**REPORT BY**

**NAME: KUSHAL SHAH**

**STUDENT ID:A20207420**

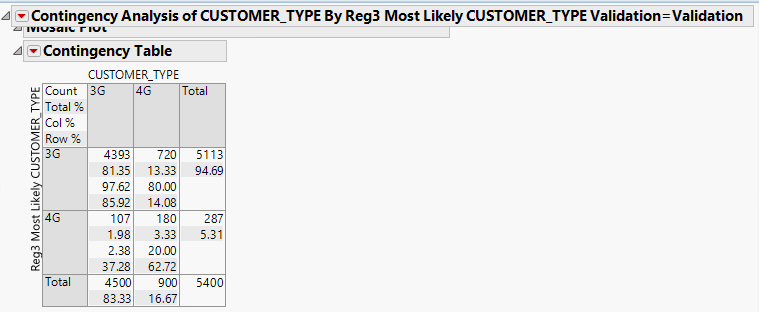
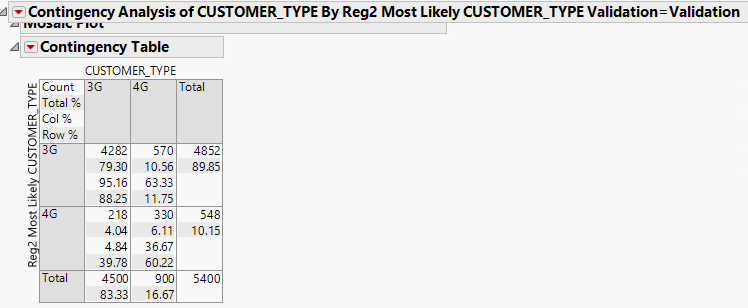
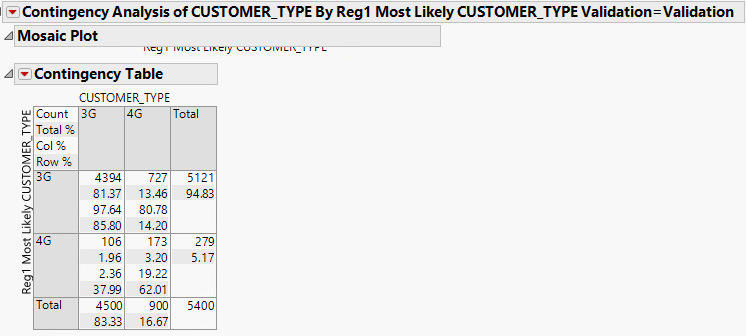
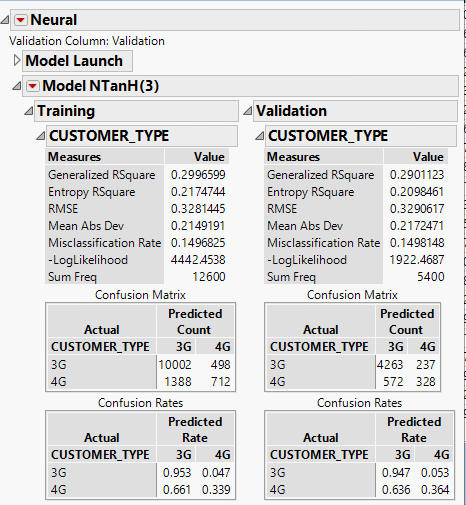
**EMAIL ID:kusshah@okstate.edu**

**Ans 1)**

1. **Yes, Imputation of missing values is needed in neural network model. If we have missing values in neural net then It will throw that observation out. In such way, we might lose important information while building and validation the model. This is why; imputation of missing values should be done before building model through neural network.**
2. **Depends, If we have extreme or unusual values (outliers) in continuous variable, we might need data transformation. Or, if we have non numeric inputs in categorical variable, we might want data transformation and convert to a flag variable. In general case, this kind of situation happens. So it is better to do data transformation for a neural network model.**

**Ans 2)**

**A)**

****

**(Neural) (R1) (R2) (R3)**

validation misclassification rate of neural net model 1: 14.98% (parameter:121)

validation misclassification rate of R1 model: 15.43%

validation misclassification rate of R2 model:14.59%

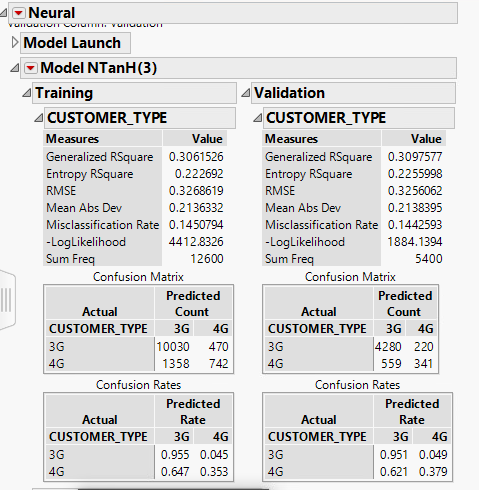
validation misclassification rate of R3 model:15.31%

Validation misclassification rate of neural net model is better than R1 and R3 model but R2 model exceeds in expectations in terms of Validation misclassification rate then neural net.

