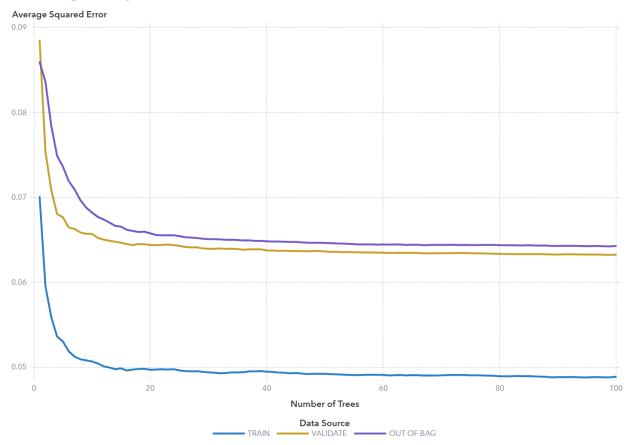


by: Jose Ivan Alejandro Cordero

Contents

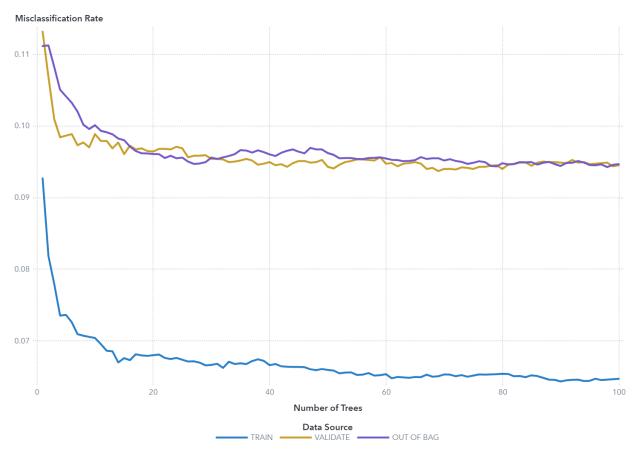
Average Squared Error	3
Misclassification Rate	4
Variable Importance	5
Score Inputs	7
Score Outputs	9
Cumulative Lift	11
Lift	13
Gain	15
Captured Response Percentage	17
Cumulative Captured Response Percentage	18
Response Percentage	20
Cumulative Response Percentage	21
ROC	22
Accuracy	24
F1 Score	25
Fit Statistics	27
Percentage Plot	28
Count Plot	29
Table	30
Properties	32
Output	36

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 error for the VALIDATE partition gives you an indication of how well your model generalizes. For this model, the minimum error for the VALIDATE partition is 0.063 and occurs for 99 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 error for the VALIDATE partition gives you an indication of how well your model generalizes. For this model, the minimum error for the VALIDATE partition is 0.094 and occurs for 69 trees.

Variable Importance

Variable Label	Role	Variable Name	Training Importance
	INPUT	duration	747.8237
	INPUT	month	322.5404
	INPUT	poutcome	283.6222
	INPUT	age	139.2879
	INPUT	job	127.7516
	INPUT	day	120.0004
	INPUT	balance	108.3888
	INPUT	pdays	66.4022
	INPUT	contact	52.2321
	INPUT	campaign	51.0411
	INPUT	housing	45.5947
	INPUT	education	42.3435
	INPUT	previous	35.3994
	INPUT	marital	30.5035
	INPUT	loan	11.8937
	INPUT	default	2.6294

Importance Standard Deviation	Relative Importance
55.7823	1
60.5141	0.4313
92.9913	0.3793
28.8970	0.1863
21.4131	0.1708
27.2676	0.1605
17.0310	0.1449
49.8738	0.0888

Importance Standard Deviation	Relative Importance
20.6726	0.0698
11.0611	0.0683
23.2176	0.0610
9.2902	0.0566
31.8216	0.0473
8.5067	0.0408
6.1761	0.0159
2.3199	0.0035

Score Inputs

Name	Role	Variable Level	Туре
age	INPUT	INTERVAL	N
balance	INPUT	INTERVAL	N
campaign	INPUT	INTERVAL	N
contact	INPUT	NOMINAL	С
day	INPUT	INTERVAL	N
default	INPUT	BINARY	С
duration	INPUT	INTERVAL	N
education	INPUT	NOMINAL	С
housing	INPUT	BINARY	С
job	INPUT	NOMINAL	С
loan	INPUT	BINARY	С
marital	INPUT	NOMINAL	С
month	INPUT	ORDINAL	С
pdays	INPUT	INTERVAL	N
poutcome	INPUT	NOMINAL	С
previous	INPUT	INTERVAL	N

