of Validation data with variance impurity is 0.7733333333333333  PrunedTree with validation data for info Gain K = 10 l= 5  The Best Accuracy of validation data for Pruned Tree is 0.78  PrunedTree with validation data, the Test data results for variance impurity K = 10 l= 5  The Accuracy of test data for Pruned Tree is 0.7216666666666667 |
|  |  |

Report the accuracy on the test set for decision trees constructed using the two heuristics mentioned above. \_ Choose 10 suitable values for L and K (not 10 values for each, just 10 combinations). For each of them, report the accuracies for the post-pruned decision trees constructed using the two heuristics.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Unpruned Accuracy | With Pruning Accuracy (Validation Data) and New tree is tested with Test data and also validation data new accuracy are reported. | | | | | | | | | |
|  |  |  | K=10  L=5 | K=10 L=10 | K=10  L=15 | K=10  L=20 | K=20  L=5 | K=20  L=10 | K=20  L=15 | K=20  L=20 | K=30  L=30 | K=5 L=5 |
| Entropy method | (V.D) | 0.759 | 0.759 | 0.764 | 0.768 | 0.768 | 0.773 | 0.759 | 0.77 | 0.774 | 0.768 | 0.759 |
| (test) | 0.752 | 0.758 | 0.7605 | 0.763 | 0.757 | 0.773 | 0.758 | 0.758 | 0.763 | 0.760 | 0.758 |
| variance impurity | (V.D) | 0.758 | 0.77 | 0.7665 | 0.767 | 0.771 | 0.762 | 0.758 | 0.770 | 0.759 | 0.764 | 0.762 |
| (test) | 0.752 | 0.766 | 0.755 | 0.761 | 0.762 | 0.752 | 0.752 | 0.764 | 0.752 | 0.756 | 0.754 |
| Entropy | (V.D) | 0.773 | 0.77 | 0.7833 | 0.785 | 0.785 | 0.778 | 0.775 | 0.79 | 0.783 | 0.781 | 0.78 |
| (test) | 0.7233 | 0.723 | 0.7833 | 0.726 | 0.72 | 0.713 | 0.73 | 0.736 | 0.721 | 0.705 | 0.726 |
| variance impurity | (V.D) | 0.7733 | 0.776 | 0.78166 | 0.785 | 0.78 | 0.786 | 0.776 | 0.791 | 0.786 | 0.8 | 0.776 |
| (test) | 0.725 | 0.76 | 0.7366 | 0.733 | 0.73 | 0.721 | 0.731 | 0.741 | 0.726 | 0.741 | 0.728 |