DATA MINING PROJECT

FALL 2019

Telcom Customer Churn Using SAS Enterprise Miner

By:
Karan Gupta

INTRODUCTION:

In today's competitive environment Organizations of all sizes must become more efficient than ever, particularly when using all available data to drive new revenue and growth. But greater organizational effectiveness isn't just about the efficiency ratio. As a business owner, it is imperative to leverage data analysis to improve the customer experience.

Improving the customer experience requires a wholesome understanding of your customers and relating to them in ways that they understand. This can only be achieved by having 360-degree vantage over your customers which in turn requires leveraging the gold mine of data available to you today, including and not limited to:

- Core customer information (including contact, KYC and location data)
- Additional experiential customer information (gathered from all stages of the customer lifecycle)
- Transaction information (including checking, savings and credit card transactions; loan draws and repayments; investment positions and balances)

Telecom companies are overflowing with data but harnessing and leveraging that organizational data for more effective customer experience has always been a challenge. In today's data driven market, it's more important than ever that you understand your customer, your products, your channels and your pricing – all to ensure you're tailoring product offerings to your customers' changing needs and maximizing ROI for customers as well increasing the revenue.

DOMAIN & GOAL OF THE PROJECT:

The project is based on a private Telecom company dataset available on the Kaggle website.

The goal of the project is "Predict behaviour to retain customers and analyse all relevant customer data and develop focused customer retention programs."

Dataset Description:

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

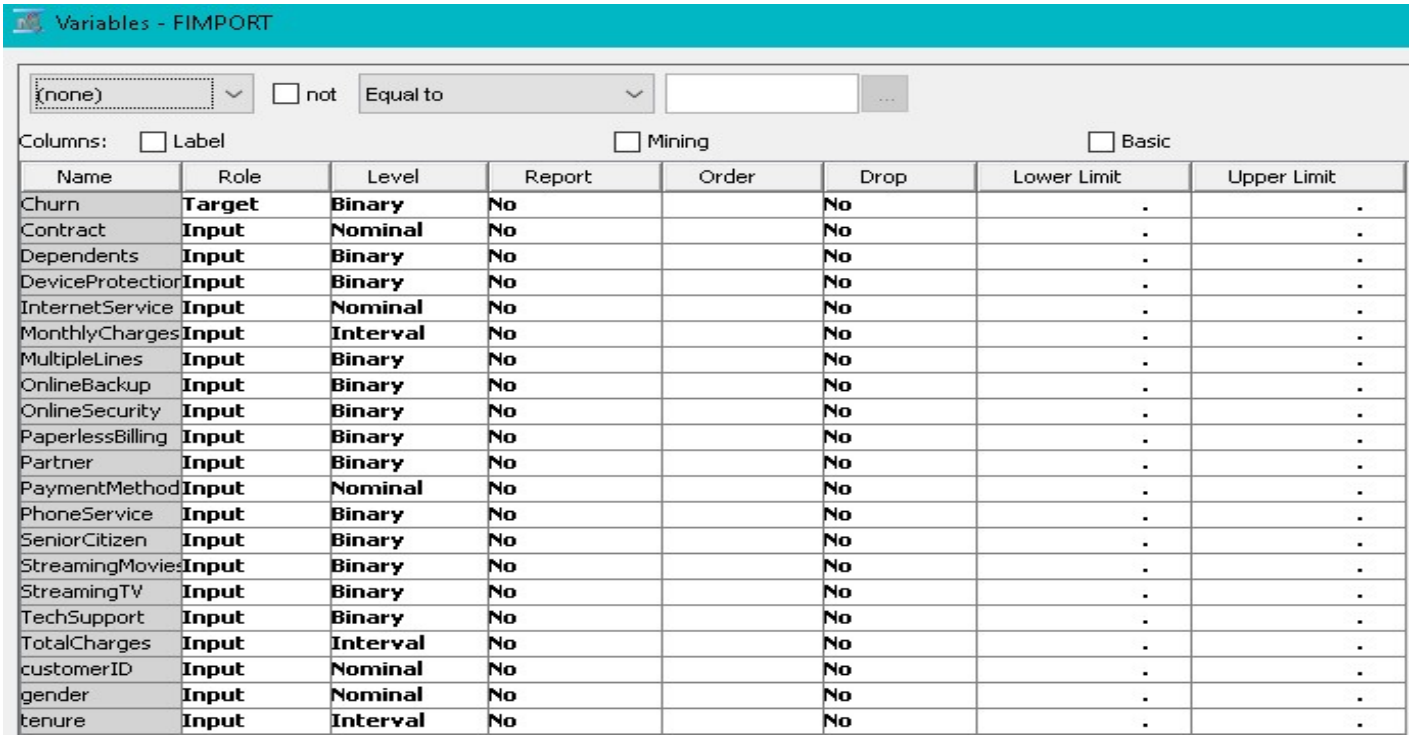
The raw data contains 7043 rows (customers) and 21 columns (features).

