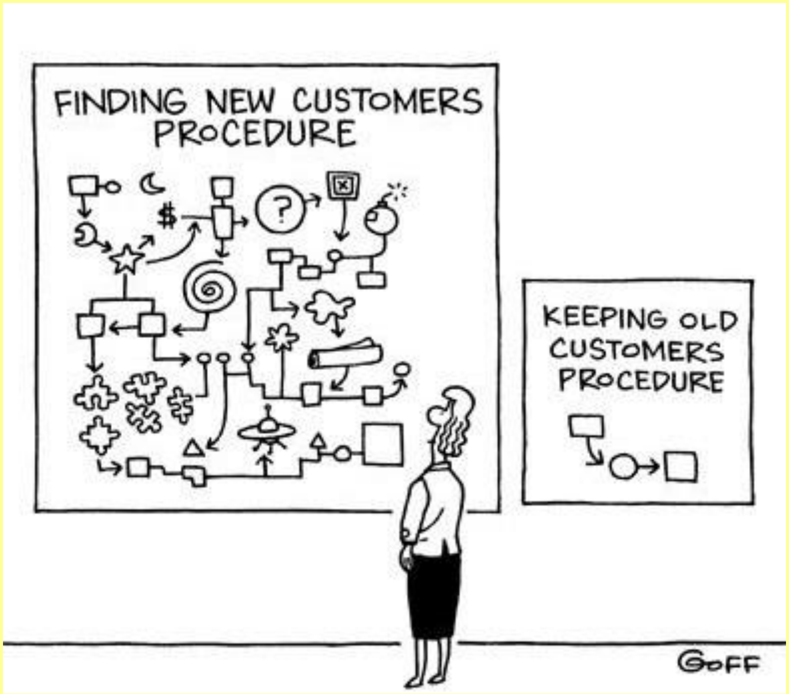


[1]

Prediction of Customer Attrition in Telecommunication

MIS 6324.001 – Spring 2019

“Since it’s so expensive to gain new customers today, what can you do to establish an unbreakable bond?”
– Chet Holmes

Group #1

- Anand Muraleedharan
- Jagruti Wagh
- Neethu Narayanan
- Srishty Kshirsagar
- Namrata Patil

TABLE OF CONTENTS

Contents

EXECUTIVE SUMMARY

Customer attrition, also known as customer churn, customer turnover, or customer defection, is the loss of clients or customers. [2] Customer attrition can be attributed to several factors such as difficult user experience, issues with the product/ service, better products offered by competitors, etc. But this impacts the revenue of the company and if there is no way to identify or minimize the attrition, then the company will probably go into bankruptcy. Also, the retention of an existing customer is much more cost effective as opposed to acquiring new customers.

As per Alex Scroxton of ComputersWeekly.com, "Canada's BCE and Telus revealed in 2017 that it cost almost 50 times less for them to keep an existing mobile customer than to acquire a new one, with retention costs of CAD11.04 and CAD11.74 respectively, while average subscriber acquisition cost weighed in at an eye watering CAD521."

Usually, attrition can be of two types:

- **Voluntary**: Here, the customer satisfaction is low with regards to the product and services standard provided by the company. So, he moves on to a competitor.
- **Involuntary**: This type of churn happens because of situational factors like relocation, death etc. There is no way the company can influence this.

When it comes to using data to understand customer churn, there are two general approaches: driver analysis and predictive modelling. Driver analysis can be as simple as correlation analysis and as sophisticated as multiple regression – but either way, the output is the same: the identification and relative strength of leading indicators (independent variables) of customer churn. This type of output is referred to as top-down: it tells you in general and in aggregate what things contribute to customer flight. It is great for high-level strategic planning because an organization can easily prioritize what to focus on and where to allocate resources.

But unfortunately, we do not live in a one-size-fits-all world where people all act the same. And in our highly fragmented, personalized and digitized society, a top-down approach to understanding customer behaviour does not provide individual customer-specific data. For this, we need predictive modelling. A predictive modelling approach to customer churn provides scores of likelihoods to leave at an individual customer level. Because values are attached to each customer, predictive modelling can be referred to as bottom-up. With bottom-up data, you can specifically identify the risk of churn for every single customer. With data like this, you can then target your customer responses at an individual level. [3]

The objective is to predict customer attrition/churn of telecom customers. The churn/attrition is usually an instrumental factor in determining future business strategies. This project is intended at building a predictive model that can identify whether a customer will discontinue the product/services of the company with some confidence. However, the model cannot predict if the attrition is voluntary or involuntary.

The company must identify the probable customer who are about to churn and provide them with lucrative offers for churn prevention. It is cost effective investment for the company to provide offers as compared to the investment in acquiring new customers.

PROJECT MOTIVATION

Customer churn is one of the mounting issues of today's rapidly growing and competitive telecom sector. The focus of the telecom sector has shifted from acquiring new customers to retaining existing customers due to the associated high cost. The retention of existing customers also leads to improved sales and reduced marketing cost as compared to new customers. These facts have ultimately resulted in customer churn prediction activity to be an indispensable part of telecom sector's strategic decision making and planning process. Customer retention is one of the main objectives of Customer Relationship Management (CRM) and its importance has led to the development of various tools that support some important tasks in predictive modelling and classification. In recent decades, organizations are increasingly focusing on long-term relationships with their customers and observing a customer's behaviour from time to time. They use various applied knowledge discovery in database (KDD) techniques to extract hidden relationships between different entities and attributes in a flood of data banks. These facts have attracted many companies to invest in CRM to maintain customer information. Customer centric approach is very common, particularly, in the telecommunication sector for predicting customers' behaviour based on historical data stored in CRM. To handle the mounting issue of customer churn, data maintained in such CRM systems can be converted into meaningful information that will help to identify customer's churn activities before the customers are lost; thereby, increasing customer strength. [4]

It is of substantial interest to both academic researchers and industry practitioners, interested in forecasting the behaviour of customers in order to differentiate the churn from non-churn customers. If these businesses want to keep growing profits, they will have to focus on and identify the value of each customer in their current database. Their success will lie in how effectively they are able to leverage the large sets of operational and customer data at their disposal to derive recommendations for those customers who may choose to leave them. As Microsoft founder Bill Gates says, “Your most unhappy customers are your greatest source of learning.” This project focuses on this very idea and aims to predict with some confidence, the likelihood of customer disloyalty. It will help organizations utilize their customer data to target at-risk customers.

Our goal is to give the company a targeted set of actions to improve customer retention. Once an accurate model has been trained, all of your past and present customers will be run through the predictive model. If the prediction matches the actual status of the customer, then there is no action necessary. However, if an active customer is predicted to be a churned customer, this will imply that this customer is at-risk of churning. Alternatively, if a customer is currently inactive (meaning that they previously churned), but the prediction is that they are an active customer, this may indicate that they are a good target with which to attempt to reinstate business.

We show how attribute-level analysis can pave the way for developing a successful customer retention policy that could form an indispensable part of strategic decision making and planning process in the telecom sector.

DATA DESCRIPTION

- The dataset is **2nd party** and is available in Kaggle [5].
- The dataset contains attrition data for telecommunication sector.
- No of rows: **7043**
- Fields: **21**

The data set provides 15+ independent variables (predictors) to predict the binary target variable (churn).

The dataset fields are as below:

1) **Customer ID**: unique identifier for Customer

Predictors

2) **Gender:** Whether the customer is a male or a female
3) **Senior Citizen**: Whether the customer is a senior citizen or not (1, 0)
4) **Partner**: Whether the customer has a partner or not (Yes, No)
5) **Dependents**: Whether the customer has dependents or not (Yes, No)
6) **Tenure**: Number of months the customer has stayed with the company
7) **Phone Service:** Whether the customer has a phone service or not (Yes, No)
8) **Multiple Lines:** Whether the customer has multiple lines or not (Yes, No, No phone service)
9) **Internet Service**: Customer's internet service provider (DSL, Fibre optic, No)
10) **Online Security**: Whether the customer has online security or not (Yes, No, No internet service)
11) **Online Backup**: Whether the customer has online backup or not (Yes, No, No internet service)
12) **Device Protection**: Whether the customer has device protection or not (Yes, No, No internet service)
13) **Tech Support**: Whether the customer has tech support or not (Yes, No, No internet service)
14) **Streaming TV**: Whether the customer has streaming TV or not (Yes, No, No internet service)

15) **Streaming Movies**: Whether the customer has streaming movies or not (Yes, No, No internet service)
16) **Contract:** The contract term of the customer (Month-to-month, One year, Two year)
17) **Paperless Billing**: Whether the customer has paperless billing or not (Yes, No)
18) **Payment Method**: The customer's payment method (Electronic check, mailed check, Bank transfer (automatic), Credit card (automatic))
19) **Monthly Charges:** The amount charged to the customer monthly
20) **Total Charges**: The total amount charged to the customer

Target Variable

21) **Churn**: Whether the customer churned or not (Yes or No)

BI MODEL - SELECTION

Predictive BI models are used to address the following types of prediction problems:

a. Classification
b. Ranking
c. Estimation

Churning is a classic classification problem where the companies try to predict the probability of customer/employee attrition probabilities. The various predictor variables in the dataset are used to figure out the pattern of churn amongst the customers. Then, a generic BI model can be built by training using the dataset. The model will then be able predict the chances of churning given a new customer record.

The various classification approaches that can be used are:

- Decision Tree
- Logistic Regression
- Neural Network

Usually, the net worth of a customer can vary, and the customers might fall into various classes or segments based on the net worth. A Clustering Analysis is suitable to identify the various segments that exist within a company. This will allow the company to predict the impact if a particular customer from a segment churn.

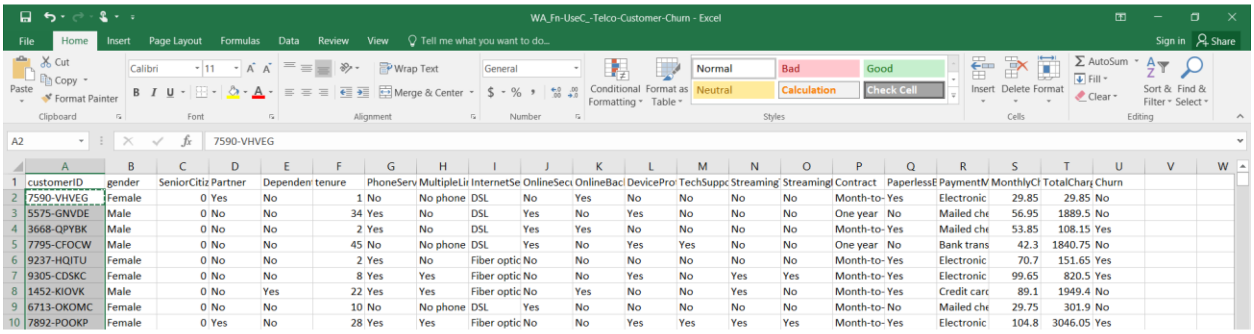
So, the dataset will be processed with four algorithms. Three of them fall into supervised learning while the other falls into unsupervised learning. The unsupervised learning/descriptive analysis is intended at adding better insights and hence better decisions. The customer churning is predicted using the above listed supervised algorithms.

DATA PREPARATION

The initial step is to import the external data set into SAS Enterprise Miner and then exploring the variables to infer if there is a need to impute or filter the variables.

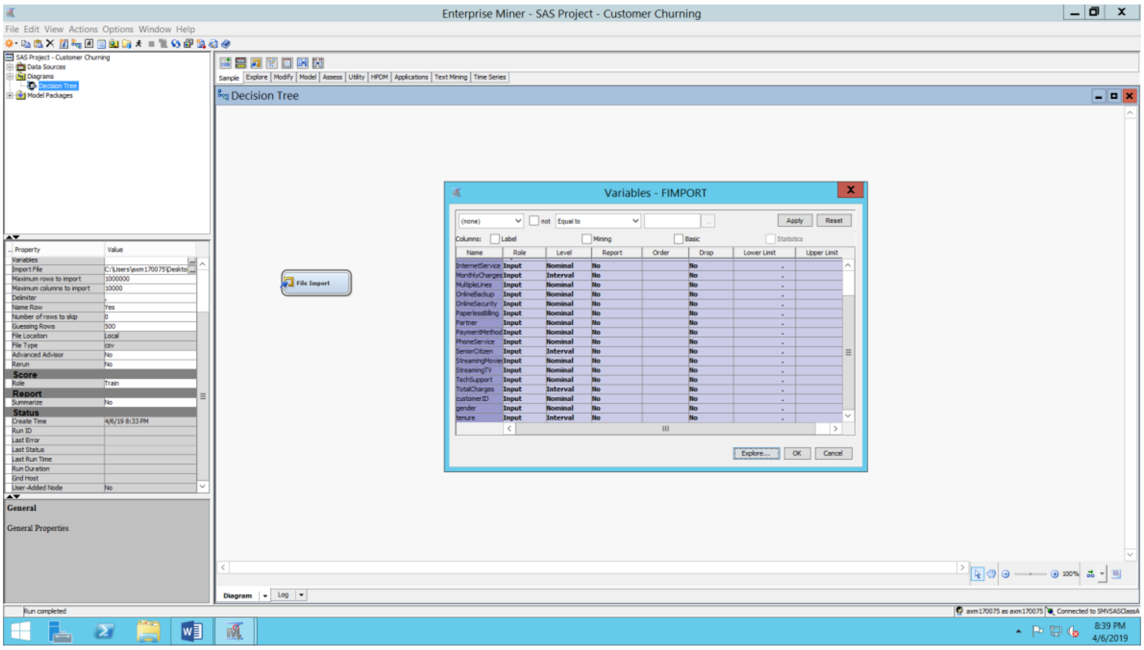
Import and Explore

An important analysis was to figure out if a customer has **unique records** or multiple records. Excel was used for this and a duplicate value check was conducted on the *customerID* field.



It was figured out that the field is unique and most probably might be the **primary key** of the dataset.

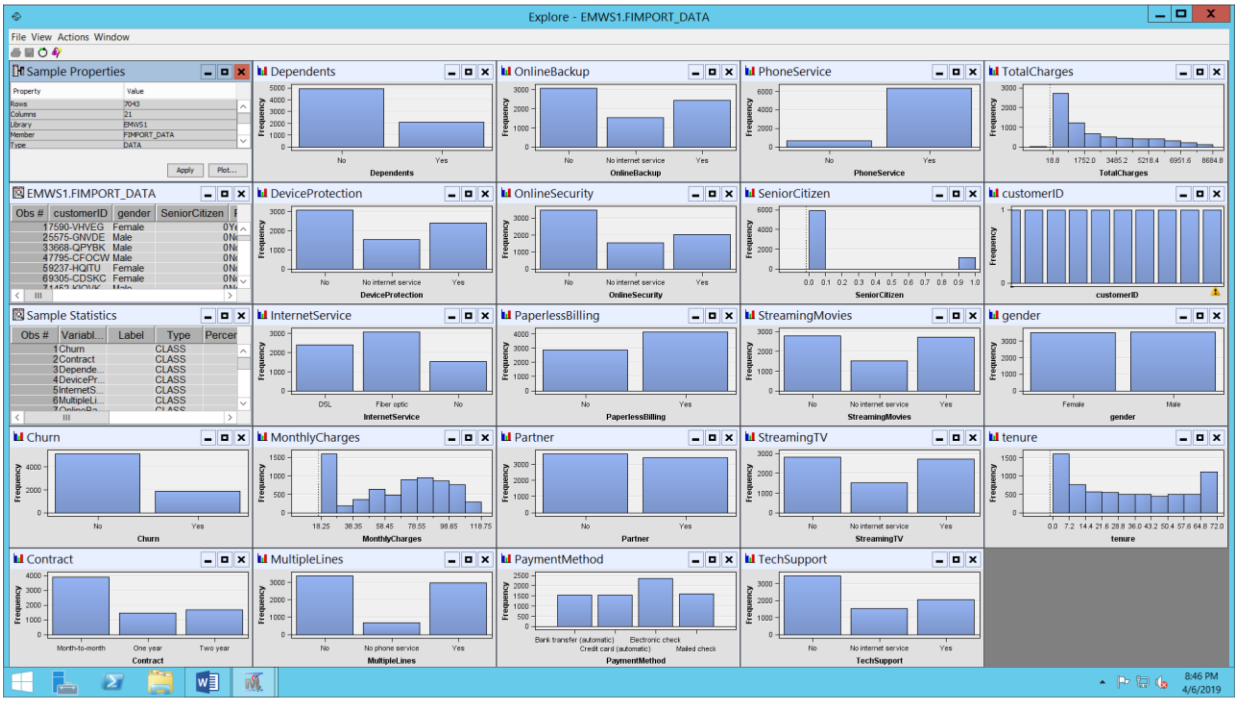
As a subsequent step, the data was imported to the SAS Enterprise Miner tool using the File Import tool.



Subsequently, the variable distribution has to be looked at to make sure if there is a need to:

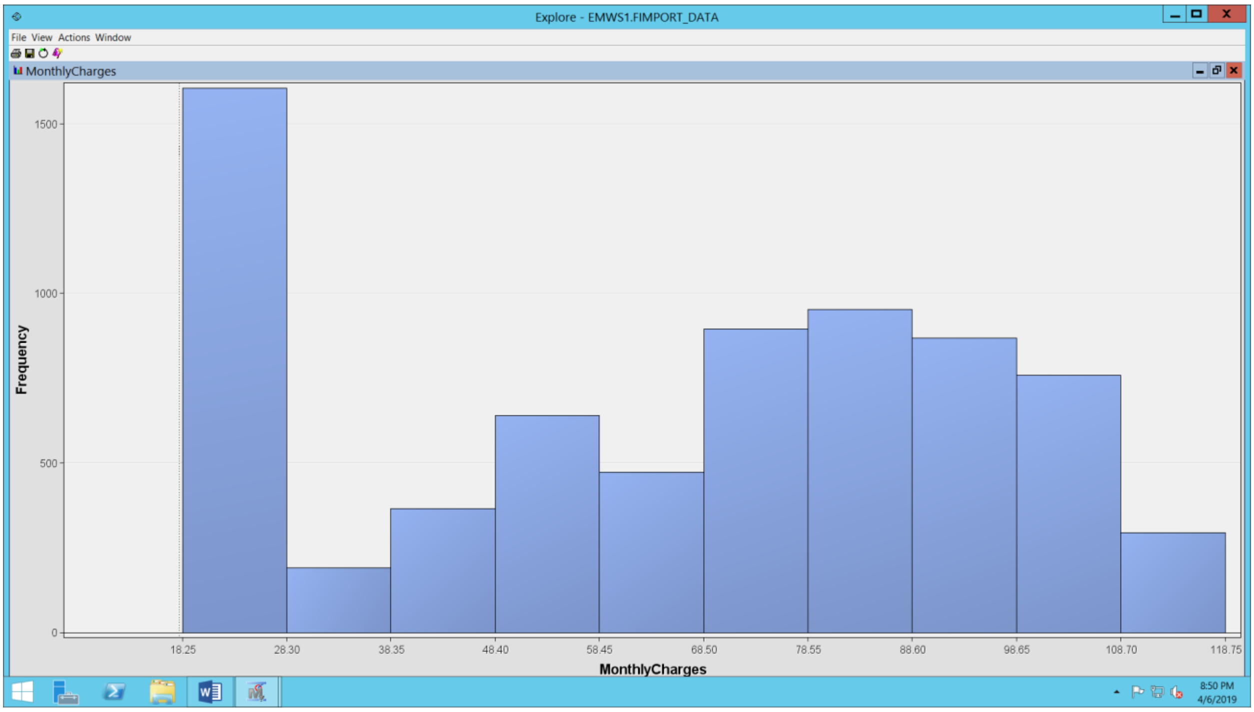
1. Impute
2. Replace
3. Filter

The exploration of variables looks like below:



The efforts were next channelized to analyze the variable value distribution. The following observations were made:

- Monthly charges had a spike in the range $18.25 – $28.30



- The overall statistics looked as below:

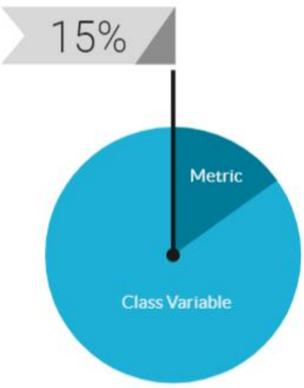
Sample Statistics

Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode Percent...	Mode
1	Churn		CLASS	0	.	.	.	2	73.46301	NO
2	Contract		CLASS	0	.	.	.	3	55.01917	MONTH-TO-MONTH
3	Dependents		CLASS	0	.	.	.	2	70.04118	NO
4	DeviceProtection		CLASS	0	.	.	.	3	43.94434	NO
5	InternetService		CLASS	0	.	.	.	3	43.95854	FIBER OPTIC
6	MultipleLines		CLASS	0	.	.	.	3	48.1329	NO
7	OnlineBackup		CLASS	0	.	.	.	3	43.84495	NO
8	OnlineSecurity		CLASS	0	.	.	.	3	49.66634	NO
9	PaperlessBilling		CLASS	0	.	.	.	2	59.22192	YES
10	Partner		CLASS	0	.	.	.	2	51.69672	NO
11	PaymentMethod		CLASS	0	.	.	.	4	33.57944	ELECTRONIC CHECK
12	PhoneService		CLASS	0	.	.	.	2	90.31663	YES
13	StreamingMovies		CLASS	0	.	.	.	3	39.54281	NO
14	StreamingTV		CLASS	0	.	.	.	3	39.89777	NO
15	TechSupport		CLASS	0	.	.	.	3	49.31137	NO
16	customerID		CLASS	0	.	.	.	128+	0.775194	0191-ZHSKZ
17	gender		CLASS	0	.	.	.	2	50.47565	MALE
18	MonthlyCharges		VAR	0	18.25	118.75	64.76169	.	.	
19	SeniorCitizen		VAR	0	0	1	0.162147	.	.	
20	TotalCharges		VAR	0.156183	18.8	8684.8	2283.3	.	.	
21	tenure		VAR	0	0	72	32.37115	.	.	

