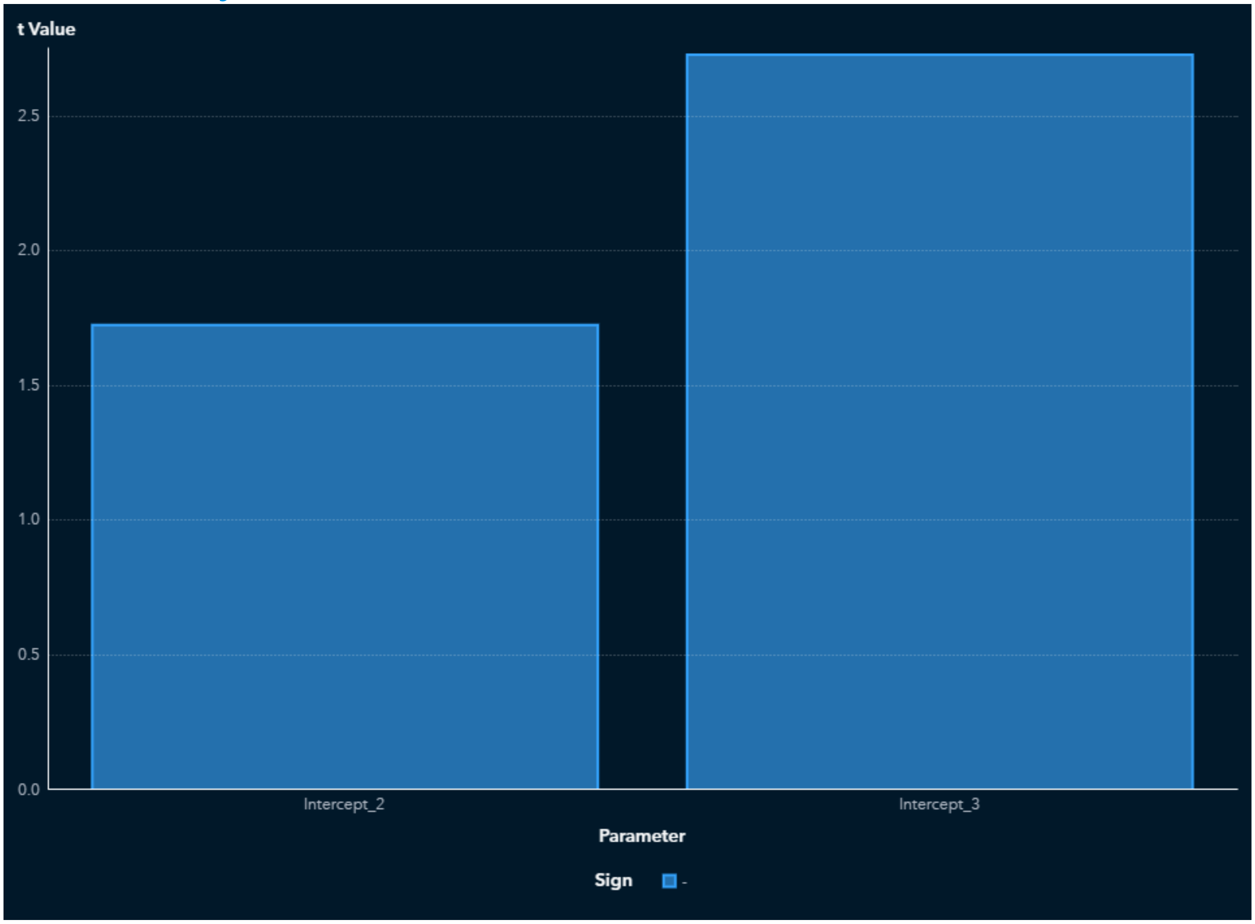
Predicting_Accidents_D

"Logistic Regression" Results

by: di00222@surrey.ac.uk

Contents

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 target level 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 the target level as the parameter value increases. The target level to which the parameter estimate corresponds is suffixed to the parameter name (for a cumulative link model, this is only true for the intercept). The most significant parameter is Intercept for the target level "3" with a t value of -2.728.

Parameter Estimates

Effect	Parameter	t Value	Sign
Intercept	Intercept_3	2.7281	-
Intercept	Intercept_2	1.7251	-

Estimate	Absolute Estimate	Standard Error	Chi-Square
-0.1345	0.1345	0.0493	7.4425
-0.0839	0.0839	0.0487	2.9758

Pr > Chi-Square	Degrees of Freedom	Predicted Outcome
0.0064	1	3
0.0845	1	2

Selection Summary

Step	Effect Entered	Effect Removed	Number of Effects
0	Intercept		1
1	WOEENC__months_		2
2		WOEENC__months_	1

SBC	Optimal SBC
5,421.9257	0
4,968.8724	0
4,928.9100	1

Regression Fit Statistics

Statistic	Description	Training	Validation
M2LL	-2 Log Likelihood	5,406.3066	2,314.7226
AIC	AIC (smaller is better)	5,410.3066	2,318.7226
AICC	AICC (smaller is better)	5,410.3115	2,318.7341
SBC	SBC (smaller is better)	5,421.9257	2,328.6452
ASE	Average Square Error	0.6656	0.6656

Score Inputs

Name	Role	Variable Level	Type
day_of_week	INPUT	NOMINAL	N
first_road_class	INPUT	NOMINAL	N
junc_detail_d	INPUT	NOMINAL	C
latitude	INPUT	INTERVAL	N
light_con_d	INPUT	NOMINAL	C
longitude	INPUT	INTERVAL	N
num_of_vehi	INPUT	NOMINAL	N
road_type_d	INPUT	NOMINAL	C
speed_limit	INPUT	NOMINAL	N
weath_con_d	INPUT	NOMINAL	C
months	INPUT	NOMINAL	C

Variable Type	Variable Label	Variable Format	Variable Length
double			8
double			8
varchar			35
double			8
varchar			23
double			8
double			8
varchar			18
double			8
varchar			21
varchar			3

Score Outputs

Name	Role	Type	Variable Type
BIN_first_road_class	REJECTED	C	char
BIN_num_of_vehi	REJECTED	C	char
BIN_road_type_d	REJECTED	C	char
EM_CLASSIFICATION	CLASSIFICATION	C	char
EM_EVENTPROBABILITY	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
I_acci_severity	CLASSIFICATION	C	char
LOG_longitude	REJECTED	N	double
P_acci_severity1	PREDICT	N	double
P_acci_severity2	PREDICT	N	double
P_acci_severity3	PREDICT	N	double
SQRT_LOG_longitude	INPUT	N	double
SQR_SQR_latitude	INPUT	N	double
SQR_latitude	REJECTED	N	double
WOEENC_BIN_first_road_class	INPUT	N	double
WOEENC_BIN_num_of_vehi	INPUT	N	double
WOEENC_BIN_road_type_d	INPUT	N	double
WOEENC__months_	INPUT	N	double
WOEENC_day_of_week	INPUT	N	double
WOEENC_junc_detail_d	INPUT	N	double

Name	Role	Type	Variable Type
WOEENC_light_con_d	INPUT	N	double
WOEENC_speed_limit	INPUT	N	double
WOEENC_weath_con_d	INPUT	N	double

Variable Label	Variable Format	Variable Length	Creator
Transformed first_road_class		12	transform
Transformed num_of_vehi		12	transform
Transformed road_type_d		18	transform
Predicted for acci_severity		12	logisticreg
Probability for acci_severity=1		8	logisticreg
Probability of Classification		8	logisticreg
Into: acci_severity		12	logisticreg
Transformed longitude		8	transform
Predicted: acci_severity=1		8	logisticreg
Predicted: acci_severity=2		8	logisticreg
Predicted: acci_severity=3		8	logisticreg
Transformed Transformed longitude		8	transform
Transformed		8	transform

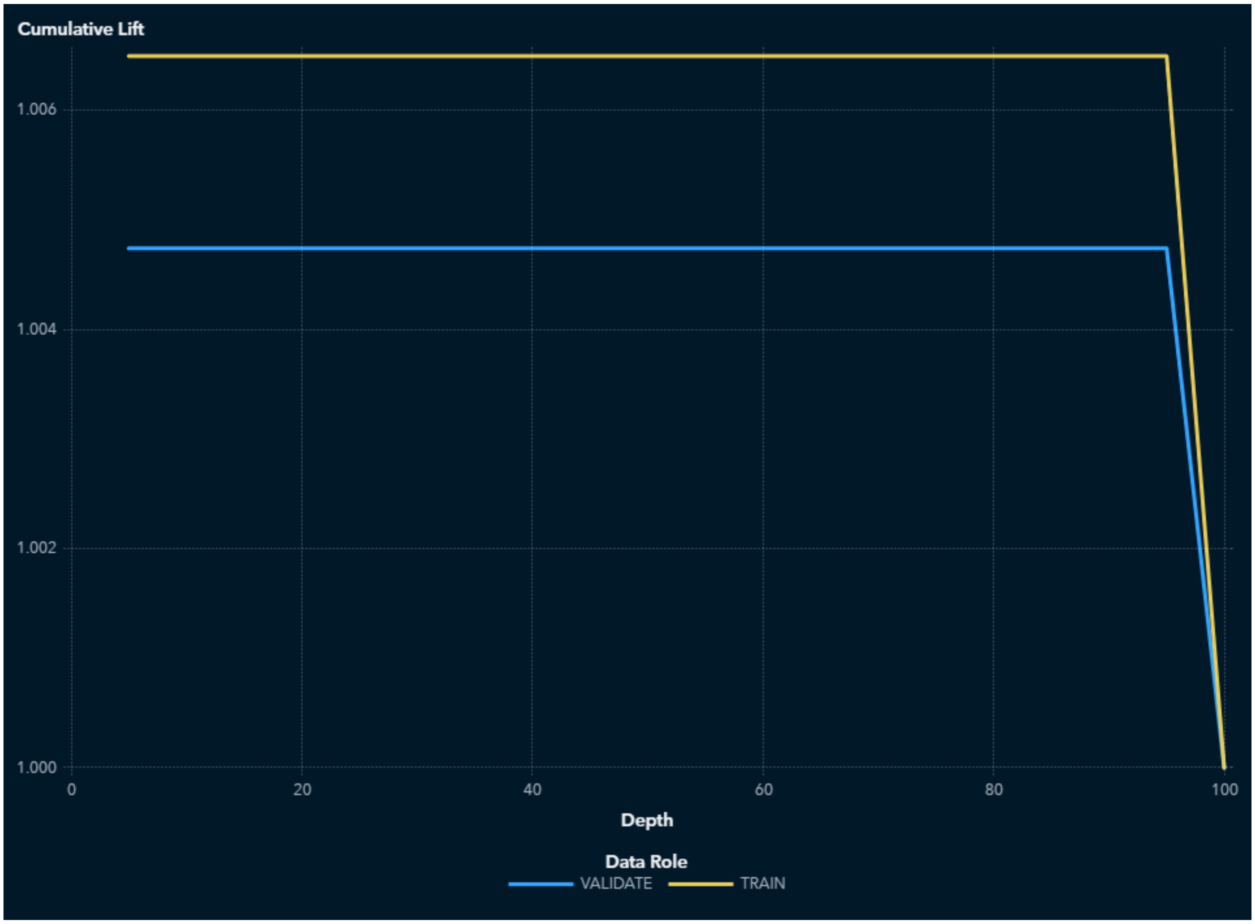
Variable Label	Variable Format	Variable Length	Creator
Transformed latitude			
Transformed latitude		8	transform
Transformed Transformed first_road_class		8	transform
Transformed Transformed num_of_vehi		8	transform
Transformed Transformed road_type_d		8	transform
Transformed _months_		8	transform
Transformed day_of_week		8	transform
Transformed junc_detail_d		8	transform
Transformed light_con_d		8	transform
Transformed speed_limit		8	transform
Transformed weath_con_d		8	transform

Function	Creator GUID
TRANSFORM	bc2a41dc-386f-4ffe - a5cb-4eed66ab6b1 9
TRANSFORM	bc2a41dc-386f-4ffe - a5cb-4eed66ab6b1

Function	Creator GUID
	9
TRANSFORM	bc2a41dc-386f-4ffe-a5cb-4eed66ab6b19
CLASSIFICATION	3c919ddf-e427-4933-ac74-dfbdeb11cdc5
PREDICT	3c919ddf-e427-4933-ac74-dfbdeb11cdc5
PREDICT	3c919ddf-e427-4933-ac74-dfbdeb11cdc5
CLASSIFICATION	3c919ddf-e427-4933-ac74-dfbdeb11cdc5
TRANSFORM	bc2a41dc-386f-4ffe-a5cb-4eed66ab6b19
PREDICT	3c919ddf-e427-4933-ac74-dfbdeb11cdc5
PREDICT	3c919ddf-e427-4933-ac74-dfbdeb11cdc5
PREDICT	3c919ddf-e427-4933-ac74-dfbdeb11cdc5
TRANSFORM	677c504a-30d6-47b9-b07e-ceca00823c09
TRANSFORM	677c504a-30d6-47b9-b07e-

Function	Creator GUID
	ceca00823c09
TRANSFORM	bc2a41dc-386f-4ffe - a5cb-4eed66ab6b1 9
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09
TRANSFORM	677c504a-30d6-47b 9-b07e- ceca00823c09

Cumulative Lift



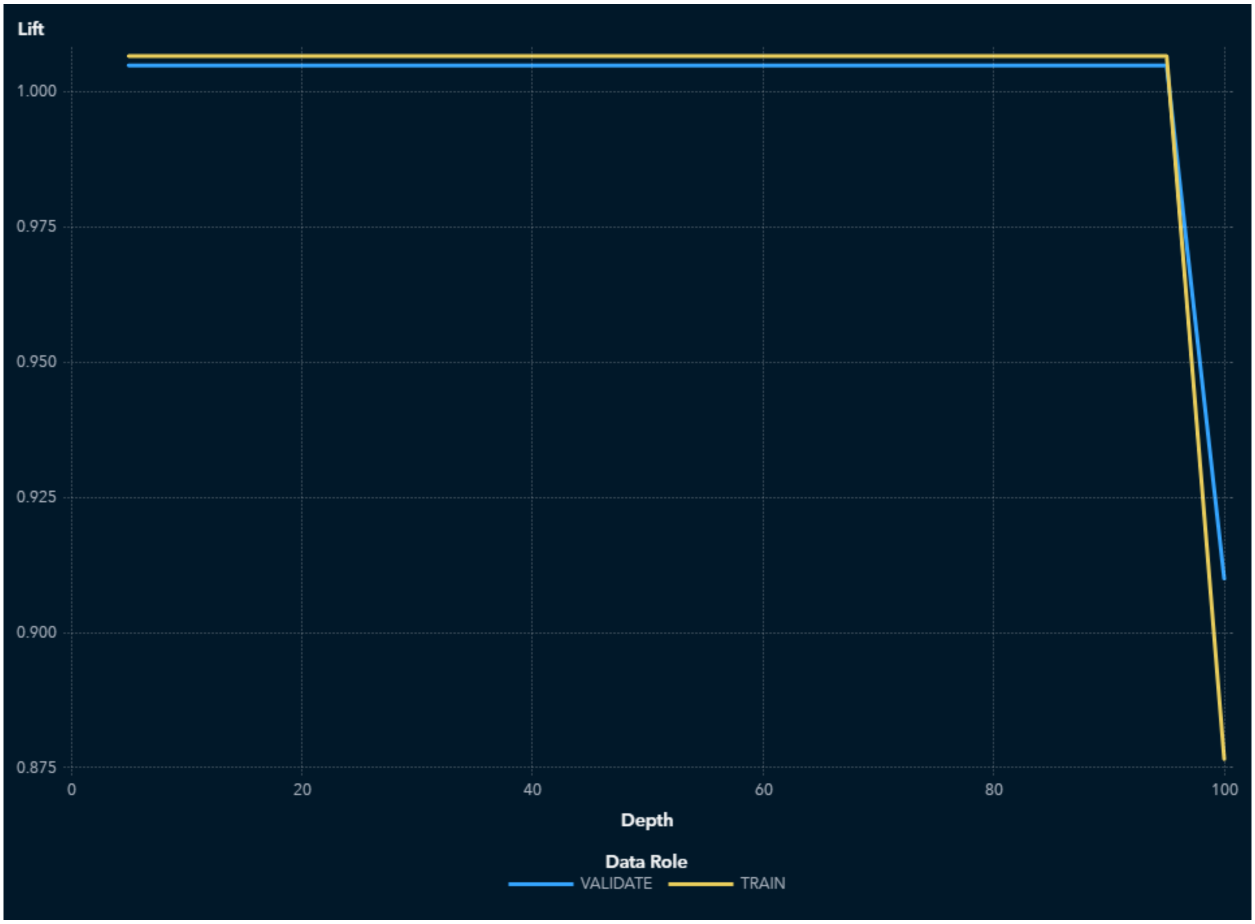
The VALIDATE partition has a Cumulative Lift of 1 in the 10% quantile (depth of 10) meaning there are 1 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders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

