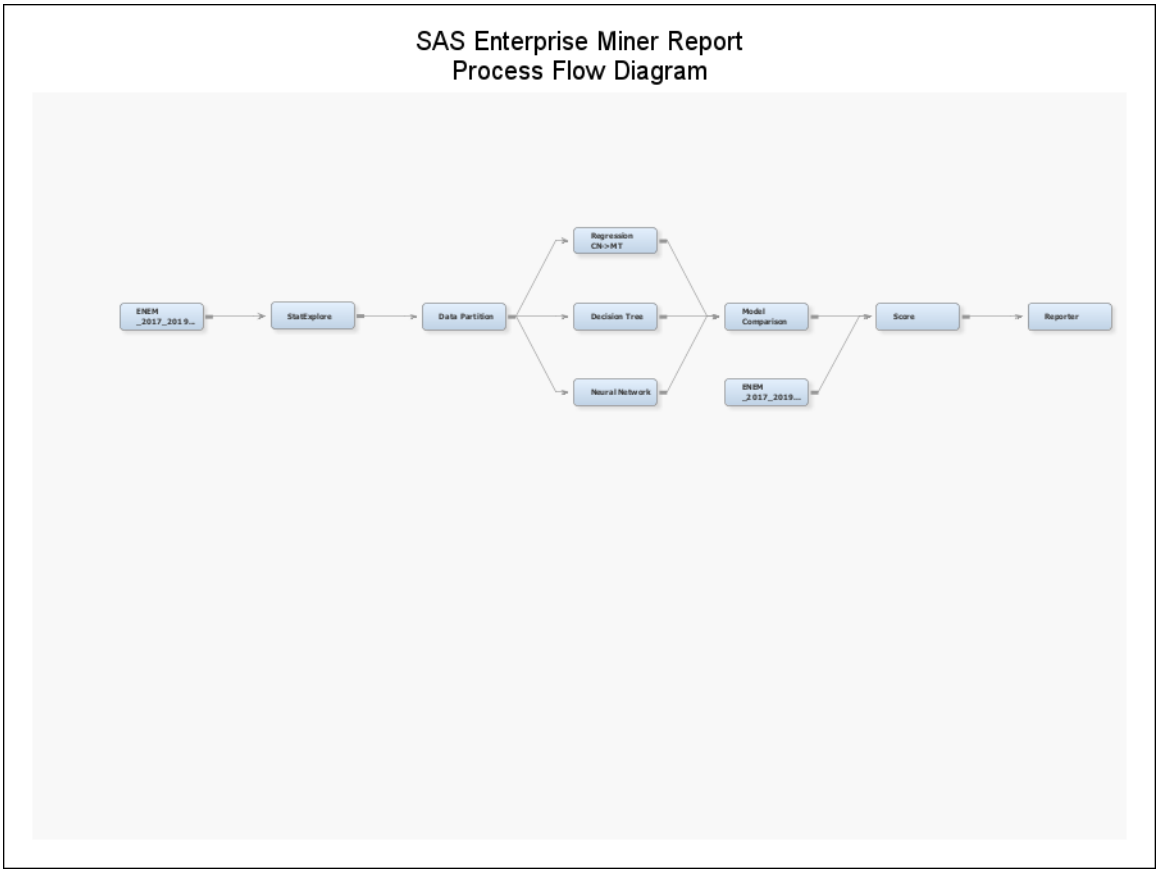
SAS Enterprise Miner Report

User = Leandro
Date = 12:24:18 September 03
Project = Predição da nota de matemática
Diagram = Sexo candidato

Start Node = Report
Node label = Reporter
Nodes = ALL
Showall = Y

Format = PDF
Style = JOURNAL



SAS Enterprise Miner Report

Node=ENEM_2017_2019_AMOSTRA_10 Summary

Node id = Ids
Node label = ENEM_2017_2019_AMOSTRA_10
Meta path = Ids
Notes =

Leitura dos dados ENEM 2017-2019

Node=ENEM_2017_2019_AMOSTRA_10 Properties

Property	Value	Default	Property	Value	Default
Component	DataSource		DsCreatedBy	Leandro	
ApplyIntervalLevelLowerLimit	Y		DsId	enemamostra	
ApplyMaxClassLevels	Y		DsModifiedBy	Leandro	
ApplyMaxPercentMissing	Y		DsModifyDate	1914661471.8	
CMeta	WORK.M02SHSKA		DsSampleName		
ComputeStatistics	N		DsSampleSize		
DBPassThrough	Y		DsSampleSizeType		
Data	ENEM.ENEM_2017_2019_AMOSTRA_10		DsScope	LOCAL	
DataSelection	DATASOURCE		IdentifyEmptyColumns	Y	
DataSource	enemamostra		IntervalLowerLimit	20	
DataSourceRole	RAW		Library	ENEM	
Description			MaxClassLevels	20	
DropMapVariables	Y		MaxPercentMissing	50	
DsCreateDate	1914661471.8		MetaAdvisor	BASIC	

Property	Value	Default
NBytes	866989056	.
NCols	145	.
NObs	1020953	.
NewTable		
NewVariableRole	REJECT	
OutputType	DATA	VIEW
Role	RAW	TRAIN
Sample	D	
SampleSizeObs	10000	
SampleSizePercent	20	
SampleSizeType	PERCENT	
Scope	LOCAL	
Segment		
Table	ENEM_2017_2019_AMOSTRA_10	

Node=ENEM_2017_2019_AMOSTRA_10 Data Attributes

Attribute	Value	Attribute	Value	Attribute	Value
Data Name	ENEM_2017_2019_AMOSTRA_10	Date Created	01Jul2020:12:38:33	Data Size	866989056
Data Type	DATA	Date Modified	01Jul2020:12:38:33	Role	RAW
Data Label		Number Rows	1020953	Segment	
Engine	BASE	Number Columns	145	Data Library	ENEM

Node=ENEM_2017_2019_AMOSTRA_10
Variable Summary

Role	Level	Frequency Count	Name
TARGET	NOMINAL	1	TP_SEXO
REJECTED	INTERVAL	27	ActualProportion AllocProportion CO_ESCOLA CO_MUNICIPIO_ESC CO_MUNICIPIO_NASCIMENTO CO_MUNICIPIO_PROVA CO_MUNICIPIO_RESIDENCIA CO_UF_ESC CO_UF_NASCIMENTO CO_UF_PROVA CO_UF_RESIDENCIA NOTA_MEDIA ...
REJECTED	NOMINAL	114	CO_PROVA_CH CO_PROVA_CN CO_PROVA_LC CO_PROVA_MT IN_ACESSO IN_AMPLIADA_18 IN_AMPLIADA_24 IN_APOIO_PERNA IN_AUTISMO IN_BAIXA_VISAO IN_BRAILLE IN_CADEIRA_ACOLCHOADA IN_CADEIRA_CANHOTO ... IN_CEGUEIRA
INPUT	INTERVAL	2	NU_NOTA_CH NU_NOTA_MT
INPUT	NOMINAL	1	TP_COR_RACA

SAS Enterprise Miner Report

Node=StatExplore
Summary

Node id = Stat
Node label = StatExplore
Meta path = Ids => Stat
Notes =

Node=StatExplore
Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	StatExplore		Correlation	Y		NObs	ALL	1000000
BySegment	N	Y	DropRejected	Y		Pearson	Y	
ChiSquare	Y		HideVariable	Y		Spearman	N	
ChiSquareInterval	N		IntervalDistribution	Y		UseScore	N	
ChiSquareIntervalNBins	5		LevelSummary	Y		UseTest	N	
ClassDistribution	Y		MaximumVars	1000		UseValidate	N	

Node=StatExplore
Variable Summary

Role	Level	Frequency Count	Name
INPUT	INTERVAL	2	NU_NOTA_CH NU_NOTA_MT
INPUT	NOMINAL	1	TP_COR_RACA

Node=StatExplore
Interval Variables

Data Role	Target	Target Level	Variable	Median	Missing	Non Missing	Minimum	Maximum	Mean	Standard Deviation	Skewness
TRAIN	TP_SEXO	F	NU_NOTA_MT	485.2	0	598399	0	993.9	489.8811	134.4182	-1.18946
TRAIN	TP_SEXO	M	NU_NOTA_MT	531.4	0	422554	0	994.4	530.8631	150.4993	-1.15881
TRAIN	TP_SEXO	F	NU_NOTA_CH	527.9	0	598399	0	828.3	510.75	127.5515	-2.13221
TRAIN	TP_SEXO	M	NU_NOTA_CH	548.5	0	422554	0	844.7	525.2237	133.0162	-2.16306

Kurtosis	Role	Label	Scaled Mean Deviation	Maximum Deviation	Level Id
4.318448	INPUT	NU_NOTA_MT	-0.033465	0.047392	1
3.432861	INPUT	NU_NOTA_MT	0.047392	0.047392	2
6.565656	INPUT	NU_NOTA_CH	-0.011593	0.016417	1
6.382119	INPUT	NU_NOTA_CH	0.016417	0.016417	2

Node=StatExplore
Class Variables

Data Role	Target	Target Level	Variable Name	Level	CODE	Frequency Count	Type	Percent Within	Level Index	Role	Label	Percent	Plot
TRAIN	TP_SEXO	F	TP_COR_RACA	0	5	10320	C	1.7246	1	INPUT	TP_COR_RACA	0.01011	1
TRAIN	TP_SEXO	M	TP_COR_RACA	0	5	8828	C	2.0892	1	INPUT	TP_COR_RACA	0.00865	1
TRAIN	TP_SEXO	F	TP_COR_RACA	1	1	210177	C	35.1232	2	INPUT	TP_COR_RACA	0.20586	1
TRAIN	TP_SEXO	M	TP_COR_RACA	1	1	154483	C	36.5594	2	INPUT	TP_COR_RACA	0.15131	1
TRAIN	TP_SEXO	F	TP_COR_RACA	2	3	75354	C	12.5926	3	INPUT	TP_COR_RACA	0.07381	1
TRAIN	TP_SEXO	M	TP_COR_RACA	2	2	57275	C	13.5545	3	INPUT	TP_COR_RACA	0.05610	1
TRAIN	TP_SEXO	F	TP_COR_RACA	3	0	284036	C	47.4660	4	INPUT	TP_COR_RACA	0.27821	1
TRAIN	TP_SEXO	M	TP_COR_RACA	3	0	191649	C	45.3549	4	INPUT	TP_COR_RACA	0.18772	1
TRAIN	TP_SEXO	F	TP_COR_RACA	4	2	15217	C	2.5430	5	INPUT	TP_COR_RACA	0.01490	1
TRAIN	TP_SEXO	M	TP_COR_RACA	4	3	7752	C	1.8346	5	INPUT	TP_COR_RACA	0.00759	1
TRAIN	TP_SEXO	F	TP_COR_RACA	5	4	3295	C	0.5506	6	INPUT	TP_COR_RACA	0.00323	1
TRAIN	TP_SEXO	M	TP_COR_RACA	5	4	2567	C	0.6075	6	INPUT	TP_COR_RACA	0.00251	1

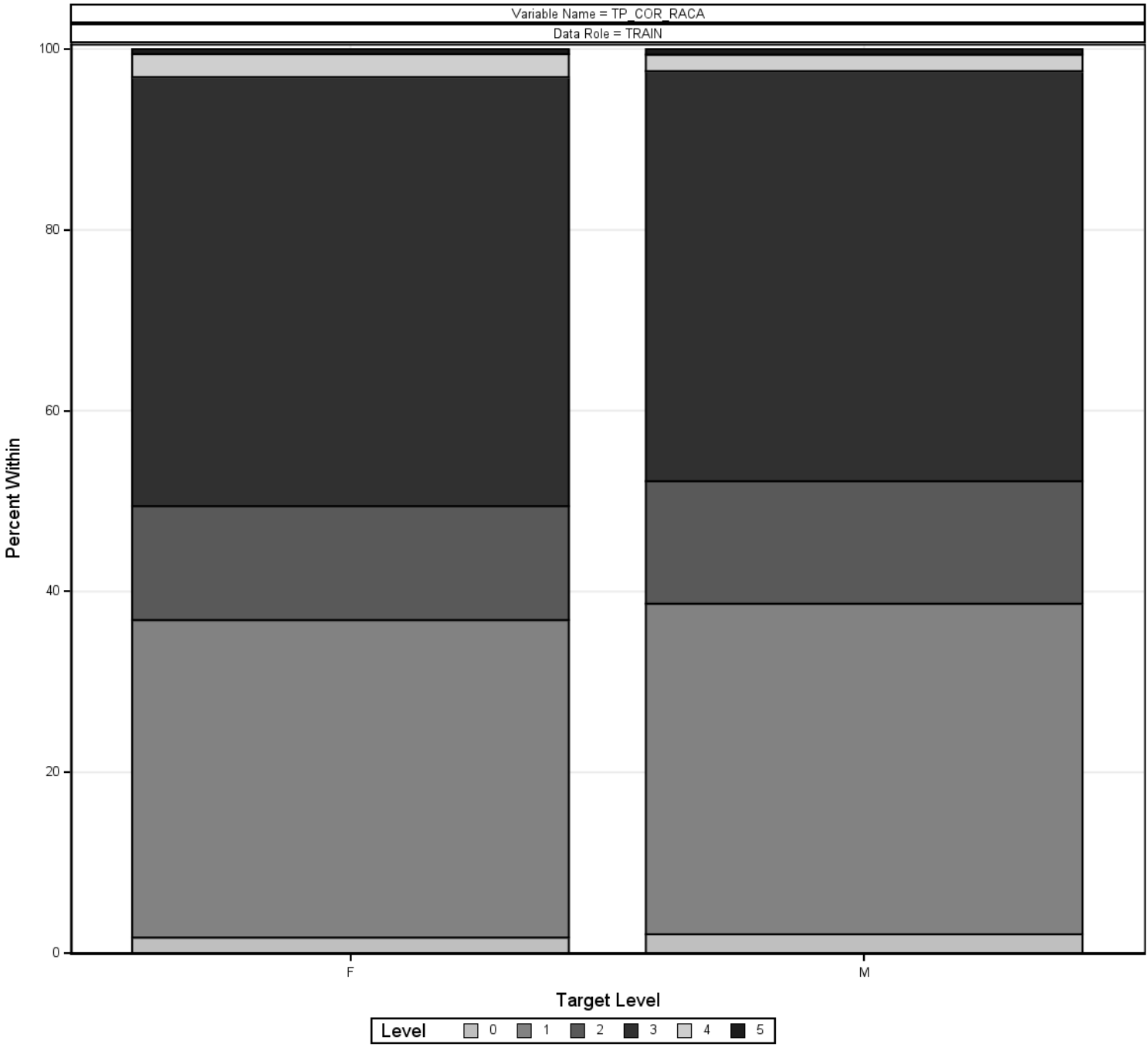
Node=StatExplore
Cell Chi-Squares

Data Role	Segment	Segment Id	Segment Name:Value	Target	Input	Target: Formatted Value	Input: Formatted Value	Frequency Count	Target: Numeric Value
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	F	0	10320	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	F	1	210177	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	F	2	75354	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	F	3	284036	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	F	4	15217	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	F	5	3295	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	M	0	8828	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	M	1	154483	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	M	2	57275	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	M	3	191649	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	M	4	7752	.
TRAIN			_OVERALL_	TP_SEXO	TP_COR_RACA	M	5	2567	.

Input: Numeric Value	Chi-Square	Log Chi Square	Role	Label
.	72.653	4.28570	INPUT	TP_COR_RACA
.	59.190	4.08075	INPUT	TP_COR_RACA
.	73.005	4.29053	INPUT	TP_COR_RACA
.	98.048	4.58545	INPUT	TP_COR_RACA
.	228.642	5.43216	INPUT	TP_COR_RACA
.	5.772	1.75301	INPUT	TP_COR_RACA
.	102.888	4.63364	INPUT	TP_COR_RACA
.	83.822	4.42870	INPUT	TP_COR_RACA
.	103.386	4.63847	INPUT	TP_COR_RACA
.	138.850	4.93339	INPUT	TP_COR_RACA
.	323.792	5.78010	INPUT	TP_COR_RACA
.	8.174	2.10095	INPUT	TP_COR_RACA

Target	Variable	Importance	Worth	Analysis Variable	Label	plot
TP_SEXO	NU_NOTA_MT	1	0.019489	1	NU_NOTA_MT	.
TP_SEXO	NU_NOTA_CH	2	0.006758	1	NU_NOTA_CH	.
TP_SEXO	TP_COR_RACA	3	0.000811	1	TP_COR_RACA	.