- Out of the 21 variables present in the dataset – 18 are class variables and 3 are metrics. The SAS tool mistakenly has classified *SeniorCitizen* as a metric instead of a class variable. This must be modified. The variable is edited and marked as a class variable.

SeniorCitizen	**Input**	**Binary**	**No**		**No**

The distribution of variable is as seen below:



- The variables do not have any missing value. This eliminates the need to have a **replacement/filter node**. There is no need to impute.

Assumptions

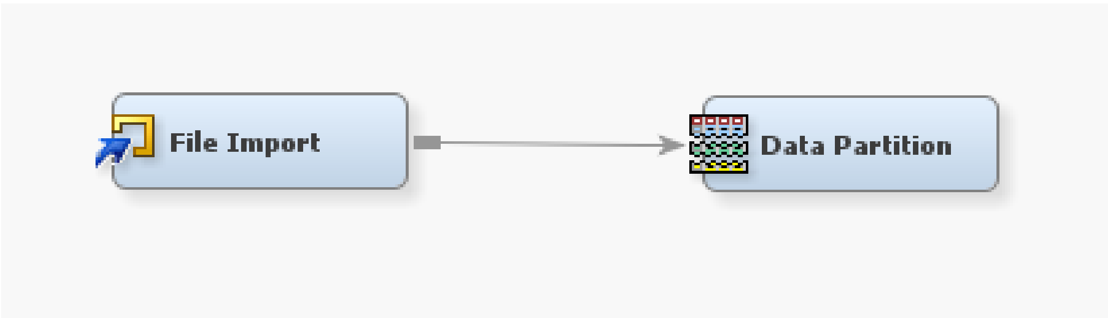
1. Data set is not merged with other datasets for improved insights.
2. Tenure has zero values. It is assumed that they have not completed a month yet.
3. Senior Citizen is a Boolean flag. It takes the values of either 0 or 1.
4. Total charges are 0 despite there are monthly charges. At a closer look, it is seen that the tenure for these entries are 0. The bill has not been generated, and the total value column is yet to be filled.

Data Partition

The objective of the project is to create a **stable predictive model** that can predict the customers who might churn. So, it is intended to be used most of the dataset to either train or validate. Hence only 10% of data is allocated for testing. The partitioning is as below:

Data Set Allocations	
Training	50.0
Validation	40.0
Test	10.0
Report	
Interval Targets	Yes
Class Targets	Yes

The status of the diagram is as below:



ENTERPRISE MINER PROCEDURE AND DIAGRAMS

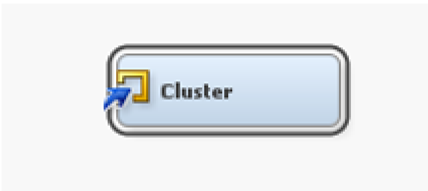
The four models – clustering, logistic regression, neural networks, and decision trees are used to conduct study on the Telco dataset. The preliminary analysis is done with Clustering and then the report delves into supervised models like Logistic regression, neural networks, and decision trees.

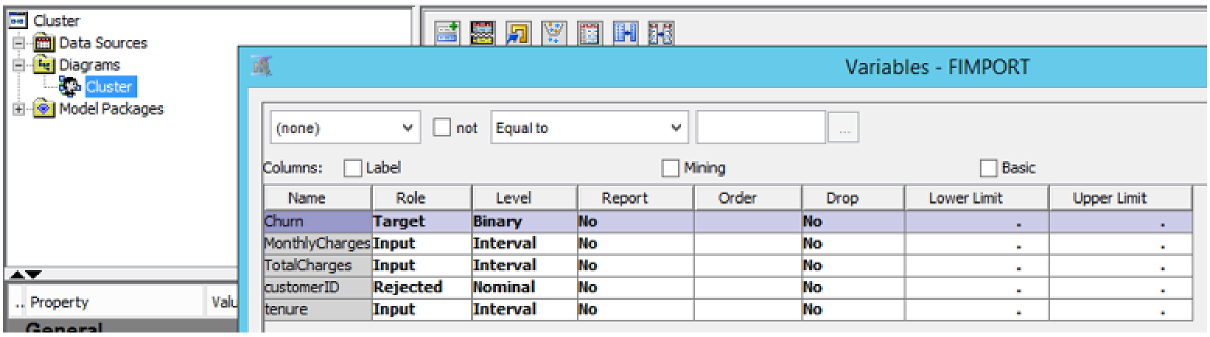
CLUSTER ANALYSIS

- Cluster analysis is an unsupervised data mining technique that is applied to determine distinct groups of customers and use this knowledge to develop targeted marketing strategies to generate increased revenue.
- Application of cluster analysis to our telecom dataset will help us in identifying unique segments of customers, the factors that define each segment and how to leverage this knowledge for customer retention.
- Cluster analysis requires interval data and the columns in the dataset that are of this type are – *Tenure*, *MonthlyCharges*, and *TotalCharges*.

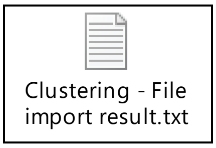
Let us perform cluster analysis in SAS EM by taking into consideration these three columns.

Step 1:

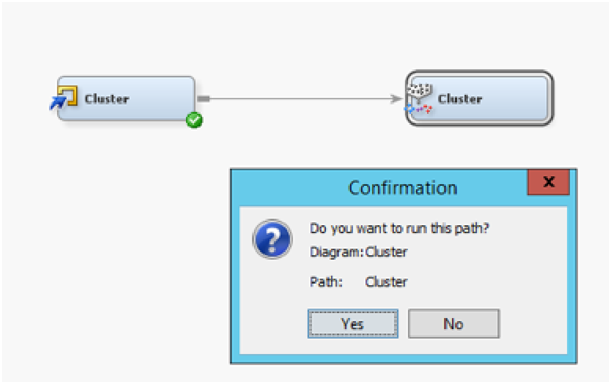




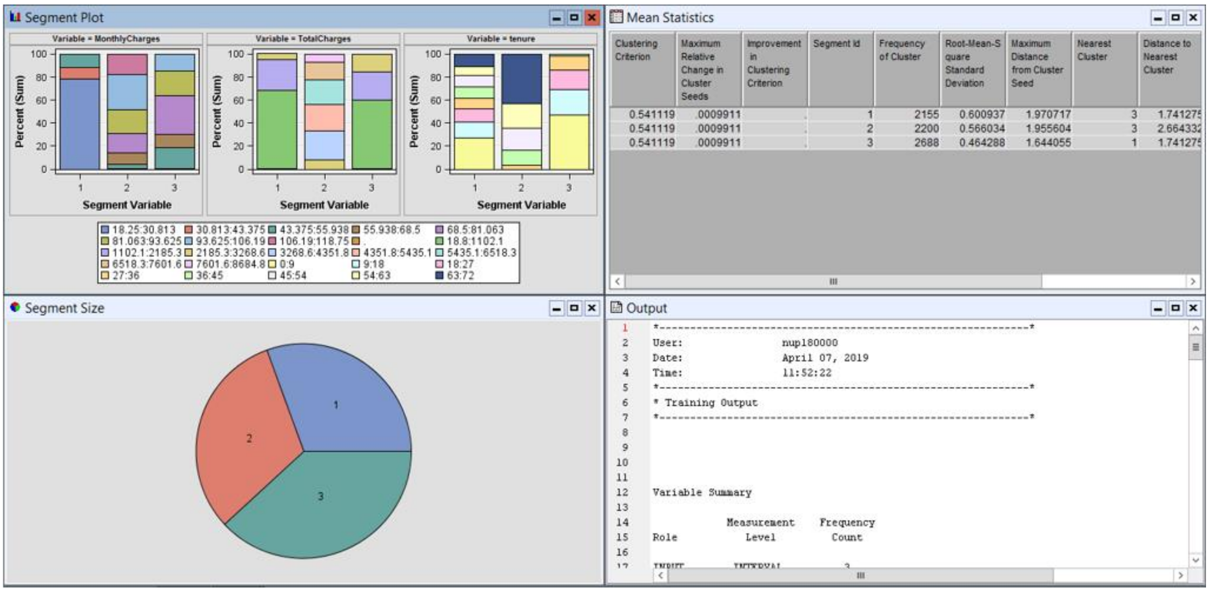
The clustering import results are as seen below:



The workspace diagram looks as below. The next step is to execute the clustering:



The segmentation results are as seen below:



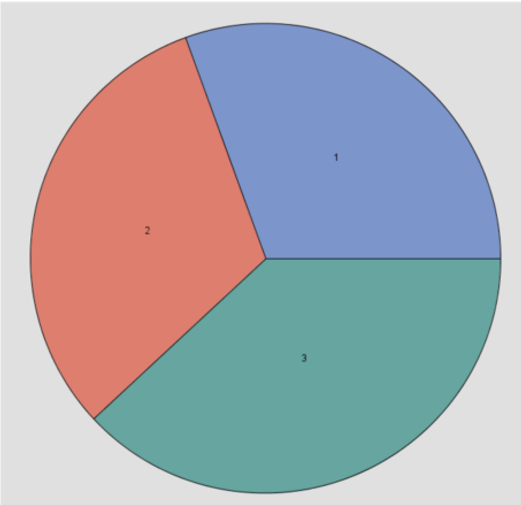
The Clustering output file is embedded below for the reference of the reader:

From the clustering results below, it is seen that three clusters are formed that are distinctly different from each other i.e., the inter-cluster distance is high and the intra-cluster difference is low. The mean statistics of the clusters are as seen below.

Mean Statistics

Clustering Criterion	Maximum Relative Change in Cluster Seeds	Improvement in Clustering Criterion	Segment Id	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	MonthlyCharges	TotalCharges	tenure
0.541119	.0009911	.	1	2155	0.600937	1.970717	3	1.741275	26.58116	811.475	29.48817
0.541119	.0009911	.	2	2200	0.566034	1.955604	3	2.664332	89.69793	5246.126	58.55955
0.541119	.0009911	.	3	2688	0.464288	1.644055	1	1.741275	74.96233	1032.736	13.24851

The cluster pie chart is as seen below. It depicts the three different clusters.



The important observations that can be made from the cluster analysis are as discussed below:

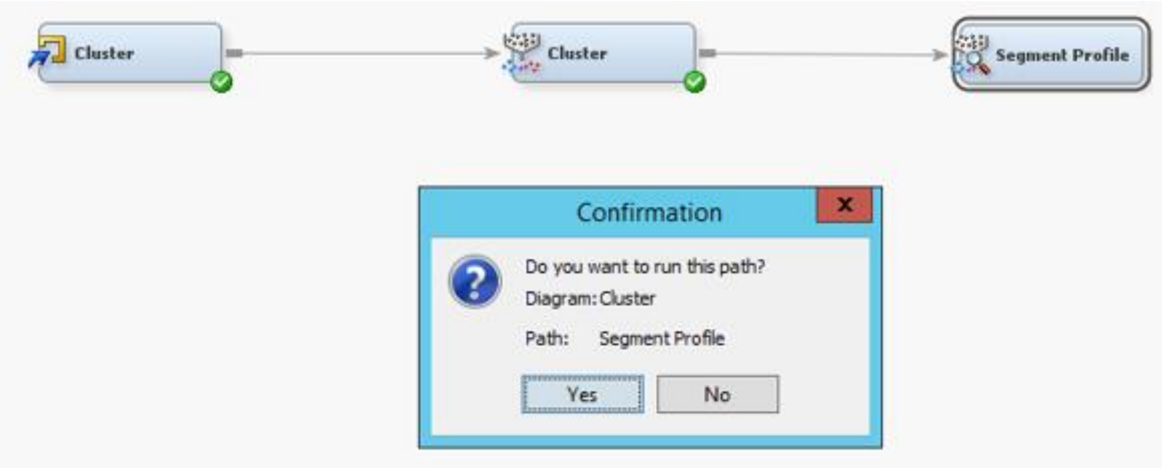
1. In the segment 1, the *tenure* value is around 30 months. This means that the segment has been with telco for some time now, and these are matured customers. Their *MonthlyCharges* is around $25 and the *totalCharges* is in and around *$800*. They are not yet loyalists and the efforts of the company should be towards making them loyalists. The effect of churning is considerable.

2. In the Segment 2, the *tenure* value is around 60 months. This means that the segment has been with telco for a long time now, and these are valuable customers. Their *MonthlyCharges* is around $90 and the *totalCharges* is in and around *$5000*. They are the loyalists. The effect of churning will impact the company.

3. In the Segment 3, the *tenure* value is around 12 months. This means that the segment has been with telco for just a short time, and these are fickle minded customers. Their *MonthlyCharges* is around $75 and the *totalCharges* is in and around *$1000*. They are adding a lot of revenue to the company as compared to the Segment 1.

As a bottom-line, churning in Segment 2 is what will hit the company worst.

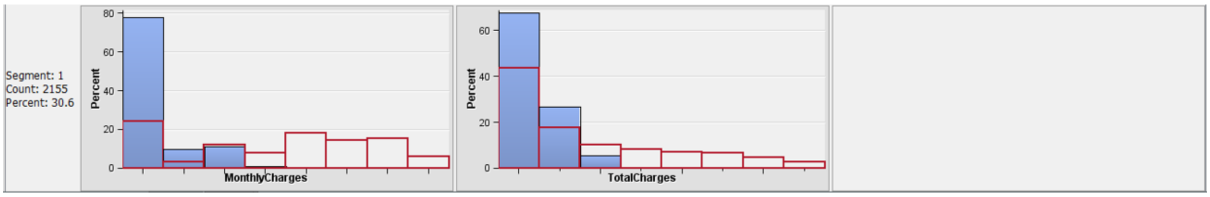
An improved study can be conducted if an analysis of the segments is conducted using segment profiling. The Segment profile node is added to the workspace diagram and connected to the cluster node. Then, the path is executed to profile the segments. The workspace looks as below:



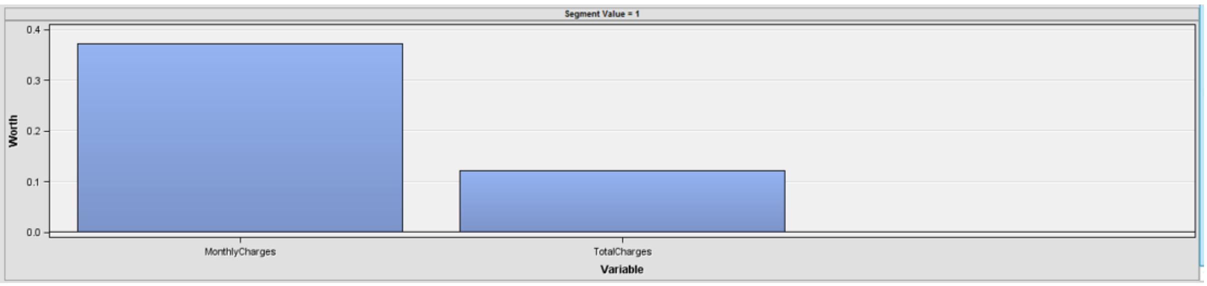
The Cluster-Segment Profile results are given below for the reference of the reader:

Below, is a study on the various segments using profiling:

<u>Segment #1:</u>

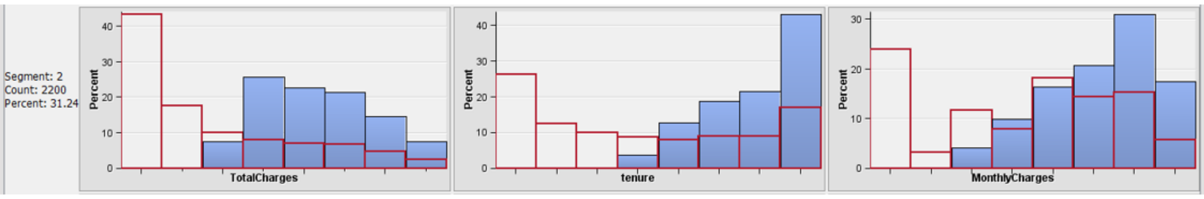


As seen in the above figure, 'MonthlyCharges' and 'TotalCharges' are relatively low as compared to the entire dataset. Variable 'tenure' doesn't influence this segment at all.

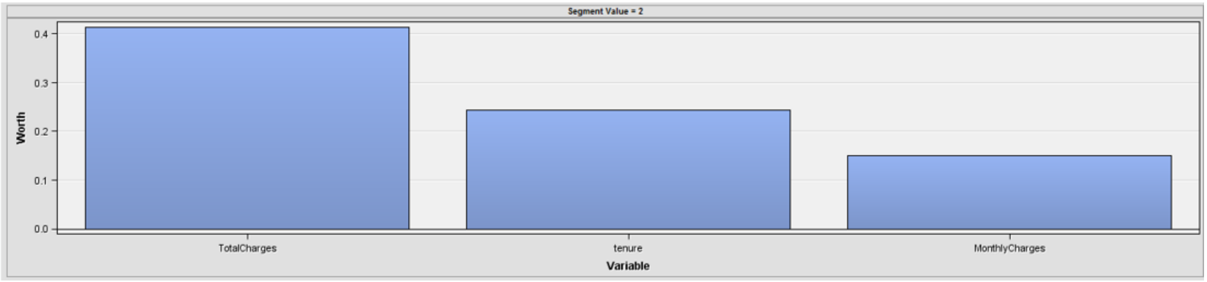


The above figure is indicative of the fact that in segment #1, 'MonthlyCharges' is the most influential variable followed by 'TotalCharges' to some extent and 'tenure' doesn't play a role.

Segment #2:

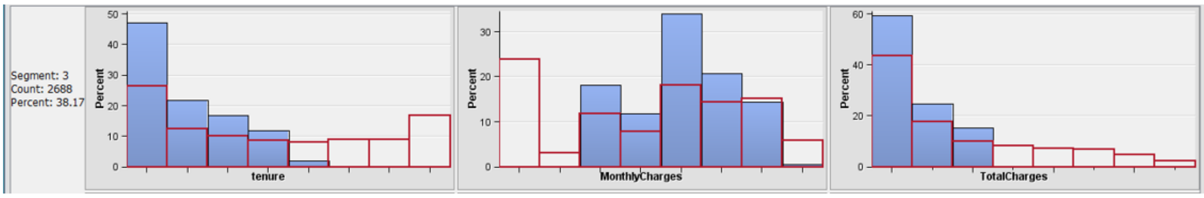


As seen in the above figure, *MonthlyCharges*, *tenure*, and *TotalCharges* are higher as compared to the entire dataset.

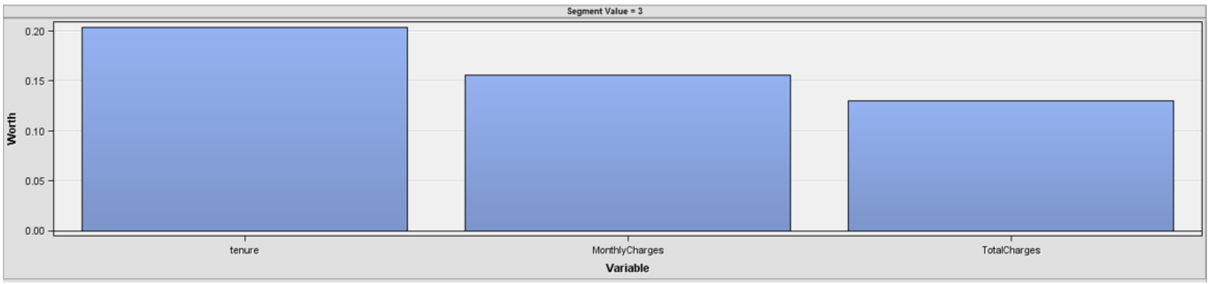


The above figure is indicative of the fact that in segment #1, *TotalCharges* is the most influential variable followed by *tenure* and *MonthlyCharges*.

Segment #3:



As seen in the above figure, *tenure*, *MonthlyCharges* and *totalCharges* are in a medium level. The metrics are not high or low but displays an average trend.



The above figure is indicative of the fact that in segment #3, *tenure* is the most influential variable followed by *MonthlyCharges* and *TotalCharges*.

