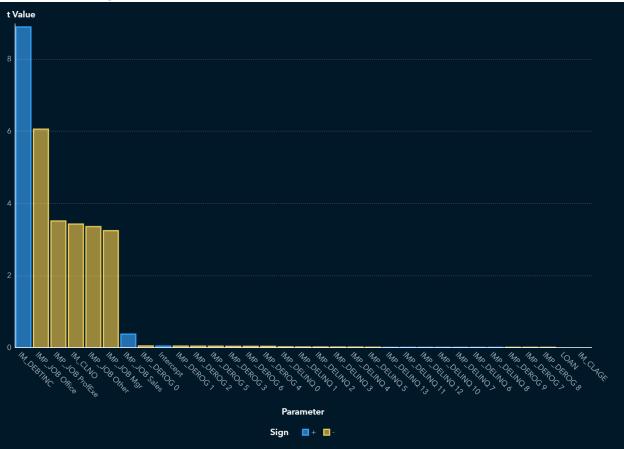


by: jbae7@ncsu.edu

Contents

t Values by Parameter	3
Parameter Estimates	4
Selection Summary	8
Regression Fit Statistics	9
Score Inputs	10
Score Outputs	11
Cumulative Lift	14
Lift	15
Gain	16
Captured Response Percentage	17
Cumulative Captured Response Percentage	18
Response Percentage	19
Cumulative Response Percentage	20
ROC	21
Accuracy	23
F1 Score	24
Fit Statistics	26
Percentage Plot	27
Count Plot	28
Table	29
Properties	31
Output	33

t Values by Parameter



This plot displays the absolute value of the t value for each parameter estimate in the logistic regression model. Larger values indicate more significant parameters. The bar that represents the parameter is colored by the sign of the estimate. Bars that are colored as positive (+) correspond to a positive parameter estimate, which indicates an increase in the predicted probability of the event as the parameter value increases. Bars that are colored as negative (-) correspond to a negative parameter estimate, which indicates a decrease in the predicted probability of event as the parameter value increases. The most significant parameter is IM_DEBTINC with a t value of 8.889.

Parameter Estimates

Effect	Parameter	t Value	Sign
IM_DEBTINC	IM_DEBTINC	8.8886	+
IMP_JOB	IMP_JOB Office	6.0592	-
IMP_JOB	IMP_JOB ProfExe	3.5162	-
IM_CLNO	IM_CLNO	3.4311	-
IMP_JOB	IMP_JOB Other	3.3603	-
IMP_JOB	IMP_JOB Mgr	3.2482	-
IMP_JOB	IMP_JOB Sales	0.3837	+
IMP_DEROG	IMP_DEROG 0	0.0557	-
Intercept	Intercept	0.0525	+
IMP_DEROG	IMP_DEROG 1	0.0519	-
IMP_DEROG	IMP_DEROG 2	0.0504	-
IMP_DEROG	IMP_DEROG 5	0.0493	-
IMP_DEROG	IMP_DEROG 3	0.0462	-
IMP_DEROG	IMP_DEROG 6	0.0462	-
IMP_DEROG	IMP_DEROG 4	0.0450	-
IMP_DELINQ	IMP_DELINQ 0	0.0311	-
IMP_DELINQ	IMP_DELINQ 1	0.0279	-
IMP_DELINQ	IMP_DELINQ 2	0.0273	-
IMP_DELINQ	IMP_DELINQ 3	0.0259	-
IMP_DELINQ	IMP_DELINQ 4	0.0246	-
IMP_DELINQ	IMP_DELINQ 5	0.0206	-
IMP_DELINQ	IMP_DELINQ 13	0.0065	+
IMP_DELINQ	IMP_DELINQ 11	0.0063	+
IMP_DELINQ	IMP_DELINQ 12	0.0060	+
IMP_DELINQ	IMP_DELINQ 10	0.0044	+
IMP_DELINQ	IMP_DELINQ 7	0.0044	+
IMP_DELINQ	IMP_DELINQ 6	0.0043	+
IMP_DELINQ	IMP_DELINQ 8	0.0042	+

