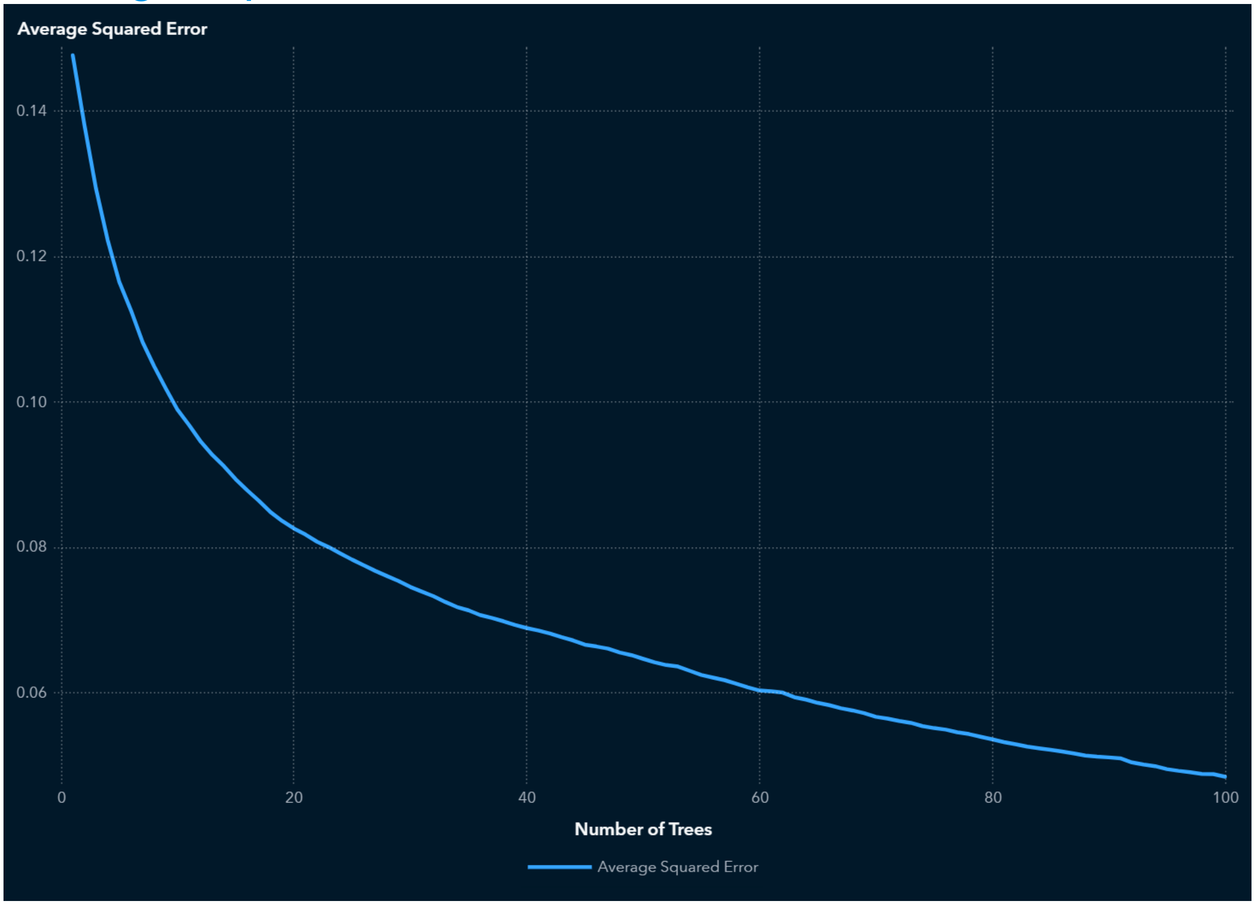
Machine Learning Analytic

"Gradient Boosting" Results

by: jbae7@ncsu.edu

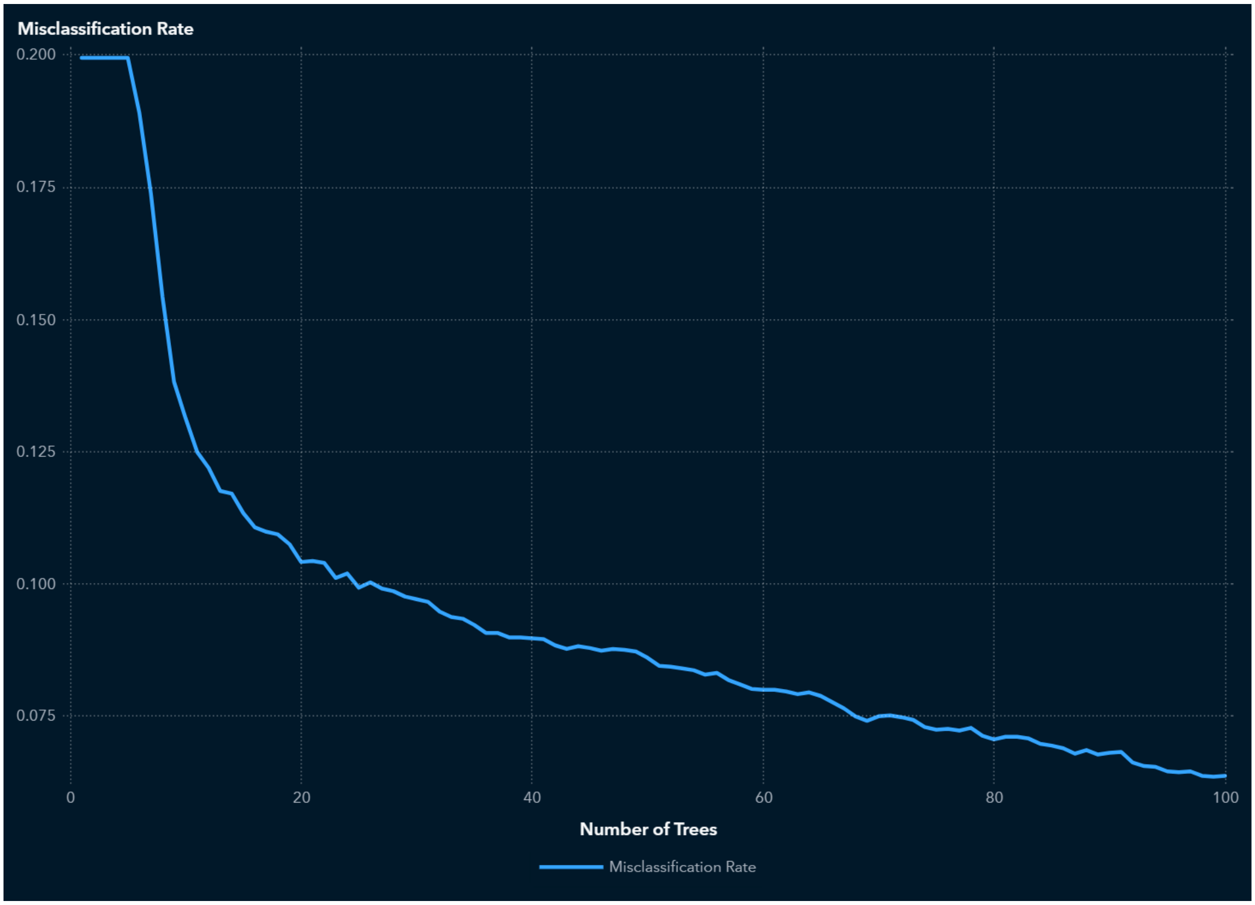
Contents

Average Squared Error



This plot shows how the average squared error changes as the number of trees in the gradient boosting model increases. The training error decreases as the number of trees increases.

Misclassification Rate



This plot shows how the misclassification rate changes as the number of trees in the gradient boosting model increases. The training error decreases as the number of trees increases.