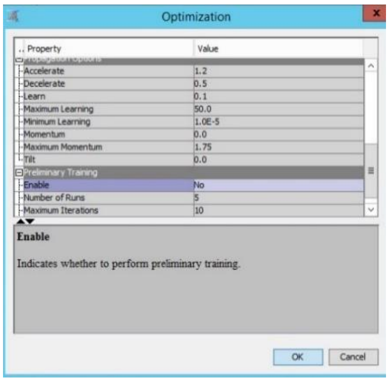
NEURAL NETWORK ANALYSIS

The intention of this analysis is to build a predictive model that can identify whether a customer will discontinue the product/ services of the company with some confidence.

- Defining the data source by file import and defining the variables.
- Target – Churn, Rejected – *CustomerID*

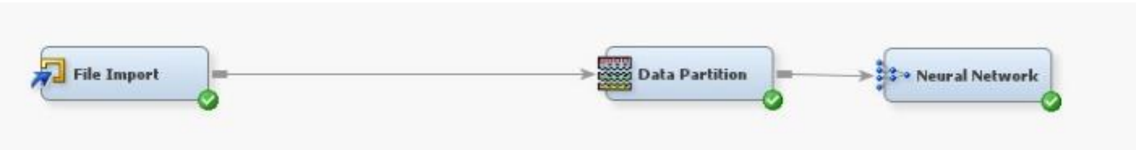
The neural network node is connected to the partitioned data and run the model with the following settings**:**

Optimization -> Preliminary Training ->No

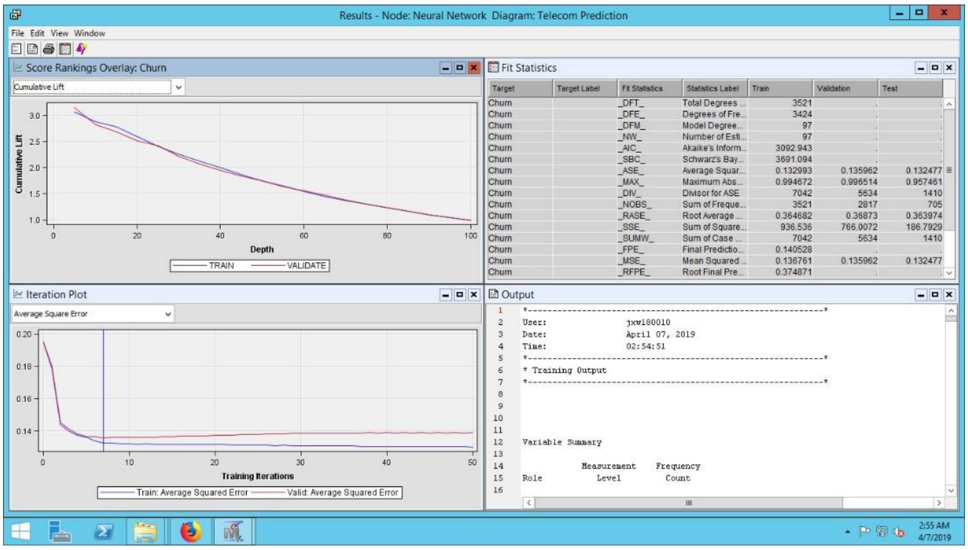


Approach #1

The workspace diagram is as follows:

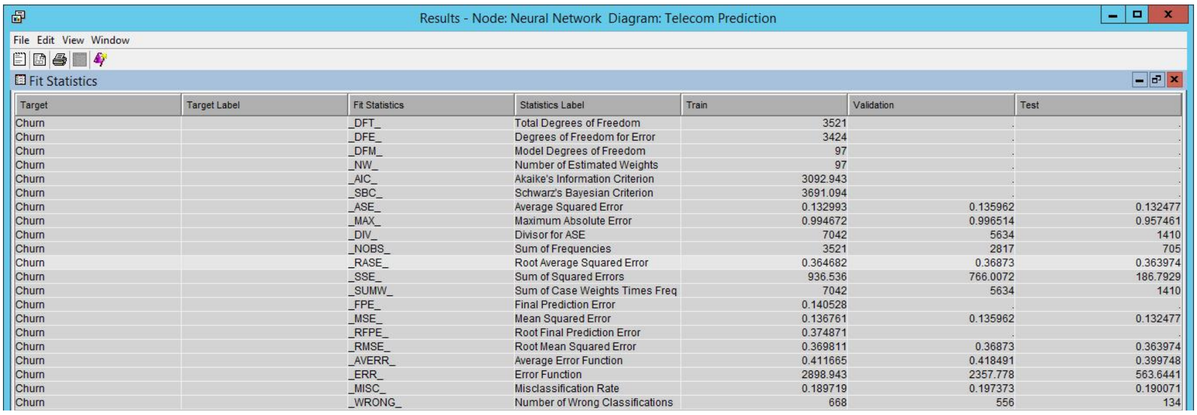


The Output is as follows:



As seen in the iteration plot above, the Iteration 8 is the optimized convergence point, and the maximum default iterations = 50. The above iteration is a failed/incomplete/insufficient one, keeping the convergence factor in mind. As a result, the output shows a warning message (The screenshot is attached at a later stage of the report)

The fit statistics is as follows:



Results - Node: Neural Network Diagram: Telecom Prediction

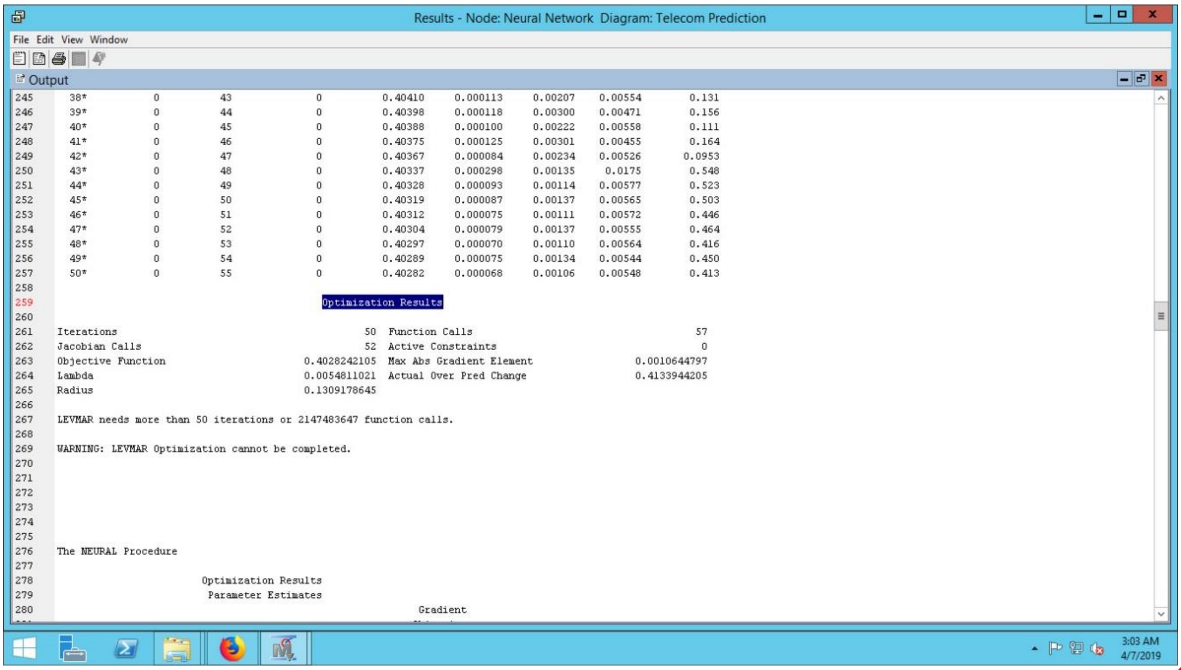
Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		_DFT_	Total Degrees of Freedom	3521	.	.
Churn		_DFE_	Degrees of Freedom for Error	3424	.	.
Churn		_DFM_	Model Degrees of Freedom	97	.	.
Churn		_NW_	Number of Estimated Weights	97	.	.
Churn		_AIC_	Akaike's Information Criterion	3092.943	.	.
Churn		_SBC_	Schwarz's Bayesian Criterion	3691.094	.	.
Churn		_ASE_	Average Squared Error	0.132993	0.135962	0.132477
Churn		_MAX_	Maximum Absolute Error	0.994672	0.996514	0.957461
Churn		_DIV_	Divisor for ASE	7042	5634	1410
Churn		_NOBS_	Sum of Frequencies	3521	2817	705
Churn		_RASE_	Root Average Squared Error	0.364682	0.36873	0.363974
Churn		_SSE_	Sum of Squared Errors	936.536	766.0072	186.7929
Churn		_SUMW_	Sum of Case Weights Times Freq	7042	5634	1410
Churn		_FPE_	Final Prediction Error	0.140528	.	.
Churn		_MSE_	Mean Squared Error	0.136761	0.135962	0.132477
Churn		_RFPE_	Root Final Prediction Error	0.374871	.	.
Churn		_RMSE_	Root Mean Squared Error	0.369811	0.36873	0.363974
Churn		_AVERR_	Average Error Function	0.411665	0.418491	0.399748
Churn		_ERR_	Error Function	2898.943	2357.778	563.6441
Churn		_MISC_	Misclassification Rate	0.189719	0.197373	0.190071
Churn		_WRONG_	Number of Wrong Classifications	668	556	134

The important metrics from the fit statistics is as follows:

	TRAIN	VALIDATE	TEST
ASE	0.1329	0.1359	0.1325
MISCLASSIFICATION RATE	0.1897	0.1974	0.1901
NO. OF ESTIMATED WEIGHTS	97		
ITERATIONS - 8			

The warning message described for convergence earlier is as seen in the below screenshot:



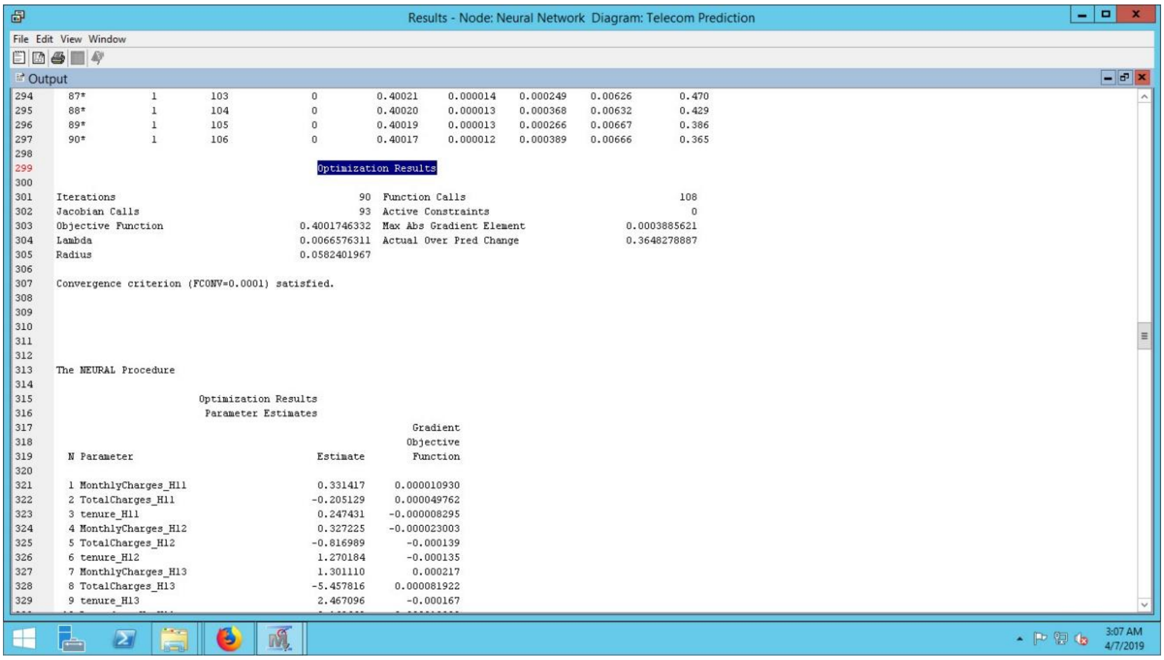
Results - Node: Neural Network Diagram: Telecom Prediction

Output

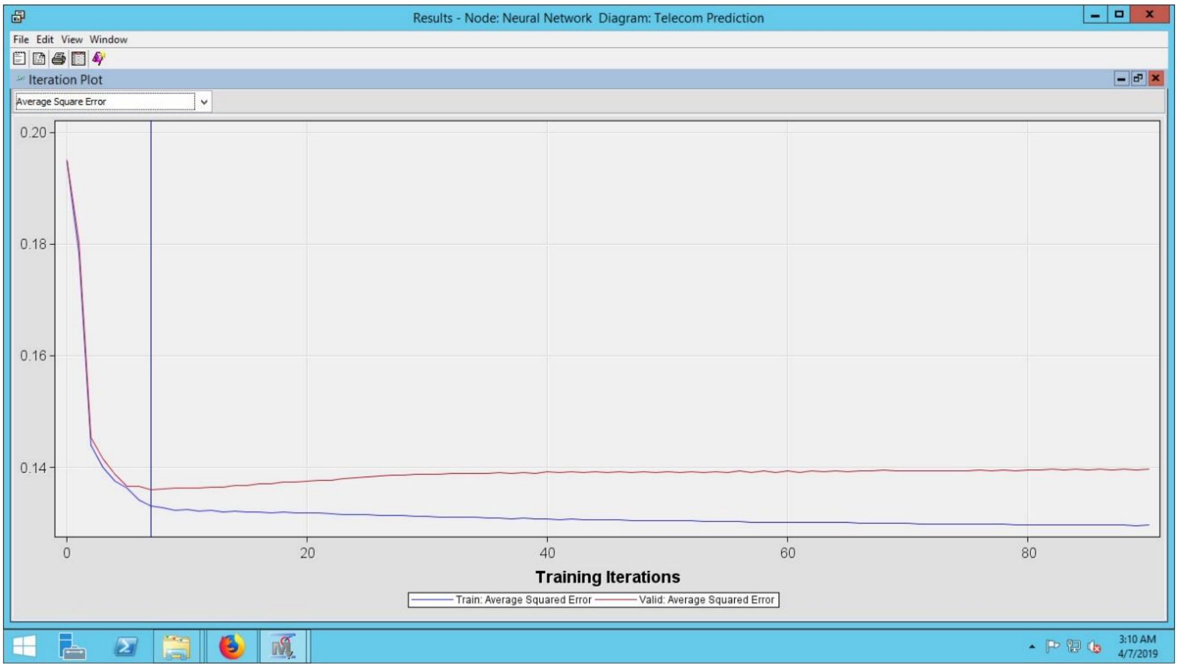
```
38*     0     43     0     0.40410   0.000113   0.00207   0.00554    0.131
39*     0     44     0     0.40398   0.000118   0.00300   0.00471    0.156
40*     0     45     0     0.40388   0.000100   0.00222   0.00558    0.111
41*     0     46     0     0.40375   0.000125   0.00301   0.00455    0.164
42*     0     47     0     0.40367   0.000084   0.00234   0.00526   0.0953
43*     0     48     0     0.40337   0.000298   0.00135    0.0175    0.548
44*     0     49     0     0.40328   0.000093   0.00114   0.00577    0.523
45*     0     50     0     0.40319   0.000087   0.00137   0.00565    0.503
46*     0     51     0     0.40312   0.000075   0.00111   0.00572    0.446
47*     0     52     0     0.40304   0.000079   0.00137   0.00555    0.464
48*     0     53     0     0.40297   0.000070   0.00110   0.00564    0.416
49*     0     54     0     0.40289   0.000075   0.00134   0.00544    0.450
50*     0     55     0     0.40282   0.000068   0.00106   0.00548    0.413

                            Optimization Results

Iterations                          50  Function Calls                      57
Jacobian Calls                      52  Active Constraints                   0
Objective Function        0.4028242105  Max Abs Gradient Element  0.0010644797
Lambda                    0.0054811021  Actual Over Pred Change   0.4133944205
Radius                    0.1309178645

LEVMAR needs more than 50 iterations or 2147483647 function calls.

WARNING: LEVMAR Optimization cannot be completed.



The NEURAL Procedure

                   Optimization Results
                    Parameter Estimates
                                              Gradient
```

3:03 AM
4/7/2019

To address this, change the maximum iterations from 50-100, and then the output is checked. Here a warning message is not seen. It is inferred that 100 iterations are sufficient to run this dataset.

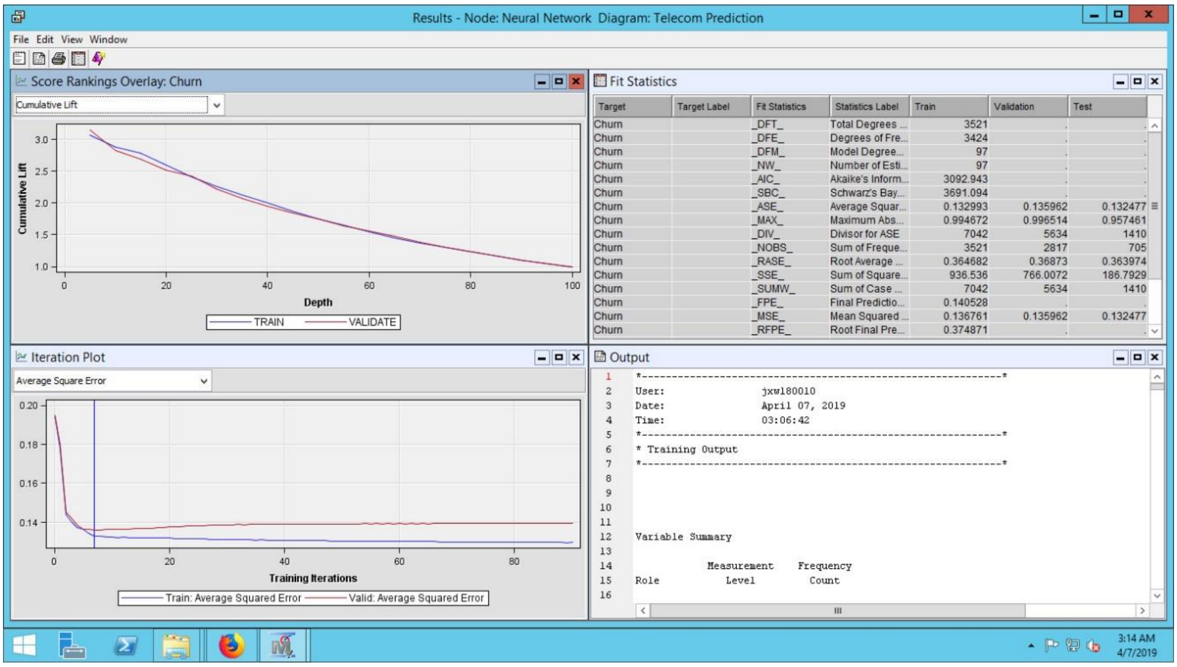


The Iteration Plot for this run is as seen below:



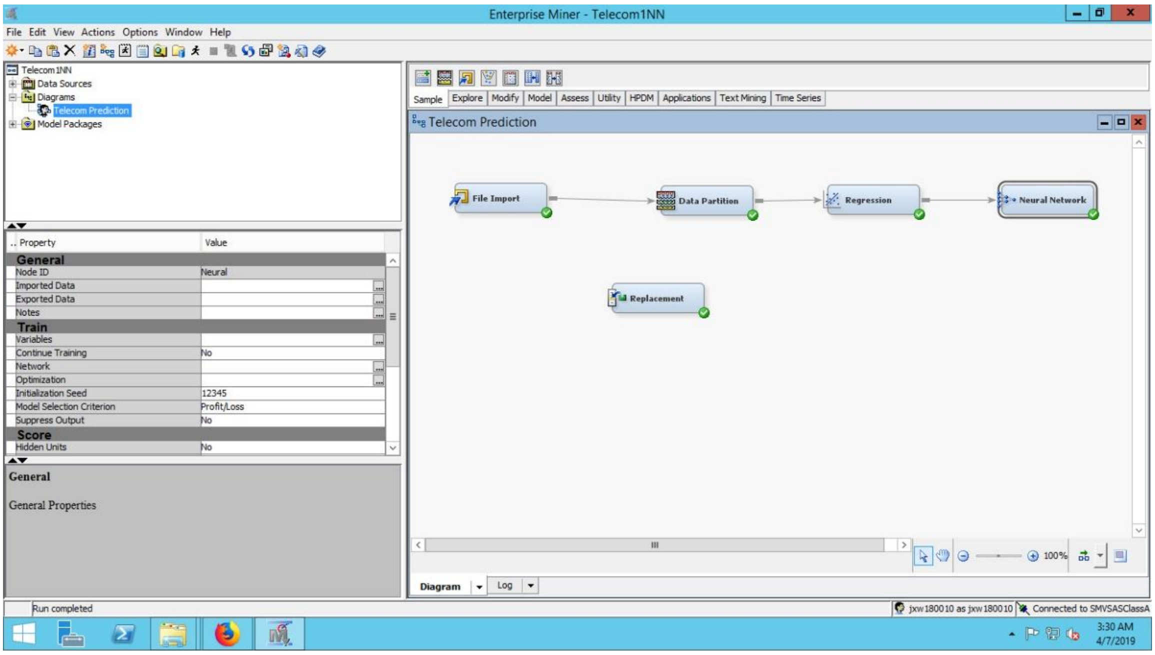
Note: *Divergence is at iteration 7, which was 8 earlier for 50 max. iterations.*

