Sawthamini Nadesapillai

Chanveer Singh

Anandha Padmanabhan Sasi

Christina Varghese

Big Data & Predictive Analysis

"Unlocking Customer Insights: Building a Predictive Model for Gold Membership Campaign at Superstore"

Table of Contents

Abstract

For the approaching year-end sale at a superstore, we give an extensive study in this abstract that focuses on exploiting prior campaign data to improve customer targeting and boost Gold Membership sales. The main goal of the study is to discover the critical variables that affect customers' reactions to the Gold Membership offer and to build a solid prediction model to precisely estimate high levels of customer satisfaction. The goal is to develop an effective plan for precisely targeting clients using the insights gained from this thorough study, maximising the effectiveness of the phone call campaign while lowering associated costs.

The main goals of the study include a detailed analysis of past campaign data to identify key variables influencing client responses. The research aims to develop a comprehensive knowledge of the dynamics at play by addressing fundamental topics, such as the discriminating traits of Gold Membership adopters and the impact of consumer demographics on engagement decisions. In order to determine the most efficient model for determining the ideal client group, the research also aims to assess various predictive models, including Decision Tree, Logistic Regression, and Neural Network. The study also looks at strategic strategies that can improve consumer targeting efforts and subsequently boost sales of Gold Memberships.

A set of specific recommendations is produced in light of the findings. The first tip is to target people with higher education levels and income groups in order to take advantage of the strong influence that education and income have on gold membership purchases. Second, taking into account a customer's marital status emerges as a crucial advice, stating that groups like those who are divorced, single, married, or in a relationship need different communication tactics. To maximise customer engagement, high-value product promotion, improved website engagement, and customised recency-based techniques are also crucial. Another important strategy is to target families with particular teen demographics. Finally, it is advised to use a neural network model with 50 iterations, 3 hidden units, and backward regression because of its precision, interpretability, and useful information.

Background

In preparation for their year-end sale, a superstore is strategizing to introduce a new promotional offer termed "Gold Membership," entailing a 20% discount on all purchases. This offer will be exclusively available to existing customers, and the store aims to optimize its outreach through a phone call campaign that targets potential offer adopters. In a bid to enhance campaign cost-effectiveness, the store aims to develop a predictive model that can categorize customers who are more inclined to take up the gold membership offer.

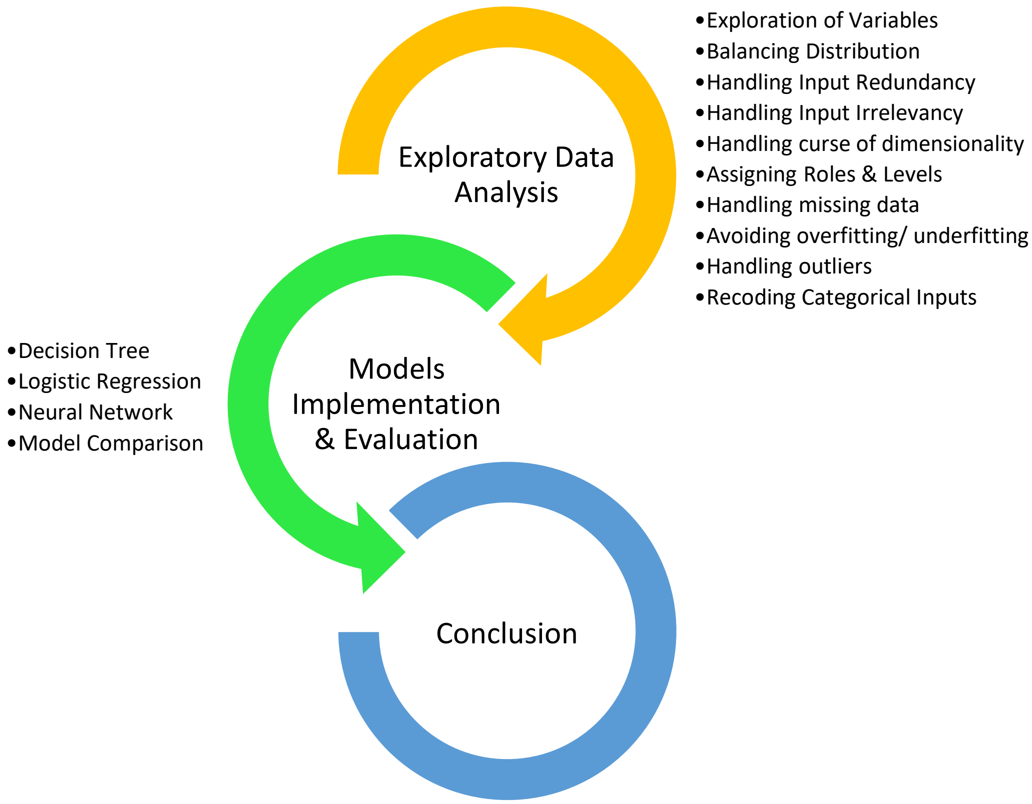
The primary objective of this project revolves around an in-depth analysis of historical data gathered from the previous year's campaign. The intention is to identify the pivotal factors that significantly impact customer responses to the Gold Membership offer during the upcoming year-end sale at the superstore. By harnessing insights derived from this comprehensive analysis, the goal is to formulate a robust predictive model capable of accurately gauging the probability of a positive customer response to the Gold Membership proposal. The anticipated outcome of this predictive model is to enable the superstore to precisely target the right set of customers, thereby optimizing the efficacy of the phone call campaign. Ultimately, this approach aims to maximize the attainment of the gold membership promotion's success while simultaneously curtailing campaign-associated expenses.

The research endeavors to address key questions in achieving the objectives. These include:

1. Identification of the discerning characteristics defining superstore customers more inclined to embrace the Gold Membership offer.
2. Exploration of whether customer demographics significantly influence the decision to engage with the membership offer or not.
3. Comparative evaluation of predictive models, including Decision Tree, Logistic Regression, and Neural Network, to ascertain the most effective model for pinpointing the right customer segment.
4. Exploration of strategic initiatives that can be adopted to enhance customer targeting and subsequently boost Gold Membership sales.

In conclusion, this initiative seeks to transform historical campaign data into actionable insights, effectively guiding the superstore's approach to its year-end sale. The envisaged predictive model's accuracy and insights are anticipated to provide a competitive edge, ensuring the superstore's optimal utilization of resources and the realization of successful Gold Membership sales during the upcoming year-end sale event.

Methodology



Tools:

The exploratory data analysis of the previous campaign data was conducted using SAS Miner 15.2. This software was employed to visualize the data and construct classification models, allowing for a comprehensive understanding of the dataset, and facilitating the creation of predictive models for customer response classification.

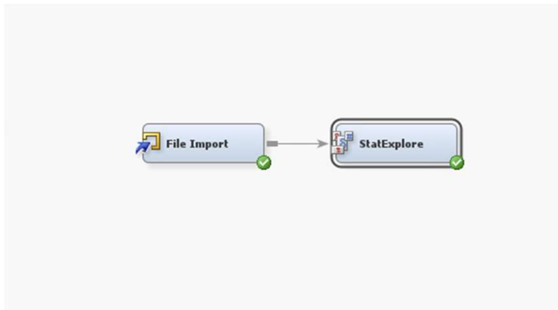
Github Sources:

https://github.com/anandpn95/Superstore_Final-Project

Exploratory Data Analysis

Exploration of Variables

Data obtained was added to the SAS miner project using the file import node. Then StatExplore node obtained from Explore connected with File Import node to investigate the variables.



The data provided contains 22 variables and 2240 Observations. Out of there are 7 nominal and 15 interval variables.

The 'Response' variable indicates whether a customer accepted an offer in the last campaign, with a value of 1 for acceptance and 0 for rejection. Each customer is identified by a unique 'ID'. The 'Year_Birth' field represents the customer's birth year. 'Complain' is a binary variable indicating whether the customer made any complaints in the last 2 years. 'Dt_Customer' represents the enrollment date of the customer with the company.

The following attributes are related to the customer's household and lifestyle: 'Education' represents the customer's level of education, while 'Marital' indicates their marital status. 'Kidhome' and 'Teenhome' represent the number of small children and teenagers in the customer's household, respectively. 'Income' represents the yearly household income.

The next set of attributes ('MntFishProducts', 'MntMeatProducts', 'MntFruits', 'MntSweetProducts', 'MntWines', and 'MntGoldProds') represent the amounts spent by the customer on various product categories in the last 2 years. The attributes 'NumDealsPurchases', 'NumCatalogPurchases', 'NumStorePurchases', and 'NumWebPurchases' represent the number of purchases made by the customer under different conditions (e.g., discounts, catalog, in-store, and web). 'NumWebVisitsMonth' represents the number of visits the customer made to the company's website in the last month, while 'Recency' represents the number of days since the customer's last purchase.

Summary of Index Variable

Variable	Description
ID	Unique ID of customers
Dt_Customer	Enrollment date of the customer with the superstore.

Summary of Class Variables

Variable	Description	Level	Categories- Count
Response	This variable indicates whether a customer accepted an offer in the last campaign, with a value of 1 for acceptance and 0 for rejection.	2	0-1906 1-334
Complain	This variable indicates whether the customer made any complaints in the last 2 years, with a value of 1 for Yes and 0 for No.	2	0-2219 1- 21
Education	This variable indicated the education level of customer (2nd cycle, Basic, Graduation, Master, PhD).	5	2nd Cycle-203 Basic-54 Graduation-1127 Master-370 PhD-486
Marital_Status	Marital status of the customer (Absurd, Alone, Divorced, Married, Single, Together, Widow, YOLO)	8	Divorced-232 Married-864 Single- 480 Together-580 Widow-77
Kidhome	Number of small children in the customer's household	3	0-1293 1-899 2-48
Teenhome	Number of teenagers in the customer's household	3	0-1158 1-1030 2-52
Year_Birth	Birth year of the customer.	54	

Summary of Interval Variables

Variable	Description	Min	Max	Mean	Std
Income	Yearly household income of customer	1730	666666	52247.25	25173.08
MntFishProducts	Amount spent by the customer on fish product category in the last 2 years.	0	259	37.52	54.62
MntMeatProducts	Amount spent by the customer on meat product category in the last 2 years.	0	1725	166.95	52.16
MntFruits	Amount spent by the customer on fruit product category in the last 2 years.	0	199	26.30	39.77
MntSweetProducts	Amount spent by the customer on sweet product category in the last 2 years.	0	263	27.06	41.28
MntWines	Amount spent by the customer on wine product category in the last 2 years.	0	1493	303.93	336.59

MntGoldProds	Amount spent by the customer on gold product category in the last 2 years.	0	362	44.02	52.16
NumDealsPurchases	Number of purchases made by the customer using discount	0	15	2.32	1.93
NumCatalogPurchases	Number of purchases made by the customer using catalog and to be shipped through the mail	0	28	2.66	2.92
NumStorePurchases	Number of purchases made by the customer directly from stores	0	13	5.79	3.25
NumWebPurchases	Number of purchases made by the customer through the superstore’s website	0	27	4.08	2.77
NumWebVisitsMonth	Number of visits the customer made to the superstore’s website in the last month	0	20	5.31	2.42
Recency	Number of days since the customer's last purchase.	0	99	49.10	28.96

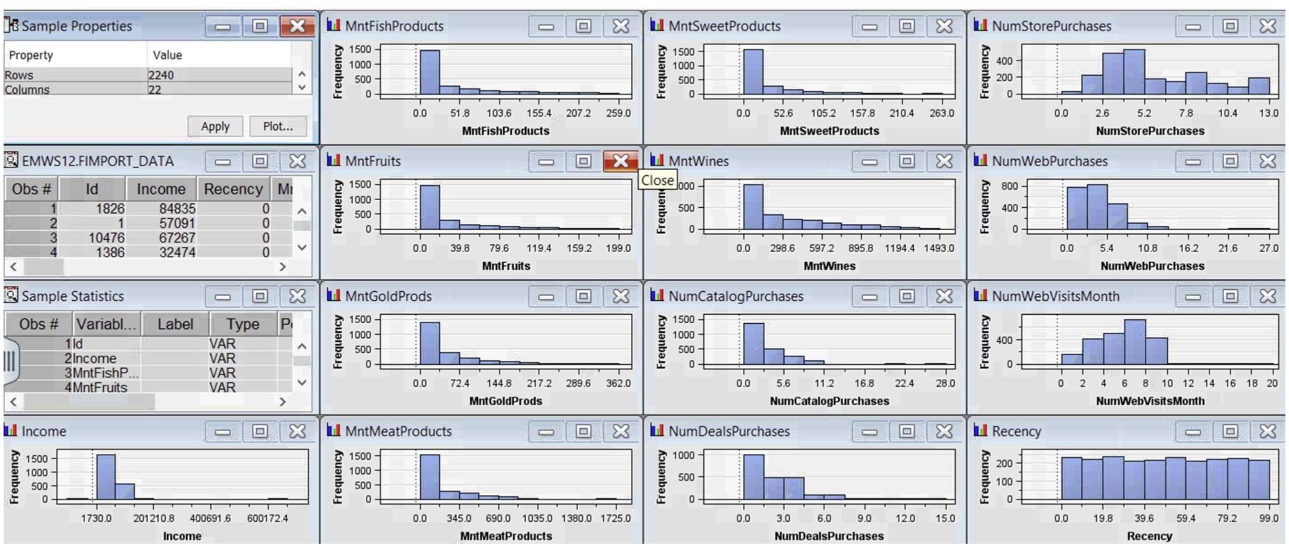


Figure 1 Exploration Results of Interval Variables

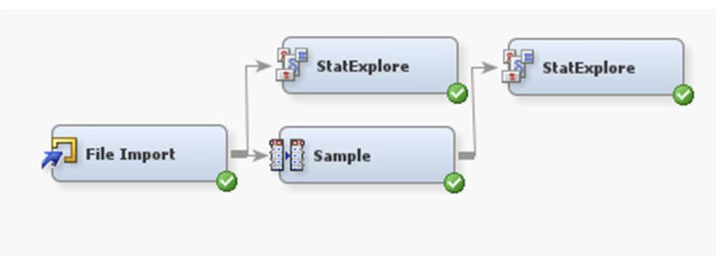
The provided diagram illustrates that several interval variables, including Income, MntFishProducts, MntFruits, MntMeatProducts, MntGoldProducts, MntSweetProcucts, MntWines, MntCatalogPurchases, MntWebPurchases, and NumWebVisits, exhibit skewness. Notably, the "Income" variable stands out with the highest skewness value of 6.76 and a corresponding kurtosis value of 159.63. In contrast, the remaining variables fall within a skewness range of 0 to 2.5 and a kurtosis range of -1.2 to 9.

The elevated skewness and kurtosis observed in the "Income" variable can be attributed to a single observation with a value of 666,666, significantly deviating from the rest of the income values in the sample, which lie between the range of 1000 and 165,000. This extreme outlier contributes to the substantial skewness and kurtosis values in the distribution of the "Income" variable. Further, only the "Income" variable exhibited 24 missing values, while the rest of the variables were complete.

Balancing Distributions

Upon meticulous examination of our target variable, we found out that the response rate in terms of 0 & 1 are imbalanced. We had 1906 observations (85%) with the response value of 0 and 334 observations (15%) with the response value of 1. To balance the class distributions, we used the sample node. The sample node can help to create balanced samples by either oversampling the minority class, under sampling the majority class, or both. This is crucial for building accurate predictive models that are not biased towards the majority class.

We added a sample node from the sample tab and connected it to the file import node. We went into the properties panel and changed the level selection to the rarest level and sample proportion to 50 so that we can get balanced distribution for our target variable. After this step, our target variable has 334 observations with the response 1 and 334 observations with the response 0.



Property	Value
Train	
Variables	...
Output Type	Data
Sample Method	Default
Random Seed	12345
Size	
Type	Percentage
Observations	.
Percentage	100.0
Alpha	0.01
PValue	0.01
Cluster Method	Random
Stratified	
Criterion	Level Based
Ignore Small Strata	No
Minimum Strata Size	5
Level Based Options	
Level Selection	Rarest Level
Level Proportion	100.0
Sample Proportion	50.0

```
Summary Statistics for Class Targets
(maximum 500 observations printed)

Data=DATA

                  Numeric    Formatted    Frequency
Variable           Value       Value        Count       Percent      Label

Response             0           0           1906       85.0893
Response             1           1            334       14.9107


Data=SAMPLE

                  Numeric    Formatted    Frequency
Variable           Value       Value        Count       Percent      Label

Response             0           0            334          50
Response             1           1            334          50
```

Upon applying target variable balancing, a significant reduction in the skewness and kurtosis of the interval variables has been observed. The following table presents a comparison between the skewness and kurtosis values of interval variables in both the original source data and the balanced sample data.

	SOURCE DATA		SAMPLE DATA	
Variable	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**
Income	6.76	159.63	-0.052	-1.088
MntFishProducts	1.92	3.09	1.60	1.82
MntMeatProducts	2.08	5.51	1.26	0.49
MntFruits	2.10	4.05	1.77	2.53
MntSweetProducts	2.14	4.37	1.83	3.03
MntWines	1.18	0.59	0.87	-0.32
MntGoldProds	1.88	3.55	1.55	1.76
NumDealsPurchases	2.42	8.93	1.98	4.92
NumCatalogPurchases	1.88	8.04	0.84	-0.27
NumStorePurchases	0.70	-0.62	0.62	-0.62
NumWebPurchases	1.38	5.70	1.24	5.61
NumWebVisitsMonth	0.20	1.82	-0.29	-1.10
Recency	-0.002	-1.20	0.15	-1.17

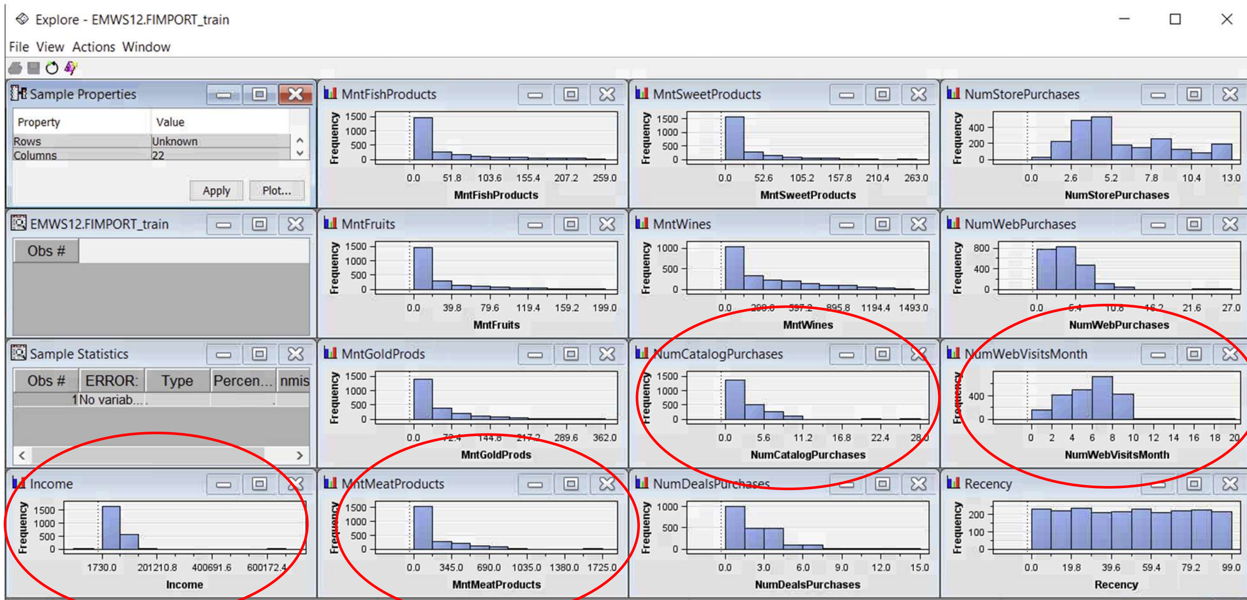


Figure 2 Exploration results of the interval variables of original data source

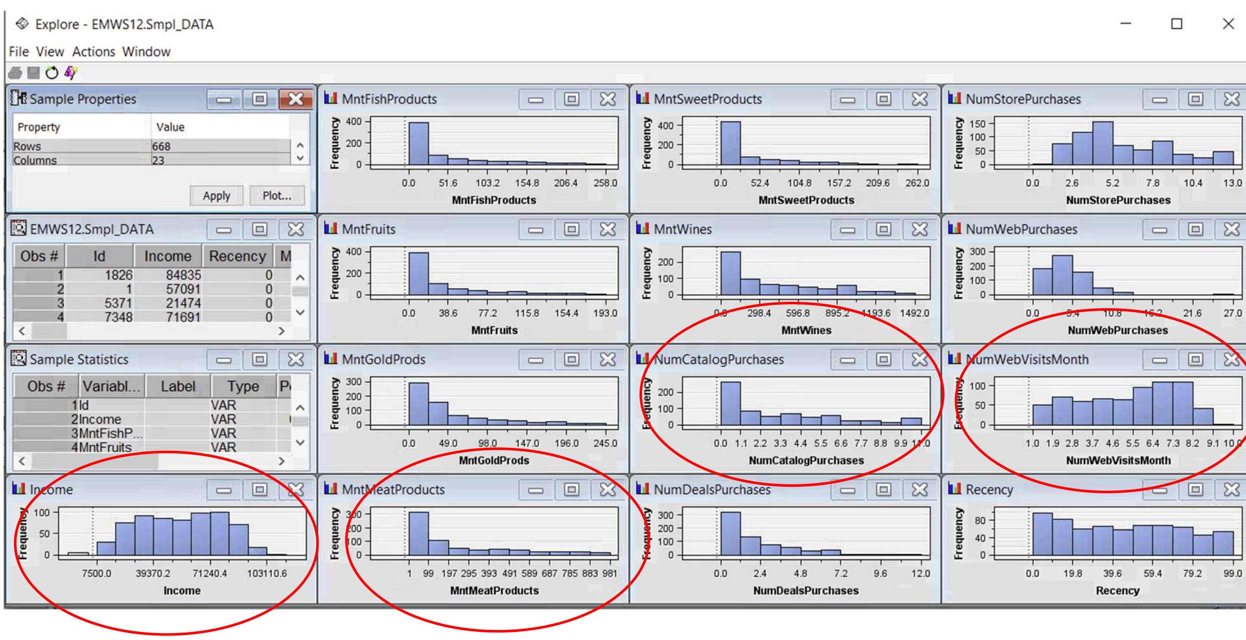
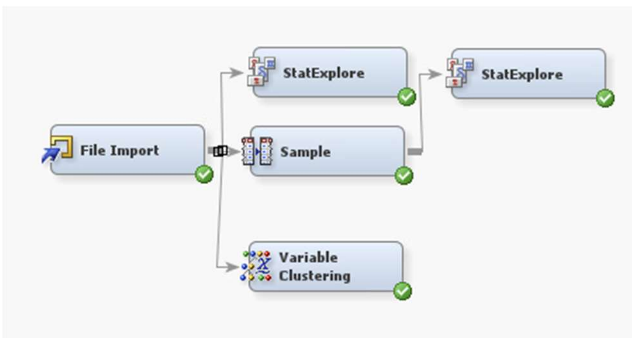


Figure 3 Exploration results of interval variables of the sample data source

The provided visuals indicate a significant contrast in the distribution curves of Income, NumCatalogPurchases, NumWebVisitMonth, and MntMeatProducts between the original source data and the sample data.

Handling Input Redundancy

An input is considered redundant when it doesn't provide additional information beyond what is already conveyed by other inputs. In the context of decision tree models, redundant inputs have a lesser impact due to the nature of the algorithm. However, when using other modeling tools such as Logistic Regression & Neural Network, dealing with input redundancy demands more sophisticated approaches. To address this, an examination of correlations among inputs was carried out by utilizing a clustering variable node. This process aids in identifying and addressing redundant inputs and rejecting them to ensure accuracy and efficiency of the predictive modeling process. Variable clustering node obtained from the Explore tab was connected to the file import node.



The model presented below elucidates the outcomes of the correlation analysis conducted using variable clustering. The results reveal that the correlation values among the interval variables fall within the range of -0.55309 to 0.734127. This pattern signifies that the correlation among these interval variables exhibits a moderate level of positivity or negativity, implying a discernible but not excessively strong relationship between them. Hence, we did not reject any input as redundant for the predictive modelling.

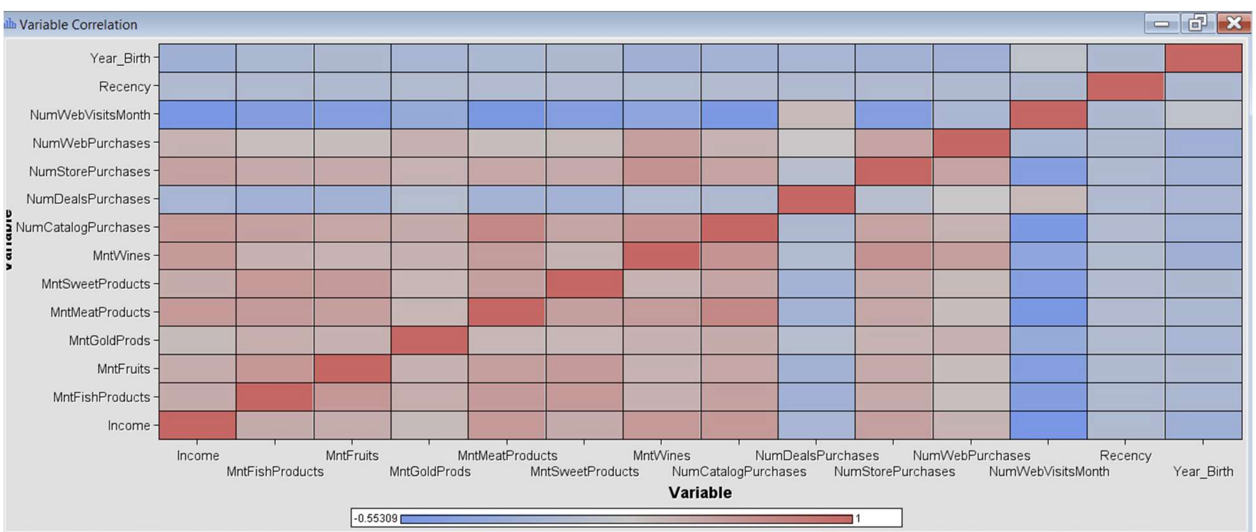


Figure 4 Pearson Correlation heatmap for interval variables from SAS miner

	Year_Birth	Income	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumDealsPurchases	NumWebPurchases	NumCatalogPurchases	NumStorePurcha	Num'
Year_Birth	1.00													
Income	-0.16	1.00												
Recency	-0.02	0.00	1.00											
MntWines	-0.16	0.58	0.02	1.00										
MntFruits	-0.02	0.43	0.00	0.39	1.00									
MntMeatProc	-0.03	0.58	0.02	0.56	0.54	1.00								
MntFishProdu	-0.04	0.44	0.00	0.40	0.59	0.57	1.00							
MntSweetPro	-0.02	0.44	0.02	0.39	0.57	0.52	0.58	1.00						
MntGoldProd	-0.06	0.32	0.02	0.39	0.39	0.35	0.42	0.37	1.00					
NumDealsPur	-0.06	-0.08	0.00	0.01	-0.13	-0.12	-0.14	-0.12	0.05	1.00				
NumWebPurc	-0.15	0.38	-0.01	0.54	0.30	0.29	0.29	0.35	0.42	0.23	1.00			
NumCatalogP	-0.12	0.59	0.03	0.64	0.49	0.72	0.53	0.49	0.44	-0.01	0.38	1.00		
NumStorePur	-0.13	0.53	0.00	0.64	0.46	0.48	0.46	0.45	0.38	0.07	0.50	0.52	1.00	
NumWebVisit	0.12	-0.55	-0.02	-0.32	-0.42	-0.54	-0.45	-0.42	-0.25	0.35	-0.06	-0.52	-0.43	1.00

Figure 5 Pearson Correlation for interval variable form excel.