Variable Importance

Variable Label	Role	Variable Name	Training Importance
	INPUT	IM_DEBTINC	22.4423
	INPUT	DELINQ	6.5489
	INPUT	VALUE	6.2246
	INPUT	IM_CLAGE	5.8827
	INPUT	DEROG	5.0021
	INPUT	JOB	3.6054
	INPUT	IM_CLNO	3.1719
	INPUT	LOAN	2.9135
	INPUT	IM_YOJ	2.5767
	INPUT	NINQ	2.3565
	INPUT	IM_MORTDUE	2.3553
	INPUT	REASON	0.4329

Importance Standard Deviation	Relative Importance
65.4393	1
9.6621	0.2918
7.1337	0.2774
6.4780	0.2621
11.5203	0.2229
3.3835	0.1607
3.1344	0.1413
3.2797	0.1298
2.8216	0.1148
2.7267	0.1050
2.6627	0.1049
1.1111	0.0193

Score Inputs

Name	Role	Variable Level	Type
DELINQ	INPUT	NOMINAL	N
DEROG	INPUT	NOMINAL	N
IM_CLAGE	INPUT	INTERVAL	N
IM_CLNO	INPUT	INTERVAL	N
IM_DEBTINC	INPUT	INTERVAL	N
IM_MORTDUE	INPUT	INTERVAL	N
IM_YOJ	INPUT	INTERVAL	N
JOB	INPUT	NOMINAL	C
LOAN	INPUT	INTERVAL	N
NINQ	INPUT	NOMINAL	N
REASON	INPUT	BINARY	C
VALUE	INPUT	INTERVAL	N

Variable Type	Variable Label	Variable Format	Variable Length
double			8
double			8
double			8
double			8
double			8
double			8
double			8
char			7
double			8
double			8
char			7
double			8

Score Outputs

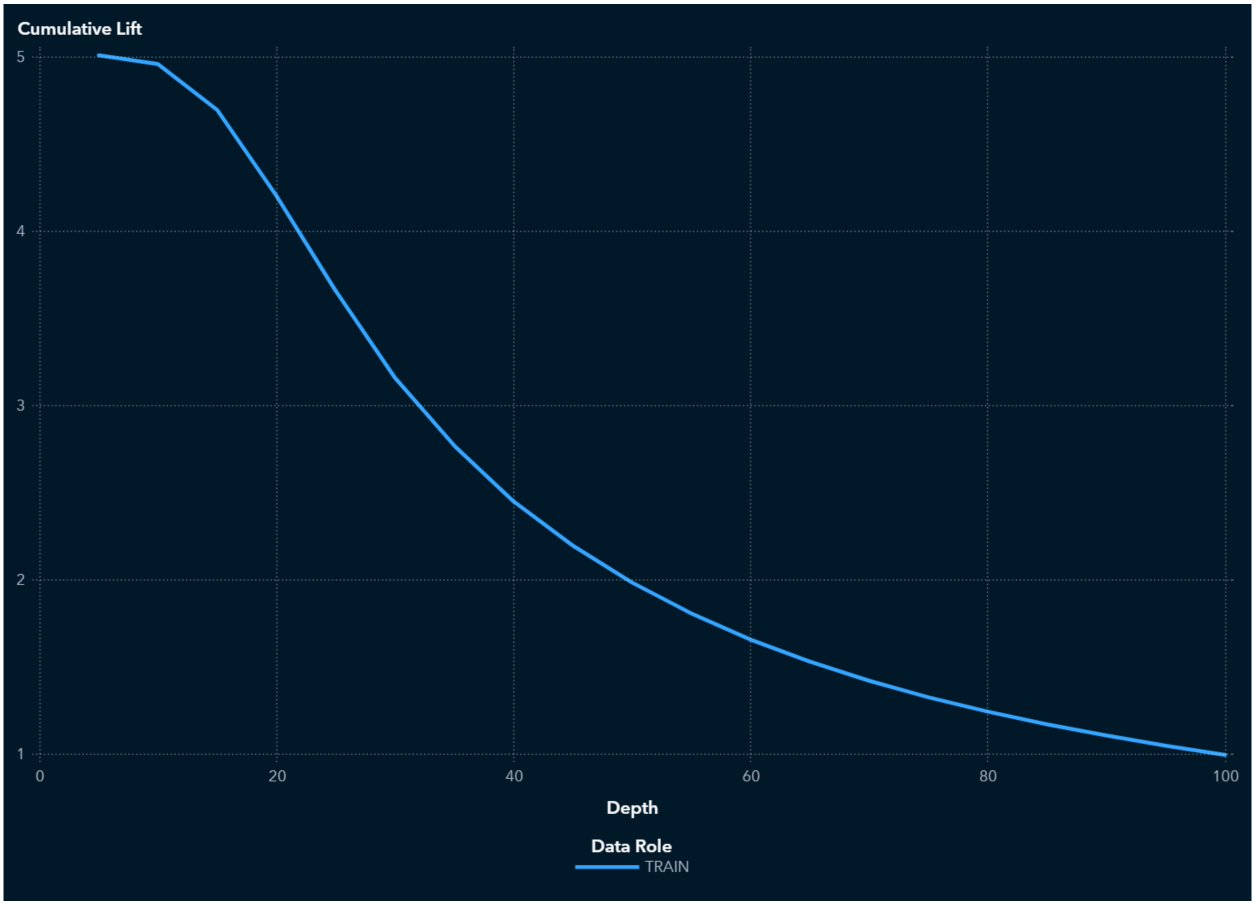
Name	Role	Type	Variable Type
EM_CLASSIFICATION	CLASSIFICATION	C	char
EM_EVENTPROBABILITY	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
I_BAD	CLASSIFICATION	C	char
P_BAD0	PREDICT	N	double
P_BAD1	PREDICT	N	double
WARN	ASSESS	C	char

Variable Label	Variable Format	Variable Length	Creator
Predicted for BAD		12	gradboost
Probability for BAD=1		8	gradboost
Probability of Classification		8	gradboost
Into: BAD		12	gradboost
Predicted: BAD=0		8	gradboost
Predicted: BAD=1		8	gradboost
Warnings		4	gradboost

Function	Creator GUID
CLASSIFICATION	429114b0-3892-4e94-96fa-41b75cd7ceff
PREDICT	429114b0-3892-4e94-96fa-41b75cd7ceff
PREDICT	429114b0-3892-4e94-96fa-41b75cd7ceff

Function	Creator GUID
CLASSIFICATION	429114b0-3892-4e94-96fa-41b75cd7ceff
PREDICT	429114b0-3892-4e94-96fa-41b75cd7ceff
PREDICT	429114b0-3892-4e94-96fa-41b75cd7ceff
ASSESS	429114b0-3892-4e94-96fa-41b75cd7ceff

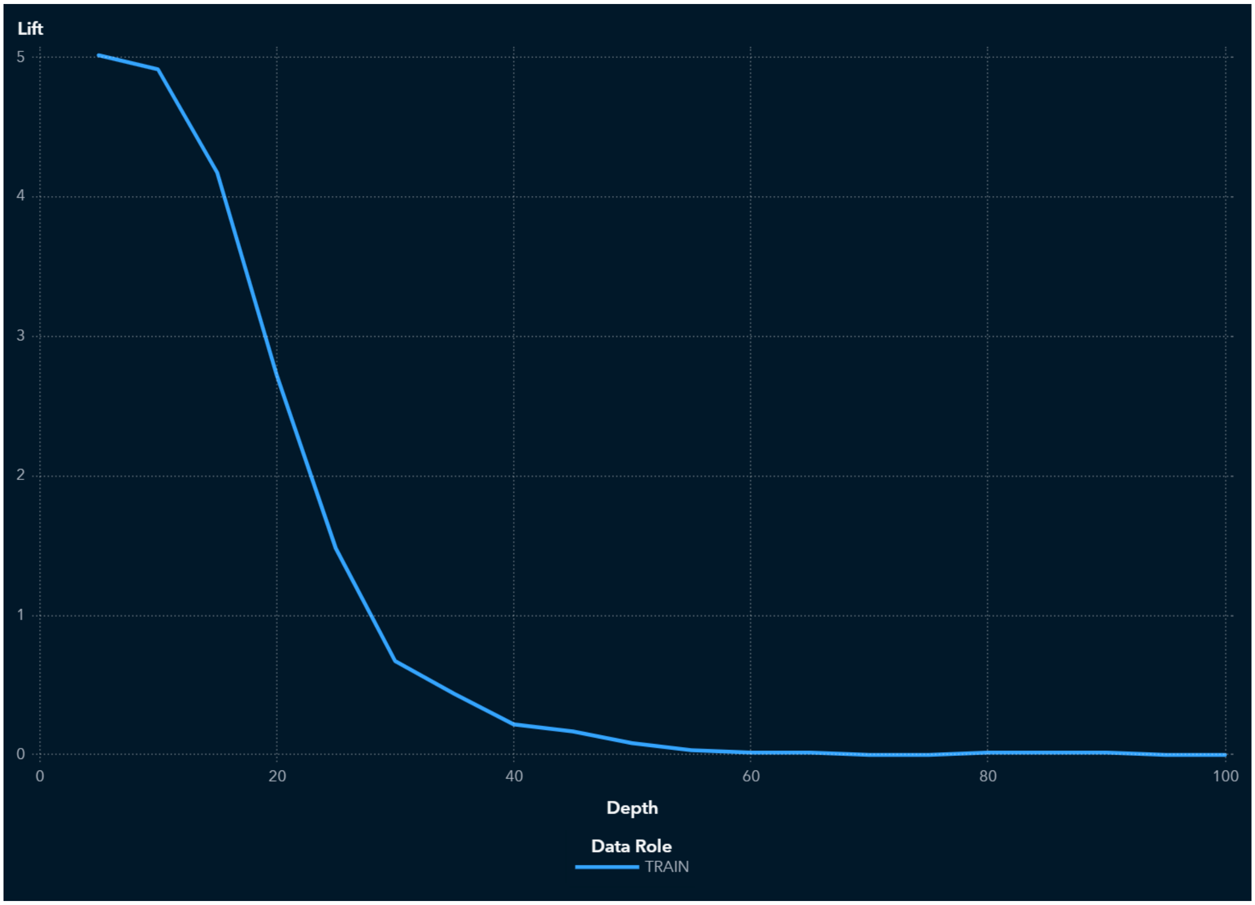
Cumulative Lift



The TRAIN partition has a Cumulative Lift of 4.96 in the 10% quantile (depth of 10) meaning there are 4.96 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

Cumulative lift is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. The cumulative lift for a particular quantile is the ratio of the number of events across all quantiles up to and including the current quantile to the number of events that would be there at random, or equivalently, the ratio of the cumulative response percentage to the baseline response percentage. The cumulative lift at depth 10 includes the top 10% of the data, which is the first 2 quantiles, which would have 10% of the events at random. Thus, cumulative lift measures how much more likely it is to observe an event in the quantiles than by selecting observations at random.

