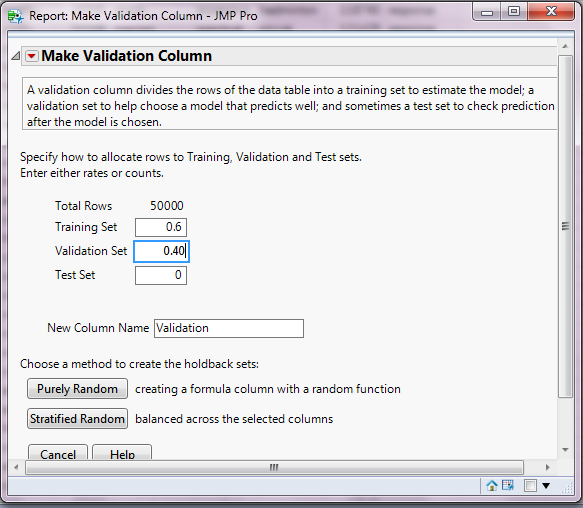
# Individual Assignment – 4

**Question 1**- Using the file “Direct Marketing” create a validation column (60 % training, 40% validation)

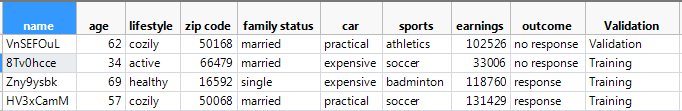
**Solution1**- Validation column could be created via following steps:

Column -> Modelling Utilities -> Make Validation Column



Select ‘Purely Random option’ above.

This leads to the creation of a Validation column as the last column of the dataset.



**Question 2**- Using the file “Direct Marketing” with “outcome as a target variable”, complete a decision tree analysis – show screen shot of small tree and the confusion matrices for the training and for the validation data set.

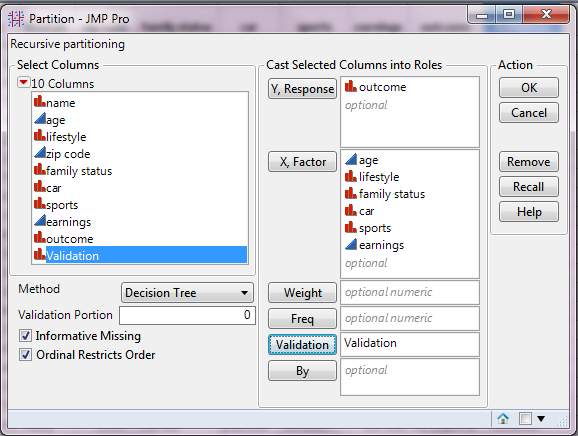
**Solution 2-**

To provide solution for this task we need to form a decision tree analysis, which is called as ‘Partition’ in JMP software.

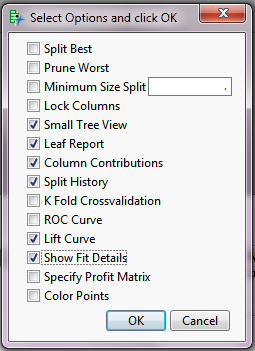
We need to goto

Analyze - > Modeling -> Partition

And select the following settings :

