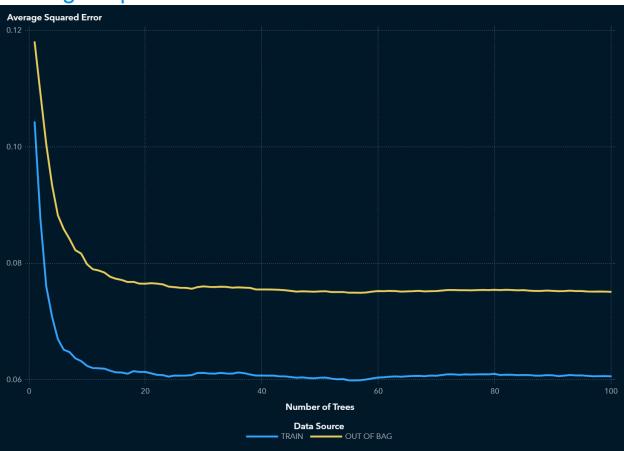


by: jbae7@ncsu.edu

Contents

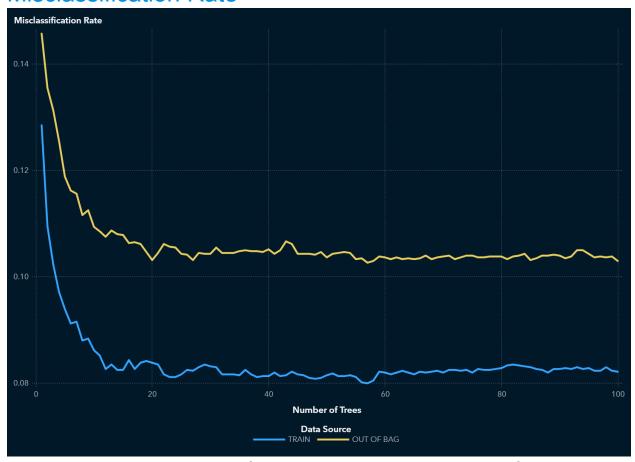
Average Squared Error	3
Misclassification Rate	4
Variable Importance	5
Score Inputs	6
Score Outputs	7
Cumulative Lift	9
Lift	10
Gain	11
Captured Response Percentage	12
Cumulative Captured Response Percentage	13
Response Percentage	14
Cumulative Response Percentage	15
ROC	16
Accuracy	18
F1 Score	19
Fit Statistics	21
Percentage Plot	22
Count Plot	23
Table	24
Properties	26
Output	30

Average Squared Error



This plot shows how the average squared error changes as the number of trees in the forest increases. The training error typically decreases as the number of trees increases, but the out-of-bag error gives you an indication of how well your model generalizes. For this model, the minimum error for the out-of-bag sample is 0.075 and occurs for 57 trees.

Misclassification Rate



This plot shows how the misclassification rate changes as the number of trees in the forest increases. The training error typically decreases as the number of trees increases, but the out-of-bag error gives you an indication of how well your model generalizes. For this model, the minimum error for the out-of-bag sample is 0.103 and occurs for 57 trees.

Variable Importance

Variable Label	Role	Variable Name	Training Importance
	INPUT	IM_DEBTINC	206.3396
	INPUT	VALUE	96.9240
	INPUT	DELINQ	82.1900
	INPUT	DEROG	61.2162
	INPUT	LOAN	46.4651
	INPUT	IM_CLAGE	44.3967
	INPUT	NINQ	37.2497
	INPUT	IM_CLNO	35.9392
	INPUT	JOB	28.0990
	INPUT	IM_MORTDUE	26.6277
	INPUT	IM_YOJ	19.2770
	INPUT	REASON	6.4995

Importance	Relative
Standard	Importance
Deviation	
67.5688	1
17.3223	0.4697
22.9028	0.3983
19.3261	0.2967
14.0975	0.2252
13.4295	0.2152
13.6989	0.1805
13.0849	0.1742
10.4693	0.1362
13.0085	0.1290
8.6298	0.0934
5.6535	0.0315

Score Inputs

Name	Role	Variable Level	Туре
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	С
LOAN	INPUT	INTERVAL	N
NINQ	INPUT	NOMINAL	N
REASON	INPUT	BINARY	С
VALUE	INPUT	INTERVAL	N