Handling Input Irrelevancy

An irrelevant input lacks the capacity to offer meaningful information about the target variable. In the context of decision tree models, the algorithm inherently disregards irrelevant inputs, ensuring their limited impact on the modeling process. However, when employing other predictive modeling techniques such as Logistic Regression and Neural Network, it is required to effectively manage irrelevant inputs and their potential influence on the model's performance.

During our examination of the provided data, we recognized the presence of the variable "Dt_Customer," which signifies the date (dd/mm/yy) when a customer enrolled with the superstore. However, we determined that this variable, being a time series data, does not contribute significantly to our understanding of the target variable. Consequently, we made the decision to exclude this variable from our analysis, as it does not yield meaningful insights in relation to our goals.

Handling Curse of Dimensionality

The concept of the "curse of dimensionality" is employed within statistics, machine learning, and data analysis to depict the hurdles encountered while dealing with datasets featuring numerous dimensions or features. This phenomenon underscores the complexities that surface as the dimensionality of data increases. We have detected that one of our nominal variables, specifically "Year Birth," consists of 54 distinct levels. Given that we performed under sampling to address target variable balance, this has led to heightened data sparsity within "Year Birth." Consequently, as a measure to alleviate the curse of dimensionality within the model, we opted to exclude the nominal variable "Year Birth."

Assigning Roles & Levels

The roles and level of the variables were chosen after careful exploratory study. The below table summarizes the roles and level of the variables.

Summary of Role & Level of Variables

No	Variable	Role	Level
01	Response	Target	Binary
02	ID	ID	Nominal
03	Year_Birth	Rejected	Nominal
04	Complain	Input	Binary
05	Dt_Customer	Rejected	Interval
06	Education	Input	Nominal
07	Marital_Status	Input	Nominal
08	Kidhome	Input	Nominal
09	Teenhome	Input	Nominal
10	Income	Input	Interval
11	MntFishProducts	Input	Interval
12	MntMeatProducts	Input	Interval
13	MntFruits	Input	Interval
14	MntSweetProducts	Input	Interval
15	MntWines	Input	Interval
16	MntGoldProds	Input	Interval
17	NumDealsPurchases	Input	Interval
18	NumCatalogPurchases	Input	Interval
19	NumStorePurchases	Input	Interval
20	NumWebPurchases	Input	Interval
21	NumWebVisitsMonth	Input	Interval
22	Recency	Input	Interval

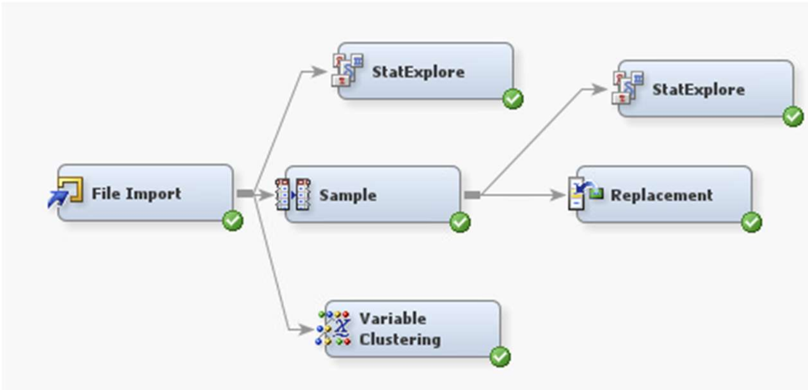
Handling Missing Values

Replacement

We have observed that the variable "Income" has 24 missing values in the source data and 6 missing values in the sample data. While Decision Trees inherently handle missing values without issue, addressing these gaps is essential for other models like Logistic Regression and Neural Networks. These models disregard training cases with missing values as inputs and are unable to predict cases with missing values.

To proactively manage the missing values, we initiated the process by labeling them using a Replacement node, obtained from the Modify tab. This Replacement node was linked to the file import process. The User-Specified Limits method was employed to identify income observations falling below the threshold of 1.

By implementing this strategy, we establish a consistent approach to handle missing values, ensuring that our models can effectively train and predict across various scenarios. While Decision Trees can natively manage missing values, this tailored methodology caters to Logistic Regression and Neural Networks, allowing us to harness their predictive capabilities while accommodating the data's missing values.



.. Property	Value
General	
Node ID	Repl
Imported Data	...
Exported Data	...
Notes	...
Train	
Interval Variables	
Replacement Editor	...
Default Limits Method	User-Specified Limits
Cutoff Values	...
Class Variables	
Replacement Editor	...
Unknown Levels	Ignore
Score	
Replacement Values	Missing
Hide	No
Report	
Replacement Report	Yes
Status	
Create Time	12/07/23 7:16 AM

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Use	Limit Method	Replacement Lower Limit	Replacement Upper Limit	Replace I
Income	**Default**	**Default**	1	.	**Default**
MntFishProduc	**Default**	**Default**	.	.	**Default**
MntFruits	**Default**	**Default**	.	.	**Default**
MntGoldProds	**Default**	**Default**	.	.	**Default**
MntMeatProdu	**Default**	**Default**	.	.	**Default**
MntSweetProd	**Default**	**Default**	.	.	**Default**
MntWines	**Default**	**Default**	.	.	**Default**
NumCatalogPu	**Default**	**Default**	.	.	**Default**
NumDealsPurc	**Default**	**Default**	.	.	**Default**
NumStorePurc	**Default**	**Default**	.	.	**Default**
NumWebPurch	**Default**	**Default**	.	.	**Default**
NumWebVisits	**Default**	**Default**	.	.	**Default**
Recency	**Default**	**Default**	.	.	**Default**

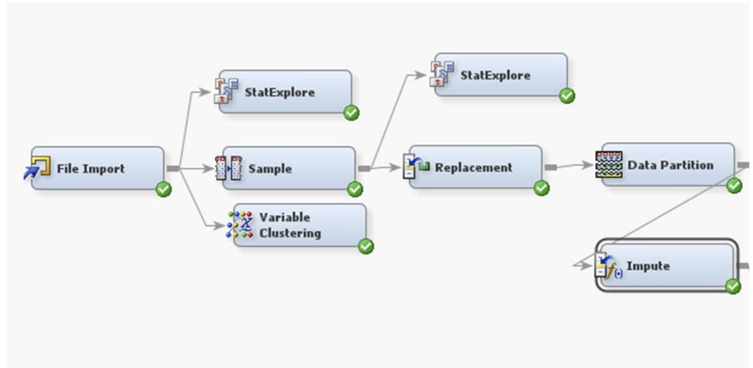
Impute

Following replacement, the subsequent phase in addressing missing values involves imputation. In our case, there were 24 missing values in the "Income" variable from the data source and 6 from the sample data. To rectify this, we chose to impute these values using the mean value. This is an synthetic estimation method.

For implementation, we obtained the "Impute" function from the "Modify" tab and integrated it into the data partition. It is important to note that in our model, imputation is primarily a preparatory step for the Regression and Neural Network models. These models necessitate complete observations, making imputation crucial. However, it is noteworthy that imputation is unnecessary for the Decision Tree model.

By performing imputation, we ensure that our dataset retains full observations, thereby enabling the successful application of Regression and Neural Network models that demand complete data. On the other hand, Decision Tree models inherently accommodate missing values and thus do not require imputation.

Property	Value
Class Variables	
Default Input Method	Count
Default Target Method	None
Normalize Values	Yes
Interval Variables	
Default Input Method	Mean
Default Target Method	None
Default Constant Value	
Default Character Valu	
Default Number Value	.
Method Options	
Random Seed	12345
Tuning Parameters	...
Tree Imputation	...
Score	
Hide Original Variables	Yes
Indicator Variables	
Type	Unique
Source	Imputed Variables
Role	Rejected



To account for potential variations in predicted responses due to missing input values, we have introduced a binary imputation indicator (1 & 0) variable into the training data. This inclusion allows a model to adapt its predictions when encountering instances where the input values are missing. The diagrams provided below visually depict the process wherein the previously replaced missing values in the "Income" variable have been substituted with the mean value.

EMWS12.Impt_VALIDATE

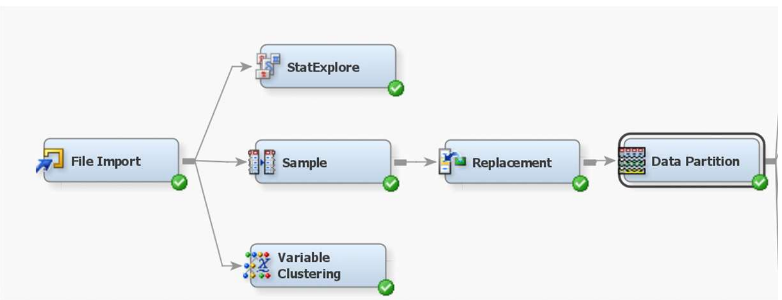
no...	Dt_Cus...	Recency	MntWi...	MntFruits	MntMe...	MntFis...	MntSw...	MntGol...	NumDe...	NumW...	NumCa...	NumSt...	NumW...	Respo...	Complain	Replac...	Impute...	Impu...
1	01/04/2013	39	187	5	65	26	20	14	2	4	2	6	5	0	0	.	54579.54	1
0	10/30/2013	75	532	126	490	164	126	126	1	5	5	11	1	1	0	.	54579.54	1
1	11/08/2013	96	231	65	196	38	71	124	1	6	5	7	4	0	0	.	54579.54	1
0	06/15/2014	0	464	5	64	7	0	37	1	7	3	7	5	1	0	57091	57091	0
0	08/04/2014	0	6	16	24	11	0	34	2	3	1	2	7	1	0	21474	21474	0
0	07/12/2013	0	431	82	441	80	20	102	1	3	6	6	1	1	0	67786	67786	0
1	07/23/2013	0	53	1	5	2	1	10	2	2	0	3	8	0	0	47823	47823	0
1	09/14/2012	0	454	0	171	8	19	32	12	9	2	8	8	0	0	54450	54450	0
0	04/27/2014	1	423	42	706	73	197	197	1	4	8	9	2	0	0	79529	79529	0
0	04/20/2013	1	12	2	17	6	1	10	2	2	0	3	8	1	0	21359	21359	0
0	09/17/2012	1	13	3	8	7	4	16	2	1	0	3	9	1	0	14796	14796	0
1	06/03/2014	2	46	0	12	0	2	23	2	2	1	2	7	1	0	47139	47139	0
1	01/01/2014	2	522	0	257	32	16	66	4	2	2	8	7	1	0	60597	60597	0
0	11/01/2013	2	1074	37	518	193	92	129	1	5	6	7	2	1	0	79174	79174	0
1	07/09/2012	2	50	4	28	6	3	26	3	3	1	2	9	1	0	34213	34213	0
1	05/28/2014	3	292	6	37	0	3	34	4	6	1	6	7	1	0	50388	50388	0
2	04/14/2014	3	520	7	154	19	0	14	2	6	3	11	3	0	0	77622	77622	0
1	04/29/2013	3	421	76	536	82	178	102	2	8	6	5	4	0	0	80011	80011	0
1	02/01/2013	3	502	19	132	0	6	26	6	6	2	11	6	0	0	60200	60200	0
0	02/11/2012	3	145	193	459	205	26	145	2	3	8	7	2	1	0	63211	63211	0
1	10/18/2012	3	322	3	50	4	3	42	5	7	1	6	8	1	0	48432	48432	0
0	03/09/2012	3	890	63	292	0	25	12	4	8	4	7	6	1	0	65220	65220	0
0	05/23/2014	4	448	21	125	52	101	62	0	7	6	8	2	1	0	96547	96547	0
0	03/23/2014	4	35	3	67	10	8	24	1	3	1	3	6	1	0	45203	45203	0
0	06/11/2013	4	179	28	520	111	123	47	1	3	8	13	1	1	0	81698	81698	0
0	09/09/2013	4	30	8	12	8	8	12	1	2	0	4	5	0	0	33039	33039	0
0	03/28/2014	5	33	4	24	4	2	5	2	3	0	4	7	0	0	31928	31928	0
1	07/09/2012	5	520	8	223	32	49	42	4	10	5	7	8	1	0	49912	49912	0
0	03/21/2014	6	1296	17	311	45	69	51	1	2	4	10	1	1	0	85683	85683	0
0	01/16/2013	6	303	23	751	82	26	191	1	6	5	13	4	1	0	70321	70321	0
1	11/25/2012	6	462	61	184	10	53	107	4	7	5	9	6	1	0	55424	55424	0
0	01/05/2014	7	530	117	678	134	44	147	1	4	10	8	1	1	0	83528	83528	0

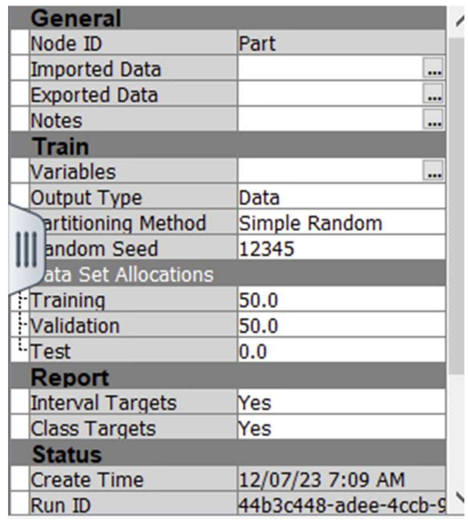
Avoiding overfitting/ underfitting

Data Partition

In SAS Enterprise Miner, the dataset can be divided into three key segments: training, test, and validation. The training data set is used for preliminary model fitting. The validation data set is used to monitor and tune the model weights during estimation and is also used for model assessment. The test data set is an additional hold-out data set that you can use for model assessment. For accurate evaluation of our model performance, we divided our dataset into two equal parts: a 50% training dataset and a 50% validation dataset. Data partition node was added from Sample tab, and we used simple random sampling to create partitioned data sets.

Although having a dedicated test dataset is generally suggested for robust model evaluation in predictive modelling, we chose not to have one due to data availability constraints. We under sampled the responses for 0 due to the target variable imbalance. As a result, we had a very little dataset to divide into training, validation, and test sets, resulting in extremely limited data for each segment.



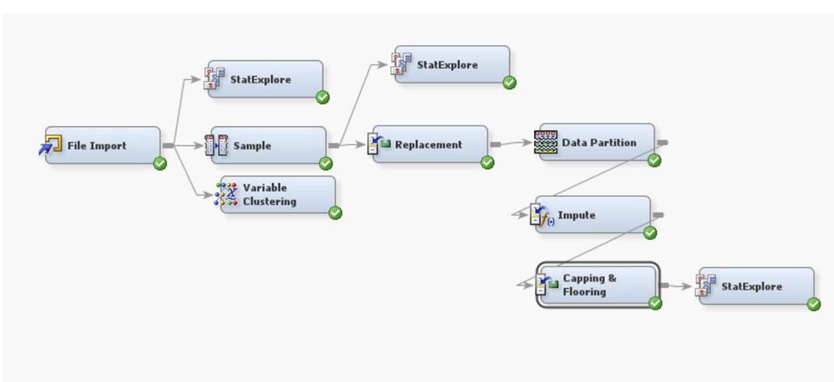


Handling Outliers

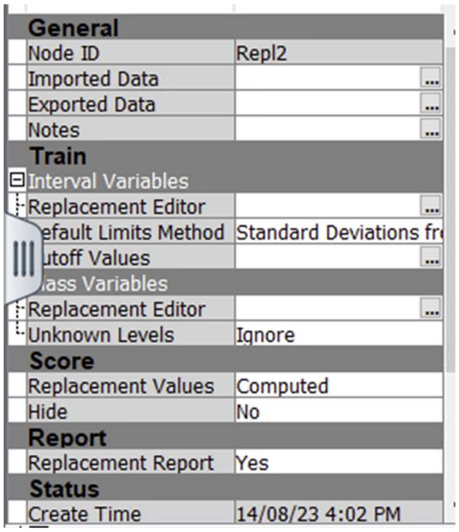
Managing outliers is a crucial step when preparing data for regression and neural network models. Outliers, which are data points that significantly deviate from the rest of the data, can have a substantial impact on model performance. However, outliers will not have any impact on the decision tree. We can handle outliers for regression and neural network models by doing Caping & Flooring or Log Transformation or both.

Caping & Flooring

Capping and flooring is a method used to mitigate the impact of extreme values (outliers) in interval variables to reduce their skewness. This approach involves setting a limit or boundary beyond which data points are constrained. In other words, values that exceed the upper limit are "capped" or set to that limit, and values below the lower limit are "floored" or adjusted to the floor value. This helps normalize the distribution and lessen the influence of extreme values on the overall data. In SAS Miner this can be done by using the replacement node. We obtained the Replacement



node from Modify tab and connected it with the data partition node to execute this process.



We defined the threshold for capping and flooring as Standard Deviation from the Mean. To monitor any alterations in the skewness of the interval variables, we linked the StatExplorer node to the Capping & Flooring node.

Below table summarizes the difference between the initial sample data skewness and kurtosis after capping and flooring.

	SAMPLE DATA before Capping & Flooring		**After Capping & Flooring**	
Variable	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**
Income	-0.052	-1.088	-0.061	-1.143
MntFishProducts	1.60	1.82	1.599	1.658
MntMeatProducts	1.26	0.49	1.258	0.449
MntFruits	1.77	2.53	1.752	2.268
MntSweetProducts	1.83	3.03	1.688	2.054
MntWines	0.87	-0.32	0.825	-0.467
MntGoldProds	1.55	1.76	1.622	1.959
NumDealsPurchases	1.98	4.92	1.407	1.340
NumCatalogPurchases	0.84	-0.27	0.875	-0.275
NumStorePurchases	0.62	-0.62	0.716	-0.553
NumWebPurchases	1.24	5.61	0.568	-0.373
NumWebVisitsMonth	-0.29	-1.10	-0.253	-1.131
Recency	0.15	-1.17	0.098	-1.203

Upon analyzing the obtained results, it is evident that the application of capping and flooring did not yield substantial changes in skewness. However, notable reductions in the kurtosis values of variables such as NumDealsPurchases, NumWebPurchases, and MntSweetProducts have been observed because of this process. . Therefore, the decision to retain the capping and flooring approach was made, even though it didn't have a substantial effect on skewness values.

Transformation

Log transformation serves as an alternative method for managing outliers. When encountering few of the interval variables with pronounced skewness, applying a log transformation can help normalize their distribution. In our case, however, all interval variables exhibit skewness within the range of -0.253 to 1.688. Consequently, the decision was made to retain these values without employing log transformation. This choice was motivated by the intention to maintain the original essence of the inputs.

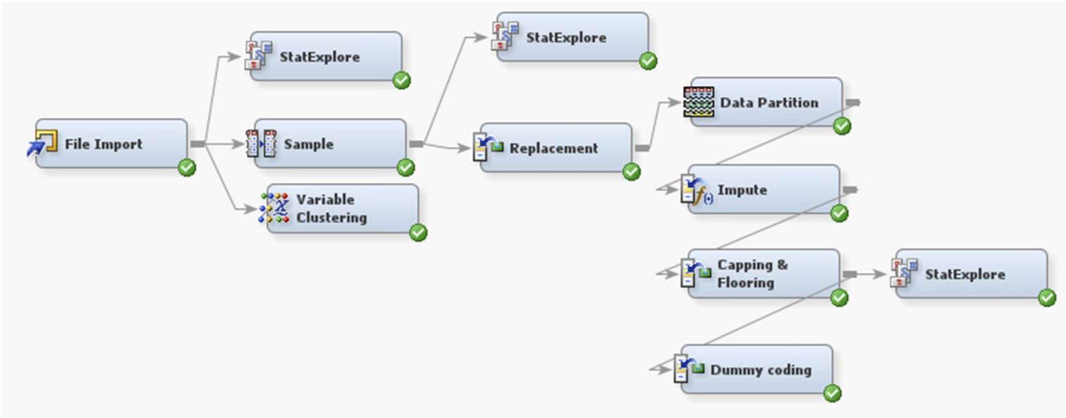
Recoding Categorical Inputs

To recoding categorical inputs (Dummy Code) we used replacement node obtained node from the modify tab to facilitate combining input levels of a categorical input. In the class variables property we selected the replacement editor to recode the categorical inputs such as “Education” & Marital_Status”. Marital Status was the input we had with the largest number of levels (8). We consolidated these levels as below.

Replacement Editor-WORK.OUTCLASS

Variable	Formatted Value	Replacement Value	Frequency Count	Type	Character Unformatted Value	Numeric Value
Complain	0		330	N		[illegible]
Complain	1		4	N		[illegible]
Complain	_UNKNOWN_	_DEFAULT_	.	N		.
Education	Graduation		163	C	Graduation	.
Education	PhD	Post_Graduate	79	C	PhD	.
Education	Master	Post_Graduate	58	C	Master	.
Education	2n Cycle	Schooling	25	C	2n Cycle	.
Education	Basic	Schooling	9	C	Basic	.
Education	_UNKNOWN_	_DEFAULT_	.	C		.
Kidhome	0		203	N		[illegible]
Kidhome	1		127	N		[illegible]
Kidhome	2		4	N		[illegible]
Kidhome	_UNKNOWN_	_DEFAULT_	.	N		.
Marital_Status	Married		115	C	Married	.
Marital_Status	Single	Single	83	C	Single	.
Marital_Status	Together	Couple	74	C	Together	.
Marital_Status	Divorced		48	C	Divorced	.
Marital_Status	Widow		11	C	Widow	.
Marital_Status	Absurd	Single	1	C	Absurd	.
Marital_Status	Alone	Single	1	C	Alone	.
Marital_Status	YOLO	Single	1	C	YOLO	.
Marital_Status	_UNKNOWN_	_DEFAULT_	.	C		.
Response	0		168	N		[illegible]
Response	1		166	N		[illegible]
Response	_UNKNOWN_	_DEFAULT_	.	N		.
Teenhome	0		209	N		[illegible]

We aggregated the categories "Absurd," "Single," "Alone," and "YOLO" under the label "Single." Similarly, we grouped "Married" and "Together" as "Couple." In the "Education" variable, we recategorized "2n Cycle" and "Basic" as "Schooling," while "PhD" and "Masters" were redefined as "Post_Graduates."



Model Implementation & Evaluation

Decision Tree

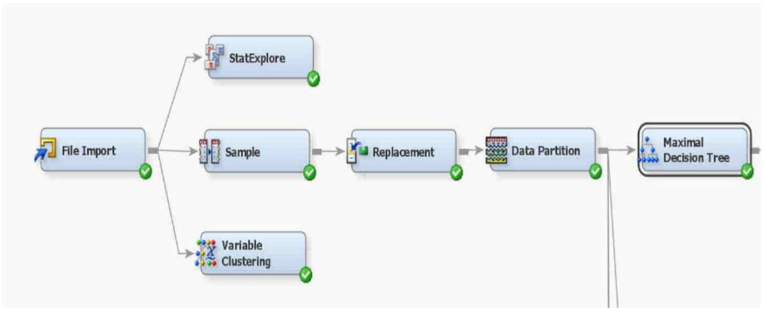
Decision trees serve as a valuable entry point into the realm of predictive modeling. They utilize prediction rules to evaluate cases, aided by a split-search algorithm for input selection. While they effectively manage the curse of dimensionality by disregarding irrelevant inputs, they lack an innate mechanism to eliminate redundant ones. However, in our specific dataset, redundant inputs are not present. Pruning is utilized to manage model complexity.

Split search in decision tee involves selecting an input variable to partition the training data. For interval inputs, unique values serve as potential split points, while categorical inputs use average target values within each category. The data is divided into left and right branches based on the selected input and division point. These branches, along with target outcomes, form a 2x2 contingency table. The Pearson chi-squared statistic quantifies the independence of column counts in the table, indicating if outcome proportions differ between branches. A higher chi-squared statistic suggests a favorable split.

The chi-squared statistic is converted into a p-value, reflecting the likelihood of observing the statistic assuming equal target proportions. The resulting p-value is used to calculate logworth, a measure of split quality represented as the negative logarithm of the p-value [*logworth* = -*log*(chi-squared *p*-value)]. For a split to occur, at least one logworth must exceed a preset threshold. This threshold, by default, corresponds to a chi-squared p-value of 0.20 or an approximate logworth of 0.7.

Maximal Tree

We added the decision tree node from the model tab and connected it to the data partition node. To create a maximal tree, we went into the properties panel and under the subtree heading, we selected the method as largest & the assessment measure as decision to make it a maximal tree. Maximal tree will help us to create the most complex model in the sequence.



Number of Rules	5
Number of Surrogate R	0
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Largest
Number of Leaves	1
Assessment Measure	Decision
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validati	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
Observation Based Imp	
Observation Based Imp	No

The below leaf statistics graphical representation contrasts the anticipated blue outcome percentages in the bars on the left (derived from the training data) with the actual red outcome percentages in the bars on the right (extracted from the validation data). The purpose of this visualization is to assess the alignment between response rates observed in the training data and those seen in the validation data. Ideally, the heights of the bars should be equal. Disparities in bar heights in this case stem from limited case counts in the corresponding leaf nodes. In this case, we get 12 leaves in the maximal tree.