**B)**

When we speak in terms of prediction, misclassification rate is the parameter that signifies % of false prediction. Lower the misclassification rate, higher would be the accuracy for prediction. This is why We chose R2 model when it comes to prediction. Otherwise, we may look at R value for model behaviour in compare to baseline model.

Ans 3)

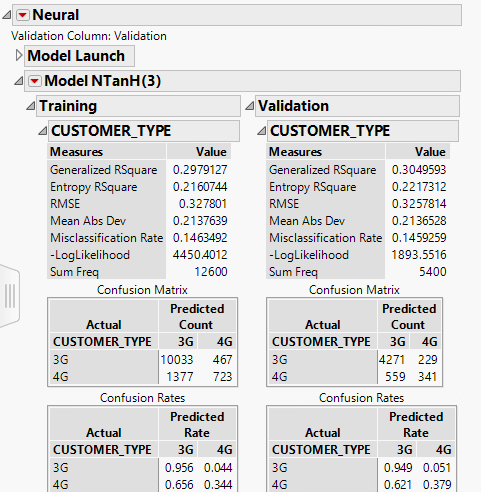


Validation misclassification rate for Neural model 2: 14.59%(parameters:37)

Validation misclassification rate for Neural model 2 is different from neural model 1(less.)

**Ans 4)**

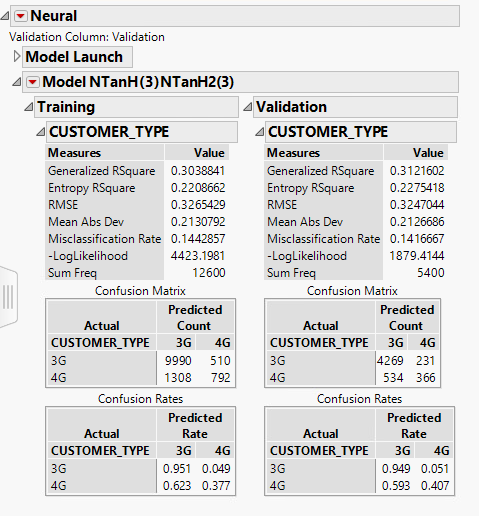
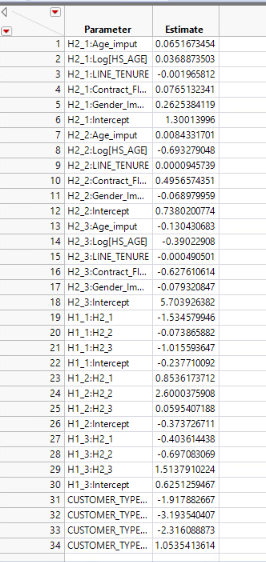
**A)**



Validation misclassification rate for Neural model 3: 14.26% (22)

Validation misclassification rate for Neural model 3 is different from neural model 1 & 2(less.)

**Ans 5)**

1. (B)

**A)**

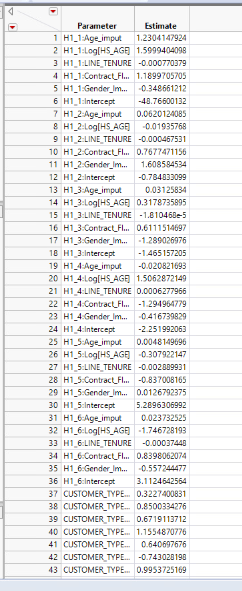
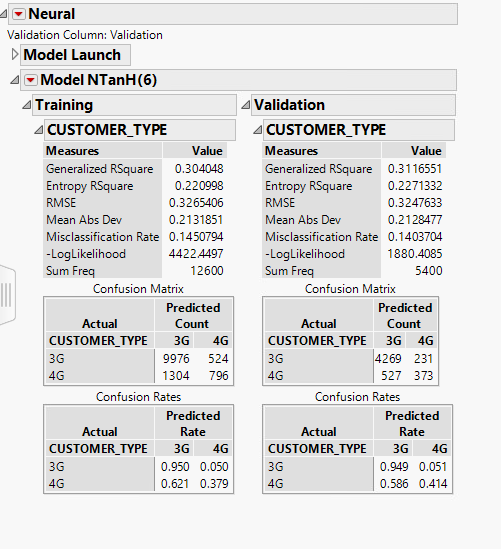
Validation misclassification rate for Neural model 4: 14.17%(parameters:34)

Validation misclassification rate for Neural model 4 is different from neural model 1, 2, 3(less.)