SAS Enterprise Miner Report
Node=StatExplore
Class Variables
TARGET='TP_SEXO' and PLOT=1



SAS Enterprise Miner Report

Node=Data Partition Summary

Node id = Part
Node label = Data Partition
Meta path = Ids => Stat => Part
Notes =

Node=Data Partition Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	Partition		Method	RANDOM	DEFAULT	TestPct	30	
ClassDistribution	Y		OutputType	DATA		TrainPct	60	40
IntervalDistribution	Y		RandomSeed	12345		ValidatePct	10	30

Node=Data Partition Variable Summary

Role	Level	Frequency Count	Name
TARGET	NOMINAL	1	TP_SEXO
INPUT	INTERVAL	2	NU_NOTA_CH NU_NOTA_MT
INPUT	NOMINAL	1	TP_COR_RACA

Node=Data Partition Class Variables

Data	Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
DATA	TP_SEXO	.	F	598399	58.6118	
DATA	TP_SEXO	.	M	422554	41.3882	
TEST	TP_SEXO	.	F	179507	58.6076	
TEST	TP_SEXO	.	M	126779	41.3924	
TRAIN	TP_SEXO	.	F	358971	58.6006	
TRAIN	TP_SEXO	.	M	253601	41.3994	
VALIDATE	TP_SEXO	.	F	59921	58.6914	
VALIDATE	TP_SEXO	.	M	42174	41.3086	

SAS Enterprise Miner Report

Node=Decision Tree Summary

Node id = Tree
Node label = Decision Tree
Meta path = Ids => Stat => Part => Tree
Notes =

Node=Decision Tree Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	DecisionTree		Kass	Y		Pred	N	
AVG	Y		KassApply	BEFORE		Predict	Y	
AssessMeasure	PROFIT/LOSS		LeafSize	5		ProfitLoss	NONE	
AssessPercentage	0.25		Leafid	Y		RASE	N	
CV	N		Maxbranch	2		SampleMethod	RANDOM	
CVNIter	10		Maxdepth	6		SampleSeed	12345	
CVRepeat	1		MinCatSize	5		SampleSize	10000	
CVSeed	12345		MissingValue	USEINSEARCH		ShowNodeId	Y	
ClassColorBy	PERCENTCORRECT		NSubtree	1		ShowValid	Y	
Count	Y		NodeRole	SEGMENT		SigLevel	0.2	
CreateSample	DEFAULT		NodeSample	20000		SplitPrecision	4	
Criterion	DEFAULT		NominalCriterion	PROBCHISQ		Splitsize	.	
Depth	Y		Nrules	5		Subtree	ASSESSMENT	
Dummy	N		Nsurrs	0		Target	ALL	
Exhaustive	5000		NumInputs	1		ToolType	MODEL	
Freeze	N		NumSingleImp	5		TrainMode	BATCH	
ImportModel	N		ObsImportance	N		UseDecision	N	
ImportedTreeData			OrdinalCriterion	ENTROPY		UseMultipleTarget	N	
Inputs	N		PercentCorrect	N		UsePriors	N	
IntColorBy	AVG		Performance	DISK		UseVarOnce	N	
IntervalCriterion	PROBF		Precision	4		VarSelection	Y	

Node=Decision Tree Variable Summary

Role	Level	Frequency Count	Name
TARGET	NOMINAL	1	TP_SEXO
INPUT	INTERVAL	2	NU_NOTA_CH NU_NOTA_MT
INPUT	NOMINAL	1	TP_COR_RACA
ID	INTERVAL	1	_dataobs_

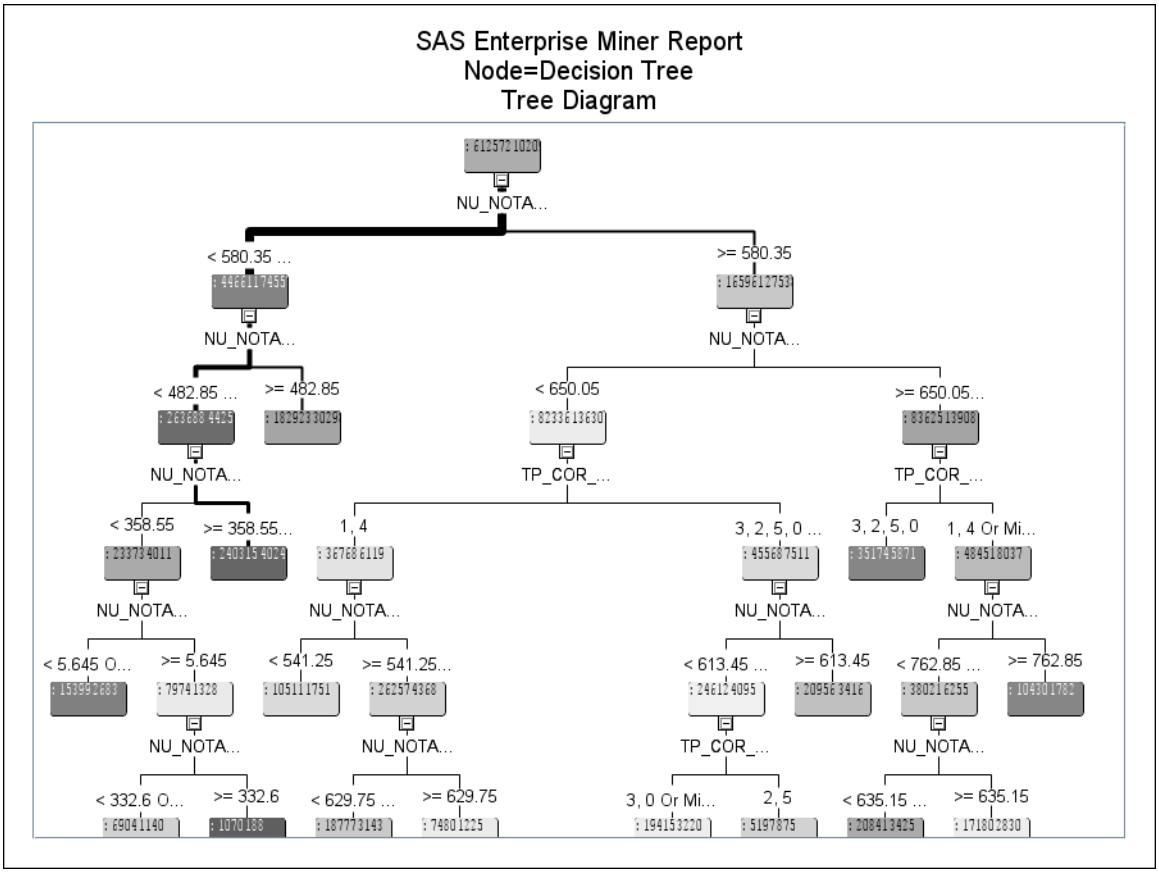
Node=Decision Tree Model Fit Statistics

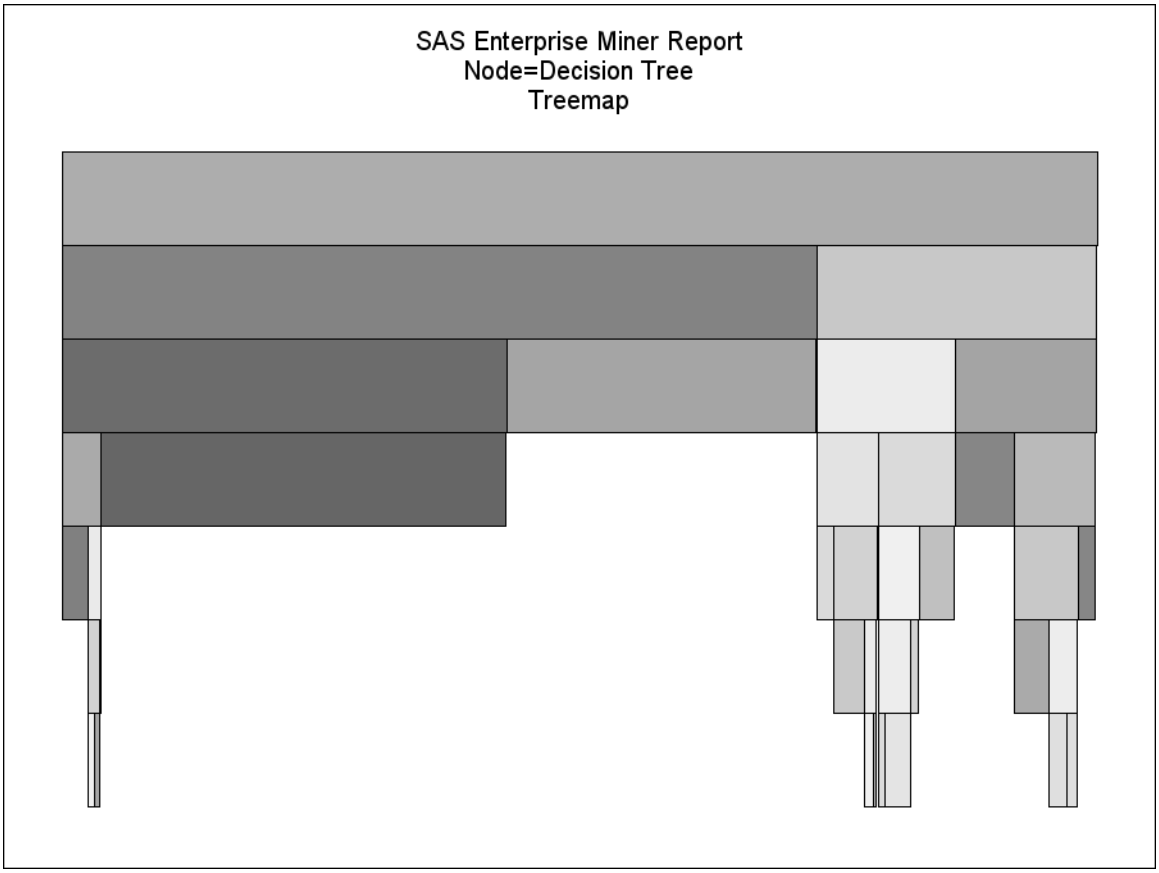
Target=TP_SEXO Target Label=' '

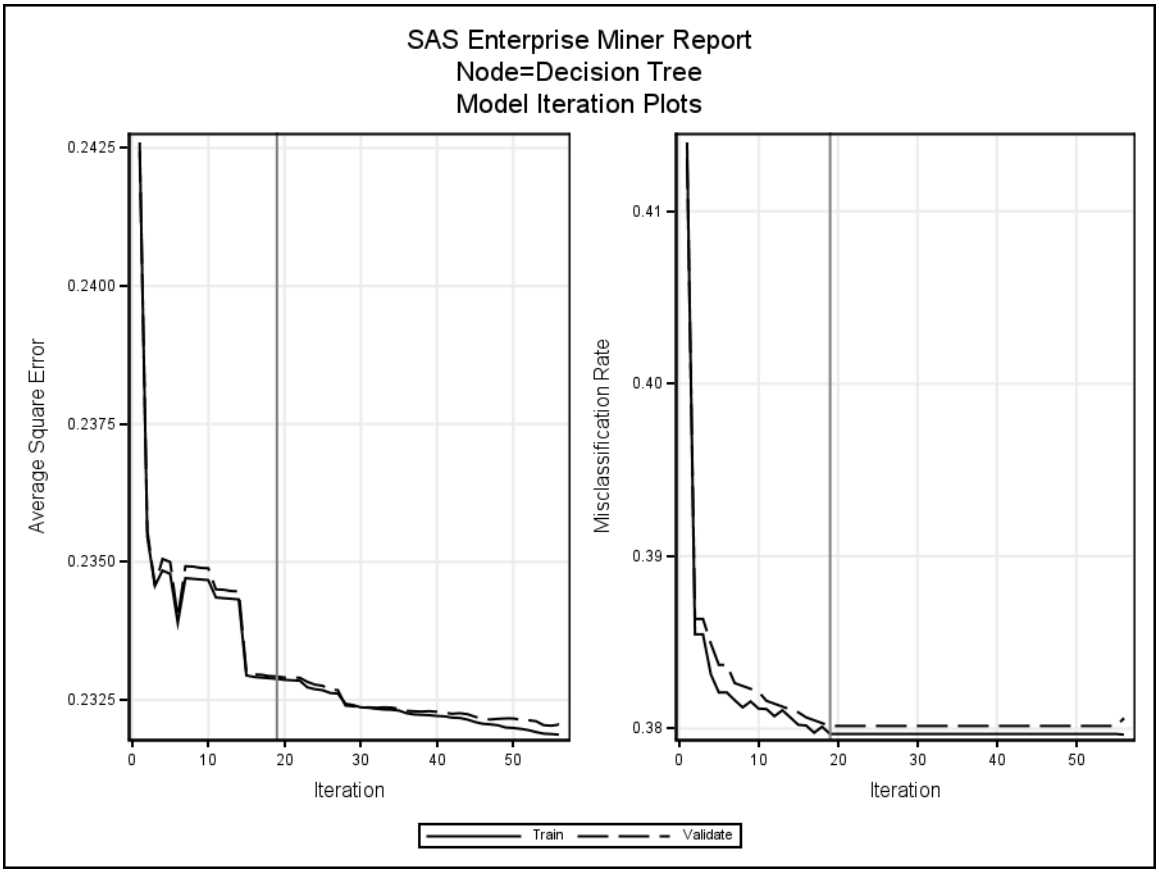
Label of Statistic	Train	Validation	Test
Sum of Frequencies	612572.00	102095.00	306286.00
Misclassification Rate	0.38	0.38	0.38
Maximum Absolute Error	0.68	0.68	0.68
Sum of Squared Errors	285313.81	47561.23	142743.22
Average Squared Error	0.23	0.23	0.23
Root Average Squared Error	0.48	0.48	0.48
Divisor for ASE	1225144.00	204190.00	612572.00
Total Degrees of Freedom	612572.00	.	.

Node=Decision Tree
Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
NU_NOTA_MT		10	1.0000	1.0000	1.0000
TP_COR_RACA		3	0.1590	0.1583	0.9951
NU_NOTA_CH		5	0.1471	0.1635	1.1121