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

Score Outputs

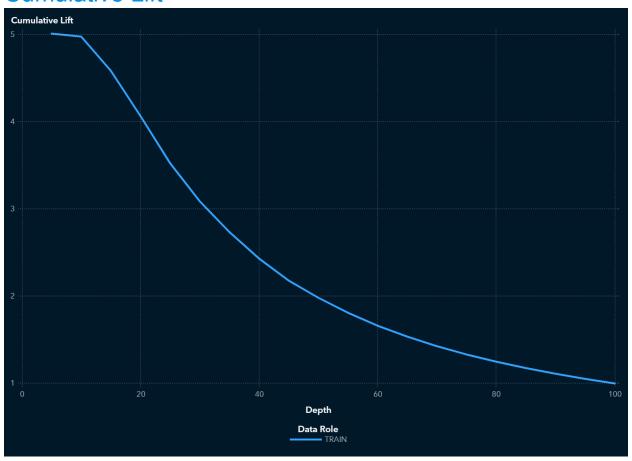
Name	Role	Туре	Variable Type
EM_CLASSIFICAT ION	CLASSIFICATION	С	char
EM_EVENTPROBAE	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
I_BAD	CLASSIFICATION	С	char
P_BAD0	PREDICT	N	double
P_BAD1	PREDICT	N	double
WARN	ASSESS	С	char

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

Function	Creator GUID
CLASSIFICATION	6de51889-3c65-40c 9-b5da- a333df38325f
PREDICT	6de51889-3c65-40c 9-b5da- a333df38325f
PREDICT	6de51889-3c65-40c 9-b5da- a333df38325f

Function	Creator GUID
CLASSIFICATION	6de51889-3c65-40c 9-b5da- a333df38325f
PREDICT	6de51889-3c65-40c 9-b5da- a333df38325f
PREDICT	6de51889-3c65-40c 9-b5da- a333df38325f
ASSESS	6de51889-3c65-40c 9-b5da- a333df38325f

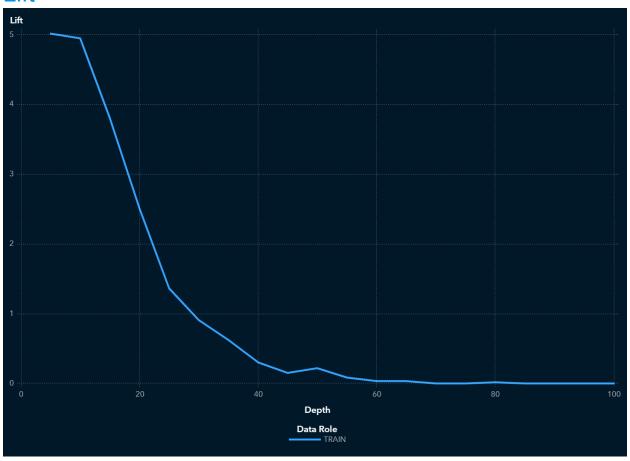
Cumulative Lift



The TRAIN partition has a Cumulative Lift of 4.98 in the 10% quantile (depth of 10) meaning there are 4.98 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.

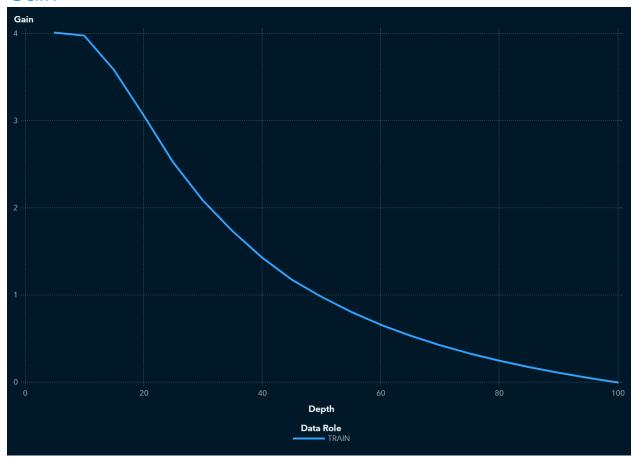
Lift



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.