**B)**

Out of 34 Parameter estimates, H2\_3:Intercept(Intercept of 3rd node in second layer) has the largest value of 5.7039.

**Ans 6)**



1. (B)

**A)**

Validation misclassification rate for Neural model 5: 14.03%

Validation misclassification rate for Neural model 5 is different from neural model 1, 2, 3, 4(less.)

**B)** There are 43 parameters created by this model.

**Ans 7)**

**A)**

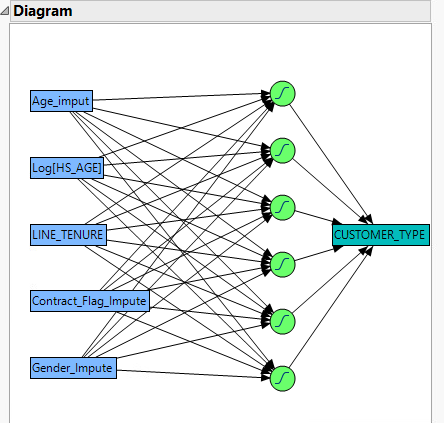
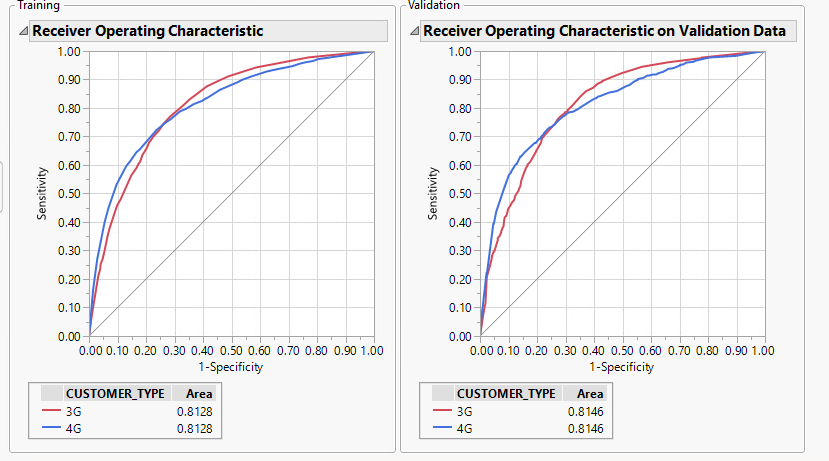
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Models | Number of Parameters | Training misclassification rate | training specificity | training sensitivity | validation misclassification | validation specificty | validation sensitivity |
| NEURAL NET 1 | 121 | 14.97% | 95.25% | 33.90% | 14.98% | 94.73% | 36.44% |
| NEURAL NET 2 | 37 | 14.63% | 95.60% | 34.4% | 14.59% | 94.9% | 37.90% |
| NEURAL NET 3 | 22 | 14.60% | 95.29% | 35.95% | 14.26% | 94.91% | 39.89% |
| NEURAL NET 4 | 34 | 14.43% | 95.14% | 37.71% | 14.17% | 94.87% | 40.67% |
| NEURAL NET 5 | 43 | 14.50% | 95.00% | 37.90% | 14.03% | 94.87% | 41.44% |
| Reg 1 | 14 | 15.31% | 98.1% | 17.57% | 15.43% | 97.64% | 19.22% |
| Reg 2 | 13 | 14.60% | 95.81% | 33.33% | 14.59% | 95.15% | 36.67% |
| Reg 3 | 18 | 15.23% | 98.08% | 18.23% | 15.31% | 97.62% | 20% |
| Decision tree | N/A | 14.15% | 96.64% | 31.8% | 13.98% | 96.57 | 33.22% |

**B)**

For this part, I have first run a MR model taking all of the input variables as independent, customer type as dependent variable and validation as validation with minimum AICC criteria, whole effect and forward selection. The variables which were selected in forward selection method, I put only them into neural network model as dependent variable and ran model. It tends to have misclassification rate of 14.22% which is less than neural network 5.