The TRAIN partition has a Cumulative Lift of 1.01 in the 10% quantile (depth of 10) meaning there are 1.01 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



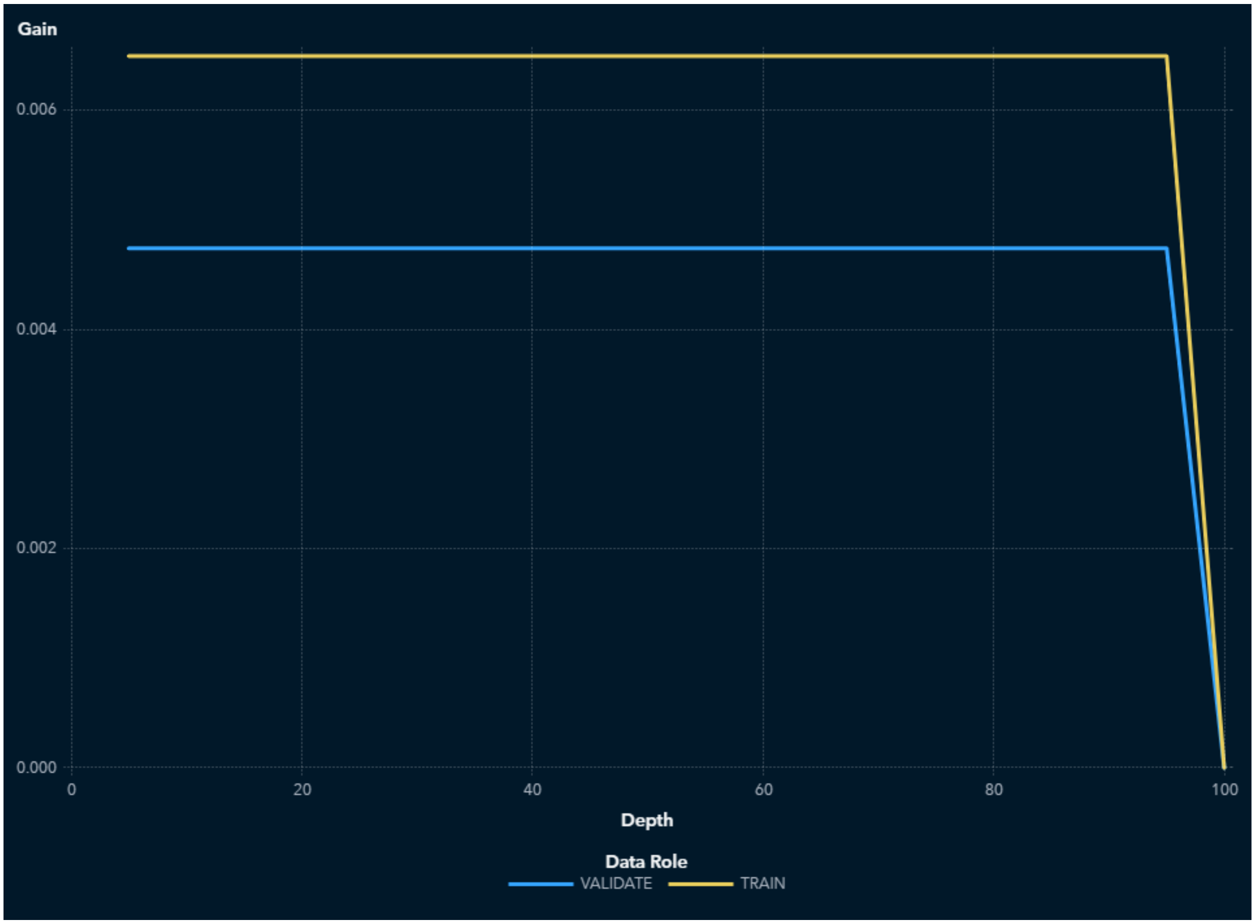
The VALIDATE partition has a Lift of 1 in the 5% quantile (depth of 5) meaning there are 1 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on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

The TRAIN partition has a Lift of 1.01 in the 5% quantile (depth of 5) meaning there are 1.01 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



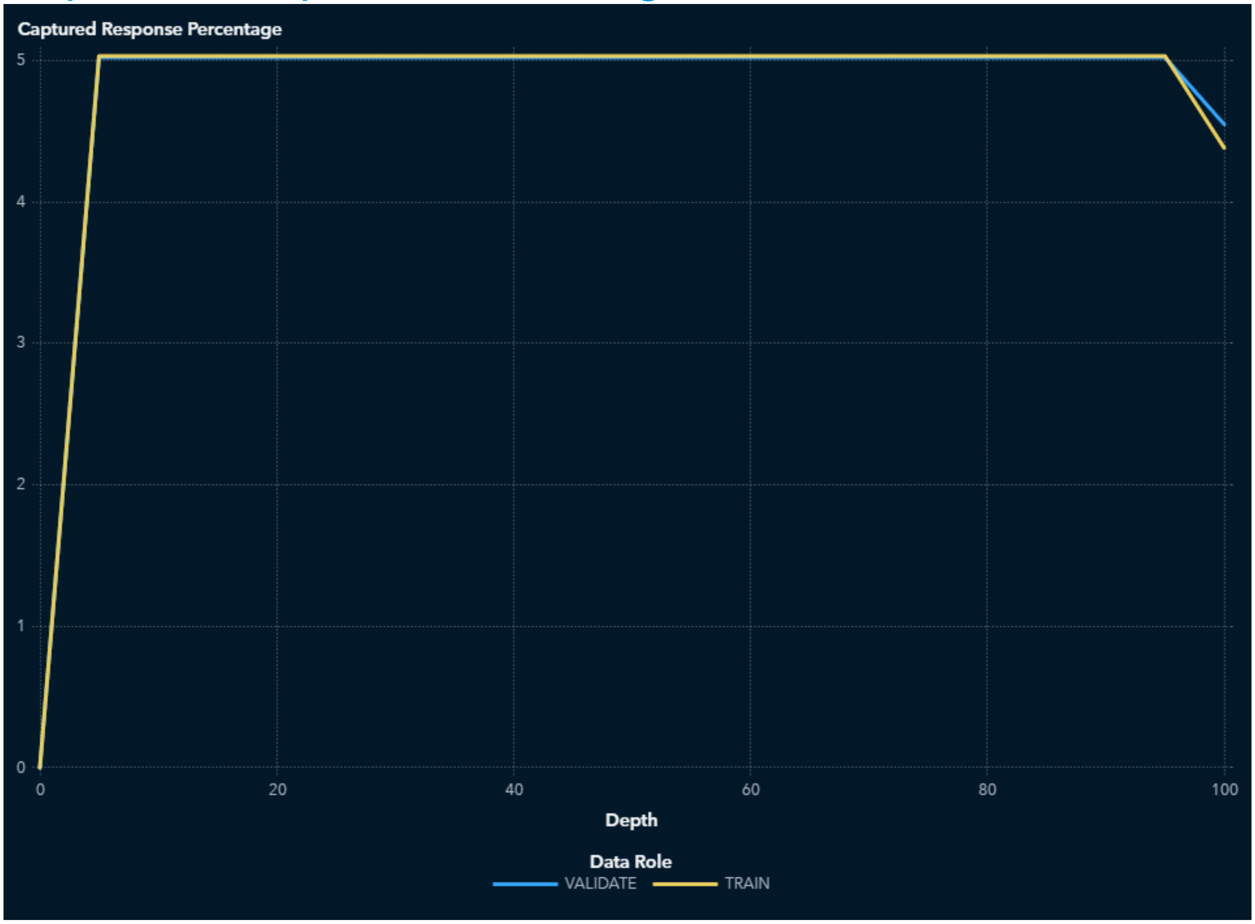
The VALIDATE partition has a Gain of 0 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 1.8.

The TRAIN partition has a Gain of 0 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 1.81.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

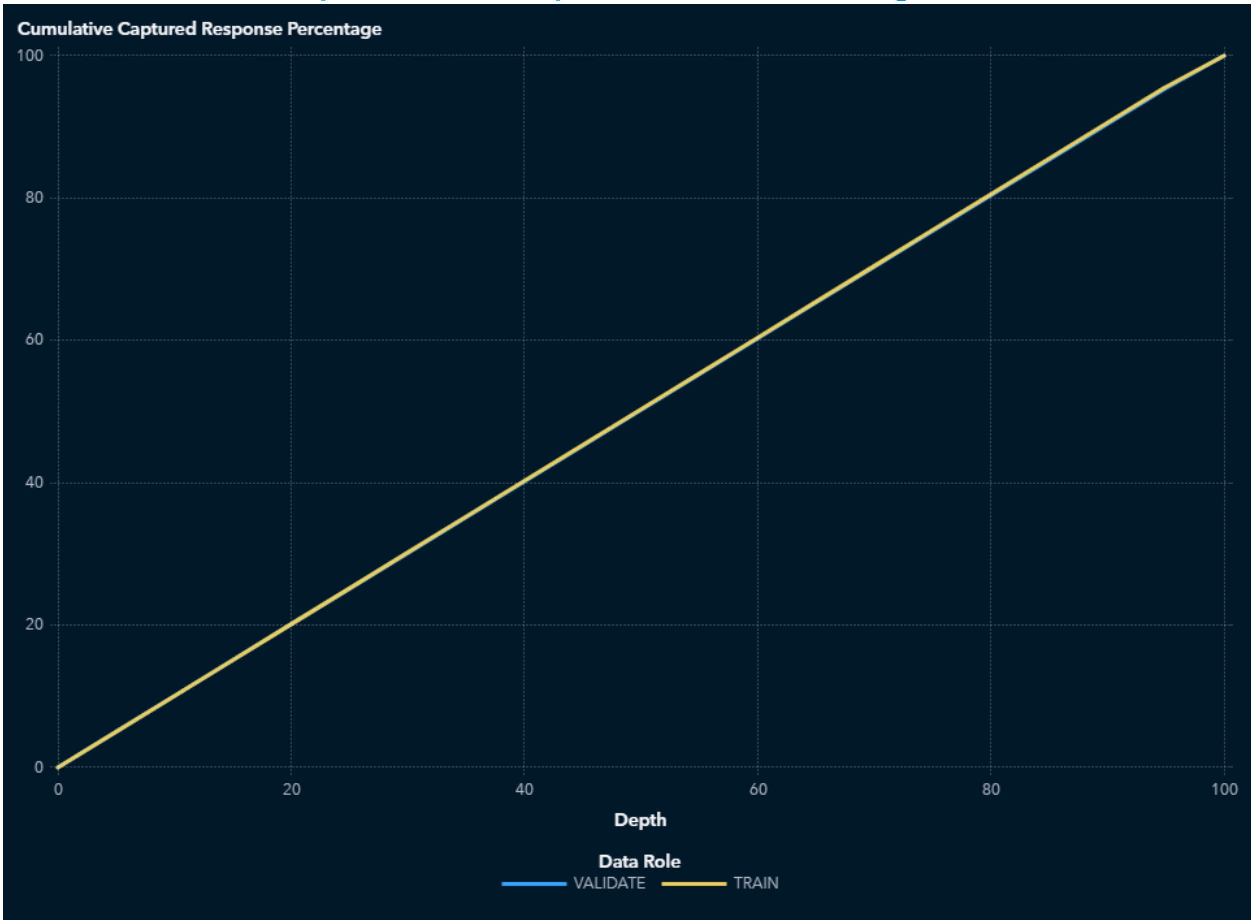


At the 5% quantile (depth of 5), the VALIDATE partition has a Captured response percentage of 5 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 14.02.

At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 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 14.06.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



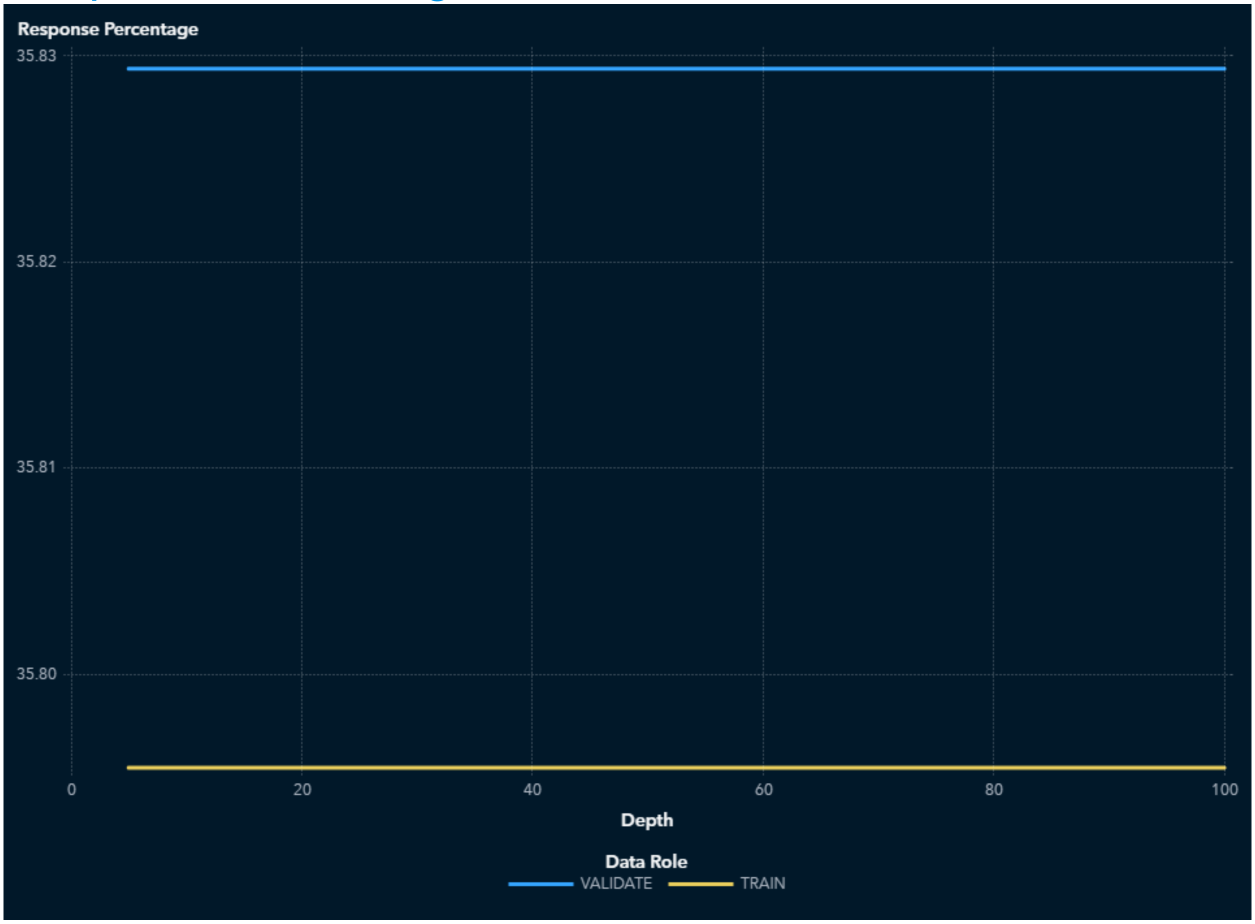
In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative captured response percentage of 10 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 28.04.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 10.1 (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 28.12.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