The Results Window is as seen below:

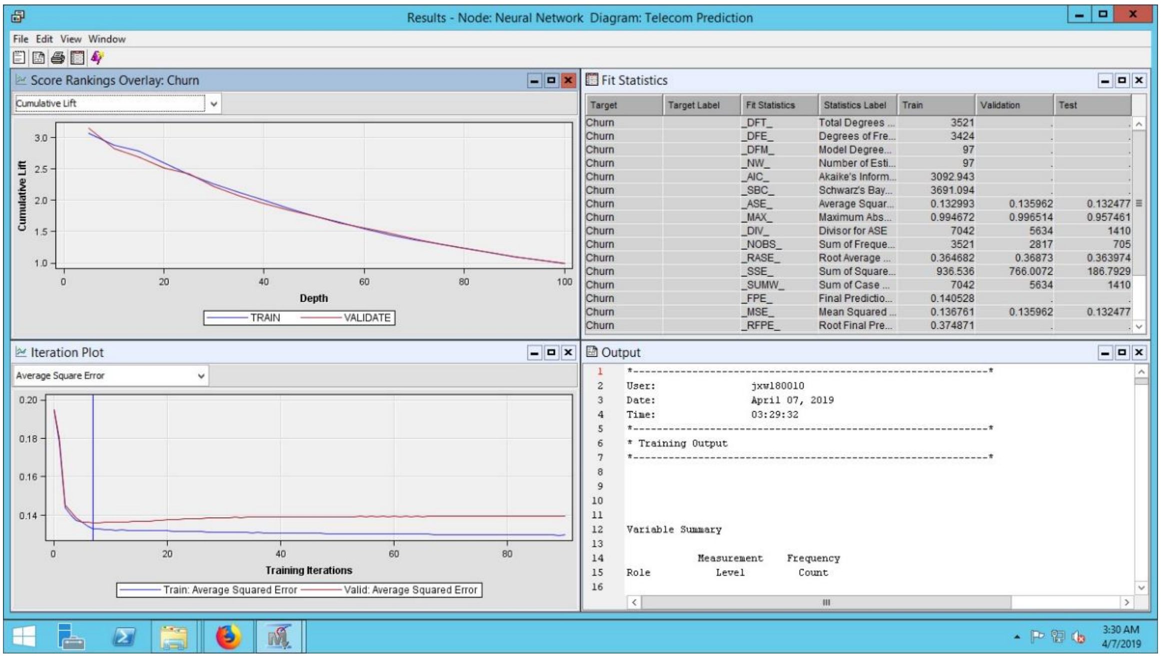


Approach #2

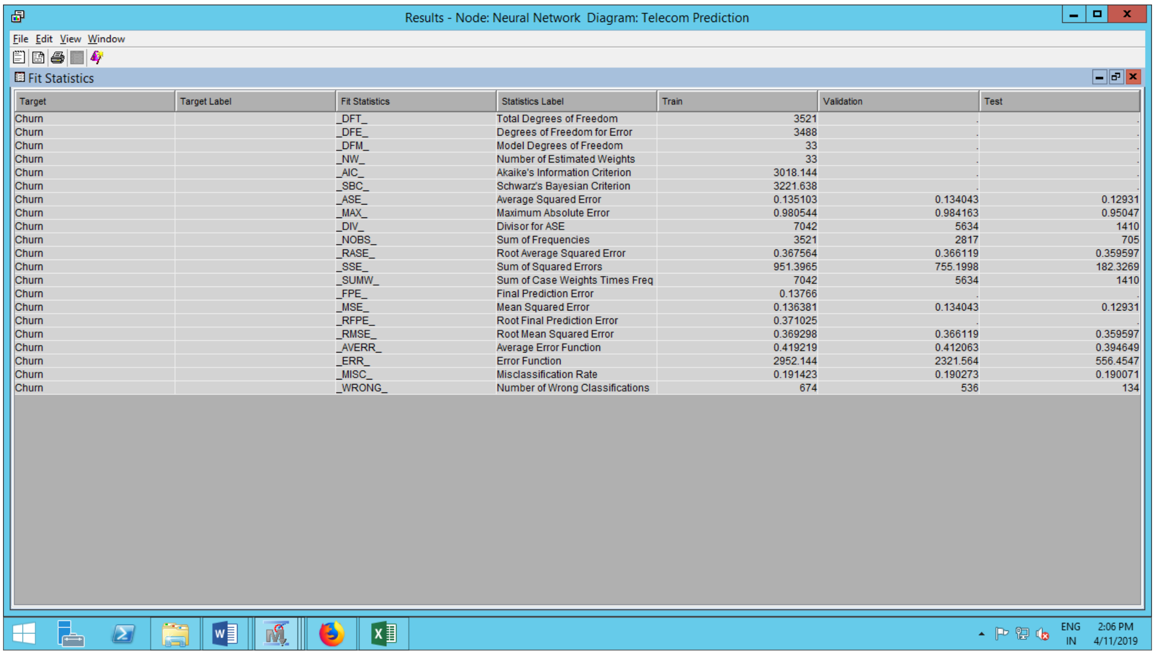
Logistic regression can be used to feed as the inputs for neural network. This improves the quality of the model by rejecting the unwanted inputs:



The output with regression to set as inputs for neural networks is as follows:



The fit Statistics is as seen below:



Results - Node: Neural Network Diagram: Telecom Prediction

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		_DFT_	Total Degrees of Freedom	3521	.	.
Churn		_DFE_	Degrees of Freedom for Error	3488	.	.
Churn		_DFM_	Model Degrees of Freedom	33	.	.
Churn		_NW_	Number of Estimated Weights	33	.	.
Churn		_AIC_	Akaike's Information Criterion	3018.144	.	.
Churn		_SBC_	Schwarz's Bayesian Criterion	3221.638	.	.
Churn		_ASE_	Average Squared Error	0.135103	0.134043	0.12931
Churn		_MAX_	Maximum Absolute Error	0.980544	0.984163	0.95047
Churn		_DIV_	Divisor for ASE	7042	5634	1410
Churn		_NOBS_	Sum of Frequencies	3521	2817	705
Churn		_RASE_	Root Average Squared Error	0.367564	0.366119	0.359597
Churn		_SSE_	Sum of Squared Errors	951.3965	755.1998	182.3269
Churn		_SUMW_	Sum of Case Weights Times Freq	7042	5634	1410
Churn		_FPE_	Final Prediction Error	0.13766	.	.
Churn		_MSE_	Mean Squared Error	0.136381	0.134043	0.12931
Churn		_RFPE_	Root Final Prediction Error	0.371025	.	.
Churn		_RMSE_	Root Mean Squared Error	0.369298	0.366119	0.359597
Churn		_AVERR_	Average Error Function	0.419219	0.412063	0.394649
Churn		_ERR_	Error Function	2952.144	2321.564	556.4547
Churn		_MISC_	Misclassification Rate	0.191423	0.190273	0.190071
Churn		_WRONG_	Number of Wrong Classifications	674	536	134

The important metrics from the fit statistics is as follows:

	Train	*Validate*	*Test*
ASE	0.1351	0.1340	0.1293
Misclassification Rate	0.1914	0.1903	0.1900
No. of Estimated Weights	33		

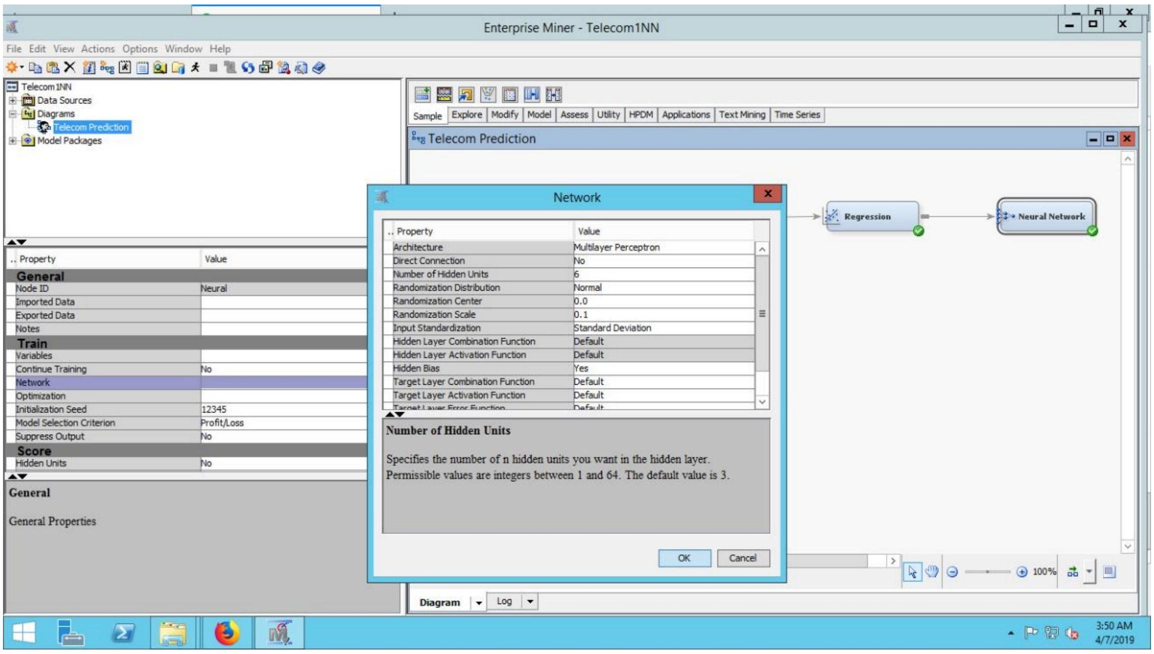
Iterations – 7 [same as the output of neural network]

From the above result, it can be concluded that although the ASE and Misclassification rate values are quite similar, the number of weights is reduced considerably, so this is a better model than the previous one.

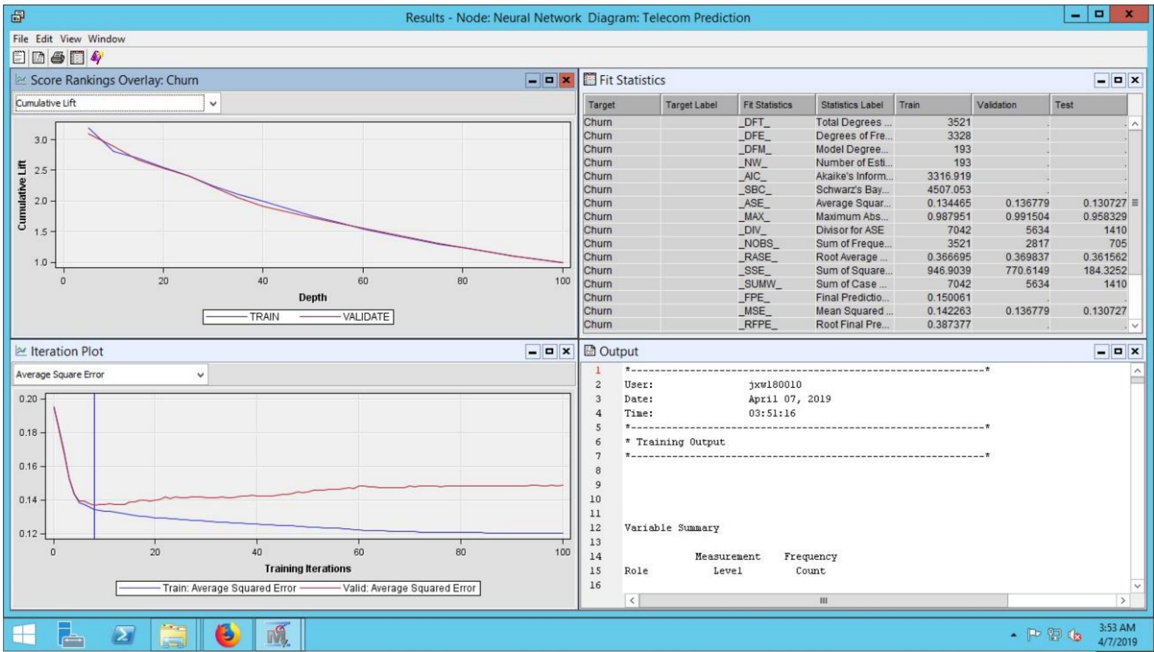
In order to increase the flexibility of the network, the hidden nodes can be adjusted by 2 methods, either **manually by trial and error** and also automatically by **auto-neural networks**.

Approach #3

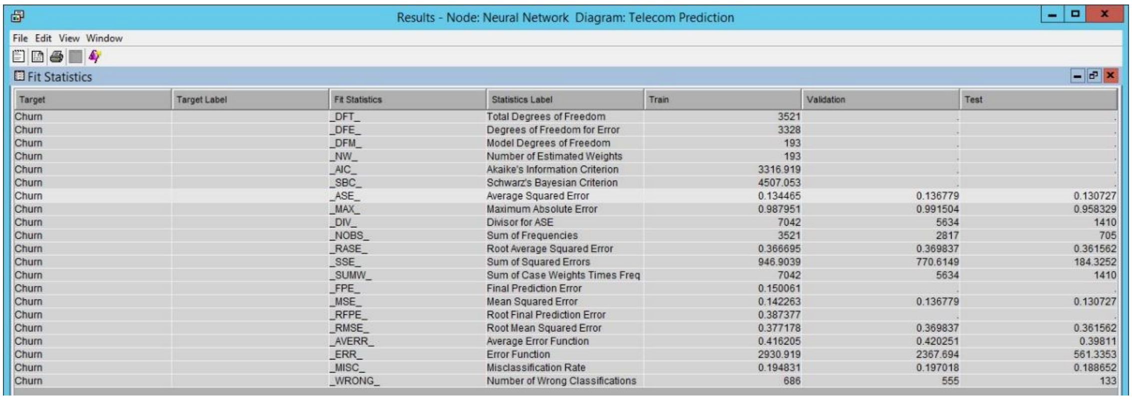
First, manual method is carried out, changing the number of hidden units from 3 [default] to 6.



The Results after running the model is as seen below:



The fit statistics is as follows:



Results - Node: Neural Network Diagram: Telecom Prediction

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		_DFT_	Total Degrees of Freedom	3521	.	.
Churn		_DFE_	Degrees of Freedom for Error	3328	.	.
Churn		_DFM_	Model Degrees of Freedom	193	.	.
Churn		_NW_	Number of Estimated Weights	193	.	.
Churn		_AIC_	Akaike's Information Criterion	3316.919	.	.
Churn		_SBC_	Schwarz's Bayesian Criterion	4507.053	.	.
Churn		_ASE_	Average Squared Error	0.134465	0.136779	0.130727
Churn		_MAX_	Maximum Absolute Error	0.987951	0.991504	0.958329
Churn		_DIV_	Divisor for ASE	7042	5634	1410
Churn		_NOBS_	Sum of Frequencies	3521	2817	705
Churn		_RASE_	Root Average Squared Error	0.366695	0.369837	0.361562
Churn		_SSE_	Sum of Squared Errors	946.9039	770.6149	184.3252
Churn		_SUMW_	Sum of Case Weights Times Freq	7042	5634	1410
Churn		_FPE_	Final Prediction Error	0.150061	.	.
Churn		_MSE_	Mean Squared Error	0.142263	0.136779	0.130727
Churn		_RFPE_	Root Final Prediction Error	0.387377	.	.
Churn		_RMSE_	Root Mean Squared Error	0.377178	0.369837	0.361562
Churn		_AVERR_	Average Error Function	0.416205	0.420251	0.39811
Churn		_ERR_	Error Function	2930.919	2367.694	561.3353
Churn		_MISC_	Misclassification Rate	0.194831	0.197018	0.188652
Churn		_WRONG_	Number of Wrong Classifications	686	555	133

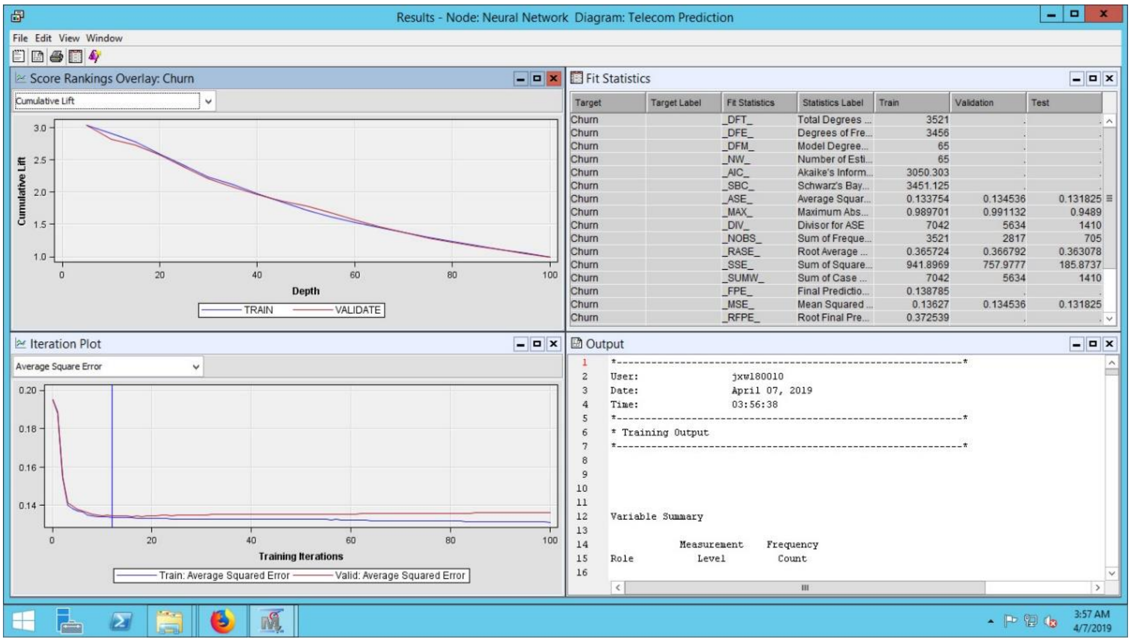
The important metrics from the fit statistics is as follows:

	Train	*Validate*	*Test*
ASE	0.1345	0.1368	0.1307
Misclassification Rate	0.1948	0.1970	0.1865
No. of Estimated Weights	193 [reduced from 97]		
	Iterations – 8		

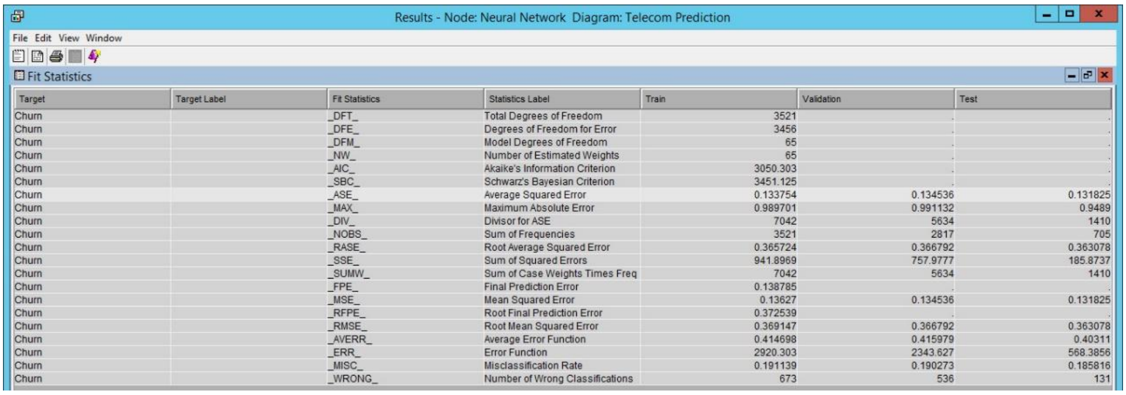
In this above model the ASE and Misclassification rate are quite similar with also the number of estimated weights increased from 97-193, which is bad and the number of iterations = 8.