Variable Type	Variable Label	Variable Format	Variable Length
double			8
double			8
double			8
varchar			9
double			8
varchar			3
double			8
varchar			9
varchar			3
varchar			13

Variable Type	Variable Label	Variable Format	Variable Length
varchar			3
varchar			8
varchar			3
double			8
varchar			7
double			8

Score Outputs

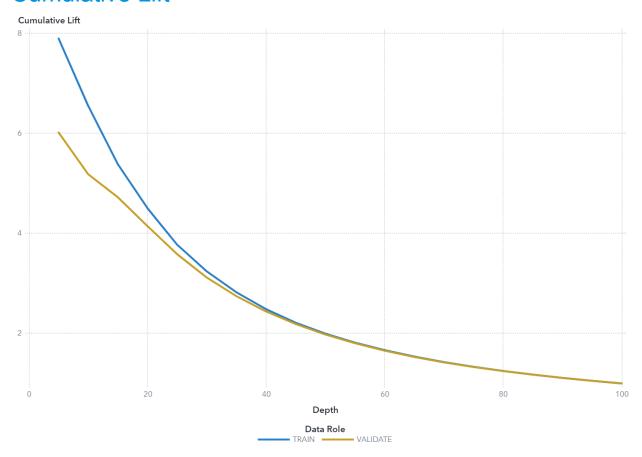
Name	Role	Туре	Variable Type
EM_CLASSIFICAT ION	CLASSIFICATION	С	char
EM_EVENTPROBAE	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
l_y	CLASSIFICATION	С	char
P_yno	PREDICT	N	double
P_yyes	PREDICT	N	double
P	ASSESS	N	double
WARN	ASSESS	С	char

Variable Label	Variable Format	Variable Length	Creator
Predicted for y		3	forest
Probability for y=yes		8	forest
Probability of Classification		8	forest
Into: y		3	forest
Predicted: y=no		8	forest
Predicted: y=yes		8	forest
		8	forest
Warnings		4	forest

Function	Creator GUID
CLASSIFICATION	246e8461- bd24-4e11-8b20- e712063af87e
PREDICT	246e8461- bd24-4e11-8b20- e712063af87e

Function	Creator GUID
PREDICT	246e8461- bd24-4e11-8b20- e712063af87e
CLASSIFICATION	246e8461- bd24-4e11-8b20- e712063af87e
PREDICT	246e8461- bd24-4e11-8b20- e712063af87e
PREDICT	246e8461- bd24-4e11-8b20- e712063af87e
PREDICT	246e8461- bd24-4e11-8b20- e712063af87e
ASSESS	246e8461- bd24-4e11-8b20- e712063af87e

Cumulative Lift

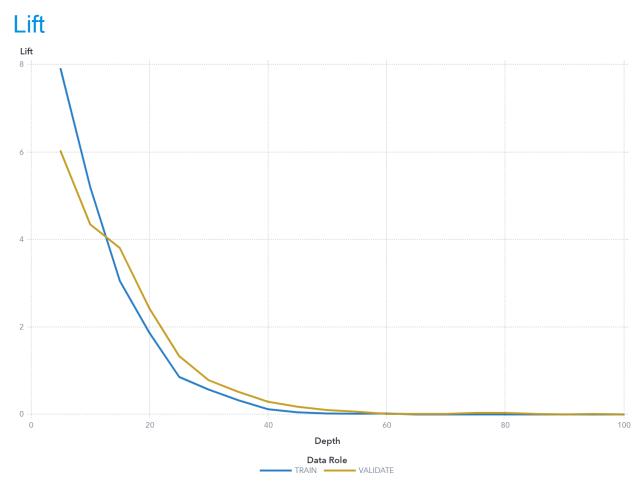


The VALIDATE partition has a Cumulative Lift of 5.19 in the 10% quantile (depth of 10) meaning there are 5.19 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Cumulative Lift of 6.55 in the 10% quantile (depth of 10) meaning there are 6.55 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_yyes, which represents the predicted probability of the event "yes" for the target y. The data is divided into 20 quantiles (demideciles, 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.