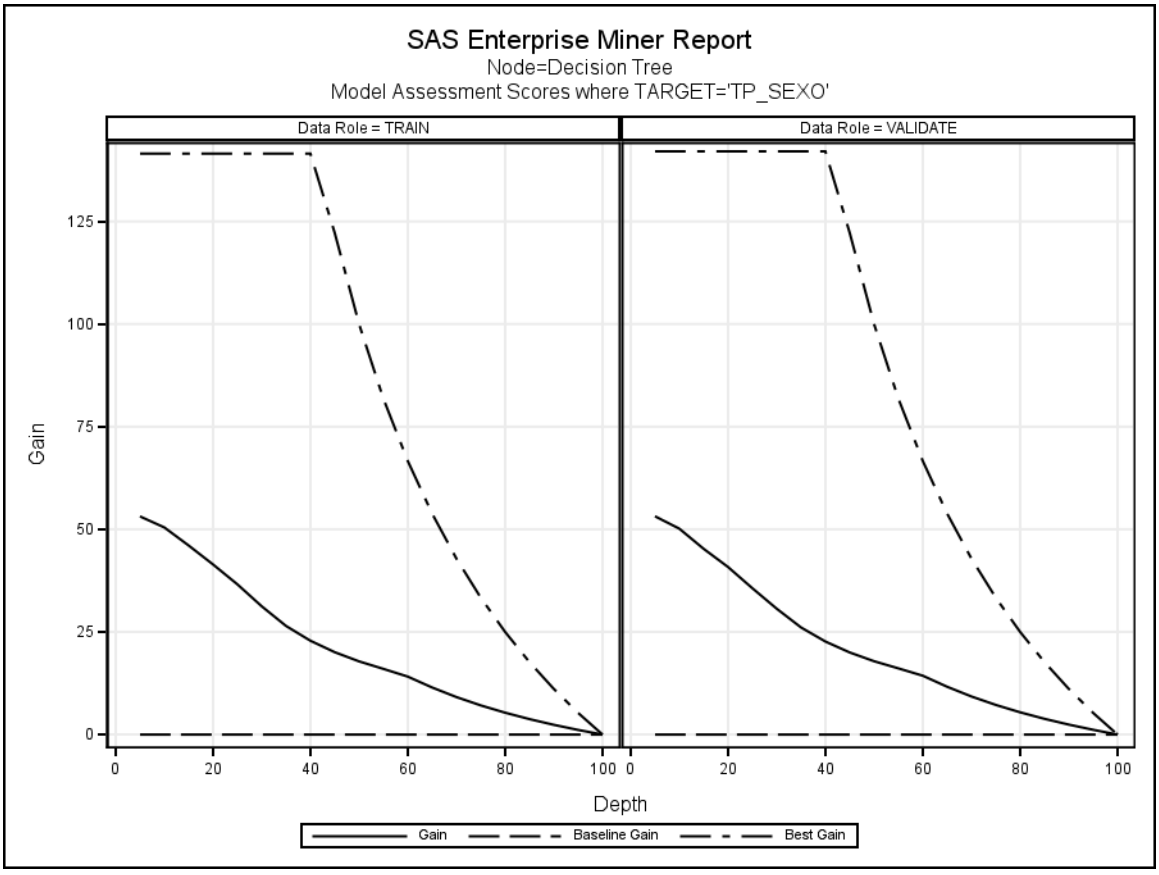
Node=Decision Tree
Event Classification

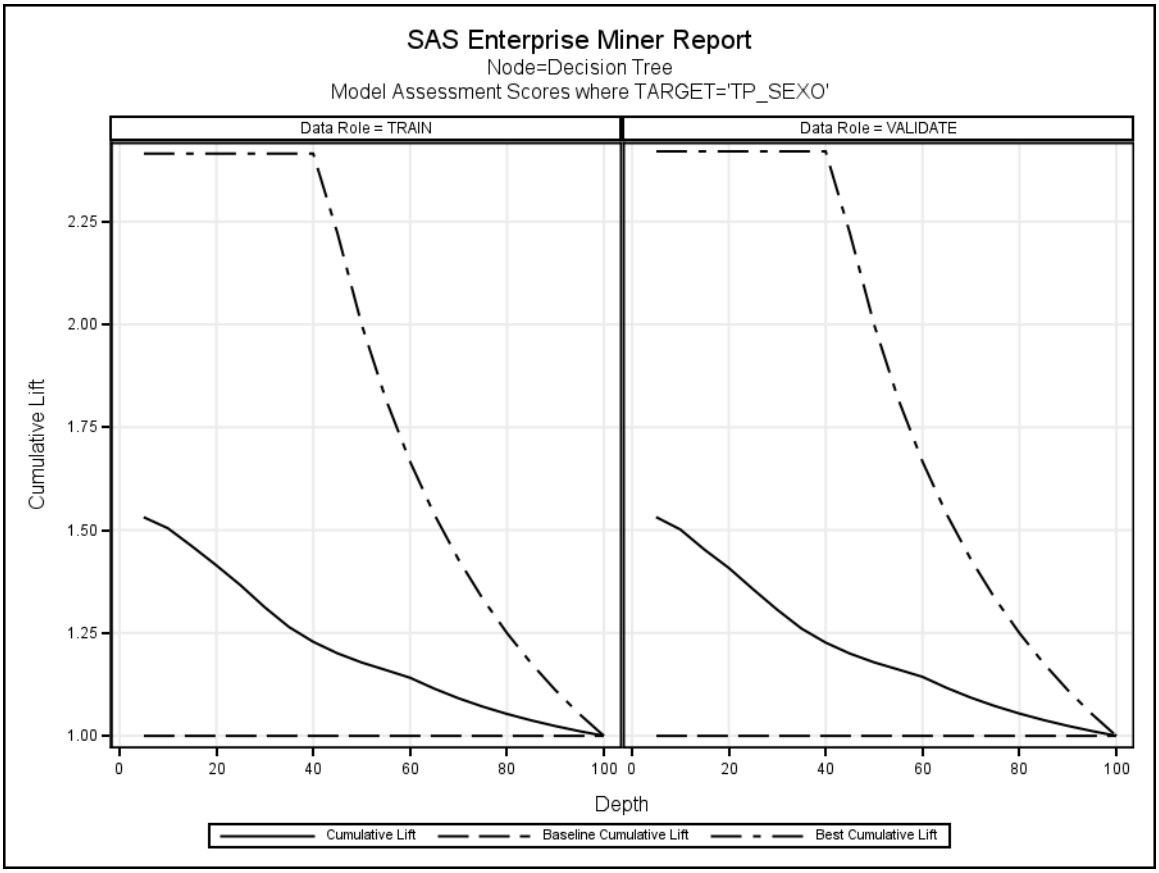
Target Variable=TP_SEXO Data Role=TRAIN

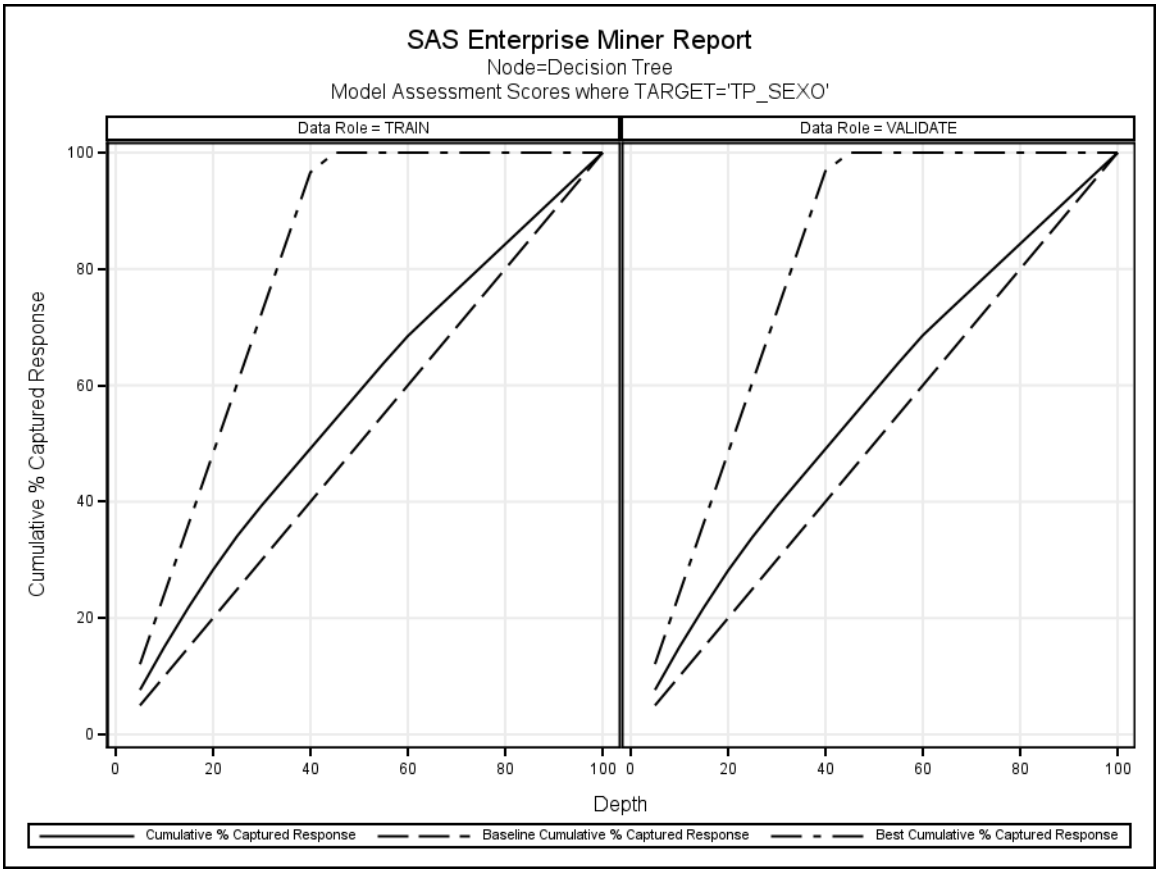
Target	Outcome	State	Frequency Count	Percent
F	F	Correct	305775	49.9166
M	F	Incorrect	179396	29.2857
F	M	Incorrect	53196	8.6840
M	M	Correct	74205	12.1137

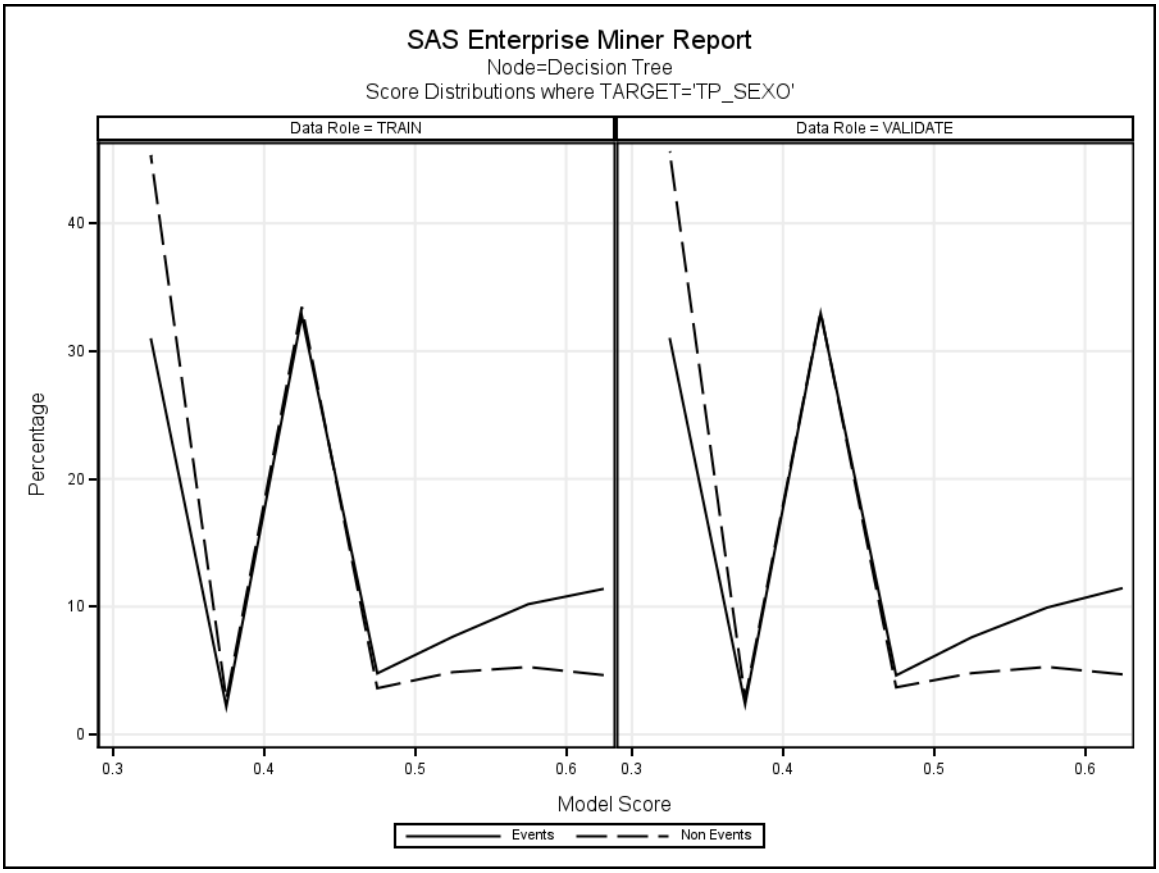
Target Variable=TP_SEXO Data Role=VALIDATE

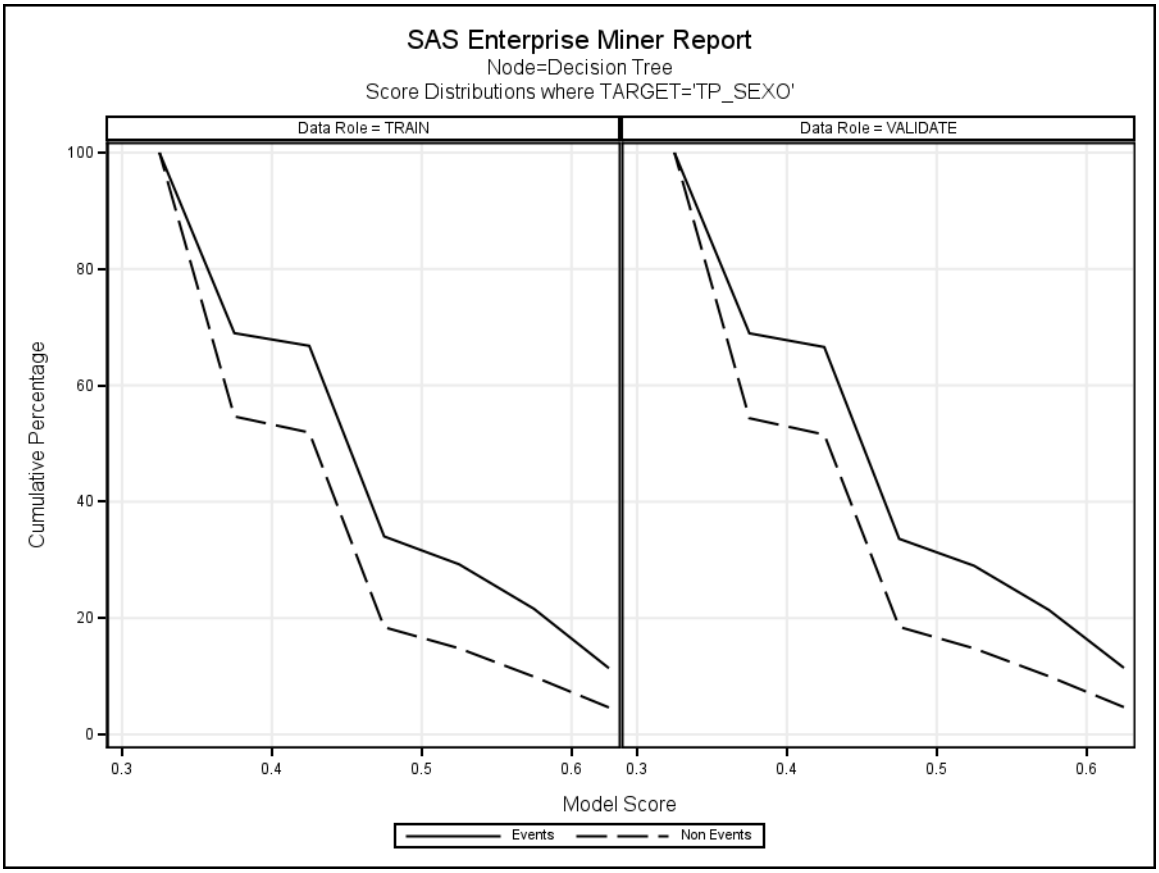
Target	Outcome	State	Frequency Count	Percent
F	F	Correct	51050	50.0024
M	F	Incorrect	29941	29.3266
F	M	Incorrect	8871	8.6890
M	M	Correct	12233	11.9820




SAS Enterprise Miner Report
Node=Decision Tree
Model Assessment Scores where TARGET='TP_SEXO'
Data Role = TRAIN
Data Role = VALIDATE
Cumulative Lift
2.25
2.00
1.75
1.50
1.25
1.00
0
20
40
60
80
100
Depth
Cumulative Lift
Baseline Cumulative Lift
Best Cumulative Lift


SAS Enterprise Miner Report
Node=Decision Tree
Model Assessment Scores where TARGET='TP_SEXO'
Data Role = TRAIN
Data Role = VALIDATE
Cumulative % Captured Response
100
80
60
40
20
0
0
20
40
60
80
100
Depth
Cumulative % Captured Response
Baseline Cumulative % Captured Response
Best Cumulative % Captured Response





Node=Decision Tree
Score Distributions

Target Variable=TP_SEXO Data Role=TRAIN

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.60-0.65	28914	11.4014	4.6494	11.401	4.649
0.55-0.60	25876	10.2034	5.2871	21.605	9.936
0.50-0.55	19415	7.6557	4.8826	29.261	14.819
0.45-0.50	12140	4.7870	3.6248	34.048	18.444
0.40-0.45	83083	32.7613	33.4712	66.809	51.915
0.35-0.40	5523	2.1778	2.7512	68.987	54.666
0.30-0.35	78650	31.0133	45.3337	100.000	100.000

Target Variable=TP_SEXO Data Role=VALIDATE

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.60-0.65	4832	11.4573	4.7079	11.457	4.708
0.55-0.60	4191	9.9374	5.2936	21.395	10.002
0.50-0.55	3210	7.6113	4.8030	29.006	14.804
0.45-0.50	1953	4.6308	3.6932	33.637	18.498
0.40-0.45	13900	32.9587	33.0535	66.596	51.551
0.35-0.40	996	2.3616	2.8154	68.957	54.367
0.30-0.35	13092	31.0428	45.6334	100.000	100.000

SAS Enterprise Miner Report

Node=Regression CN->MT
Summary

Node id = Reg
Node label = Regression CN->MT
Meta path = Ids => Stat => Part => Reg
Notes =

Node=Regression CN->MT
Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	Regression		Force	0		PolynomialDegree	2	
AbsConvValue	-1.34078E154	-7.237006E75	GConvTimes	1		PrintDesignMatrix	N	
AbsFTime	1		GConvValue	1E-6		Rule	NONE	
AbsFValue	0		Hierarchy	CLASS		SASSPDS	N	
AbsGTime	1		InputCoding	DEVIATION		SelectionCriterion	DEFAULT	
AbsGValue	0.00001		Interactions			SelectionDefault	Y	
AbsXTime	1		LinkFunction	LOGIT		Sequential	N	
AbsXValue	1E-8		MainEffect	Y		Simple	N	
CIParm	N		MaxCPUTime	1 HOUR		SIEntry	0.05	
ConvDefaults	Y		MaxFunctionCalls	.		SIStay	0.05	
CorB	N		MaxIterations	.		Start	0	
CovB	N		MaxStep	.		StepOutput	N	
Covout	N		MinResourceUse	N		Stop	0	
Details	N		ModelDefaults	Y		SuppressIntercept	N	
Error	NORMAL	LOGISTIC	ModelSelection	NONE		SuppressOutput	N	
ExcludedVariable	REJECT		OptimizationTechnique	DEFAULT		Terms	N	
FConvTimes	1		Performance	N		TwoFactor	N	
FConvValue	0		Polynomial	N				

Node=Regression CN->MT
Variable Summary

Role	Level	Frequency Count	Name
TARGET	NOMINAL	1	TP_SEXO
INPUT	INTERVAL	2	NU_NOTA_CH NU_NOTA_MT
INPUT	NOMINAL	1	TP_COR_RACA