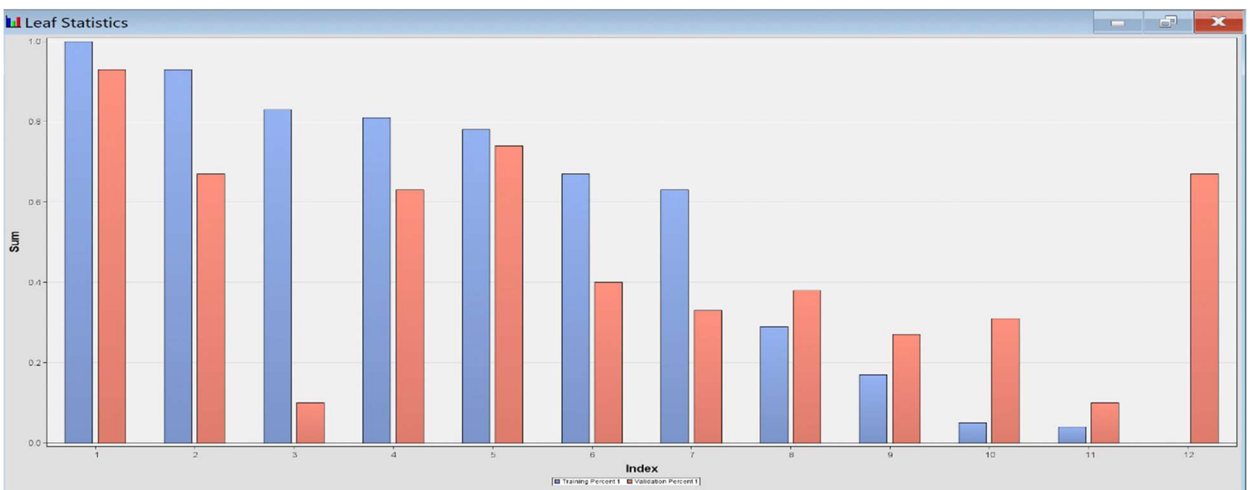


Figure 6 Leaf Statistics of Maximal Tree

The variable importance display offers a perspective on the significance of input variables within the decision tree. The value of the importance statistic indicates the extent to which a specific input accounts for variations in the target outcome, relative to the input featured at the table's uppermost position. In this case, NumWebVisits month explains 86.56 % of the variability explained by NumCatalogPurchases.

Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
NumCatalogPurchases		1	1.0000	1.0000	1.0000
NumWebVisitsMonth		3	0.8656	0.4700	0.5430
Recency		2	0.7478	0.8950	1.1967
REP_Income	Replacement: Income	1	0.6530	0.0553	0.0846
Education		2	0.6373	0.0000	0.0000
Teenhome		1	0.5745	0.5371	0.9348
Marital_Status		1	0.3015	0.0000	0.0000
MntSweetProducts		0	0.0000	0.0000	.
NumWebPurchases		0	0.0000	0.0000	.
NumStorePurchases		0	0.0000	0.0000	.
NumDealsPurchases		0	0.0000	0.0000	.
MntFishProducts		0	0.0000	0.0000	.
Kidhome		0	0.0000	0.0000	.
MntGoldProds		0	0.0000	0.0000	.
MntMeatProducts		0	0.0000	0.0000	.
MntFruits		0	0.0000	0.0000	.
MntWines		0	0.0000	0.0000	.
Complain		0	0.0000	0.0000	.
Year_Birth		0	0.0000	0.0000	.

Figure 7 Variable Importance of Maximal Tree

In the fit statistics table, average squared error for validation comes out to 0.219598. Misclassification rate for validation is 0.305389.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		NOBS	Sum of Frequencies	334	334	.
Response		MISC	Misclassification Rate	0.164671	0.305389	.
Response		MAX	Maximum Absolute Error	0.963636	1	.
Response		SSE	Sum of Squared Errors	85.51944	146.6911	.
Response		ASE	Average Squared Error	0.128023	0.219598	.
Response		RASE	Root Average Squared Error	0.357803	0.468612	.
Response		DIV	Divisor for ASE	668	668	.
Response		DFT	Total Degrees of Freedom	334	.	.

Figure 8 Fit Statistics of Maximal Tree

The maximal tree has 12 leaves, and the root node is determined by the variable NumCatalogPurchases. Among the instances that satisfy the condition (greater than 0.5 or missing), there is a higher count and percentage that belong to the desired target response. Following the root node, the 1st competing nodes identified are NumWebVisitsMonths & Teenhome.

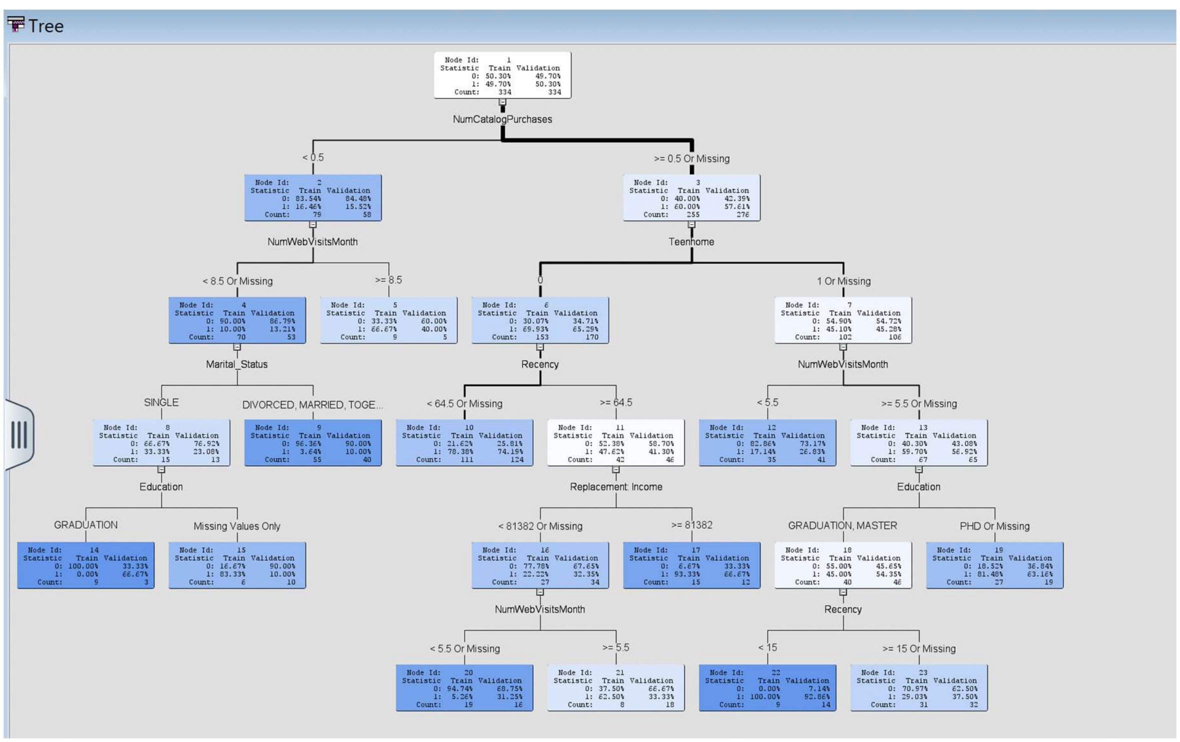
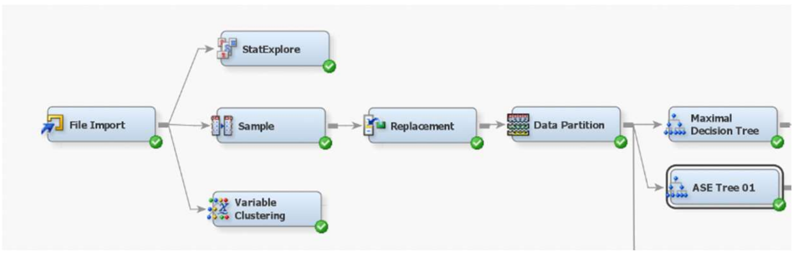
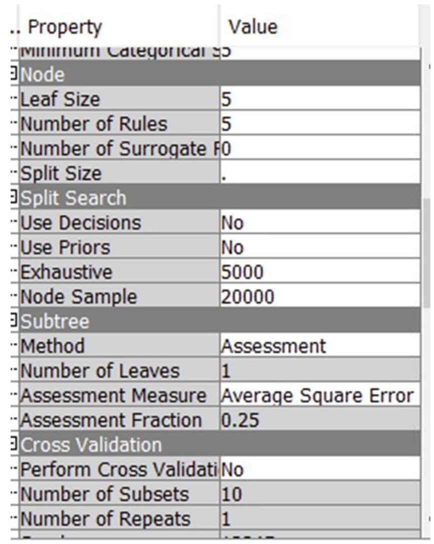


Figure 9 Maximal Tree

Average Squared Error Tree (ASE Tree 01)

The main purpose to create an average squared error tree is to prune the decision tree to explore how different modeling objectives can change the optimal model specification. We added the decision tree node from the model tab and connected it to the data partition node. To create an Average Squared Error Tree, we went into the properties panel and under the subtree heading, we selected the method as assessment & the assessment measure as average squared error.





The ASE tree has 8 leaves, and there is better alignment between training and the validation data as compared to the maximal tree.

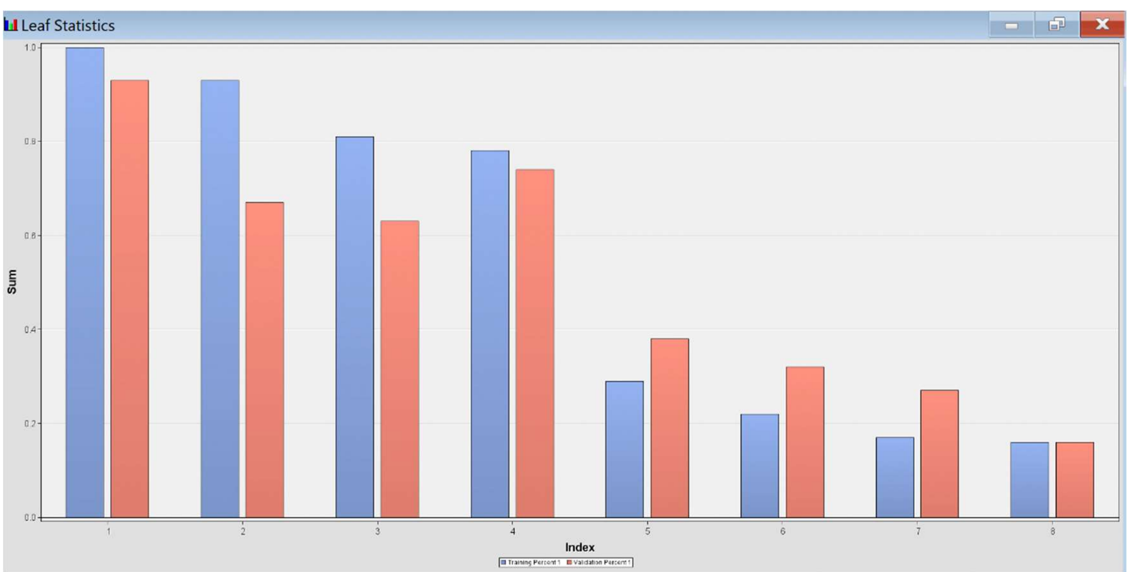


Figure 10 Leaf Statistics of ASE Tree 01

In the ASE tree, the variable with the highest importance is same as Maximal tree but the variable with the second highest importance is Recency.

Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
NumCatalogPurchases		1	1.0000	1.0000	1.0000
Recency		2	0.7478	0.8950	1.1967
REP_Income	Replacement: Income	1	0.6530	0.0553	0.0846
NumWebVisitsMonth		1	0.6034	0.4700	0.7789
Teenhome		1	0.5745	0.5371	0.9348
Education		1	0.4331	0.0000	0.0000
NumWebPurchases		0	0.0000	0.0000	.
MntSweetProducts		0	0.0000	0.0000	.
MntWines		0	0.0000	0.0000	.
NumStorePurchases		0	0.0000	0.0000	.
NumDealsPurchases		0	0.0000	0.0000	.
MntFishProducts		0	0.0000	0.0000	.
Kidhome		0	0.0000	0.0000	.
MntGoldProds		0	0.0000	0.0000	.
MntMeatProducts		0	0.0000	0.0000	.
MntFruits		0	0.0000	0.0000	.
Marital_Status		0	0.0000	0.0000	.
Complain		0	0.0000	0.0000	.
Year_Birth		0	0.0000	0.0000	.

Figure 11 Variable Importance of ASE tree 01

The average squared error is 0.194909 which is less than the maximal tree. Also, the misclassification rate is 0.260479 which is less and means that the incorrectly classified instances in relation to the total number of instances in the dataset are less.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		NOBS	Sum of Frequencies	334	334	.
Response		MISC	Misclassification Rate	0.191617	0.260479	.
Response		MAX	Maximum Absolute Error	0.933333	1	.
Response		SSE	Sum of Squared Errors	101.4083	130.1994	.
Response		ASE	Average Squared Error	0.151809	0.194909	.
Response		RASE	Root Average Squared Error	0.389627	0.441485	.
Response		DIV	Divisor for ASE	668	668	.
Response		DFT	Total Degrees of Freedom	334	.	.

Figure 12 Fit Statistics of ASE tree 01

The ASE tree has 8 leaves, and the first split is for NumCatalogPurchases followed by teenhome. Like the maximal tree, in the ASE tree the condition (greater than 0.5 or missing), there is a higher count and percentage that belong to the desired target response. Our target customer according to this tree should have purchased more than or equal to 0.5 products from the catalog, should have 0 teens at home and the days since last purchase should be less than 64.5 days. After the root node, the competing split is teenhome & then recency and NumVisitsMonth.

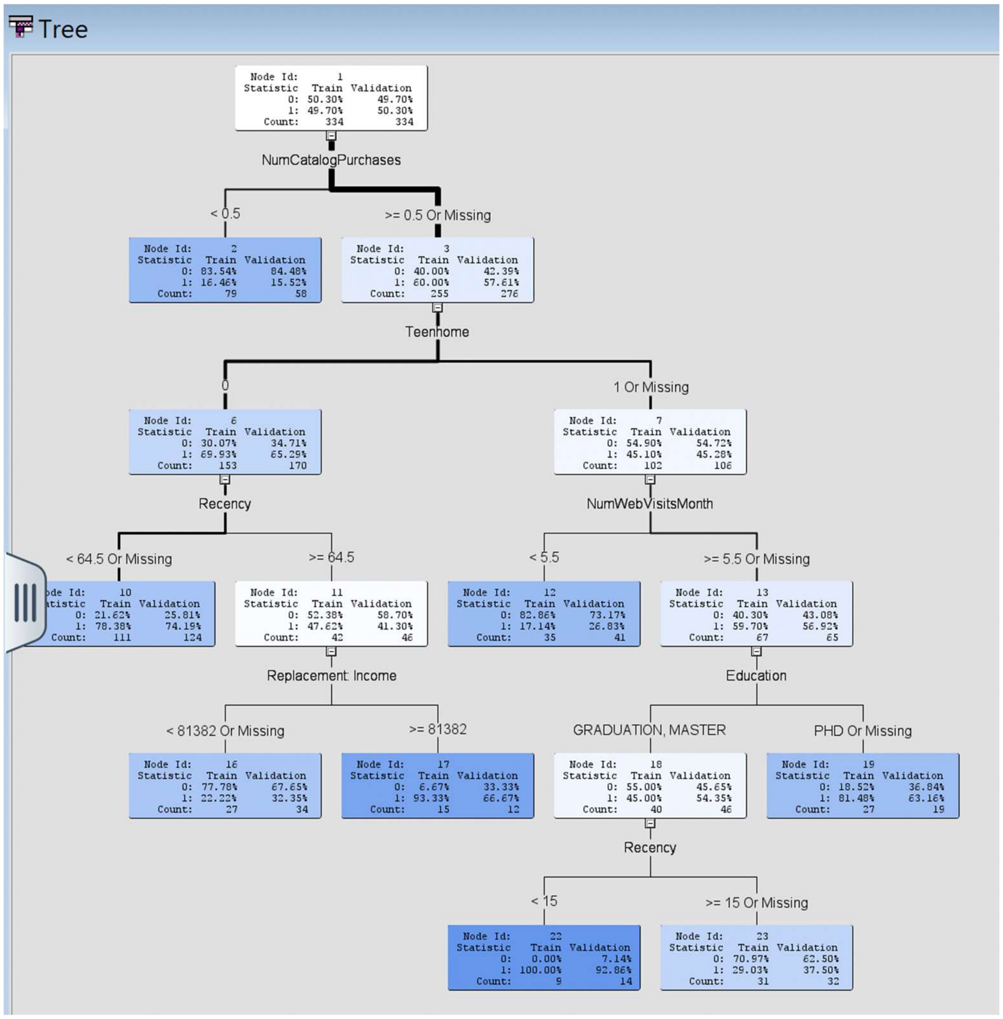
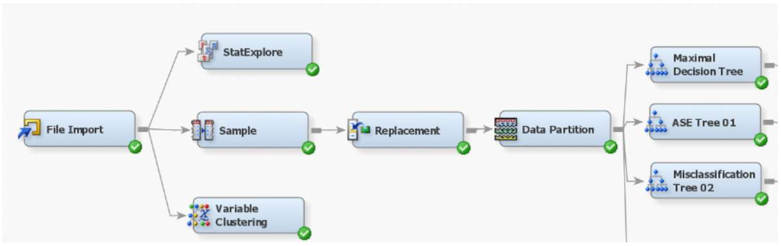
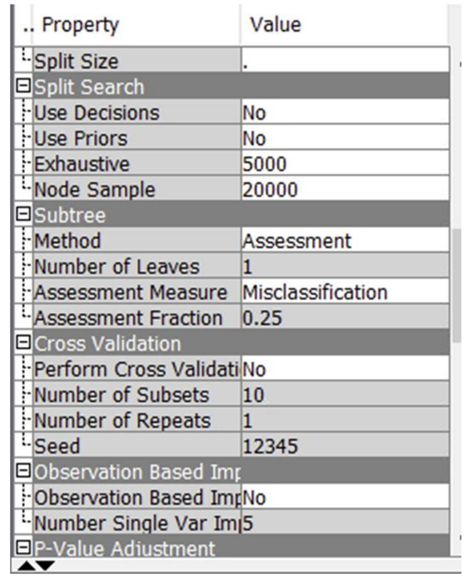


Figure 13 ASE Tree 01

Misclassification Tree (Misclassification Tree 02)

The misclassification tree can provide a clear and interpretable framework for analyzing the factors contributing to misclassifications and guiding strategies for model improvement. We added the decision tree node from the model tab and connected it to the data partition node. To create an Average Squared Error Tree, we went into the properties panel and under the subtree heading, we selected the method as assessment & the assessment measure as misclassification.





.. Property	Value
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Misclassification
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validati	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
Observation Based Imp	
Observation Based Imp	No
Number Single Var Im	5
P-Value Adjustment	

The misclassification tree also has 8 leaves which is like the ASE tree. The alignment between the train and validation data is also like ASE tree.

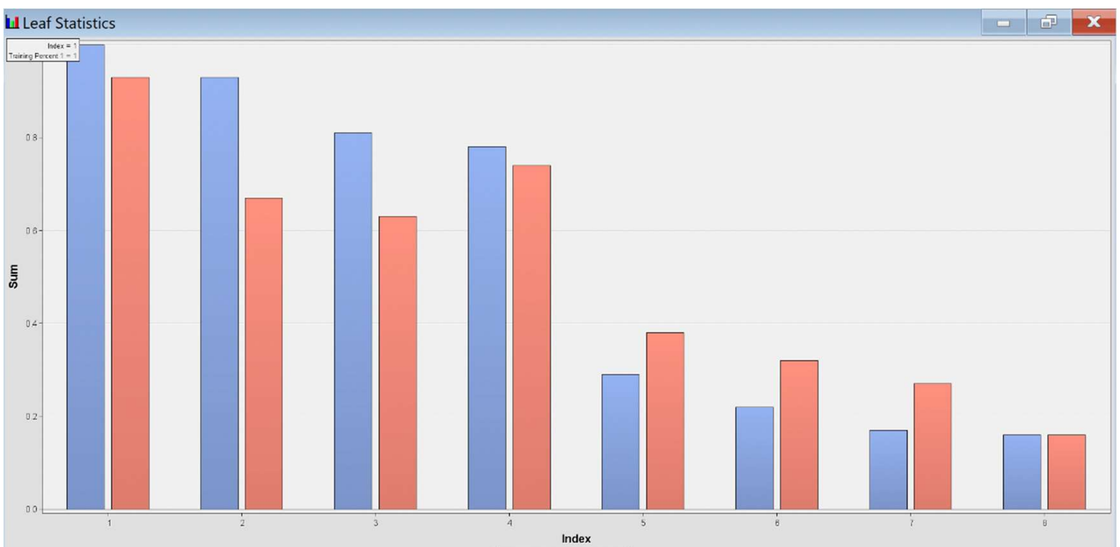


Figure 14 Leaf Statistics of Misclassification tree 02

The variable importance is like the ASE tree. NumCatalogPurchases has the highest importance followed by Recency with 74.78% variability explained by NumCatalogPurchases.

Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
NumCatalogPurchases		1	1.0000	1.0000	1.0000
Recency		2	0.7478	0.8950	1.1967
REP_Income	Replacement: Income	1	0.6530	0.0553	0.0846
NumWebVisitsMonth		1	0.6034	0.4700	0.7789
Teenhome		1	0.5745	0.5371	0.9348
Education		1	0.4331	0.0000	0.0000
NumWebPurchases		0	0.0000	0.0000	.
MntSweetProducts		0	0.0000	0.0000	.
MntWines		0	0.0000	0.0000	.
NumStorePurchases		0	0.0000	0.0000	.
NumDealsPurchases		0	0.0000	0.0000	.
MntFishProducts		0	0.0000	0.0000	.
Kidhome		0	0.0000	0.0000	.
MntGoldProds		0	0.0000	0.0000	.
MntMeatProducts		0	0.0000	0.0000	.
MntFruits		0	0.0000	0.0000	.
Marital_Status		0	0.0000	0.0000	.
Complain		0	0.0000	0.0000	.
Year_Birth		0	0.0000	0.0000	.

Figure 15 Variable Importance of Misclassification tree 02

The average squared error is 0.194909 and the misclassification rate is 0.260479 which is similar to the ASE tree.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		_NOBS_	Sum of Frequencies	334	334	.
Response		_MISC_	Misclassification Rate	0.191617	0.260479	.
Response		_MAX_	Maximum Absolute Error	0.933333	1	.
Response		_SSE_	Sum of Squared Errors	101.4083	130.1994	.
Response		_ASE_	Average Squared Error	0.151809	0.194909	.
Response		_RASE_	Root Average Squared Error	0.389627	0.441485	.
Response		_DIV_	Divisor for ASE	668	668	.
Response		_DFT_	Total Degrees of Freedom	334	.	.

Figure 16 Fit Statistics of Misclassification tree 02

The tree structure of the misclassification tree is also similar to the ASE tree. After the root node, the competing split is teenhome & then recency and NumVisitsMonth.

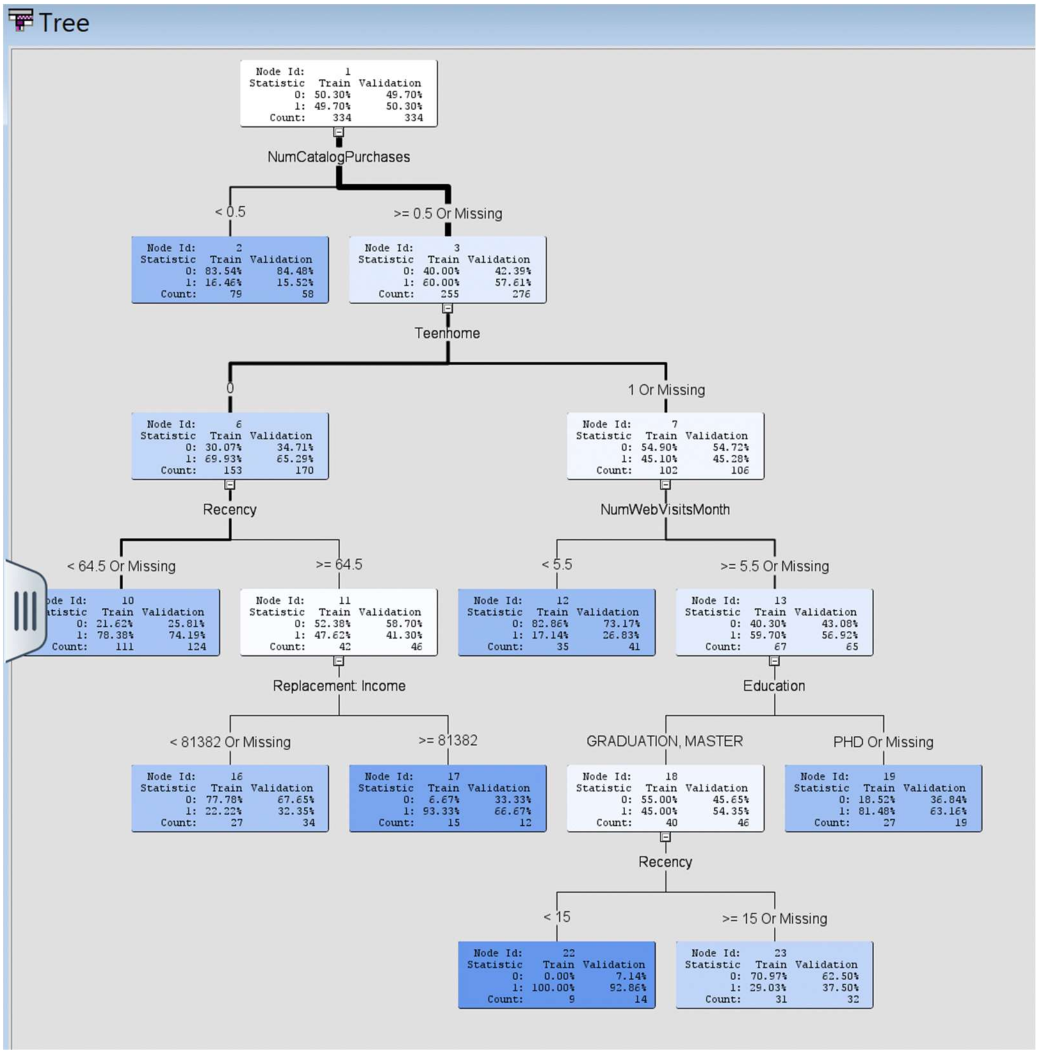


Figure 17 Misclassification tree

In summary, given that our target variable is a binary numeric variable, an Average Square Error (ASE) tree appears to be a more appropriate choice between the two options. However, it's important to note that we will conduct a comprehensive model comparison using the model comparison node to determine the most suitable model.