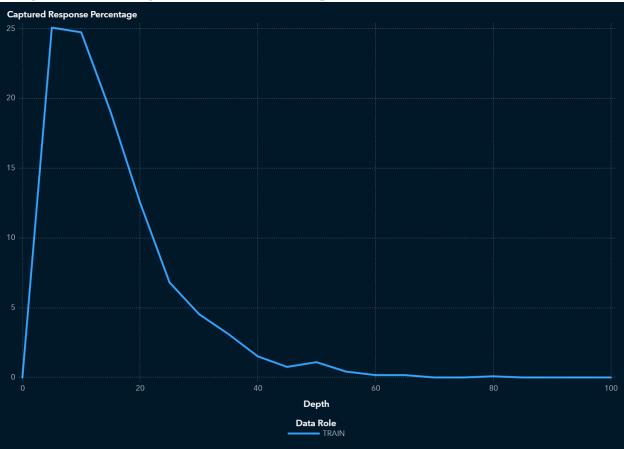
Gain



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.

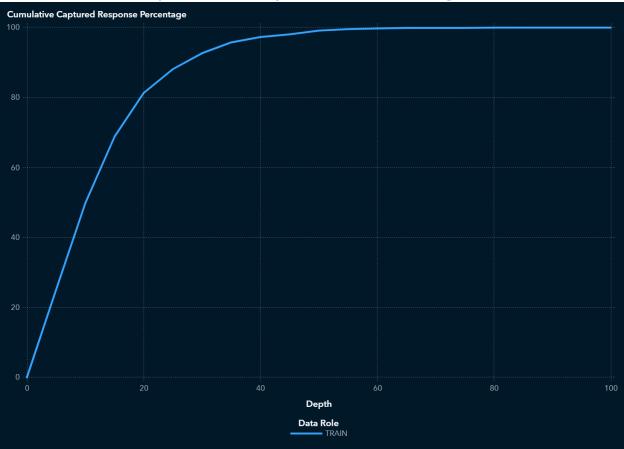
Captured Response Percentage



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.

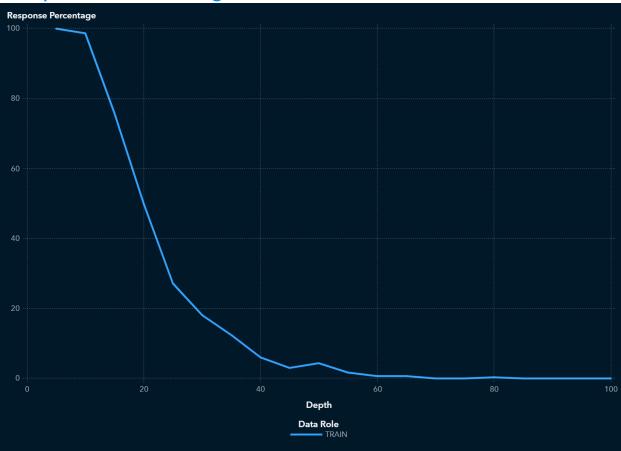
Cumulative Captured Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 49.8 (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%.

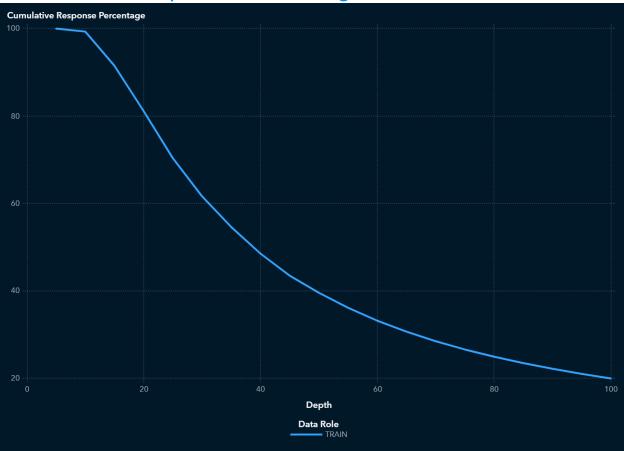
Response Percentage



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.

