**Assignment 8 - SAS #2: Decision Trees**

**NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

***Instructions: Follow the steps and answer the questions below. Then email this document to your instructor.***

You’ll be working on the project you created in the previous assignment. Remember, this project used the “Organics” data set. When you open SAS Enterprise Miner, you should be able to find your work under the File/Recent Projects. If you can’t find it there, go to File/Open Projects… and search for your project.

*Make sure that the variable roles and levels in the Organics data set are still properly set up from the last assignment. You can verify this by right-clicking the Organics node in your diagram and selecting “Edit Variables.” Check those settings against page 1 of “Introduction to Working with SAS.”*

**Create a Decision Tree based on the Organics Data Set**

1. Add a **Data Partition** node to the diagram and connect it to the **Data Source** node. Assign 50% of the data for training and 50% for validation (and 0% for test). Run it.  
     
   Add a **Decision Tree** node to the workspace and connect it to the **Data Partition** node.
2. Create a decision tree model autonomously (i.e., just run the Decision Tree node).   
     
   **Answer the two questions below and attach the screenshot(s) where you found the answer below each question. Make sure you’re looking at misclassification rate in the subtree assessment plot.**
   * 1. How many leaves are in the optimal tree?
     2. Which variable was used for the first split?
3. Add a second **Decision Tree** node to the diagram and connect it to the **Data Partition** node.  
     
   In the Properties panel of the new Decision Tree node, change the maximum number of branches (Maximum Branch, under Splitting Rule) to **3** to allow for three-way splits.  
   * 1. How many leaves are in the optimal tree?   
        *(HINT: In your iteration plot, you can click near the “Number of leaves” label and drag right to zoom in)***Attach the screenshot in your solution document where you found the answer.**
     2. Based on misclassification rate, which of the decision two tree models appears to be better (the first [autonomous] one or the second one)? ­­­­­­­­­­­­­­­­­­­­­

**<<GO TO THE NEXT PAGE!!>>**

1. Start at the top of the second decision tree you created and work your way down to answer the following questions. You don’t need to include screen shots for these questions – just provide the answer:  
     
   ***(Remember, use the validation set to find the probabilities, and a “1” means they buy.)***

|  |  |
| --- | --- |
| Question | Answer |
| What is the probability that a 33 year old man with affluence grade 10 buys Organics? |  |
| What is the probability that a 23 year old woman with affluence grade 3 buys Organics? |  |
| What is the probability that a 55 year old man with affluence grade 6 buys Organics? |  |
| What is the probability that a 64 year old woman with affluence grade 4 buys Organics? |  |
| What is the probability that a 65 year old woman with affluence grade 20 buys Organics? |  |