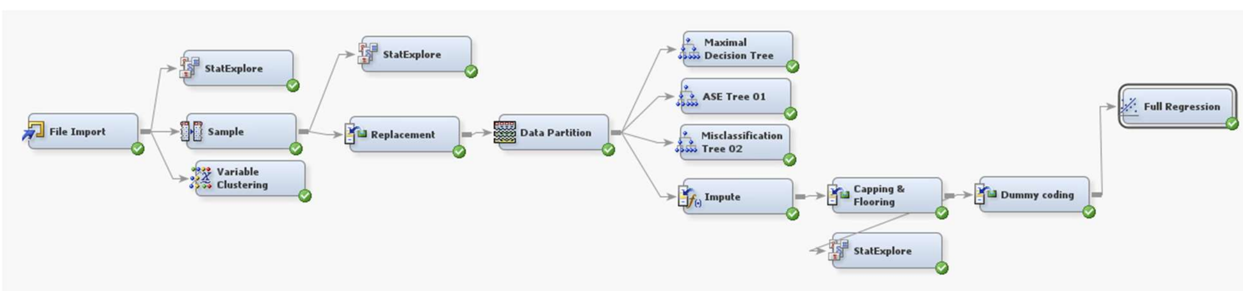
Logistic Regression

Full Regression

In SAS Enterprise Miner, "Full Regression" refers to the process of building a regression model using all available predictor variables in the dataset. In other words, it involves including every potential independent variable in the model simultaneously, without any variable selection or exclusion.

When performing a full regression in SAS Enterprise Miner, we are essentially using all the predictor variables to predict the target or response variable. This can provide a comprehensive understanding of how each variable contributes to the model's predictive power. However, including all available variables without any selection can lead to overfitting, especially if some of the variables are irrelevant, redundant, or noisy.

Given the potential concerns with overfitting, SAS Enterprise Miner provides a mechanism to determine the optimal inputs for the regression model using a method known as sequential selection. The Regression node within SAS Enterprise Miner offers three distinct sequential selection techniques: Forward, Backward, and Stepwise. These methods aid in systematically identifying the most influential predictor variables, striking a balance between model complexity and predictive performance.



Property	Value
Variables	...
Equation	
Main Effects	Yes
Two-Factor Interaction	No
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	...
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	None
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	...
Optimization Options	

To execute a full regression analysis, the Regression node was obtained from the model tab and subsequently linked to the dummy coding node. The chosen approach for model selection was set as "none," while the default selection criteria were adopted to guide the model-building process. This approach involves incorporating all available variables without engaging in any specific variable selection procedure.

Variable Summary

Role	Measurement Level	Frequency Count
INPUT	BINARY	1
INPUT	INTERVAL	13
INPUT	NOMINAL	4
REJECTED	BINARY	1
REJECTED	INTERVAL	14
REJECTED	NOMINAL	3
TARGET	BINARY	1

The screenshot of the initial lines of the output window summarizes the roles of variables used (or not) by the Regression node. According to this the fit model has 18 inputs that predict a binary target.

Figure 18 Variable Summary of Full Regression

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Complain	1	0.0116	0.9142
Kidhome	2	0.0824	0.9596
REP_Education	2	7.4683	0.0239
REP_IMP_REP_Income	1	8.2638	0.0040
REP_Marital_Status	4	12.3680	0.0148
REP_MntFishProducts	1	1.0125	0.3143
REP_MntFruits	1	1.1513	0.2833
REP_MntGoldProds	1	0.0592	0.8078
REP_MntMeatProducts	1	6.2998	0.0121
REP_MntSweetProducts	1	0.9830	0.3215
REP_MntWines	1	0.0368	0.8478
REP_NumCatalogPurchases	1	11.2323	0.0008
REP_NumDealsPurchases	1	5.0361	0.0248
REP_NumStorePurchases	1	13.3314	0.0003
REP_NumWebPurchases	1	0.0000	0.9966
REP_NumWebVisitsMonth	1	20.1111	<.0001
REP_Recency	1	25.0944	<.0001
Teenhome	2	10.8813	0.0043

The Type 3 Analysis tests the statistical significance of adding the indicated input to a model that already contains other listed inputs. A value near 0 in the Pr > ChiSq column approximately indicates a significant input; a value near 1 indicates an extraneous input. In our model the statistical significance measures a range from <0.0001 (highly significant) to 0.9963 (highly dubious). Results such as this suggest that certain inputs can be eliminated without affecting the predictive capabilities of the model.

Figure 19 Type 3 Analysis of Effects of Full Regression

The validation average squared error of the full regression is 0.172122.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		AIC	Akaike's Information Criterion	323.6843	.	.
Response		ASE	Average Squared Error	0.13341	0.172122	.
Response		AVERR	Average Error Function	0.409707	0.598558	.
Response		DFE	Degrees of Freedom for Error	309	.	.
Response		DFM	Model Degrees of Freedom	25	.	.
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DIV	Divisor for ASE	668	668	.
Response		ERR	Error Function	273.6843	399.8368	.
Response		FPE	Final Prediction Error	0.154998	.	.
Response		MAX	Maximum Absolute Error	0.968857	0.999998	.
Response		MSE	Mean Square Error	0.144204	0.172122	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		NW	Number of Estimate Weights	25	.	.
Response		RASE	Root Average Sum of Squares	0.365254	0.414876	.
Response		RFPE	Root Final Prediction Error	0.393697	.	.
Response		RMSE	Root Mean Squared Error	0.379742	0.414876	.
Response		SBC	Schwarz's Bayesian Criterion	418.9629	.	.
Response		SSE	Sum of Squared Errors	89.11803	114.9777	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		MISC	Misclassification Rate	0.203593	0.233533	.

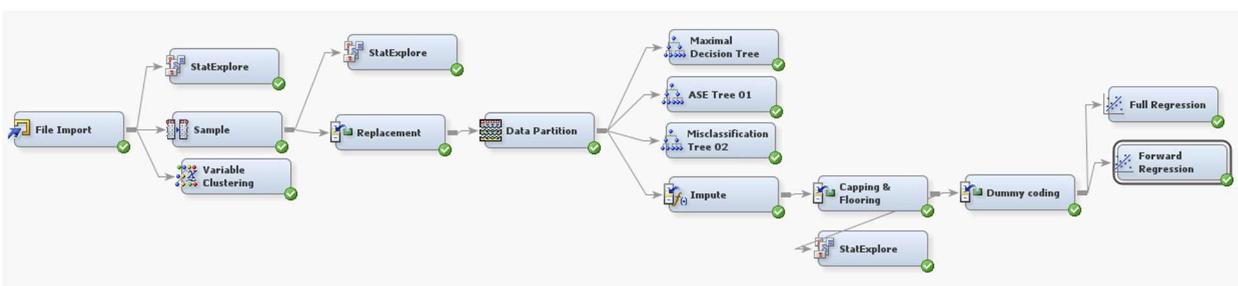
Figure 20 Fit Statistics of Full Regression

Forward Regression

Forward selection iteratively builds models, starting with a baseline prediction. It selects inputs that significantly improve the model based on p-values. The process stops when no more improvements are significant. This systematic method enhances model complexity while prioritizing significant inputs.

To execute the forward sequential selection process, the Regression node from the model tab is linked to the dummy coding. In the Regression node's property panel, the selection method is set to "forward," and the selection criteria is defined as validation error. This approach systematically builds the model by iteratively adding inputs based on their significance in reducing validation error.

.. Property	Value
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	...
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
uppress Intercept	No
put Coding	Deviation
odel Selection	
Selection Model	Forward
Selection Criterion	Validation Error
Use Selection Defaults	Yes
Selection Options	...
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0



The fit statistics in the results reveal a validation average squared error of 0.169496 for the forward regression. Notably, this value is lower than the average squared error observed in the full regression.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		AIC	Akaike's Information Criterion	320.536	.	.
Response		ASE	Average Squared Error	0.141874	0.169496	.
Response		AVERR	Average Error Function	0.434934	0.532181	.
Response		DFE	Degrees of Freedom for Error	319	.	.
Response		DFM	Model Degrees of Freedom	15	.	.
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DIV	Divisor for ASE	668	668	.
Response		ERR	Error Function	290.536	355.4967	.
Response		FPE	Final Prediction Error	0.155217	.	.
Response		MAX	Maximum Absolute Error	0.973842	0.997756	.
Response		MSE	Mean Square Error	0.148545	0.169496	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		NW	Number of Estimate Weights	15	.	.
Response		RASE	Root Average Sum of Squares	0.376662	0.411699	.
Response		RFPE	Root Final Prediction Error	0.393975	.	.
Response		RMSE	Root Mean Squared Error	0.385416	0.411699	.
Response		SBC	Schwarz's Bayesian Criterion	377.7031	.	.
Response		SSE	Sum of Squared Errors	94.77192	113.2232	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		MISC	Misclassification Rate	0.206587	0.248503	.

Figure 21 Fit Statistics results of Forward Regression

The below summary of forward selection shows the variables considered for the 9-step model.

Summary of Forward Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Validation Error Rate
1	REP_NumCatalogPurchases	1	1	42.0221	<.0001	439.9
2	REP_Recency	1	2	26.7324	<.0001	389.1
3	REP_NumWebVisitsMonth	1	3	20.7607	<.0001	379.2
4	REP_MntMeatProducts	1	4	26.3790	<.0001	374.2
5	REP_Marital_Status	4	5	16.4548	0.0025	370.8
6	REP_NumStorePurchases	1	6	7.1748	0.0074	374.8
7	REP_IMP_REP_Income	1	7	11.8949	0.0006	381.1
8	REP_Education	2	8	6.9020	0.0317	370.3
9	Teenhome	2	9	7.7758	0.0205	355.5

Figure 22 Summary of Forward Regression Selection

The variables included in the forward selection regression model was Intercept, REP_Education, REP-IMP_REP_Income, REP_Marital-Status, REP_MntMeatProducts, REP_NumCatalogPurchases, REP_NumStorePurchases, REP_NumWebVIsitsMonth, REP_Recency, Teenhome.

The selected model, based on the error rate for the validation data, is the model trained in Step 9. It consists of the following effects:

Intercept REP_Education REP_IMP_REP_Income REP_Marital_Status REP_MntMeatProducts REP_NumCatalogPurchases REP_NumStorePurchases REP_NumWebVisitsMonth REP_Recency Teenhome

Odds Ratio Estimates

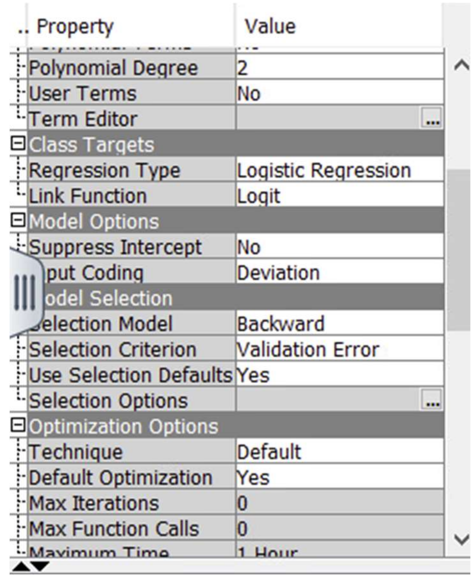
Effect		Point Estimate
REP_Education	Graduation vs Schooling	1.965
REP_Education	Post_Graduate vs Schooling	3.905
REP_IMP_REP_Income		1.000
REP_Marital_Status	Couple vs Widow	0.824
REP_Marital_Status	Divorced vs Widow	2.799
REP_Marital_Status	Married vs Widow	0.936
REP_Marital_Status	Single vs Widow	2.621
REP_MntMeatProducts		1.003
REP_NumCatalogPurchases		1.375
REP_NumStorePurchases		0.767
REP_NumWebVisitsMonth		1.918
REP_Recency		0.974
Teenhome	0 vs 2	4.377
Teenhome	1 vs 2	1.616

Figure 23 Odd Ratio Estimates of Forward Regression

The Odd Ratio Estimate results from the forward regression output led to the following conclusions:

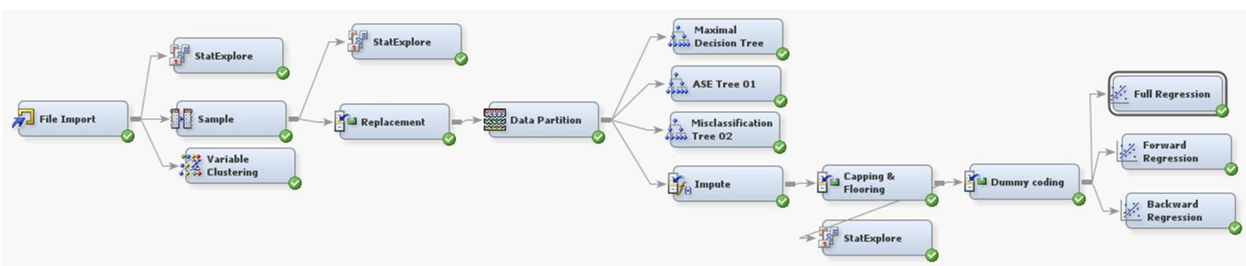
- A one-unit increase in education from basic to graduate is associated with an approximately 1.96 times higher likelihood of purchasing a gold membership.
- A one-unit increase in education from basic to postgraduate is linked to around 3.9 times higher odds of purchasing a gold membership.
- Each unit increase in income corresponds to a 100% increase in the likelihood of purchasing a gold membership.
- Transitioning from being a widow to being in a couple is associated with an 18% decrease in the odds of purchasing gold membership.
- Going from widow to divorced increases the odds of gold membership purchase by approximately 2.8 times.
- Moving from being widow to being a married decreases the odds of gold membership purchases by 7%.
- Transitioning from widow to single increases the odds of gold membership purchases by around 2.6 times.
- A unit increase in meat purchases is linked to a 100% increase in the odds of membership purchases.
- Every unit increase in catalog purchases is associated with 1.37 times increase in the odds of membership purchases.
- An increase in store purchases by a unit corresponds to a 24% decrease in the odds of membership purchases.
- A unit increase in website visits per month raises the odds of gold membership purchases by approximately 1.9 times.
- Each unit increase in recency is linked to a 3% decrease in the odds of gold membership purchases.
- Changing from a household with two teens to having no teen increases the odds of gold membership purchase by around 4.37 times.
- Transitioning from a household with two teens to having one teen increases the odds of gold membership purchase by approximately 1.6 times.

Backward Regression

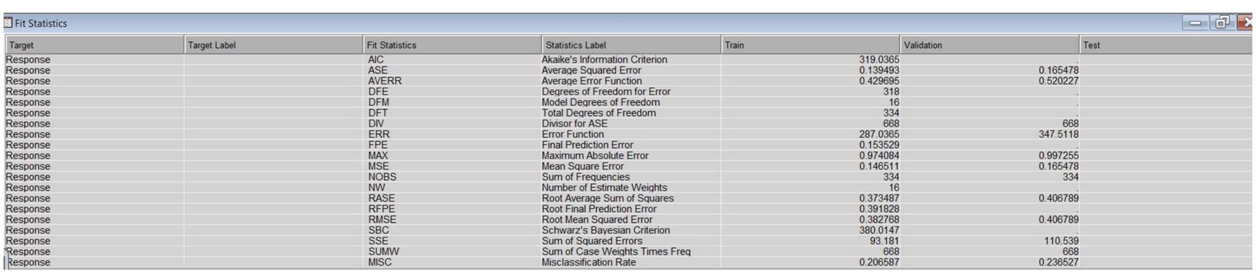


In contrast to forward selection, backward selection generates a sequence of models with diminishing complexity. This sequence originates from a saturated model, encompassing all available inputs for the highest possible fit statistic. Subsequently, inputs are systematically excluded. In each step, the input chosen for removal causes the least reduction in the overall model fit statistic, equivalent to removing the input with the highest p-value. The process concludes when all remaining inputs exhibit a p-value below the predefined stay cutoff which is 0.05.

To execute the backward sequential selection process, the Regression node from the model tab is linked to the dummy coding. In the Regression node's property panel, the selection method is set to "backward," and the selection criteria is defined as validation error.



The fit statistics in the results reveal a validation average squared error of 0.165478 for the backward regression. Notably, this value is lower compared to the average squared error observed in the full regression & forward regression.



Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		AIC	Akaike's Information Criterion	319.0365	.	
Response		ASE	Average Squared Error	0.139493	0.165478	
Response		AVERR	Average Error Function	0.429695	0.520227	
Response		DFE	Degrees of Freedom for Error	318	.	
Response		DFM	Model Degrees of Freedom	16	.	
Response		DFT	Total Degrees of Freedom	334	.	
Response		DIV	Divisor for ASE	668	668	
Response		ERR	Error Function	287.0365	347.5118	
Response		FPE	Final Prediction Error	0.153529	.	
Response		MAX	Maximum Absolute Error	0.974084	0.997255	
Response		MSE	Mean Square Error	0.146511	0.165478	
Response		NOBS	Sum of Frequencies	334	334	
Response		NW	Number of Estimate Weights	16	.	
Response		RASE	Root Average Sum of Squares	0.373487	0.406789	
Response		RFPE	Root Final Prediction Error	0.391828	.	
Response		RMSE	Root Mean Squared Error	0.382768	0.406789	
Response		SBC	Schwarz's Bayesian Criterion	380.0147	.	
Response		SSE	Sum of Squared Errors	93.181	110.539	
Response		SUMW	Sum of Case Weights Times Freq	668	668	
Response		MISC	Misclassification Rate	0.206587	0.236527	

Figure 24 Fit Statistics of Backward Regression

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Validation Error Rate
1	REP_NumWebPurchases	1	17	0.0000	0.9966	399.8
2	Kidhome	2	16	0.0860	0.9579	352.0
3	REP_MntGoldProds	1	15	0.0001	0.9926	352.0
4	Complain	1	14	0.0111	0.9162	352.2
5	REP_MntWines	1	13	0.4454	0.5045	357.6
6	REP_MntFruits	1	12	0.5256	0.4685	356.3
7	REP_MntFishProducts	1	11	1.1613	0.2812	352.0
8	REP_MntSweetProducts	1	10	0.7056	0.4009	347.5
9	REP_NumDealsPurchases	1	9	3.4228	0.0643	355.5

Figure 25 Summary of Backward Eliminations

The above summary of backward elimination shows the variables that have been eliminated from the saturated model.

Variables such as REP_Education, REP_IMP_REP_Income, REP_Marital_Status, REP_MntMeatProducts, REP_MntCatalogPurchases, REP_NumDealsPurchases, REP_NumStorePurchases, REP_NumWebVisitsMonth, REP_Recency, Teenhouse

The selected model, based on the error rate for the validation data, is the model trained in Step 8. It consists of the following effects:

Intercept REP_Education REP_IMP_REP_Income REP_Marital_Status REP_MntMeatProducts REP_NumCatalogPurchases REP_NumDealsPurchases REP_NumStorePurchases REP_NumWebVisitsMonth REP_Recency Teenhome

Odds Ratio Estimates

Effect		Point Estimate
REP_Education	Graduation vs Schooling	1.765
REP_Education	Post_Graduate vs Schooling	3.526
REP_IMP_REP_Income		1.000
REP_Marital_Status	Couple vs Widow	0.655
REP_Marital_Status	Divorced vs Widow	2.329
REP_Marital_Status	Married vs Widow	0.758
REP_Marital_Status	Single vs Widow	2.203
REP_MntMeatProducts		1.003
REP_NumCatalogPurchases		1.383
REP_NumDealsPurchases		1.247
REP_NumStorePurchases		0.746
REP_NumWebVisitsMonth		1.805
REP_Recency		0.974
Teenhome	0 vs 2	8.944
Teenhome	1 vs 2	2.305

Figure 26 Odd Ratio Estimates of Backward Regression

The Odd Ratio Estimate results from the backward regression output led to the following conclusions:

- A one-unit increase in education from basic to graduate is associated with an approximately 1.76 times higher likelihood of purchasing a gold membership.
- A one-unit increase in education from basic to postgraduate is linked to around 3.5 times higher odds of purchasing a gold membership.
- Each unit increase in income corresponds to a 100% increase in the likelihood of purchasing a gold membership.
- Transitioning from being a widow to being in a couple is associated with a 35% decrease in the odds of purchasing gold membership.
- Going from widow to divorced increases the odds of gold membership purchase by approximately 2.3 times.
- Moving from being widow to being a married decreases the odds of gold membership purchases by 25%.
- Transitioning from widow to single increases the odds of gold membership purchases by around 2.2 times.
- A unit increase in meat purchases is linked to a 100% increase in the odds of membership purchases.
- Every unit increase in catalog purchases is associated with 1.37 times increase in the odds of membership purchases.
- Every unit increase in deals purchases is associated with 1.24 times increase in the odds of membership purchases.
- An increase in store purchases by a unit corresponds to a 26% decrease in the odds of membership purchases.
- A unit increase in website visits per month raises the odds of gold membership purchases by approximately 1.8 times.
- Each unit increase in recency is linked to a 3% decrease in the odds of gold membership purchases.
- Changing from a household with two teens to having no teen increases the odds of gold membership purchase by around 8.94 times.
- Transitioning from a household with two teens to having one teen increases the odds of gold membership purchase by approximately 2.3 times.

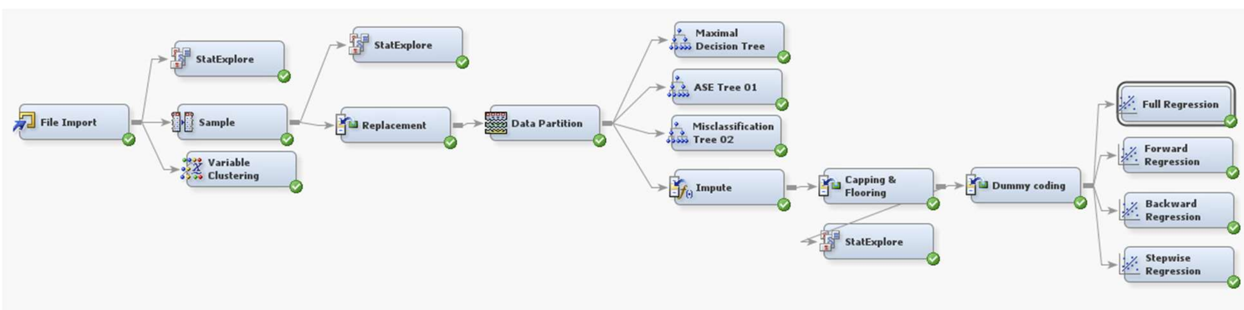
Stepwise Regression

Stepwise selection combines aspects of both forward and backward selection methods. It starts like forward selection, progressively adding inputs with p-values below the entry cutoff. However, after adding an input, the algorithm reassesses the statistical significance of all included inputs. If any input's p-value surpasses the stay cutoff, it's removed from the model and becomes available for future steps. The process concludes when all available inputs for inclusion have p-values above the entry cutoff, and inputs already in the model have p-values below the stay cutoff which is 0.05.

To execute the stepwise sequential selection process, the Regression node from the model tab is linked to the dummy coding. In the Regression node's property panel, the selection method is set to "Stepwise," and the selection criteria is defined as validation error.

Property	Value
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Stepwise
Selection Criterion	Validation Error
Use Selection Defaults	Yes
Selection Options	...
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour
Convergence Criteria	
Uses Defaults	Yes
Options	...
Output Options	
Confidence Limits	No

General



The validation average squared error for the stepwise regression is 0.169496. This value is lower than that of full regression, higher than backward regression, and identical to forward regression.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		AIC	Akaike's Information Criterion	320.536	.	.
Response		ASE	Average Squared Error	0.141874	0.169496	.
Response		AVERR	Average Error Function	0.434934	0.532181	.
Response		DFE	Degrees of Freedom for Error	319	.	.
Response		DFM	Model Degrees of Freedom	15	.	.
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DIV	Divisor for ASE	668	668	.
Response		ERR	Error Function	290.536	355.4967	.
Response		FPE	Final Prediction Error	0.155217	.	.
Response		MAX	Maximum Absolute Error	0.973842	0.997756	.
Response		MSE	Mean Square Error	0.148545	0.169496	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		NW	Number of Estimate Weights	15	.	.
Response		RASE	Root Average Sum of Squares	0.376662	0.411699	.
Response		RFPE	Root Final Prediction Error	0.393975	.	.
Response		RMSE	Root Mean Squared Error	0.385416	0.411699	.
Response		SBC	Schwarz's Bayesian Criterion	377.7031	.	.
Response		SSE	Sum of Squared Errors	94.77192	113.2232	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		MISC	Misclassification Rate	0.206587	0.248503	.

Figure 25 Average Squared error of Stepwise Regression

Summary of Stepwise Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Validation Error Rate
1	REP_NumCatalogPurchases	1	1	42.0221		<.0001	439.9
2	REP_Recency	1	2	26.7324		<.0001	389.1
3	REP_NumWebVisitsMonth	1	3	20.7607		<.0001	379.2
4	REP_MntMeatProducts	1	4	26.3790		<.0001	374.2
5	REP_Marital_Status	4	5	16.4548		0.0025	370.8
6	REP_NumStorePurchases	1	6	7.1748		0.0074	374.8
7	REP_IMP_REP_Income	1	7	11.8949		0.0006	381.1
8	REP_Education	2	8	6.9020		0.0317	370.3
9	Teenhome	2	9	7.7758		0.0205	355.5

Figure 26 Summary of Stepwise Regression

The table above provides a summary of the selected inputs for the model. These include REP_Education, REP_IMP_REP_Income, REP_Marital_Status, REP_NumCatalogPurchases, REP_NumStorePurchase, REP_NumWebVisitsMont, REP_Recency, and Teenhome.

