Workflow

First we added the data then I put graph explore node to see the data graphically because the data is big. Then, I put data partition node to set the data to training and validation to weight the data. Then I added Impute node to replace the missing values for interval variables. Then I added replacement node to replace selected values for variables. Then I created 2 paths one with transform variables node and the other one without transform variables node. In the path has transform variables node I created 3 new variables INDELINQ: I created Boolean logic for DELINQ if the values is missing, it will treat as 0, INDEROG I created Boolean logic for DEROG if the values is missing, it will treat as 0, and log (YOJ): log the variable YOJ. Then I added Interactive binning node to create a group for the value bigger than 2. Then I added 3 neural network nodes to select the how many hidden layers, and I have 1, 3 and 5 hidden layer node. In the path without transform variables node, I added 3 neural network nodes with 1, 3 and 5 hidden layer node. Then I added model comparison node to compare the 6 neural network nodes. then reporter to get a pdf file of the result.

In ROC chart, the variables chart show that NN without transform variables and 5 hidden layers is the best model then NN with transform variables and 5 hidden layers, in the same level the NN with transform variables and 3 hidden layers. The worst one is NN with transform variables and 1 hidden layer.

NN model without transform variables and 5 hidden layers is the best model.

I think without transform variables is better because it has more correct values than with transform variables node.

No because the best model is without transform variables node.

Yes, the high the line is the better.

	good	bad
good	357	157
bad	238	2229

overall correct classification rate = 357+2229/357+157+238+2229=0.86

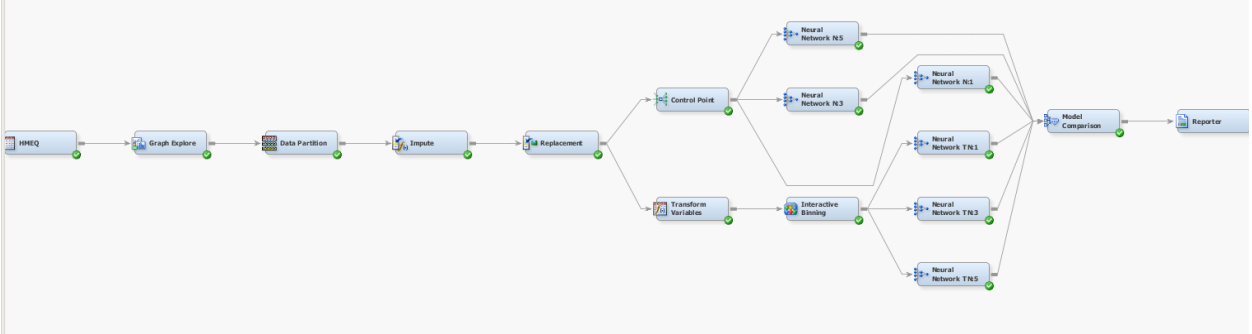
correct classification accuracy rate of good = 357/357+157=0.69