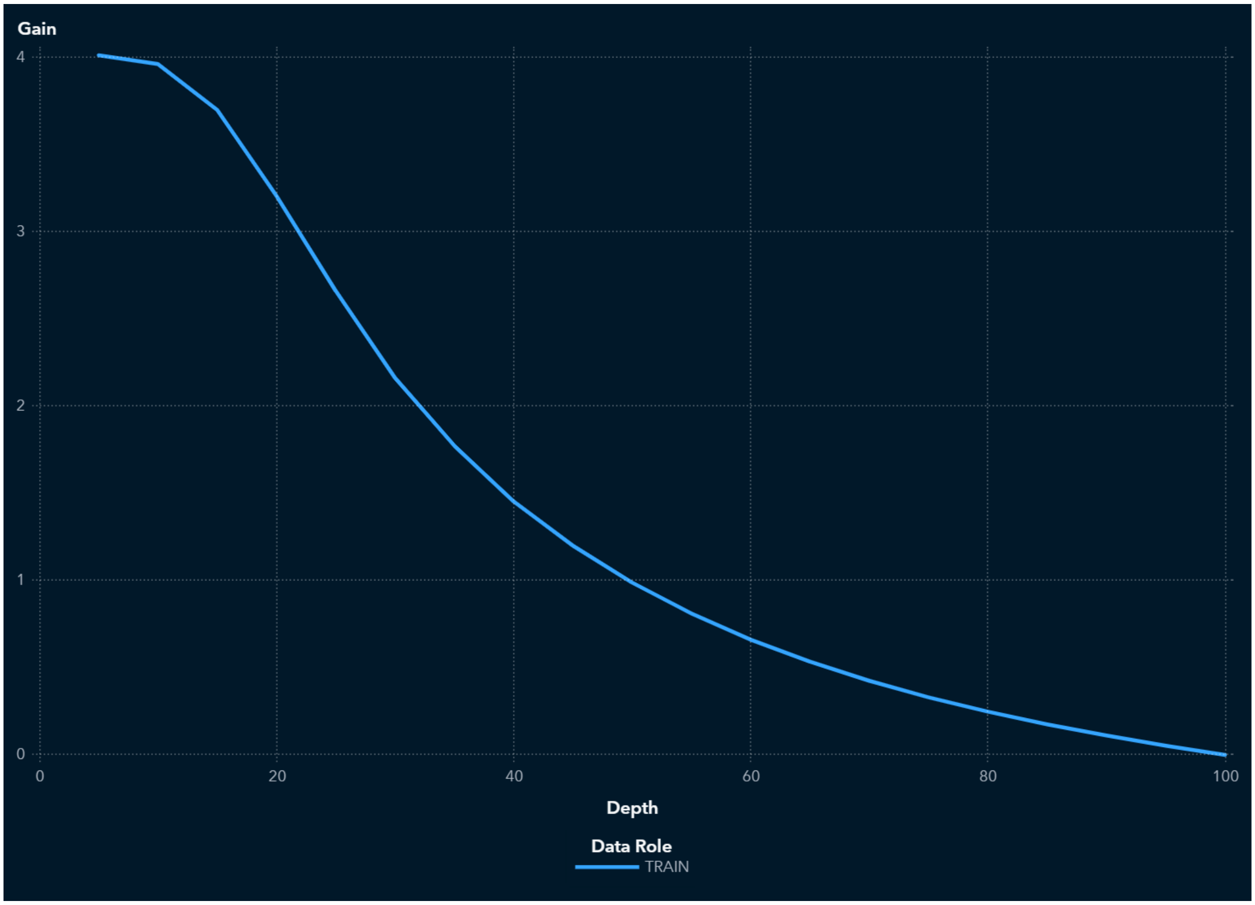
Lift



The TRAIN partition has a Lift of 5.01 in the 5% quantile (depth of 5) meaning there are 5.01 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

Lift is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Lift is the ratio of the number of events in that quantile to the number of events that would be there at random, or equivalently, the ratio of the response percentage to the baseline response percentage. With 20 quantiles, it is expected that 5% of the events occur in each quantile. Thus, Lift measures how much more likely it is to observe an event in each quantile than by selecting observations at random.

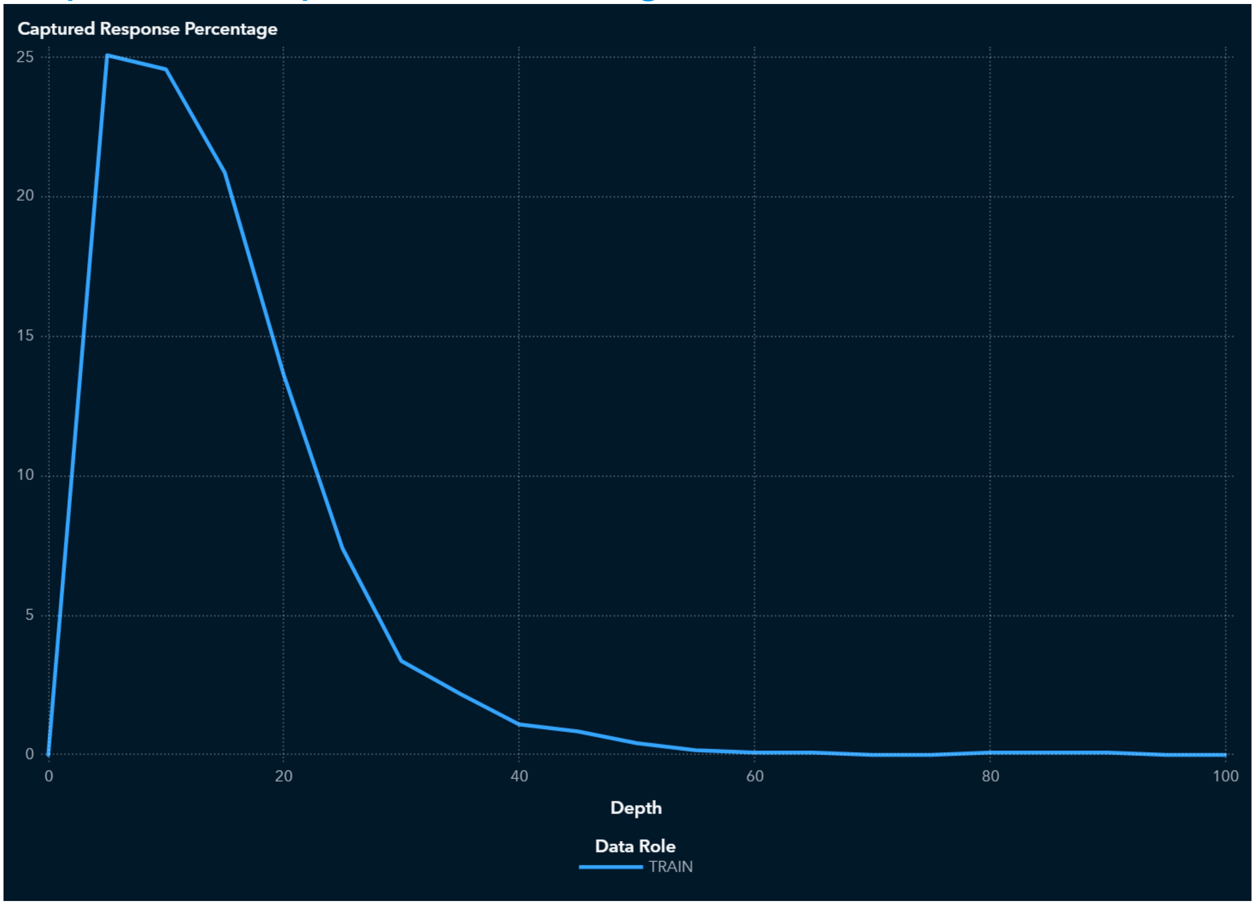
Gain



The TRAIN partition has a Gain of 4 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 4.01.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Gain is a cumulative measure for the quantiles up to an including the current one and is calculated as (number of events in the quantiles) / (number of events expected by random) - 1. With 20 quantiles, it is expected that 5% of the events occur in each quantile. Note that the value of Gain is the same as the value of Cumulative Lift - 1. If the value of Gain is greater than 0, then your model is better at identifying events than using no model.