Effect	Parameter	t Value	Sign
IMP_DEROG	IMP_DEROG 9	0.0033	-
IMP_DEROG	IMP_DEROG 7	0.0023	-
IMP_DEROG	IMP_DEROG 8	0.0008	-
IMP_DEROG	IMP_DEROG 10		+
LOAN	LOAN		-
IM_CLAGE	IM_CLAGE		-
IMP_DELINQ	IMP_DELINQ 15		+
IMP_JOB	IMP_JOB Self		+

Estimate	Absolute Estimate	Standard Error	Chi-Square
0.0537	0.0537	0.0060	79.0075
-1.3222	1.3222	0.2182	36.7135
-0.7282	0.7282	0.2071	12.3637
-0.0144	0.0144	0.0042	11.7722
-0.6588	0.6588	0.1960	11.2917
-0.6914	0.6914	0.2129	10.5507
0.1175	0.1175	0.3061	0.1472
-13.3768	13.3768	240.3652	0.0031
23.7916	23.7916	453.2066	0.0028
-12.4722	12.4722	240.3652	0.0027
-12.1215	12.1215	240.3652	0.0025
-11.8504	11.8504	240.3658	0.0024
-11.1055	11.1055	240.3654	0.0021
-11.0948	11.0948	240.3659	0.0021
-10.8240	10.8240	240.3658	0.0020
-11.9384	11.9384	384.2145	0.0010
-10.7349	10.7349	384.2145	0.0008
-10.4981	10.4981	384.2146	0.0007
-9.9430	9.9430	384.2146	0.0007
-9.4699	9.4699	384.2146	0.0006

Estimate	Absolute Estimate	Standard Error	Chi-Square
-7.9038	7.9038	384.2148	0.0004
3.5132	3.5132	543.3614	0.0000
2.9385	2.9385	463.4176	0.0000
3.2732	3.2732	543.3614	0.0000
2.0916	2.0916	470.5377	0.0000
1.7303	1.7303	397.0008	0.0000
1.6625	1.6625	389.2518	0.0000
1.7588	1.7588	420.2692	0.0000
-1.0187	1.0187	312.6217	0.0000
-0.5903	0.5903	260.7527	0.0000
-0.2271	0.2271	278.1948	0.0000
0	0		
0.0000	0.0000		
-0.0061	0.0061		
0	0		
0	0		

Pr > Chi-Square	Degrees of Freedom
0.0000	1
0.0000	1
0.0004	1
0.0006	1
0.0008	1
0.0012	1
0.7012	1
0.9556	1
0.9581	1
0.9586	1
0.9598	1

Pr > Chi-Square	Degrees of
	Freedom
0.9607	1
0.9631	1
0.9632	1
0.9641	1
0.9752	1
0.9777	1
0.9782	1
0.9794	1
0.9803	1
0.9836	1
0.9948	1
0.9949	1
0.9952	1
0.9965	1
0.9965	1
0.9966	1
0.9967	1
0.9974	1
0.9982	1
0.9993	1
	0
	0
	0
	0
	0

Selection Summary

Step	Effect Entered	Number of Effects	SBC
0	Intercept	1	5,965.1625
1	IMP_DELINQ	2	5,314.3655
2	IMP_DEROG	3	5,235.5867
3	IM_CLAGE	4	5,099.4411
4	IM_DEBTINC	5	4,996.3894
5	LOAN	6	4,980.1606
6	IMP_JOB	7	4,962.6008
7	IM_CLNO	8	4,950.7771

Optimal SBC
0
0
0
0
0
0
0
1

Regression Fit Statistics

Statistic	Description	Value
M2LL	-2 Log Likelihood	4,681.1655
AIC	AIC (smaller is better)	4,743.1655
AICC	AICC (smaller is better)	4,743.5002
SBC	SBC (smaller is better)	4,950.6431

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
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
IMP_DELINQ	INPUT	N	double
IMP_DEROG	INPUT	N	double
IMP_JOB	INPUT	С	char
IMP_NINQ	INPUT	N	double
IMP_REASON	INPUT	С	char
IMP_VALUE	REJECTED	N	double
I_BAD	CLASSIFICATION	С	char
P_BAD0	PREDICT	N	double
P_BAD1	PREDICT	N	double