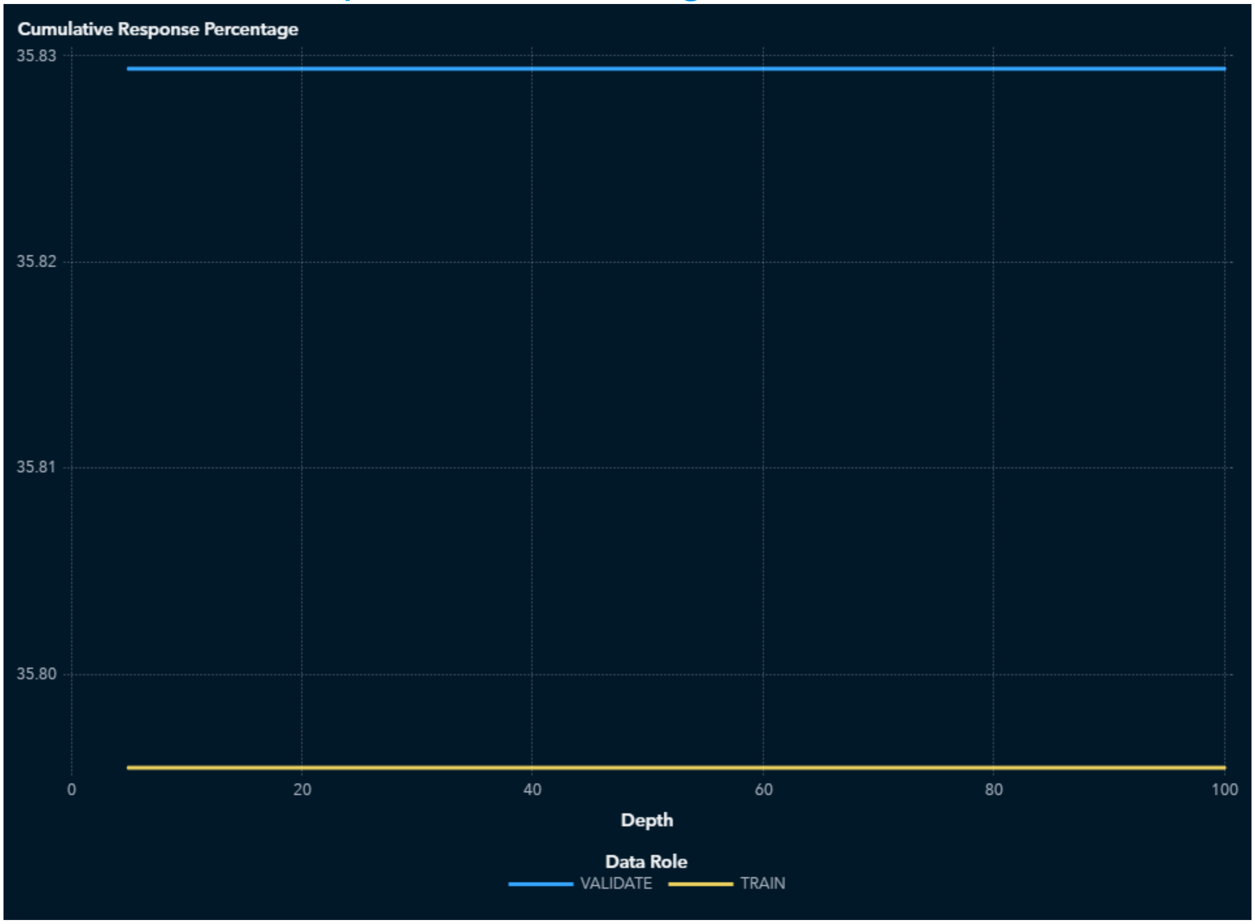


At the 5% quantile (depth of 5), the VALIDATE partition has a Response percentage of 35.8. The best possible value of Response percentage for this partition at depth 5 is 100.

At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 35.8. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

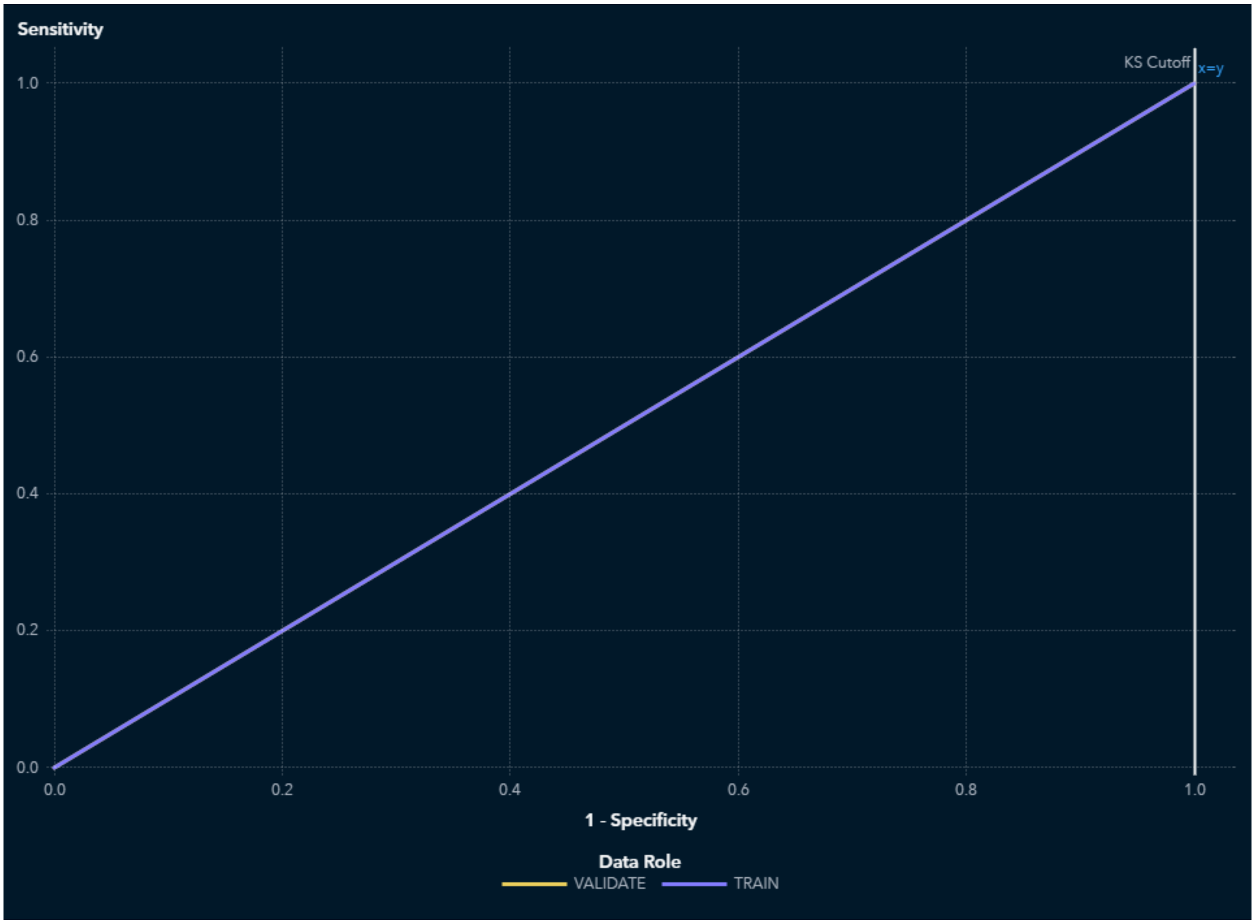


In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative response percentage of 35.8. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 35.8. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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 VALIDATE partition. The KS Cutoff line is drawn at the cutoff value 0, where the 1-specificity value is 1 and the sensitivity value is 1.

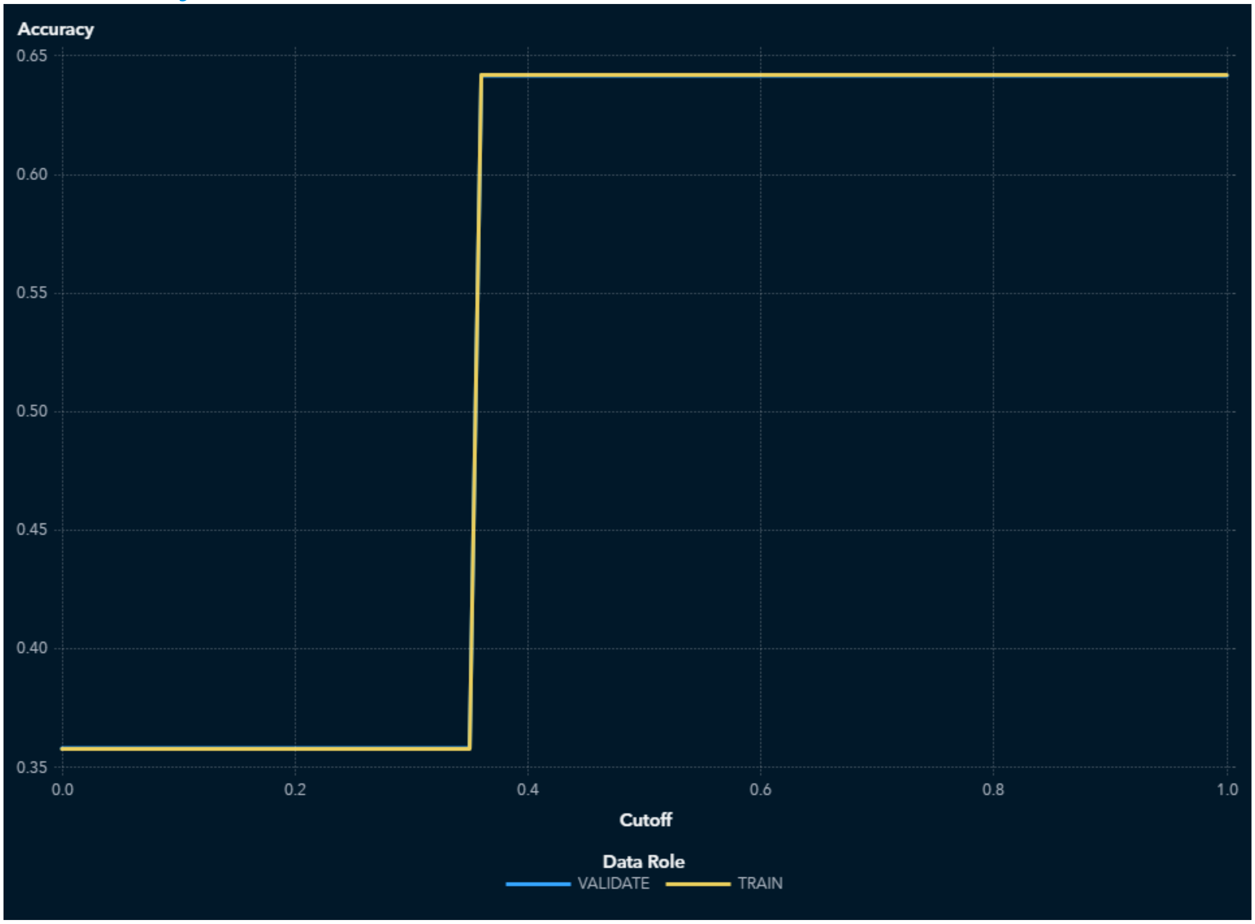
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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

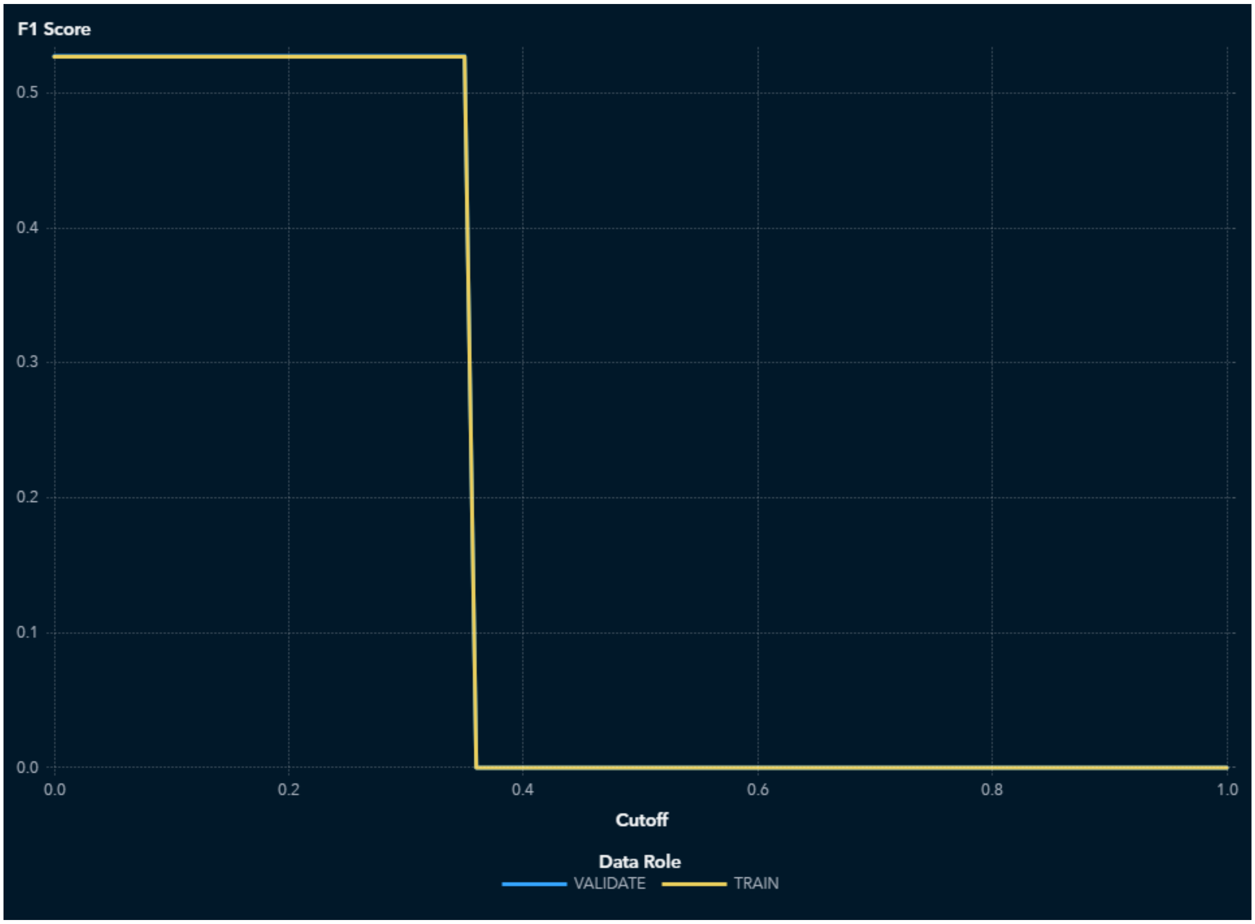


For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.642.

For this model, the accuracy in the VALIDATE partition at the cutoff of 0.5 is 0.642.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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.