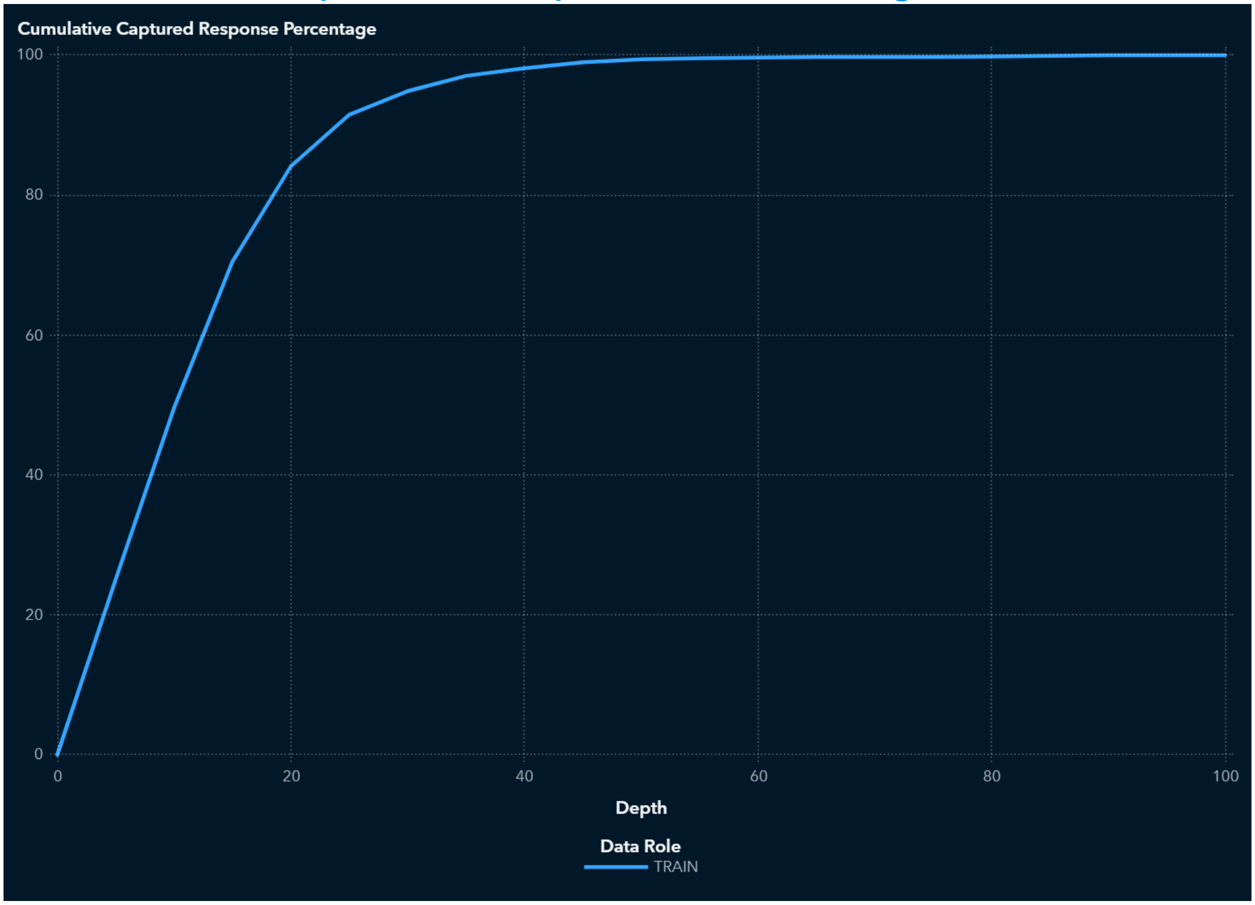
Captured Response Percentage



At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 25.1 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 25.06.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Captured response percentage is the percentage of the total number of events that are in that quantile. With no model, it is expected that 5% of the events are in each quantile.

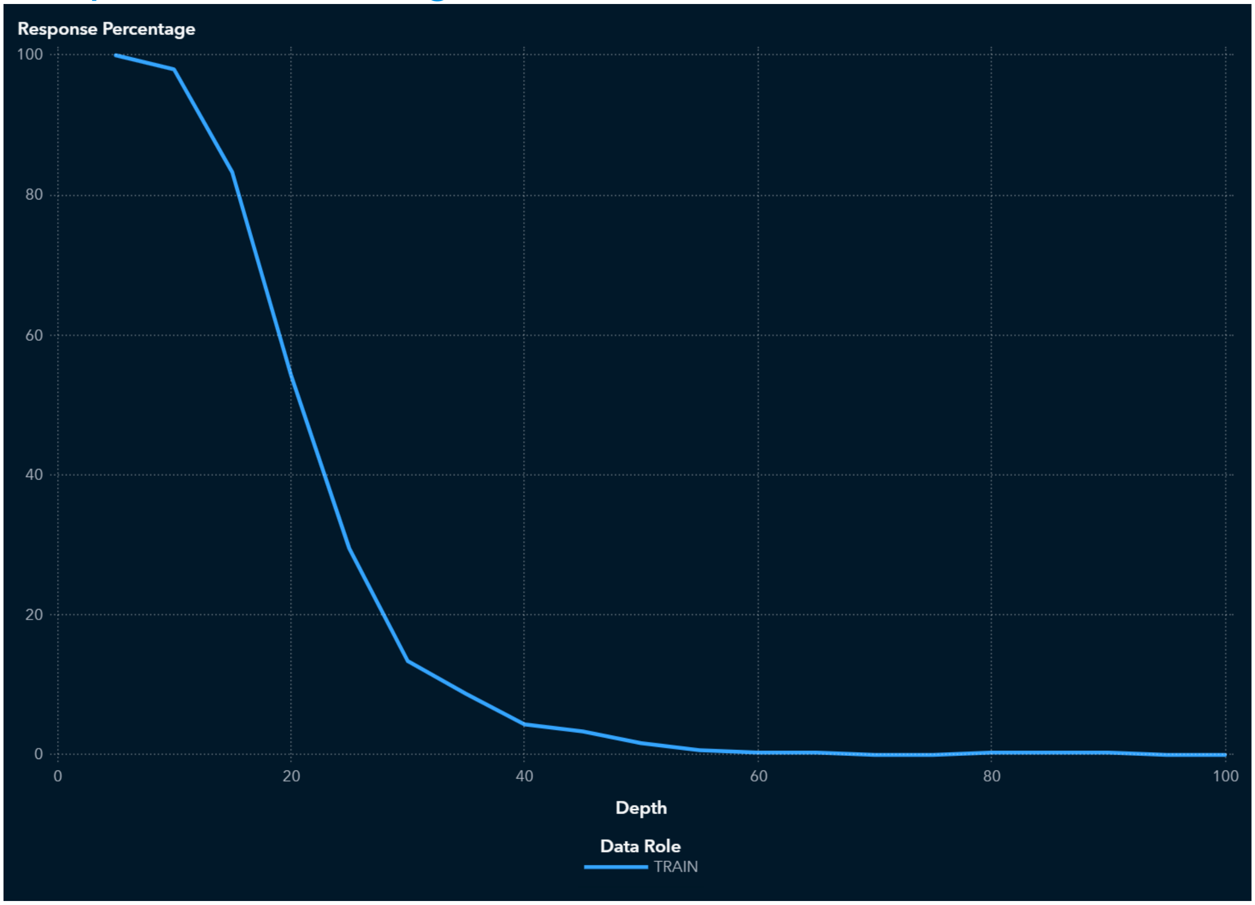
Cumulative Captured Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 49.6 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 50.13.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. The cumulative captured response percentage for a particular quantile is the percentage of the total number of events that are in the quantiles up to and including the current quantile. With no model, it is expected that 5% of the events are in each quantile, so the cumulative captured response percentage at depth 10 would be 10%.