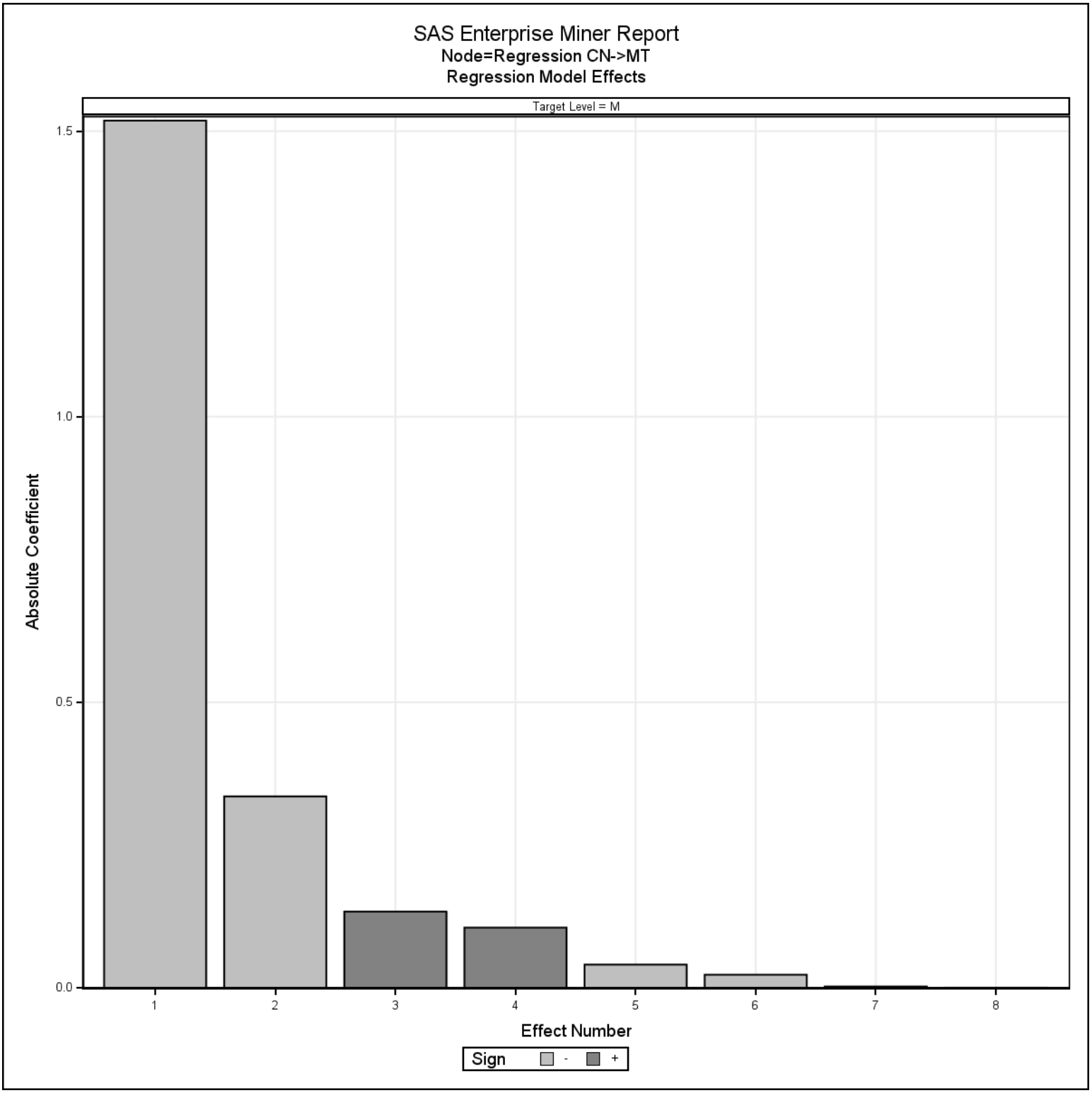
Node=Regression CN->MT
Model Fit Statistics

Target=TP_SEXO Target Label=' '

Label of Statistic	Train	Validation	Test
Akaike's Information Criterion	817320.05	.	.
Average Squared Error	0.24	0.24	0.24
Average Error Function	0.67	0.67	0.67
Degrees of Freedom for Error	612564.00	.	.
Model Degrees of Freedom	8.00	.	.
Total Degrees of Freedom	612572.00	.	.

Target=TP_SEXO Target Label=' '

Label of Statistic	Train	Validation	Test
Divisor for ASE	1225144.00	204190.00	612572.00
Error Function	817304.05	136216.27	408694.41
Final Prediction Error	0.24	.	.
Maximum Absolute Error	0.86	0.86	0.86
Mean Square Error	0.24	0.24	0.24
Sum of Frequencies	612572.00	102095.00	306286.00
Number of Estimate Weights	8.00	.	.
Root Average Sum of Squares	0.49	0.49	0.49
Root Final Prediction Error	0.49	.	.
Root Mean Squared Error	0.49	0.49	0.49
Schwarz's Bayesian Criterion	817410.66	.	.
Sum of Squared Errors	289904.71	48311.94	144960.16
Sum of Case Weights Times Freq	1225144.00	204190.00	612572.00
Misclassification Rate	0.39	0.39	0.39



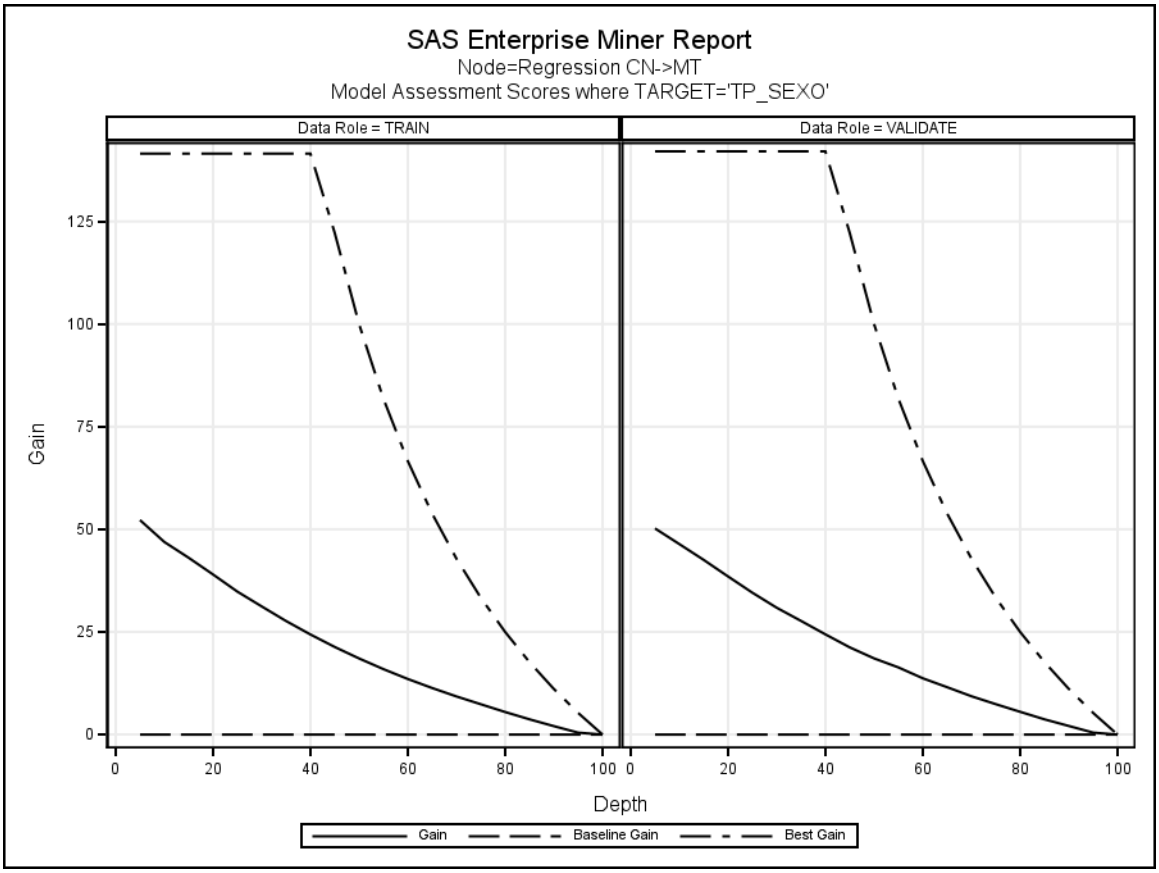
Node=Regression CN->MT
Event Classification

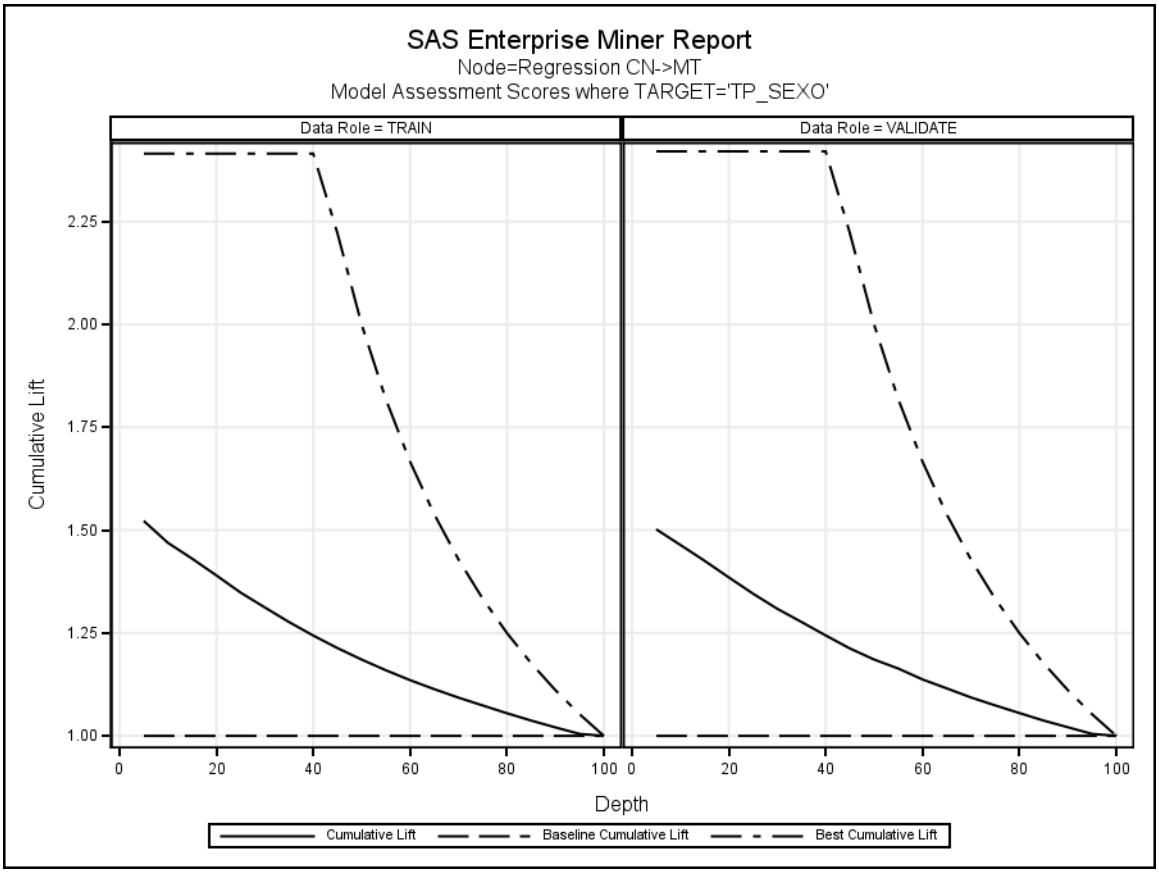
Target Variable=TP_SEXO Data Role=TRAIN

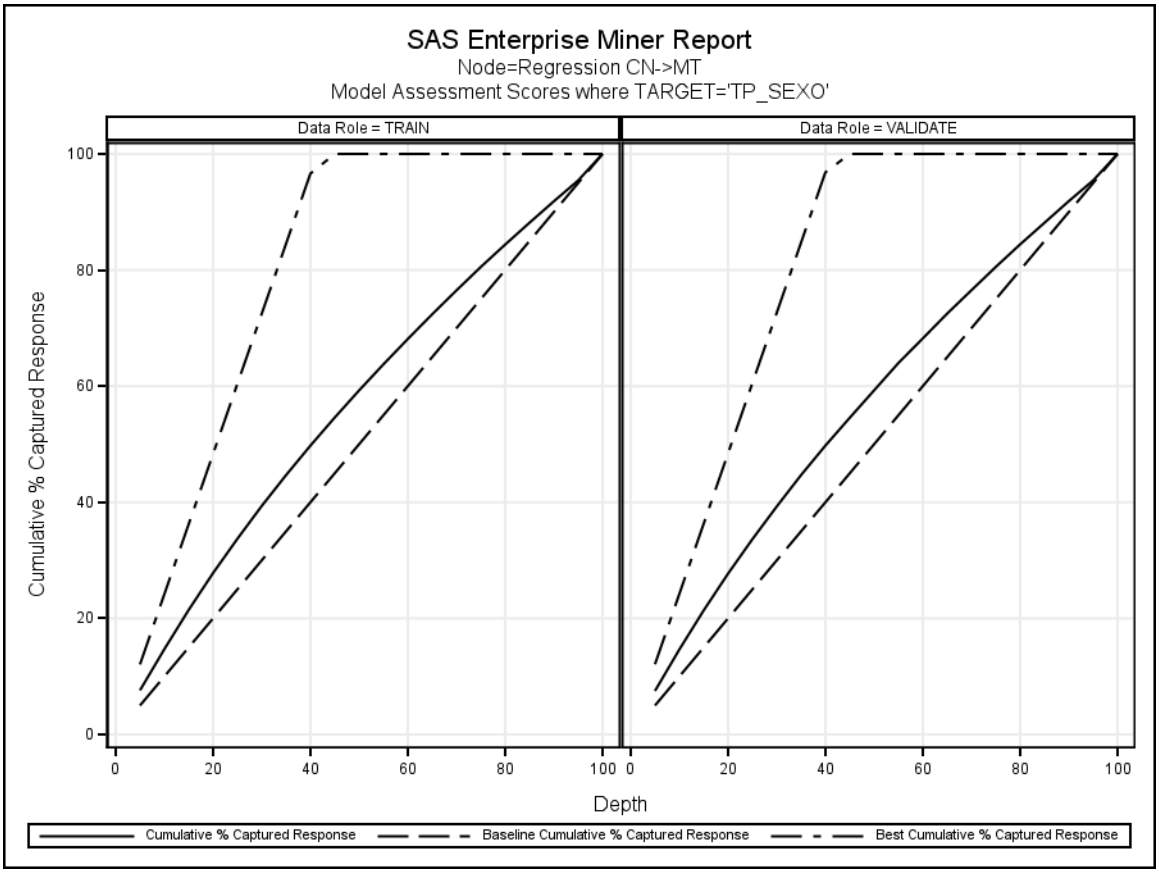
Target	Outcome	State	Frequency Count	Percent
F	F	Correct	331728	54.1533
M	F	Incorrect	212154	34.6333
F	M	Incorrect	27243	4.4473
M	M	Correct	41447	6.7661

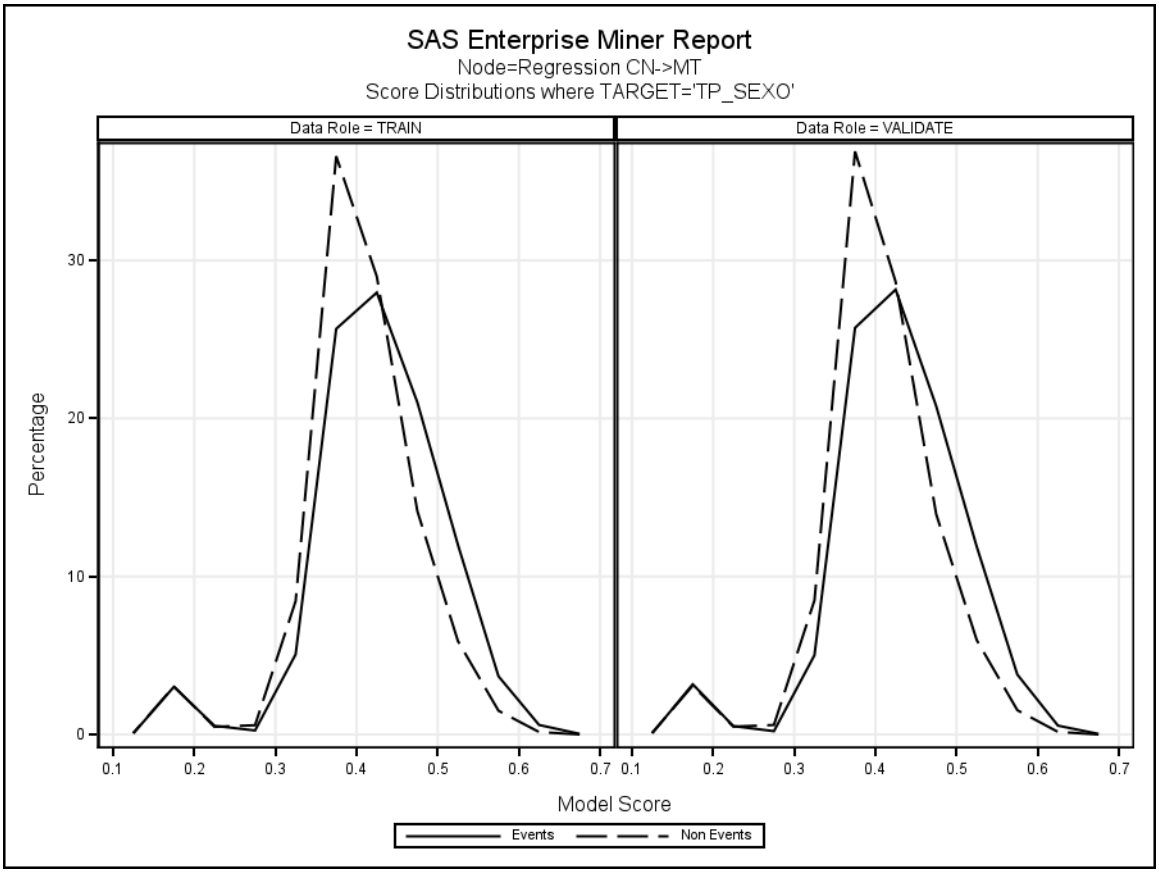
Target Variable=TP_SEXO Data Role=VALIDATE