For this model, the F1 score in the VALIDATE partition at the cutoff of 0.5 is 0.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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
acci_severity	TRAIN	1	1
acci_severity	VALIDATE	0	0

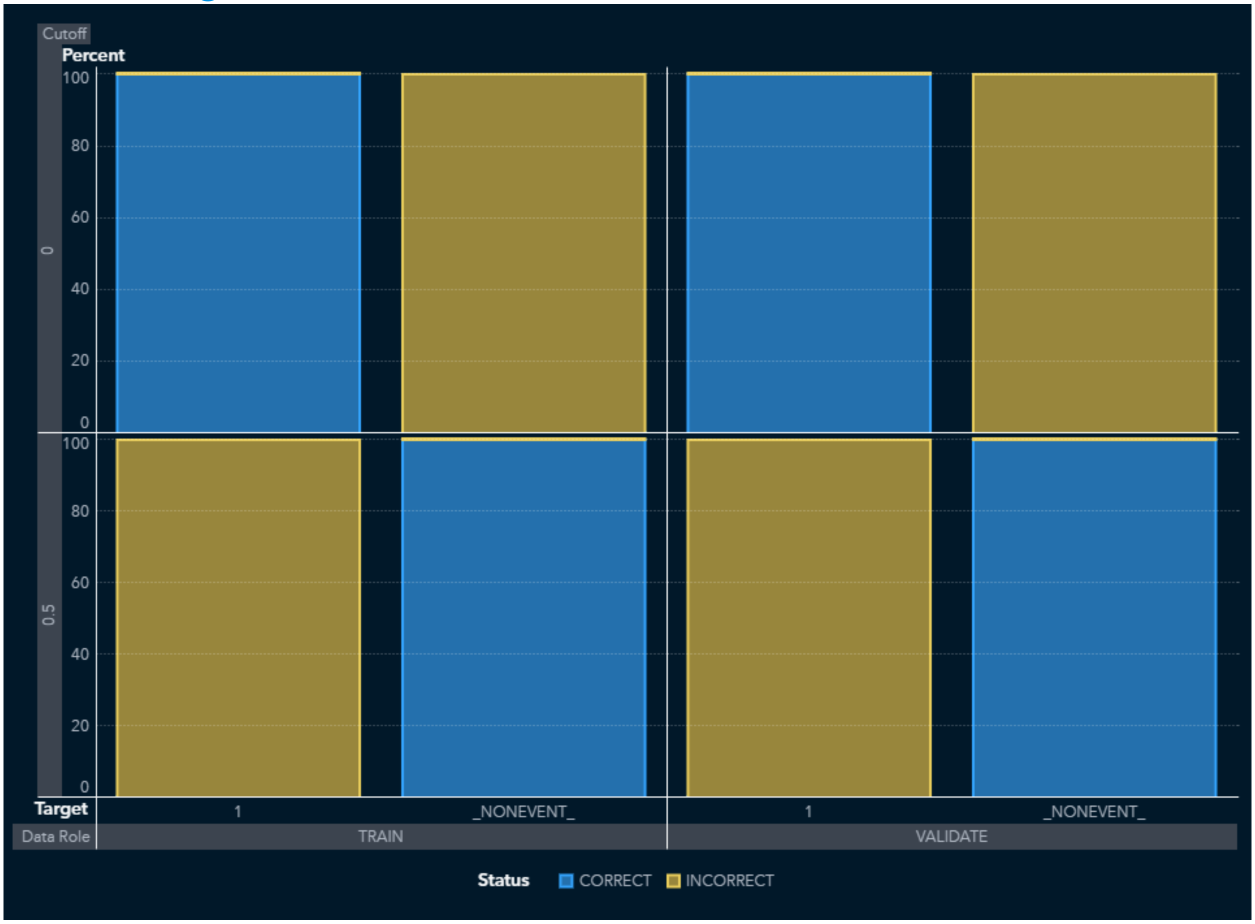
Number of Observations	Average Squared Error	Divisor for ASE	Root Average Squared Error
2,464	0.2219	2,464	0.4710
1,055	0.2219	1,055	0.4710

Misclassification Rate	Multi-Class Log Loss	KS (Youden)	Area Under ROC
0.6420	1.0971	0	0.5000
0.6417	1.0970	0	0.5000

Gini Coefficient	Gamma	Tau	KS Cutoff
0		0	0
0		0	0

KS at Default Cutoff	Misclassification Rate at KS Cutoff (Event)	Misclassification Rate (Event)
0	0.6420	0.3580
0	0.6417	0.3583

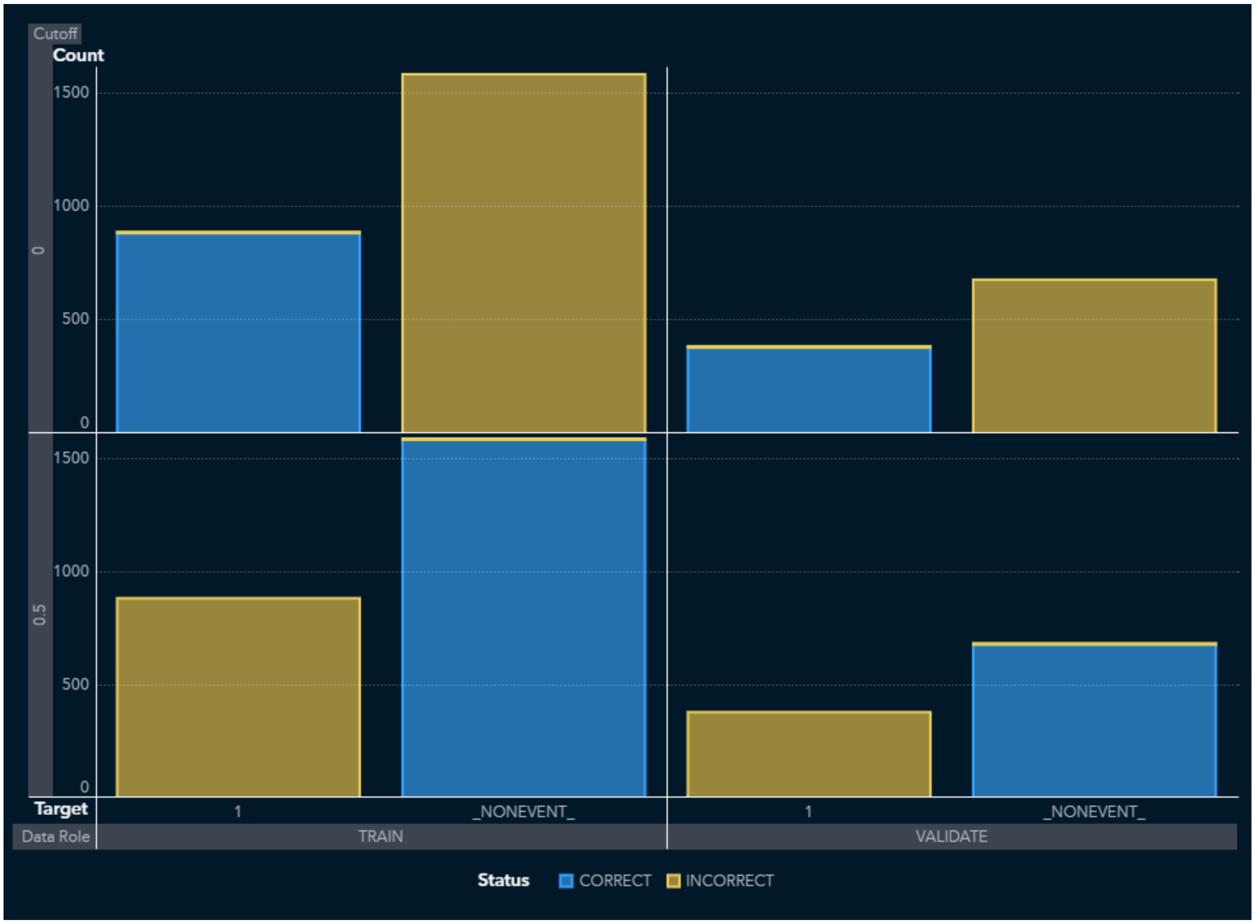
Percentage Plot



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 (TRAIN), 0 (VALIDATE).

For this data, for the bar corresponding to the event level of acci_severity, "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 these KS cutoff values for existing partitions: 0 (TRAIN), 0 (VALIDATE).

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

Table

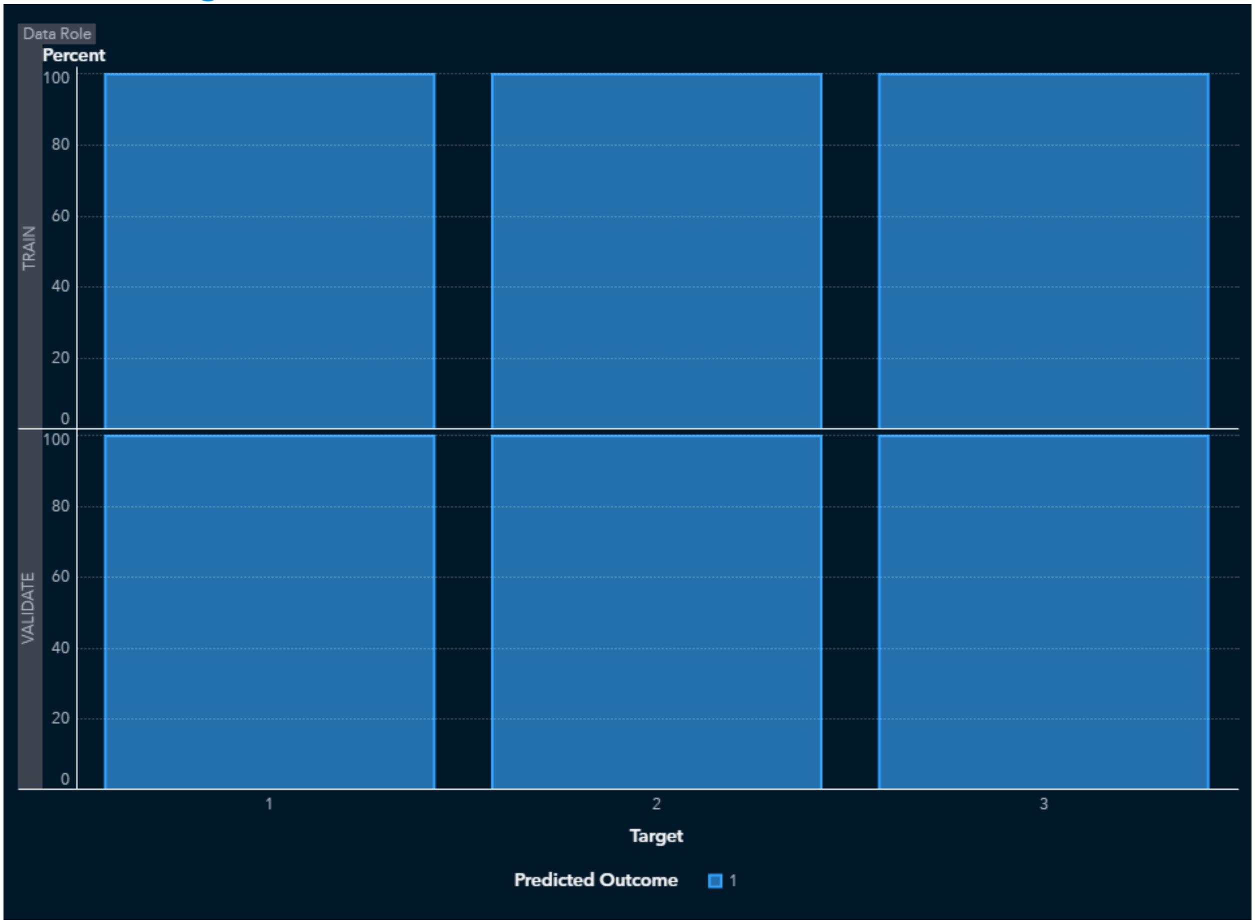
Cutoff	Cutoff Source	Target Name	Response
0	KS	acci_severity	CORRECT
0	KS	acci_severity	INCORRECT
0	KS	acci_severity	CORRECT
0	KS	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT

Event	Value	Training Frequency	Validation Frequency
1	True Positive	882	378
1	False Negative	0	0
NONEVENT	True Negative	0	0
NONEVENT	False Positive	1,582	677
1	True Positive	0	0
1	False Negative	882	378
NONEVENT	True Negative	1,582	677
NONEVENT	False Positive	0	0

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	100	100	
	0	0	
	0	0	
	100	100	
	0	0	
	100	100	
	100	100	

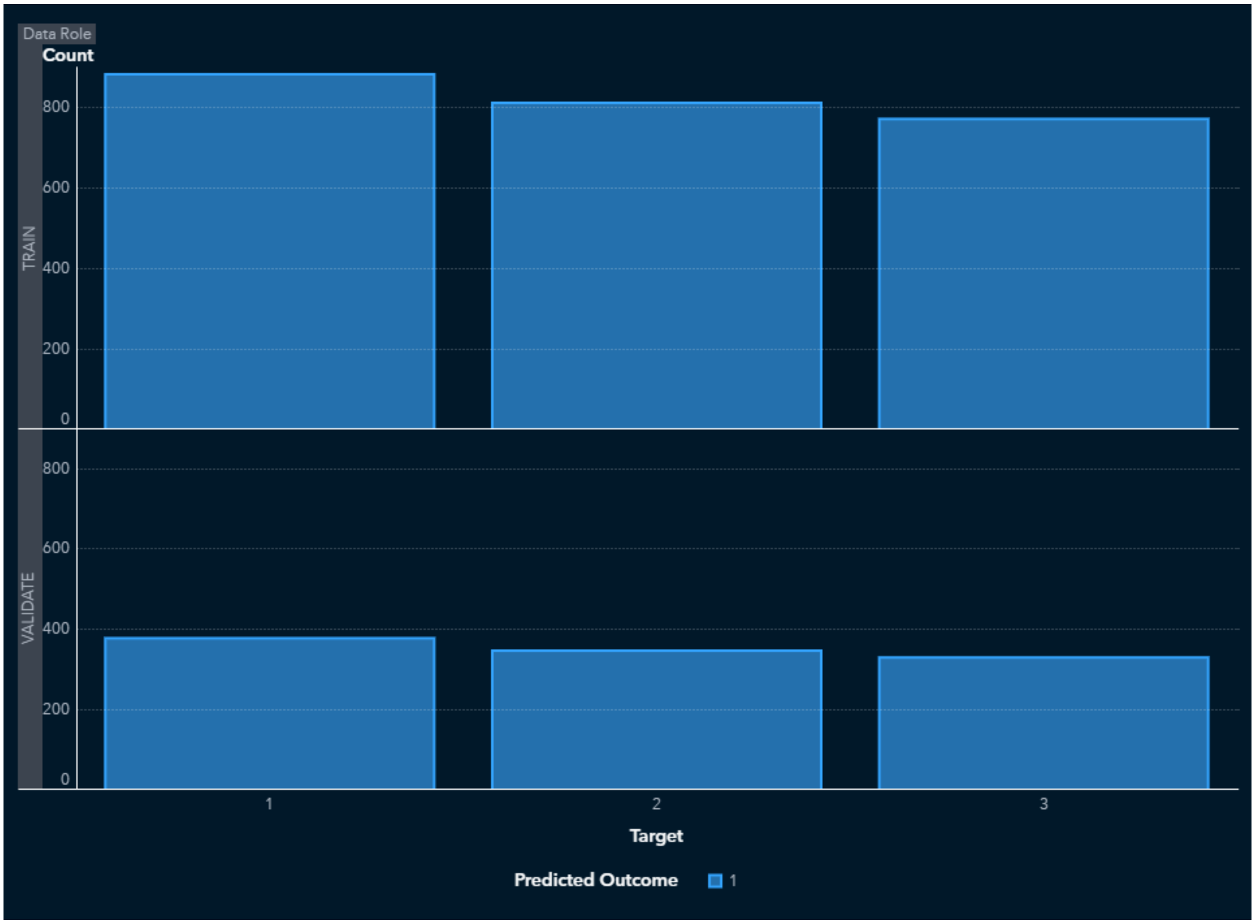
Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	0	0	

Percentage Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Count Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Table

Target Name	Data Role	Target	Unformatted Target
acci_severity	VALIDATE	1	1
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	3	3
acci_severity	TRAIN	1	1
acci_severity	TRAIN	2	2
acci_severity	TRAIN	3	3

Predicted Outcome	Count	Percent	Status
1	378	100	CORRECT
1	347	100	INCORRECT
1	330	100	INCORRECT
1	882	100	CORRECT
1	811	100	INCORRECT
1	771	100	INCORRECT

Properties

Property Name	Property Value
binaryProbCutoff	0.5000
chooseCriterion	SBC
classCoding	GLM
classOrder	FMTASC
codeLocation	mlearning
dataMiningVersion	V2024.09
exactPctlLift	true
explainFidelity	false
explainInfo	false
factorInteractions	false
factorSplit	false
fullDatasetReconstitution	false
hierarchy	NONE
icePlots	false
informativeMiss	false
linkFunction	LOGIT
maxEffects	0
maxNumShapVars	20
maxSteps	0
minEffects	0
missAsLvl	false
nBins	50
nomlinkFunction	GLOGIT
normalize	true
pdNumImportantInputs	5
pdObsSamples	1,000

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
truncateLI	5
truncateUI	95
usePolynomial	false
useSpline	false
useSplineSplit	false
userProbCutoff	false

Output

The SAS System

The GENSELECT Procedure

Model Information	
Data Source	_INPUT_3L381EISND27YT0HJAOJZAPC5
Response Variable	acci_severity
Number of Response Levels	3
Distribution	Multinomial
Link Type	Generalized
Link Function	Logit
Optimization Technique	Newton-Raphson with Ridging
Predicted Response Level	I_acci_severity

Number of Observations

Description	Total	Training	Validation
Number of Observations Read	3519	2464	1055
Number of Observations Used	3519	2464	1055

Response Profile

Ordered Value	acci_severity	Total Frequency	Training	Validation
1	3	1101	771	330
2	2	1158	811	347
3	1	1260	882	378

Probabilities modeled use acci_severity = 1 as the reference category.