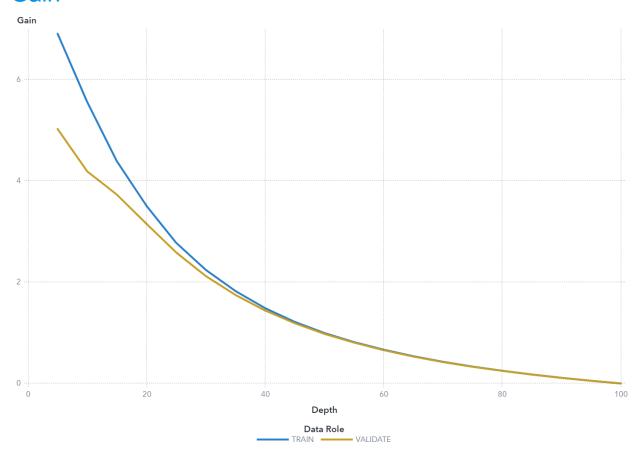
The VALIDATE partition has a Lift of 6.02 in the 5% quantile (depth of 5) meaning there are 6.02 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Lift of 7.9 in the 5% quantile (depth of 5) meaning there are 7.9 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_yyes, which represents the predicted probability of the event "yes" for the target y. 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



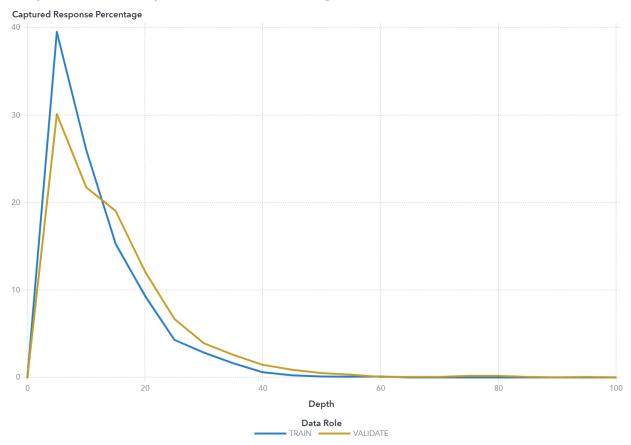
The VALIDATE partition has a Gain of 4.2 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 7.56.

The TRAIN partition has a Gain of 5.6 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 7.55.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_yyes, which represents the predicted probability of the event "yes" for the target y. 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

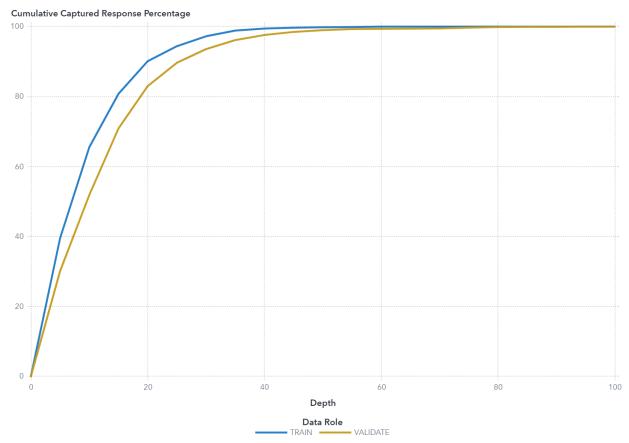


At the 5% quantile (depth of 5), the VALIDATE partition has a Captured response percentage of 30.1 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 42.79.

At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 39.5 (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 42.76.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_yyes, which represents the predicted probability of the event "yes" for the target y. 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



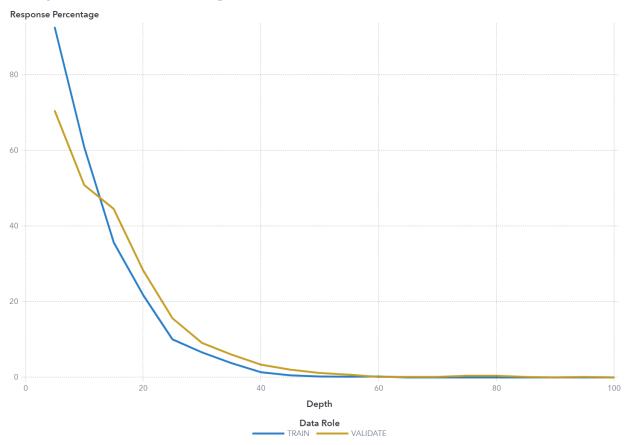
In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative captured response percentage of 51.9 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 85.57.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 65.5 (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 85.52.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_yyes, which represents the predicted probability of the event "yes" for the target y. 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

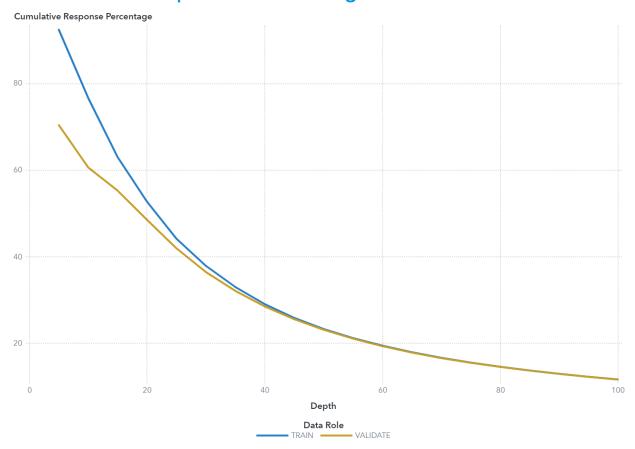


At the 5% quantile (depth of 5), the VALIDATE partition has a Response percentage of 70.4. The best possible value of Response percentage for this partition at depth 5 is 100.