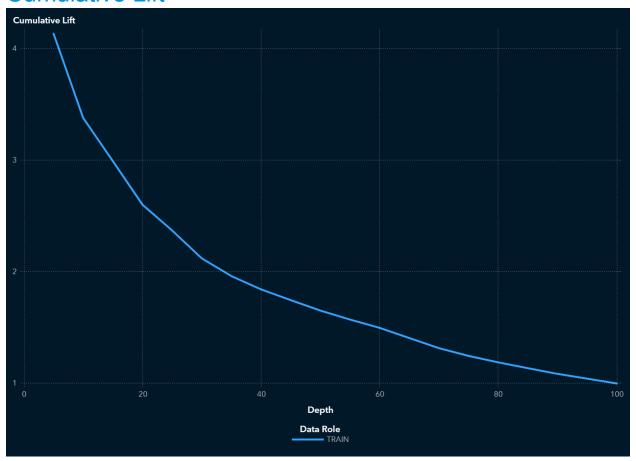
Variable Label	Variable Format	Variable Length	Creator
Predicted for BAD		12	logisticreg
Probability for BAD=1		8	logisticreg
Probability of Classification		8	logisticreg
Imputed DELINQ		8	impute
Imputed DEROG		8	impute
Imputed JOB		7	impute
Imputed NINQ		8	impute
Imputed REASON		7	impute
Imputed VALUE		8	impute
Into: BAD		12	logisticreg
Predicted: BAD=0		8	logisticreg

Variable Label	Variable Format	Variable Length	Creator
Predicted: BAD=1		8	logisticreg

Function	Creator GUID
CLASSIFICATION	fbbe14ac-4e1c-4b6 2-
	a472-0edafbf6f3f2
PREDICT	fbbe14ac-4e1c-4b6 2-
	a472-0edafbf6f3f2
PREDICT	fbbe14ac-4e1c-4b6 2-
	a472-0edafbf6f3f2
TRANSFORM	b727c11b-1371-4c 1d-8746-51b94c23 7069
CLASSIFICATION	fbbe14ac-4e1c-4b6 2-
	a472-0edafbf6f3f2
PREDICT	fbbe14ac-4e1c-4b6 2-

Function	Creator GUID
	a472-0edafbf6f3f2
PREDICT	fbbe14ac-4e1c-4b6 2-
	a472-0edafbf6f3f2

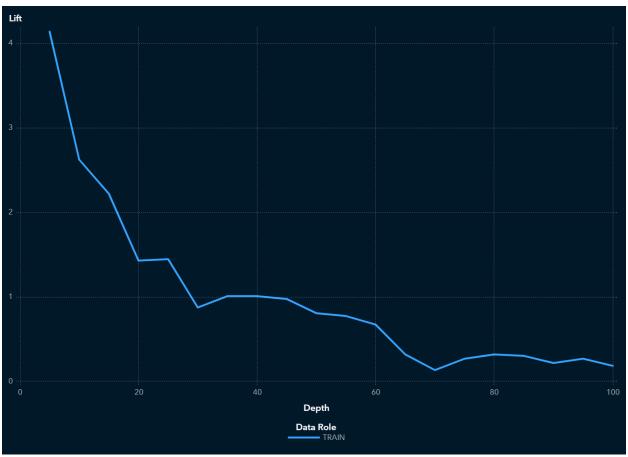
Cumulative Lift



The TRAIN partition has a Cumulative Lift of 3.38 in the 10% quantile (depth of 10) meaning there are 3.38 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.

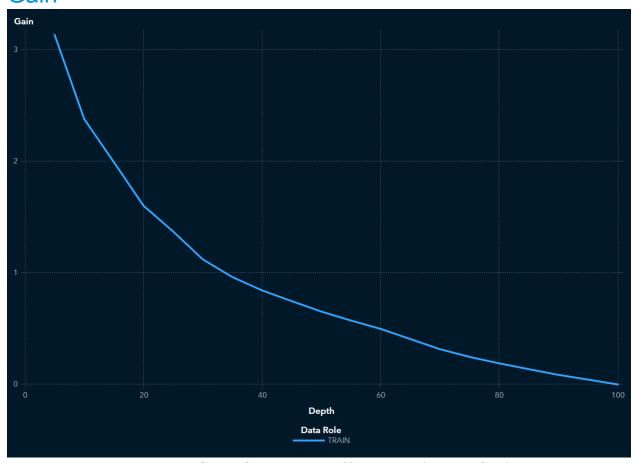
Lift



The TRAIN partition has a Lift of 4.14 in the 5% quantile (depth of 5) meaning there are 4.14 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.