The "Churn" column is our target.

COLUMNS

- **customerID**
- **gender:** Whether the customer is a male or a female
- **SeniorCitizen:** Whether the customer is a senior citizen or not (1, 0)
- **Partner:** Whether the customer has a partner or not (Yes, No)
- **Dependents:** Whether the customer has dependents or not (Yes, No)
- **Tenure:** Number of months the customer has stayed with the company
- **PhoneService:** Whether the customer has a phone service or not (Yes, No)
- **MultipleLines:** Whether the customer has multiple lines or not (Yes, No, No phone service)
- **InternetService:** Customer's internet service provider (DSL, Fiber optic, No)
- **OnlineBackup:** Whether the customer has online backup or not (Yes, No, No internet service)
- **DeviceProtection:** Whether the customer has device protection or not (Yes, No, No internet service)
- **TechSupport:** Whether the customer has tech support or not (Yes, No, No internet service)
- **StreamingTV:** Whether the customer has streaming TV or not (Yes, No, No internet service)
- **StreamingMovies:** Whether the customer has streaming movies or not (Yes, No, No internet service)
- **Contract:** The contract term of the customer (Month-to-month, One year, Two year)
- **PaperlessBilling:** Whether the customer has paperless billing or not (Yes, No)
- **PaymentMethod:** The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
- **MonthlyCharges:** The amount charged to the customer monthly
- **TotalCharges:** The total amount charged to the customer
- **Churn:** Whether the customer churned or not (Yes or No)

After loading the dataset, the **Roles** of the variables are being assigned where the **Churn** variable is our "Target" variable and the rest are the input variables. Furthermore, the **Level** the variables are selected where we chose from the Binary, Nominal, Interval, Unary, etc depending upon the variable data type.



Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Churn	Target	Binary	No		No	.	.
Contract	Input	Nominal	No		No	.	.
Dependents	Input	Binary	No		No	.	.
DeviceProtectior	Input	Binary	No		No	.	.
InternetService	Input	Nominal	No		No	.	.
MonthlyCharges	Input	Interval	No		No	.	.
MultipleLines	Input	Binary	No		No	.	.
OnlineBackup	Input	Binary	No		No	.	.
OnlineSecurity	Input	Binary	No		No	.	.
PaperlessBilling	Input	Binary	No		No	.	.
Partner	Input	Binary	No		No	.	.
PaymentMethod	Input	Nominal	No		No	.	.
PhoneService	Input	Binary	No		No	.	.
SeniorCitizen	Input	Binary	No		No	.	.
StreamingMovies	Input	Binary	No		No	.	.
StreamingTV	Input	Binary	No		No	.	.
TechSupport	Input	Binary	No		No	.	.
TotalCharges	Input	Interval	No		No	.	.
customerID	Input	Nominal	No		No	.	.
gender	Input	Nominal	No		No	.	.
tenure	Input	Interval	No		No	.	.

DATA PREPROCESSING

The data pre-processing mainly included:

i) All the "Yes" and "No" values were replaced by 1 and 0 respectively to make the desired data set variables in binary form.

