**Trevor Tumas, CIS 375, MW 3:05-4:20**

**Assignment 2**

**Part 1**

* Any missing values imputation needed? **Yes**
  + Should you do this before generating the decision tree models? **No**
  + Why? **Because decision tree models are able to manage missing values on their own.**
* Run the regression node
  + Which variables are included in the final model? **IMP\_DemAffl, IMP\_DemAge, IMP\_DemGender, M\_DemAffl, M\_DemAge, M\_DemGender**
  + Which are important? **IMP\_DemAffl, IMP\_DemAge, IMP\_DemGender, M\_DemGender**
  + Validation ASE? **0.137156**
* are any transformations of the data warranted? **Yes**
  + Why? **So that we can mitigate outlying values**
* Run the transform variables node
  + Less skewed distributions? **Yes**
* Rerun the regression node
  + Selected variables change? **Yes, DemAffl becomes LOG\_DemAffl**
  + Validation ASE? **0.138204**

**Part 2**

* Is imputation of missing values needed? **Yes**
  + Why? **Like regression, neural networks face prediction problems when there are missing values. There are still missing values.**
  + **However, the neural network node can take advantage of the imputations we did for regression.**
* Is data transformation generally needed? **Yes**
  + Why? **So that we can mitigate outliers, like in regression.**
  + **However, the neural network node can take advantage of the transformations we did for regression.**
* Run the neural network node
  + How does the ASE compare to other models? **It is 0.132933, the smallest of all the other models.**

**Part 3**

* Run the model comparison node
  + Which model was selected and with what criteria? **Neural network, misclassification rate**
  + Which model has the best ROC curve? **Neural network**
  + What is the corresponding ROC index? **0.823**