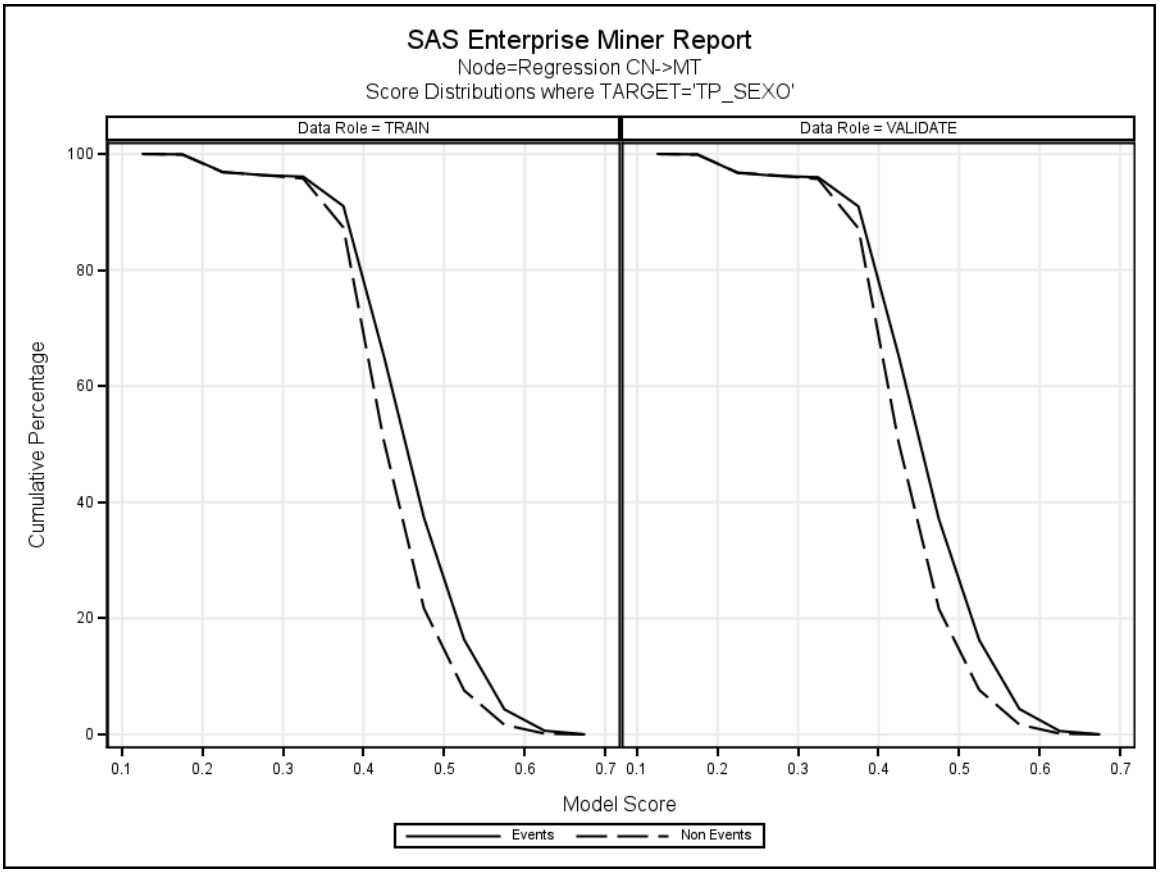
Target	Outcome	State	Frequency Count	Percent
F	F	Correct	55315	54.1799
M	F	Incorrect	35306	34.5815
F	M	Incorrect	4606	4.5115
M	M	Correct	6868	6.7271




SAS Enterprise Miner Report
Node=Regression CN->MT
Model Assessment Scores where TARGET='TP_SEXO'
Data Role = TRAIN
Data Role = VALIDATE
Cumulative Lift
2.25
2.00
1.75
1.50
1.25
1.00
0
20
40
60
80
100
Depth
Cumulative Lift
Baseline Cumulative Lift
Best Cumulative Lift


SAS Enterprise Miner Report
Node=Regression CN->MT
Model Assessment Scores where TARGET='TP_SEXO'
Data Role = TRAIN
Data Role = VALIDATE
Cumulative % Captured Response
100
80
60
40
20
0
0
20
40
60
80
100
Depth
Cumulative % Captured Response
Baseline Cumulative % Captured Response
Best Cumulative % Captured Response





Node=Regression CN->MT
Score Distributions

Target Variable=TP_SEXO Data Role=TRAIN

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.65-0.70	110	0.0434	0.0053	0.043	0.005
0.60-0.65	1539	0.6069	0.1596	0.650	0.165
0.55-0.60	9348	3.6861	1.5071	4.336	1.672
0.50-0.55	30450	12.0071	5.9172	16.343	7.589
0.45-0.50	53320	21.0252	14.1432	37.369	21.732
0.40-0.45	70939	27.9727	28.9929	65.341	50.725
0.35-0.40	65120	25.6781	36.6041	91.019	87.329
0.30-0.35	12871	5.0753	8.4447	96.095	95.774
0.25-0.30	655	0.2583	0.5917	96.353	96.366
0.20-0.25	1404	0.5536	0.4883	96.907	96.854
0.15-0.20	7662	3.0213	3.0490	99.928	99.903
0.10-0.15	183	0.0722	0.0969	100.000	100.000

Target Variable=TP_SEXO Data Role=VALIDATE

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.65-0.70	17	0.0403	0.0050	0.040	0.005
0.60-0.65	238	0.5643	0.1686	0.605	0.174
0.55-0.60	1601	3.7962	1.5404	4.401	1.714
0.50-0.55	5012	11.8841	5.9729	16.285	7.687
0.45-0.50	8764	20.7806	13.9417	37.065	21.628
0.40-0.45	11877	28.1619	28.6544	65.227	50.283
0.35-0.40	10856	25.7410	36.9053	90.968	87.188
0.30-0.35	2119	5.0244	8.4979	95.993	95.686
0.25-0.30	90	0.2134	0.6058	96.206	96.292
0.20-0.25	226	0.5359	0.5007	96.742	96.792
0.15-0.20	1342	3.1821	3.0991	99.924	99.892
0.10-0.15	32	0.0759	0.1085	100.000	100.000

SAS Enterprise Miner Report

Node=Neural Network Summary

Node id = Neural
Node label = Neural Network
Meta path = Ids => Stat => Part => Neural
Notes =

Node=Neural Network Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	NeuralNetwork		Hidden	3		Prelim	Y	
AbsConvValue	-1.34078E154	-7.237006E75	HiddenActivation	DEFAULT		PrelimMaxTime	1 HOUR	
AbsFTime	1		HiddenBias	Y		PrelimMaxiter	10	
AbsFValue	0		HiddenCombFunction	DEFAULT		PrelimOutest		
AbsGTime	1		HiddenUnits	N		PreliminaryRuns	5	
AbsGValue	0.00001		InitialDs			RandDist	NORMAL	
AbsXTime	1		InitialSeed	12345		RandLoc	0	
AbsXValue	1E-8		InputStandardization	STD		RandScale	0.1	
Accelerate	1.2		Learn	0.1		Residuals	Y	
AddHidden	Y		MaxLearn	50		Standardizations	N	
CodefileNoRes			MaxMomentum	1.75		SuppressOutput	N	
CodefileRes			Maxiter	50		TargetActivation	DEFAULT	
ConvDefaults	Y		Maxtime	4 HOURS		TargetBias	Y	
Decelerate	0.5		MinLearn	0.00001		TargetCombFunction	DEFAULT	
DirectConnection	N		ModelSelectionCriterion	PROFIT/LOSS		TargetError	DEFAULT	
FConvTime	1		Momentum	0		Tilt	0	
FConvValue	0		NetworkArchitecture	MLP		TrainCode		
GConvTime	1		Outest			TrainingTechnique	DEFAULT	
GConvValue	1E-6		Outfit			UseEstimates	N	

Node=Neural Network Variable Summary

Role	Level	Frequency Count	Name
TARGET	NOMINAL	1	TP_SEXO
INPUT	INTERVAL	2	NU_NOTA_CH NU_NOTA_MT
INPUT	NOMINAL	1	TP_COR_RACA

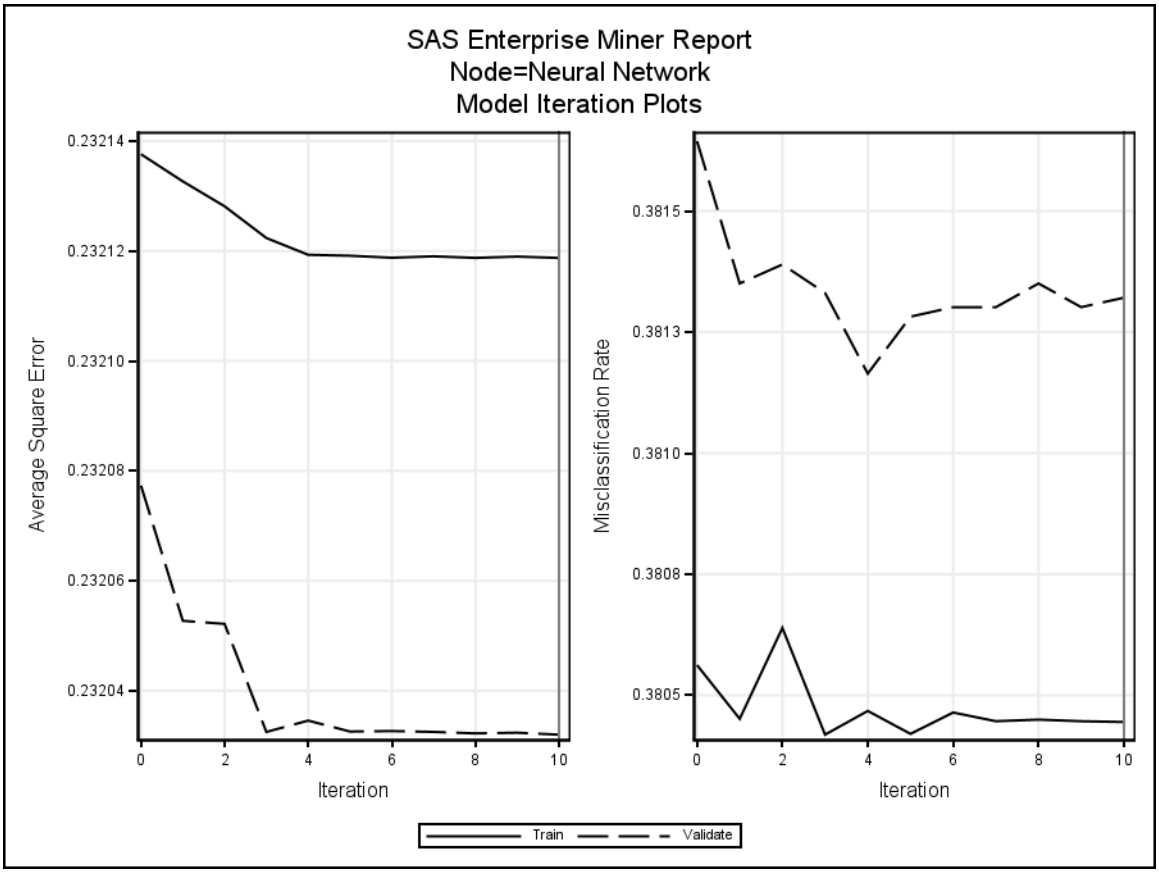
Node=Neural Network Model Fit Statistics

Target=TP_SEXO Target Label=' '

Label of Statistic	Train	Validation	Test
Total Degrees of Freedom	612572.00	.	.
Degrees of Freedom for Error	612544.00	.	.
Model Degrees of Freedom	28.00	.	.
Number of Estimated Weights	28.00	.	.
Akaike's Information Criterion	804568.00	.	.

Target=TP_SEXO Target Label=' '

Label of Statistic	Train	Validation	Test
Schwarz's Bayesian Criterion	804885.11	.	.
Average Squared Error	0.23	0.23	0.23
Maximum Absolute Error	0.93	0.91	0.94
Divisor for ASE	1225144.00	204190.00	612572.00
Sum of Frequencies	612572.00	102095.00	306286.00
Root Average Squared Error	0.48	0.48	0.48
Sum of Squared Errors	284378.90	47378.62	142181.65
Sum of Case Weights Times Freq	1225144.00	204190.00	612572.00
Final Prediction Error	0.23	.	.
Mean Squared Error	0.23	0.23	0.23
Root Final Prediction Error	0.48	.	.
Root Mean Squared Error	0.48	0.48	0.48
Average Error Function	0.66	0.66	0.66
Error Function	804512.00	134042.63	402234.24
Misclassification Rate	0.38	0.38	0.38
Number of Wrong Classifications	233050.00	38931.00	116741.00



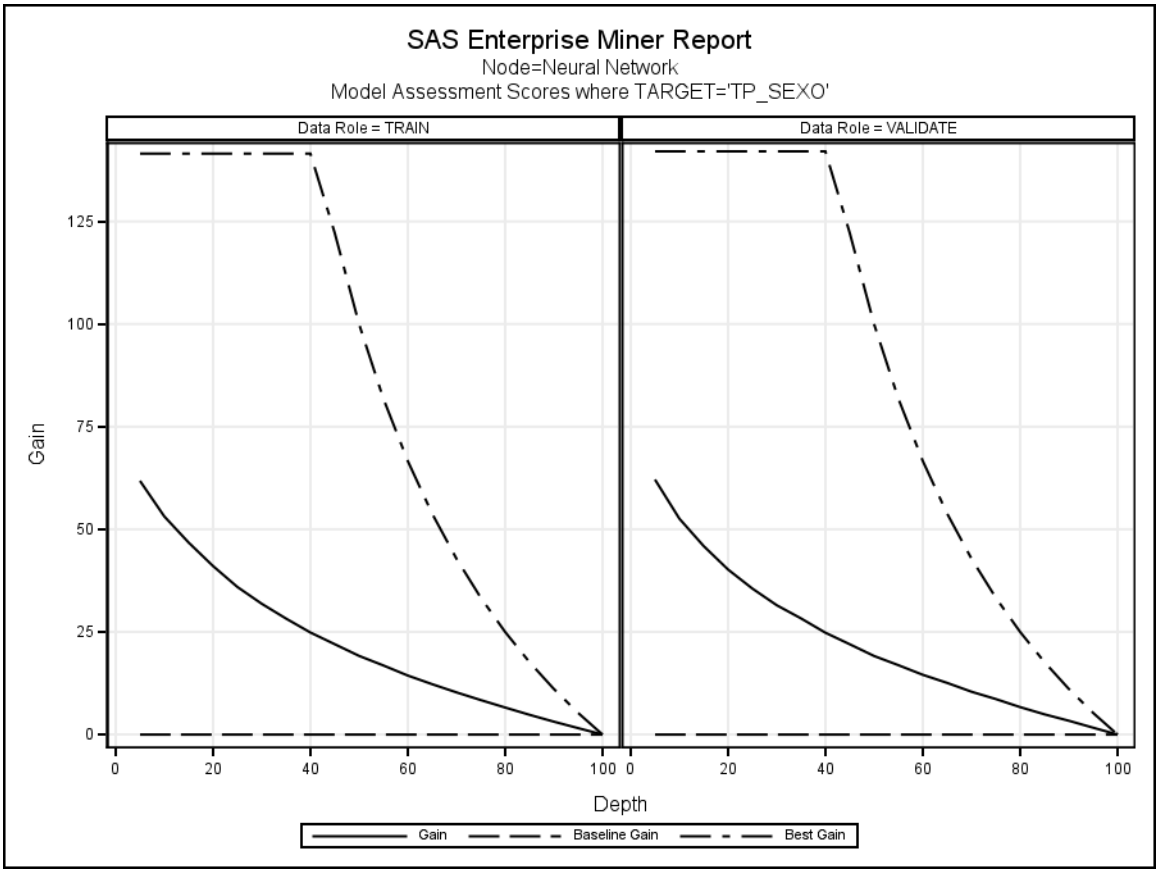
Node=Neural Network
Event Classification