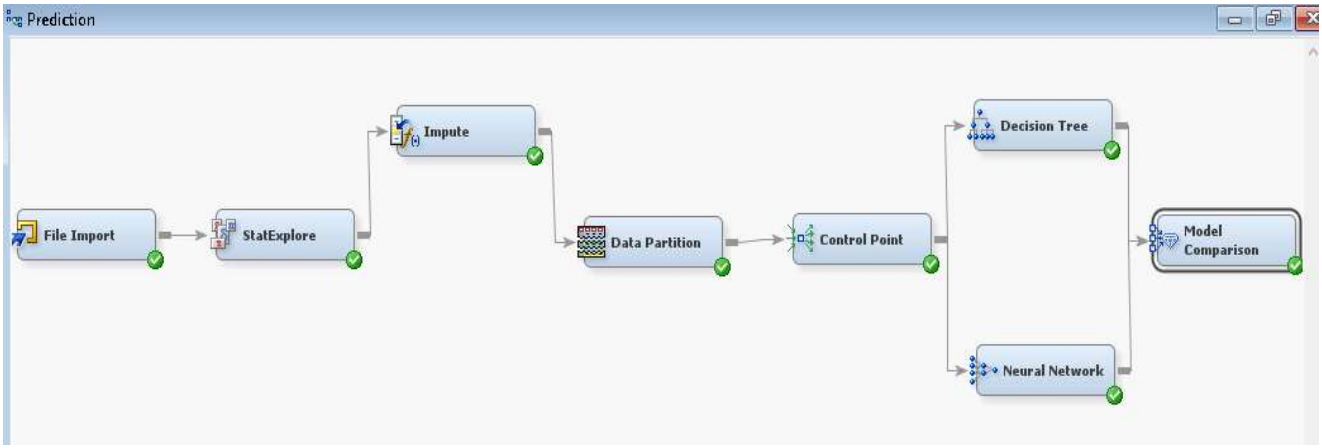
ii) Also, all the missing were replaced by the mean of the variable through Imputation method in the SAS Enterprise Miner.

Evaluation Metric for Model Analysis:

The **Churn** variable is a two-class output (either 0 / 1) i.e. binary variable so **Misclassification Rate** is chosen as the evaluation metric for the predictive model since the target variable is binary**.**

I chose to stick with a 60:20:20 split of Training, Validation and Testing data for building the predictive models.

Predictive Model Workflow:



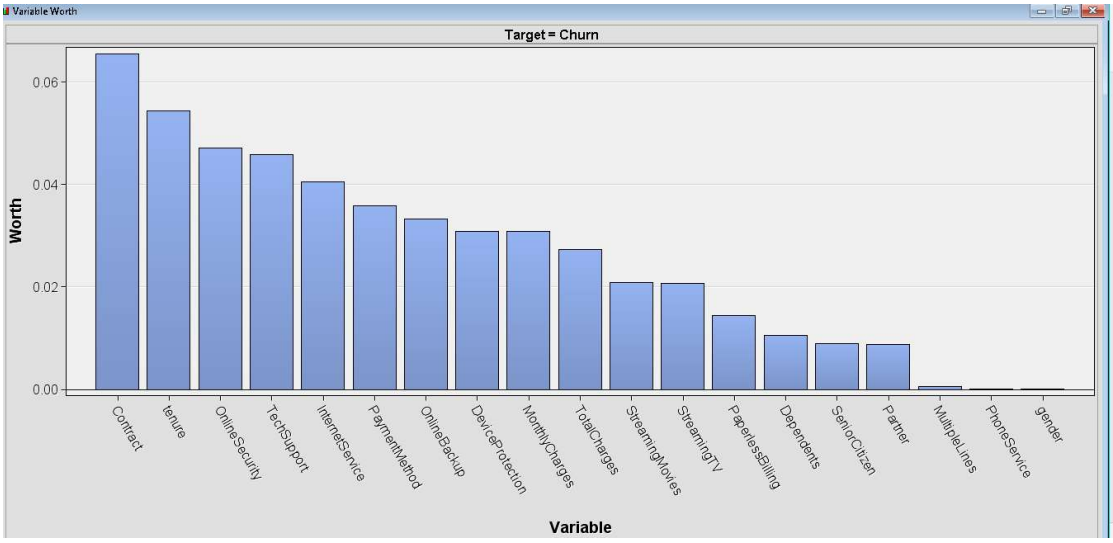
Feature selection for Predictive Model Build:

I have built a workflow, that predicts whether a customer will stay with the current telecom company or not with the help of comparing two models which are decision tree and neural network.

Predicting whether a customer will stay with the company or not
TARGET VARIABLE – churn

After the initial data cleaning and pre-processing, the final dataset was loaded into SAS Enterprise Miner, for feature analysis using the **StatExplore** module.