The selected model, based on the error rate for the validation data, is the model trained in Step 9. It consists of the following effects:

Intercept REP_Education REP_IMP_REP_Income REP_Marital_Status REP_MntMeatProducts REP_NumCatalogPurchases REP_NumStorePurchases REP_NumWebVisitsMonth REP_Recency Teenhome

Odds Ratio Estimates

Effect		Point Estimate
REP_Education	Graduation vs Schooling	1.965
REP_Education	Post_Graduate vs Schooling	3.905
REP_IMP_REP_Income		1.000
REP_Marital_Status	Couple vs Widow	0.824
REP_Marital_Status	Divorced vs Widow	2.799
REP_Marital_Status	Married vs Widow	0.936
REP_Marital_Status	Single vs Widow	2.621
REP_MntMeatProducts		1.003
REP_NumCatalogPurchases		1.375
REP_NumStorePurchases		0.767
REP_NumWebVisitsMonth		1.918
REP_Recency		0.974
Teenhome	0 vs 2	4.377
Teenhome	1 vs 2	1.616

Figure 27 Odd Ratio Estimates of Stepwise regression

The Odd Ratio Estimate results from the stepwise regression output led to the following conclusions:

- A one-unit increase in education from basic to graduate is associated with an approximately 1.96 times higher likelihood of purchasing a gold membership.
- A one-unit increase in education from basic to postgraduate is linked to around 3.9 times higher odds of purchasing a gold membership.
- Each unit increase in income corresponds to a 100% increase in the likelihood of purchasing a gold membership.
- Transitioning from being a widow to being in a couple is associated with a 18% decrease in the odds of purchasing gold membership.
- Going from widow to divorced increases the odds of gold membership purchase by approximately 2.8 times.
- Moving from being widow to being a married decreases the odds of gold membership purchases by 7%.
- Transitioning from widow to single increases the odds of gold membership purchases by around 2.6 times.
- A unit increase in meat purchases is linked to a 100% increase in the odds of membership purchases.
- Every unit increase in catalog purchases is associated with 1.37 times increase in the odds of membership purchases.
- An increase in store purchases by a unit corresponds to a 24% decrease in the odds of membership purchases.
- A unit increase in website visits per month raises the odds of gold membership purchases by approximately 1.9 times.
- Each unit increase in recency is linked to a 3% decrease in the odds of gold membership purchases.
- Changing from a household with two teens to having no teen increases the odds of gold membership purchase by around 4.37 times.
- Transitioning from a household with two teens to having one teen increases the odds of gold membership purchase by approximately 1.61 times.

In summary, among all the regression models, backward regression yielded the lowest validation average squared error, while full regression exhibited the highest. The validation average squared error for both full and stepwise regression models was identical. Consequently, we have exclusively connected neural networks to the backward regression model for subsequent iterations.

Neural Network

Regression modelling on a set of derived inputs, referred to as hidden units, is how a neural network might be conceptualised. The names of a neural network's components differ from those of similar components in a regression model. A bias term is used in neural networks in place of an intercept term. A neural network has weight estimations rather than parameter estimates. The ability of neural networks to simulate almost any continuous link between the inputs and the target is what makes them intriguing. Simply choose the appropriate number of hidden units and establish fair weight values. There will be some trial and error in determining the ideal number of hidden units.

Implementation of Neural Network

A neural network node was linked sequentially with an impute node, a dummy coding node, a cap and flooring node and a stepwise regression node in the SAS Enterprise Miner workflow. Among various regression models, the neural network node was exclusively connected to the backward regression node due to its consistently demonstrating the lowest ASE (Average Squared Error) across the range of regression models considered. Subsequently, a series of neural network iterations were conducted, commencing with a configuration of 3 hidden units and 50 iterations. The iteration count and the number of hidden units were then systematically adjusted with incremental steps, aiming to identify the configuration that yielded the lowest ASE. This iterative process was carried out in conjunction with the aforementioned nodes, persistently seeking optimal neural network configurations, and halting adjustments when an increase in ASE was observed.

Neural Network ran on Impute Node

Neural Network with 3 Hidden Units and 50 Iterations

An iteration of a neural network, characterized by 3 hidden units and 50 iterations, was executed while opting for the average square error as the model assessment criterion. Moreover, the optimization option was chosen from the properties panel, and the preliminary training phase was deactivated.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	240	.	.
Response		DFM	Model Degrees of Freedom	94	.	.
Response		NW	Number of Estimated Weights	94	.	.
Response		AIC	Akaike's Information Criterion	485.6252	.	.
Response		SBC	Schwarz's Bayesian Criterion	843.8725	.	.
Response		ASE	Average Squared Error	0.140938	0.167573	.
Response		MAX	Maximum Absolute Error	0.907071	0.928977	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.375418	0.409357	.
Response		SSE	Sum of Squared Errors	94.14681	111.9387	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.25134	.	.
Response		MSE	Mean Squared Error	0.196139	0.167573	.
Response		RFPE	Root Final Prediction Error	0.501338	.	.
Response		RMSE	Root Mean Squared Error	0.442876	0.409357	.
Response		AVERR	Average Error Function	0.445547	0.518006	.
Response		ERR	Error Function	297.6252	346.0277	.
Response		MISC	Misclassification Rate	0.197605	0.212575	.
Response		WRONG	Number of Wrong Classifications	66	71	.

Figure 28 Fit Statistics - Neural Network ran on Impute Node 3 Hidden Units 50 Iterations (No Preliminary Training)

Based on the model, The average squared error, which was chosen as the selection criteria, was observed to be 0.167573, and the misclassification rate 0.212575.

Weights Final Plot

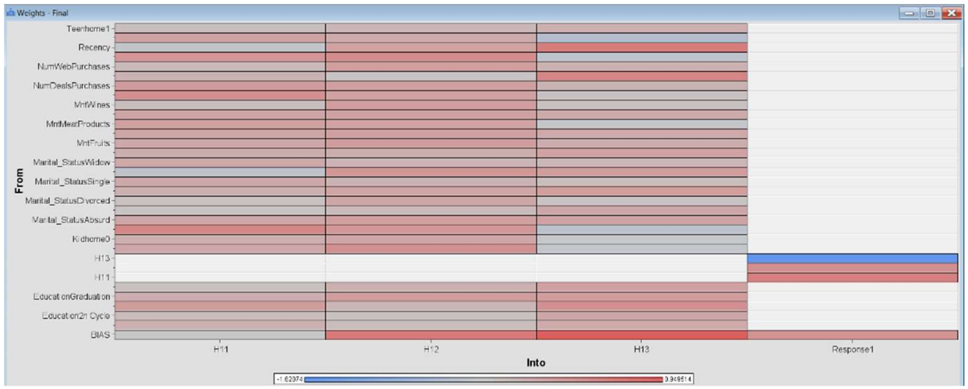


Figure 29 Weights Final Plot- Neural Network ran on Impute Node 3 Hidden Units 50 Iterations (No Preliminary Training)

The Weight Final Plot shows that the weights corresponding to each variable play a notably significant role in predicting customer responses. We can see the variables that went into the model were only Teenhome1, Recency, NumWebPurchases, NumDealPurchases, MntWines, MntMeatProducts, MntFruits, Marital_StatusWidow, Marital_StatusSingle, Marital_StatusDivorced, Marital_StatusAbsurd, Kidhome0, EducationGraduation and Education2nCycle.

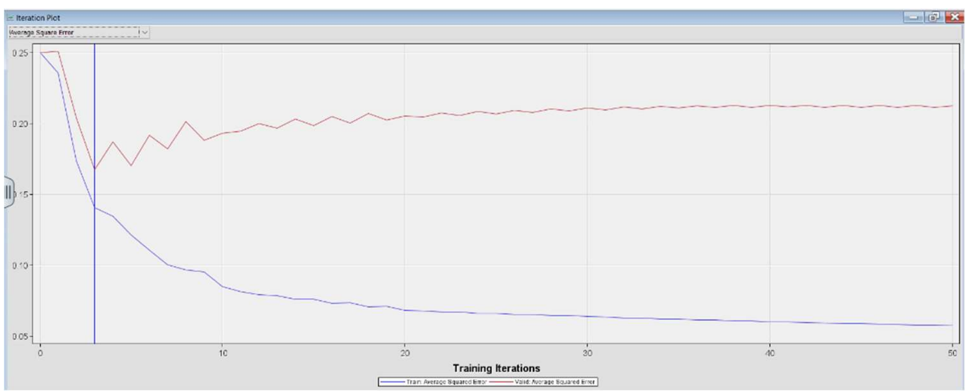


Figure 30 Iterations Plot- Neural Network ran on Impute Node 3 Hidden Units 50 Iterations (No Preliminary Training)

As illustrated in the subsequent diagram, it is clear that the model achieves convergence by approximately the third iteration. This suggests that a substantial amount of extra iterations might have been avoidable, and it is plausible that the optimal model could have been recognized as early as the third iteration.

Neural Network with 3 Hidden Units and 100 Iterations

A neural network with 3 Hidden Units and 100 Iterations was now run.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	240	.	.
Response		DFM	Model Degrees of Freedom	94	.	.
Response		NW	Number of Estimated Weights	94	.	.
Response		AIC	Akaike's Information Criterion	485.6252	.	.
Response		SBC	Schwarz's Bayesian Criterion	843.8725	.	.
Response		ASE	Average Squared Error	0.140938	0.167573	.
Response		MAX	Maximum Absolute Error	0.907071	0.928977	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.375418	0.409357	.
Response		SSE	Sum of Squared Errors	94.14681	111.9387	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.25134	.	.
Response		MSE	Mean Squared Error	0.196139	0.167573	.
Response		RFPE	Root Final Prediction Error	0.501338	.	.
Response		RMSE	Root Mean Squared Error	0.442876	0.409357	.
Response		AVERR	Average Error Function	0.445547	0.518006	.
Response		ERR	Error Function	297.6252	346.0277	.
Response		MISC	Misclassification Rate	0.197605	0.212575	.
Response		WRONG	Number of Wrong Classifications	66	71	.

Figure 31 Fit Statistics - Neural Network ran on Impute 3 Hidden Units 100 Iterations (No Preliminary Training)

For the model, we identified that the average squared error is 0.167573 and the misclassification rate is 0.212575.

Figure 32 Weights Final Plot- Neural Network ran on Impute 3 Hidden Units 100 Iterations (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, Recency, NumWebPurchases, NumDealPurchases, MntWines, MntMeatProducts, MntFruits, Marital_StatusWidow, Marital_StatusSingle, Marital_StatusDivorced, Marital_StatusAbsurd, Kidhome0, EducationGraduation and Education2nCycle.

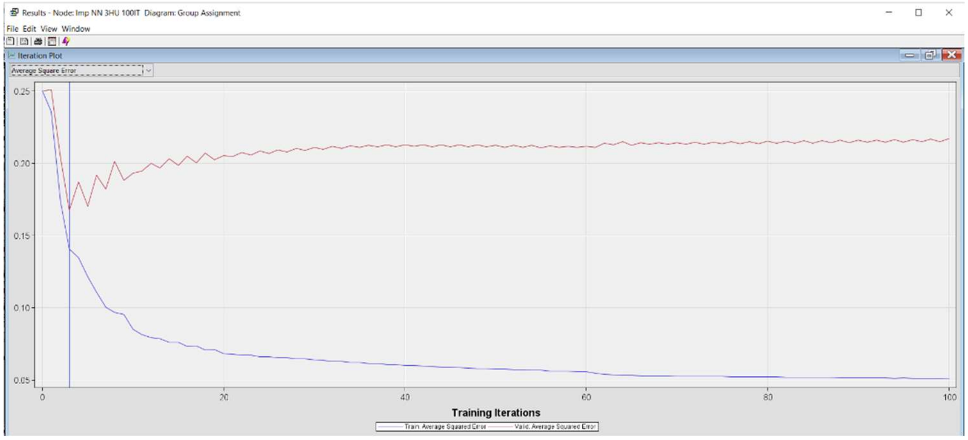
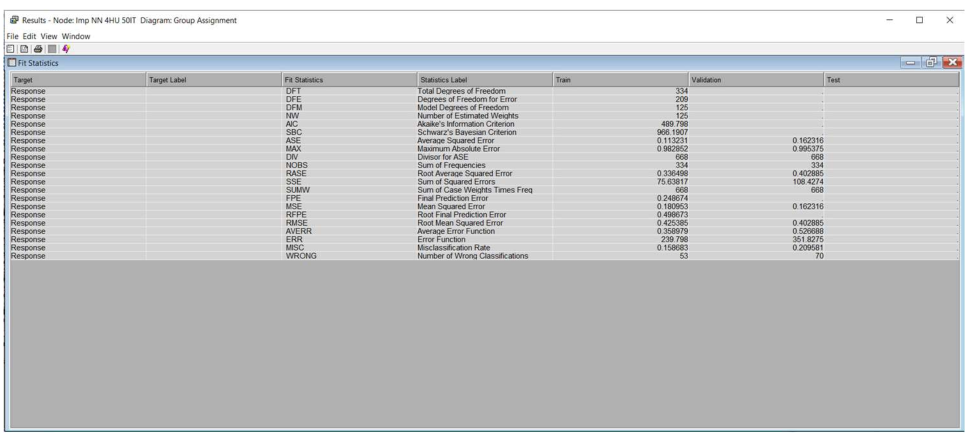


Figure 33 Iteration Plot- Neural Network ran on Impute 3 Hidden Units 100 Iterations (No Preliminary Training)

As depicted in the following figure, it becomes evident that the model reaches convergence around the 3rd iteration. This implies that a significant number of additional iterations might not have been necessary, and the optimal model could likely have been identified by the 3rd iteration.

Neural network with 4 Hidden Units and 50 Iterations

Next trial of neural network has characterized by 4 Hidden Units and 50 Iterations



Results - Node: Imp NN 4HU 50IT Diagram: Group Assignment

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	209	.	.
Response		DFM	Model Degrees of Freedom	125	.	.
Response		NW	Number of Estimated Weights	125	.	.
Response		AIC	Akaike's Information Criterion	489.798	.	.
Response		SBC	Schwarz's Bayesian Criterion	966.1907	.	.
Response		ASE	Average Squared Error	0.113231	0.162316	.
Response		MAX	Maximum Absolute Error	0.982852	0.995375	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.336498	0.402885	.
Response		SSE	Sum of Squared Errors	75.63817	108.4274	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.248674	.	.
Response		MSE	Mean Squared Error	0.180953	0.162316	.
Response		RFPE	Root Final Prediction Error	0.498673	.	.
Response		RMSE	Root Mean Squared Error	0.425385	0.402885	.
Response		AVERR	Average Error Function	0.358979	0.526688	.
Response		ERR	Error Function	239.798	351.8275	.
Response		MISC	Misclassification Rate	0.158683	0.209581	.
Response		WRONG	Number of Wrong Classifications	53	70	.

Figure 34 Fit Statistics - Neural Network ran on Impute Node 4 Hidden Units 50 Iterations (No Preliminary Training)

For the model, it is seen from the figure that the average squared error for the particular model is 0.162316 and the misclassification rate is 0.209581.

Figure 35 Weights Final Plot- Neural Network ran on Impute 4 Hidden Units 50 Iterations (No Preliminary Training)

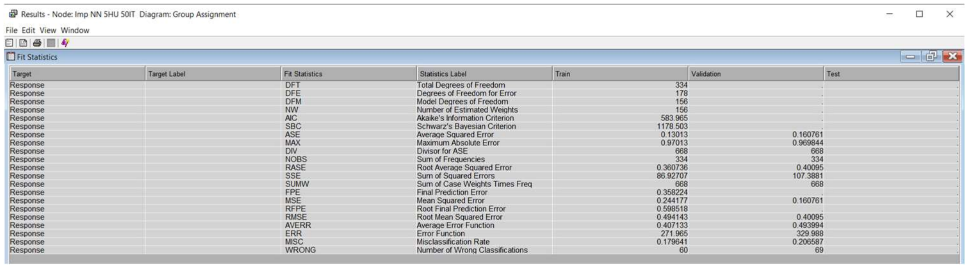
We can see the variables that went into the model were only Teenhome1, NumWebVisitsMonth, NumStorePurchases, NumCatalogPurchases, MntSweetProducts, MntGOldProds, MntFishProducts, Marital_StatusTogether, NumWebPurchases, NumDealPurchases, MntWines, MntMeatProducts, MntFruits, Marital_StatusWidow, Marital_StatusSingle, Marital_StatusDivorced, Marital_StatusAbsurd, Kidhome0, EducationGraduation and Education2nCycle.

Figure 36 Iterations Plot- Neural Network ran on Impute Node 4 Hidden Units 50 Iterations (No Preliminary Training)

Illustrated in the diagram, it becomes apparent that the model achieves convergence approximately by the 10th iteration. This suggests that a notable quantity of further iterations might have been unnecessary, and it is probable that the optimal model could have been discerned by the 10th iteration.

Neural Network with 5 Hidden Units and 50 Iterations

The next neural network was characterized by 5 Hidden Units and 50 Iterations


Results - Node: Imp NN 5HU 50IT Diagram: Group Assignment

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	178	.	.
Response		DFM	Model Degrees of Freedom	156	.	.
Response		NW	Number of Estimated Weights	156	.	.
Response		AIC	Akaike's Information Criterion	583.965	.	.
Response		SBC	Schwarz's Bayesian Criterion	1178.503	.	.
Response		ASE	Average Squared Error	0.13013	0.160761	.
Response		MAX	Maximum Absolute Error	0.97013	0.969844	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.360736	0.40095	.
Response		SSE	Sum of Squared Errors	86.92707	107.3881	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.358224	.	.
Response		MSE	Mean Squared Error	0.244177	0.160761	.
Response		RFPE	Root Final Prediction Error	0.598518	.	.
Response		RMSE	Root Mean Squared Error	0.494143	0.40095	.
Response		AVERR	Average Error Function	0.407133	0.493994	.
Response		ERR	Error Function	271.965	329.988	.
Response		MISC	Misclassification Rate	0.179641	0.206587	.
Response		WRONG	Number of Wrong Classifications	60	69	.

Figure 37 Fit Statistics -Impute Neural Network 5 Hidden Units 50 Iterations (No Preliminary Training)

For the model, it is seen from the figure that the average squared error is 0.160761 and the misclassification rate is 0.206587.

Figure 38 Weights Final Plot- Neural Network ran on Impute 5 Hidden Units 50 Iterations (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, Recency, NumWebPurchases, NumDealPurchases, MntWines, MntMeatProducts, MntFruits, Marital_StatusWidow, Marital_StatusSingle, Marital_StatusDivorced, Marital_StatusAbsurd, Kidhome0, EducationGraduation and Education2nCycle.

Figure 39 Iterations Plot- Neural Network ran on Impute Node 5 Hidden Units 50 Iterations (No Preliminary Training)

As illustrated in the subsequent diagram, it is clear that the model achieves convergence by approximately the sixth iteration. This suggests that a substantial amount of extra iterations might have been avoidable, and it is plausible that the optimal model could have been recognized as early as the sixth iteration.

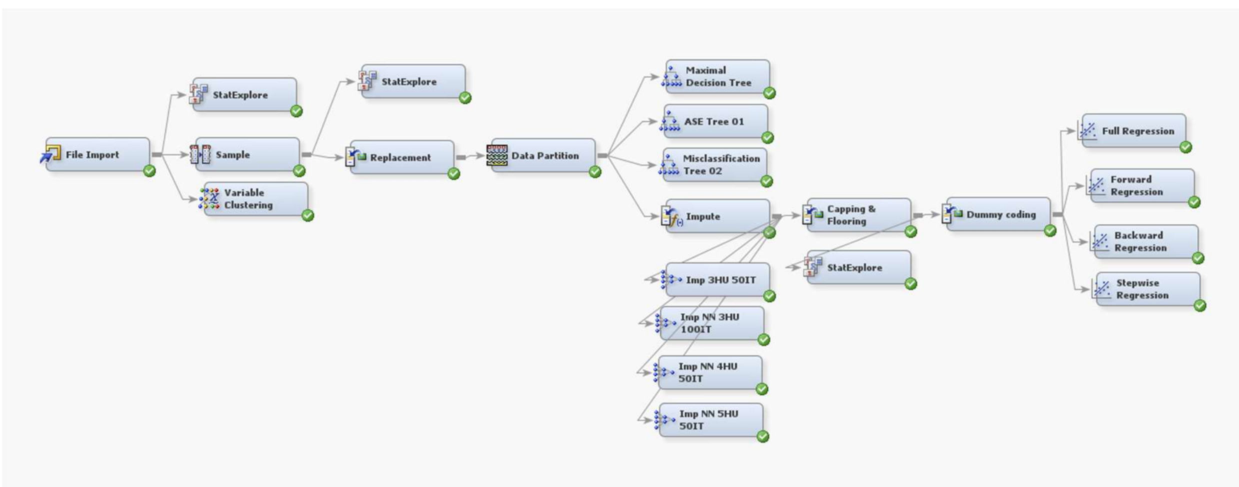
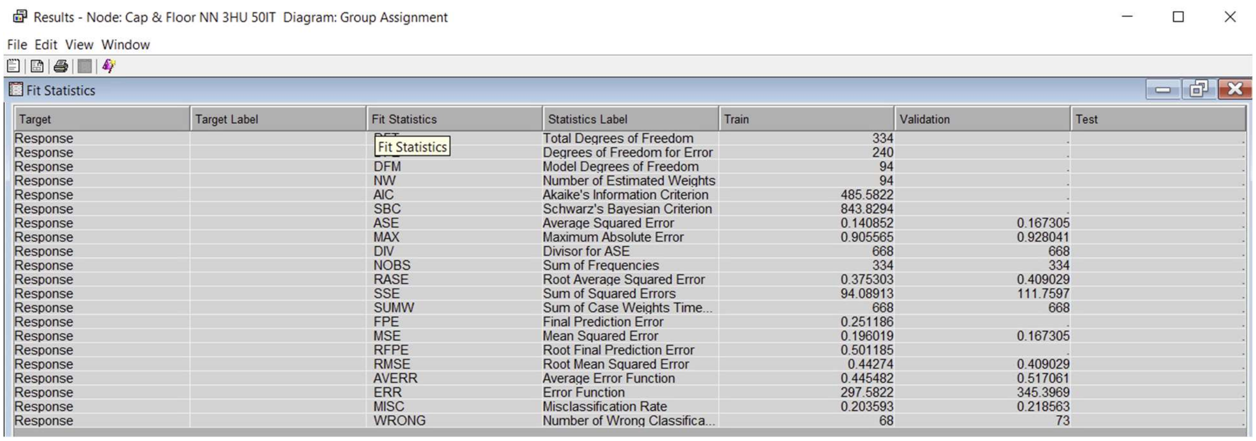


Figure 40 Ran Dummy Coding

Neural Network ran on Capping & Flooring.

Neural Network with 3 Hidden units & 50 Iterations



Results - Node: Cap & Floor NN 3HU 50IT Diagram: Group Assignment

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		Fit Statistics	Total Degrees of Freedom	334	.	.
Response			Degrees of Freedom for Error	240	.	.
Response		DFM	Model Degrees of Freedom	94	.	.
Response		NW	Number of Estimated Weights	94	.	.
Response		AIC	Akaike's Information Criterion	485.5822	.	.
Response		SBC	Schwarz's Bayesian Criterion	843.8294	.	.
Response		ASE	Average Squared Error	0.140852	0.167305	.
Response		MAX	Maximum Absolute Error	0.905565	0.928041	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.375303	0.409029	.
Response		SSE	Sum of Squared Errors	94.08913	111.7597	.
Response		SUMW	Sum of Case Weights Time...	668	668	.
Response		FPE	Final Prediction Error	0.251186	.	.
Response		MSE	Mean Squared Error	0.196019	0.167305	.
Response		RFPE	Root Final Prediction Error	0.501185	.	.
Response		RMSE	Root Mean Squared Error	0.44274	0.409029	.
Response		AVERR	Average Error Function	0.445482	0.517061	.
Response		ERR	Error Function	297.5822	345.3969	.
Response		MISC	Misclassification Rate	0.203593	0.218563	.
Response		WRONG	Number of Wrong Classifica...	68	73	.

Figure 41 Fit Statistics –Capping and Flooring Node -Neural Network 3 Hidden Units 50 Iterations (No Preliminary Training)

For the neural network model ran on Capping and Flooring, with 3 hidden units and 50 iterations have an average squared error of 0.167305 and misclassification rate of 0.218563.

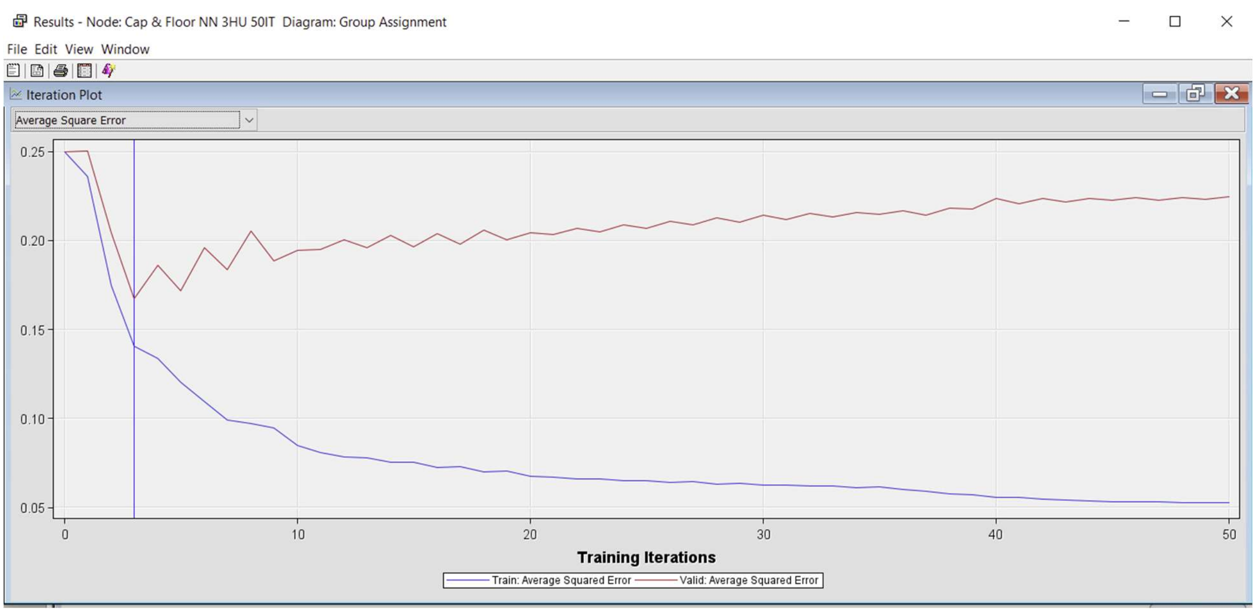


Figure 42 Iterations Plot- Neural Network ran on Cap& Floor Node 3 Hidden Units 50 Iterations (No Preliminary Training)

It can be seen that the model reaches convergence at the third iteration while keeping average squared error as selection criteria.

Figure 43 Weights Final Plot- Neural Network ran on capping and flooring 3 Hidden Units 50 Iterations (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, REP_Recency, REP_NumWebPurchases, REP_NumDealPurchases, REP_MntWines, REP_MntMeatProducts, REP_MntFruits, REP_IMP_REP_Income, Marital_StatusTogether, Marital_StatusMarried, Marital_StatusAlone, Kidhome1, EducationGraduation and Education2nCycle.