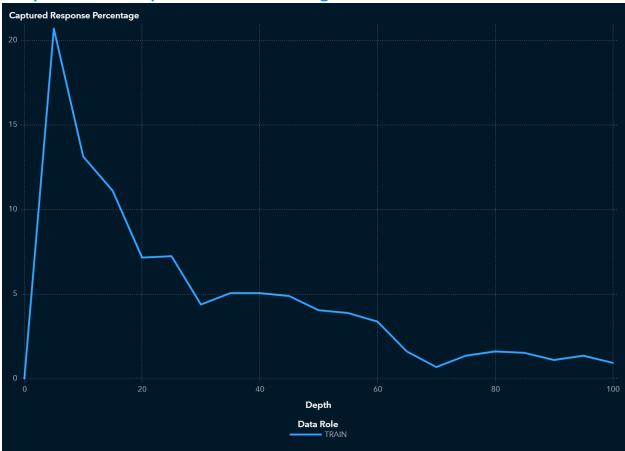
Gain



The TRAIN partition has a Gain of 2.4 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.

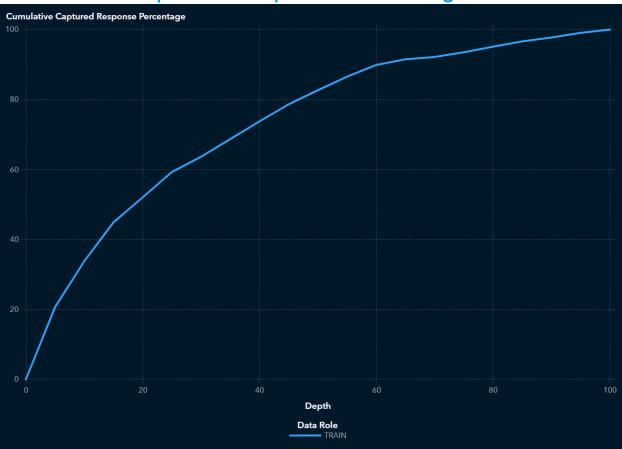
Captured Response Percentage



At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 20.7 (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.

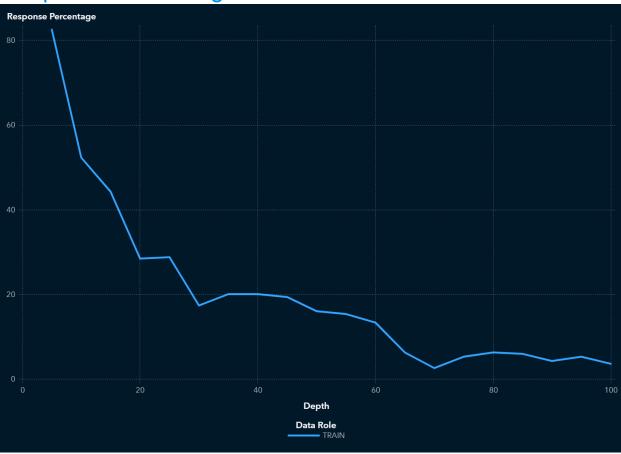
Cumulative Captured Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 33.8 (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%.

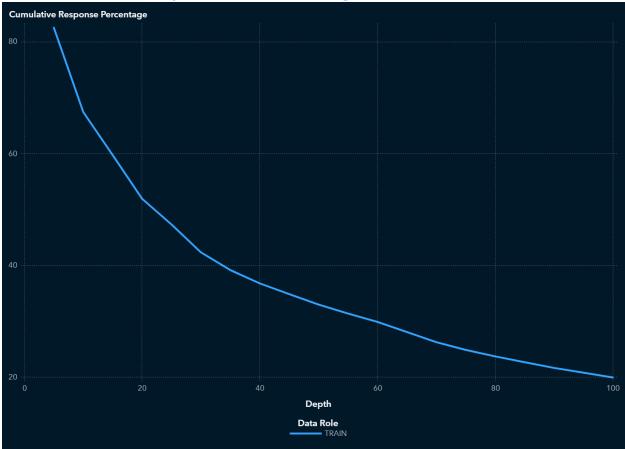
Response Percentage



At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 82.6. 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.