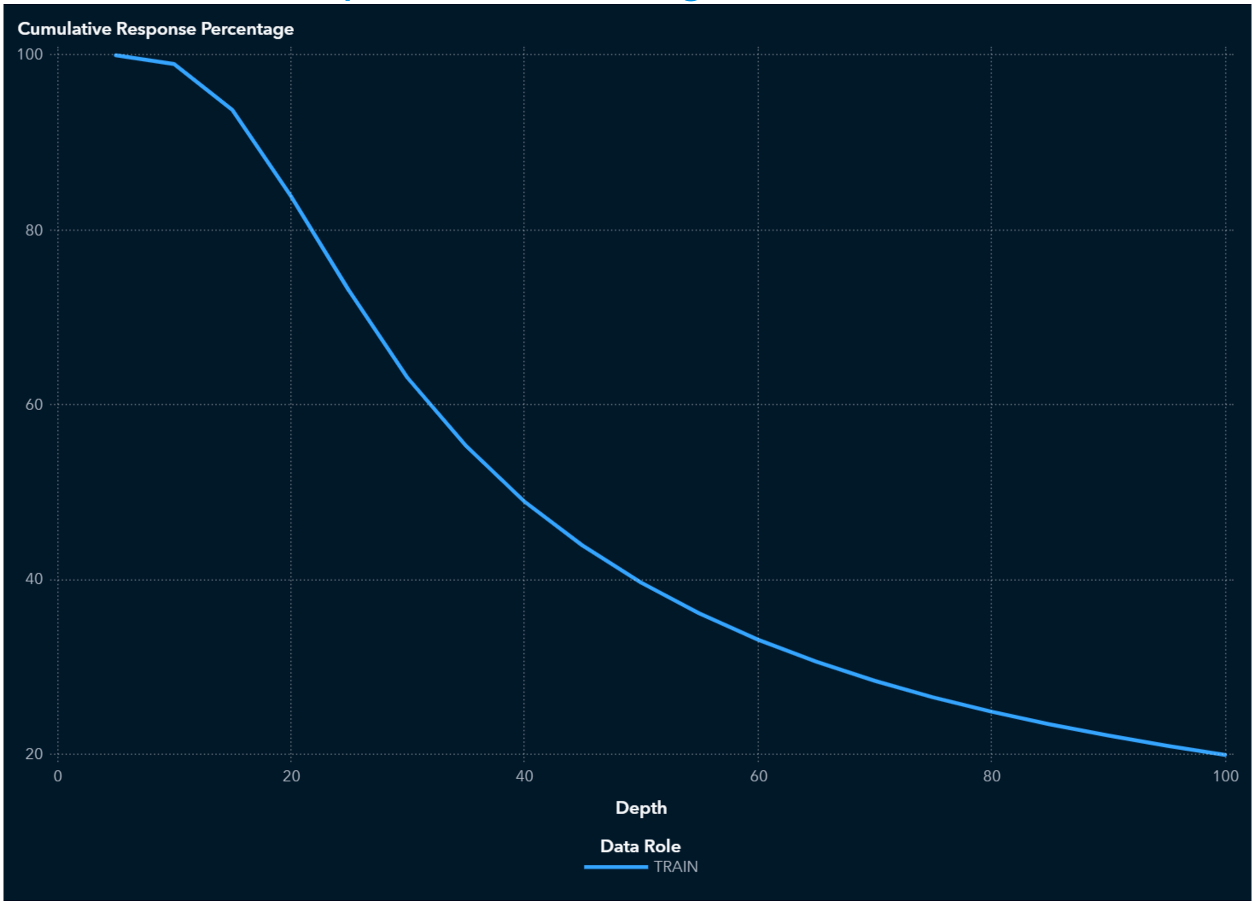
Response Percentage



At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 100. The best possible value of Response percentage for this partition at depth 5 is 100.

Response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Response percentage is the percentage of observations that are events in that quantile. With no model, it is expected that the response percentage is constant across quantiles, 100*overall-event-rate. This is also called the baseline response percentage.

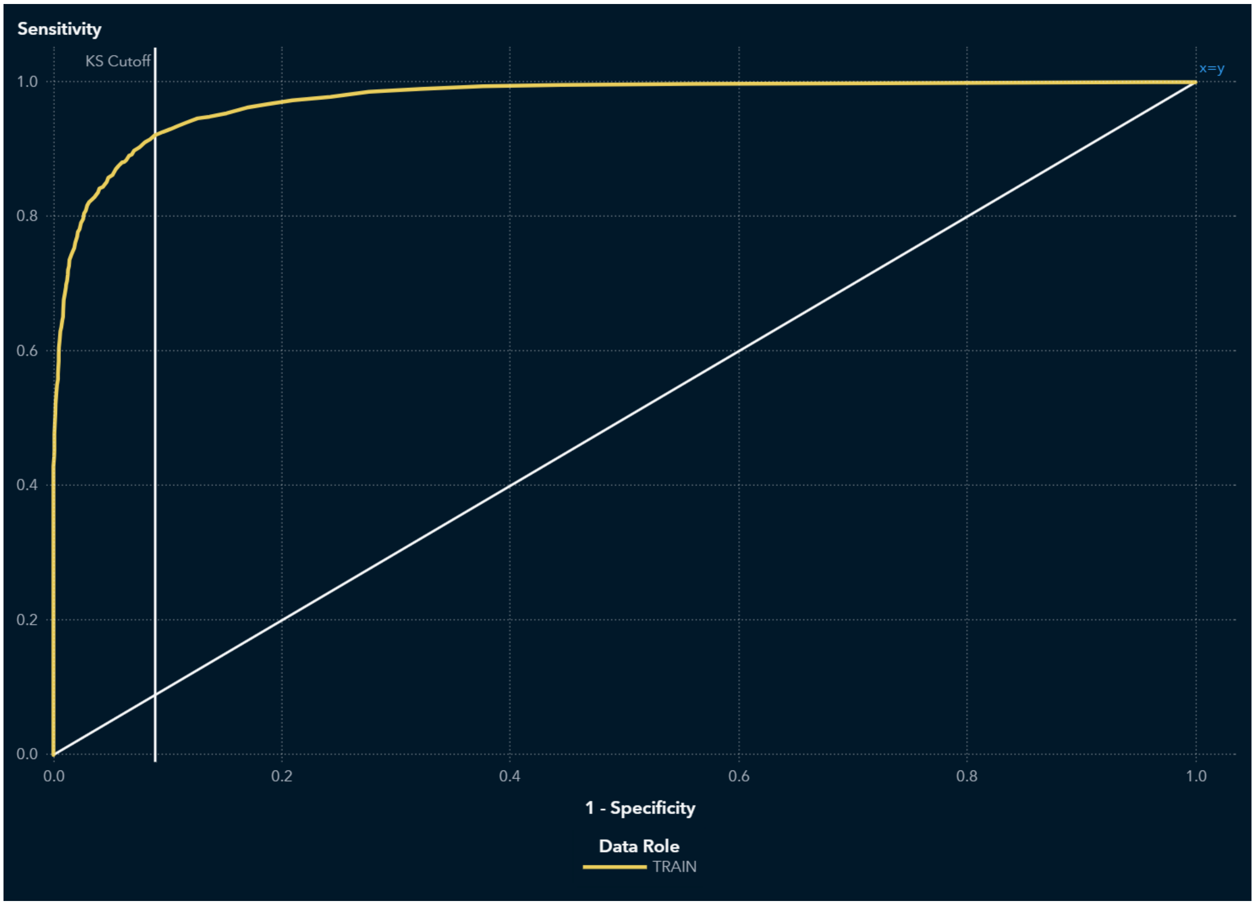
Cumulative Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 99. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

Cumulative response percentage is calculated by sorting in descending order each partition of the data by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. The cumulative response percentage for a particular quantile is the percentage of observations that are events in the quantiles up to and including the current quantile. With no model, it is expected that the response percentage is constant across quantiles, 100*overall-event-rate. This is also called the baseline response percentage.