Approach #4

Here, the number of hidden units is changed from 6 to 2. The results window is as below:



The fit statistics is as follows:



Results - Node: Neural Network Diagram: Telecom Prediction

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		_DFT_	Total Degrees of Freedom	3521	.	.
Churn		_DFE_	Degrees of Freedom for Error	3456	.	.
Churn		_DFM_	Model Degrees of Freedom	65	.	.
Churn		_NW_	Number of Estimated Weights	65	.	.
Churn		_AIC_	Akaike's Information Criterion	3050.303	.	.
Churn		_SBC_	Schwarz's Bayesian Criterion	3451.125	.	.
Churn		_ASE_	Average Squared Error	0.133754	0.134536	0.131825
Churn		_MAX_	Maximum Absolute Error	0.989701	0.991132	0.9489
Churn		_DIV_	Divisor for ASE	7042	5634	1410
Churn		_NOBS_	Sum of Frequencies	3521	2817	705
Churn		_RASE_	Root Average Squared Error	0.365724	0.366792	0.363078
Churn		_SSE_	Sum of Squared Errors	941.8969	757.9777	185.8737
Churn		_SUMW_	Sum of Case Weights Times Freq	7042	5634	1410
Churn		_FPE_	Final Prediction Error	0.138785	.	.
Churn		_MSE_	Mean Squared Error	0.13627	0.134536	0.131825
Churn		_RFPE_	Root Final Prediction Error	0.372539	.	.
Churn		_RMSE_	Root Mean Squared Error	0.369147	0.366792	0.363078
Churn		_AVERR_	Average Error Function	0.414698	0.415979	0.40311
Churn		_ERR_	Error Function	2920.303	2343.627	568.3856
Churn		_MISC_	Misclassification Rate	0.191139	0.190273	0.185816
Churn		_WRONG_	Number of Wrong Classifications	673	536	131

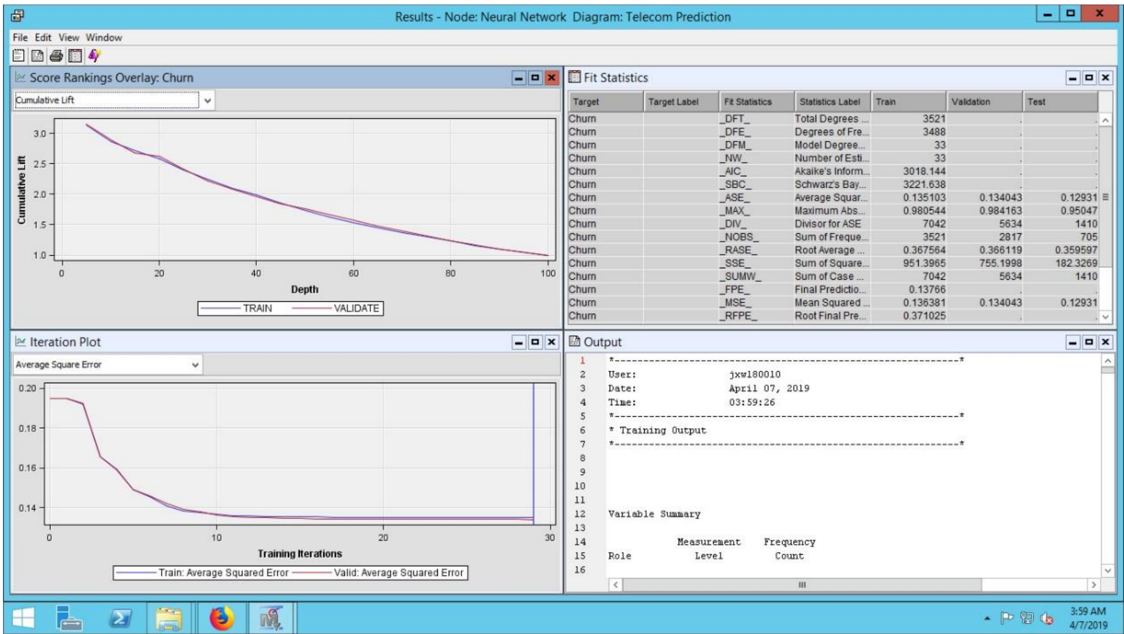
The important metrics from the fit statistics is as follows:

	Train	*Validate*	*Test*
ASE	0.1337	0.1345	0.1318
Misclassification Rate	0.1911	0.1902	0.1856
No. of Estimated Weights	65 [same as 6 hidden units]		
	Iterations – 12		

In this above model the ASE and Misclassification rate are quite similar with also the number of estimated weights is 65, which is better, and the number of iterations increased to 12. This is a better model.

Approach #5

Here, the number of hidden units is changed from 2 to 1. The results window is as below:



The fit statistics is as follows:

Results - Node: Neural Network Diagram: Telecom Prediction

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		_DFT_	Total Degrees of Freedom	3521	.	.
Churn		_DFE_	Degrees of Freedom for Error	3488	.	.
Churn		_DFM_	Model Degrees of Freedom	33	.	.
Churn		_NW_	Number of Estimated Weights	33	.	.
Churn		_AIC_	Akaike's Information Criterion	3018.144	.	.
Churn		_SBC_	Schwarz's Bayesian Criterion	3221.638	.	.
Churn		_ASE_	Average Squared Error	0.135103	0.134043	0.12931
Churn		_MAX_	Maximum Absolute Error	0.980544	0.984163	0.95047
Churn		_DIV_	Divisor for ASE	7042	5634	1410
Churn		_NOBS_	Sum of Frequencies	3521	2817	705
Churn		_RASE_	Root Average Squared Error	0.367564	0.366119	0.359597
Churn		_SSE_	Sum of Squared Errors	951.3965	755.1998	182.3269
Churn		_SUMW_	Sum of Case Weights Times Freq	7042	5634	1410
Churn		_FPE_	Final Prediction Error	0.13766	.	.
Churn		_MSE_	Mean Squared Error	0.136381	0.134043	0.12931
Churn		_RFPE_	Root Final Prediction Error	0.371025	.	.
Churn		_RMSE_	Root Mean Squared Error	0.369298	0.366119	0.359597
Churn		_AVERR_	Average Error Function	0.419219	0.412063	0.394649
Churn		_ERR_	Error Function	2952.144	2321.564	556.4547
Churn		_MISC_	Misclassification Rate	0.191423	0.190273	0.190071
Churn		_WRONG_	Number of Wrong Classifications	674	536	134

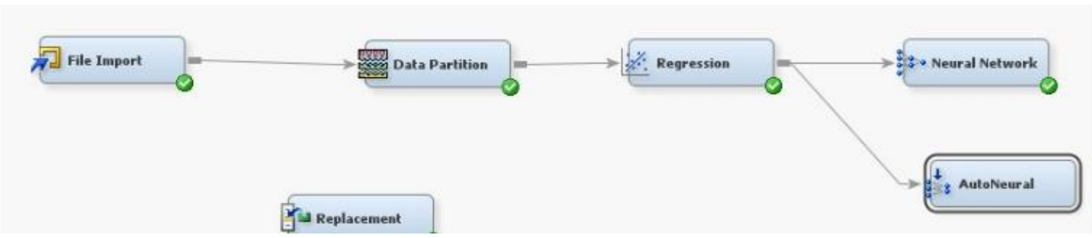
The important metrics from the fit statistics is as follows:

	Train	*Validate*	*Test*
ASE	0.1351	0.1340	0.1293
Misclassification Rate	0.1914	0.1902	0.1900
No. of Estimated Weights	33 [reduced from 65]		
Iterations – 29			

In this above model the ASE and Misclassification rate are quite similar and also the number of estimated weights is 33. It is better than 65 and the number of iterations increased to 29 from 12. This is a model with average performance.

Approach #6

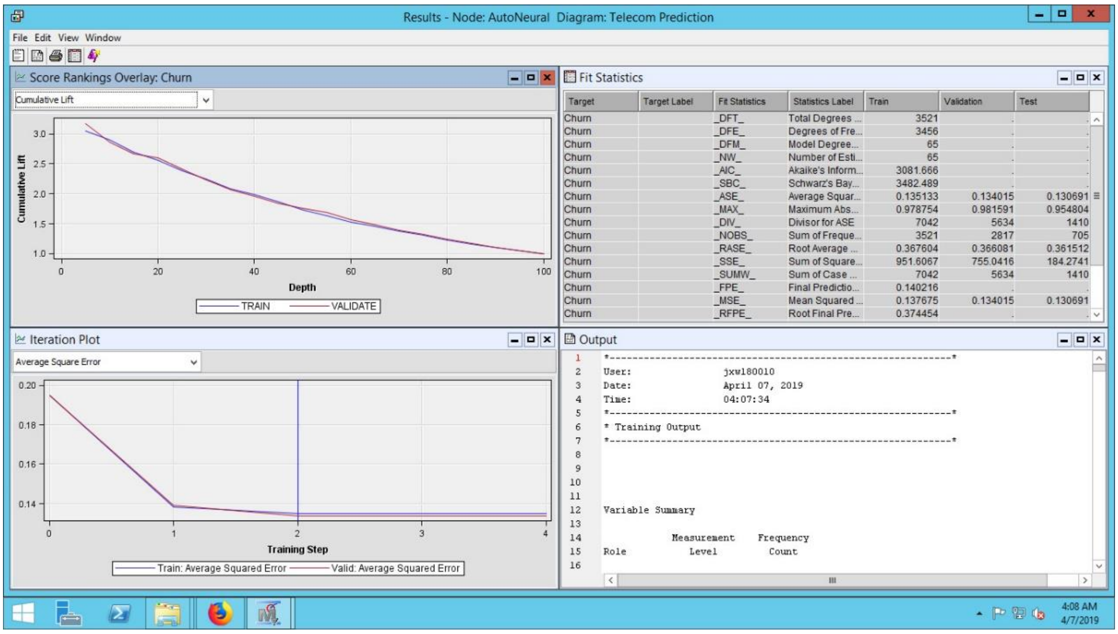
The analysis delves into Auto neural network. The workspace looks as seen below:



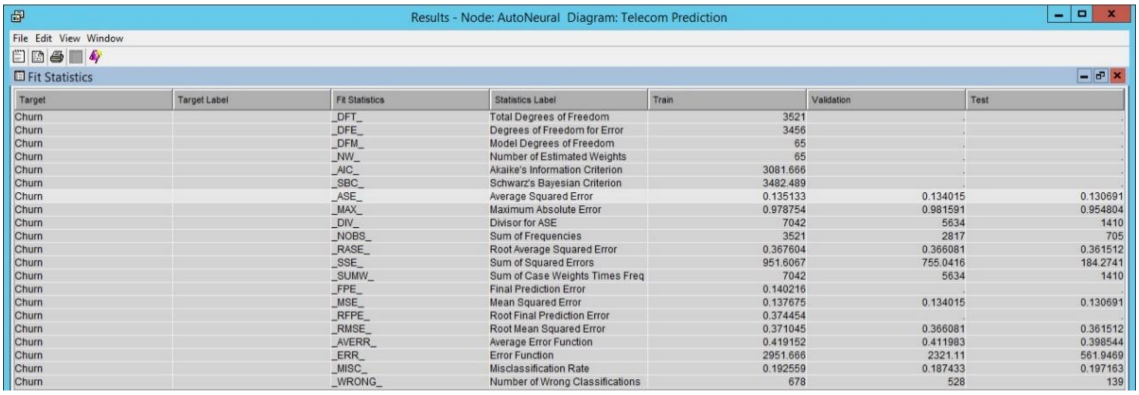
The important properties are as listed below:

- Training Action -> search
- No. of hidden units -> 1,
- Changing Direct, sine, normal settings of activation function to No.

The results window is as below:



The fit statistics is as follows:


Results - Node: AutoNeural Diagram: Telecom Prediction

File Edit View Window

Fit Statistics

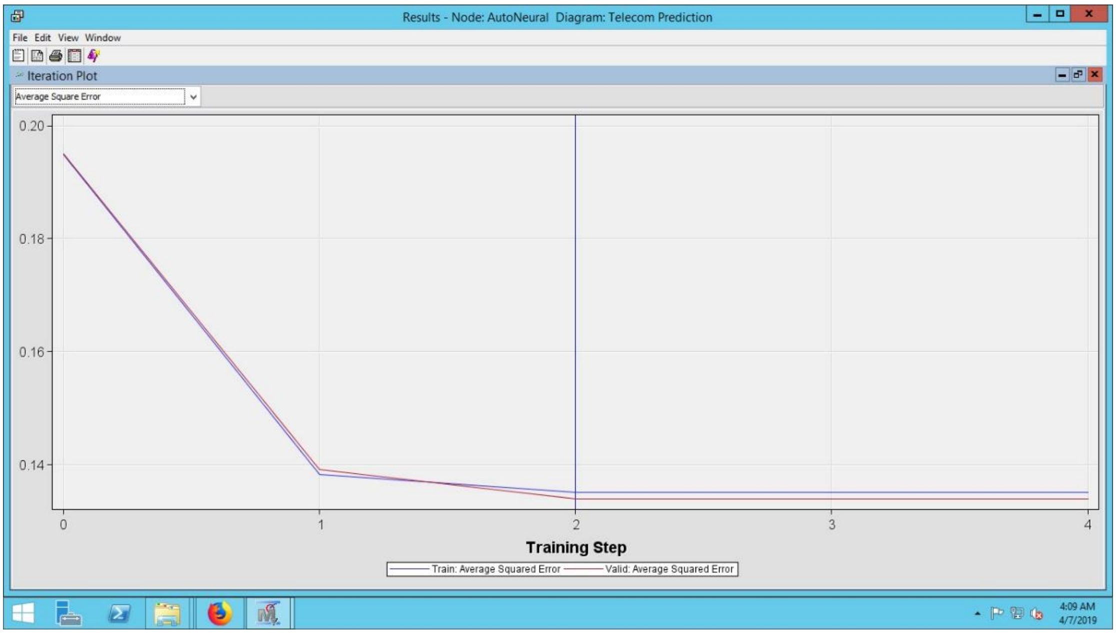
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		_DFT_	Total Degrees of Freedom	3521	.	.
Churn		_DFE_	Degrees of Freedom for Error	3456	.	.
Churn		_DFM_	Model Degrees of Freedom	65	.	.
Churn		_NW_	Number of Estimated Weights	65	.	.
Churn		_AIC_	Akaike's Information Criterion	3081.666	.	.
Churn		_SBC_	Schwarz's Bayesian Criterion	3482.489	.	.
Churn		_ASE_	Average Squared Error	0.135133	0.134015	0.130691
Churn		_MAX_	Maximum Absolute Error	0.978754	0.981591	0.954804
Churn		_DIV_	Divisor for ASE	7042	5634	1410
Churn		_NOBS_	Sum of Frequencies	3521	2817	705
Churn		_RASE_	Root Average Squared Error	0.367604	0.366081	0.361512
Churn		_SSE_	Sum of Squared Errors	951.6067	755.0416	184.2741
Churn		_SUMW_	Sum of Case Weights Times Freq	7042	5634	1410
Churn		_FPE_	Final Prediction Error	0.140216	.	.
Churn		_MSE_	Mean Squared Error	0.137675	0.134015	0.130691
Churn		_RFPE_	Root Final Prediction Error	0.374454	.	.
Churn		_RMSE_	Root Mean Squared Error	0.371045	0.366081	0.361512
Churn		_AVERR_	Average Error Function	0.419152	0.411983	0.398544
Churn		_ERR_	Error Function	2951.666	2321.11	561.9469
Churn		_MISC_	Misclassification Rate	0.192559	0.187433	0.197163
Churn		_WRONG_	Number of Wrong Classifications	678	528	139

The important metrics from the fit statistics is as follows:

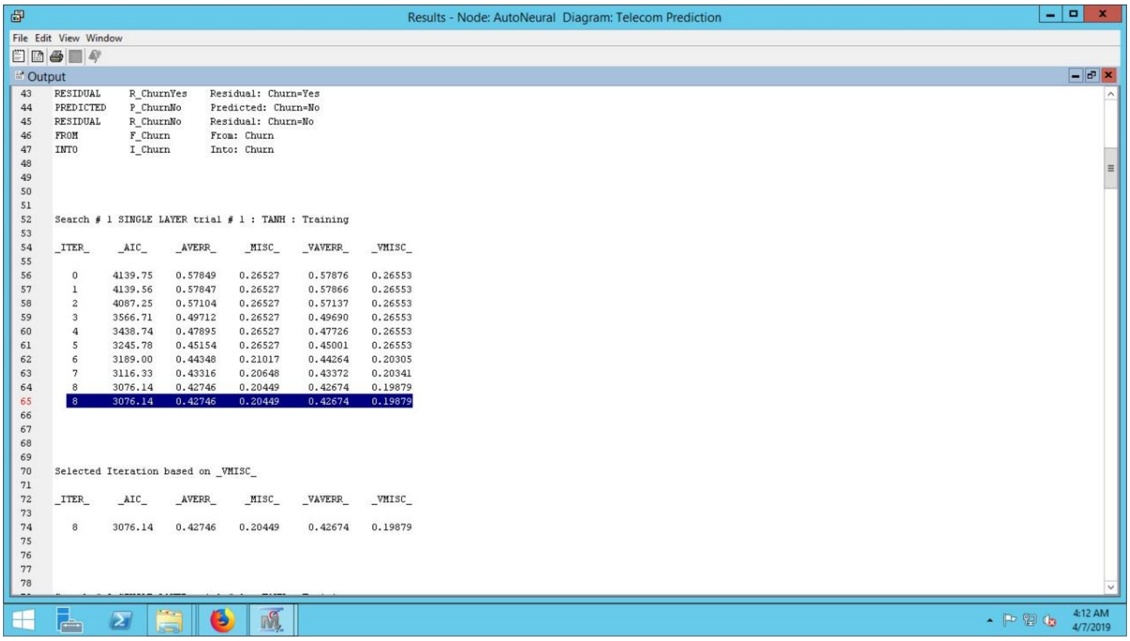
	Train	*Validate*	*Test*
ASE	0.1351	0.1340	0.1306
Misclassification Rate	0.1925	0.1874	0.1971
No. of Estimated Weights	65		
Iterations – Representation is different in auto-neural networks			

In this above auto neural model the ASE and Misclassification rate are quite similar with also the number of estimated weights is 65.

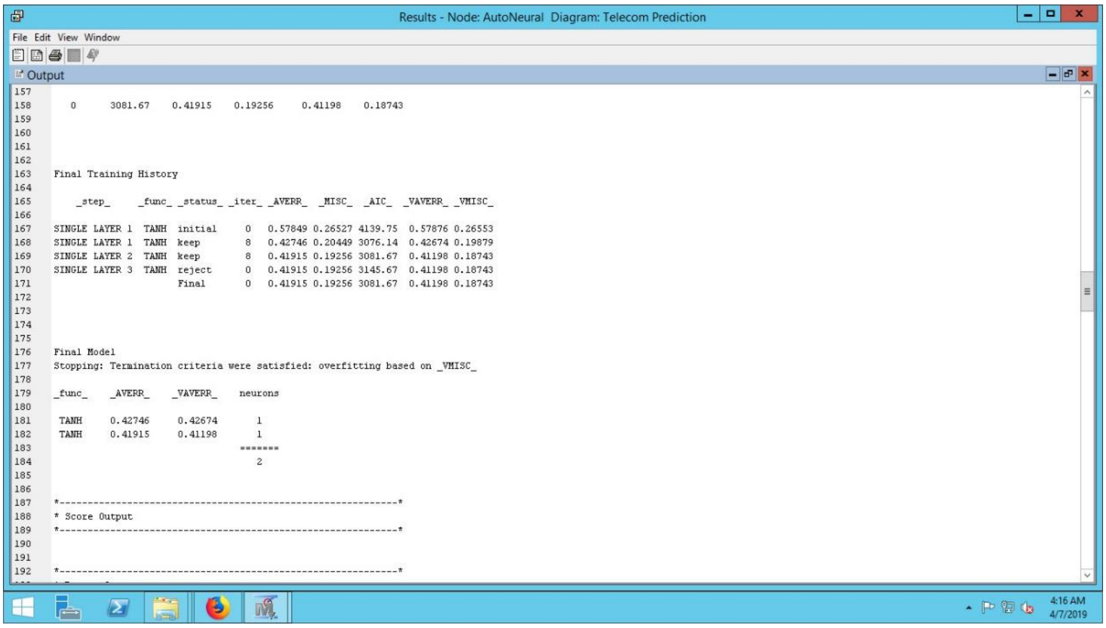
The Auto-Neural and Neural Network node's iteration plots differ. The Auto-Neural node's iteration plot shows the final fit statistic versus the number of hidden units in the neural network.