At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 92.4. 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_yyes, which represents the predicted probability of the event "yes" for the target y. The data is divided into 20 quantiles (demideciles, 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

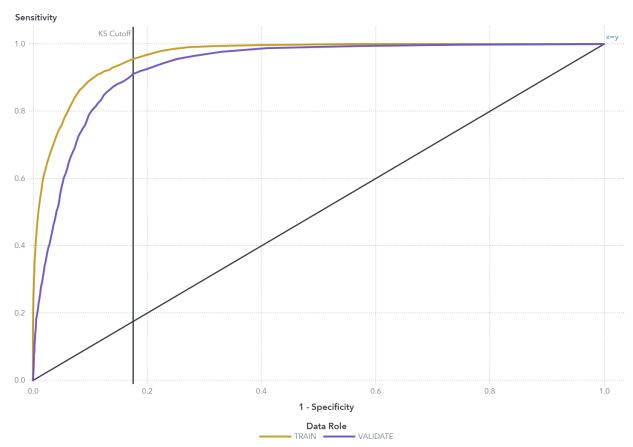


In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative response percentage of 60.6. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 76.6. 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_yyes, which represents the predicted probability of the event "yes" for the target y. 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.





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 VALIDATE partition. The KS Cutoff line is drawn at the cutoff value 0.11, where the 1-specificity value is 0.176 and the sensitivity value is 0.911.

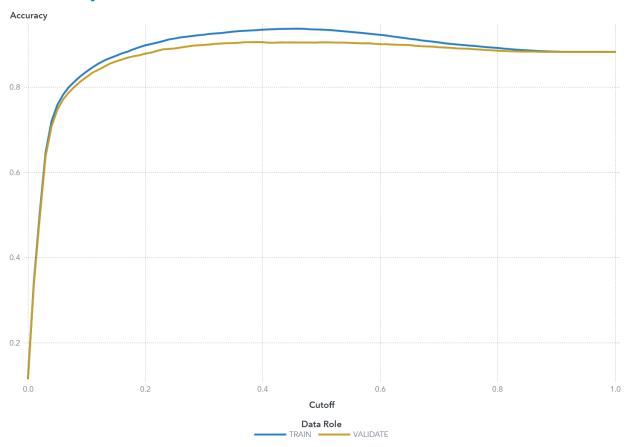
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_yyes, which is the predicted probability of the event "yes" for the target y, is greater than or equal to the cutoff value. When P_yyes is greater than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

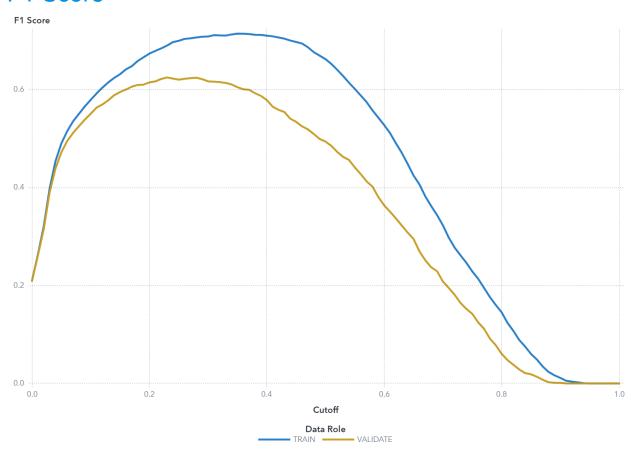


For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.935.

For this model, the accuracy in the VALIDATE partition at the cutoff of 0.5 is 0.905.

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_yyes, which is the predicted probability of the event "yes" for the target y, is greater than or equal to the cutoff value. When P_yyes 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 0.662.

For this model, the F1 score in the VALIDATE partition at the cutoff of 0.5 is 0.494.