correct classification accuracy rate of bad= 2229/2229+238=0.90

false positive= 157/357+157=0.30

false negative= 0.096

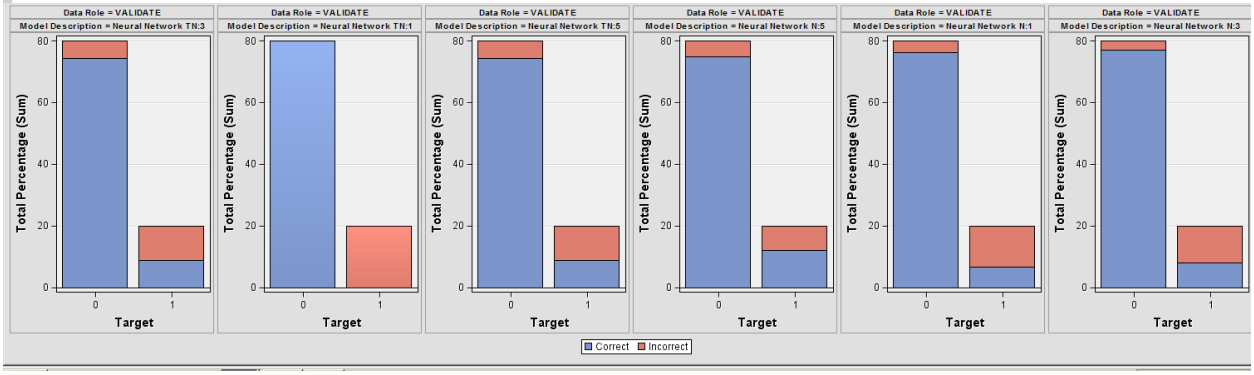
No



Fit Statistics

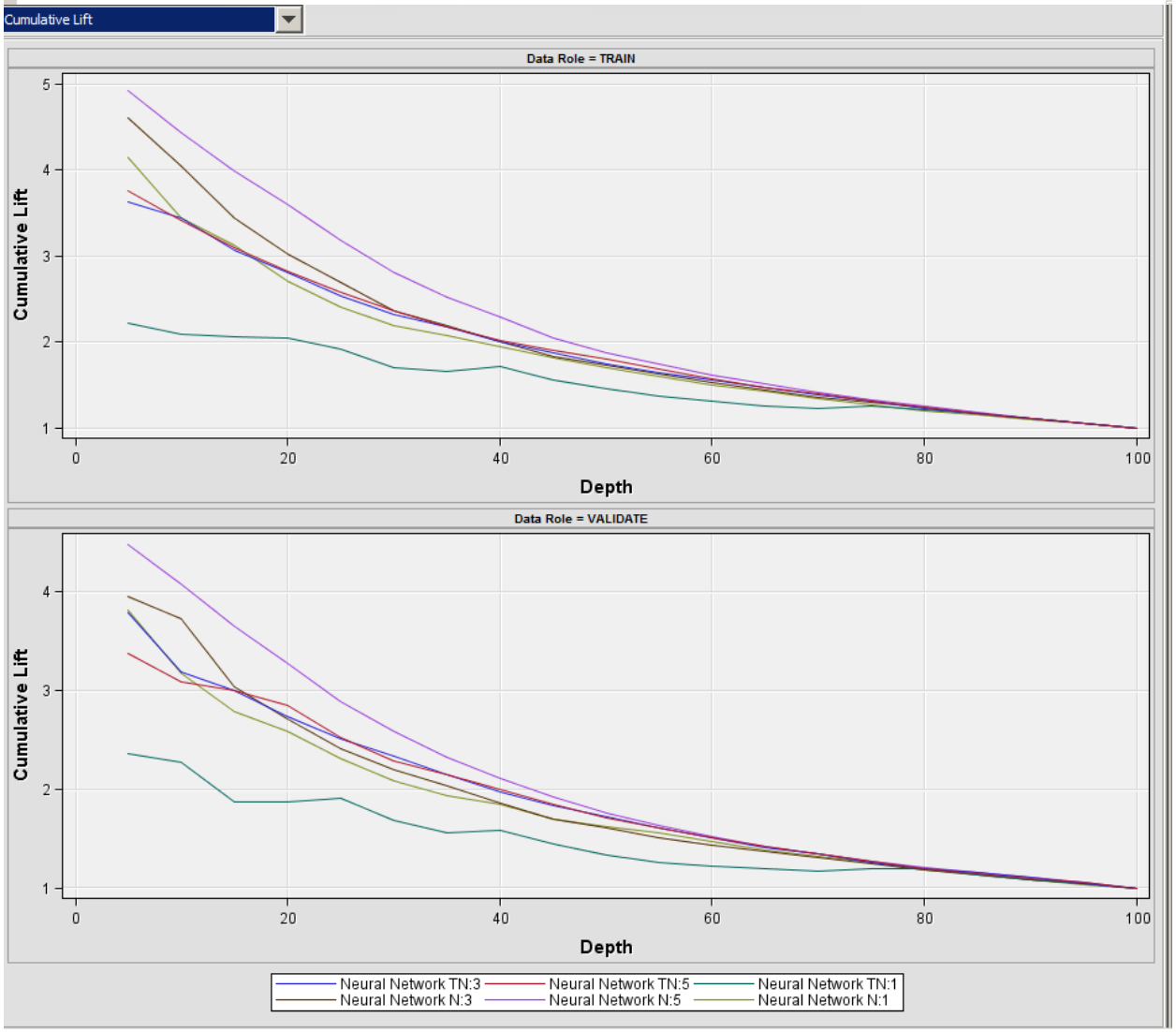
Model Selection based on Valid: Misclassification Rate (_VMISC_)