First iteration output, the min is taken as the input for next step, in the below screenshot it is iteration 8:

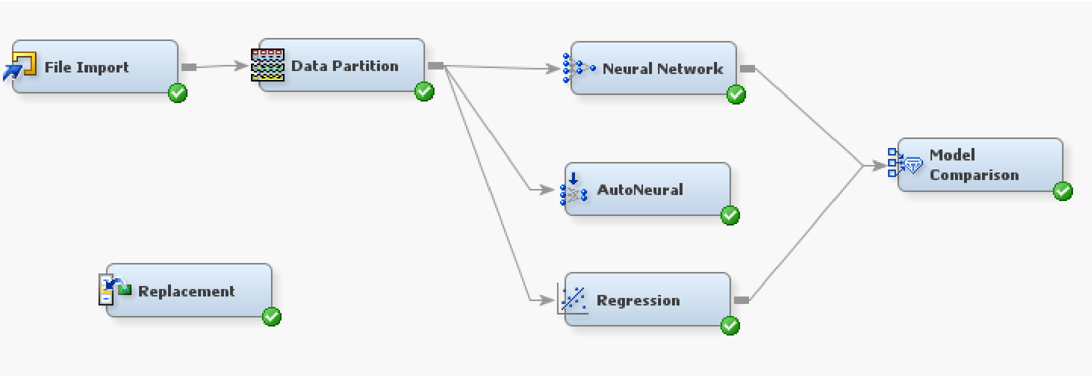


The final output of the auto-neural network is as seen below:

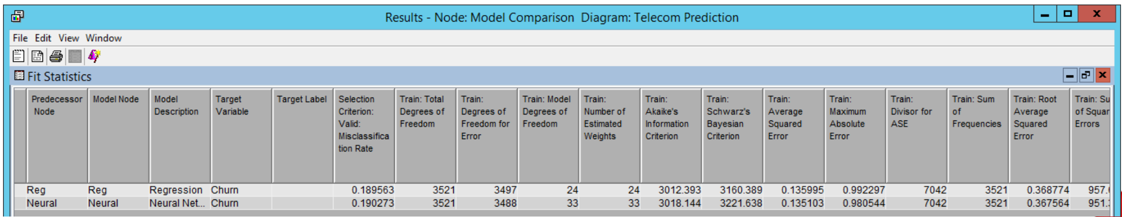


Model Comparison

To conduct a model comparison, the Model Comparison node is inserted and connected with Neural Network and Regression nodes. The workspace diagram looks as below:

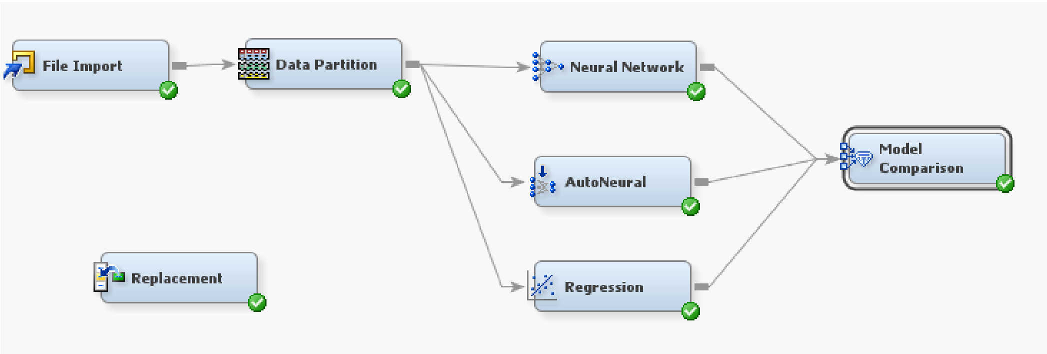


From the results it is seen that the results from both the models are very similar to each other as the ASE values are nearly same [0.1359 for Logistic regression and 0.1351 for neural network], but the number of estimated weights is 24 and 33. A model with less weights is preferred; so logistic regression is preferred.

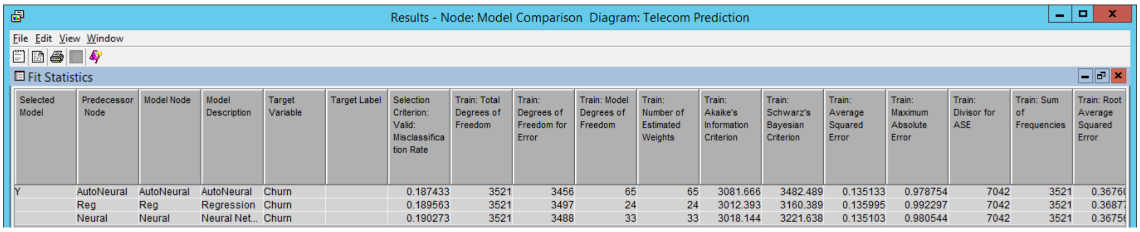


Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Total Degrees of Freedom	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Number of Estimated Weights	Train: Akaike's Information Criterion	Train: Schwarz's Bayesian Criterion	Train: Average Squared Error	Train: Maximum Absolute Error	Train: Divisor for ASE	Train: Sum of Frequencies	Train: Root Average Squared Error	Train: Su of Squar Errors
Reg	Reg	Regression	Churn		0.189563	3521	3497	24	24	3012.393	3160.389	0.135995	0.992297	7042	3521	0.368774	957.(
Neural	Neural	Neural Net...	Churn		0.190273	3521	3488	33	33	3018.144	3221.638	0.135103	0.980544	7042	3521	0.367564	951.:

Further on, model comparison is conducted between Auto-neural, neural and Logistic. The workspace is as seen below:



ASE values are almost similar, but estimated weights for logistic is least, so logistic regression is preferred. The results are as below:


Results - Node: Model Comparison Diagram: Telecom Prediction

File Edit View Window

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Total Degrees of Freedom	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Number of Estimated Weights	Train: Akaike's Information Criterion	Train: Schwarz's Bayesian Criterion	Train: Average Squared Error	Train: Maximum Absolute Error	Train: Divisor for ASE	Train: Sum of Frequencies	Train: Root Average Squared Error
Y	AutoNeural	AutoNeural	AutoNeural	Churn		0.187433	3521	3456	65	65	3081.666	3482.489	0.135133	0.978754	7042	3521	0.3676
	Reg	Reg	Regression	Churn		0.189563	3521	3497	24	24	3012.393	3160.389	0.135995	0.992297	7042	3521	0.3687
	Neural	Neural	Neural Net...	Churn		0.190273	3521	3488	33	33	3018.144	3221.638	0.135103	0.980544	7042	3521	0.3675

The model comparison summary is as below:

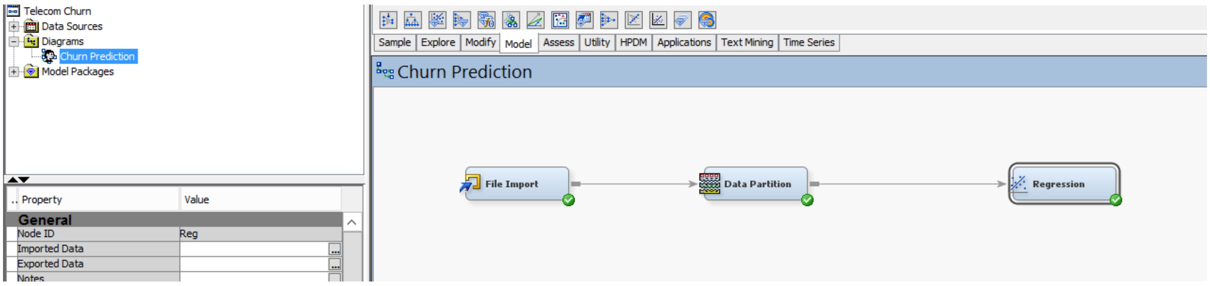
Model	**ASE**			**Misclassification Rate**			**No. of Estimated Weights**	**Iterations**
	Train	Validate	Test	Train	Validate	Test		
NN	0.1329	0.1359	0.1325	0.1897	0.1974	0.1901	97	8
NN with Logistic regression as input	**0.1351**	**0.1340**	**0.1293**	**0.1914**	**0.1903**	**0.1900**	**33**	**7**
NN with hidden networks = 6	0.1345	0.1368	0.1307	0.1948	0.1970	0.1865	193	8
NN with hidden networks = 2	0.1337	0.1345	0.1318	0.1911	0.1902	0.1856	65	12
NN with hidden networks = 1	0.1351	0.1340	0.1293	0.1914	0.1902	0.1900	33	29
Auto-Neural Network	0.1351	0.1340	0.1306	0.1925	0.1874	0.1971	65	Representation is different for auto-neural

From the above analysis of all types of neural networks, it is concluded that all the models have similar ASE and Misclassification rates in comparison to the values and also with the Train, Validate and Test data. But the number of estimated weights and the number of iterations is considerably low as compared to others for the Neural network with logistic regression as the input. So, this can be considered the best model from amongst the neural network models.

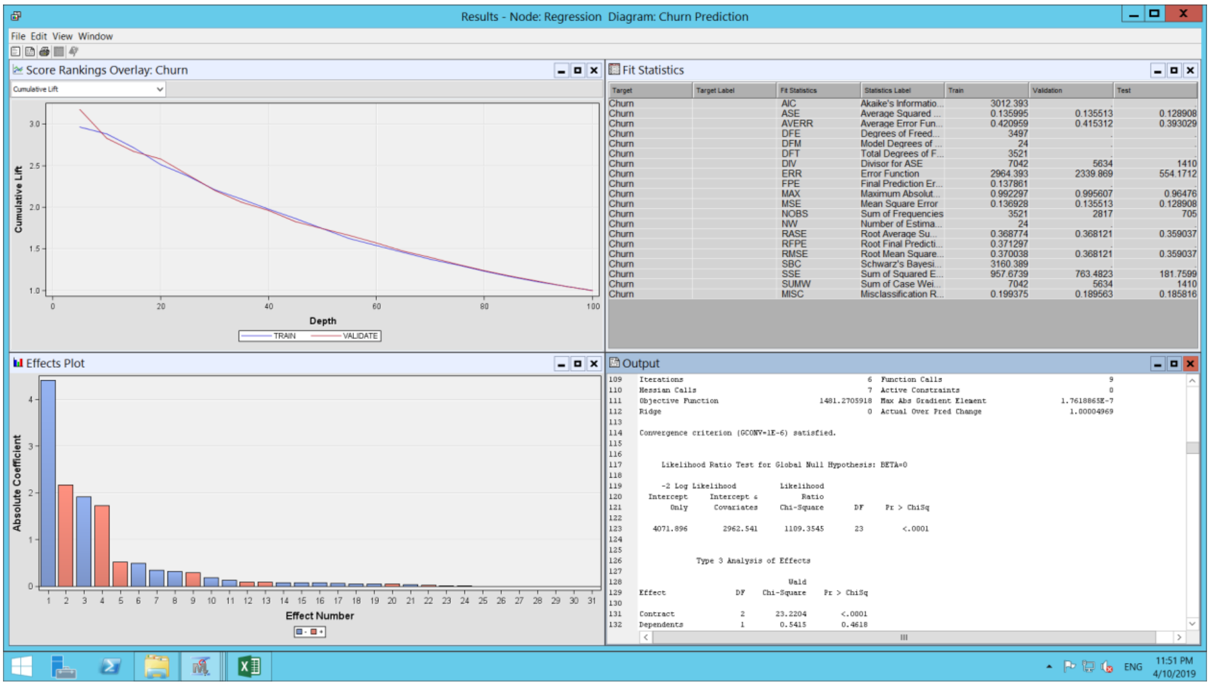
LOGISTIC REGRESSION

Telco Churn is a classification problem which classifies the users into users who churn and users who do not churn. Regression model approaches the prediction in a different way compared to decision trees. Regression assumes a specific association between input and target. Decision trees seek to isolate concentration of cases with like-valued target measurements.

As mentioned above data partition is done as 'training:validation:test' as 50:40:10.



Run the regression node without any specific input selection method and view the results. Fit statistics of the validation helps us to choose the best model.



View the Output window in Results. Initial lines of the output window give the Variable summary. Regression model for Telco churn consists of 6 Binary Inputs, 3 Interval Inputs, 10 Nominal Inputs and 1 Binary Target Variable.

```
Variable Summary

            Measurement    Frequency
 Role          Level         Count

INPUT       BINARY              6
INPUT       INTERVAL            3
INPUT       NOMINAL            10
TARGET      BINARY              1
```

Below the Variable Summary is the Model Information. This gives the details of the model. Regression model for Telco churn has target variable: Churn. As it is logistic regression, Link Function is mentioned as Logit.

Model Information

Training Data Set	WORK.EM_DMREG.VIEW
DMDB Catalog	WORK.REG_DMDB
Target Variable	Churn
Target Measurement Level	Ordinal
Number of Target Categories	2
Error	MBernoulli
Link Function	Logit
Number of Model Parameters	31
Number of Observations	3518

Type 3 analysis in the output window provides the statistical significance of the input variables and a value near 0 in the Pr > ChiSq column approximately indicates a significant input; a value near 1 indicates an extraneous input.

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Contract	2	23.2204	<.0001
Dependents	1	0.5415	0.4618
DeviceProtection	2	7.9269	0.0190
InternetService	1	2.7338	0.0982
MonthlyCharges	1	1.0252	0.3113
MultipleLines	2	17.1376	0.0002
OnlineBackup	1	0.3423	0.5585
OnlineSecurity	1	0.5221	0.4699
PaperlessBilling	1	12.3685	0.0004
Partner	1	0.2646	0.6070
PaymentMethod	3	15.8402	0.0012
PhoneService	0	0.0000	.
SeniorCitizen	1	4.2162	0.0400
StreamingMovies	1	2.0688	0.1503
StreamingTV	1	1.8303	0.1761
TechSupport	1	0.0319	0.8583
TotalCharges	1	2.0987	0.1474
gender	1	0.0236	0.8780
tenure	1	32.3837	<.0001

Input selection methods:

There are three methods of selecting the inputs for running a regression model:

1. Forward Selection: Forward selection creates a sequence of models of increasing complexity. The sequence starts with the baseline model, a model predicting the overall average target value for all cases. The algorithm searches the set of one-input models and selects the model that most improves on the baseline model. It then searches the set of two-input models that contain the input selected in the previous step and selects the model showing the most significant improvement. By adding a new input to those

selected in the previous step, a nested sequence of increasingly complex models is generated. The sequence terminates when no significant improvement can be made.

2. Backward Selection: Backward selection creates a sequence of models of decreasing complexity. The sequence starts with a saturated model, which is a model that contains all available inputs, and therefore, has the highest possible fit statistic. Inputs are sequentially removed from the model. At each step, the input chosen for removal least reduces the overall model fit statistic. This is equivalent to removing the input with the highest p-value. The sequence terminates when all remaining inputs have a p-value that is less than the predetermined stay cutoff.

3. Stepwise Selection: Stepwise Selection combines elements from both the forward and backward selection procedures. The method begins in the same way as the forward procedure, sequentially adding inputs with the smallest p-value below the entry cut-off. After each input is added, however, the algorithm re-evaluates the statistical significance of all included inputs. If the p-value of any of the included inputs exceeds the stay cut-off, the input is removed from the model and re-entered into the pool of inputs that are available for inclusion in a subsequent step. The process terminates when all inputs available for inclusion in the model have p-values in excess of the entry cut-off and all inputs already included in the model have p-values below the stay cut-off.

Stepwise Selection

Regression model for Telco churn is run with input selection method as stepwise selection. See the output window for each step.

```
Stepwise Selection Procedure


Step 0: Intercept entered.





The DMREG Procedure

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Parameter Estimates               1

                                    Optimization Start

Active Constraints                          0  Objective Function                  2035.947861
Max Abs Gradient Element         6.228351E-13

                                   Optimization Results

Iterations                                  0  Function Calls                                3
Hessian Calls                               1  Active Constraints                            0
Objective Function                2035.947861  Max Abs Gradient Element           6.228351E-13
Ridge                                       0  Actual Over Pred Change                       0

Convergence criterion (ABSGCONV=0.00001) satisfied.


     Likelihood Ratio Test for Global Null Hypothesis: BETA=0

     -2 Log Likelihood          Likelihood
  Intercept     Intercept &          Ratio
       Only      Covariates     Chi-Square       DF     Pr > ChiSq

   4071.896        4071.896         0.0000        0              .




                                   Analysis of Maximum Likelihood Estimates

                                                 Standard          Wald                      Standardized
Parameter                          DF    Estimate       Error    Chi-Square    Pr > ChiSq        Estimate    Exp(Est)

Intercept                           1     -1.0176      0.0382        710.42        <.0001                       0.361

```

After step 10, regression model is finalized with the following effects as significant:

1. Intercept
2. Contract
3. InternetService
4. MultipleLines
5. OnlineSecurity
6. PaperlessBilling
7. PaymentMethod
8. SeniorCitizen
9. StreamingMovies
10. TechSupport
11. tenure

The selected model is the model trained in the last step (Step 10). It consists of the following effects:

Intercept Contract InternetService MultipleLines OnlineSecurity PaperlessBilling PaymentMethod SeniorCitizen StreamingMovies TechSupport tenure

The type 3 analysis is follows:

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Contract	2	24.4044	<.0001
InternetService	2	92.4632	<.0001
MultipleLines	2	17.8271	0.0001
OnlineSecurity	1	12.0125	0.0005
PaperlessBilling	1	13.7359	0.0002
PaymentMethod	3	16.1435	0.0011
SeniorCitizen	1	5.3778	0.0204
StreamingMovies	1	8.1985	0.0042
TechSupport	1	4.4924	0.0340
tenure	1	136.8397	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept		1	-0.7503	0.1512	24.62	<.0001		0.472
Contract	Month-to-month	1	0.5226	0.1065	24.06	<.0001		1.686
Contract	One year	1	0.0830	0.1065	0.61	0.4361		1.087
InternetService	DSL	1	-0.0103	0.0862	0.01	0.9045		0.990
InternetService	Fiber optic	1	0.8308	0.0865	92.24	<.0001		2.295
MultipleLines	No	1	-0.2908	0.0758	14.71	0.0001		0.748
MultipleLines	No phone service	1	0.1602	0.1161	1.90	0.1678		1.174
OnlineSecurity	No	1	0.2024	0.0584	12.01	0.0005		1.224
OnlineSecurity	No internet service	0	0	.	.	.	.	.
PaperlessBilling	No	1	-0.1936	0.0522	13.74	0.0002		0.824
PaymentMethod	Bank transfer (automatic)	1	-0.0780	0.0929	0.70	0.4011		0.925
PaymentMethod	Credit card (automatic)	1	-0.0911	0.0936	0.95	0.3302		0.913
PaymentMethod	Electronic check	1	0.2932	0.0731	16.08	<.0001		1.341
SeniorCitizen	0	1	-0.1370	0.0591	5.38	0.0204		0.872
StreamingMovies	No	1	-0.1490	0.0520	8.20	0.0042		0.862
StreamingMovies	No internet service	0	0	.	.	.	.	.
TechSupport	No	1	0.1255	0.0592	4.49	0.0340		1.134
TechSupport	No internet service	0	0	.	.	.	.	.
tenure		1	-0.0367	0.00314	136.84	<.0001	-0.4940	0.964

As seen in the above figure, Estimate is maximum for Internet Service with Fiber Optic. Similarly, contract as Month-to-Month also has significant effect.

Odds Ratio Estimates

Effect		Point Estimate
Contract	Month-to-month vs Two year	3.090
Contract	One year vs Two year	1.991
InternetService	DSL vs No	2.248
InternetService	Fiber optic vs No	5.213
MultipleLines	No vs Yes	0.656
MultipleLines	No phone service vs Yes	1.030
OnlineSecurity	No vs Yes	1.499
OnlineSecurity	No internet service vs Yes	.
PaperlessBilling	No vs Yes	0.679
PaymentMethod	Bank transfer (automatic) vs Mailed check	1.047
PaymentMethod	Credit card (automatic) vs Mailed check	1.034
PaymentMethod	Electronic check vs Mailed check	1.518
SeniorCitizen	0 vs 1	0.760
StreamingMovies	No vs Yes	0.742
StreamingMovies	No internet service vs Yes	.
TechSupport	No vs Yes	1.285
TechSupport	No internet service vs Yes	.
tenure		0.964

Odds ratio estimate shows that churn rate for a month-to-month contract to two-year contract is approximately 3 to 1. Similarly, churn rate for Fiber Optic to No internet service is approximately 5 to 1. Similar observations can be deduced from the above figure.