Selection Information	
Selection Method	Stepwise
Select Criterion	SBC
Choose Criterion	SBC
Stop Criterion	SBC
Effect Hierarchy Enforced	None
Stop Horizon	3

Selection Details

Convergence criterion (ABSGCONV=1E-7) satisfied.

Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	SBC
0	Intercept		1	5421.9257
1	WOEENC__months_		2	4968.8724
2		WOEENC__months_	1	4928.9100*

* Optimal Value Of Criterion

Stepwise selection stopped because adding or removing an effect does not improve the SBC criterion

The model at step 2 is selected where SBC is 4928.91.

Selected Effects:	Intercept

Selected Model

Dimensions	
Columns in Design	2
Number of Effects	1
Max Effect Columns	2
Rank of Design	2
Parameters in Optimization	2

Fit Statistics

Description	Training	Validation
-2 Log Likelihood	5406.30661	2314.72264
AIC (smaller is better)	5410.30661	2318.72264
AICC (smaller is better)	5410.31148	2318.73405
SBC (smaller is better)	5421.92569	2328.64524
Average Square Error	0.66563	0.66560

Parameter Estimates

Parameter	acci_severity	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	3	1	-0.134504	0.049303	7.4425	0.0064
Intercept	2	1	-0.083924	0.048650	2.9758	0.0845

Score Code Variables for Predicted Probability

acci_severity	Variable
1	P_acci_severity1
3	P_acci_severity3
2	P_acci_severity2

Task Timing

Task	Seconds	Percent
Setup and Parsing	0.01	11.89%
Levelization	0.00	6.03%
Model Initialization	0.00	4.01%
SSCP Computation	0.01	15.19%
Model Selection	0.04	60.88%
Producing Score Code	0.00	1.54%
Display	0.00	0.36%
Cleanup	0.00	0.01%
Total	0.07	100.00%

Predicting_Accidents_D

"Transformations" Results

by: di00222@surrey.ac.uk

Contents

Input Variable Statistics

Name	Variable Level	Number of Missing Values	Percent Missing
day_of_week	NOMINAL	0	0
first_road_class	NOMINAL	0	0
junc_detail_d	NOMINAL	0	0
latitude	INTERVAL	0	0
light_con_d	NOMINAL	0	0
longitude	INTERVAL	0	0
num_of_vehi	NOMINAL	0	0
road_type_d	NOMINAL	0	0
speed_limit	NOMINAL	0	0
weath_con_d	NOMINAL	0	0
months	NOMINAL	0	0

Minimum	Maximum	Mean	Midrange
51.0832	51.4664	51.2918	51.2748
-0.8231	0.0571	-0.4148	-0.3830

Importance Standard Deviation	Skewness	Kurtosis	Variable Label
0.0857	-0.2663	-0.5195	
0.2298	0.2659	-1.0181	

Transformed Variables Summary

Transformed Variable	Method	Input Variable	Formula
BIN_first_road_class	BINRARE	first_road_class	BINRARE
SQR_latitude	SQUARE	latitude	('latitude'n)**2
LOG_longitude	LOG	longitude	log('longitude'n + 1.82311)
BIN_num_of_vehi	BINRARE	num_of_vehi	BINRARE
BIN_road_type_d	BINRARE	road_type_d	BINRARE

Variable Level	Type	Variable Label
NOMINAL	N	Transformed first_road_class
INTERVAL	N	Transformed latitude
INTERVAL	N	Transformed longitude
NOMINAL	N	Transformed num_of_vehi
NOMINAL	C	Transformed road_type_d

Properties

Property Name	Property Value
bestRankBinTarg	CLASSICALSKEW
bestRankIntTarg	CLASSICALSKEW
bestRankNomTarg	CLASSICALSKEW
codeLocation	mlearning
cutoffPercent	0.5000
dataMiningVersion	V2024.09
defaultMethod	BEST
defltClassMethod	BINRARE
fullDatasetReconstitution	false
ignoreMetadata	false
minObsinBin	5
missingTreatment	SEPARATE
numBins	4
outlierBins	false
rejectVariable	true
reportingOnly	false
summaryStatistics	false
templateRevision	2
woeAdjust	0.5000

Output

Best Variable Transformation

Obs	Variable	Transformation Name	N	NMiss	Moment Skewness
1	latitude	Square	2464	0	-0.2628
2	longitude	Log	2464	0	0.04286

Variable Transformation Ranking

Obs	Variable	Transformation Name	Best Transform Rank	N	NMiss	Moment Skewness
1	latitude	Square	1	2464	0	-0.2628
2	latitude	Center	2	2464	0	-0.2663
3	latitude	Range	2	2464	0	-0.2663
4	latitude	Standardize	2	2464	0	-0.2663
5	latitude	None	5	2464	0	-0.2663
6	latitude	Inverse Square	6	2464	0	0.2769
7	latitude	Square Root	7	2464	0	-0.3438
8	latitude	Log	8	2464	0	-0.4231
9	latitude	Log10	8	2464	0	-0.4231
10	latitude	Inverse Square Root	10	2464	0	0.5044
11	latitude	Inverse	11	2464	0	0.5875
12	longitude	Log	1	2464	0	0.04286
13	longitude	Log10	1	2464	0	0.04286
14	longitude	Inverse Square Root	3	2464	0	0.06984
15	longitude	Inverse Square	4	2464	0	-0.1366
16	longitude	Square Root	5	2464	0	0.1550
17	longitude	None	6	2464	0	0.2659
18	longitude	Center	6	2464	0	0.2659
19	longitude	Range	6	2464	0	0.2659
20	longitude	Standardize	6	2464	0	0.2659
21	longitude	Square	10	2464	0	0.4721

Input Variable Statistics

Obs	Input Variable	Measurement Level	Number of Missing Values	Percentage Missing	Minimum	Maximum	Mean	Midrange	Standard Deviation	Skewness	Kurtosis	Label
1	day_of_week	NOMINAL	0	0	.	.	.	.	.	.	.	
2	first_road_class	NOMINAL	0	0	.	.	.	.	.	.	.	
3	junc_detail_d	NOMINAL	0	0	.	.	.	.	.	.	.	
4	latitude	INTERVAL	0	0	51.083212	51.466373	51.291816514	51.2748	0.0856850913	-0.266329518	-0.519544193	
5	light_con_d	NOMINAL	0	0	.	.	.	.	.	.	.	
6	longitude	INTERVAL	0	0	-0.82311	0.057074	-0.414829599	-0.3830	0.2297913081	0.2659481665	-1.01812926	
7	num_of_vehi	NOMINAL	0	0	.	.	.	.	.	.	.	
8	road_type_d	NOMINAL	0	0	.	.	.	.	.	.	.	
9	speed_limit	NOMINAL	0	0	.	.	.	.	.	.	.	
10	weath_con_d	NOMINAL	0	0	.	.	.	.	.	.	.	
11	_months_	NOMINAL	0	0	.	.	.	.	.	.	.	

Predicting_Accidents_D

"Transformations_1" Results

by: di00222@surrey.ac.uk

Contents

Input Variable Statistics

Name	Variable Level	Number of Missing Values	Percent Missing
BIN_first_road_class	NOMINAL	0	0
BIN_num_of_vehi	NOMINAL	0	0
BIN_road_type_d	NOMINAL	0	0
day_of_week	NOMINAL	0	0
junc_detail_d	NOMINAL	0	0
light_con_d	NOMINAL	0	0
LOG_longitude	INTERVAL	0	0
speed_limit	NOMINAL	0	0
SQR_latitude	INTERVAL	0	0
weath_con_d	NOMINAL	0	0
months	NOMINAL	0	0

Minimum	Maximum	Mean	Midrange
0	0.6314	0.3291	0.3157
2,609.4945	2,648.7875	2,630.8578	2,629.1410

Importance Standard Deviation	Skewness	Kurtosis	Variable Label
			Transformed first_road_class
			Transformed num_of_vehi
			Transformed road_type_d

Importance Standard Deviation	Skewness	Kurtosis	Variable Label
0.1629	0.0429	-1.0776	Transformed longitude
8.7879	-0.2628	-0.5231	Transformed latitude

Transformation Summary

Transformed Variable	Method	Ranking Criterion for Best Transformation	Input Variable
SQRT_LOG_longitude	SQRT	Moment Skewness	LOG_longitude
SQR_SQR_latitude	SQUARE	Moment Skewness	SQR_latitude
WOEENC_BIN_first_road_class	WOEENCODE		BIN_first_road_class
WOEENC_BIN_num_of_vehi	WOEENCODE		BIN_num_of_vehi
WOEENC_BIN_road_type_d	WOEENCODE		BIN_road_type_d
WOEENC__months_	WOEENCODE		_months_
WOEENC_day_of_week	WOEENCODE		day_of_week
WOEENC_junc_detail_d	WOEENCODE		junc_detail_d
WOEENC_light_con_d	WOEENCODE		light_con_d
WOEENC_speed_limit	WOEENCODE		speed_limit
WOEENC_weath_con_d	WOEENCODE		weath_con_d

Formula	Number of Missing Values	Minimum	Maximum
SQRT('LOG_longitude'n + 1)	0	1	1.2773
('SQR_latitude'n)**2	0	6,809,461.7973	7,016,075.4838
WOEENC	0	-0.5311	2.3602
WOEENC	0	-1.8080	1.8136

Formula	Number of Missing Values	Minimum	Maximum
WOEENC	0	-0.5676	4.4330
WOEENC	0	-1.3042	5.2593
WOEENC	0	-0.5042	1.1244
WOEENC	0	-0.2950	4.9879
WOEENC	0	-0.3936	3.3860
WOEENC	0	-0.8753	4.7485
WOEENC	0	-0.1685	4.0697

Mean	Standard Deviation	Skewness	Kurtosis
1.1507	0.0708	-0.0428	-1.0690
6,921,489.8566	46,219.4444	-0.2558	-0.5299
0.0295	0.4474	0.7008	1.0097
0.0401	0.7168	-1.1540	0.2403
0.1243	0.9198	3.6668	13.9075
0.7619	2.2655	1.2349	-0.1873
0.0347	0.4568	1.3354	1.3087
0.2431	1.2864	2.9847	7.6979
0.0348	0.4753	5.2017	33.3837
0.2163	1.2295	2.4739	6.5290
0.1100	0.8309	3.4776	11.8409

Variable Label
Transformed Transformed longitude
Transformed Transformed latitude
Transformed

Variable Label
Transformed first_road_class
Transformed Transformed num_of_vehi
Transformed Transformed road_type_d
Transformed _months_
Transformed day_of_week
Transformed junc_detail_d
Transformed light_con_d
Transformed speed_limit
Transformed weath_con_d

Properties

Property Name	Property Value
bestRankBinTarg	CLASSICALSKEW
bestRankIntTarg	CLASSICALSKEW
bestRankNomTarg	CLASSICALSKEW
codeLocation	mlearning
cutoffPercent	0.5000
dataMiningVersion	V2024.09
defaultMethod	BEST
defltClassMethod	WOEENCODE
fullDatasetReconstitution	false
ignoreMetadata	true
minObsinBin	5
missingTreatment	SEPARATE
numBins	4
outlierBins	false
rejectVariable	true
reportingOnly	false
summaryStatistics	true
templateRevision	2
woeAdjust	0.5000

Output

Best Variable Transformation

Obs	Variable	Transformation Name	N	NMiss	Moment Skewness
1	LOG_longitude	Square Root	2464	0	-0.04281
2	SQR_latitude	Square	2464	0	-0.2558

Variable Transformation Ranking

Obs	Variable	Transformation Name	Best Transform Rank	N	NMiss	Moment Skewness
1	LOG_longitude	Square Root	1	2464	0	-0.04281
2	LOG_longitude	None	2	2464	0	0.04286
3	LOG_longitude	Center	2	2464	0	0.04286
4	LOG_longitude	Range	2	2464	0	0.04286
5	LOG_longitude	Standardize	2	2464	0	0.04286
6	LOG_longitude	Log	6	2464	0	-0.1294
7	LOG_longitude	Log10	6	2464	0	-0.1294
8	LOG_longitude	Inverse Square Root	8	2464	0	0.2168
9	LOG_longitude	Inverse	9	2464	0	0.3046
10	LOG_longitude	Inverse Square	10	2464	0	-0.4435
11	LOG_longitude	Square	11	2464	0	0.6696
12	SQR_latitude	Square	1	2464	0	-0.2558
13	SQR_latitude	Center	2	2464	0	-0.2628
14	SQR_latitude	Range	2	2464	0	-0.2628
15	SQR_latitude	Standardize	2	2464	0	-0.2628
16	SQR_latitude	None	5	2464	0	-0.2628
17	SQR_latitude	Inverse Square	6	2464	0	0.2840
18	SQR_latitude	Square Root	7	2464	0	-0.9638
19	SQR_latitude	Log	8	2464	0	-2.1241
20	SQR_latitude	Log10	8	2464	0	-2.1241
21	SQR_latitude	Inverse Square Root	10	2464	0	3.7848
22	SQR_latitude	Inverse	11	2464	0	5.6092

Input Variable Statistics

Obs	Input Variable	Measurement Level	Number of Missing Values	Percentage Missing	Minimum	Maximum	Mean	Midrange	Standard Deviation	Skewness	Kurtosis	Label
1	BIN_first_road_class	NOMINAL	0	0	.	.	.	.	.	.	.	Transformed first_road_class
2	BIN_num_of_vehi	NOMINAL	0	0	.	.	.	.	.	.	.	Transformed num_of_vehi
3	BIN_road_type_d	NOMINAL	0	0	.	.	.	.	.	.	.	Transformed road_type_d
4	day_of_week	NOMINAL	0	0	.	.	.	.	.	.	.	
5	junc_detail_d	NOMINAL	0	0	.	.	.	.	.	.	.	
6	light_con_d	NOMINAL	0	0	.	.	.	.	.	.	.	
7	LOG_longitude	INTERVAL	0	0	0	0.6313696444	0.329109044	0.32	0.1628901619	0.0428636523	-1.077583367	Transformed longitude
8	speed_limit	NOMINAL	0	0	.	.	.	.	.	.	.	
9	SQR_latitude	INTERVAL	0	0	2609.4945482	2648.7875498	2630.8577803	2629.14	8.7879384931	-0.262799664	-0.523127481	Transformed latitude
10	weath_con_d	NOMINAL	0	0	.	.	.	.	.	.	.	
11	_months_	NOMINAL	0	0	.	.	.	.	.	.	.	