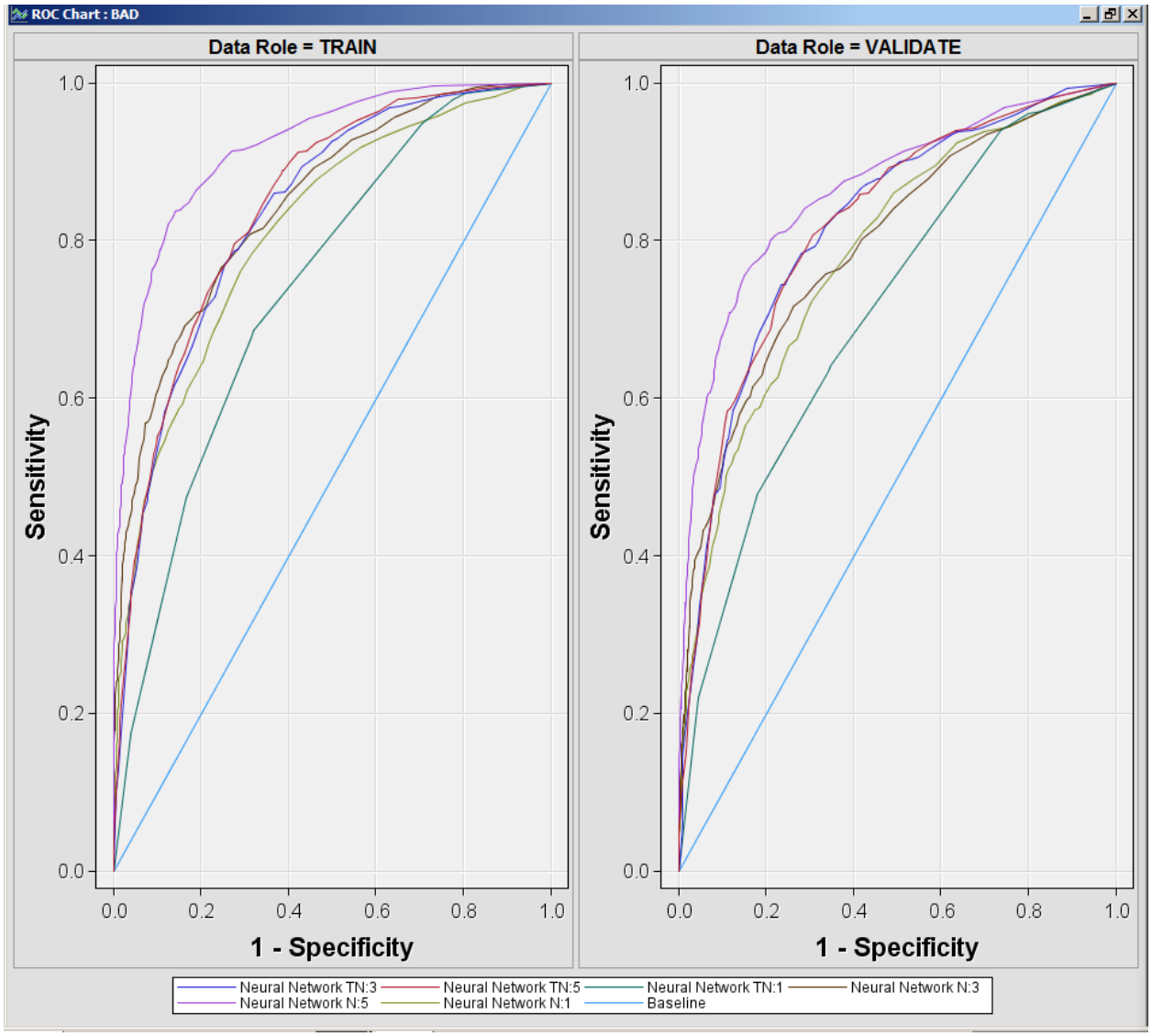
Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Neural2	Neural Network N:5	0.13251	0.07983	0.10809	0.10013
	Neural3	Neural Network N:3	0.15196	0.10631	0.13763	0.12147
	Neural5	Neural Network TN:5	0.17075	0.11567	0.16415	0.12132
	Neural6	Neural Network TN:3	0.17075	0.11758	0.16482	0.12193
	Neural	Neural Network N:1	0.17310	0.11901	0.16213	0.12747
	Neural4	Neural Network TN:1	0.19960	0.14013	0.19940	0.14419



Event Classification Table
Model Selection based on Valid: Misclassification Rate (_VMISC_)

Model Node	Model Description	Data Role	Target	Target Label	False Negative	True Negative	False Positive	True Positive
Neural6	Neural Network TN:3	TRAIN	BAD		342	2236	149	252
Neural6	Neural Network TN:3	VALIDATE	BAD		334	2211	175	261
Neural4	Neural Network TN:1	TRAIN	BAD		594	2385	0	0
Neural4	Neural Network TN:1	VALIDATE	BAD		595	2386	0	0
Neural5	Neural Network TN:5	TRAIN	BAD		340	2236	149	254
Neural5	Neural Network TN:5	VALIDATE	BAD		334	2211	175	261
Neural2	Neural Network N:5	TRAIN	BAD		221	2284	101	373
Neural2	Neural Network N:5	VALIDATE	BAD		238	2229	157	357
Neural	Neural Network N:1	TRAIN	BAD		389	2291	94	205
Neural	Neural Network N:1	VALIDATE	BAD		397	2267	119	198
Neural3	Neural Network N:3	TRAIN	BAD		343	2318	67	251
Neural3	Neural Network N:3	VALIDATE	BAD		361	2294	92	234




ROC Chart : BAD
Data Role = TRAIN
Data Role = VALIDATE
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Neural Network TN:3
Neural Network TN:5
Neural Network TN:1
Neural Network N:3
Neural Network N:5
Neural Network N:1
Baseline