ROC



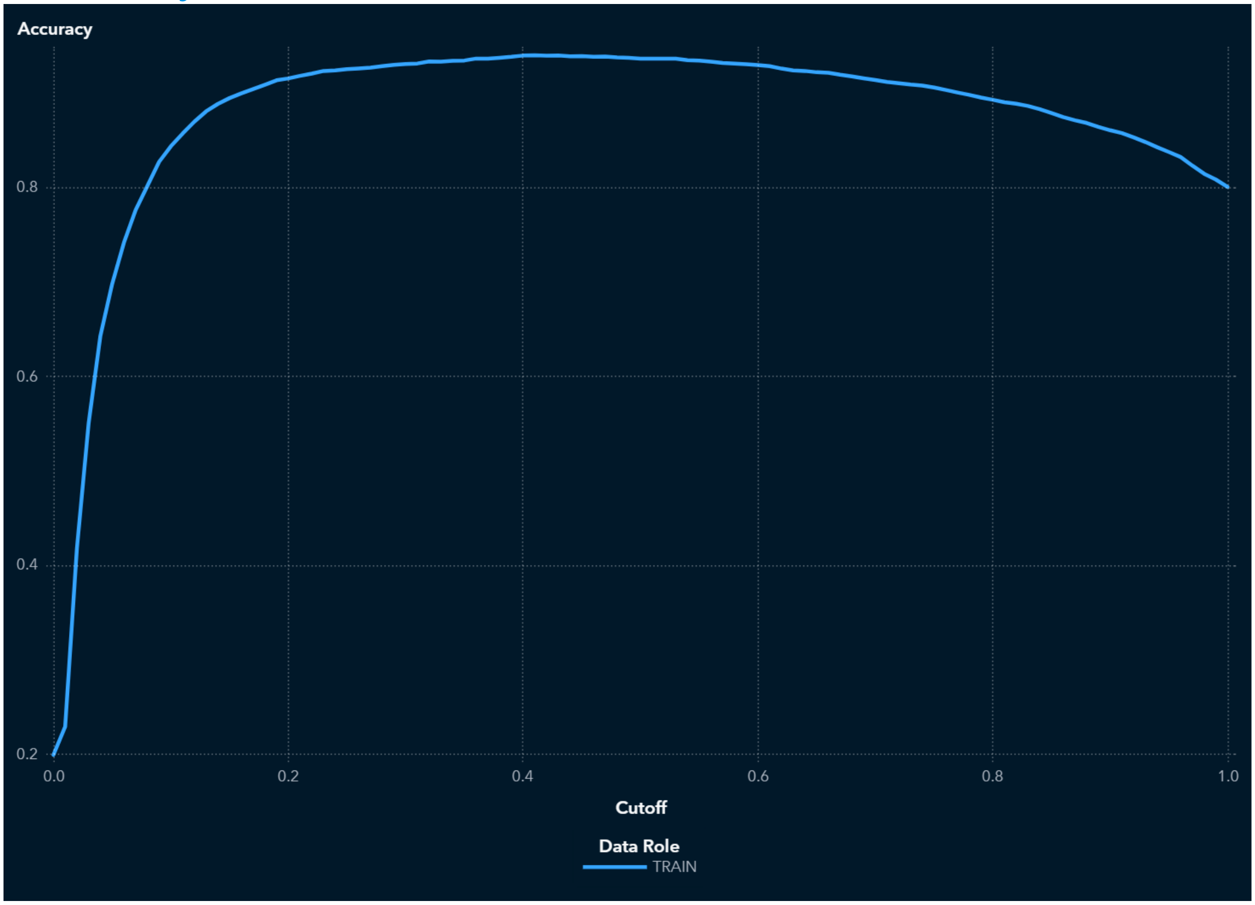
The ROC curve is a plot of sensitivity (the true positive rate) against 1-specificity (the false positive rate), which are both measures of classification based on the confusion matrix. These measures are calculated at various cutoff values. To help identify the best cutoff to use when scoring your data, the KS Cutoff reference line is drawn at the value of 1-specificity where the greatest difference between sensitivity and 1-specificity is observed for the TRAIN partition. The KS Cutoff line is drawn at the cutoff value 0.19, where the 1-specificity value is 0.089 and the sensitivity value is 0.921.

Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than or equal to the cutoff value, then the predicted classification is the event, otherwise it is a non-event.

The confusion matrix for each cutoff value contains four cells that display the true positives for events that are correctly classified (TP), false positives for non-events that are classified as events (FP), false negatives for events that are classified as non-events (FN), and true negatives for non-events that are classified as non-events (TN). True negatives include non-event classifications that specify a different non-event. Sensitivity is calculated as TP / (TP + FN). Specificity, the true negative rate, is calculated as TN / (TN + FP), so 1-specificity is FP / (TN + FP). The values of sensitivity and 1-specificity are plotted at each cutoff value.

A ROC curve that rapidly approaches the upper-left corner of the graph, where the difference between sensitivity and 1-specificity is the greatest, indicates a more accurate model. A diagonal line where sensitivity = 1-specificity indicates a random model.

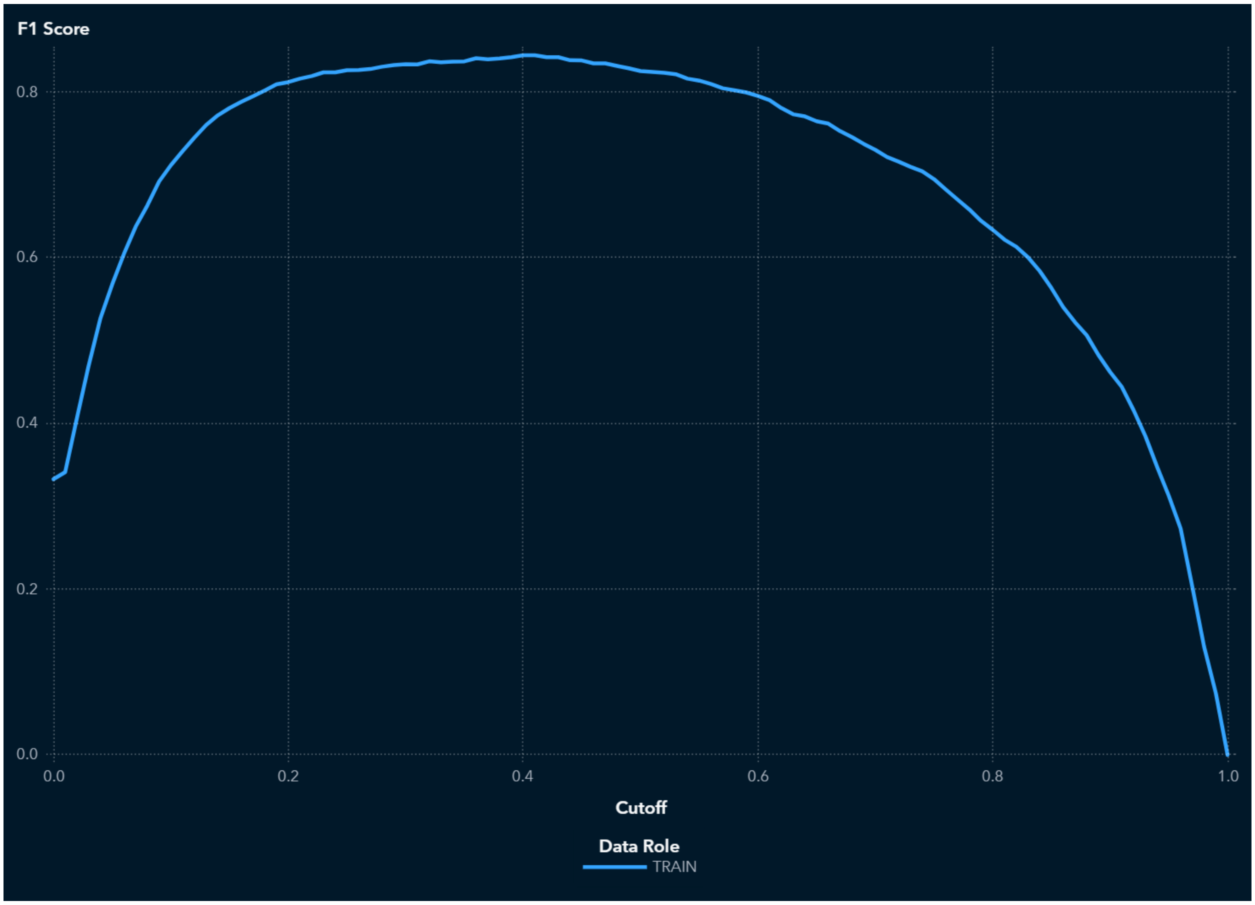
Accuracy



For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.936.

Accuracy is the proportion of observations that are correctly classified as either an event or non-event, calculated at various cutoff values. Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than or equal to the cutoff value, then the predicted classification is the event, otherwise it is a non-event. When the predicted classification and the actual classification are both events (true positives) or both non-events (true negatives), the observation is correctly classified. If the predicted classification and actual classification disagree, then the observation is incorrectly classified. Accuracy is calculated as (true positives + true negatives) / (total observations).

F1 Score



For this model, the F1 score in the TRAIN partition at the cutoff of 0.5 is 0.825.

The F1 score combines the measures of precision and recall (or sensitivity), which are measures of classification based on the confusion matrix that are calculated at various cutoff values. Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than or equal to the cutoff value, then the predicted classification is the event, otherwise it is a non-event.

The confusion matrix for each cutoff value contains four cells that display the true positives for events that are correctly classified (TP), false positives for non-events that are classified as events (FP), false negatives for events that are classified as non-events (FN), and true negatives for non-events that are classified as non-events (TN). True negatives include non-event classifications that specify a different non-event.

Precision is calculated as TP / (TP + FP), and recall (or sensitivity) is calculated as TP /

(TP + FN). The F1 score is calculated as 2*Precision*Recall / (Precision + Recall), which is the harmonic mean of Precision and Recall. Larger F1 scores indicate a more accurate model.