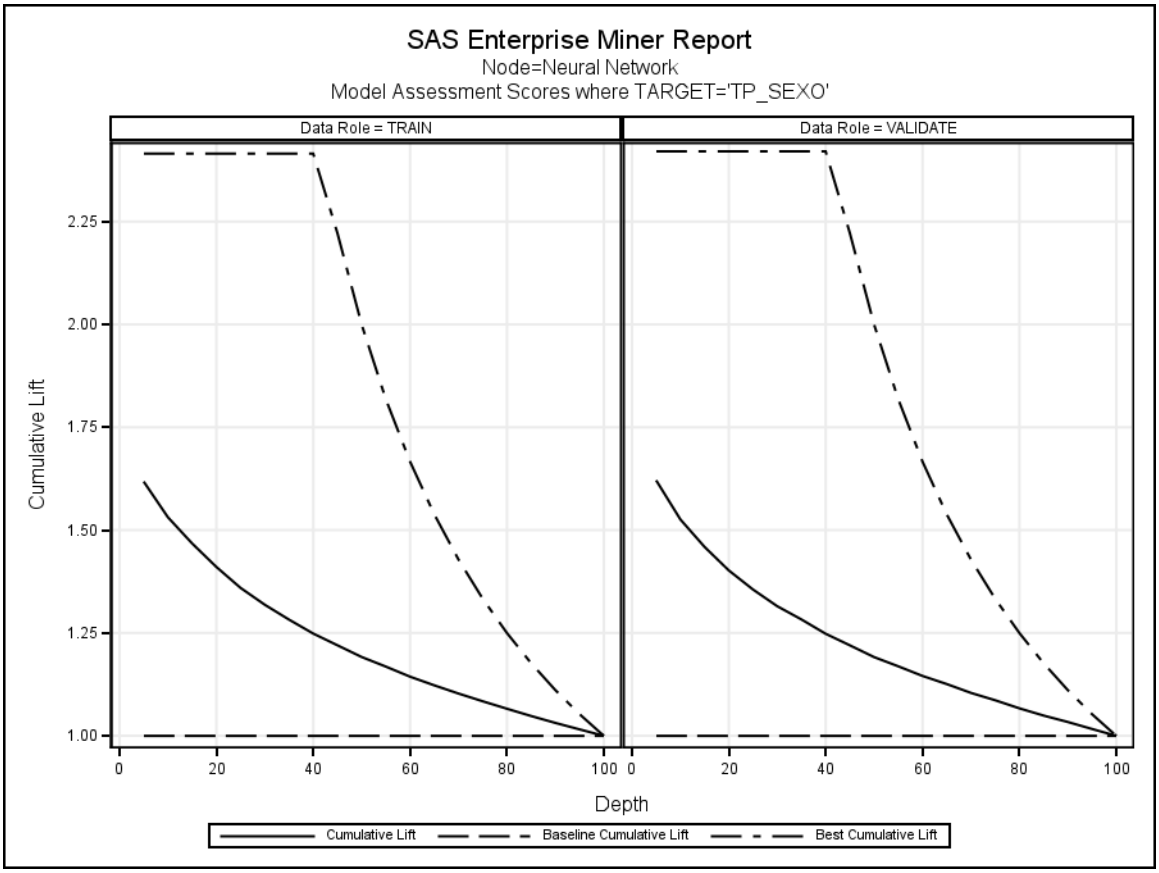
Target Variable=TP_SEXO Data Role=TRAIN

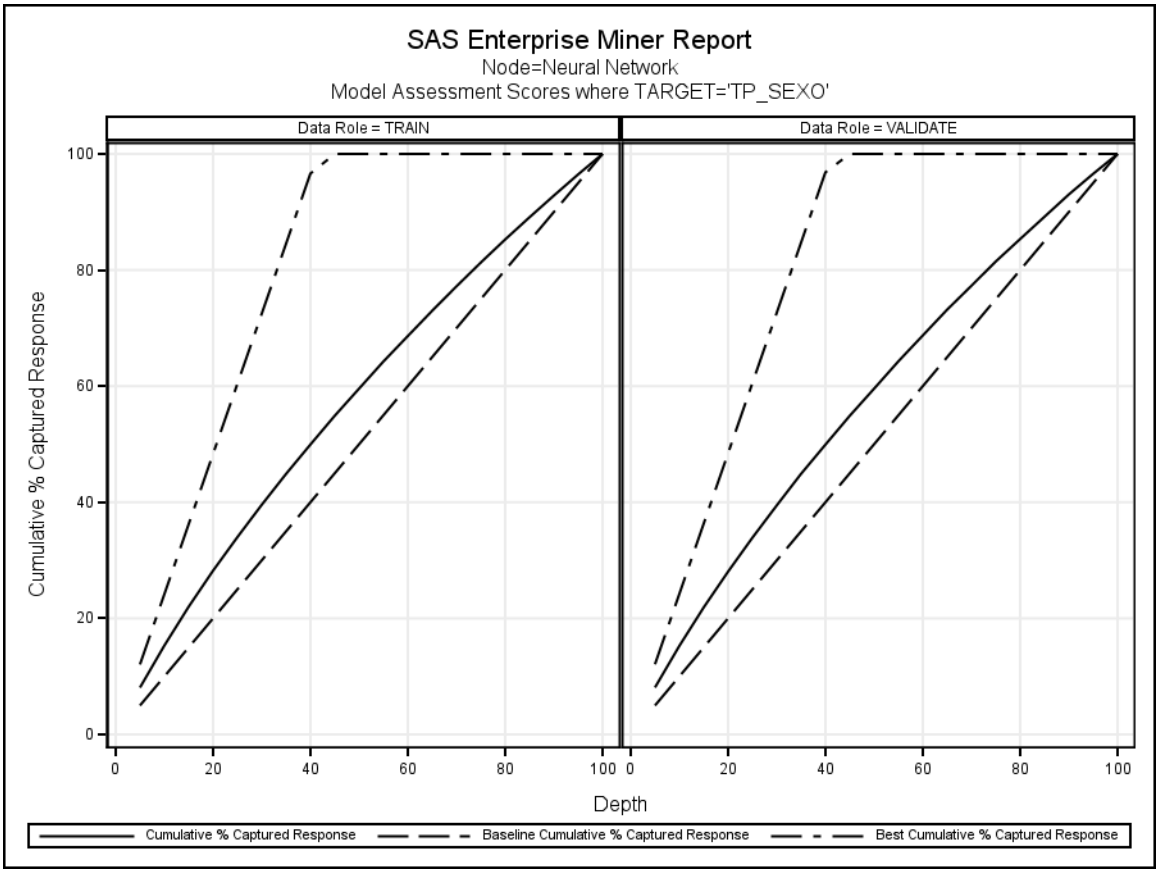
Target	Outcome	State	Frequency Count	Percent
F	F	Correct	307117	50.1357
M	F	Incorrect	181196	29.5795
F	M	Incorrect	51854	8.4650
M	M	Correct	72405	11.8198

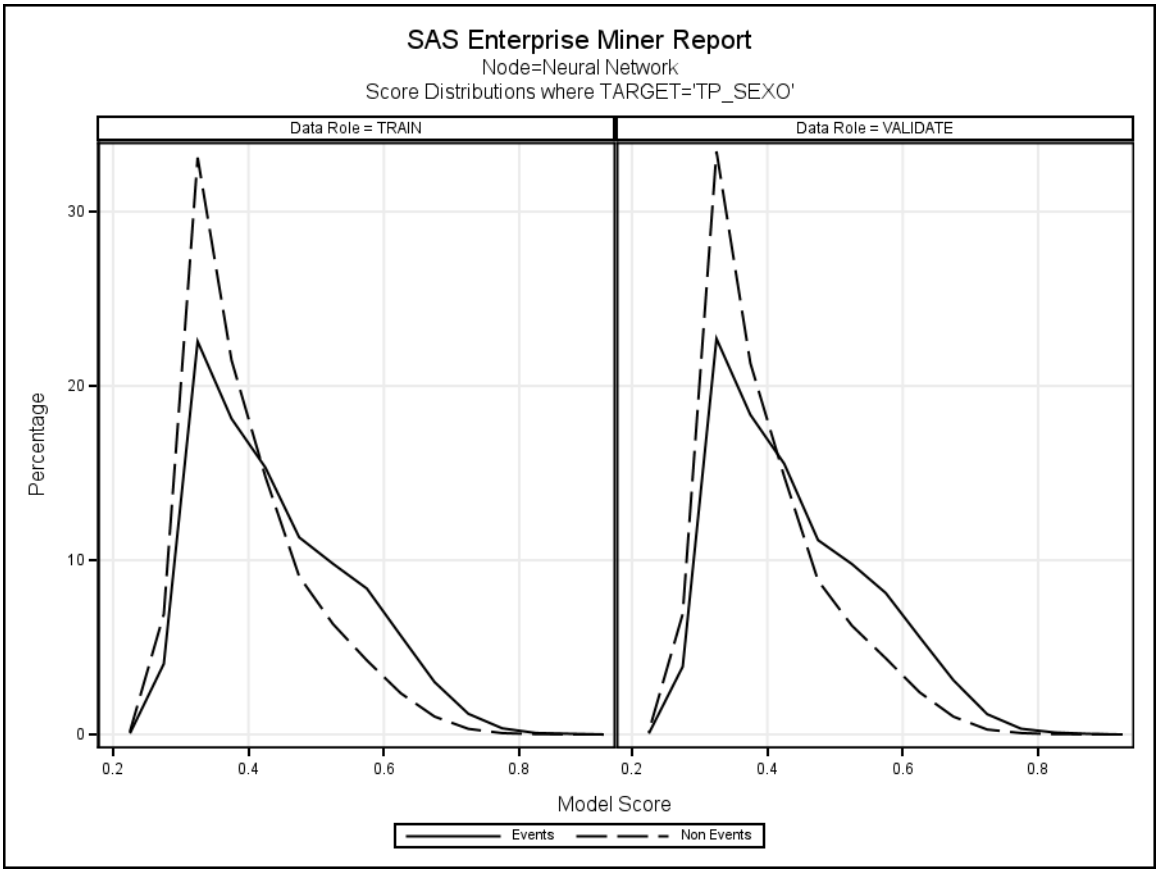
Target Variable=TP_SEXO Data Role=VALIDATE

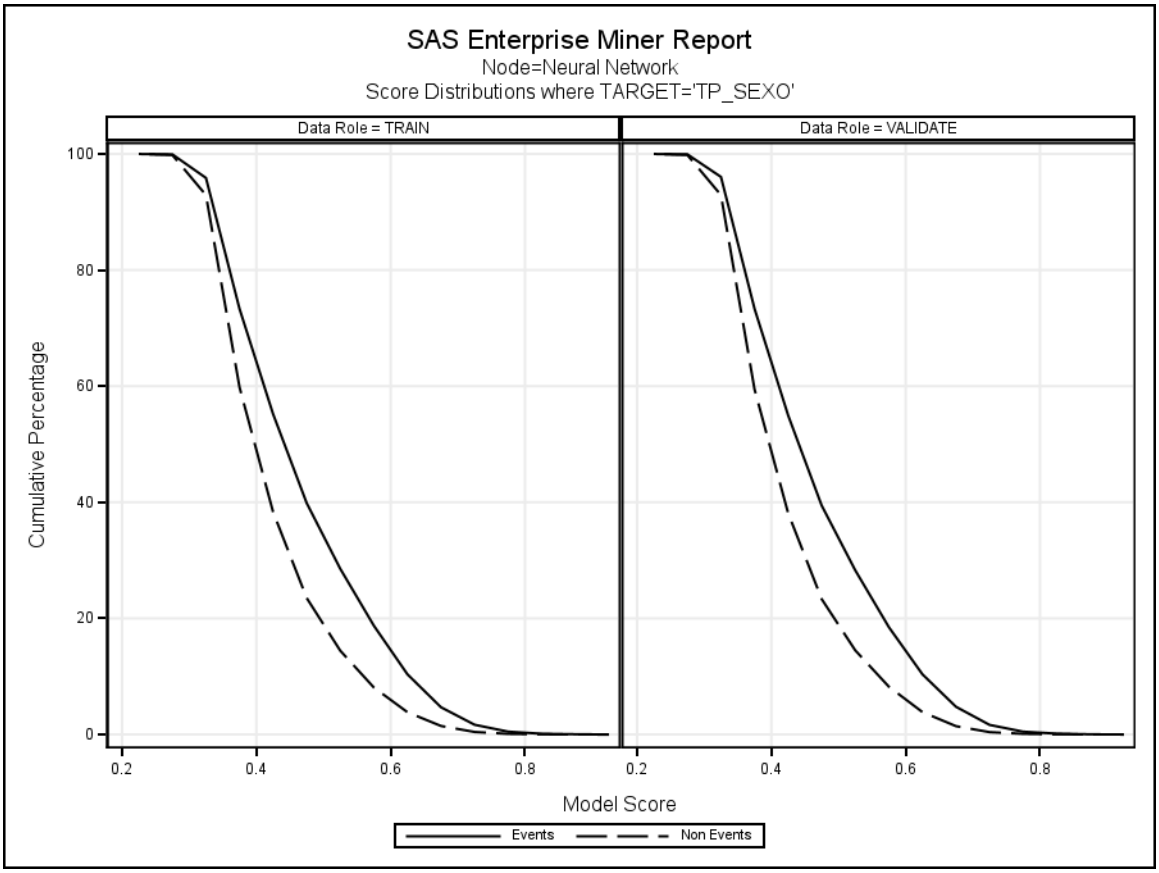
Target	Outcome	State	Frequency Count	Percent
F	F	Correct	51235	50.1837
M	F	Incorrect	30245	29.6244
F	M	Incorrect	8686	8.5078
M	M	Correct	11929	11.6842




SAS Enterprise Miner Report
Node=Neural Network
Model Assessment Scores where TARGET='TP_SEXO'
Data Role = TRAIN
Data Role = VALIDATE
Cumulative Lift
2.25
2.00
1.75
1.50
1.25
1.00
0
20
40
60
80
100
Depth
Cumulative Lift
Baseline Cumulative Lift
Best Cumulative Lift


SAS Enterprise Miner Report
Node=Neural Network
Model Assessment Scores where TARGET='TP_SEXO'
Data Role = TRAIN
Data Role = VALIDATE
Cumulative % Captured Response
100
80
60
40
20
0
0
20
40
60
80
100
Depth
Cumulative % Captured Response
Baseline Cumulative % Captured Response
Best Cumulative % Captured Response





Node=Neural Network
Score Distributions

Target Variable=TP_SEXO Data Role=TRAIN

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.90-0.95	24	0.0095	0.0011	0.009	0.001
0.85-0.90	116	0.0457	0.0117	0.055	0.013
0.80-0.85	222	0.0875	0.0259	0.143	0.039
0.75-0.80	905	0.3569	0.0822	0.500	0.121
0.70-0.75	3014	1.1885	0.3265	1.688	0.447
0.65-0.70	7637	3.0114	1.0304	4.700	1.478
0.60-0.65	14400	5.6782	2.3718	10.378	3.850
0.55-0.60	21240	8.3754	4.2563	18.753	8.106
0.50-0.55	24847	9.7977	6.3392	28.551	14.445
0.45-0.50	28696	11.3154	9.0737	39.866	23.519
0.40-0.45	38866	15.3256	14.8020	55.192	38.321
0.35-0.40	45973	18.1281	21.4859	73.320	59.807
0.30-0.35	57200	22.5551	33.1080	95.875	92.915
0.25-0.30	10289	4.0572	6.9061	99.932	99.821
0.20-0.25	172	0.0678	0.1791	100.000	100.000

Target Variable=TP_SEXO Data Role=VALIDATE

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.90-0.95	2	0.0047	0.0033	0.005	0.003
0.85-0.90	17	0.0403	0.0167	0.045	0.020
0.80-0.85	49	0.1162	0.0234	0.161	0.043
0.75-0.80	141	0.3343	0.0851	0.496	0.129
0.70-0.75	492	1.1666	0.2870	1.662	0.416
0.65-0.70	1315	3.1180	1.0280	4.780	1.444
0.60-0.65	2360	5.5959	2.4249	10.376	3.868
0.55-0.60	3426	8.1235	4.3774	18.500	8.246
0.50-0.55	4127	9.7856	6.2499	28.285	14.496
0.45-0.50	4705	11.1562	8.8366	39.441	23.332
0.40-0.45	6555	15.5428	14.8112	54.984	38.144
0.35-0.40	7740	18.3525	21.3164	73.337	59.460
0.30-0.35	9579	22.7130	33.4607	96.050	92.921
0.25-0.30	1641	3.8910	6.9041	99.941	99.825
0.20-0.25	25	0.0593	0.1752	100.000	100.000

SAS Enterprise Miner Report

Node=Model Comparison Summary

Node id = MdlComp
Node label = Model Comparison
Meta path = Ids => Stat => Part => Neural => MdlComp
Notes =

