Assignment 5 – Classification Tree 45 Points

1. The “vacation-trip-classification.csv” dataset holds data household income, family size and result of buying vacation package or not. We want to classify these household on the purchasing or the vacation package or not. (10 Points)
   1. *What is the target attribute?*
   2. *What are classifier attributes?*
   3. *Create a classification tree with the following configuration:*
   4. *Dataset is partitioned 70% training and 30% test*
   5. *For control parameters use: ignore minbucket, minsplit=2 and cp = 0*
   6. *Create a confusion matrix for test dataset. Interpret the accuracy, sensitivity, and specifity*
   7. *Plot the tree. Interpret the tree leaves data*
   8. *Apply your model on new data in the “New vacation-trip.csv” file.*

Submit validation data confusion matrix, and all necessary plots and new data classification in a csv file.

1. The objective of this exercise is to classify the Sales of the child car seats. The available data set is stored in *careseat.csv* file (20 Points)

The dataset attributes are defined below.

1. Sales: unit sales in thousands
2. CompPrice: price charged by competitor at each location
3. Income: community income level in 1000s of dollars
4. Advertising: local ad budget at each location in 1000s of dollars
5. Population: regional pop in thousands
6. Price: price for car seats at each site
7. ShelveLoc: Bad, Good or Medium indicates quality of shelving location
8. Age: age level of the population
9. Education: ed level at location
10. Urban: Yes/No
11. US: Yes/No

You Observe that Sales is a quantitative variable. You want to demonstrate it using a decision tree with a binary response. To do so, you turn Sales into a binary variable, which will be called High\_Sales. If the sales is less than or equal to 8 (8k), it will be not high. Otherwise, it will be high. Then you can put that new variable High\_Sales back into the data frame and use it as the model response!

You job is to document analysis of every attributes as well as dataset then, create a decision tree model and evaluate its performance.

1. The Wine.csv includes characteristics of quality wines. These characteristics are used to classify wines into three classes A, B, C. Apply tree algorithm and predict classification of validation data. Then evaluate your model using confusion matrix. (15 Points)