Fit Statistics

Target Name	Data Role	Number of Observations	Average Squared Error
BAD	TRAIN	5,960	0.0485

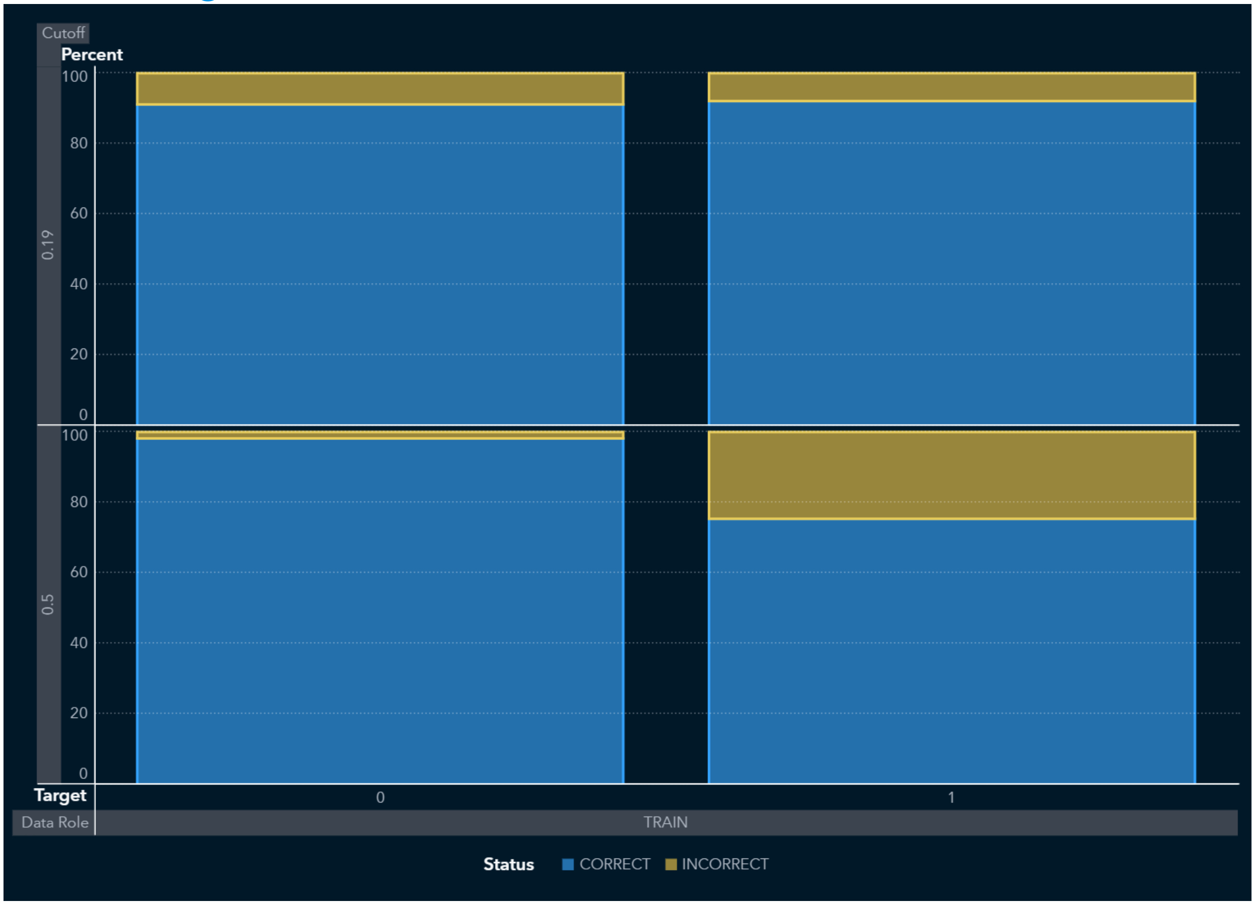
Divisor for ASE	Root Average Squared Error	Misclassification Rate	Multi-Class Log Loss
5,960	0.2202	0.0638	0.1744

KS (Youden)	Area Under ROC	Gini Coefficient	Gamma
0.8323	0.9743	0.9486	0.9513

Tau	KS Cutoff	KS at User-Specified Cutoff	Misclassification Rate at KS Cutoff (Event)
0.3030	0.1900	0.7353	0.0867

Misclassification Rate (Event)
0.0638

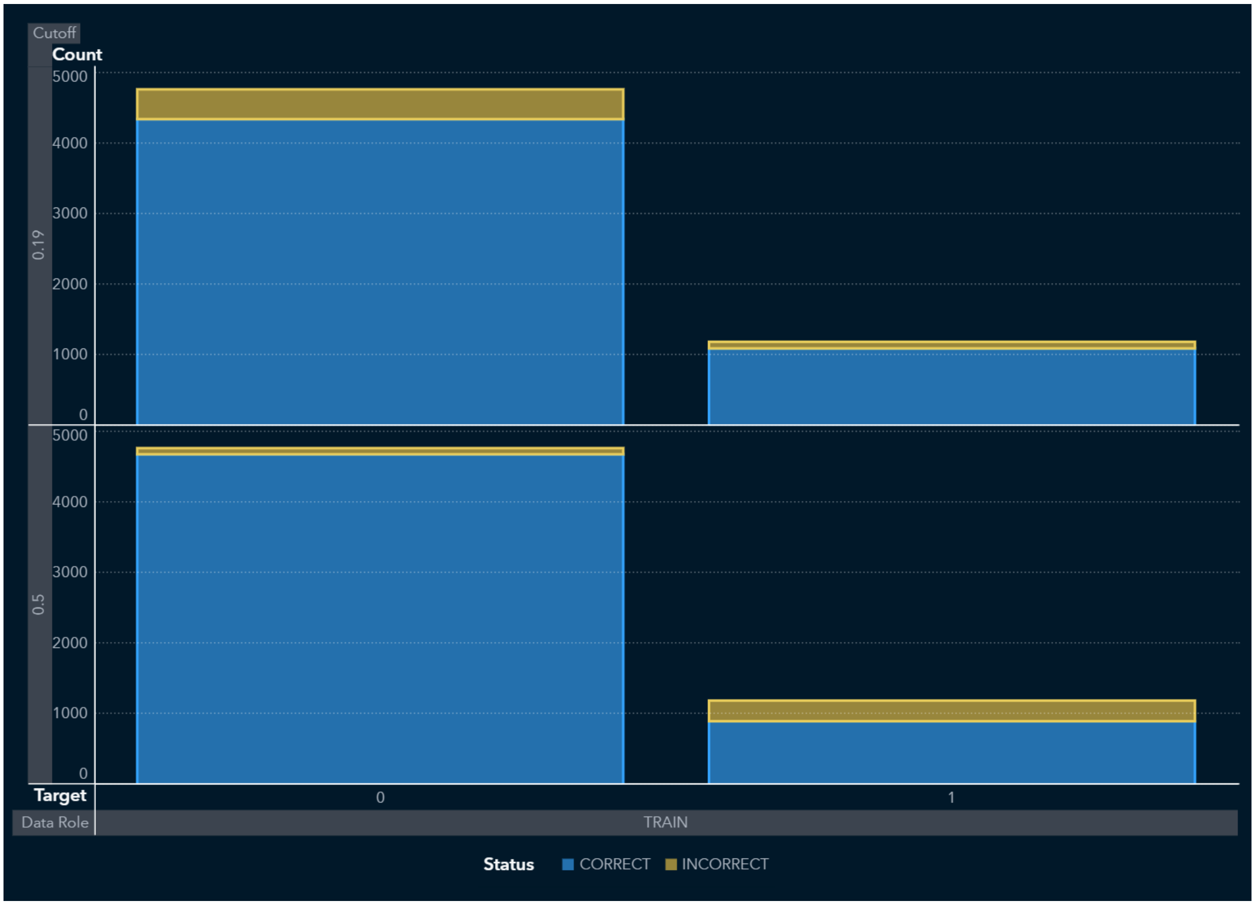
Percentage Plot



The Event Classification report is a visual representation of the confusion matrix at various cutoff values for each partition. The classification cutoffs used in the plot are the default (0.5) and the KS cutoff value 0.19 for the TRAIN partition.

For this data, for the bar corresponding to the event level of BAD, "   1", the segment of the bar colored as "CORRECT" corresponds to true positives.

Count Plot



The Event Classification report is a visual representation of the confusion matrix at various cutoff values for each partition. The classification cutoffs used in the plot are the default (0.5) and the KS cutoff value 0.19 for the TRAIN partition.

For this data, for the bar corresponding to the event level of BAD, "   1", the segment of the bar colored as "CORRECT" corresponds to true positives.