Node=Model Comparison Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	ModelCompare		NumberOfReportedLevels	1E-6		SelectionData	DEFAULT	
AssessAllTargetLevels	N		NumberofBins	20		SelectionDepth	10	
DecileBin	20		ProfitEpsilon	1E-6		SelectionTable	TRAIN	TABLE
HPCriteria	DEFAULT		RecomputeAssess	N		StatisticUsed	_AUR_	
LiftEpsilon	1E-6		RocChart	Y		TargetLabel		
ModelCriteria	Train: Roc Index		RocEpsilon	0.01		TargetName	TP_SEXO	
ModelDescription	Neural Network		RoiEpsilon	1E-6		classViyaCriteria	DEFAULT	
ModelId	Neural		ScoreDistBin	20		intervalViyaCriteria	DEFAULT	
NormalizeReportingVariables	Y		SelectionCriteria	_AUR_	DEFAULT			

Node=Model Comparison Variable Summary

Role	Level	Frequency Count	Name
TARGET	NOMINAL	1	TP_SEXO

Node=Model Comparison Fit Statistics Table

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Train: Roc Index	Train: Average Squared Error	Train: Misclassification Rate	Train: Kolmogorov-Smirnov Statistic
Y	Neural	Neural	Neural Network	TP_SEXO		0.615	0.23212	0.38045	0.170
	Tree	Tree	Decision Tree	TP_SEXO		0.607	0.23288	0.37970	0.161
	Reg	Reg	Regression CN->MT	TP_SEXO		0.606	0.23663	0.39081	0.167

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Train: Roc Index	Valid: Average Squared Error	Valid: Misclassification Rate	Valid: Kolmogorov-Smirnov Statistic
Y	Neural	Neural	Neural Network	TP_SEXO		0.615	0.23203	0.38132	0.169
	Tree	Tree	Decision Tree	TP_SEXO		0.607	0.23293	0.38016	0.157
	Reg	Reg	Regression CN->MT	TP_SEXO		0.606	0.23660	0.39093	0.166

Node=Model Comparison Model Fit Comparison

Target Variable=TP_SEXO Data Role=Test

Target Label	Statistic	Neural	Tree	Reg
	Test: Kolmogorov-Smirnov Statistic	0.17	0.16	0.17
	Test: Average Squared Error	0.23	0.23	0.24
	Test: Roc Index	0.62	0.61	0.61
	Test: Average Error Function	0.66	.	0.67
	Test: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff	0.42	0.43	0.43
	Test: Cumulative Percent Captured Response	15.23	14.95	14.65
	Test: Percent Captured Response	7.15	7.38	7.07
	Test: Divisor for TASE	612572.00	612572.00	612572.00
	Test: Error Function	402234.24	.	408694.41
	Test: Gain	52.32	49.47	46.53
	Test: Gini Coefficient	0.23	0.21	0.21
	Test: Bin-Based Two-Way Kolmogorov-Smirnov Statistic	0.17	0.16	0.17
	Test: Kolmogorov-Smirnov Probability Cutoff	0.41	0.41	0.43
	Test: Cumulative Lift	1.52	1.49	1.47
	Test: Lift	1.43	1.48	1.41
	Test: Maximum Absolute Error	0.94	0.68	0.86
	Test: Misclassification Rate	0.38	0.38	0.39
	Test: Lower 95% Conf. Limit for TMISC	.	.	0.39
	Test: Upper 95% Conf. Limit for TMISC	.	.	0.39
	Test: Mean Squared Error	0.23	.	0.24
	Test: Sum of Frequencies	306286.00	306286.00	306286.00
	Test: Root Average Squared Error	0.48	0.48	0.49
	Test: Cumulative Percent Response	63.05	61.87	60.65
	Test: Percent Response	59.17	61.08	58.53
	Test: Root Mean Squared Error	0.48	.	0.49
	Test: Sum of Squared Errors	142181.65	142743.22	144960.16
	Test: Sum of Case Weights Times Freq	612572.00	612572.00	612572.00
	Test: Number of Wrong Classifications	116741.00	.	.

Target Variable=TP_SEXO Data Role=Train

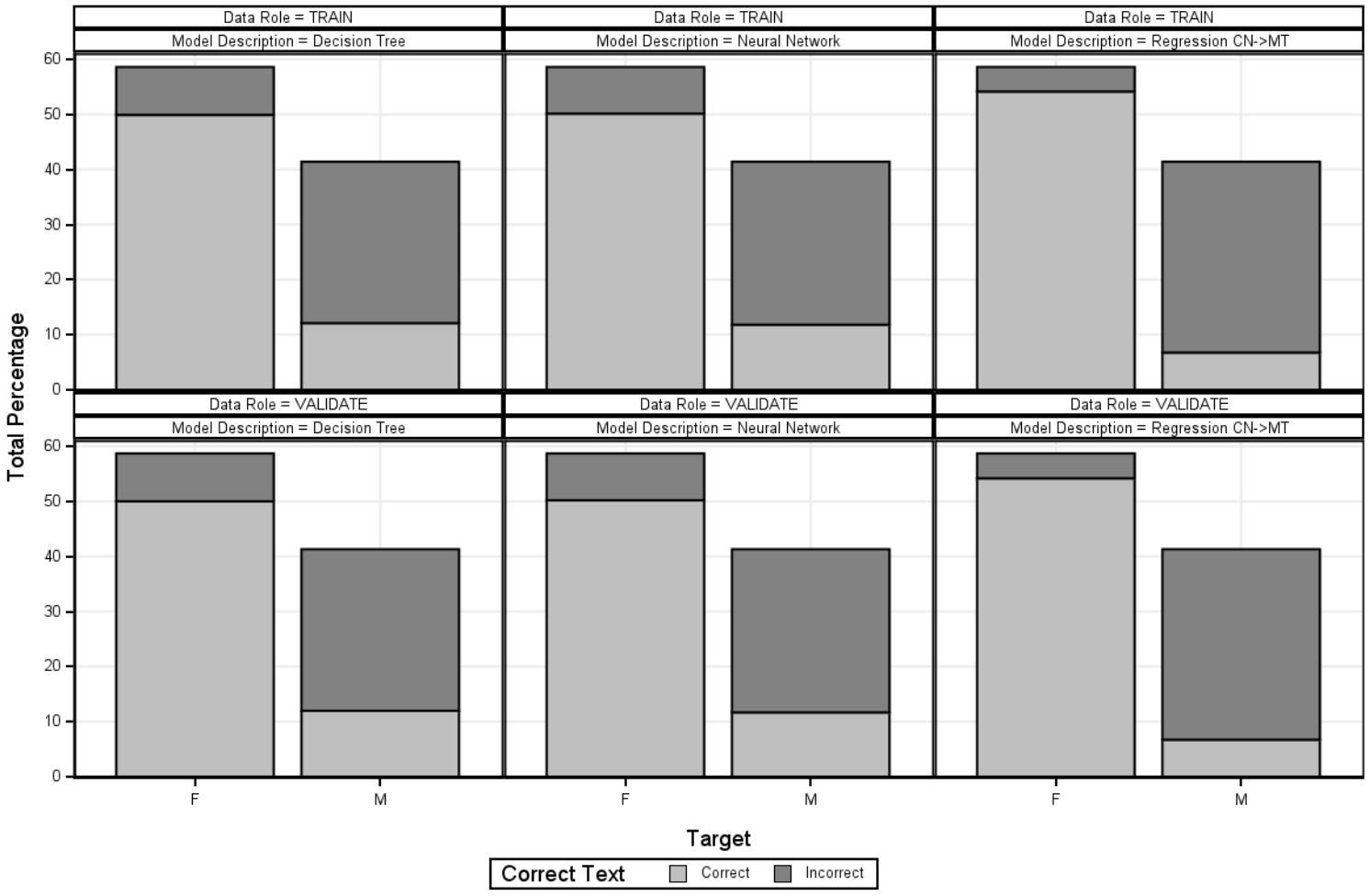
Target Label	Statistic	Neural	Tree	Reg
	Train: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff	0.42	0.43	0.43
	Train: Kolmogorov-Smirnov Statistic	0.17	0.16	0.17
	Train: Akaike's Information Criterion	804568.00	.	817320.05
	Train: Average Squared Error	0.23	0.23	0.24
	Train: Roc Index	0.62	0.61	0.61
	Train: Average Error Function	0.66	.	0.67
	Train: Cumulative Percent Captured Response	15.31	15.05	14.69
	Train: Percent Captured Response	7.22	7.39	7.08
	Selection Criterion: Train: Roc Index	0.62	0.61	0.61
	Train: Degrees of Freedom for Error	612544.00	.	612564.00
	Train: Model Degrees of Freedom	28.00	.	8.00
	Train: Total Degrees of Freedom	612572.00	612572.00	612572.00
	Train: Divisor for ASE	1225144.00	1225144.00	1225144.00
	Train: Error Function	804512.00	.	817304.05
	Train: Final Prediction Error	0.23	.	0.24
	Train: Gain	53.14	50.47	46.93
	Train: Gini Coefficient	0.23	0.21	0.21

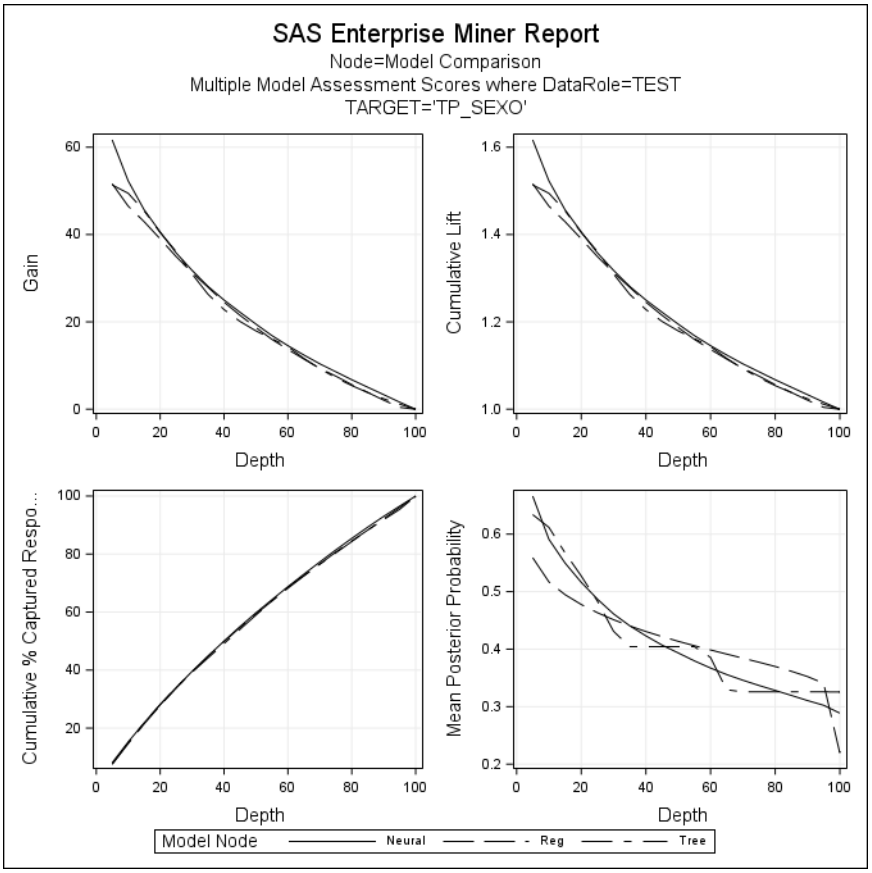
Target Variable=TP_SEXO Data Role=Train

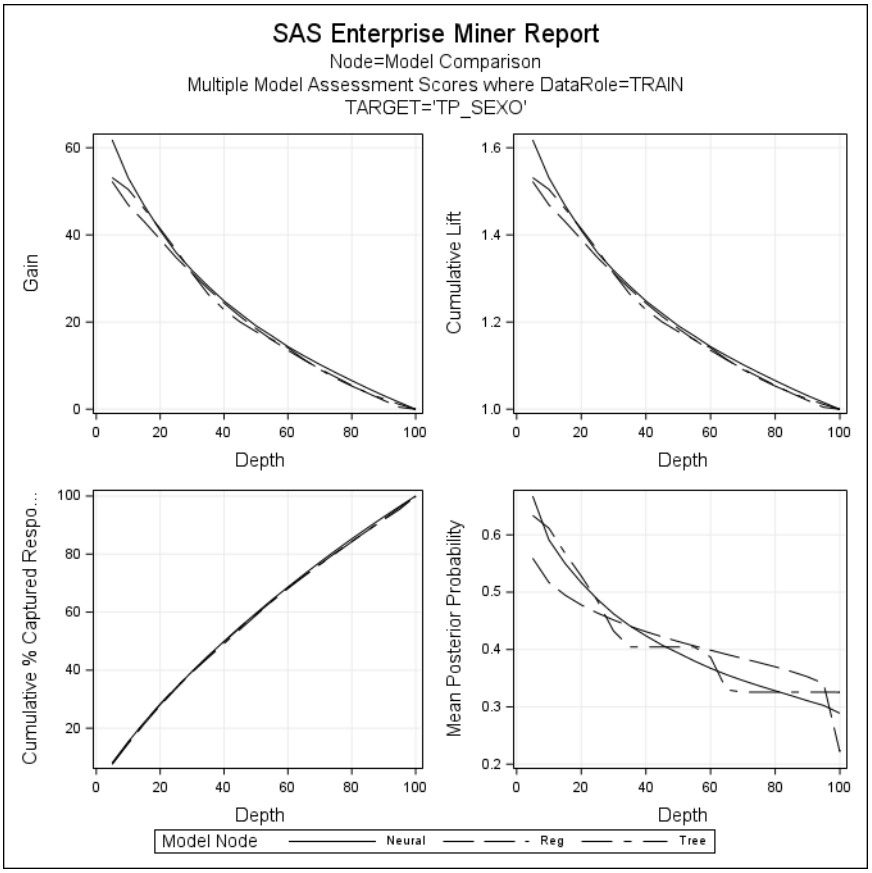
Target Label	Statistic	Neural	Tree	Reg
	Train: Bin-Based Two-Way Kolmogorov-Smirnov Statistic	0.17	0.16	0.17
	Train: Kolmogorov-Smirnov Probability Cutoff	0.41	0.41	0.43
	Train: Cumulative Lift	1.53	1.50	1.47
	Train: Lift	1.44	1.48	1.42
	Train: Maximum Absolute Error	0.93	0.68	0.86
	Train: Misclassification Rate	0.38	0.38	0.39
	Train: Mean Squared Error	0.23	.	0.24
	Train: Sum of Frequencies	612572.00	612572.00	612572.00
	Train: Number of Estimated Weights	28.00	.	8.00
	Train: Root Average Squared Error	0.48	0.48	0.49
	Train: Cumulative Percent Response	63.40	62.29	60.83
	Train: Percent Response	59.81	61.18	58.62
	Train: Root Final Prediction Error	0.48	.	0.49
	Train: Root Mean Squared Error	0.48	.	0.49
	Train: Schwarz's Bayesian Criterion	804885.11	.	817410.66
	Train: Sum of Squared Errors	284378.90	285313.81	289904.71
	Train: Sum of Case Weights Times Freq	1225144.00	.	1225144.00
	Train: Number of Wrong Classifications	233050.00	.	.

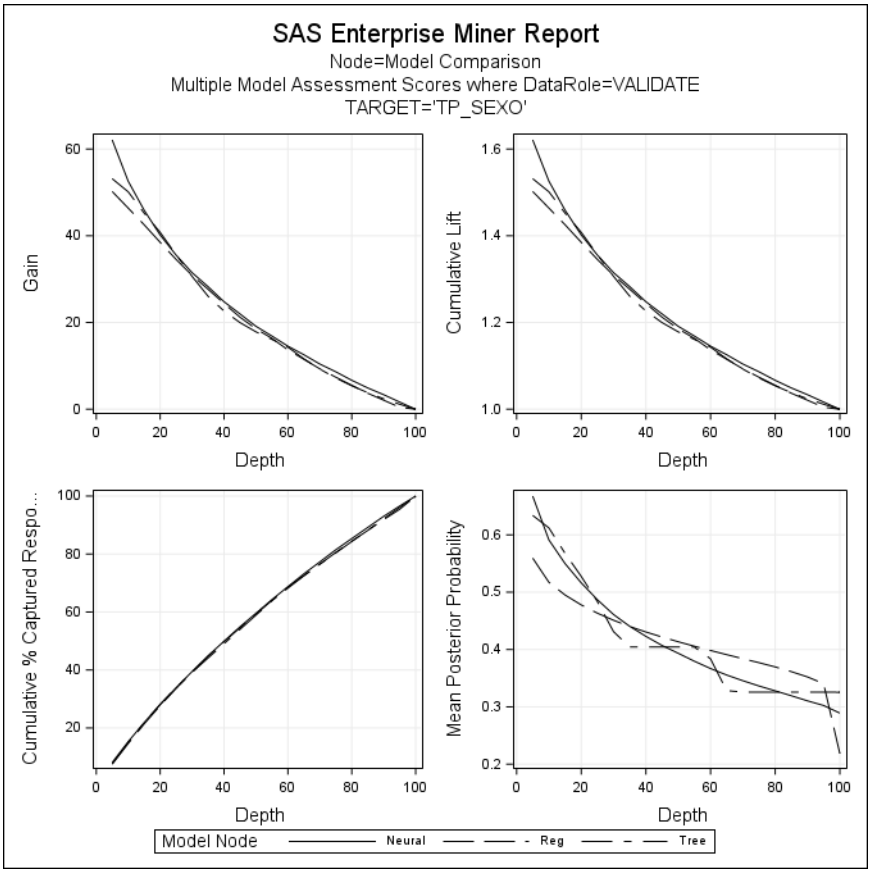
Target Variable=TP_SEXO Data Role=Valid

Target Label	Statistic	Neural	Tree	Reg
	Valid: Kolmogorov-Smirnov Statistic	0.17	0.16	0.17
	Valid: Average Squared Error	0.23	0.23	0.24
	Valid: Roc Index	0.62	0.61	0.61
	Valid: Average Error Function	0.66	.	0.67
	Valid: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff	0.42	0.43	0.43
	Valid: Cumulative Percent Captured Response	15.26	15.02	14.64
	Valid: Percent Captured Response	7.16	7.36	7.13
	Valid: Divisor for VASE	204190.00	204190.00	204190.00
	Valid: Error Function	134042.63	.	136216.27
	Valid: Gain	52.62	50.16	46.41
	Valid: Gini Coefficient	0.23	0.21	0.21
	Valid: Bin-Based Two-Way Kolmogorov-Smirnov Statistic	0.17	0.16	0.17
	Valid: Kolmogorov-Smirnov Probability Cutoff	0.42	0.41	0.43
	Valid: Cumulative Lift	1.53	1.50	1.46
	Valid: Lift	1.43	1.47	1.43
	Valid: Maximum Absolute Error	0.91	0.68	0.86
	Valid: Misclassification Rate	0.38	0.38	0.39
	Valid: Mean Squared Error	0.23	.	0.24
	Valid: Sum of Frequencies	102095.00	102095.00	102095.00
	Valid: Root Average Squared Error	0.48	0.48	0.49
	Valid: Cumulative Percent Response	63.05	62.03	60.48
	Valid: Percent Response	59.12	60.78	58.90
	Valid: Root Mean Squared Error	0.48	.	0.49
	Valid: Sum of Squared Errors	47378.62	47561.23	48311.94
	Valid: Sum of Case Weights Times Freq	204190.00	.	204190.00
	Valid: Number of Wrong Classifications	38931.00	.	.