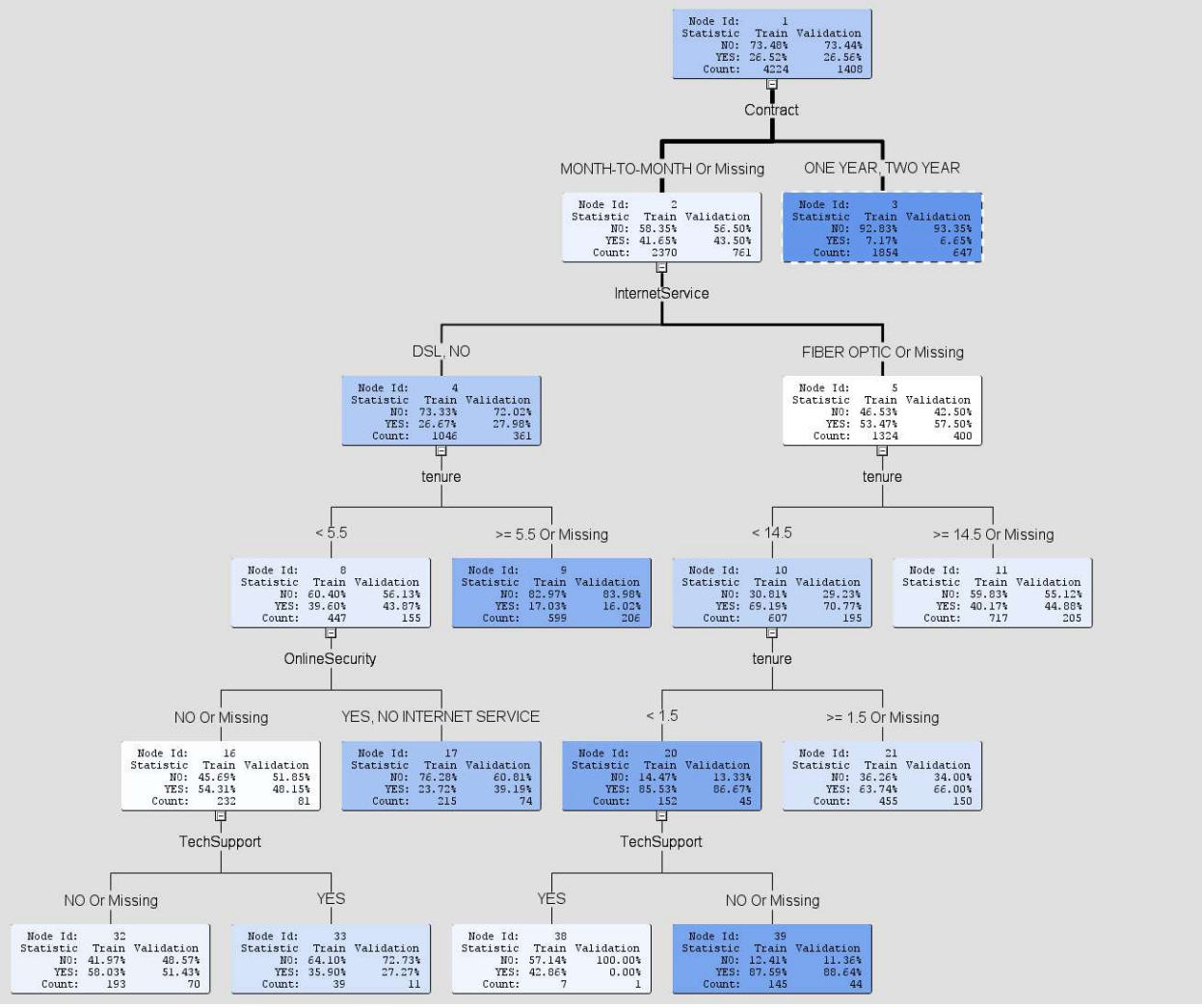
It is clearly evident from the StatExplore that the variables gender, PhoneService and MultipleLines are insignificant and do not add any value to the predictive model. Hence, these values are rejected and only 18 columns are taken out of 21.



DECISION TREE:

An empirical tree represents a segmentation of the data that is created by applying a series of simple rules. Each rule assigns an observation to a segment based on the value of one input. One rule is applied after another, resulting in a hierarchy of segments within segments. The hierarchy is called a tree, and each segment is called a node. The original segment contains the entire data set and is called the root node of the tree. A node with all its successors forms a branch of the node that created it. The final nodes are called leaves. For each leaf, a decision is made and applied to all observations in the leaf. The type of decision depends on the context. In predictive modeling, the decision is the predicted value.

