The University of Texas at Dallas
Machine Learning – Prof. Anjum Chida
Assignment #01

#01. Implement the decision tree learning algorithm.

#02. Implement the post pruning algorithm given as Algorithm 1.

#03. Implement a function to print the decision tree to standard output.

#04. Report the accuracy on the test set for decision trees constructed using the two heuristics mentioned above.

#05. 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.

Ans.
Please, refer DecisionTrees.java file attached herewith for entire design.
For DataSet 01:

[L,K]	Information Gain		Variance Impurity	
	Non-Pruned	Pruned	Non-Pruned	Pruned
[30,50]	74.54	74.04	74.20	75.04
[70,100]	75.16	75.21	76.11	76.31
[90,150]	75.16	74.86	76.11	76.21
[100,200]	75.16	76.51	76.11	75.41
[150,250]	75.16	75.16	76.11	76.11
[200,300]	75.16	75.26	76.11	76.21
[300,400]	75.16	75.81	76.11	76.11
[400,500]	75.16	75.01	76.11	76.66
[500,550]	75.16	75.31	76.11	76.11
[600,600]	75.16	75.46	76.11	75.11

Similarly, a table can be constructed for DataSet 02. I've took same 10 [L,K] pair-values for DataSet 02 too.

Please, find all the outputs in the **DSXY.log** files, and **DSXY.png** screenshots.
where X = {1,2} representing DATASET and Y ={1,2,3,4,...,10} representing 10 different [L,K] values.