SAS Enterprise Miner Report
Node=Model Comparison
Classification Chart
TARGET='TP_SEXO' and _TYPE_='PREDICTION'
Data Role = TRAIN
Model Description = Decision Tree
Data Role = TRAIN
Model Description = Neural Network
Data Role = TRAIN
Model Description = Regression CN->MT
Data Role = VALIDATE
Model Description = Decision Tree
Data Role = VALIDATE
Model Description = Neural Network
Data Role = VALIDATE
Model Description = Regression CN->MT
Total Percentage
0
10
20
30
40
50
60
F
M
Target
Correct Text
Correct
Incorrect


SAS Enterprise Miner Report
Node=Model Comparison
Multiple Model Assessment Scores where DataRole=TEST
TARGET='TP_SEXO'
Gain
Depth
Cumulative Lift
Depth
Cumulative % Captured Respo...
Depth
Mean Posterior Probability
Depth
Model Node
Neural
Reg
Tree


SAS Enterprise Miner Report
Node=Model Comparison
Multiple Model Assessment Scores where DataRole=TRAIN
TARGET='TP_SEXO'
Gain
Depth
Cumulative Lift
Depth
Cumulative % Captured Respo...
Depth
Mean Posterior Probability
Depth
Model Node
Neural
Reg
Tree


SAS Enterprise Miner Report
Node=Model Comparison
Multiple Model Assessment Scores where DataRole=VALIDATE
TARGET='TP_SEXO'
Gain
Depth
Cumulative Lift
Depth
Cumulative % Captured Respo...
Depth
Mean Posterior Probability
Depth
Model Node
Neural
Reg
Tree

SAS Enterprise Miner Report

Node=Score
Summary

Node id = Score
Node label = Score
Meta path = Ids => Stat => Part => Neural => MdlComp => Score
Notes =

Node=Score
Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	Score		HideInput	Y		JScore	N	
CScore	N		HideOther	Y		OptimizedCode	Y	
FixedOutputNames	Y		HidePredict	Y		OutputType	DATA	VIEW
GraphReports	Y		HideRejected	Y		PackageName	DEFAULT	
HideAssess	Y		HideResidual	Y		PreferenceName		
HideClassification	Y		HideTarget	Y		ScoreTest	N	
HideFreq	Y		HideVariables	N		ScoreValidate	N	

Node=Score
Variable Summary

Role	Level	Frequency Count	Name
TARGET	NOMINAL	1	TP_SEXO
SEGMENT	NOMINAL	1	b_TP_SEXO

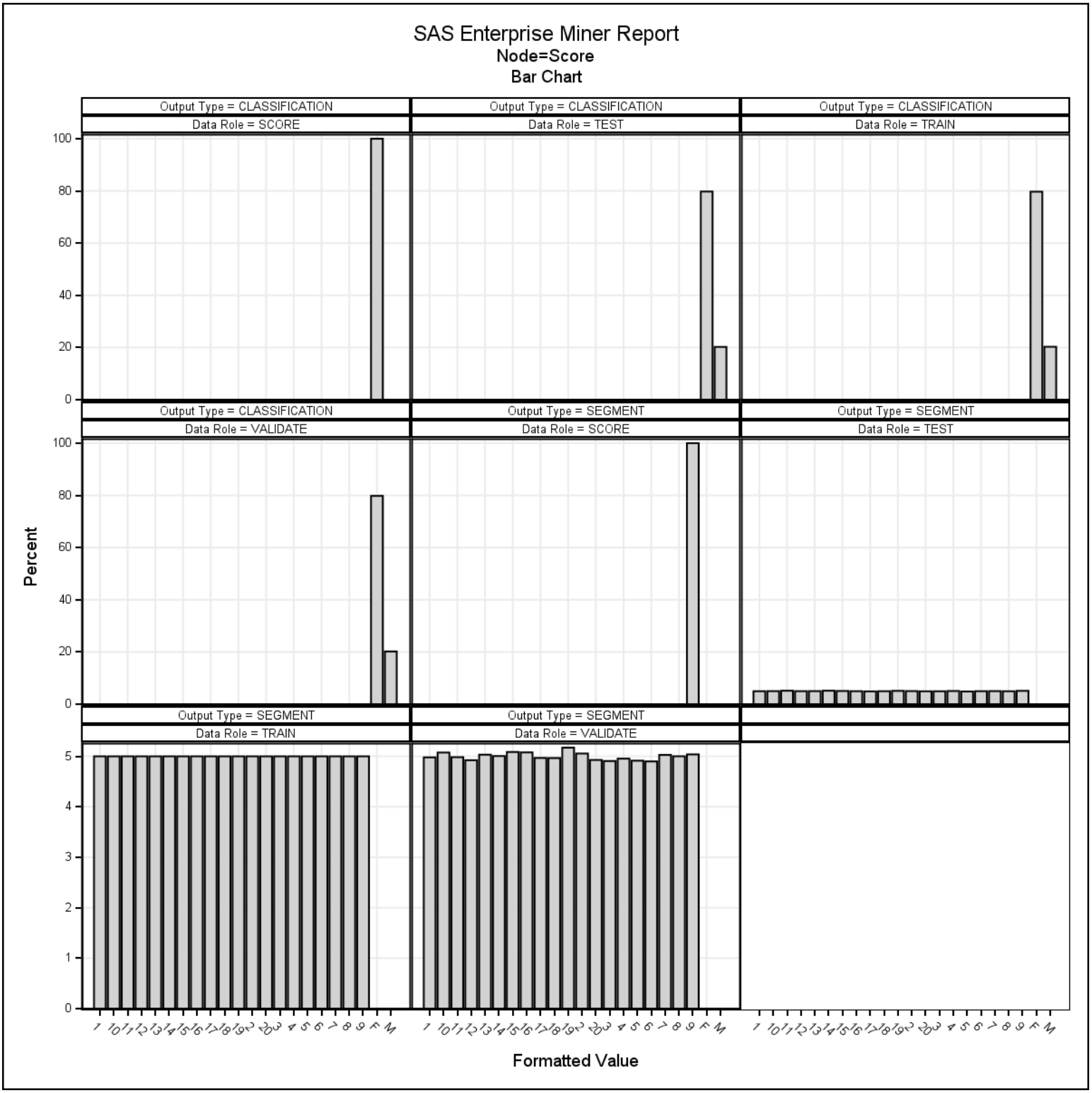
Node=Score
Input Variables

Variable Name	Role	Creator	Comment	Label	Variable Hidden	Used in Score Code
NU_NOTA_CH	INPUT				N	Y
NU_NOTA_MT	INPUT				N	Y
TP_COR_RACA	INPUT				N	Y
TP_SEXO	TARGET				N	N

Node=Score
Output Variables

Variable Name	Creator	Variable Label	Function	Type
EM_CLASSIFICATION	Score	Prediction for TP_SEXO	CLASSIFICATION	C
EM_EVENTPROBABILITY	Score	Probability for level M of TP_SEXO	PREDICT	N
EM_PROBABILITY	Score	Probability of Classification	PREDICT	N
EM_SEGMENT	Score	Segment	TRANSFORM	N
I_TP_SEXO	Neural	Into: TP_SEXO	CLASSIFICATION	C
P_TP_SEXOF	Neural	Predicted: TP_SEXO=F	PREDICT	N
P_TP_SEXOM	Neural	Predicted: TP_SEXO=M	PREDICT	N
U_TP_SEXO	Neural	Unnormalized Into: TP_SEXO	CLASSIFICATION	C

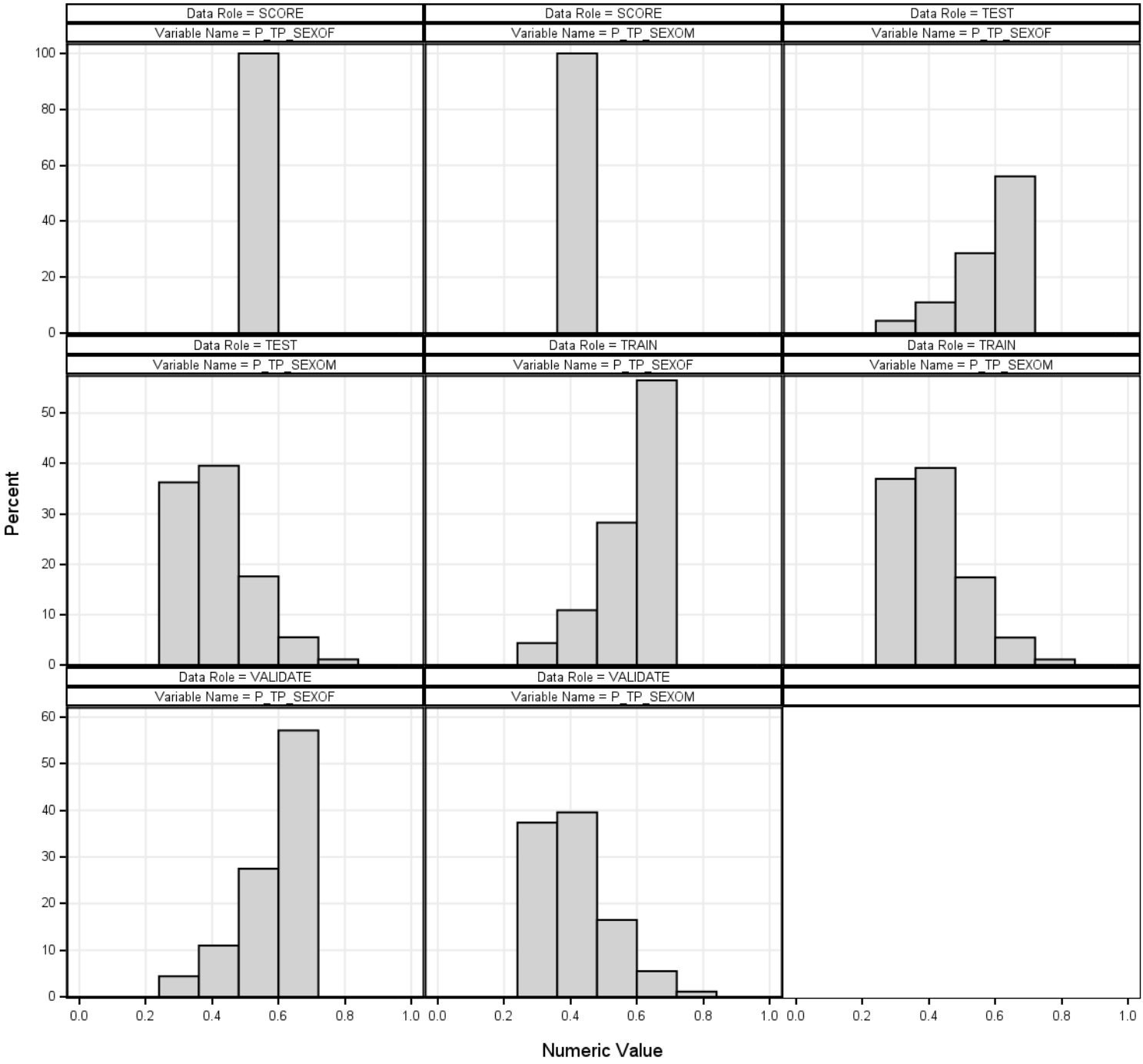
Variable Name	Creator	Variable Label	Function	Type
WARN	Neural	Warnings	ASSESS	C
b_TP_SEXO	MdlComp		TRANSFORM	N



SAS Enterprise Miner Report

Node=Score

Histogram



SAS Enterprise Miner Report

Node=ENEM_2017_2019_SEM_VARIAVEIS Summary

Node id = Ids2
Node label = ENEM_2017_2019_SEM_VARIAVEIS
Meta path = Ids2
Notes =

Node=ENEM_2017_2019_SEM_VARIAVEIS Properties

Property	Value	Default	Property	Value	Default
Component	DataSource		DsCreatedBy	Leandro	
ApplyIntervalLevelLowerLimit	Y		DsId	enemsemvariaveis	
ApplyMaxClassLevels	Y		DsModifiedBy	Leandro	
ApplyMaxPercentMissing	Y		DsModifyDate	1914664948.1	
CMeta	WORK.M1JIIXM4		DsSampleName		
ComputeStatistics	N		DsSampleSize		
DBPassThrough	Y		DsSampleSizeType		
Data	ENEM.ENEM_2017_2019_SEM_VARIAVEIS		DsScope	LOCAL	
DataSelection	DATASOURCE		IdentifyEmptyColumns	Y	
DataSource	enemsemvariaveis		IntervalLowerLimit	20	
DataSourceRole	SCORE		Library	ENEM	
Description			MaxClassLevels	20	
DropMapVariables	Y		MaxPercentMissing	50	
DsCreateDate	1914664948.1		MetaAdvisor	BASIC	

Property	Value	Default
NBytes	12321792	.
NCols	19	.
NObs	102096	.
NewTable		
NewVariableRole	REJECT	
OutputType	DATA	VIEW
Role	SCORE	TRAIN
Sample	D	
SampleSizeObs	10000	
SampleSizePercent	20	
SampleSizeType	PERCENT	
Scope	LOCAL	
Segment		
Table	ENEM_2017_2019_SEM_VARIAVEIS	

Node=ENEM_2017_2019_SEM_VARIAVEIS Data Attributes

Attribute	Value	Attribute	Value	Attribute	Value
Data Name	ENEM_2017_2019_SEM_VARIAVEIS	Date Created	15Jul2020:18:43:03	Data Size	12321792
Data Type	DATA	Date Modified	15Jul2020:18:43:03	Role	SCORE
Data Label		Number Rows	102096	Segment	
Engine	BASE	Number Columns	19	Data Library	ENEM

Node=ENEM_2017_2019_SEM_VARIAVEIS
Variables List

Name	Label	Role	Level	Type	Length	Format	Creator
NOTA_MEDIA		INPUT	INTERVAL	N	8		
NO_MUNICIPIO_RESIDENCIA		INPUT	NOMINAL	C	32		
NU_ANO		INPUT	INTERVAL	N	8		
NU_IDADE		INPUT	INTERVAL	N	8		
NU_INSCRICAO		INPUT	INTERVAL	N	8		
NU_NOTA_CH		INPUT	INTERVAL	N	8		
NU_NOTA_CN		INPUT	INTERVAL	N	8		
NU_NOTA_LC_OLD		INPUT	INTERVAL	N	8		
NU_NOTA_MT_OLD		INPUT	INTERVAL	N	8		
NU_NOTA_REDACAO		INPUT	INTERVAL	N	8		
SG_UF_RESIDENCIA		INPUT	NOMINAL	C	2		
TP_ANO_CONCLUIU		INPUT	NOMINAL	C	2		
TP_COR_RACA_OLD		INPUT	NOMINAL	C	1		
TP_ENSINO		INPUT	NOMINAL	C	1		
TP_ESCOLA		INPUT	NOMINAL	C	1		
TP_ESTADO_CIVIL		INPUT	NOMINAL	C	1		
TP_NACIONALIDADE		INPUT	NOMINAL	C	2		
TP_ST_CONCLUSAO		INPUT	NOMINAL	C	1		
TP_Sexo_OLD		INPUT	NOMINAL	C	1		

End of Report