The fit statistics of the model is as given below:

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		AIC	Akaike's Information Criterion	3003.637	.	.
Churn		ASE	Average Squared Error	0.136522	0.13617	0.130005
Churn		AVERR	Average Error Function	0.421988	0.41757	0.395775
Churn		DFE	Degrees of Freedom for Error	3505	.	.
Churn		DFM	Model Degrees of Freedom	16	.	.
Churn		DFT	Total Degrees of Freedom	3521	.	.
Churn		DIV	Divisor for ASE	7042	5634	1410
Churn		ERR	Error Function	2971.637	2352.591	558.0425
Churn		FPE	Final Prediction Error	0.137769	.	.
Churn		MAX	Maximum Absolute Error	0.991027	0.993869	0.957357
Churn		MSE	Mean Square Error	0.137145	0.13617	0.130005
Churn		NOBS	Sum of Frequencies	3521	2817	705
Churn		NW	Number of Estimate Weights	16	.	.
Churn		RASE	Root Average Sum of Squares	0.369489	0.369013	0.360561
Churn		RFPE	Root Final Prediction Error	0.371172	.	.
Churn		RMSE	Root Mean Squared Error	0.370332	0.369013	0.360561
Churn		SBC	Schwarz's Bayesian Criterion	3102.301	.	.
Churn		SSE	Sum of Squared Errors	961.3895	767.1846	183.3064
Churn		SUMW	Sum of Case Weights Times Freq	7042	5634	1410
Churn		MISC	Misclassification Rate	0.195967	0.191338	0.194326

DECISION TREE

The above problem is a classic classification problem. The problem at hand is to classify the customer records into 2 classes:

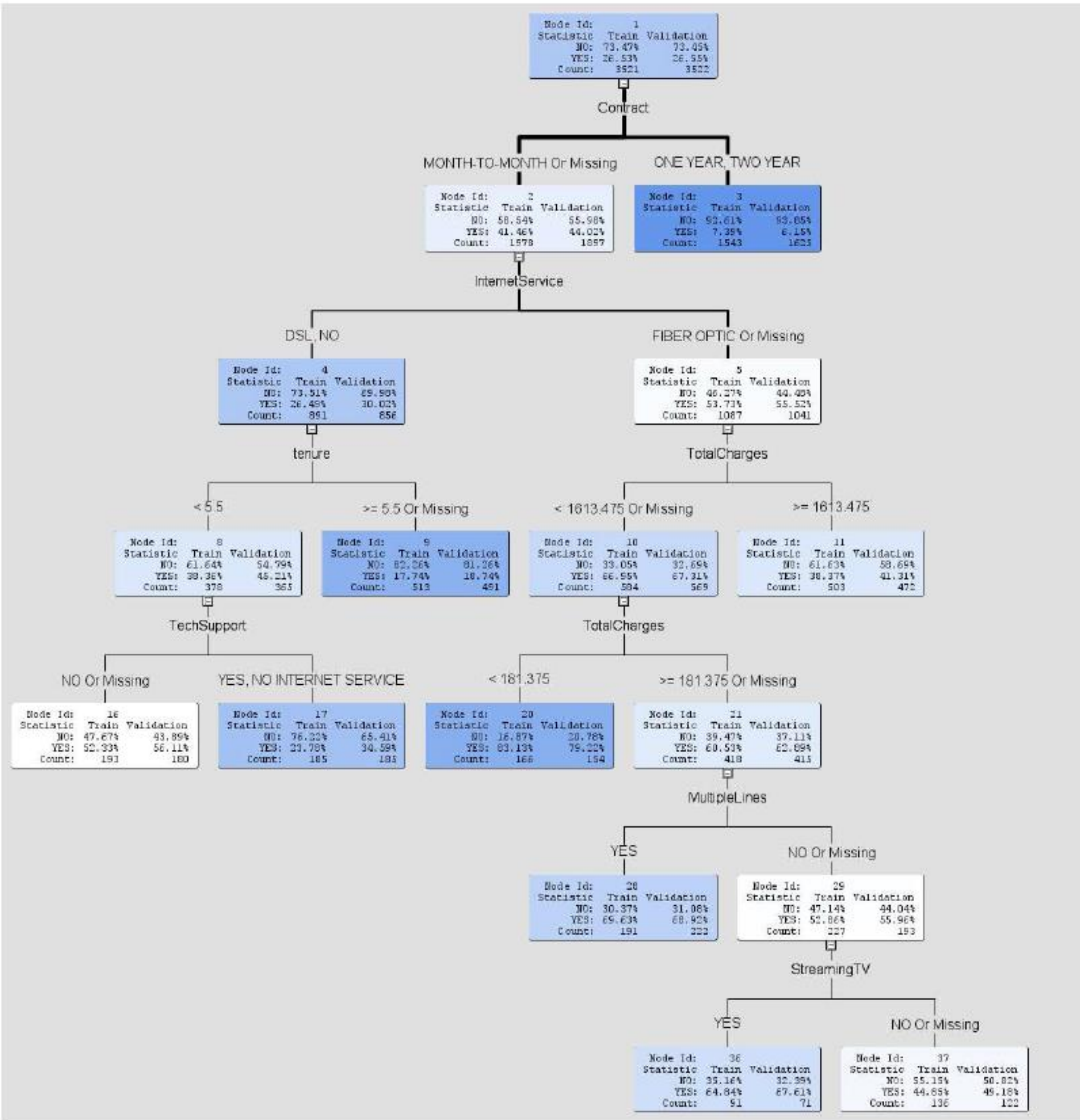
- Ones who churn
- Ones who would stay

So, the choice of the model decision tree is obvious.

Approach #1: Autonomous

Decision trees can be cultivated by two methods: autonomously and interactively. To begin with and have clarity, the decision tree is cultivated autonomously.

The autonomously build model is as shown below:

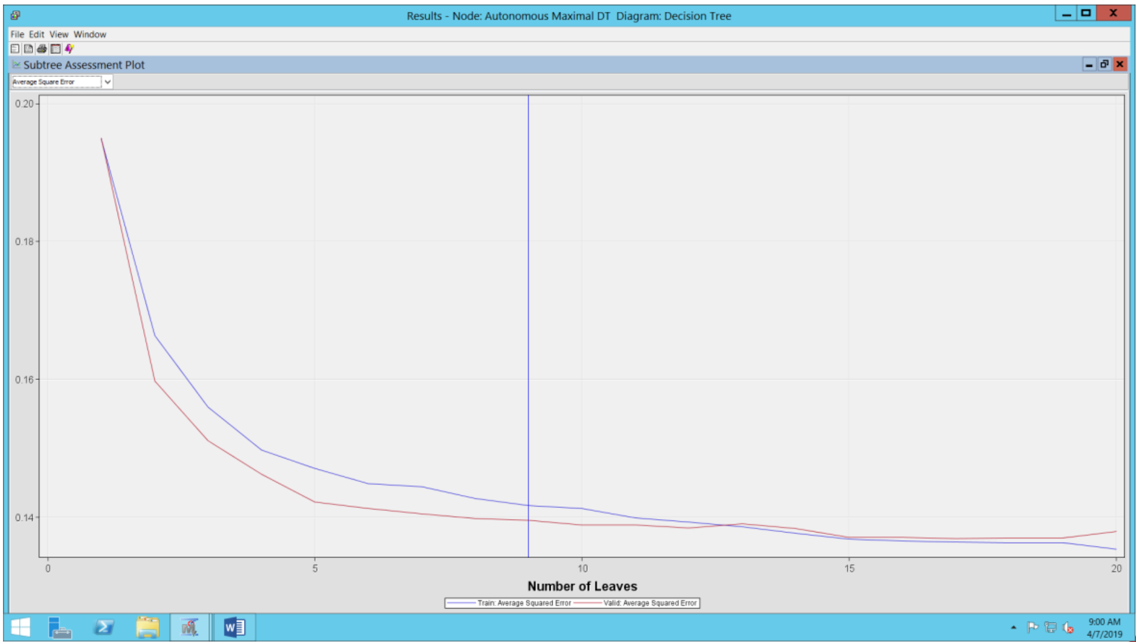


The highlight of decision trees is that they are easy to comprehend. The above model is the maximal tree and hence it is a bit complex. So, there is a need to pre-prune or post-prune the trees.

The maximal decision trees work well with the train data and misclassifies the validation data usually. This indicates that it might not work well with a new dataset. Given below is the study on misclassification rate of the maximal tree that was created by autonomous decision tree cultivation.

Much of a misclassification here is not seen here. So, the autonomously build model explains the data well. It may be also concluded that there is no model overfitting issue. But the SAS tool suggests that an optimal tree would be the one with 9 leaves.

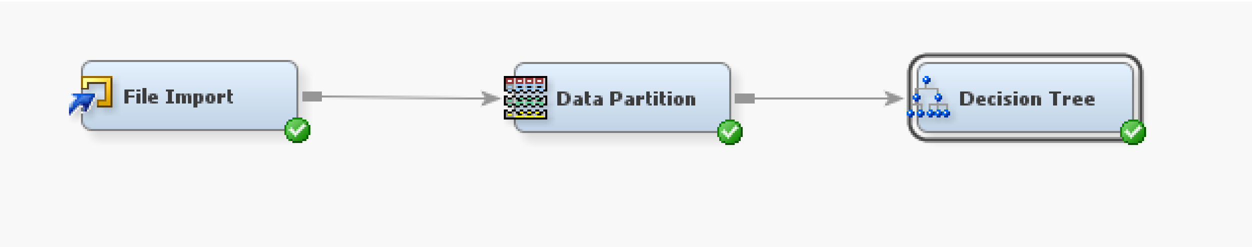
The Average Square Error plot is as seen below.



The model fit is good. But the optimal tree would be one with 9 leaves. Model fit can be explored by using the fit statistics. The fit statistics is as below:

Fit Statistics	Statistics Label	Train	Validation
NOBS	Sum of Frequencies	3521	3522
MISC	Misclassification Rate	0.202499	0.202726
MAX	Maximum Absolute Error	0.926118	0.926118
SSE	Sum of Squared Errors	998.2264	983.6685
ASE	Average Squared Error	0.141753	0.139646
RASE	Root Average Squared Error	0.376501	0.373693
DIV	Divisor for ASE	7042	7044
DFT	Total Degrees of Freedom	3521	.

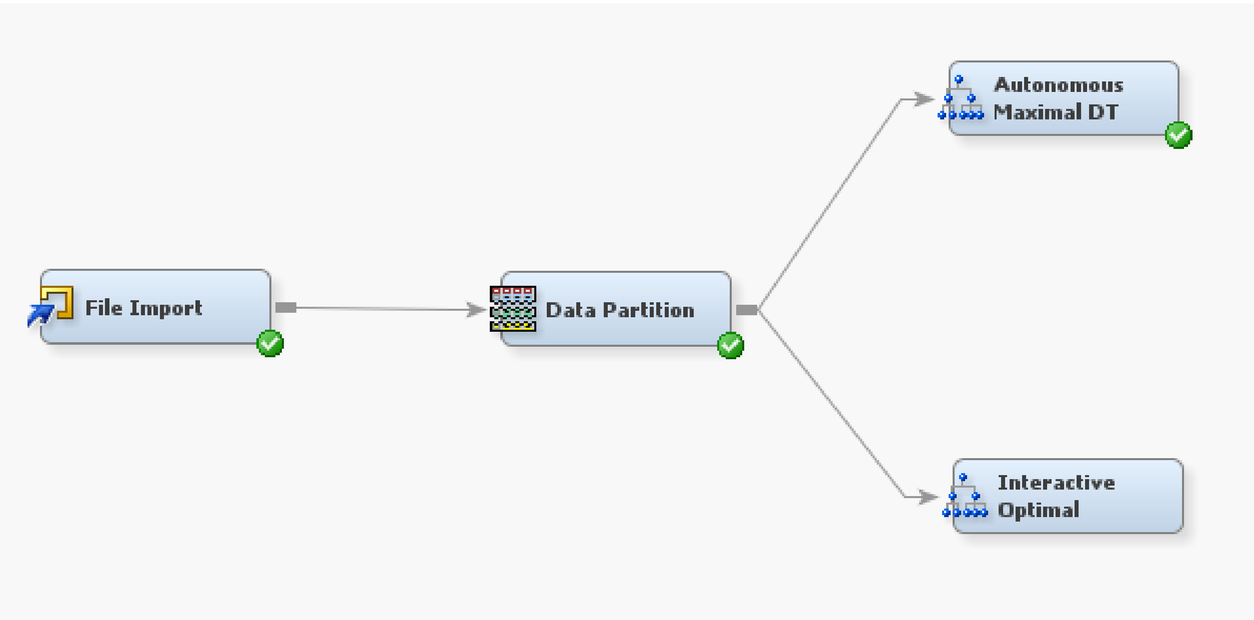
The fit statistics proves excellent fit of data. The status of the diagram is as seen below:



The next part of the analysis delves into building lesser complex and easily comprehensible model that has an excellent fit, low misclassification rate and good purity splits. For this, an interactive decision tree cultivation method will be made use of.

Approach #2: 2-way Interactive Optimal

The diagram looks as below after the addition of the new node:



Unlike the previous autonomous maximal tree, the number of splits here to have to be fixed to attain a simpler model that can be explained better and easily comprehended. Such a model will help implement the business decisions easier. So, the settings of the splitting rule look as below:

Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	2
Maximum Depth	2
Minimum Categorical Size	5

The main change here is that the maximum depth is made '2' The maximum depth is kept as '2', i.e., 2 generations.

The –log(p) value or logWorth is made use of to choose the splitting criterion. The logWorth value table looks as below:

Split Node 1

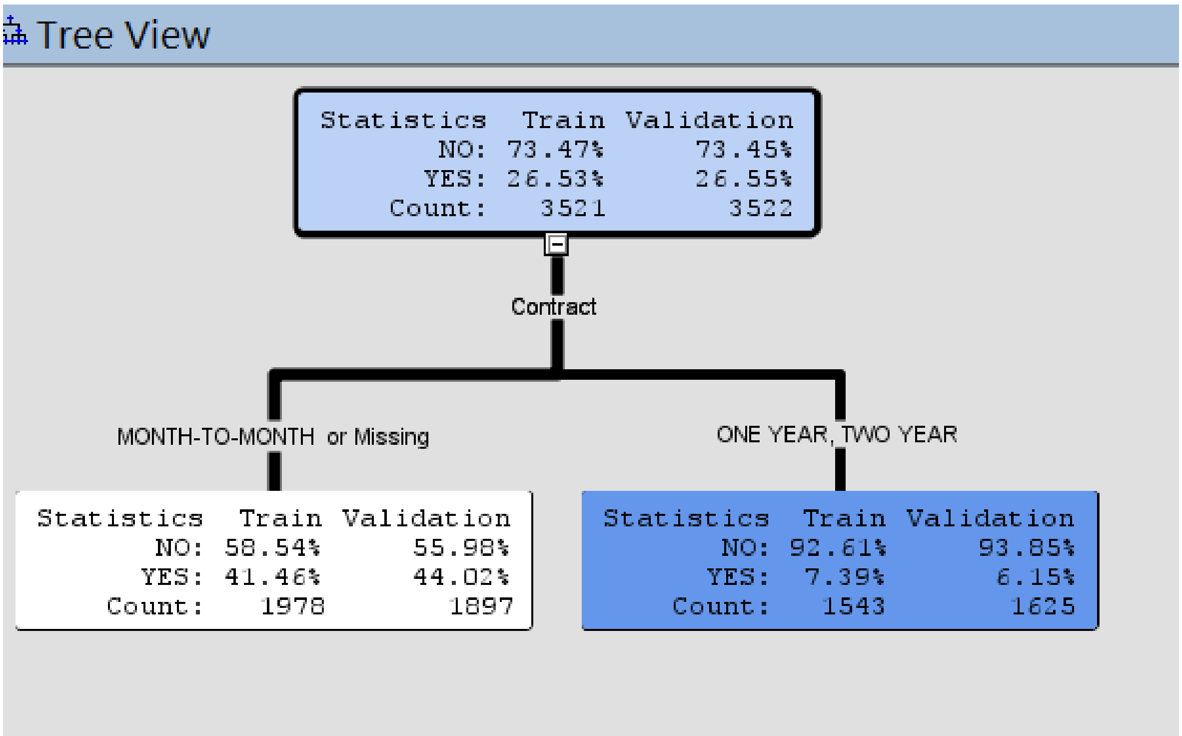
Target Variable: Churn

Variable	Variable Description	-Log(p)	Branches
Contract	Contract	113.0658	2
OnlineSecurity	OnlineSecurity	88.0719	2
TechSupport	TechSupport	87.4071	2
tenure	tenure	75.5167	2
InternetService	InternetService	72.6616	2
PaymentMethod	PaymentMethod	70.2713	2
OnlineBackup	OnlineBackup	54.0441	2
DeviceProtection	DeviceProtection	53.7843	2
StreamingMovies	StreamingMovies	43.5919	2
StreamingTV	StreamingTV	43.5919	2
MonthlyCharges	MonthlyCharges	37.6701	2
TotalCharges	TotalCharges	34.2042	2
PaperlessBilling	PaperlessBilling	29.9328	2
Dependents	Dependents	20.5425	2
SeniorCitizen	SeniorCitizen	18.7215	2
Partner	Partner	13.2178	2
MultipleLines	MultipleLines	1.7939	2
PhoneService	PhoneService	0.4525	2
gender	gender	0.1622	2

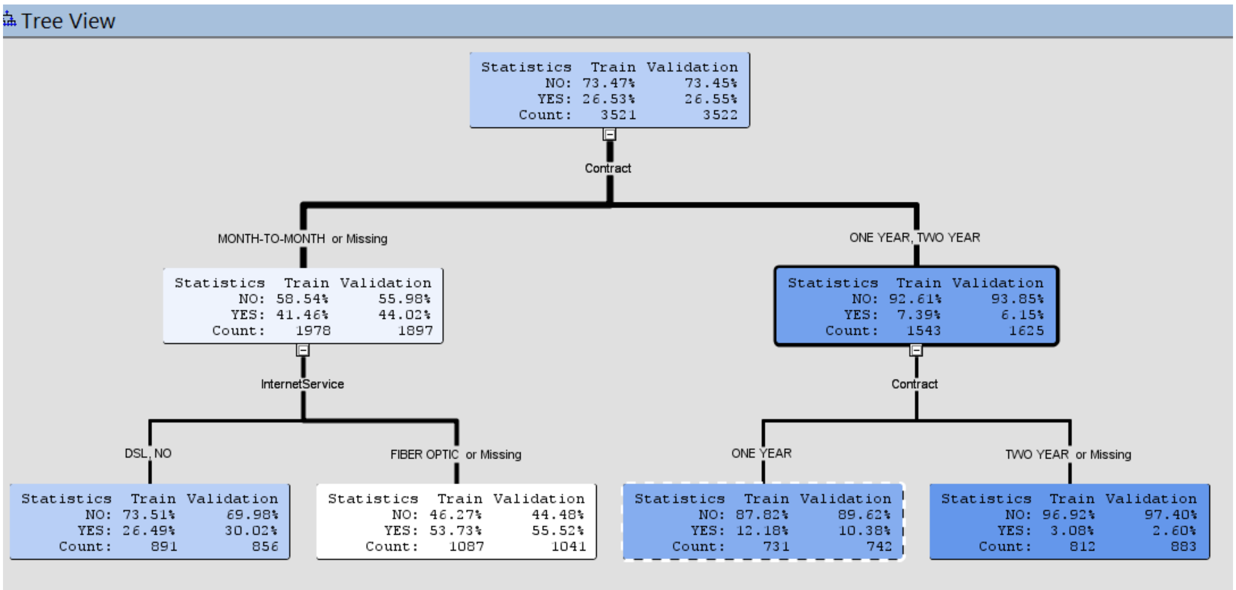
Edit Rule...

OK Cancel Apply Refresh

So, decision tree is split with contract type and the decision tree looks as below:

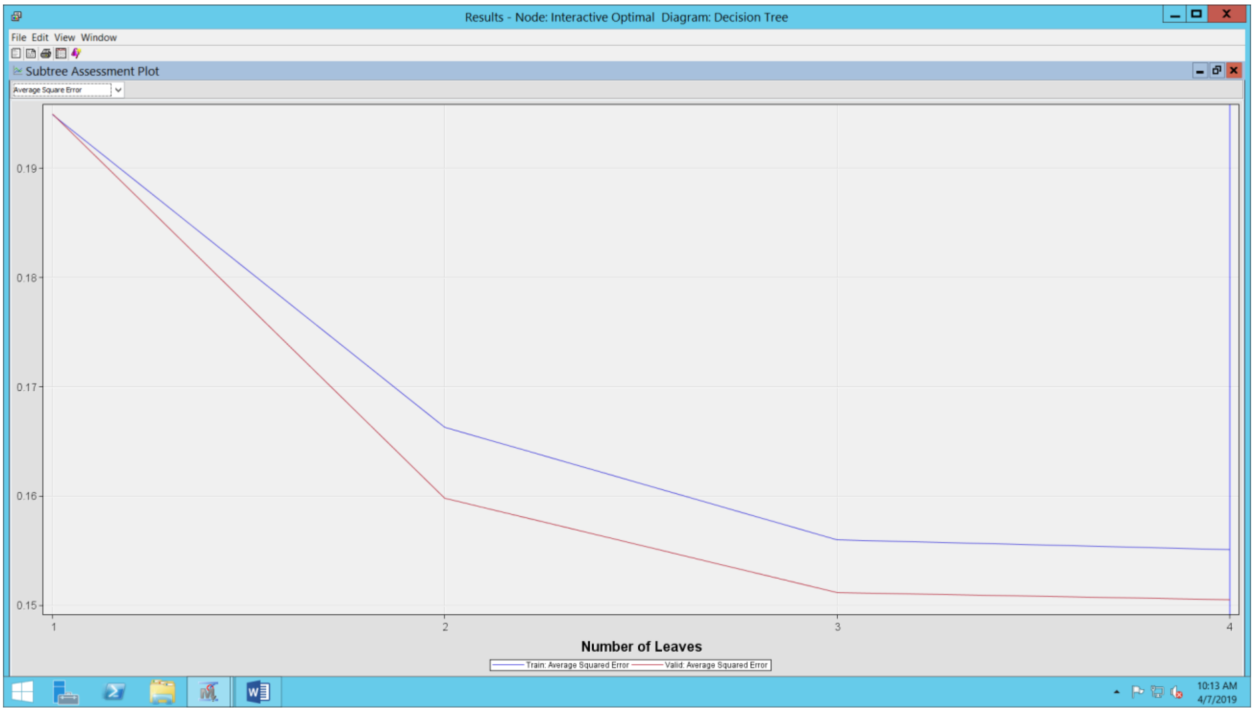


Similarly, subsequent splits are carried out for the Generation 1 nodes. The final result is as below. As decided, cultivation is stopped at 2nd generation level.