Neural Network with 3 Hidden units & 100 Iterations

Results - Node: Cap & Floor NN 3HU 100IT Diagram: Group Assignment

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	240	.	.
Response		DFM	Model Degrees of Freedom	94	.	.
Response		NW	Number of Estimated Weights	94	.	.
Response		AIC	Akaike's Information Criterion	485.5822	.	.
Response		SBC	Schwarz's Bayesian Criterion	843.8294	.	.
Response		ASE	Average Squared Error	0.140852	0.167305	.
Response		MAX	Maximum Absolute Error	0.905565	0.928041	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.375303	0.409029	.
Response		SSE	Sum of Squared Errors	94.08913	111.7597	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.251186	.	.
Response		MSE	Mean Squared Error	0.196019	0.167305	.
Response		RFPE	Root Final Prediction Error	0.501185	.	.
Response		RMSE	Root Mean Squared Error	0.44274	0.409029	.
Response		AVERR	Average Error Function	0.445482	0.517061	.
Response		ERR	Error Function	297.5822	345.3969	.
Response		MISC	Misclassification Rate	0.203593	0.218563	.
Response		WRONG	Number of Wrong Classifications	68	73	.

Figure 44 Fit Statistics -Capping and Flooring - Neural Network 3 Hidden Units 100 Iterations (No Preliminary Training)

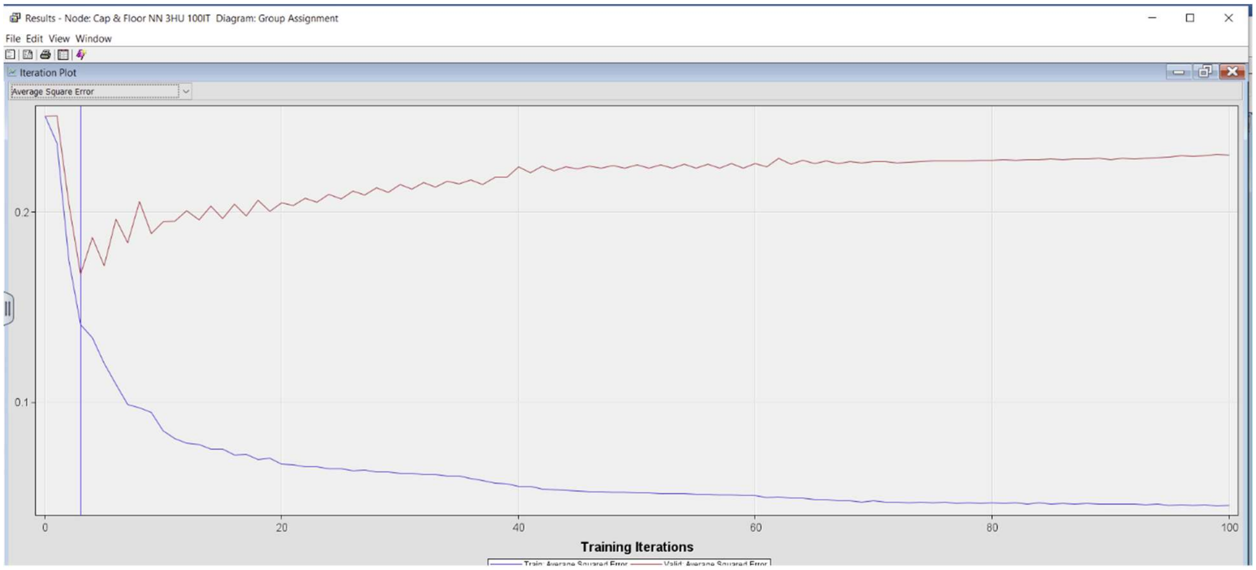


Figure 45 Iterations Plot- Neural Network ran on Impute Node 3 Hidden Units 100 Iterations (No Preliminary Training)

The model reaches convergence at the third iteration while keeping average squared error as selection criteria.

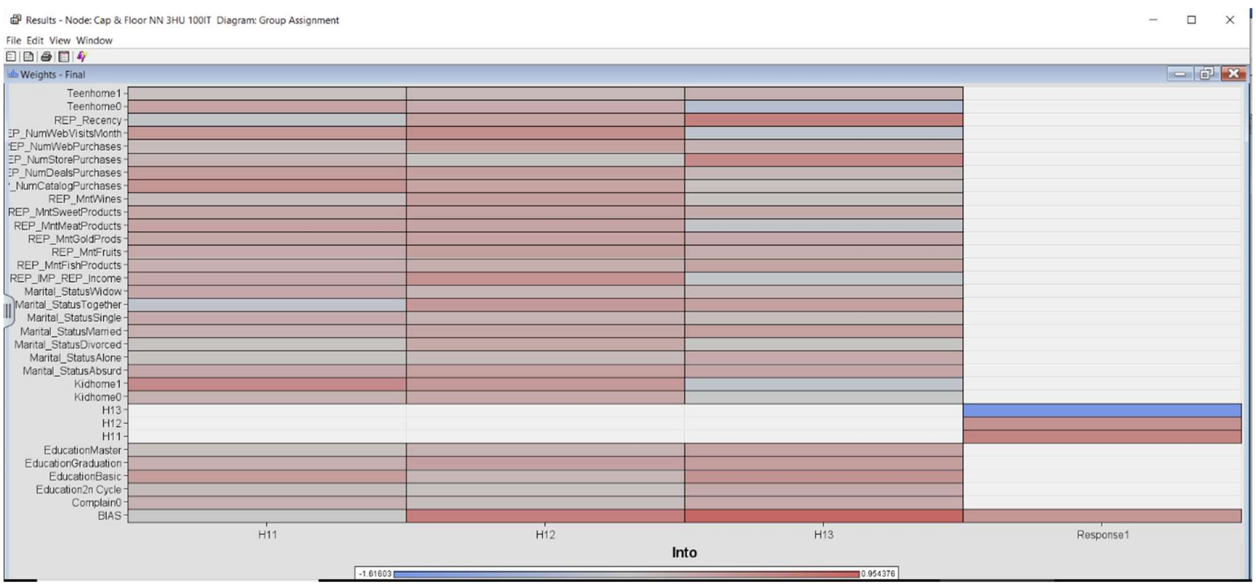


Figure 46 Weights Final Plot- Neural Network ran on Capping and Flooring 3 Hidden Units 100 Iterations (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, REP_Recency, REP_NumWebPurchases, REP_NumDealPurchases, REP_MntWines, REP_MntMeatProducts, REP_MntFruits, REP_IMP_REP_Income, Marital_StatusTogether, Marital_StatusMarried, Marital_StatusAlone, Kidhome1, EducationGraduation and Education2nCycle.

Neural Network with 4 Hidden units & 50 Iterations

Results - Node: Cap & Floor NN 4HU 50IT Diagram: Group Assignment

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	209	.	.
Response		DFM	Model Degrees of Freedom	125	.	.
Response		NW	Number of Estimated Weights	125	.	.
Response		AIC	Akaike's Information Criterion	526.2428	.	.
Response		SBC	Schwarz's Bayesian Criterion	1002.635	.	.
Response		ASE	Average Squared Error	0.132161	0.162252	.
Response		MAX	Maximum Absolute Error	0.972986	0.981074	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.363539	0.402805	.
Response		SSE	Sum of Squared Errors	88.28329	108.384	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.290247	.	.
Response		MSE	Mean Squared Error	0.211204	0.162252	.
Response		RFPE	Root Final Prediction Error	0.538746	.	.
Response		RMSE	Root Mean Squared Error	0.459569	0.402805	.
Response		AVERR	Average Error Function	0.413537	0.511803	.
Response		ERR	Error Function	276.2428	341.8844	.
Response		MISC	Misclassification Rate	0.182635	0.239521	.
Response		WRONG	Number of Wrong Classifications	61	80	.

Figure 47 Fit Statistics -Capping and Flooring - Neural Network 4 Hidden Units 50 Iterations (No Preliminary Training)

For the neural network model connected to the capping and flooring node, with 4 hidden units and 50 iterations, the average squared error was seen to be 0.162252 and the misclassification rate as 0.239521.

Figure 48 Iterations Plot- Neural Network ran on Capping and Flooring Node 4 Hidden Units 50 Iterations (No Preliminary Training)

The model reaches convergence at the 7^{th} iteration while keeping average squared error as selection criteria.

Figure 49 Weights Final Plot- Neural Network ran on capping and flooring 4 Hidden Units 50 Iterations (No Preliminary Training)

We can see the variables that went into the model were Teenhome1, Teenhome0 REP_Recency, REP_NumWebVisitsMonth, REP_NumWebPurchases, REP_NumStorePurchases, REP_NumDealsPurchases, REP_NumCatalogPurchases, REP_MntWines, REP_MntSweetProducts, REP_MntMeatProducts, REP_MntGoldProds, REP_MntFruits, REP_MntFishProducts, REP_IMP_REP_Income,Marital_StatusWidow, Marital_StatusTogether, Marital_StatusSingle, Marital_StatusMarried, Marital_StatusDivorced, Marital_StatusAlone, Marital_StatusAbsurd, KidHome1, Kidhome0, EducationMaster, EducationGraduation, EducationBasic, Education Cycle,Complain0. This may not be considered an ideal model as there are so many variables going into the model.

Neural Network with 5 Hidden units & 50 Iterations

Results - Node: Cap & Floor NN 5HU 50IT Diagram: Group Assignment

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	178	.	.
Response		DFM	Model Degrees of Freedom	156	.	.
Response		NW	Number of Estimated Weights	156	.	.
Response		AIC	Akaike's Information Criterion	585.1612	.	.
Response		SBC	Schwarz's Bayesian Criterion	1179.699	.	.
Response		ASE	Average Squared Error	0.130587	0.161202	.
Response		MAX	Maximum Absolute Error	0.973597	0.961332	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.361368	0.401499	.
Response		SSE	Sum of Squared Errors	87.23219	107.6828	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.359481	.	.
Response		MSE	Mean Squared Error	0.245034	0.161202	.
Response		RFPE	Root Final Prediction Error	0.599568	.	.
Response		RMSE	Root Mean Squared Error	0.495009	0.401499	.
Response		AVERR	Average Error Function	0.408924	0.493071	.
Response		ERR	Error Function	273.1612	329.3716	.
Response		MISC	Misclassification Rate	0.170659	0.221557	.
Response		WRONG	Number of Wrong Classifications	57	74	.

Figure 50 Fit Statistics -Capping and Flooring - Neural Network 5 Hidden Units 50 Iterations (No Preliminary Training)

On this particular run of Neural Network with 5 hidden units and 50 iterations, connected to the cap and floor node, it can be seen from the above figure that the average squared error value is coming to 0.161202 and the misclassification rate is 0.221557.

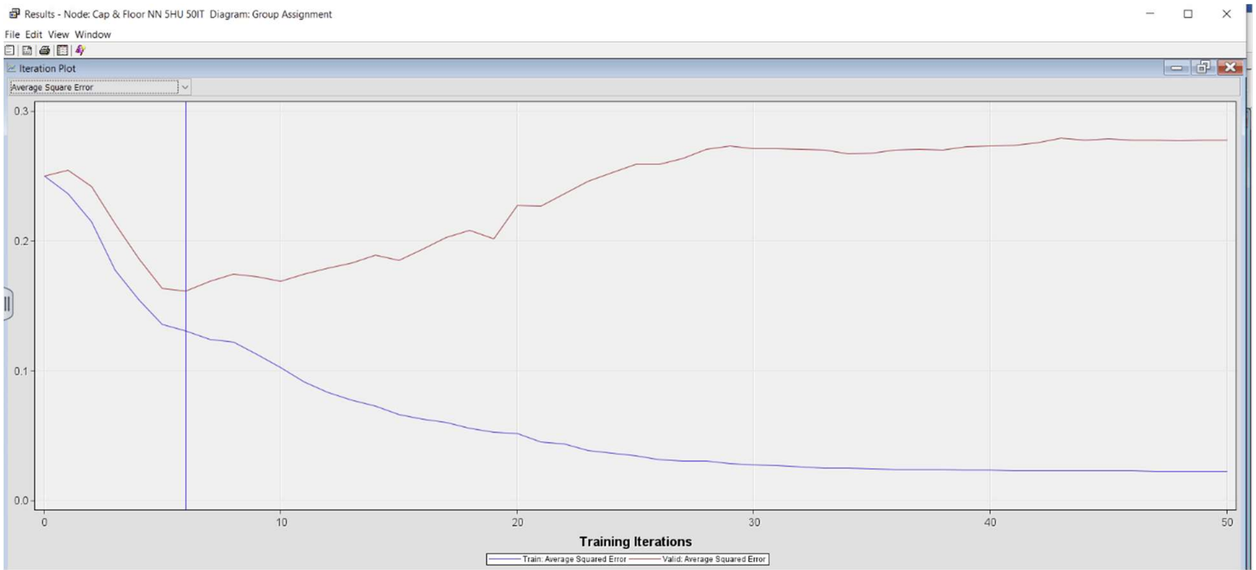


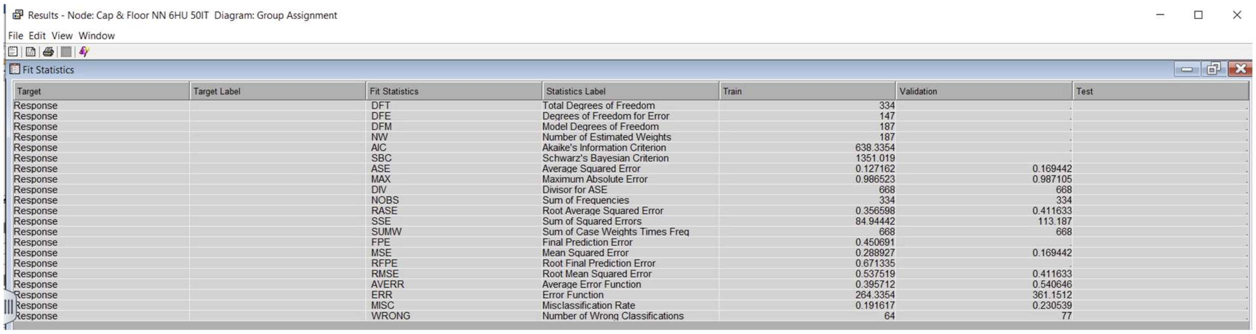
Figure 51 Iterations Plot- Neural Network ran on Capping and Flooring Node 5 Hidden Units 50 Iterations (No Preliminary Training)

The model reaches convergence at the 6th iteration while keeping average squared error as selection criteria.

Figure 52 Weights Final Plot- Neural Network ran on capping and flooring 5 Hidden Units 50 Iterations (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, REP_Recency, REP_NumWebPurchases, REP_NumDealPurchases, REP_MntWines, REP_MntMeatProducts, REP_MntFruits, REP_IMP_REP_Income, Marital_StatusTogether, Marital_StatusMarried, Marital_StatusAlone, Kidhome1, EducationGraduation and Education2nCycle.

Neural Network with 6 Hidden units & 50 Iterations


Results - Node: Cap & Floor NN 6HU 50IT Diagram: Group Assignment
File Edit View Window
Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	147	.	.
Response		DFM	Model Degrees of Freedom	187	.	.
Response		NW	Number of Estimated Weights	187	.	.
Response		AIC	Akaike's Information Criterion	638.3354	.	.
Response		SBC	Schwarz's Bayesian Criterion	1351.019	.	.
Response		ASE	Average Squared Error	0.127162	0.169442	.
Response		MAX	Maximum Absolute Error	0.986523	0.987105	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.356598	0.411633	.
Response		SSE	Sum of Squared Errors	84.94442	113.187	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.450691	.	.
Response		MSE	Mean Squared Error	0.288927	0.169442	.
Response		RFPE	Root Final Prediction Error	0.671335	.	.
Response		RMSE	Root Mean Squared Error	0.537519	0.411633	.
Response		AVERR	Average Error Function	0.395712	0.540646	.
Response		ERR	Error Function	264.3354	361.1512	.
Response		MISC	Misclassification Rate	0.191617	0.230539	.
Response		WRONG	Number of Wrong Classifications	64	77	.

Figure 53 Fit Statistics -capping and Flooring - Neural Network 6 Hidden Units 50 Iterations (No Preliminary Training)

The next neural network node with 6 hidden units and 60 iterations, connected to the cap and floor node, the average squared error is seen as 0.169442 and the misclassification of the validation data is seen as 0.230539.

Figure 54 Iterations Plot- Neural Network ran on capping and flooring Node 6 Hidden Units 50 Iterations (No Preliminary Training)

The model reaches convergence at the 6th iteration while keeping average squared error as selection criteria.

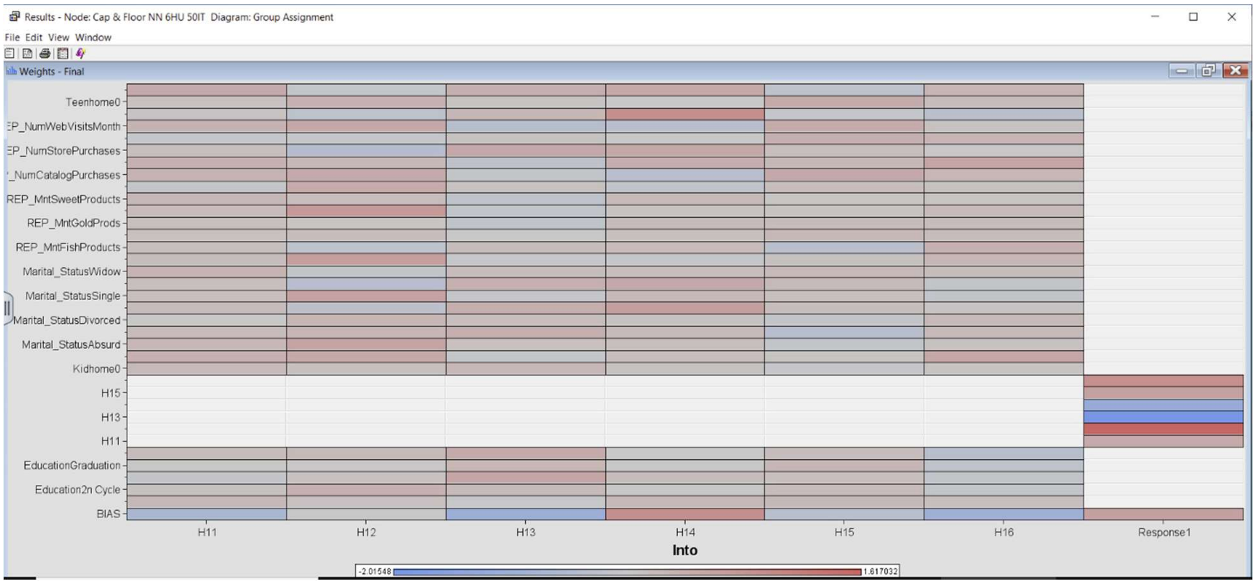


Figure 55 Weights Final Plot- Neural Network ran on capping and flooring 6 Hidden Units 50 Iterations (No Preliminary Training)

We can see the variables that went into the model were only Teenhome0, REP_NumWebVisitsMonth, REP_NumStorePurchases, REP_NumCatalogPurchases, REP_MntSweetProducts, REP_MntGoldProds, REP_MntFishProducts, Marital_StatusWidow,Marital_StatusSingle, Marital_StatusDivorced, Marital_StatusAbsurd, Kidhome0, EducationGraduation, Education2n Cycle.

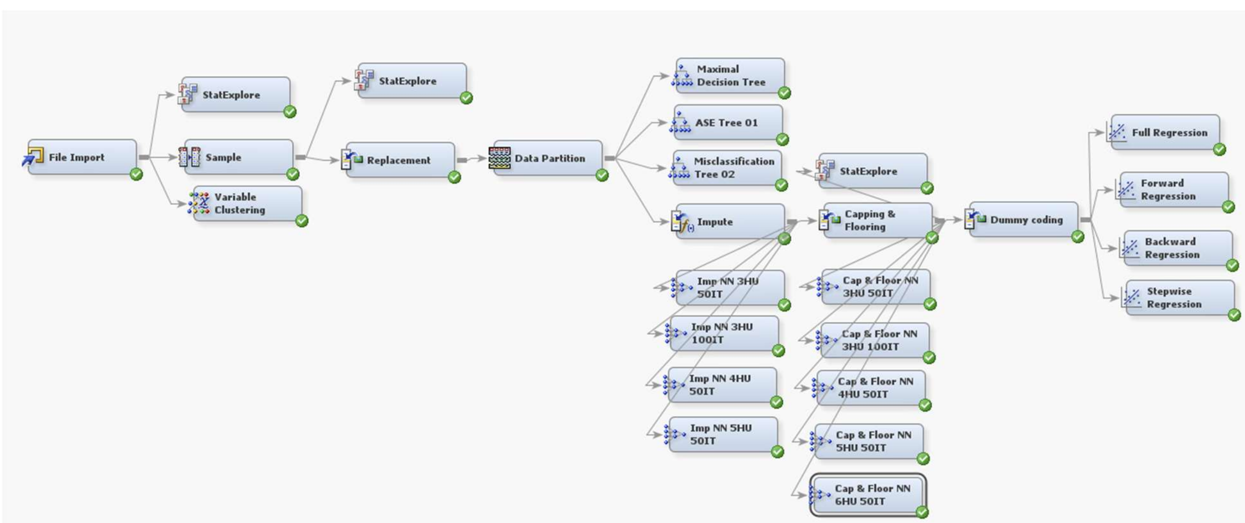


Figure 56 Diagram after connecting Neural Network to Cap & Floor node

Neural Network ran on Dummy Coding Node

Neural network nodes were connected to the Dummy Coding Node. Like the steps performed with the impute node, the hidden units and iterations started at 3 and 50 respectively, with steady increments to find the optimum value of ASE.

Neural Network with 3 Hidden units & 50 Iterations

Results - Node: DC NN 3HU 50IT Diagram: Group Assignment

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	246	.	.
Response		DFM	Model Degrees of Freedom	88	.	.
Response		NW	Number of Estimated Weights	88	.	.
Response		AIC	Akaike's Information Criterion	465.2416	.	.
Response		SBC	Schwarz's Bayesian Criterion	800.622	.	.
Response		ASE	Average Squared Error	0.137051	0.163005	.
Response		MAX	Maximum Absolute Error	0.941081	0.886452	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.370203	0.403739	.
Response		SSE	Sum of Squared Errors	91.54976	108.8873	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.235103	.	.
Response		MSE	Mean Squared Error	0.186077	0.163005	.
Response		RFPE	Root Final Prediction Error	0.484874	.	.
Response		RMSE	Root Mean Squared Error	0.431366	0.403739	.
Response		AVERR	Average Error Function	0.432996	0.497986	.
Response		ERR	Error Function	289.2416	332.6544	.
Response		MISC	Misclassification Rate	0.188623	0.218563	.
Response		WRONG	Number of Wrong Classifications	63	73	.

Figure 57 Fit Statistics – Neural Network ran on Dummy Coding 3HU 50IT (No preliminary Training)

From the Figure above, we can see that the average squared error is 0.163005 and the misclassification rate is for the validation data is 0.218563.

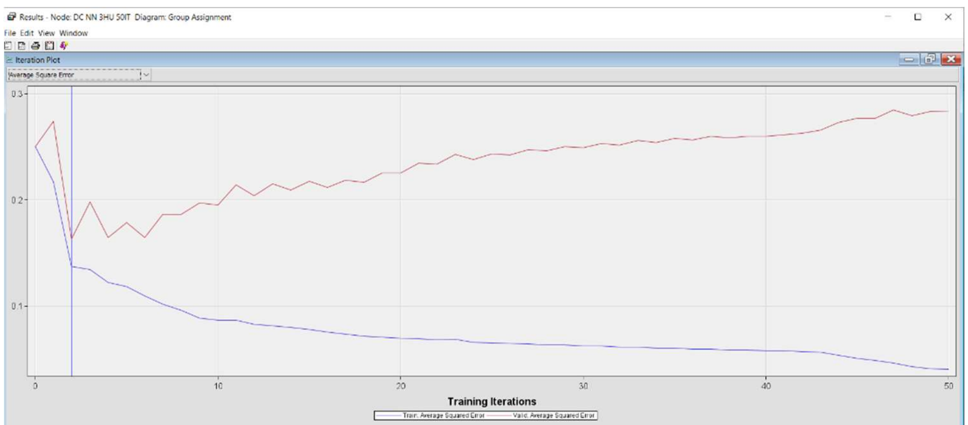


Figure 58 Iteration Plot – Neural network ran on Dummy Coding 3HU 50IT (No Preliminary Training)

Depicted in the above figure, it is evident that the model attains convergence around the second iteration. This implies that a notable number of surplus iterations could have been circumvented, and it is conceivable that the optimal model might have been identified as soon as the second iteration.

Figure 59 Weights Final Plot- Neural Network ran on Dummy Coding (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, Teenhome0 REP_Recency, REP_NumWebVisitsMonth, REP_NumWebPurchases, REP_NumStorePurchases, REP_NumDealsPurchases, REP_NumCatalogPurchases, REP_MntWines, REP_MntSweetProducts, REP_MntMeatProducts, REP_MntGoldProds, REP_MntFruits, REP_MntFishProducts, Marital_StaWidow, Marital_StatusMarried/Together, Marital_StaMarried, Marital_StatusDivorce, Marital_StatusSingle/Absurd REP_IMP_REP_Income, KidHome1, Kidhome0, EducationMaster, EducationGraduation, EducationBasic, Education2n Cycle,Complain0. This is again, too many variables going into the model and is not seen as ideal.

Neural Network with 3 Hidden Units and 100 Iterations

Results - Node: DC NN 3HU 100IT Diagram: Group Assignment

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	246	.	.
Response		DFM	Model Degrees of Freedom	88	.	.
Response		NW	Number of Estimated Weights	88	.	.
Response		AIC	Akaike's Information Criterion	465.2416	.	.
Response		SBC	Schwarz's Bayesian Criterion	800.622	.	.
Response		ASE	Average Squared Error	0.137051	0.163005	.
Response		MAX	Maximum Absolute Error	0.941081	0.886452	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.370203	0.403739	.
Response		SSE	Sum of Squared Errors	91.54976	108.8873	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.235103	.	.
Response		MSE	Mean Squared Error	0.186077	0.163005	.
Response		RFPE	Root Final Prediction Error	0.484874	.	.
Response		RMSE	Root Mean Squared Error	0.431366	0.403739	.
Response		AVERR	Average Error Function	0.432996	0.497986	.
Response		ERR	Error Function	289.2416	332.6544	.
Response		MISC	Misclassification Rate	0.188623	0.218563	.
Response		WRONG	Number of Wrong Classifications	63	73	.