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_yyes, which is the predicted probability of the event "yes" for the target y, is greater than or equal to the cutoff value. When P_yyes 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	Partition Indicator	Formatted Partition
у	TRAIN	1	1
у	VALIDATE	0	0

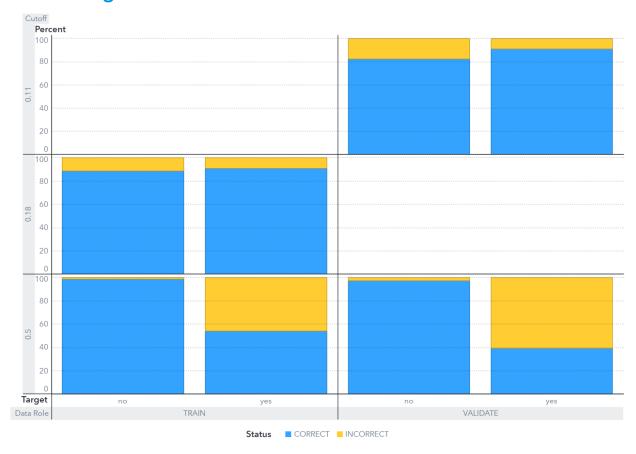
Number of Observations	Average Squared Error		Root Average Squared Error
31,647	0.0489	31,647	0.2211
13,564	0.0633	13,564	0.2515

Misclassification Rate	Multi-Class Log Loss	KS (Youden)	Area Under ROC
0.0647	0.1660	0.7961	0.9637
0.0945	0.2044	0.7355	0.9309

Gini Coefficient	Gamma	Tau	KS Cutoff
0.9274	0.9303	0.1916	0.1800
0.8618	0.8677	0.1781	0.1100

KS at User-	Misclassification	Misclassification
Specified Cutoff	Rate at KS Cutoff	Rate (Event)
	(Event)	
0.5293	0.1105	0.0647
0.3671	0.1655	0.0945

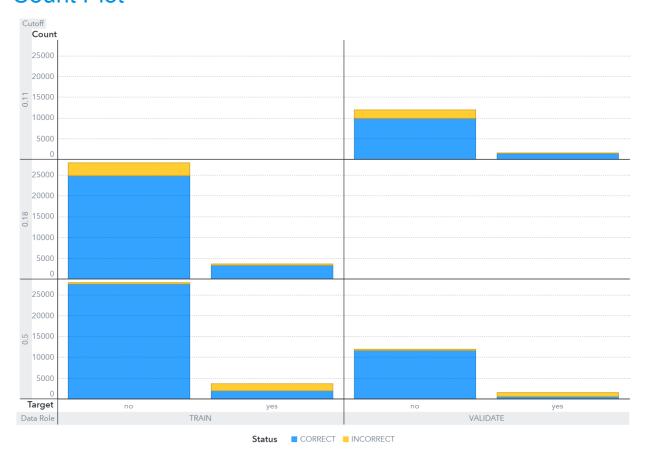
Percentage Plot



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 these KS cutoff values for existing partitions: 0.18 (TRAIN), 0.11 (VALIDATE).

For this data, for the bar corresponding to the event level of y, "yes", 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 these KS cutoff values for existing partitions: 0.18 (TRAIN), 0.11 (VALIDATE).

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

Table

Cutoff	Cutoff Source	Target Name	Response
0.1100	KS	у	CORRECT
0.1100	KS	у	INCORRECT
0.1100	KS	у	CORRECT
0.1100	KS	у	INCORRECT
0.1800	KS	у	CORRECT
0.1800	KS	у	INCORRECT
0.1800	KS	у	CORRECT
0.1800	KS	у	INCORRECT
0.5000	Default	у	CORRECT
0.5000	Default	у	INCORRECT
0.5000	Default	у	CORRECT
0.5000	Default	у	INCORRECT

Event	Value	Training Frequency	Validation Frequency
yes	True Positive		1,446
yes	False Negative		141
no	True Negative		9,873
no	False Positive		2,104
yes	True Positive	3,366	
yes	False Negative	336	
no	True Negative	24,784	
no	False Positive	3,161	
yes	True Positive	2,006	625
yes	False Negative	1,696	962
no	True Negative	27,593	11,657
no	False Positive	352	320

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
		91.1153	
		8.8847	
		82.4330	
		17.5670	
	90.9238		
	9.0762		
	88.6885		
	11.3115		
	54.1869	39.3825	
	45.8131	60.6175	
	98.7404	97.3282	
	1.2596	2.6718	

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	V2023.10
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
seed	12,345
seedId	12,345
specifyRows	RANDOM
templateRevision	4
train	true
trainFraction	0.6000
truncateLl	5

Property Name	Property Value
truncateUI	95
userProbCutoff	false
varsToTry	100
voteMethod	PROBABILITY

Output