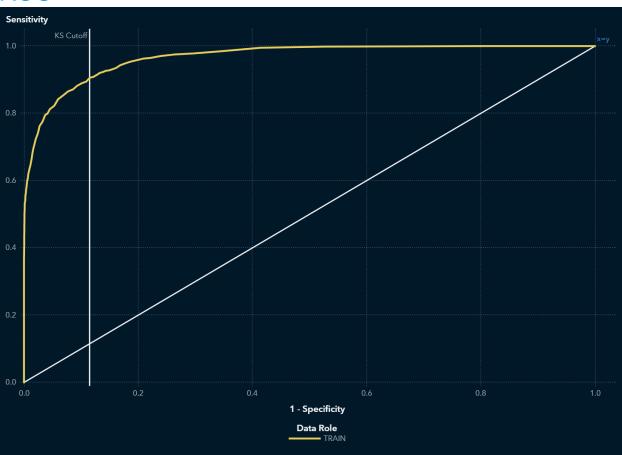
Cumulative Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 99.3. 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 0.25, where the 1-specificity value is 0.115 and the sensitivity value is 0.906.

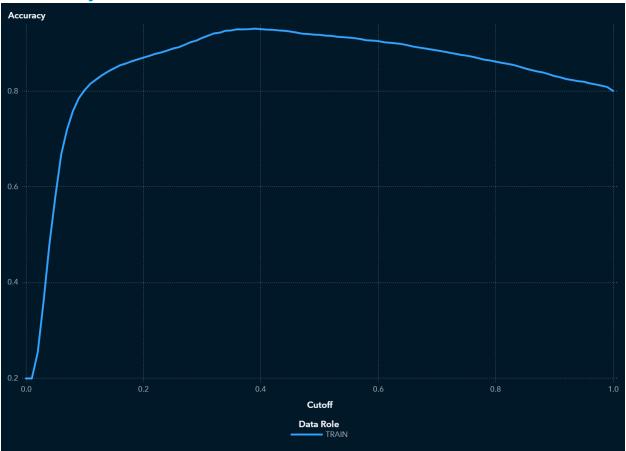
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than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

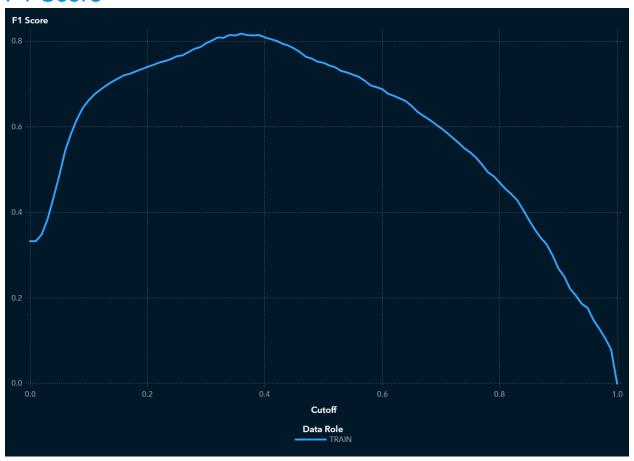
Accuracy



For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.918.

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 0.75.

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 /

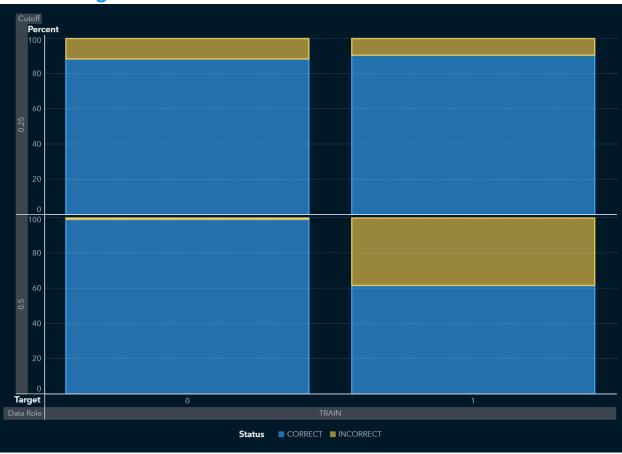
(TP + FN). The F1 score is calculated as 2*Precision*Recall / (Precision + Recall), which is the harmonic mean of Precision and Recall. Larger F1 scores indicate a more accurate model.