Table

Cutoff	Cutoff Source	Target Name	Response
0.1900	KS	BAD	CORRECT
0.1900	KS	BAD	INCORRECT
0.1900	KS	BAD	CORRECT
0.1900	KS	BAD	INCORRECT
0.5000	Default	BAD	CORRECT
0.5000	Default	BAD	INCORRECT
0.5000	Default	BAD	CORRECT
0.5000	Default	BAD	INCORRECT

Event	Value	Training Frequency	Validation Frequency
1	True Positive	1,095	
1	False Negative	94	
0	True Negative	4,348	
0	False Positive	423	
1	True Positive	896	
1	False Negative	293	
0	True Negative	4,684	
0	False Positive	87	

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	92.0942		
	7.9058		
	91.1339		
	8.8661		
	75.3574		
	24.6426		
	98.1765		

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	1.8235		

Properties

Property Name	Property Value
atAppendLookup	false
atCreateHistory	false
atHistoryLibUri	
atHistoryTblName	
atLeaveAutotuneOn	false
atLookupTableUri	
atMaxBayes	100
atMaxEval	50
atMaxIter	5
atMaxTime	60
atObjectiveInt	ASE
atObjectiveNom	KS
atPopSize	10
atSampleSize	50
atSearchMethod	GA
atTrainProp	0.7000
atUpdateProperties	false
atUseLookup	false
atValidFold	5
atValidMethod	PARTITION
atValidProp	0.3000
atbagFreqInitLgbm	0
atbagFreqLBLgbm	0
atbagFreqLgbm	true
atbagFreqUBLgbm	7
atbagPctInitLgbm	0.5000
atbagPctLBLgbm	0.2000
atbagPctLgbm	true

Property Name	Property Value
atbagPctUBLgbm	0.9500
atinputPctInitLgbm	1
atinputPctLBLgbm	0.1000
atinputPctLgbm	true
atinputPctUBLgbm	1
atintervalBins	true
atintervalBinsInit	50
atintervalBinsLB	20
atintervalBinsUB	100
atlasso	true
atlassoInit	0
atlassoLB	0
atlassoUB	10
atleafSize	false
atleafSizeInit	5
atleafSizeLB	1
atleafSizeUB	100
atlearnrt	true
atlearnrtInit	0.1000
atlearnrtLB	0.0100
atlearnrtUB	1
atmaxdepth	true
atmaxdepthInit	4
atmaxdepthLB	1
atmaxdepthUB	6
atntrees	true
atntreesInit	100
atntreesLB	20
atntreesUB	150

Property Name	Property Value
atridge	true
atridgeInit	1
atridgeLB	0
atridgeUB	10
atsamprt	true
atsamprtInit	0.5000
atsamprtLB	0.1000
atsamprtUB	1
atvarsToTry	true
atvarsToTryInit	100
atvarsToTryLB	1
atvarsToTryUB	100
autotune_enabled	false
bagFractionLgbm	0.5000
bagFreqLgbm	0
binaryProbCutoff	0.5000
boostingLgbm	GBDT
classDistrLgbm	MULTICLASS
codeLocation	mlearning
dataMiningVersion	V2024.03
defaultVarsPerTree	true
deterministicLgbm	false
distribution	GAUSSIAN
earlyStop	true
earlyStopMethod	STAGNATION
esMetric	MCR
esMinimum	false
esThreshold	0
esThresholdIter	0

Property Name	Property Value
exactPctlLift	true
explainFidelity	false
explainInfo	false
fullDatasetReconstitution	false
icePlots	false
inputFractionLgbm	1
intBinMethod	QUANTILE
intervalBins	50
intervalDistrLgbm	REGRESSION
lasso	0
learningRate	0.1000
lightGBM_enabled	false
maxBranch	2
maxCategories	128
maxDepth	4
maxNumShapVars	20
minLeafSize	5
minUseInSearch	1
missingLgbm	true
missingValue	USEINSEARCH
nBins	50
ntrees	100
pdNumImportantInputs	5
pdObsSamples	1,000
pdPlots	false
performKernelShap	false
performLime	false
performVI	false

Property Name	Property Value
power	1.5000
reportingOnly	false
ridge	1
seed	12,345
seedId	12,345
specifyRows	RANDOM
stagnation	5
subsampleRate	0.5000
templateRevision	5
tolerance	0
train	true
truncateLI	5
truncateUI	95
userProbCutoff	false
varsToTry	100

Output

The SAS System

The GRADBOOST Procedure

Model Information	
Number of Trees	100
Learning Rate	0.1
Subsampling Rate	0.5
Number of Variables Per Split	12
Number of Bins	50
Number of Input Variables	12
Maximum Number of Tree Nodes	31
Minimum Number of Tree Nodes	17
Maximum Number of Branches	2
Minimum Number of Branches	2
Maximum Depth	4
Minimum Depth	4
Maximum Number of Leaves	16
Minimum Number of Leaves	9
Maximum Leaf Size	2643
Minimum Leaf Size	5
Seed	12345
Lasso (L1) penalty	0
Ridge (L2) penalty	1
Actual Number of Trees	100
Average Number of Leaves	14.43

	Training
Number of Observations Read	5960
Number of Observations Used	5960

Variable Importance			
Variable	Importance	Std Dev Importance	Relative Importance
IM_DEBTINC	22.4423	65.4393	1.0000
DELINQ	6.5489	9.6621	0.2918
VALUE	6.2246	7.1337	0.2774
IM_CLAGE	5.8827	6.4780	0.2621
DEROG	5.0021	11.5203	0.2229
JOB	3.6054	3.3835	0.1607
IM_CLNO	3.1719	3.1344	0.1413
LOAN	2.9135	3.2797	0.1298
IM_YOJ	2.5767	2.8216	0.1148
NINQ	2.3565	2.7267	0.1050
IM_MORTDUE	2.3553	2.6627	0.1049
REASON	0.4329	1.1111	0.0193

Fit Statistics			
Number of Trees	Training Average Square Error	Training Misclassification Rate	Training Log Loss
1	0.1477	0.1995	0.464
2	0.1381	0.1995	0.438
3	0.1293	0.1995	0.415
4	0.1223	0.1995	0.396
5	0.1165	0.1995	0.382
6	0.1126	0.1891	0.371
7	0.1082	0.1738	0.360
8	0.1049	0.1544	0.351
9	0.1018	0.1383	0.343
10	0.0989	0.1314	0.335
11	0.0968	0.1250	0.329
12	0.0945	0.1220	0.323
13	0.0927	0.1176	0.317
14	0.0912	0.1171	0.313
15	0.0894	0.1134	0.308
16	0.0879	0.1107	0.303
17	0.0864	0.1099	0.299
18	0.0849	0.1094	0.294
19	0.0836	0.1076	0.290
20	0.0826	0.1042	0.286
21	0.0818	0.1044	0.283
22	0.0808	0.1040	0.280
23	0.0800	0.1012	0.278
24	0.0792	0.1020	0.275
25	0.0783	0.0993	0.272
26	0.0776	0.1003	0.270
27	0.0768	0.0992	0.267
28	0.0761	0.0987	0.265
29	0.0754	0.0977	0.263
30	0.0746	0.0971	0.260
31	0.0739	0.0966	0.258
32	0.0733	0.0948	0.256
33	0.0725	0.0938	0.254
34	0.0718	0.0935	0.251
35	0.0714	0.0923	0.250
36	0.0707	0.0908	0.248
37	0.0703	0.0908	0.246
38	0.0699	0.0899	0.245
39	0.0694	0.0899	0.243
40	0.0689	0.0898	0.241
41	0.0686	0.0896	0.240
42	0.0682	0.0884	0.239
43	0.0677	0.0878	0.237
44	0.0672	0.0883	0.235
45	0.0667	0.0879	0.234
46	0.0664	0.0874	0.232
47	0.0661	0.0878	0.231
48	0.0656	0.0876	0.230
49	0.0652	0.0872	0.229
50	0.0647	0.0861	0.227
51	0.0642	0.0846	0.225
52	0.0638	0.0844	0.224
53	0.0636	0.0841	0.223
54	0.0631	0.0837	0.221
55	0.0625	0.0829	0.219
56	0.0621	0.0832	0.218
57	0.0618	0.0819	0.217
58	0.0613	0.0810	0.215
59	0.0608	0.0802	0.214
60	0.0604	0.0800	0.213
61	0.0603	0.0800	0.212
62	0.0601	0.0797	0.211
63	0.0594	0.0792	0.209
64	0.0591	0.0795	0.208
65	0.0587	0.0789	0.207
66	0.0583	0.0777	0.206
67	0.0579	0.0765	0.205
68	0.0576	0.0750	0.204
69	0.0572	0.0742	0.203
70	0.0567	0.0750	0.201
71	0.0565	0.0752	0.200
72	0.0562	0.0748	0.199
73	0.0559	0.0743	0.198
74	0.0555	0.0730	0.197
75	0.0552	0.0725	0.196
76	0.0550	0.0727	0.195
77	0.0546	0.0723	0.194
78	0.0544	0.0728	0.193
79	0.0540	0.0713	0.192
80	0.0536	0.0706	0.191
81	0.0533	0.0711	0.190
82	0.0530	0.0711	0.189
83	0.0526	0.0708	0.187
84	0.0524	0.0698	0.187
85	0.0522	0.0695	0.186
86	0.0520	0.0690	0.185
87	0.0517	0.0680	0.185
88	0.0514	0.0686	0.184
89	0.0513	0.0678	0.183
90	0.0512	0.0681	0.182
91	0.0510	0.0683	0.182
92	0.0505	0.0663	0.180
93	0.0502	0.0656	0.179
94	0.0500	0.0654	0.179
95	0.0495	0.0646	0.178
96	0.0493	0.0644	0.177
97	0.0491	0.0646	0.176
98	0.0489	0.0638	0.176
99	0.0489	0.0636	0.175
100	0.0485	0.0638	0.174

Predicted Probability Variables	
BAD	Variable
1	P_BAD1
0	P_BAD0

Predicted Target Variable	
Level Index	Variable
	I_BAD