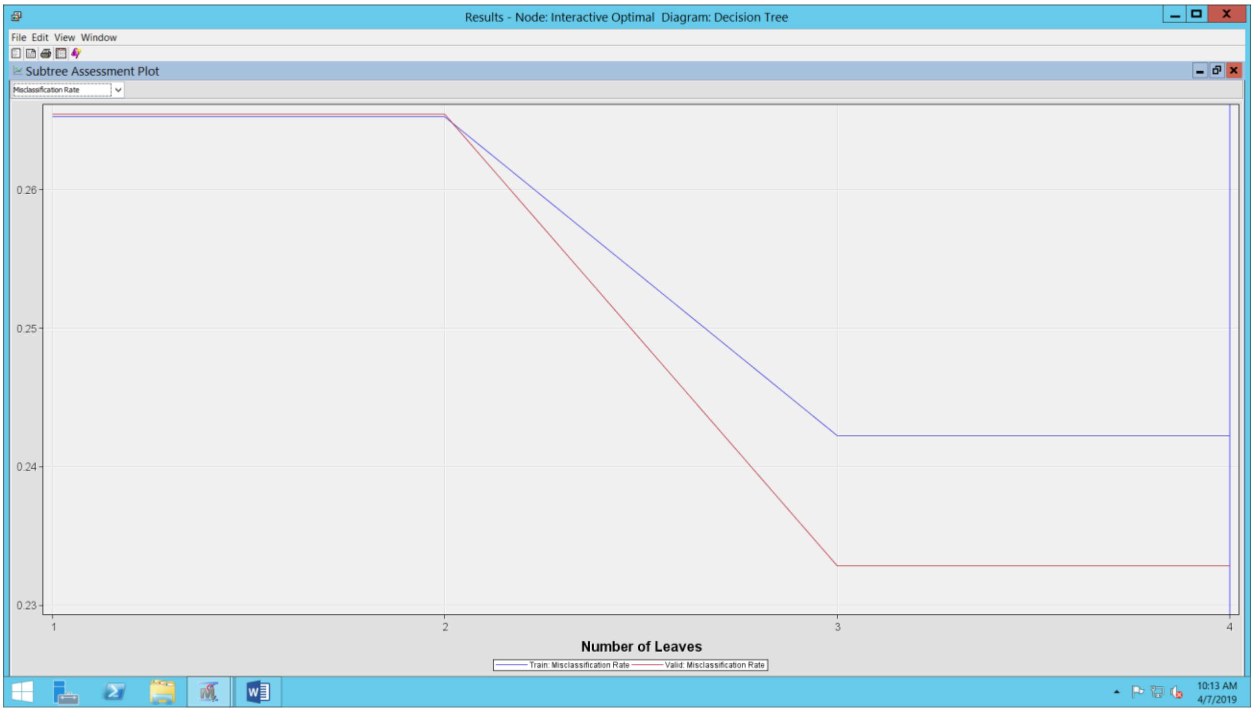


The tree is frozen at this level and model assessment is carried out.

The Average Square Error plot is as seen below.



The Misclassification rate plot is as seen below.



The misclassification is beginning slowly. Hence the tree construction is stopped at this point.

The Fit Statistics is as seen below.

Fit Statistics	Statistics Label	Train	Validation
NOBS	Sum of Frequencies	3521	3522
MISC	Misclassification R...	0.242261	0.232822
MAX	Maximum Absolut...	0.969212	0.969212
SSE	Sum of Squared E...	1092.252	1059.982
ASE	Average Squared ...	0.155105	0.15048
RASE	Root Average Squ...	0.393834	0.387918
DIV	Divisor for ASE	7042	7044
DFT	Total Degrees of F...	3521	.

Approach #3: 3-way Interactive decision tree

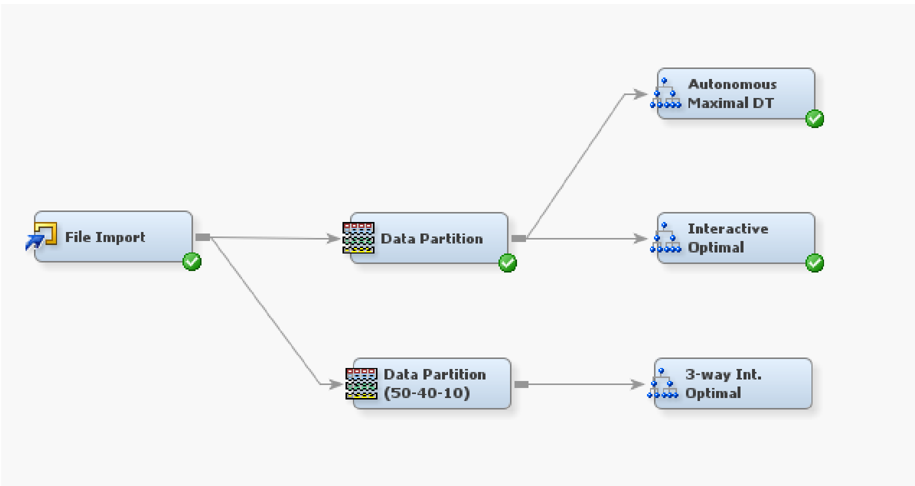
Subsequent discussion raised the following concern:

"Each categorical variable has 3 possible values. A 3-way decision tree suits the dataset"

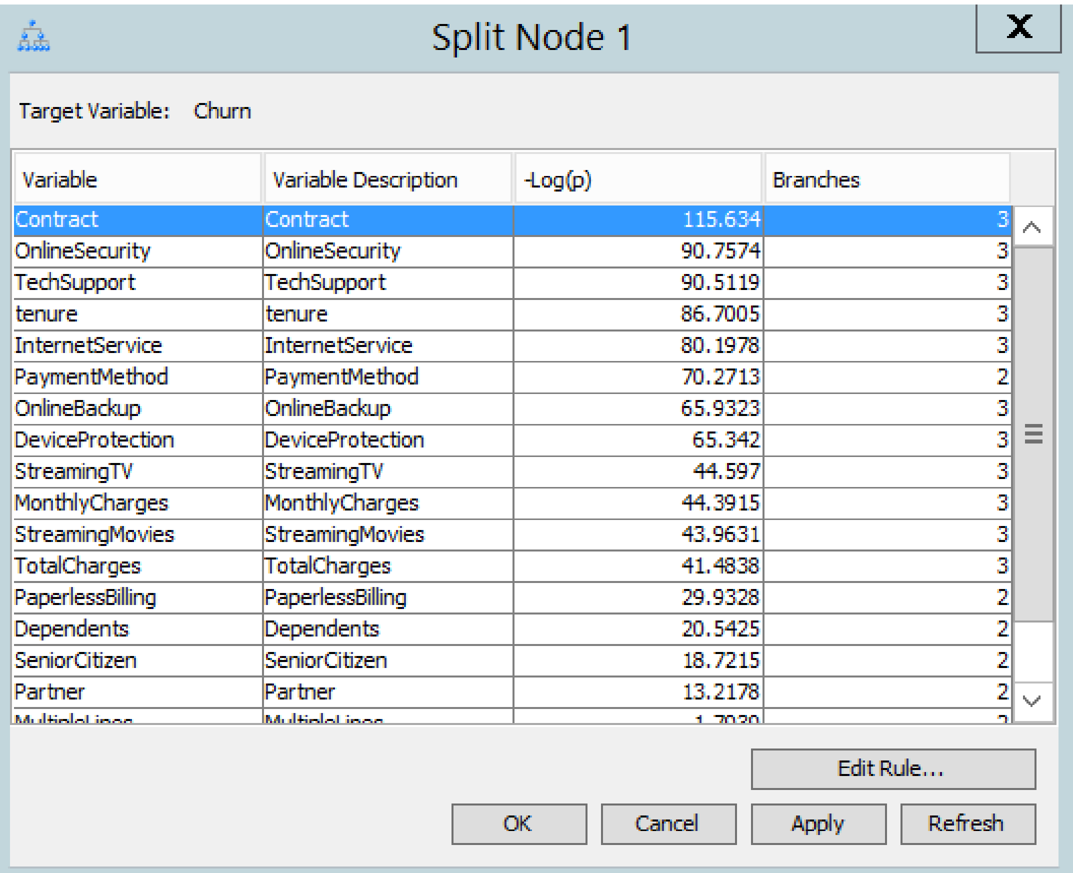
Then, the Decision tree node is added, and the maximum depth is modified as 3 generations and the maximum branch is modified as 3. See the changes below:

Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	3
Maximum Depth	3
Minimum Categorical Size	5

The workspace diagram looks as below.

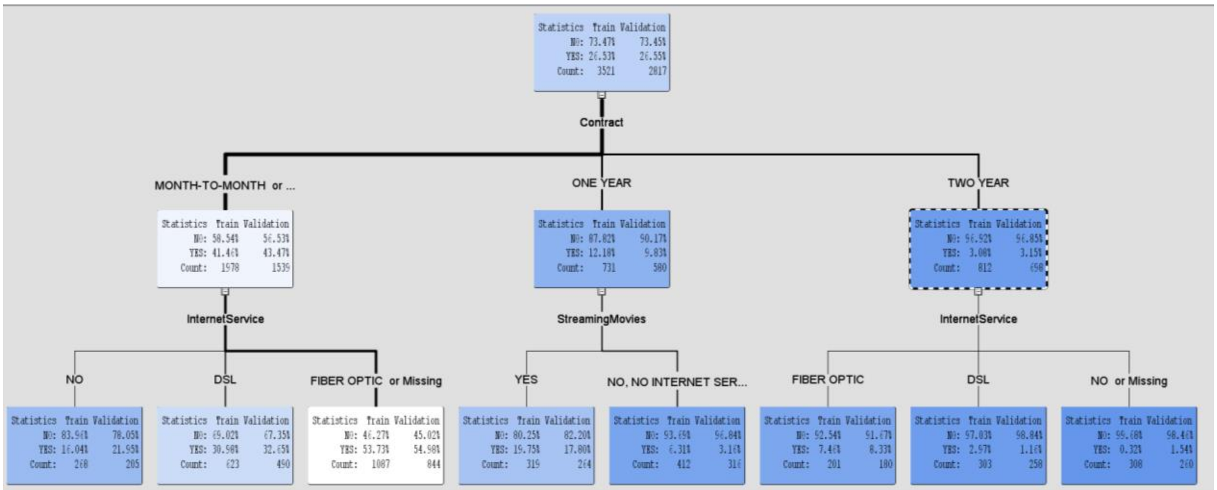


After opening the interactive mode, check the LogWorth of the splits and then create the split. The LogWorth of the first split is as seen below:



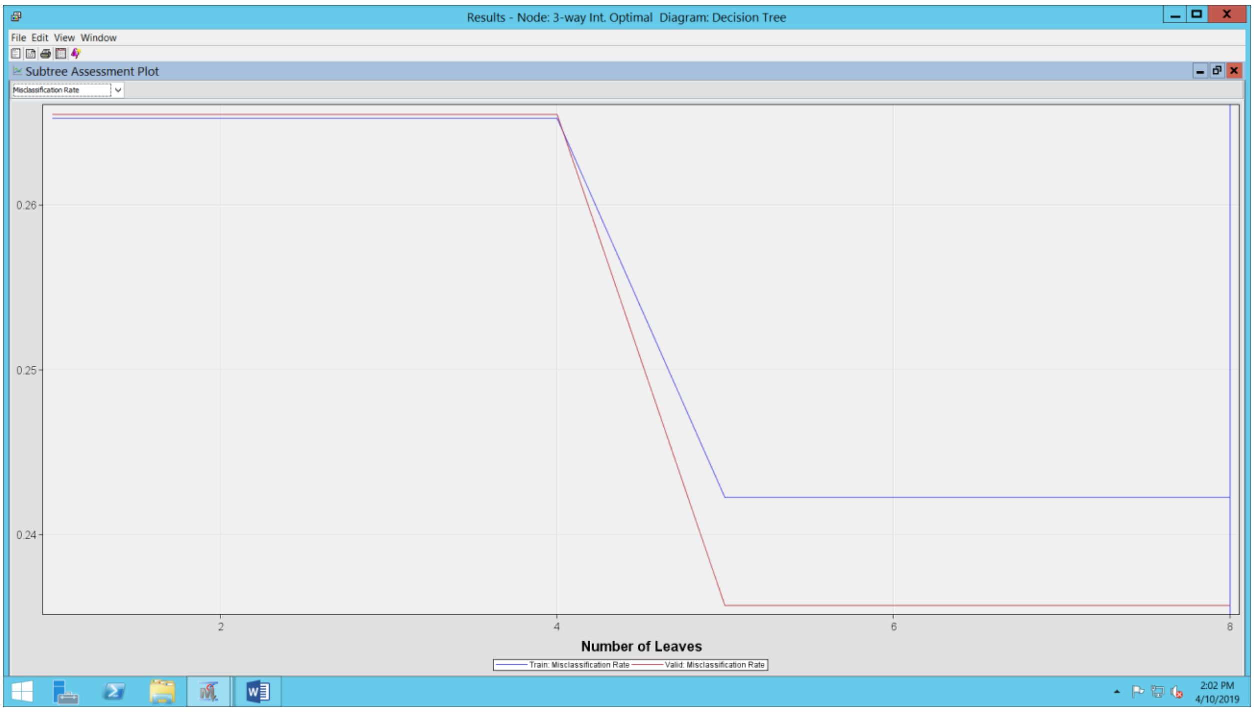
Variable	Variable Description	-Log(p)	Branches
Contract	Contract	115.634	3
OnlineSecurity	OnlineSecurity	90.7574	3
TechSupport	TechSupport	90.5119	3
tenure	tenure	86.7005	3
InternetService	InternetService	80.1978	3
PaymentMethod	PaymentMethod	70.2713	2
OnlineBackup	OnlineBackup	65.9323	3
DeviceProtection	DeviceProtection	65.342	3
StreamingTV	StreamingTV	44.597	3
MonthlyCharges	MonthlyCharges	44.3915	3
StreamingMovies	StreamingMovies	43.9631	3
TotalCharges	TotalCharges	41.4838	3
PaperlessBilling	PaperlessBilling	29.9328	2
Dependents	Dependents	20.5425	2
SeniorCitizen	SeniorCitizen	18.7215	2
Partner	Partner	13.2178	2

At this second generation, the misclassification has begun. So, the tree cultivation is stopped at this point. The tree looks as below:

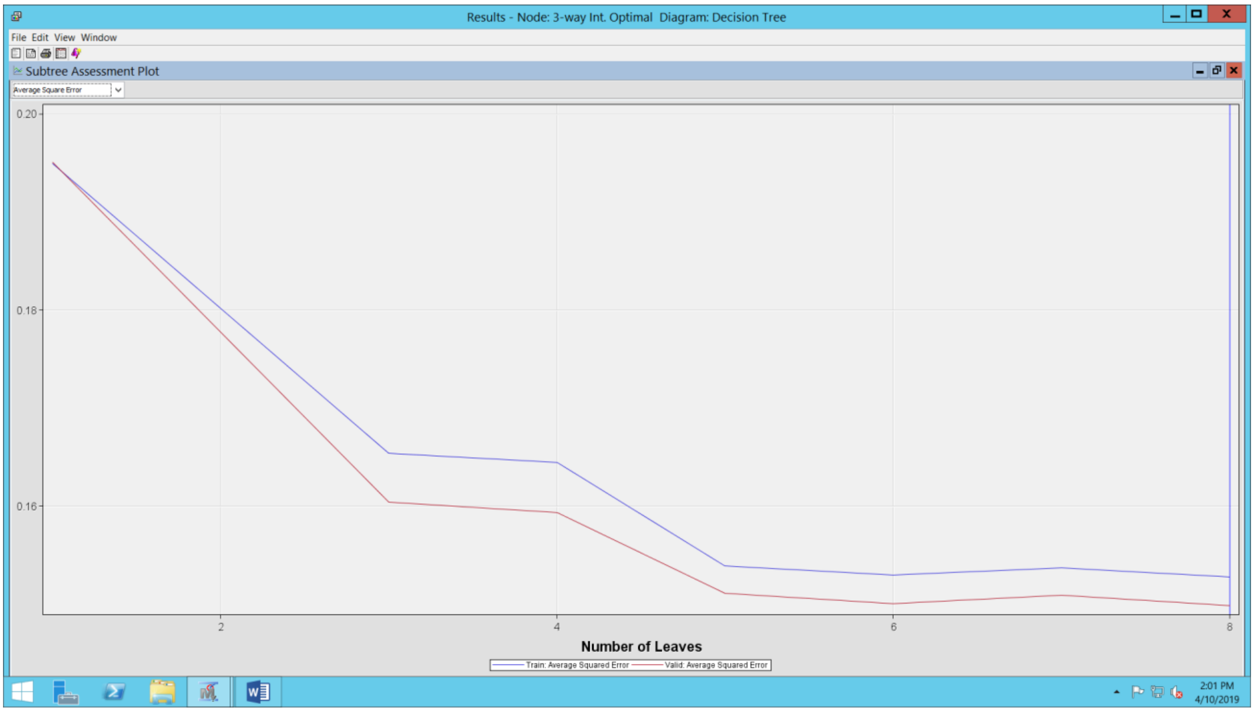


The Model assessment plot are as below:

a) Misclassification Rate



b) Average Square Error



c) Fit Statistics

STAT	LABEL	TRAIN ▼	VALIDATE	TEST
DIV	Divisor for ASE	7042	5634	1410
NOBS	Sum of Frequencies	3521	2817	705
DFT	Total Degrees of Freedom	3521	.	.
SSE	Sum of Squared Errors	1076.159	844.5803	201.066
MAX	Maximum Absolute Error	0.996753	0.996753	0.936893
RASE	Root Average Squared Error	0.390922	0.387179	0.377624
MISC	Misclassification Rate	0.242261	0.235712	0.221277
ASE	Average Squared Error	0.15282	0.149908	0.1426

Observations

1. After the decision tree cultivation, the following observations are made
2. Contracts have significant impact on how people churn.
3. The Internet technology type significantly impacts the model.
4. Streaming Movies is an important factor that decides churning in 1-year contracts.

BUSINESS INSIGHTS

From the conducted analysis, the following business insights are very evident.

1. The customers with long-term contracts tend to be loyalists. They stay with company longer and has a low churn.
2. Among the customers with month-to-month contract, the ones with 'no internet service' will stay longer.
3. The customer who have adopted new technologies (Fibre optics) have a high probability of churning.
4. The general trend is that people prone to older technology stay longer and vice-versa.

MANAGERIAL IMPLICATIONS

From a Sales Manager's standpoint the following strategies are warranted.

1. Seal Long terms contracts with customers by selling gadgets and telco connections together.
2. Keep the customers with relatively new technology under close supervision; provide them customized offers. Try to sell branded products to this segment and slowly turn them into loyalists. Make them progress the ladder of loyalty.
3. Even though, the people with older technology stay longer, if they do not venture into the new products, the company cannot progress. The company must strike a balance here and slowly transition them into newer technology.
4. Customers who are onboarding the new technologies should be awarded with lucrative offers that will make them stay. The clustering analysis done will provide more insights into choice of segment.

CONCLUSION

As a conclusion, Customer churn can never be completely avoided. But, iterative improvements can reduce this phenomenon significantly and reduce the revenue loss. In a telecommunication giant, decision making, and implementation must be done with utmost caution. The supervised learning conducted with Telco dataset gives us a picture that Contracts, and type of technology has a lot of say in the phenomenon of Customer Churning. Adequate measures must be taken to avoid this.

REFERENCES

[1] Shah, Rishi. “3 Simple Steps to Increase Customer Retention.” Getting More Awesome, 23 Dec. 2011, www.gettingmoreawesome.com/2011/05/04/3-simple-steps-to-increase-customer-retention/.

[2] Customer attrition. (n.d.). Retrieved from https://www.wikiwand.com/en/Customer_attrition#/Management

[3] “How to Build a Prediction-Based Customer Churn Program.” Qualtrics, 25 June 2018, www.qualtrics.com/blog/customer-churn/.

[4] Amin, Adnan & Anwar, Sajid & Adnan, Awais & Nawaz, Muhammad & Alawfi, Khalid & Hussain, Amir & Huang, Kaizhu. (2016). Customer Churn Prediction in Telecommunication Sector using Rough Set Approach. Neurocomputing. 10.1016/j.neucom.2016.12.009.

[5] BlastChar. (2018, February 23). Telco Customer Churn. Retrieved from https://www.kaggle.com/blastchar/telco-customer-churn