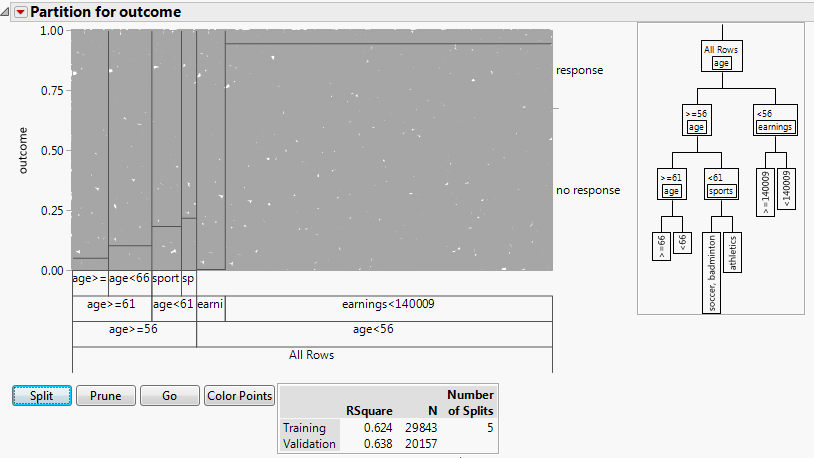


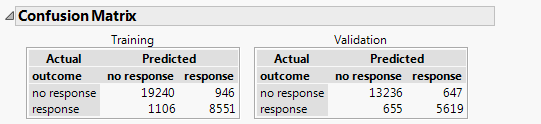
Here for modeling purpose we are excluding ‘name’ and ‘zip code’ column because they don’t bring any modeling-significant information with them. Although Zip codes could be useful if we can from their state wise groups.

After clicking OK we get the decision tree screen where we need to make following selections after clicking the red triangular button on top.

Now we need to make splits till RSquare value do not show any significant improvement between two splits. For our dataset the no. of splits comes to **5** after which there is no improvement in the RSquare value.

Small Tree Report:



Confusion Matrix:

Post this step we need to save the **prediction formula** and **script to the data table.** We need to do this so that we can use these **predictions as in input for ensemble model.**

**Question 3**- Using the file “Direct Marketing” complete a Neural Net analysis using the same training and validation data sets; show screen shot of small tree and the confusion matrices for the training and for the validation data set.

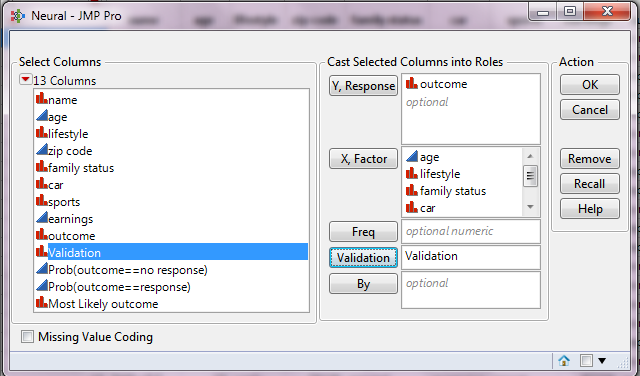
**Solution 3-**

To provide solution for this task we need to form Neural Net analysis, which is called as ‘Neural’ in JMP software.

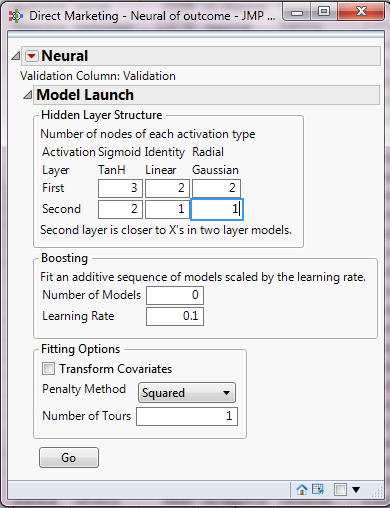
We need to goto

Analyze - > Modeling -> Neural

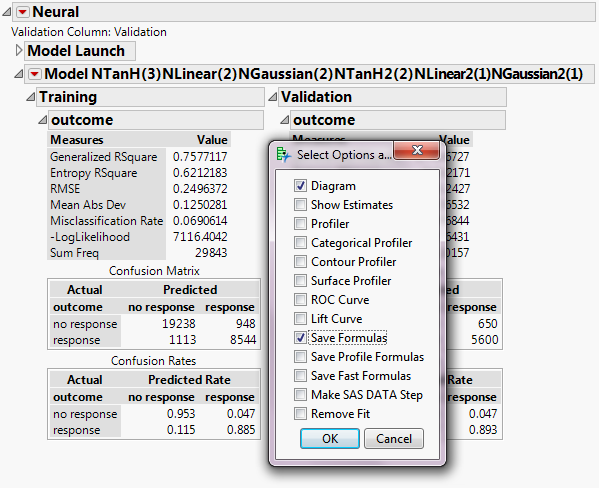
And select the following settings:

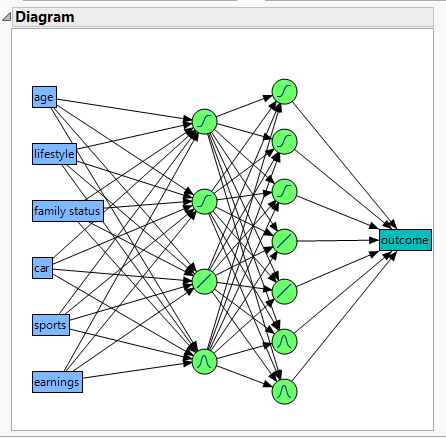


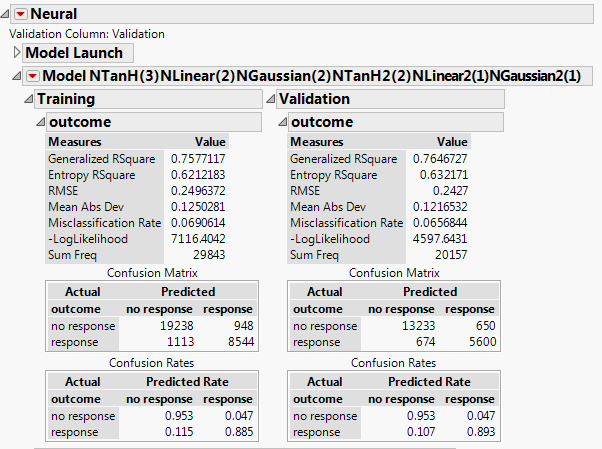
Post this we need to select the number of functions for the Neural Network Analysis. The selection here is totally subjective. **I have also tried with other different combination of numbers for the two layers**.



After clicking GO we get the decision tree screen where we need to make following selections after clicking the red triangular button on top.



**Diagram:**

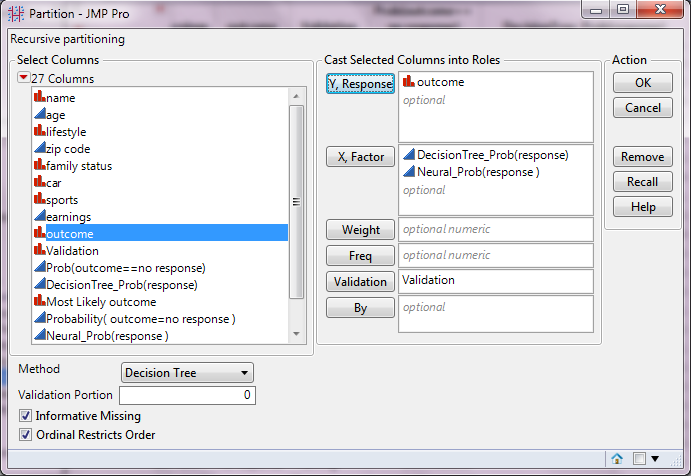
**Confusion Matrix:**

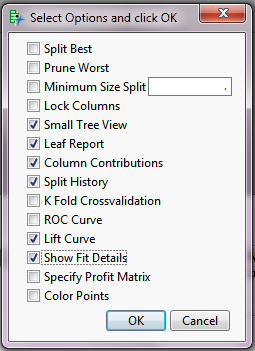
Post this step we need to save the **script to the data table.** We need to do this so that we can use these **predictions as in input for ensemble model.**

**Question 4**- complete an ensemble decision tree analysis using the output from the first decision tree and from the neural net.