Figure 60 Fit Statistics- Neural Network ran on Dummy Coding 3HU 100IT (No Preliminary Training)

For the model, we can see that the average squared error is 0.163005 and the misclassification rate is 0.218563. Since the ASE value is same, we go back to 50 iterations and with 4 hidden units for the next run.

Figure 61 Weights Final Plot- Neural Network ran on Dummy Coding 3HU 100 IT (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, Teenhome0 REP_Recency, REP_NumWebVisitsMonth, REP_NumWebPurchases, REP_NumStorePurchases, REP_NumDealsPurchases, REP_NumCatalogPurchases, REP_MntWines, REP_MntSweetProducts, REP_MntMeatProducts, REP_MntGoldProds, REP_MntFruits, REP_MntFishProducts, REP_Marital_StaWidow, REP_Marital_StaMarried/Together, REP_Marital_StaMarried, REP_Marital_StaDivorce, REP_Marital_StaSingle/Absurd, REP_IMP_REP_Income, KidHome1, Kidhome0, EducationMaster, EducationGraduation, EducationBasic, Education2n Cycle,Complain0.

Figure 62 Iterations Plot – Neural Network ran on Dummy Coding 3HU 100IT (No Preliminary Training)

Depicted in the following diagram, it is evident that the model attains convergence around the second iteration. This implies that a notable number of surplus iterations could have been circumvented, and it is conceivable that the optimal model might have been identified as soon as the second iteration.

Neural Network with 4 Hidden Unit and 50 Iterations

Results - Node: DC NN 4HU 50IT Diagram: Group Assignment

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	217	.	.
Response		DFM	Model Degrees of Freedom	117	.	.
Response		NW	Number of Estimated Weights	117	.	.
Response		AIC	Akaike's Information Criterion	500.54	.	.
Response		SBC	Schwarz's Bayesian Criterion	946.4435	.	.
Response		ASE	Average Squared Error	0.125322	0.168301	.
Response		MAX	Maximum Absolute Error	0.941027	0.966432	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.354008	0.410244	.
Response		SSE	Sum of Squared Errors	83.71478	112.4247	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.260461	.	.
Response		MSE	Mean Squared Error	0.192891	0.168301	.
Response		RFPE	Root Final Prediction Error	0.510354	.	.
Response		RMSE	Root Mean Squared Error	0.439194	0.410244	.
Response		AVERR	Average Error Function	0.399012	0.523088	.
Response		ERR	Error Function	266.54	349.4227	.
Response		MISC	Misclassification Rate	0.167665	0.224551	.
Response		WRONG	Number of Wrong Classifications	56	75	.

Figure 63 Fit Statistics – Neural Network ran on Dummy Coding 4HU 50IT (No Preliminary Training)

For the model, we can see that the average squared error is *0.168301* and the misclassification rate is *0.224551*

Figure 64 Weights Final Plot – Neural Network ran on Dummy Coding 4HU 50IT (No Preliminary Training)

We can see the variables that went into the model were only Teenhome0, REP_NumWebVisitsMonth, REP_NumStorePurchases, REP_NumCatalogPurchases, REP_MntSweetProducts, REP_MntGoldProds, REP_MntFishProducts, REP_Marital_StaMarried/Together, REP_Marital_StaDivorce, REP_IMP_REP_Income, Kidhome0, EducationGraduation, Education2n Cycle.

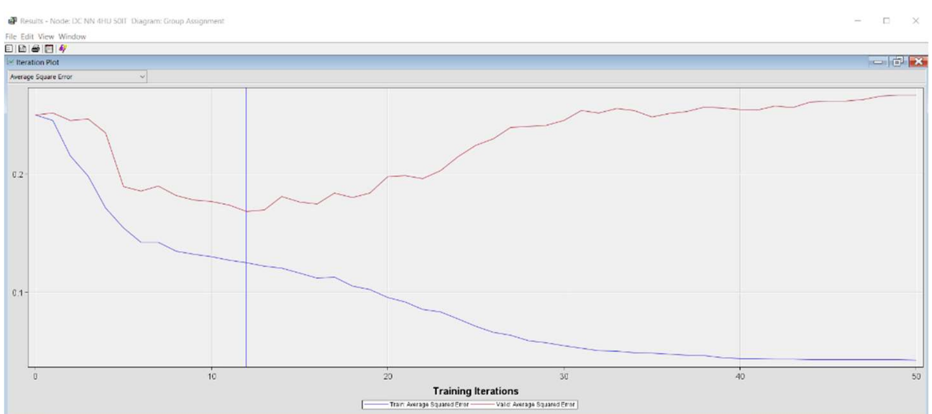


Figure 65 Iteration Plot- Neural Network ran on Dummy Coding 4HU 50IT (No Preliminary Training)

It is clear from the following figure that the model reaches convergence at the 12th iteration. This suggests that a significant number of excess iterations may have been avoided, and it is possible that the best model may have been found as early as the second iteration.

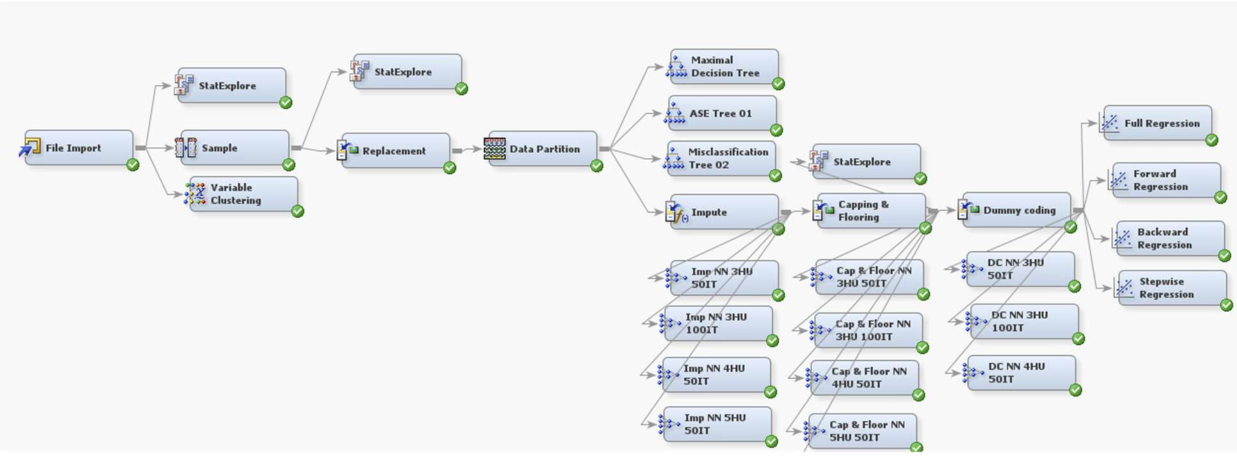


Figure 66 Diagram after linking Neural networks to Dummy Coding

Neural Network Ran on Backward Regression

Neural Network with 3 Hidden Units and 50 Iterations

Results - Node: BWReg NN 3HU 50IT Diagram: Group Assignment

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	282	.	.
Response		DFM	Model Degrees of Freedom	52	.	.
Response		NW	Number of Estimated Weights	52	.	.
Response		AIC	Akaike's Information Criterion	404.3847	.	.
Response		SBC	Schwarz's Bayesian Criterion	602.564	.	.
Response		ASE	Average Squared Error	0.148372	0.162428	.
Response		MAX	Maximum Absolute Error	0.945835	0.967192	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.385191	0.403024	.
Response		SSE	Sum of Squared Errors	99.11275	108.5019	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.203091	.	.
Response		MSE	Mean Squared Error	0.175732	0.162428	.
Response		RFPE	Root Final Prediction Error	0.450656	.	.
Response		RMSE	Root Mean Squared Error	0.419204	0.403024	.
Response		AVERR	Average Error Function	0.449678	0.496328	.
Response		ERR	Error Function	300.3847	331.5469	.
Response		MISC	Misclassification Rate	0.221557	0.233533	.
Response		WRONG	Number of Wrong Classifications	74	78	.

Figure 67 Fit Statistics- Neural Network ran on Backward Regression 3HU 50 IT (No Preliminary Training)

For the model, we can see that the average squared error is *0.162428* and the misclassification rate is *0.233533*

Figure 68 Weights Final Plot- Neural Network ran on Backward Regression 3HU 50IT (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, Teenhome0, REP_Recency, REP_NumWebVisitsMonth, REP_NumStorePurchases, REP_NumDealsPurchases, REP_NumCatalogPurchases, REP_MntMeatProducts, REP_Marital_StatusSingle, REP_Marital_StatusCouple, REP_IMP_REP_Income, REP_EducationPost_Graduate, REP_EducationGraduation.

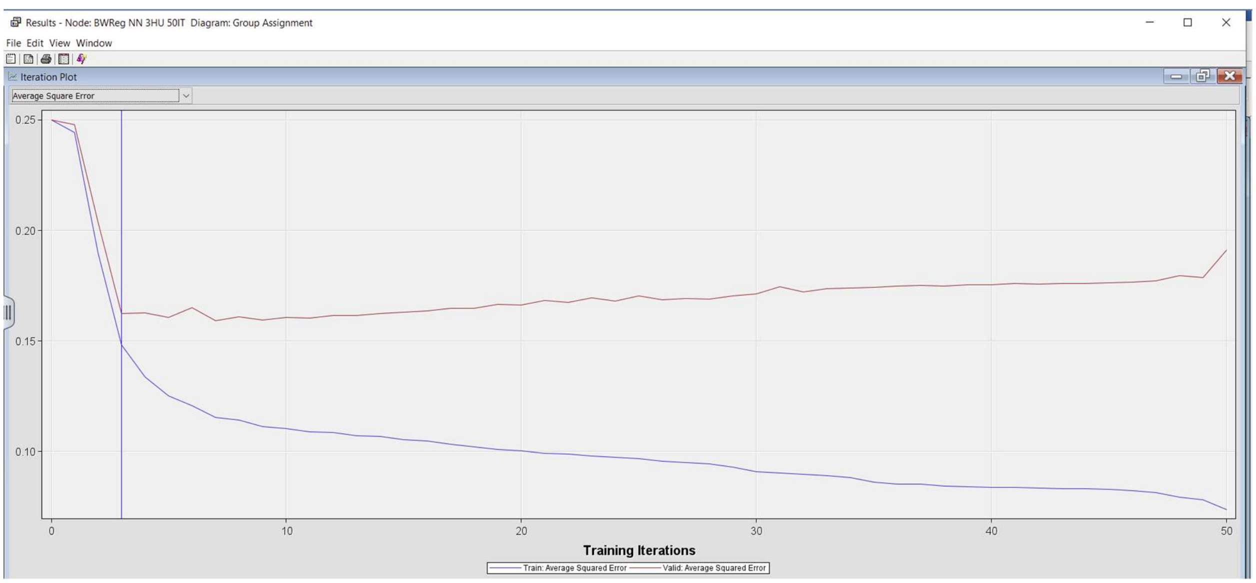


Figure 69 Iterations Plot- Neural Network ran on Backward Regression 3HU 50IT (No Preliminary Training)

As shown in the subsequent diagram, it becomes clear that the model reaches convergence by the third iteration. This indicates that a significant count of excess iterations could have been avoided, and it is conceivable that the optimal model could have been recognized by the third iteration.

Neural Network with 3 Hidden Units and 100 Iterations

Results - Node: BWReg NN 3HU 100IT Diagram: Group Assignment

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	282	.	.
Response		DFM	Model Degrees of Freedom	52	.	.
Response		NW	Number of Estimated Weights	52	.	.
Response		AIC	Akaike's Information Criterion	404.3847	.	.
Response		SBC	Schwarz's Bayesian Criterion	602.564	.	.
Response		ASE	Average Squared Error	0.148372	0.162428	.
Response		MAX	Maximum Absolute Error	0.945835	0.967192	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.385191	0.403024	.
Response		SSE	Sum of Squared Errors	99.11275	108.5019	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.203091	.	.
Response		MSE	Mean Squared Error	0.175732	0.162428	.
Response		RFPE	Root Final Prediction Error	0.450656	.	.
Response		RMSE	Root Mean Squared Error	0.419204	0.403024	.
Response		AVERR	Average Error Function	0.449678	0.496328	.
Response		ERR	Error Function	300.3847	331.5469	.
Response		MISC	Misclassification Rate	0.221557	0.233533	.
Response		WRONG	Number of Wrong Classifications	74	78	.

Figure 70 Fit Statistics- Neural Network ran on Backward Regression 3HU 100 IT (No Preliminary Training)

For the model, we can see that the average squared error is *0.162428* and the misclassification rate is *0.233533*

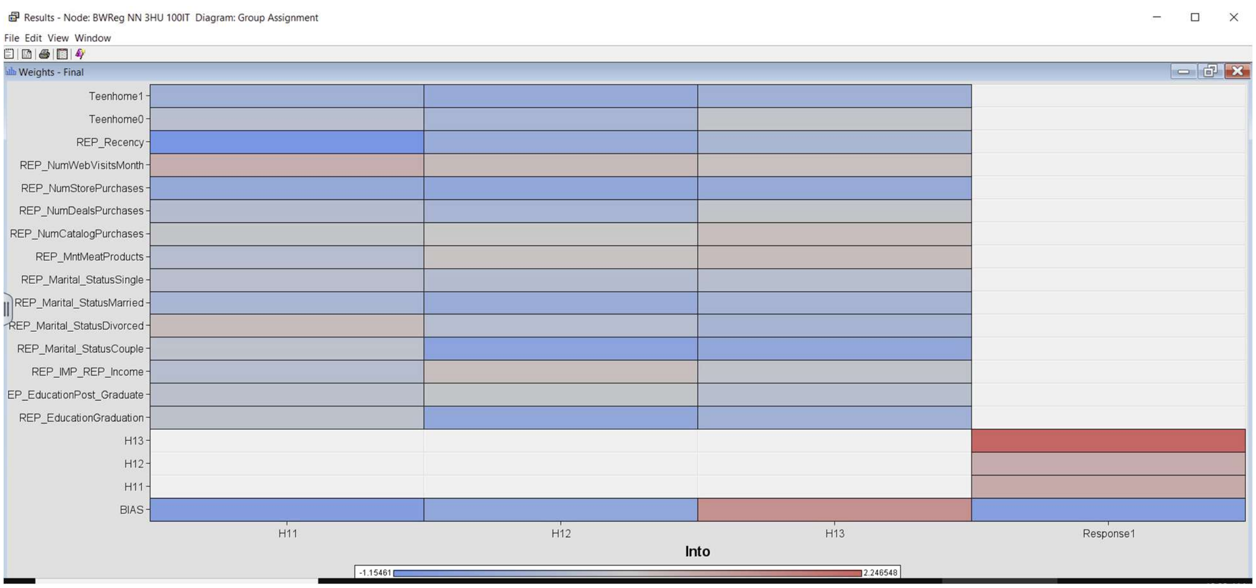


Figure 71 Weights Final Plot- Neural Network ran on Backward Regression 3HU 100IT (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, Teenhome0, REP_Recency, REP_NumWebVisitsMonth, REP_NumStorePurchases, REP_NumDealsPurchases, REP_NumCatalogPurchases, REP_MntMeatProducts, REP_Marital_StatusSingle, REP_Marital_StatusMarried, REP_Marital_StatusDivored, REP_Marital_StatusCouple, REP_IMP_REP_Income, REP_EducationPost_Graduate, REP_EducationGraduation.

Figure 72 Iterations Plot- Neural Network ran on Backward Regression 3HU 100IT (No Preliminary Training)

As shown in the subsequent diagram, it becomes clear that the model reaches convergence by the third iteration. This indicates that a significant count of excess iterations could have been avoided, and it is conceivable that the optimal model could have been recognized by the third iteration.

Neural Network on Backward Regression with 4 Hidden Units and 50 Iterations

Results - Node: BWReg NN 4HU 50IT Diagram: Group Assignment

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Response		DFT	Total Degrees of Freedom	334	.	.
Response		DFE	Degrees of Freedom for Error	265	.	.
Response		DFM	Model Degrees of Freedom	69	.	.
Response		NW	Number of Estimated Weights	69	.	.
Response		AIC	Akaike's Information Criterion	406.534	.	.
Response		SBC	Schwarz's Bayesian Criterion	669.5027	.	.
Response		ASE	Average Squared Error	0.129038	0.169164	.
Response		MAX	Maximum Absolute Error	0.960427	0.994397	.
Response		DIV	Divisor for ASE	668	668	.
Response		NOBS	Sum of Frequencies	334	334	.
Response		RASE	Root Average Squared Error	0.359218	0.411296	.
Response		SSE	Sum of Squared Errors	86.19713	113.0018	.
Response		SUMW	Sum of Case Weights Times Freq	668	668	.
Response		FPE	Final Prediction Error	0.196235	.	.
Response		MSE	Mean Squared Error	0.162636	0.169164	.
Response		RFPE	Root Final Prediction Error	0.442984	.	.
Response		RMSE	Root Mean Squared Error	0.403282	0.411296	.
Response		AVERR	Average Error Function	0.401997	0.53516	.
Response		ERR	Error Function	268.534	357.4866	.
Response		MISC	Misclassification Rate	0.185629	0.245509	.
Response		WRONG	Number of Wrong Classifications	62	82	.

Figure 73 Fit Statistics- Neural Network ran on Backward Regression 4HU 50 IT (No Preliminary Training)

For the model, we can see that the average squared error is *0.169164* and the misclassification rate is *0.245509*

Figure 74 Weights Final Plot- Neural Network ran on Backward Regression 4HU 50IT (No Preliminary Training)

We can see the variables that went into the model were only Teenhome1, Teenhome0, REP_Recency, REP_NumWebVisitsMonth, REP_NumStorePurchases, REP_NumDealsPurchases, REP_NumCatalogPurchases, REP_MntMeatProducts, REP_Marital_StatusSingle, REP_Marital_StatusMarried, REP_Marital_StatusDivored, REP_Marital_StatusCouple, REP_IMP_REP_Income, REP_EducationPost_Graduate, REP_EducationGraduation.

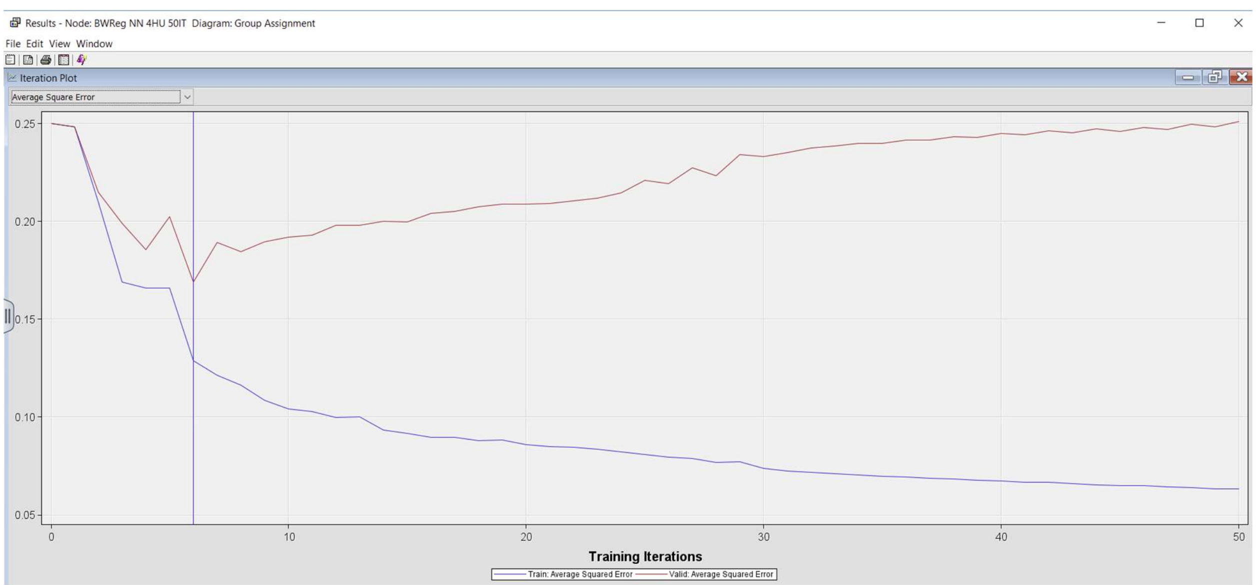


Figure 75 Iterations Plot- Neural Network ran on Backward Regression 4HU 50IT (No Preliminary Training)

As shown in the subsequent diagram, it becomes clear that the model reaches convergence by the sixth iteration. This indicates that a significant count of excess iterations could have been avoided, and it is conceivable that the optimal model could have been recognized by the sixth iteration.

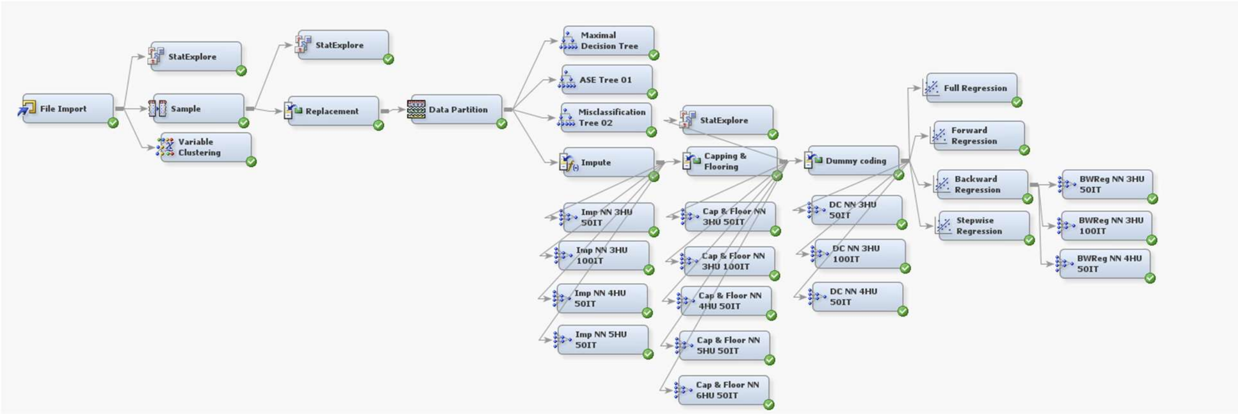


Figure 76 Diagram after connecting all the Neural Networks

Model Comparison

Users can evaluate and compare numerous predictive models using SAS Enterprise Miner's Model Comparison node depending on various models, such as the ASE (average squared error) selection criteria. This method enables a thorough assessment of model performance because the average squared discrepancies between predicted and actual values are quantified by the ASE metric. Analysts can identify the model that minimizes prediction errors the best by using ASE as a selection criterion, ensuring that the model chosen will produce accurate and dependable results for their data-driven pursuits.