The inherent advantage of a decision tree is that it provides a visualization of the algorithm and the rules on which the data have been split. The tree diagram can be seen below,



NEURAL NETWORK:

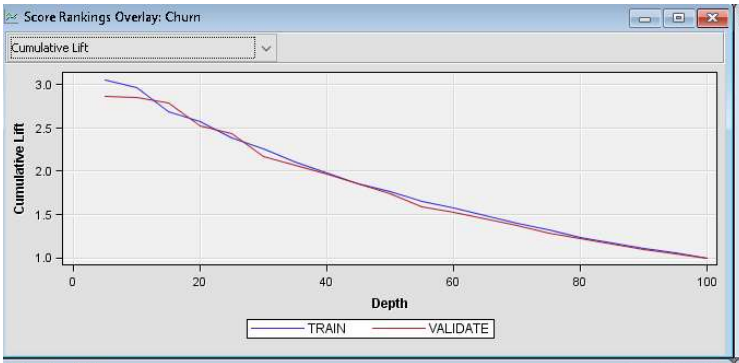
I have also applied Neural Network model so that we can compare the results with the Decision Tree and see which model has performed better.

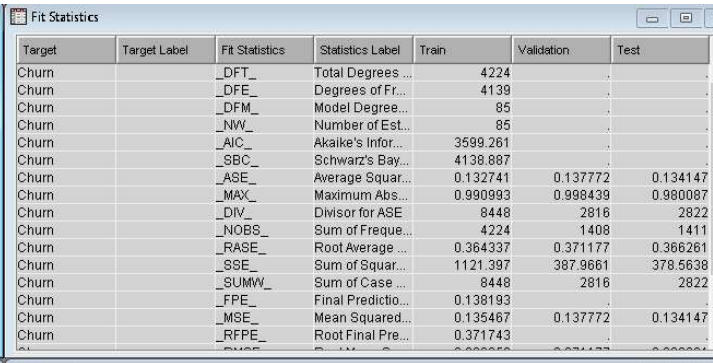
Neural networks are especially useful for prediction problems where:

- no mathematical formula is known that relates inputs to outputs.
- prediction is more important than explanation.
- there is a lot of training data.

Common applications of neural networks include credit risk assessment, direct marketing, and sales prediction.
The Neural Network node provides a variety of feedforward networks that are commonly called **backpropagation or backprop networks.**
Strictly speaking, backpropagation refers to the method for computing the error gradient for a feedforward network, a straightforward application of the chain rule of elementary calculus. By extension, backprop refers to various training methods that use backpropagation to compute the gradient. By further extension, a backprop network is a feedforward network trained by any of various gradient-descent techniques. Standard backprop is also one of the most difficult to use, tedious, and unreliable training methods. Unlike the other training methods in the Neural Network node, standard backprop comes in two varieties.

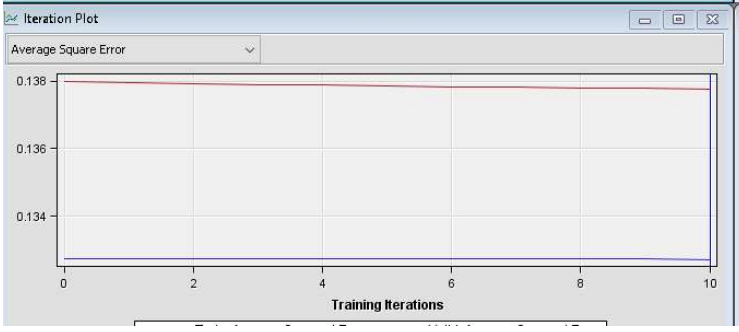
- Batch backprop, like conventional optimization techniques, reads the entire data set, updates the weights, reads the entire data set, updates the weights, and so on.
- Incremental backprop reads one case, updates the weights, reads one case, updates the weights, and so on.