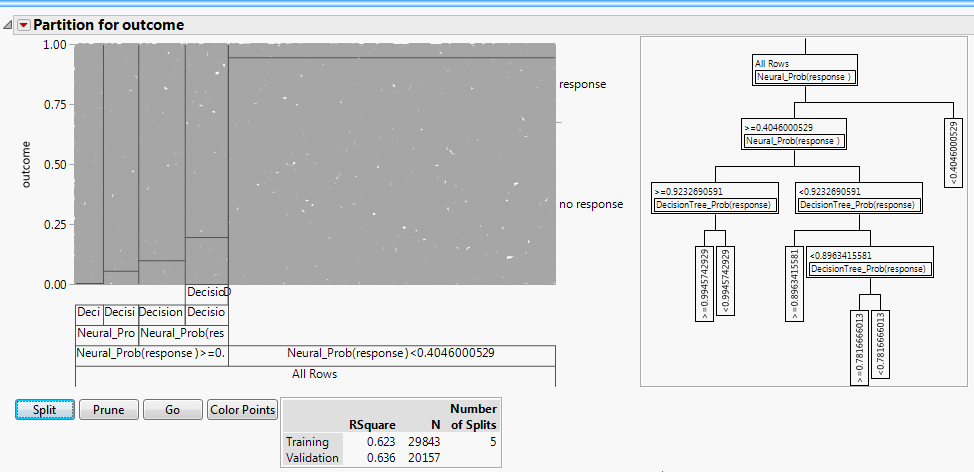
**Solution 4 –** To create this ensemble model we need to have inputs from the decision tree and from the neural net.

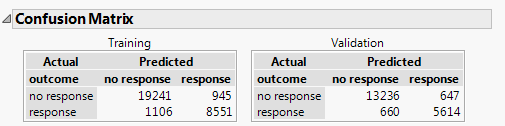
As per ask of the question we need to create a Decision Tree.

But this time our input is the probabilities (of response) i.e **DecisionTree\_Prob(response)** and **Neural\_Prob(response)** from previous 2 models.

After clicking OK we get the decision tree screen where we need to make following selections after clicking the red triangular button on top.

Now we need to make splits till RSquare value do not show any significant improvement between two splits. For our dataset the no. of splits comes to **5** after which there is no improvement in the RSquare value.

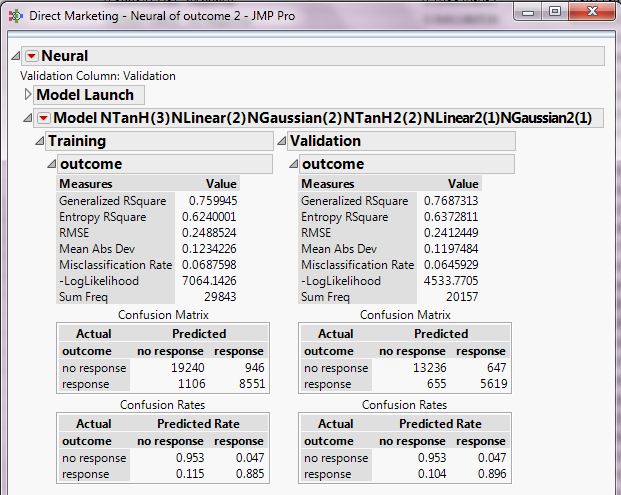
**Small Tree Report**:

**Confusion Matrix**:

**Summary :**

If we analyse Confusion Matrix of all the 3 models in this assignment, we can conclude that each of the three models provide almost the same performance. This can be said because the ratio of sum of false positives and false negatives to the total outcome is almost same in all the three models.

In order to make a choice regarding which model to choose, we might need to do a Cost (Impact) Analysis for false positives and false negatives and then make an informed decision for the most optimum model.

I also tried this problem differently by creating an ensemble **Neural Network** (using the output from the first decision tree and from the neural net) but still there is no significant difference in the performance from other models. Attached below is the outcome from that model.