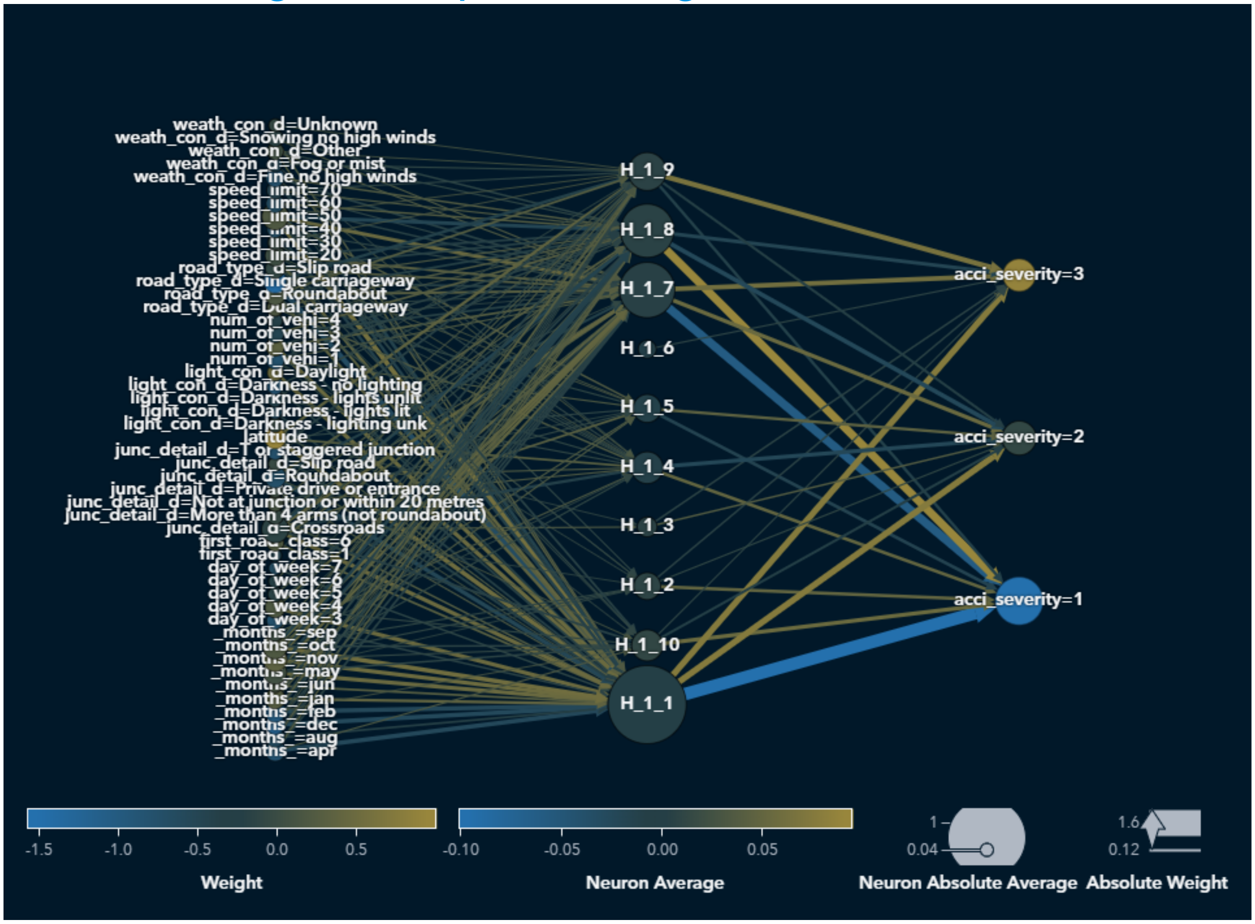
Predicting_Accidents_D

"Neural Network" Results

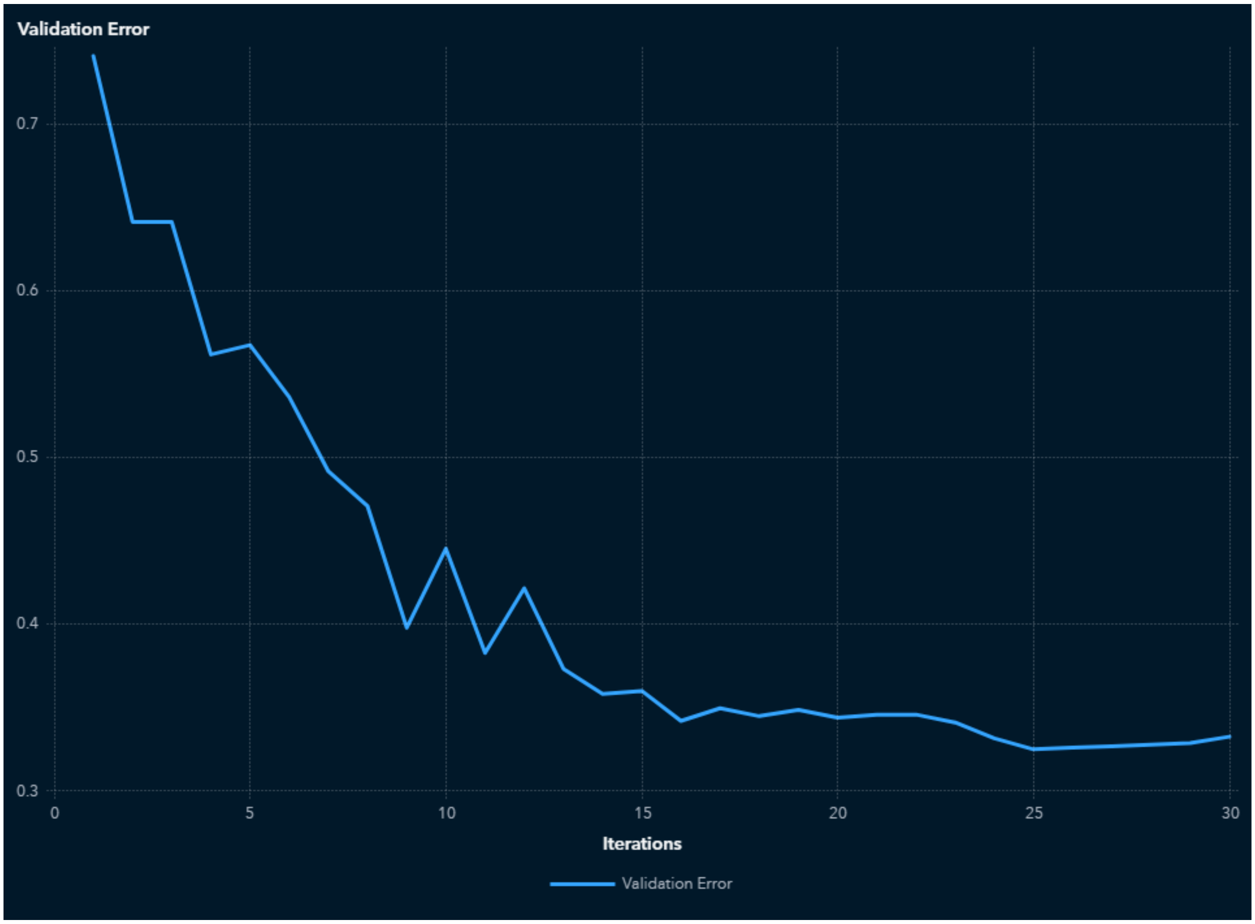
by: di00222@surrey.ac.uk

Contents

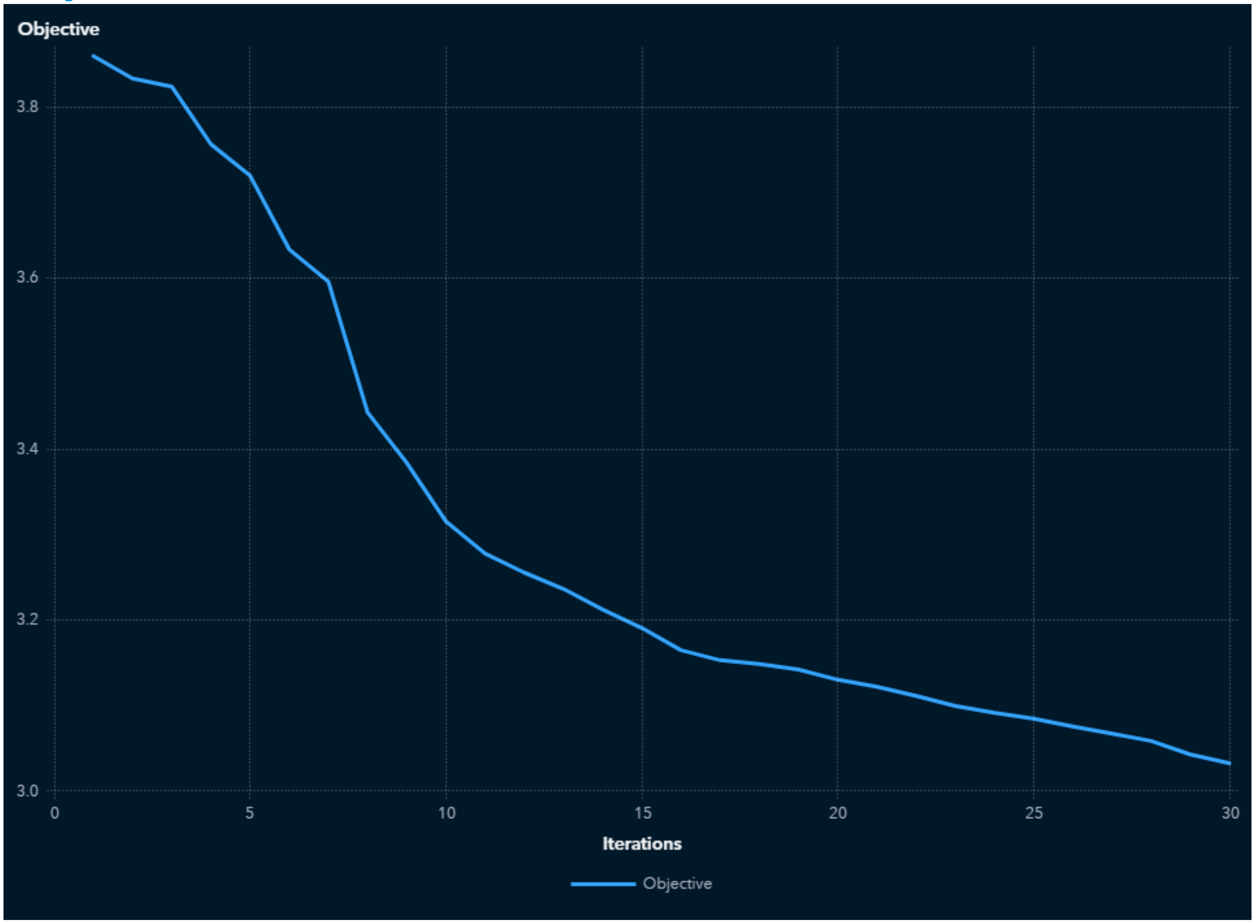
Network Diagram: Top 200 Weights



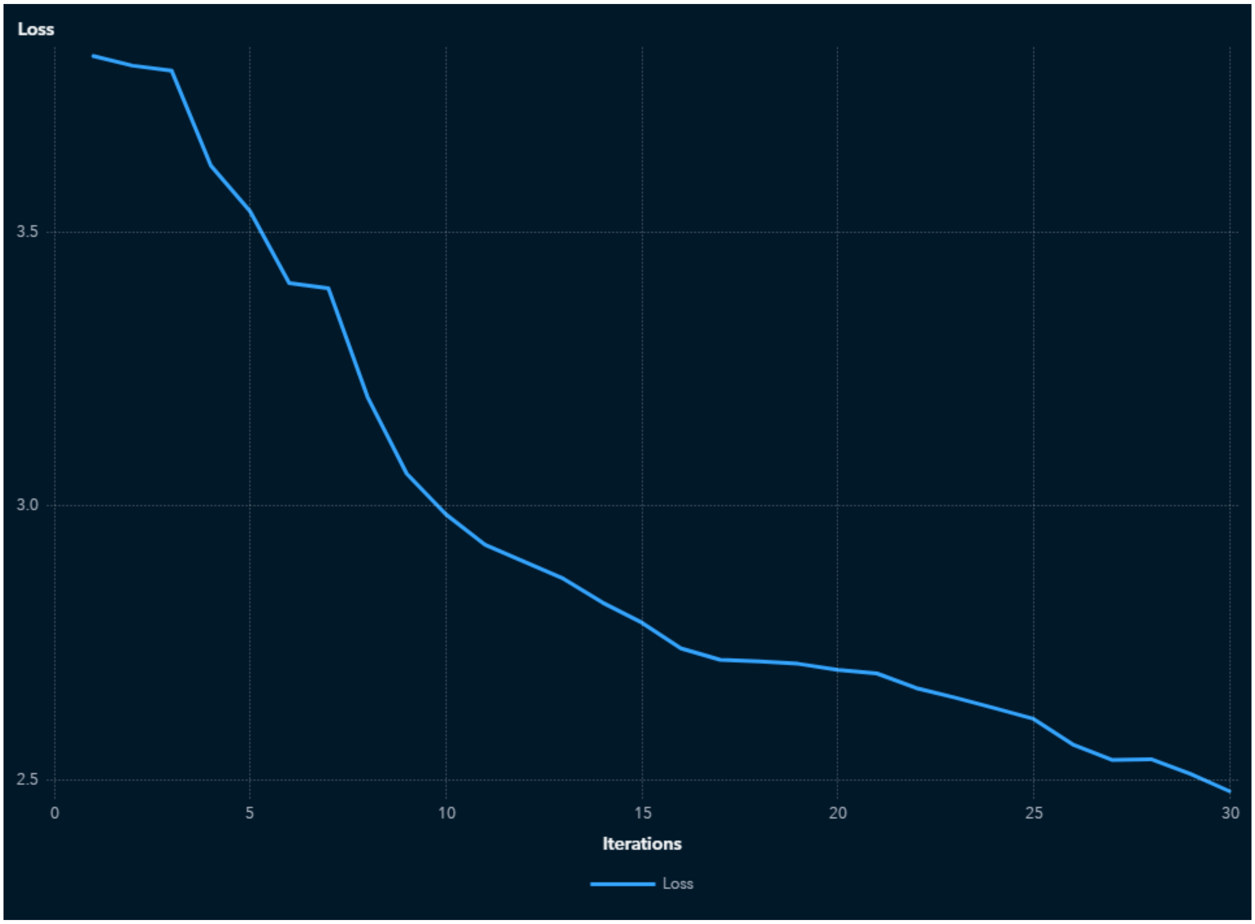
Validation Error



Objective


Objective
3.8
3.6
3.4
3.2
3.0
0
5
10
15
20
25
30
Iterations
Objective

Loss



Score Inputs

Name	Role	Variable Level	Type
day_of_week	INPUT	NOMINAL	N
first_road_class	INPUT	NOMINAL	N
junc_detail_d	INPUT	NOMINAL	C
latitude	INPUT	INTERVAL	N
light_con_d	INPUT	NOMINAL	C
num_of_vehi	INPUT	NOMINAL	N
road_type_d	INPUT	NOMINAL	C
speed_limit	INPUT	NOMINAL	N
weath_con_d	INPUT	NOMINAL	C
months	INPUT	NOMINAL	C