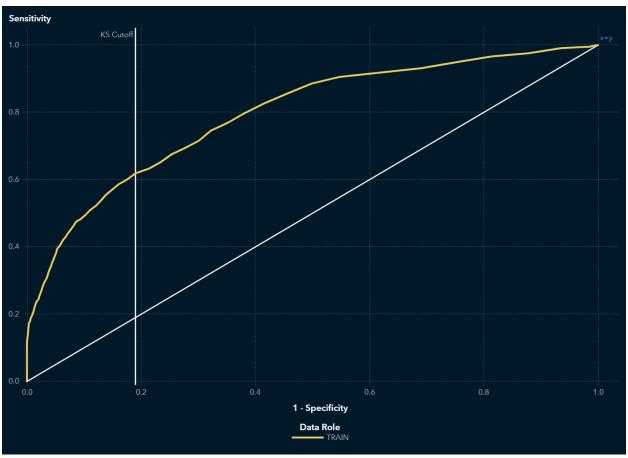
Cumulative Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 67.4. 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 0.22, where the 1-specificity value is 0.19 and the sensitivity value is 0.619.

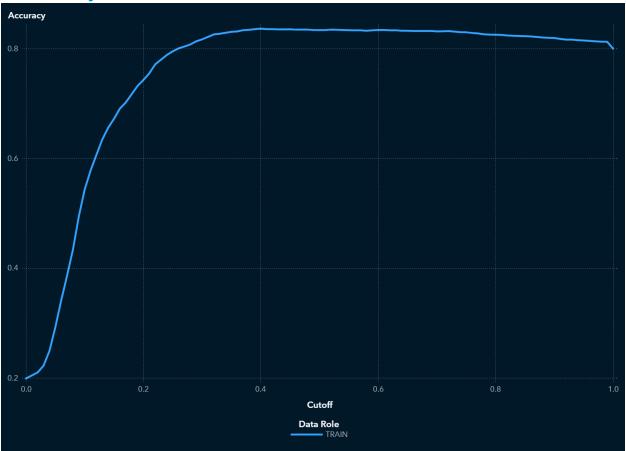
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than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

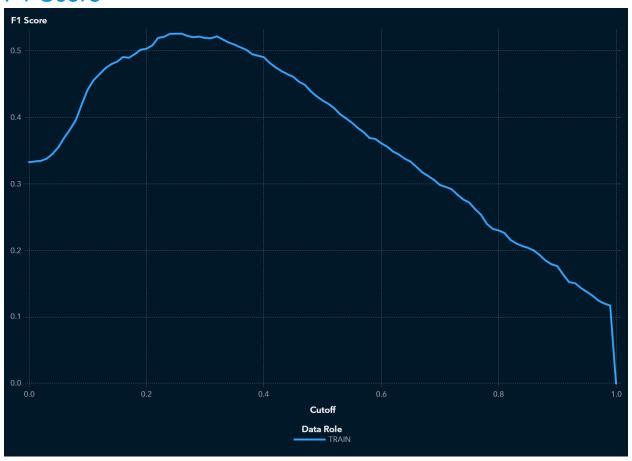
Accuracy



For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.834.

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 0.426.