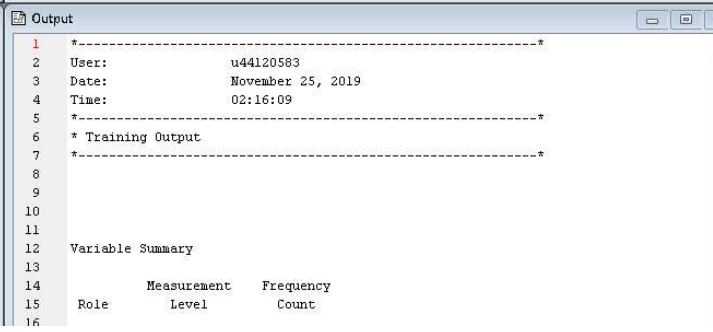




Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		_DFT_	Total Degrees ...	4224	.	.
Churn		_DFE_	Degrees of Fr...	4139	.	.
Churn		_DFM_	Model Degree...	85	.	.
Churn		_NW_	Number of Est...	85	.	.
Churn		_AIC_	Akaike's Infor...	3599.261	.	.
Churn		_SBC_	Schwarz's Bay...	4138.887	.	.
Churn		_ASE_	Average Squar...	0.132741	0.137772	0.134147
Churn		_MAX_	Maximum Abs...	0.990993	0.998439	0.980087
Churn		_DIV_	Divisor for ASE	8448	2816	2822
Churn		_NOBS_	Sum of Freque...	4224	1408	1411
Churn		_RASE_	Root Average ...	0.364337	0.371177	0.366261
Churn		_SSE_	Sum of Squar...	1121.397	387.9661	378.5638
Churn		_SUMW_	Sum of Case ...	8448	2816	2822
Churn		_FPE_	Final Predictio...	0.138193	.	.
Churn		_MSE_	Mean Squared...	0.135467	0.137772	0.134147
Churn		_RFPE_	Root Final Pre...	0.371743	.	.



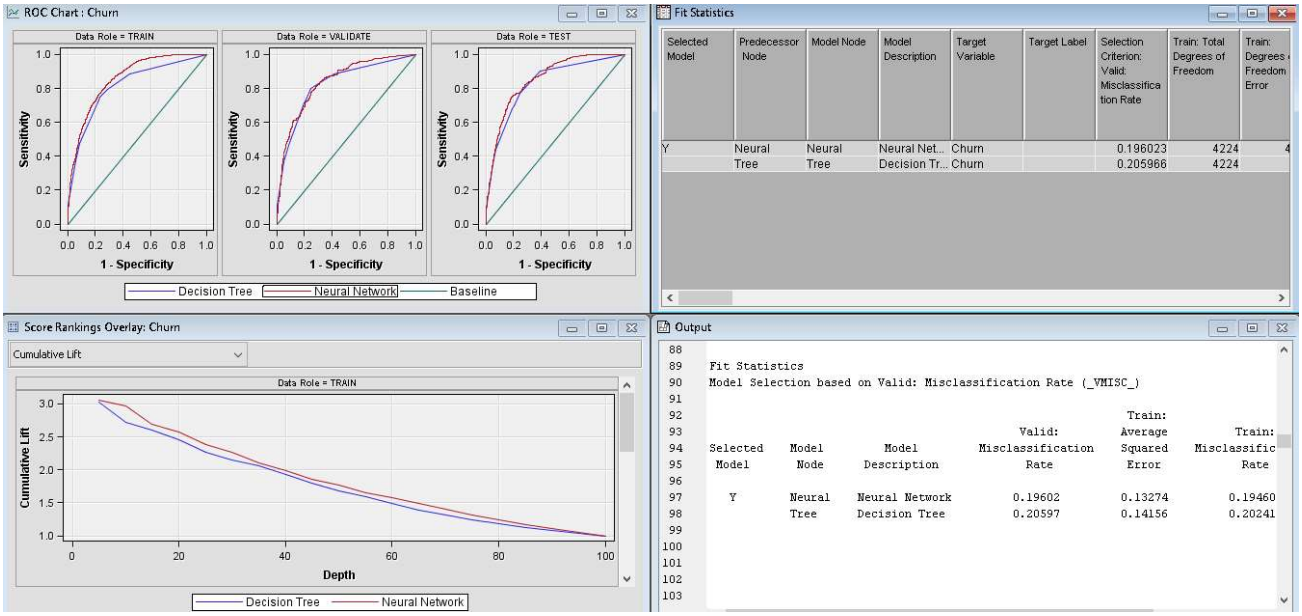


Output

```
*------------------------------------------------------------*
User:                  u44120583
Date:                  November 25, 2019
Time:                  02:16:09
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*



Variable Summary

             Measurement    Frequency
 Role          Level         Count
```

MODEL COMPARISON:

The image below represents the ROC, AUC and evaluation metrics for the two models. From the results of the fit statistics, it is obvious that the Neural Network is the best performer with an accuracy of approximately 87% and Misclassification Rate of approx. 0.19. Also, the Average Squared error of the Neural Network model is 0.13.



REFERENCES:

1. SAS Enterprise miner - Reference help: https://documentation.sas.com/
2. Dataset Reference - https://www.kaggle.com/blastchar/telco-customer-churn/data#