Variable Type	Variable Label	Variable Format	Variable Length
double			8
double			8
varchar			35
double			8
varchar			23
double			8
varchar			18
double			8
varchar			21
varchar			3

Score Outputs

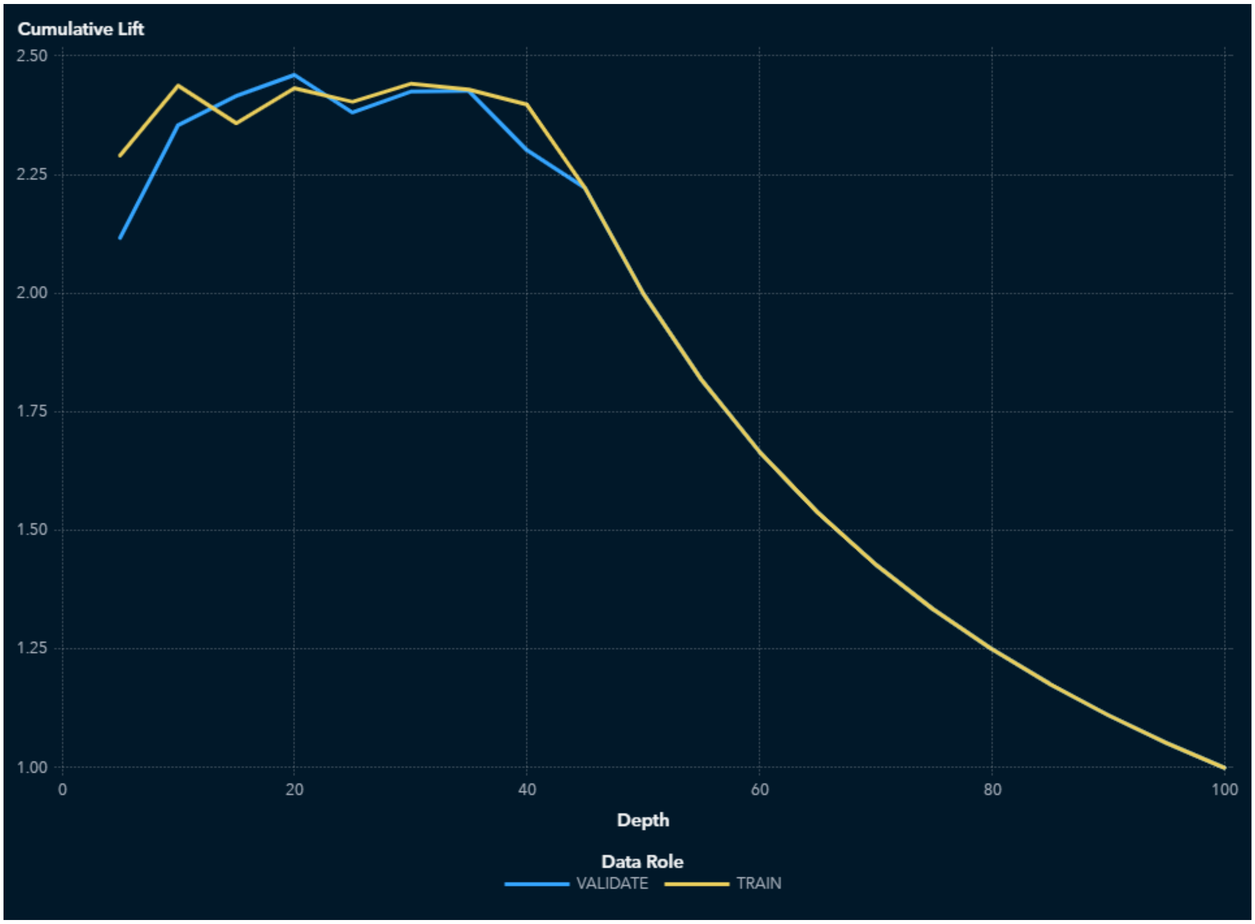
Name	Role	Type	Variable Type
EM_CLASSIFICATION	CLASSIFICATION	C	char
EM_EVENTPROBABILITY	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
I_acci_severity	CLASSIFICATION	C	char
P_acci_severity1	PREDICT	N	double
P_acci_severity2	PREDICT	N	double
P_acci_severity3	PREDICT	N	double

Variable Label	Variable Format	Variable Length	Creator
Predicted for acci_severity		12	neural
Probability for acci_severity=1		8	neural
Probability of Classification		8	neural
Into: acci_severity		13	neural
Predicted: acci_severity=1		8	neural
Predicted: acci_severity=2		8	neural
Predicted: acci_severity=3		8	neural

Function	Creator GUID
CLASSIFICATION	38dc5b6c-2d9a-4da3-8489-dfe611ea2ab3
PREDICT	38dc5b6c-2d9a-4da3-8489-

Function	Creator GUID
	dfe611ea2ab3
PREDICT	38dc5b6c-2d9a-4da3-8489-dfe611ea2ab3
CLASSIFICATION	38dc5b6c-2d9a-4da3-8489-dfe611ea2ab3
PREDICT	38dc5b6c-2d9a-4da3-8489-dfe611ea2ab3
PREDICT	38dc5b6c-2d9a-4da3-8489-dfe611ea2ab3
PREDICT	38dc5b6c-2d9a-4da3-8489-dfe611ea2ab3

Cumulative Lift



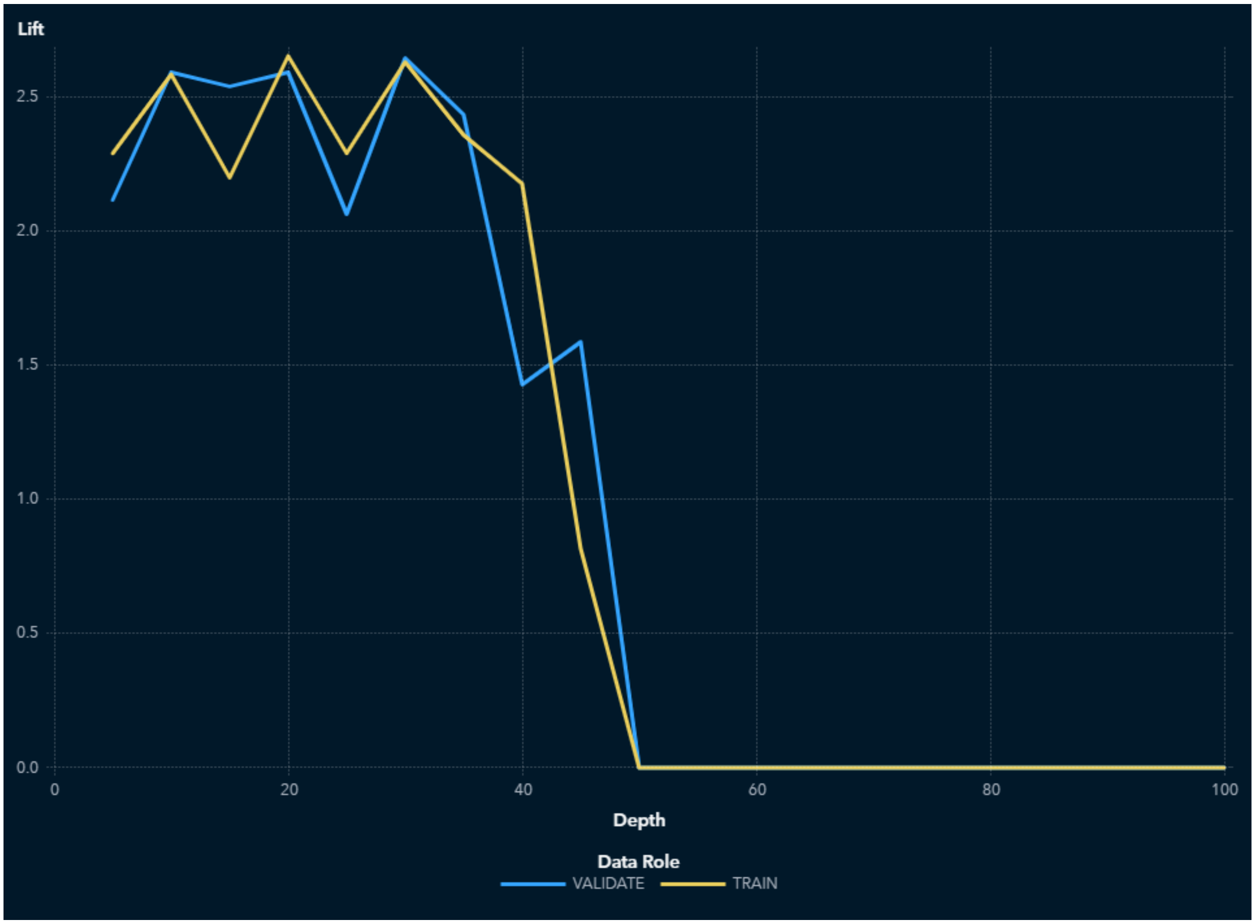
The VALIDATE partition has a Cumulative Lift of 2.35 in the 10% quantile (depth of 10) meaning there are 2.35 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Cumulative Lift of 2.44 in the 10% quantile (depth of 10) meaning there are 2.44 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



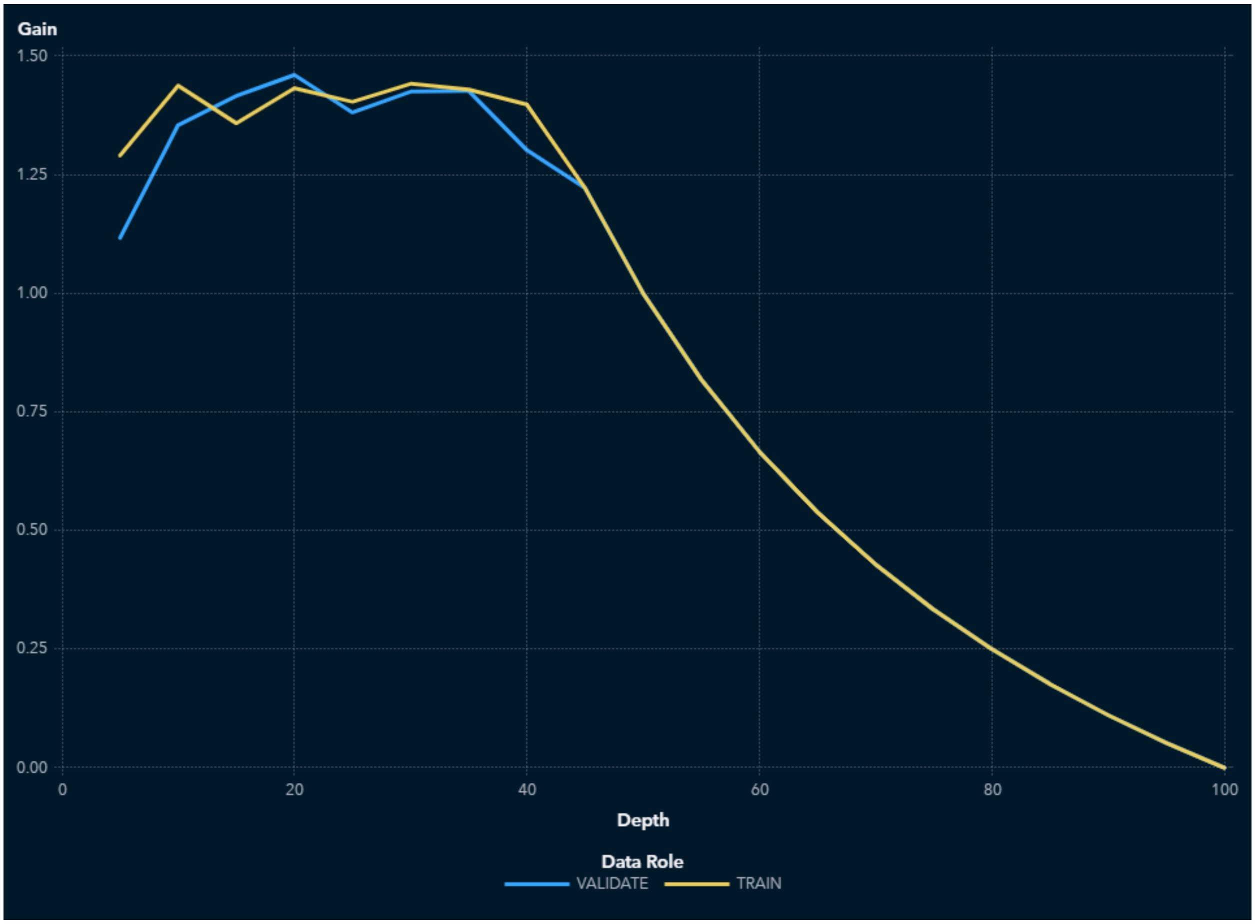
The VALIDATE partition has a Lift of 2.12 in the 5% quantile (depth of 5) meaning there are 2.12 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Lift of 2.29 in the 5% quantile (depth of 5) meaning there are 2.29 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



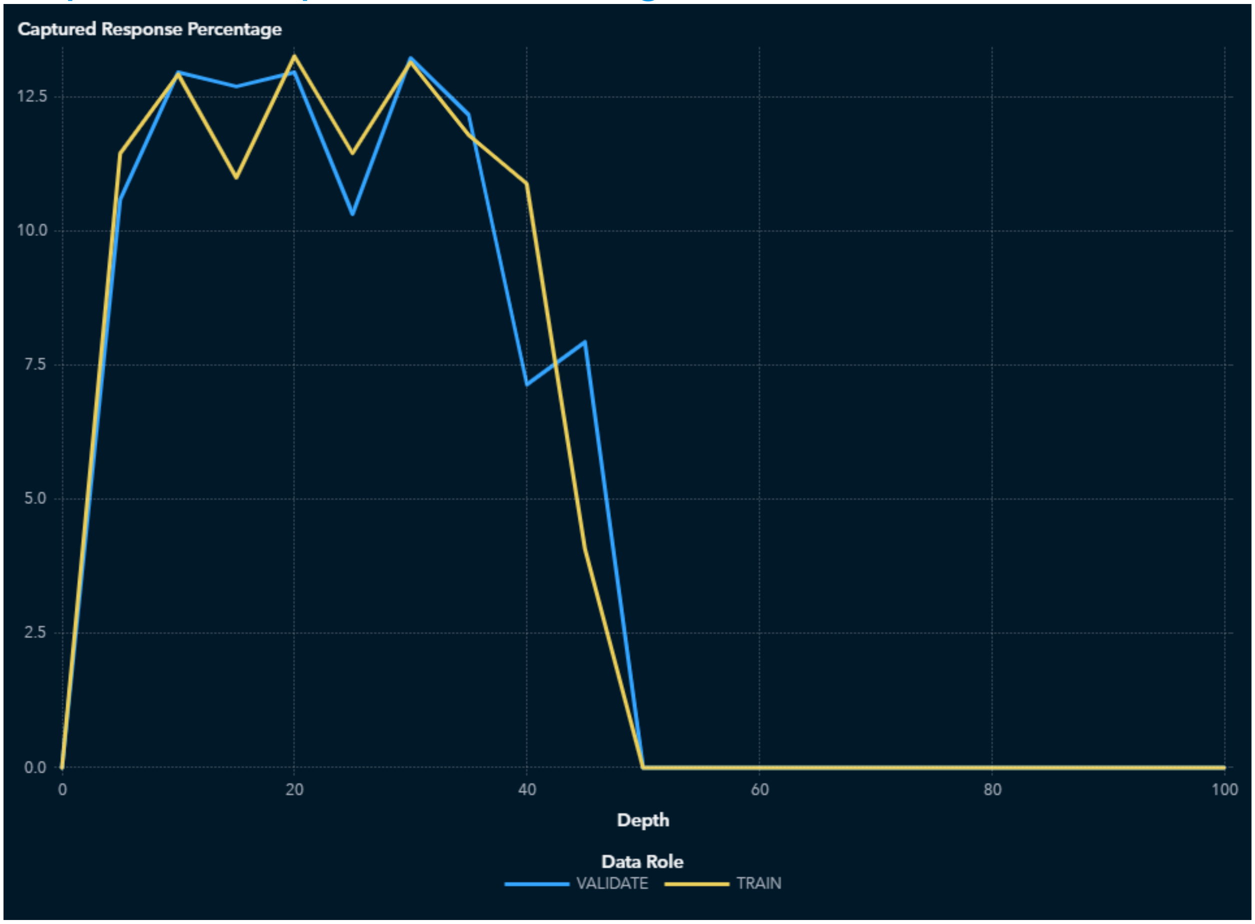
The VALIDATE partition has a Gain of 1.4 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 1.8.