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 /

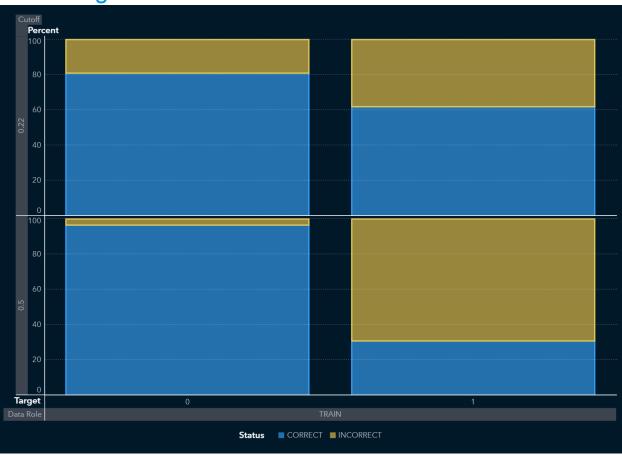
(TP + FN). The F1 score is calculated as 2*Precision*Recall / (Precision + Recall), which is the harmonic mean of Precision and Recall. Larger F1 scores indicate a more accurate model.

Fit Statistics

Target Name	Data Role	Number of Observations	Average Squared Error
BAD	TRAIN	5,960	0.1212
Divisor for ASE	Root Average	Misclassification	Multi-Class Log
	Squared Error	Rate	Loss
5,960	0.3482	0.1658	0.3927
KS (Youden)	Area Under ROC	Gini Coefficient	Gamma
0.4287	0.7949	0.5898	0.5998
Tau	KS Cutoff	KS at User- Specified Cutoff	Misclassification Rate at KS Cutoff (Event)
0.1884	0.2200	0.2732	0.2284

Misclassification	
Rate (Event)	
0.1658	

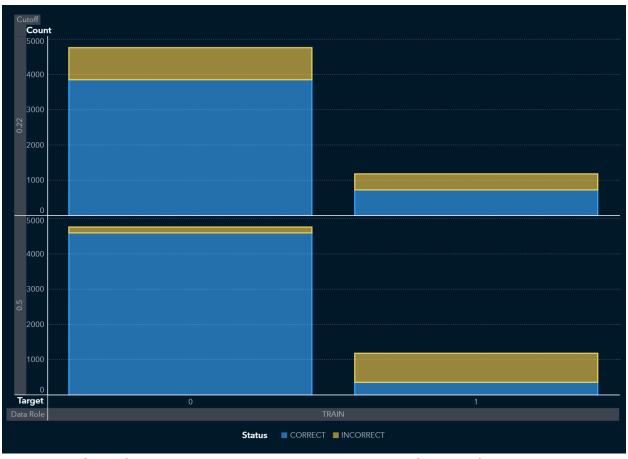
Percentage Plot



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 0.22 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 0.22 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
0.2200	KS	BAD	CORRECT
0.2200	KS	BAD	INCORRECT
0.2200	KS	BAD	CORRECT
0.2200	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	736	. ,
1	False Negative	453	
0	True Negative	3,863	
0	False Positive	908	
1	True Positive	366	
1	False Negative	823	
0	True Negative	4,606	
0	False Positive	165	

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	61.9008		
	38.0992		
	80.9684		
	19.0316		
	30.7822		
	69.2178		
	96.5416		

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	3.4584		

Properties

Property Name	Property Value
binaryProbCutoff	0.5000
chooseCriterion	SBC
classCoding	GLM
classOrder	FMTASC
codeLocation	mlearning
dataMiningVersion	V2024.03
exactPctlLift	true
explainFidelity	false
explainInfo	false
factorInteractions	false
factorSplit	false
fullDatasetReconstit ution	false
hierarchy	NONE
icePlots	false
informativeMiss	false
linkFunction	LOGIT
maxEffects	0
maxNumShapVars	20
maxSteps	0
minEffects	0
missAsLvI	false
nBins	50
nomlinkFunction	GLOGIT
normalize	true
pdNumImportantInp uts	5
pdObsSamples	1,000

Droporty Namo	Droporty Value
Property Name	Property Value
pdPlots	false
performKernelShap	false
performLime	false
performVI	false
polynomialDegree	2
reportingOnly	false
seedId	12,345
selectCriterion	SBC
selectMethod	STEPWISE
slEntry	0.0500
slStay	0.0500
specifyRows	RANDOM
stopCriterion	SBC
suppressIntercept	false
tech	NRRIDG
templateRevision	2
train	true
truncateLl	5
truncateUl	95
usePolynomial	false
useSpline	false
useSplineSplit	false
userProbCutoff	false

Output