Fit Statistics

Target Name	Data Role	Number of Observations	Average Squared Error
BAD	TRAIN	5,960	0.0606
5 (105	<u> </u>	1.4. 1. 16. 1	
Divisor for ASE	Root Average	Misclassification	Multi-Class Log
	Squared Error	Rate	Loss
5,960	0.2461	0.0822	0.2136
		T	·
KS (Youden)	Area Under ROC	Gini Coefficient	Gamma
0.7905	0.9667	0.9334	0.9368
	•	·	·
Tau	KS Cutoff	KS at User-	Misclassification
		Specified Cutoff	Rate at KS Cutoff
			(Event)
0.2982	0.2500	0.6100	0.1111

Misclassification
Rate (Event)
0.0822

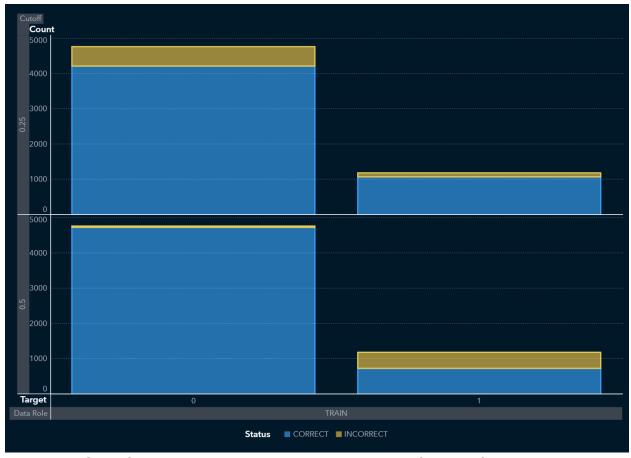
Percentage Plot



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 0.25 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 0.25 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
0.2500	KS	BAD	CORRECT
0.2500	KS	BAD	INCORRECT
0.2500	KS	BAD	CORRECT
0.2500	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	1,077	
1	False Negative	112	
0	True Negative	4,221	
0	False Positive	550	
1	True Positive	734	
1	False Negative	455	
0	True Negative	4,736	
0	False Positive	35	

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	90.5803		
	9.4197		
	88.4720		
	11.5280		
	61.7325		
	38.2675		
	99.2664		

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	0.7336		

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
atintervalBins	true
atintervalBinsInit	50
atintervalBinsLB	20
atintervalBinsUB	100
atleafSize	false
atleafSizeInit	5
atleafSizeLB	1

Property Name	Property Value
atleafSizeUB	100
atmaxDepth	true
atmaxDepthInit	20
atmaxDepthLB	1
atmaxDepthUB	29
atmaxTrees	true
atmaxTreesInit	100
atmaxTreesLB	20
atmaxTreesUB	150
attrainFraction	true
attrainFractionInit	0.6000
attrainFractionLB	0.1000
attrainFractionUB	0.9000
atvarsToTry	true
atvarsToTryInit	100
atvarsToTryLB	1
atvarsToTryUB	100
autotune_enabled	false
binaryProbCutoff	0.5000
codeLocation	mlearning
criterionMethod	IGR
dataMiningVersion	V2024.03
defaultVarsPerTree	true
exactPctlLift	true
explainFidelity	false
explainInfo	false
fullDatasetReconstit ution	false
iCriterionMethod	VARIANCE

Property Name	Property Value
icePlots	false
intBinMethod	QUANTILE
intervalBins	50
leafProp	0.0001
leafSize	5
leafSpec	COUNT
loh	0
maxBranch	2
maxCategories	128
maxDepth	20
maxNumShapVars	20
maxTrees	100
minUseInSearch	1
missingValue	USEINSEARCH
nBins	50
pdNumImportantInp uts	5
pdObsSamples	1,000
pdPlots	false
performKernelShap	false
performLime	false
performVI	false
reportingOnly	false
seed	12,345
seedId	12,345
specifyRows	RANDOM
templateRevision	4
train	true
trainFraction	0.6000

Property Name	Property Value
truncateLl	5
truncateUl	95
userProbCutoff	false
varsToTry	100
voteMethod	PROBABILITY

Output