According to validation misclassification rate, neural network 5 model is the best model out of all neural models created here(including one with independent variables selected through forward method in MR model).

1. **NEURAL NETWROK DIAGRAM:** **2)ROC CURVE:**

Above Diagram shows there is 1 hidden layer with 6 nodes and Here, ROC curve has been given above for training and validation data set

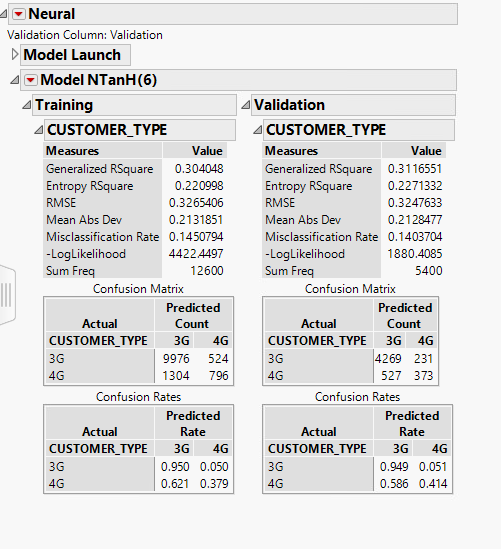
Each of the dependent variables are connected to each of the AUC for training and validation data set is 0.8128 and 0.8146 respectively.

Nodes are connected to target output(Customer\_Type). There which relates model performance to be excellent

Are total 5 dependent variables here. Here, TANH has

been used as transfer function for to calculate activation of nodes

3) **CONFUSION MATRIX, SENSITIVITY AND SPECIFICITY OF BEST NEURAL MODEL**.



In the above figure, we can see misclassification rate, and confusion matrix.Misclassification rate for training and validation data set is 14.5 and 14.03 respectively. Sensitivity of above model is 37.9% for training and 41.44% for validation. Specificity of above model is 95.00% for training and 94.87% for validation data respectively.

**C)**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Models | Number of Parameters | Training misclassification rate | training specificity | training sensitivity | validation misclassification | validation specificty | validation sensitivity | TRAINING AUC | VALIDATION AUC |
| NEURAL NET 1 | 121 | 14.97% | 95.25% | 33.90% | 14.98% | 94.73% | 36.44% | 0.81 | 0.8066 |
| NEURAL NET 2 | 37 | 14.63% | 95.55% | 34.43% | 14.59% | 94.91% | 37.89% | 0.8094 | 0.8107 |
| NEURAL NET 3 | 22 | 14.60% | 95.29% | 35.95% | 14.26% | 94.91% | 39.89% | 0.811 | 0.8149 |
| NEURAL NET 4 | 34 | 14.43% | 95.14% | 37.71% | 14.17% | 94.87% | 40.67% | 0.8126 | 0.8154 |
| NEURAL NET 5 | 43 | 14.50% | 95.00% | 37.90% | 14.03% | 94.87% | 41.44% | 0.8128 | 0.8146 |
| Reg 1 | 14 | 15.31% | 98.1% | 17.57% | 15.43% | 97.64% | 19.22% | 0.7927 | 0.7878 |
| Reg 2 | 13 | 14.60% | 95.81% | 33.33% | 14.59% | 95.15% | 36.67% | 0.8122 | 0.8101 |
| Reg 3 | 18 | 15.23% | 98.08% | 18.23% | 15.31% | 97.62% | 20% | 0.7929 | 0.7879 |
| Decision tree | N/A | 14.15% | 96.64% | 31.8% | 13.98% | 96.57 | 33.22% | N/A | N/A |

Here, in the above figure, all the different possible values for all the models are given above. Number of Parameters, Training AUC and validation AUC cannot be derived for decision tree model(not given). Put of all neural network models, neural network 5 has 2nd highest complexity(moderately complex with 43 parameters).

With respect to other model, neural network 5 tends to perform better. Only decision tree model has better validation and training misclassification rate( 13.92).

When Sensitivity % is concerned, neural network 5 works better both on training and validation data set with respect to all the regression models and neural models

When Specificity is concerned, all neural network(including neural net 5) has almost same specificity on both training and validation while those quantity tends to be lower with respect to all regression and decision tree model(slightly).

All neural network model has almost same AUC both on training and validation data as shown above which tends to show that this models shows excellent prediction. Same can be said for regression models.

This shows that when AUC and specificity criteria is concerned, neural net tends to behave same as other networks.