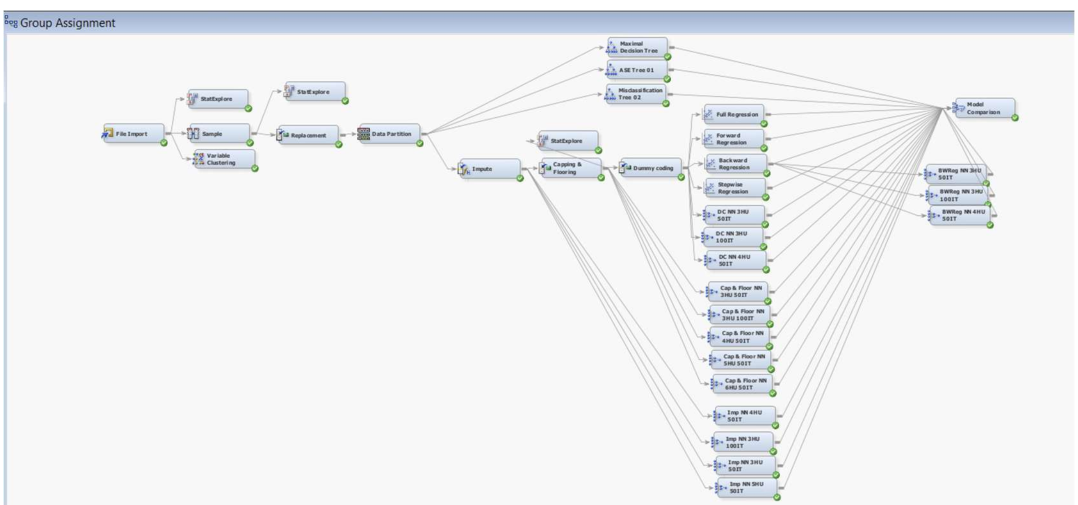
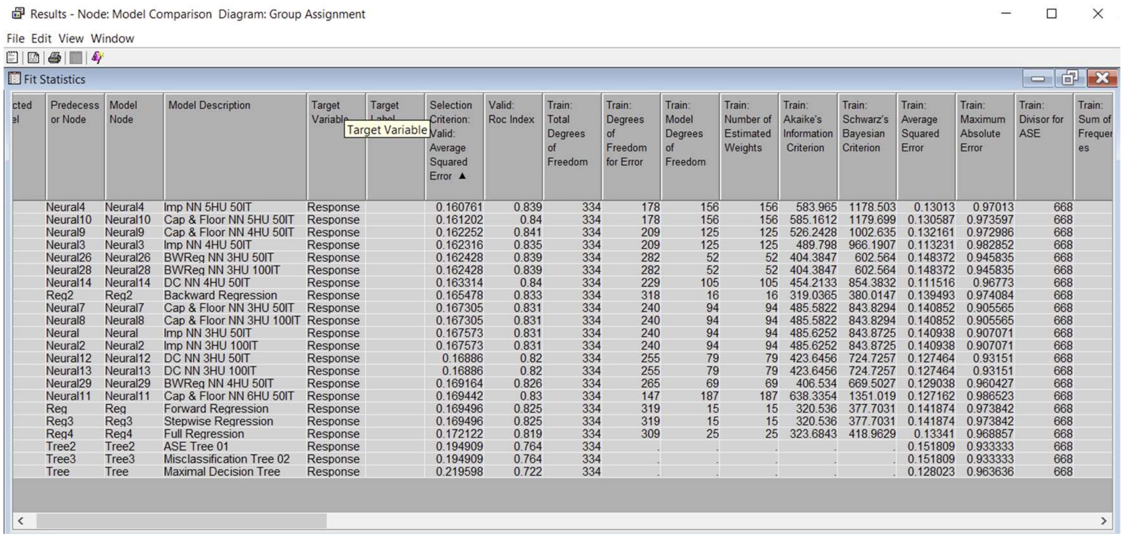


Fig 77 Final Diagram after including Model Comparison

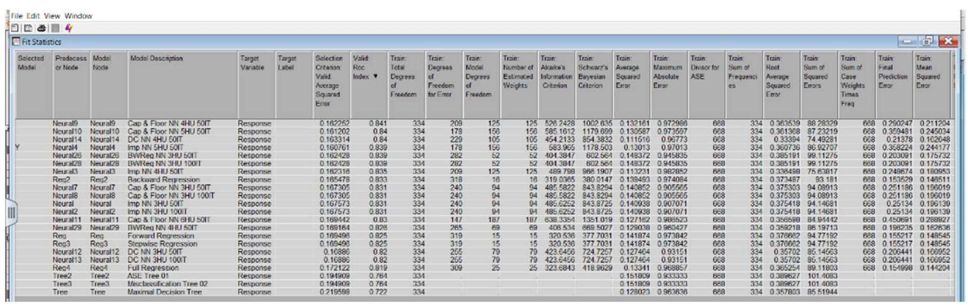
The Results of the Model comparison was observed as shown below:



cted el	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error ▲	Valid: Roc Index	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
	Neural4	Neural4	Imp NN 5HU 50IT	Response		0.160761	0.839	334	178	156	156	583.965	1178.503	0.13013	0.97013	668	
	Neural10	Neural10	Cap & Floor NN 5HU 50IT	Response		0.161202	0.84	334	178	156	156	585.1612	1179.699	0.130587	0.973597	668	
	Neural9	Neural9	Cap & Floor NN 4HU 50IT	Response		0.162252	0.841	334	209	125	125	526.2428	1002.635	0.132161	0.972986	668	
	Neural3	Neural3	Imp NN 4HU 50IT	Response		0.162316	0.835	334	209	125	125	489.798	966.1907	0.113231	0.982852	668	
	Neural26	Neural26	BWReg NN 3HU 50IT	Response		0.162428	0.839	334	282	52	52	404.3847	602.564	0.148372	0.945835	668	
	Neural28	Neural28	BWReg NN 3HU 100IT	Response		0.162428	0.839	334	282	52	52	404.3847	602.564	0.148372	0.945835	668	
	Neural14	Neural14	DC NN 4HU 50IT	Response		0.163314	0.84	334	229	105	105	454.2133	854.3832	0.111516	0.96773	668	
	Reg2	Reg2	Backward Regression	Response		0.165478	0.833	334	318	16	16	319.0365	380.0147	0.139493	0.974084	668	
	Neural7	Neural7	Cap & Floor NN 3HU 50IT	Response		0.167305	0.831	334	240	94	94	485.5822	843.8294	0.140852	0.905565	668	
	Neural8	Neural8	Cap & Floor NN 3HU 100IT	Response		0.167305	0.831	334	240	94	94	485.5822	843.8294	0.140852	0.905565	668	
	Neural	Neural	Imp NN 3HU 50IT	Response		0.167573	0.831	334	240	94	94	485.6252	843.8725	0.140938	0.907071	668	
	Neural2	Neural2	Imp NN 3HU 100IT	Response		0.167573	0.831	334	240	94	94	485.6252	843.8725	0.140938	0.907071	668	
	Neural12	Neural12	DC NN 3HU 50IT	Response		0.16886	0.82	334	255	79	79	423.6456	724.7257	0.127464	0.93151	668	
	Neural13	Neural13	DC NN 3HU 100IT	Response		0.16886	0.82	334	255	79	79	423.6456	724.7257	0.127464	0.93151	668	
	Neural29	Neural29	BWReg NN 4HU 50IT	Response		0.169164	0.826	334	265	69	69	406.534	669.5027	0.129038	0.960427	668	
	Neural11	Neural11	Cap & Floor NN 6HU 50IT	Response		0.169442	0.83	334	147	187	187	638.3354	1351.019	0.127162	0.986523	668	
	Reg	Reg	Forward Regression	Response		0.169496	0.825	334	319	15	15	320.536	377.7031	0.141874	0.973842	668	
	Reg3	Reg3	Stepwise Regression	Response		0.169496	0.825	334	319	15	15	320.536	377.7031	0.141874	0.973842	668	
	Reg4	Reg4	Full Regression	Response		0.172122	0.819	334	309	25	25	323.6843	418.9629	0.13341	0.968857	668	
	Tree2	Tree2	ASE Tree 01	Response		0.194909	0.764	334	.	.	.	.	.	0.151809	0.933333	668	
	Tree3	Tree3	Misclassification Tree 02	Response		0.194909	0.764	334	.	.	.	.	.	0.151809	0.933333	668	
	Tree	Tree	Maximal Decision Tree	Response		0.219598	0.722	334	.	.	.	.	.	0.128023	0.963636	668	

Figure 78 Model Comparison – Selection Criteria Valid: Average Squared Error

It can be observed that the neural network connected to the Imputation node, which ran with 5 hidden units and 50 iterations is the model with the lowest average squared error. The value of the ASE was observed to 0.160716. This model has an ROC index of 0.839, which is the 4th highest ROC index value among all the models.



Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error	Valid: Roc Index ▼	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: Sum of Squared Errors	Train: Sum of Case Weights Times Freq	Train: Final Prediction Error	Train: Mean Squared Error
	Neural9	Neural9	Cap & Floor NN 4HU 50IT	Response		0.162252	0.841	334	209	125	125	526.2428	1002.635	0.132161	0.972986	668	334	0.363539	88.28329	668	0.290247	0.211204
	Neural10	Neural10	Cap & Floor NN 5HU 50IT	Response		0.161202	0.84	334	178	156	156	585.1612	1179.699	0.130587	0.973597	668	334	0.361368	87.23219	668	0.359481	0.245034
	Neural14	Neural14	DC NN 4HU 50IT	Response		0.163314	0.84	334	229	105	105	454.2133	854.3832	0.111516	0.96773	668	334	0.33394	74.49281	668	0.21378	0.162648
Y	Neural4	Neural4	Imp NN 5HU 50IT	Response		0.160761	0.839	334	178	156	156	583.965	1178.503	0.13013	0.97013	668	334	0.360736	86.92707	668	0.358224	0.244177
	Neural26	Neural26	BWReg NN 3HU 50IT	Response		0.162428	0.839	334	282	52	52	404.3847	602.564	0.148372	0.945835	668	334	0.385191	99.11275	668	0.203091	0.175732
	Neural28	Neural28	BWReg NN 3HU 100IT	Response		0.162428	0.839	334	282	52	52	404.3847	602.564	0.148372	0.945835	668	334	0.385191	99.11275	668	0.203091	0.175732
	Neural3	Neural3	Imp NN 4HU 50IT	Response		0.162316	0.835	334	209	125	125	489.798	966.1907	0.113231	0.982852	668	334	0.336498	75.63817	668	0.248674	0.180953
	Reg2	Reg2	Backward Regression	Response		0.165478	0.833	334	318	16	16	319.0365	380.0147	0.139493	0.974084	668	334	0.373487	93.181	668	0.153529	0.146511
	Neural7	Neural7	Cap & Floor NN 3HU 50IT	Response		0.167305	0.831	334	240	94	94	485.5822	843.8294	0.140852	0.905565	668	334	0.375303	94.08913	668	0.251186	0.196019
	Neural8	Neural8	Cap & Floor NN 3HU 100IT	Response		0.167305	0.831	334	240	94	94	485.5822	843.8294	0.140852	0.905565	668	334	0.375303	94.08913	668	0.251186	0.196019
	Neural	Neural	Imp NN 3HU 50IT	Response		0.167573	0.831	334	240	94	94	485.6252	843.8725	0.140938	0.907071	668	334	0.375418	94.14681	668	0.25134	0.196139
	Neural2	Neural2	Imp NN 3HU 100IT	Response		0.167573	0.831	334	240	94	94	485.6252	843.8725	0.140938	0.907071	668	334	0.375418	94.14681	668	0.25134	0.196139
	Neural11	Neural11	Cap & Floor NN 6HU 50IT	Response		0.169442	0.83	334	147	187	187	638.3354	1351.019	0.127162	0.986523	668	334	0.356598	84.94442	668	0.450691	0.288927
	Neural29	Neural29	BWReg NN 4HU 50IT	Response		0.169164	0.826	334	265	69	69	406.534	669.5027	0.129038	0.960427	668	334	0.359218	86.19713	668	0.198235	0.162638
	Reg	Reg	Forward Regression	Response		0.169496	0.825	334	319	15	15	320.536	377.7031	0.141874	0.973842	668	334	0.376662	94.77192	668	0.155217	0.148545
	Reg3	Reg3	Stepwise Regression	Response		0.169496	0.825	334	319	15	15	320.536	377.7031	0.141874	0.973842	668	334	0.376662	94.77192	668	0.155217	0.148545
	Neural12	Neural12	DC NN 3HU 50IT	Response		0.16886	0.82	334	255	79	79	423.6456	724.7257	0.127464	0.93151	668	334	0.35702	85.14563	668	0.206441	0.166952
	Neural13	Neural13	DC NN 3HU 100IT	Response		0.16886	0.82	334	255	79	79	423.6456	724.7257	0.127464	0.93151	668	334	0.35702	85.14563	668	0.206441	0.166952
	Reg4	Reg4	Full Regression	Response		0.172122	0.819	334	309	25	25	323.6843	418.9629	0.13341	0.968857	668	334	0.365254	89.11803	668	0.154998	0.144204
	Tree2	Tree2	ASE Tree 01	Response		0.194909	0.764	334	.	.	.	.	.	0.151809	0.933333	668	334	0.389627	101.4083	.	.	.
	Tree3	Tree3	Misclassification Tree 02	Response		0.194909	0.764	334	.	.	.	.	.	0.151809	0.933333	668	334	0.389627	101.4083	.	.	.
	Tree	Tree	Maximal Decision Tree	Response		0.219598	0.722	334	.	.	.	.	.	0.128023	0.963636	668	334	0.357803	85.51944	.	.	.

Figure 79 Model Comparison – Validation: ROC Index

From the same Fit Statistics Window, we can see that once we sort the ROC index in ascending order, the neural network connected to the Cap & Floor node, with 4 hidden units and 50 iterations, had the largest ROC index value of 0.841. This model however, had an average squared error of 0.162252. This ASE is the third lowest of all models.

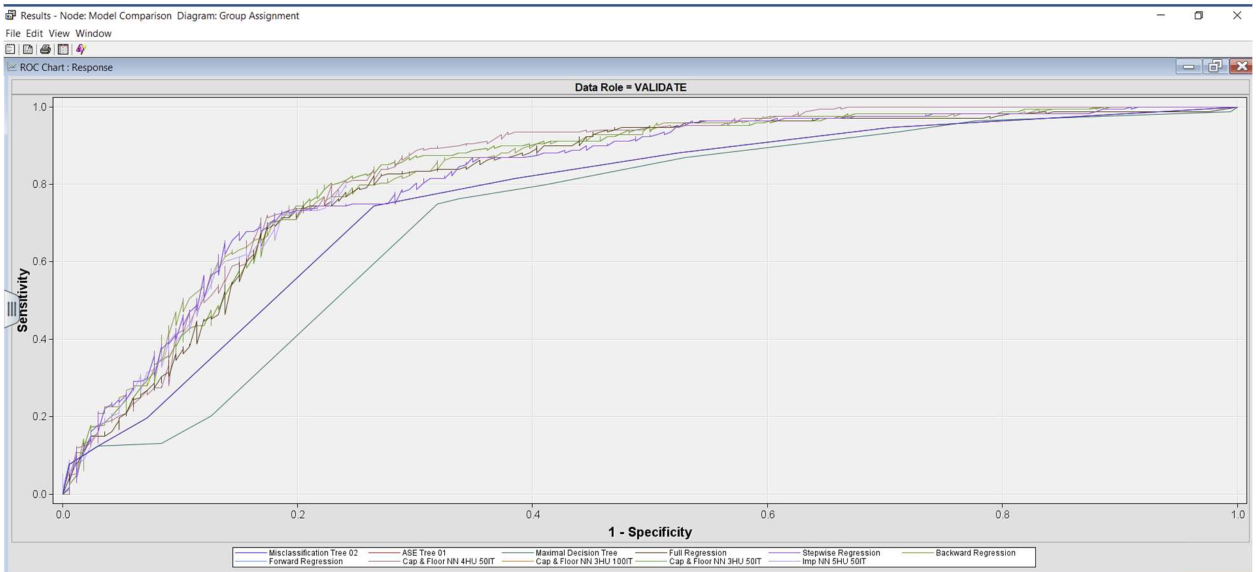


Figure 80 Model Comparison – Validation ROC Curve

The ROC curves of the different models can be seen in the above graph. We already know from the ROC index, that the neural network, with 4 hidden units and 50 iterations, connected to the capping and flooring node, has the highest ROC index.

These results highlight the efficiency of this specific neural network configuration in reducing prediction errors. The model's rapid convergence after a relatively low number of iterations underscores its effectiveness. This outcome accentuates the importance of meticulous configuration selection in neural network modeling, as even slight adjustments can result in significant enhancements in prediction accuracy.

The visual depiction of the ROC curve illustrates the balance between true positive and false positive rates at different threshold levels. The evident consistency between the previously determined optimal network and its validation phase performance reaffirms the robustness of our model selection approach and its ability to effectively generalize to unfamiliar data.

In summary, we have a range of models presenting varying ASE values and misclassification rates. There is no single model that stands out definitively from the rest. Therefore, it becomes essential to meticulously compare the ASE values and misclassification rates of a selection of the top-performing models to determine the most suitable choice that would best cater to our clients' needs.

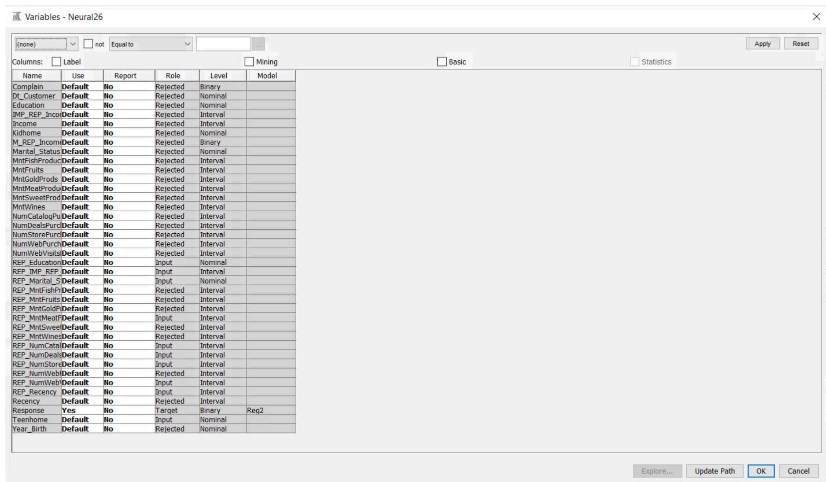
Data Role=Valid

Statistics	Neural4	Neural10	Neural9	Neural3	Neural26	Neural28	Neural14	Reg2	Neural7	Neural8	Neural	Neural2	Neural12	Neural13	Neural29	Neural11	Reg	Reg3	Reg4	Tree2	Tree3	Tree
Valid: Kolmogorov-Smirnov Statistic	0.586	0.574	0.586	0.604	0.550	0.550	0.587	0.545	0.580	0.580	0.593	0.593	0.545	0.545	0.556	0.563	0.545	0.545	0.556	0.479	0.479	0.431
Valid: Average Squared Error	0.161	0.161	0.162	0.162	0.162	0.162	0.163	0.165	0.167	0.167	0.168	0.168	0.169	0.169	0.169	0.169	0.169	0.169	0.172	0.195	0.195	0.220
Valid: Roc Index	0.839	0.840	0.841	0.835	0.839	0.839	0.840	0.833	0.831	0.831	0.831	0.831	0.820	0.820	0.826	0.830	0.825	0.825	0.819	0.764	0.764	0.722
Valid: Average Error Function	0.494	0.493	0.512	0.527	0.496	0.496	0.512	0.520	0.517	0.517	0.518	0.518	0.523	0.523	0.535	0.541	0.532	0.532	0.599	.	.	.
Valid: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff	0.510	0.502	0.387	0.501	0.551	0.551	0.667	0.563	0.528	0.528	0.535	0.535	0.632	0.632	0.433	0.607	0.604	0.604	0.495	0.784	0.784	0.725
Valid: Cumulative Percent Captured Response	17.262	17.262	17.262	16.667	17.857	17.857	17.262	17.262	17.857	17.857	17.262	17.262	16.071	16.071	16.071	17.262	17.262	17.262	16.071	15.508	15.508	12.976
Valid: Percent Captured Response	8.333	8.333	8.333	7.738	8.333	8.333	7.143	8.333	9.524	9.524	8.929	8.929	8.333	8.333	7.738	8.333	8.333	8.333	7.738	6.579	6.579	4.048
Valid: Divisor for VASE	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000
Valid: Error Function	329.988	329.372	341.884	351.828	331.547	331.547	342.175	347.512	345.397	345.397	346.028	346.028	349.345	349.345	357.487	361.151	355.497	355.497	399.837	.	.	.
Valid: Gain	69.573	69.573	69.573	63.725	75.420	75.420	69.573	69.573	75.420	75.420	69.573	69.573	57.878	57.878	57.878	69.573	69.573	69.573	57.878	52.339	52.339	27.472
Valid: Gini Coefficient	0.678	0.680	0.682	0.671	0.677	0.677	0.680	0.667	0.661	0.661	0.661	0.661	0.641	0.641	0.652	0.661	0.651	0.651	0.639	0.528	0.528	0.445
Valid: Bin-Based Two-Way Kolmogorov-Smirnov Statistic	0.556	0.556	0.586	0.592	0.544	0.544	0.563	0.539	0.568	0.568	0.568	0.568	0.528	0.528	0.550	0.563	0.540	0.540	0.556	0.473	0.473	0.425
Valid: Kolmogorov-Smirnov Probability Cutoff	0.498	0.488	0.372	0.460	0.518	0.518	0.535	0.551	0.488	0.488	0.532	0.532	0.531	0.531	0.434	0.575	0.566	0.566	0.475	0.291	0.291	0.667
Valid: Cumulative Lift	1.696	1.696	1.696	1.637	1.754	1.754	1.696	1.696	1.754	1.754	1.696	1.696	1.579	1.579	1.579	1.696	1.696	1.696	1.579	1.523	1.523	1.275
Valid: Lift	1.637	1.637	1.637	1.520	1.637	1.637	1.403	1.637	1.871	1.871	1.754	1.754	1.637	1.637	1.520	1.637	1.637	1.637	1.520	1.293	1.293	0.795
Valid: Maximum Absolute Error	0.970	0.961	0.981	0.995	0.967	0.967	0.980	0.997	0.928	0.928	0.929	0.929	0.977	0.977	0.994	0.987	0.998	0.998	1.000	1.000	1.000	1.000
Valid: Misclassification Rate	0.207	0.222	0.240	0.210	0.234	0.234	0.210	0.237	0.219	0.219	0.213	0.213	0.240	0.240	0.246	0.231	0.249	0.249	0.234	0.260	0.260	0.305
Valid: Mean Squared Error	0.161	0.161	0.162	0.162	0.162	0.162	0.163	0.165	0.167	0.167	0.168	0.168	0.169	0.169	0.169	0.169	0.169	0.169	0.172	.	.	.
Valid: Sum of Frequencies	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000	334.000
Valid: Root Average Squared Error	0.401	0.401	0.403	0.403	0.403	0.403	0.404	0.407	0.409	0.409	0.409	0.409	0.411	0.411	0.411	0.412	0.412	0.412	0.415	0.441	0.441	0.469
Valid: Cumulative Percent Response	85.294	85.294	85.294	82.353	88.235	88.235	85.294	85.294	88.235	88.235	85.294	85.294	79.412	79.412	79.412	85.294	85.294	85.294	79.412	76.625	76.625	64.118
Valid: Percent Response	82.353	82.353	82.353	76.471	82.353	82.353	70.588	82.353	94.118	94.118	88.235	88.235	82.353	82.353	76.471	82.353	82.353	82.353	76.471	65.015	65.015	40.000
Valid: Root Mean Squared Error	0.401	0.401	0.403	0.403	0.403	0.403	0.404	0.407	0.409	0.409	0.409	0.409	0.411	0.411	0.411	0.412	0.412	0.412	0.415	.	.	.
Valid: Sum of Squared Errors	107.388	107.683	108.384	108.427	108.502	108.502	109.094	110.539	111.760	111.760	111.939	111.939	112.799	112.799	113.002	113.187	113.223	113.223	114.978	130.199	130.199	146.691
Valid: Sum of Case Weights Times Freq	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	668.000	.	.	.
Valid: Number of Wrong Classifications	69.000	74.000	80.000	70.000	78.000	78.000	70.000	.	73.000	73.000	71.000	71.000	80.000	80.000	82.000	77.000	.	.	.	.	.	.

Figure 81 Model Comparison – ROC Index

Conclusion

The neural network model, integrated with backward regression, was executed using 3 hidden units and 50 iterations, yielding an average squared error of 0.162428 and a misclassification rate of 0.839. This model emerges as a standout contender among the compared models, excelling in terms of both minimal ASE and elevated ROC values. Additionally, its linkage to a backward regression enhances its interpretability, rendering it a more appealing recommendation for our clients. With these factors in mind, we conclude that this model stands as the most suitable choice and hence, we opt for its selection as the final model.



Name	Use	Report	Role	Level	Model
Complain	Default	No	Rejected	Binary	
Dt_Customer	Default	No	Rejected	Nominal	
Education	Default	No	Rejected	Nominal	
IMP_REP_Incor	Default	No	Rejected	Interval	
Income	Default	No	Rejected	Interval	
Kidhome	Default	No	Rejected	Nominal	
M_REP_Income	Default	No	Rejected	Binary	
Marital_Status	Default	No	Rejected	Nominal	
MntFishProduc	Default	No	Rejected	Interval	
MntFruits	Default	No	Rejected	Interval	
MntGoldProds	Default	No	Rejected	Interval	
MntMeatProdu	Default	No	Rejected	Interval	
MntSweetProd	Default	No	Rejected	Interval	
MntWines	Default	No	Rejected	Interval	
NumCatalogPu	Default	No	Rejected	Interval	
NumDealsPurc	Default	No	Rejected	Interval	
NumStorePurc	Default	No	Rejected	Interval	
NumWebPurch	Default	No	Rejected	Interval	
NumWebVisits	Default	No	Rejected	Interval	
REP_Education	Default	No	Input	Nominal	
REP_IMP_REP_	Default	No	Input	Interval	
REP_Marital_S	Default	No	Input	Nominal	
REP_MntFishPr	Default	No	Rejected	Interval	
REP_MntFruits	Default	No	Rejected	Interval	
REP_MntGoldP	Default	No	Rejected	Interval	
REP_MntMeatF	Default	No	Input	Interval	
REP_MntSweet	Default	No	Rejected	Interval	
REP_MntWines	Default	No	Rejected	Interval	
REP_NumCatal	Default	No	Input	Interval	
REP_NumDeals	Default	No	Input	Interval	
REP_NumStore	Default	No	Input	Interval	
REP_NumWebI	Default	No	Rejected	Interval	
REP_NumWebV	Default	No	Input	Interval	
REP_Recency	Default	No	Input	Interval	
Recency	Default	No	Rejected	Interval	
Response	Yes	No	Target	Binary	Reg2
Teenhome	Default	No	Input	Nominal	
Year_Birth	Default	No	Rejected	Nominal	

We also know, from the results of the backward regression that we ran a few steps prior, the variables that came out as significant were REP_Education, REP_IMP_REP_Income, REP_Marital_Status, REP_MntMeatProducts, REP_MntCatalogPurchases, REP_NumDealsPurchases, REP_NumStorePurchases, REP_NumWebVisitsMonth, REP_Recency, Teenhouse. These variables show significance in the predictive model that we created.

```
The selected model, based on the error rate for the validation data, is the model trained in Step 8. It consists of the following effects:

Intercept  REP_Education  REP_IMP_REP_Income  REP_Marital_Status  REP_MntMeatProducts  REP_NumCatalogPurchases  REP_NumDealsPurchases  REP_NumStorePurchases  REP_NumWebVisitsMonth  REP_Recency  Teenhome
```

Conversely, the model has judiciously excluded other variables, suggesting their diminished influence or redundancy in forecasting the target outcome.

This holds paramount importance, showcasing the model's adeptness at utilizing crucial information while excluding less influential or possibly misleading factors. A thorough examination of both the ROC index and ASE ensures a well-rounded evaluation of the model's performance, thus validating the credibility of this selection. Through the adoption of a forward regression approach with the specified hidden units and iterations, the model demonstrates its prowess in distilling valuable insights from intricate data, thereby paving the path for heightened predictive precision and informed decision-making.

Recommendations

We gained important insights that can considerably influence the superstore's strategy for its impending year-end sale in light of the extensive investigation carried out to satisfy our project objectives. These suggestions, which were developed after carefully analysing historical campaign data, are intended to improve customer targeting and raise the profile of Gold Memberships. The superstore may strengthen its competitive edge, make the best use of resources, and increase Gold Membership sales during the year-end sale event by integrating these information with your marketing initiatives. These are the main suggestions to think about:

1. **Leverage Education and Income**: Given that education and income have a substantial impact on gold membership purchases, we recommend tailoring marketing efforts to individuals with higher education levels and income brackets. This could involve crafting messages that resonate with their preferences and aspirations.
2. **Consider Marital Status**: Consider segmenting your target audience based on marital status. Divorced and single individuals exhibit a higher likelihood of purchasing gold memberships, while married and coupled individuals show different responses. Tailoring your communication to these diverse segments could yield more personalized and effective outreach.
3. **Promote High-Value Products**: As an increase in meat and catalog purchases influences gold membership purchases, focus on promoting these products to potential customers. Offering incentives related to these high-value categories could encourage more positive responses to the membership offer.
4. **Enhance Website Engagement**: The positive correlation between website visits and gold membership purchases highlights the importance of an engaging online presence. Invest in user-friendly interfaces, interactive content, and convenient online shopping experiences to further boost customer interest.
5. **Tailor Recency Strategies**: The inverse relationship between recency and gold membership purchases suggests that re-engaging customers who have not interacted recently could be beneficial. Craft targeted campaigns aimed at rekindling interest and driving more recent interactions.
6. **Target Teen-Related Households**: Households with no teens or only one teen exhibit a higher likelihood of purchasing gold memberships. Tailor your outreach to these demographics, considering their specific preferences and needs when promoting the membership offer.

APPENDIX

1. Data Source: Kaggle

https://www.kaggle.com/datasets/ahsan81/superstore-marketing-campaign-dataset