The TRAIN partition has a Gain of 1.4 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 1.81.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

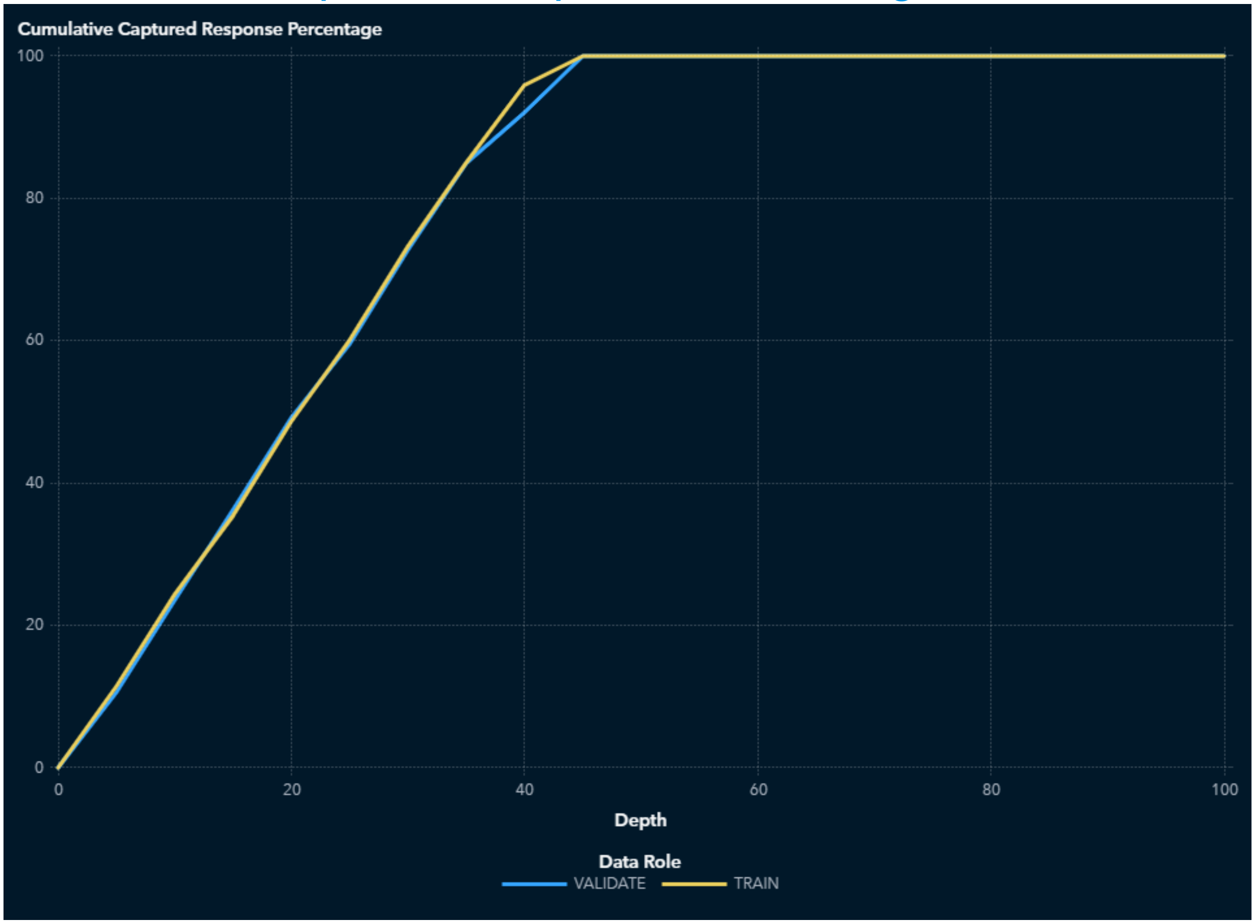


At the 5% quantile (depth of 5), the VALIDATE partition has a Captured response percentage of 10.6 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 14.02.

At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 11.5 (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 14.06.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



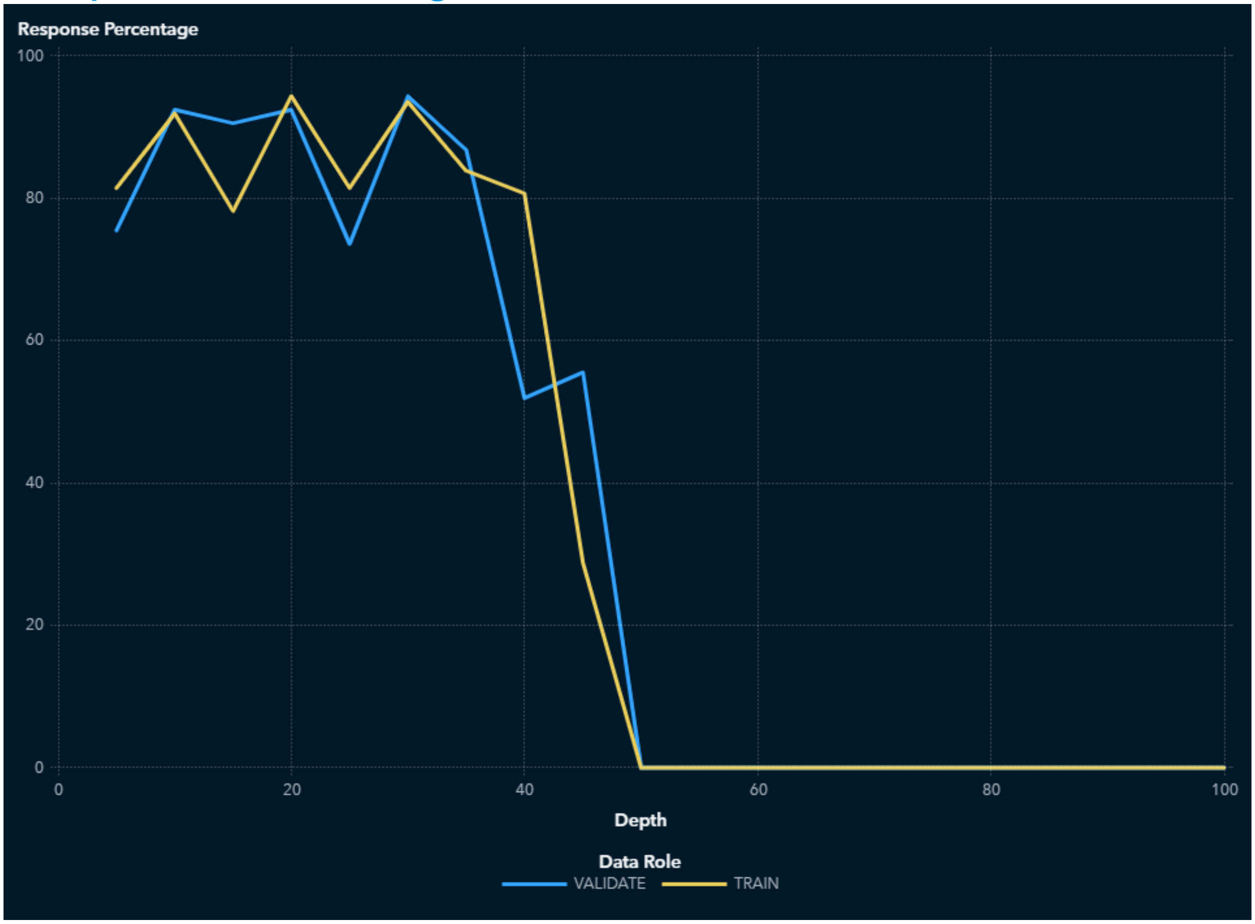
In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative captured response percentage of 23.5 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 28.04.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 24.4 (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 28.12.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

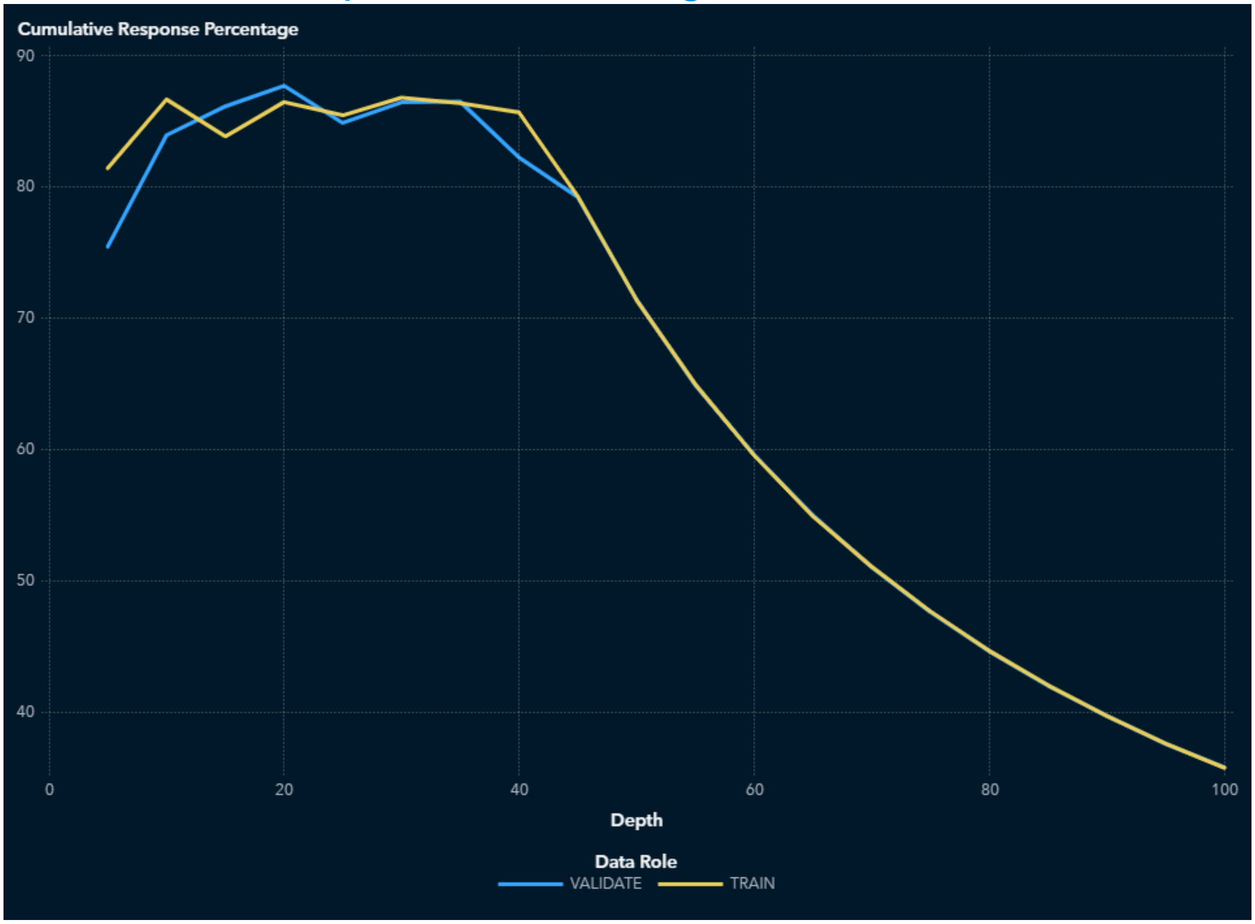


At the 5% quantile (depth of 5), the VALIDATE partition has a Response percentage of 75.5. The best possible value of Response percentage for this partition at depth 5 is 100.

At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 81.5. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

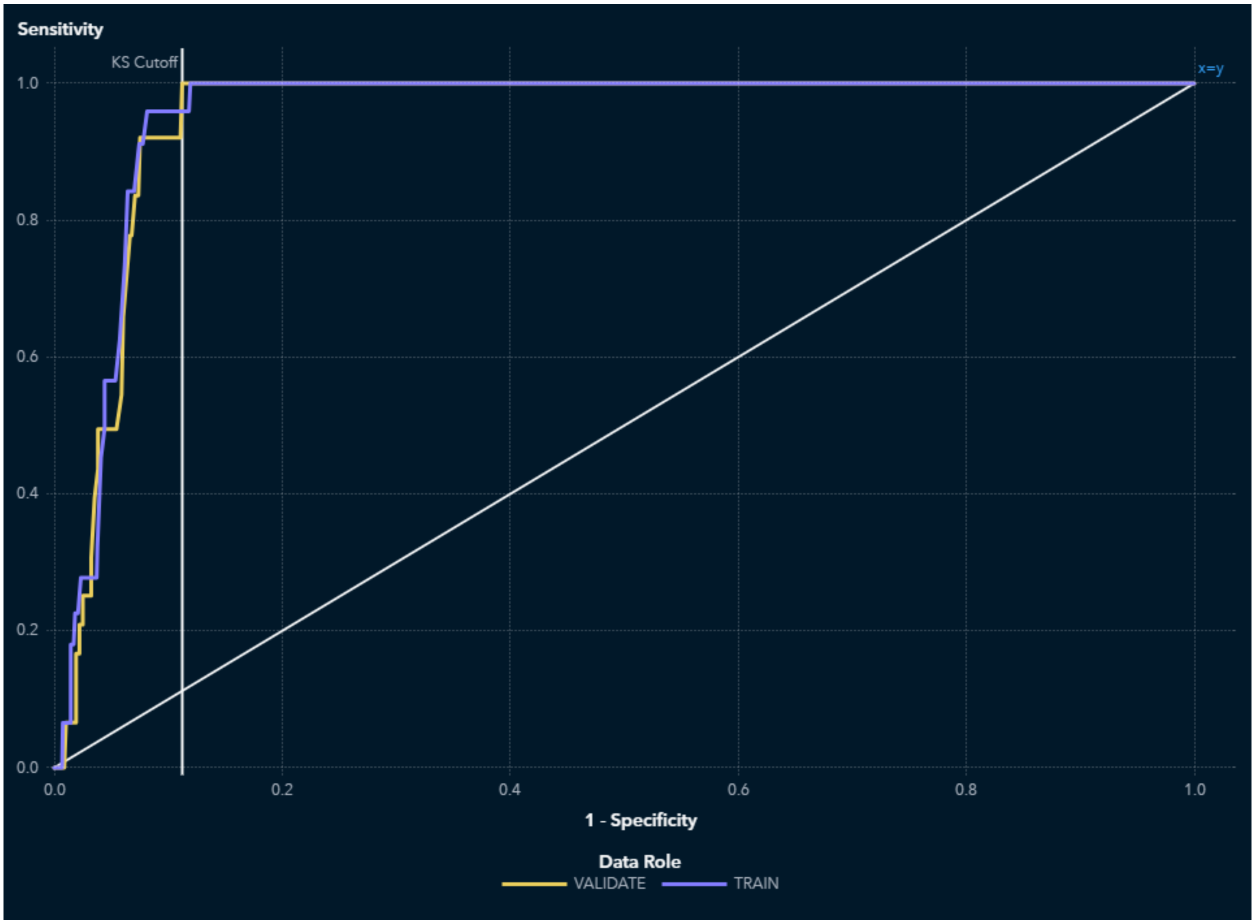


In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative response percentage of 84. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 86.7. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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 VALIDATE partition. The KS Cutoff line is drawn at the cutoff value 0.48, where the 1-specificity value is 0.112 and the sensitivity value is 1.

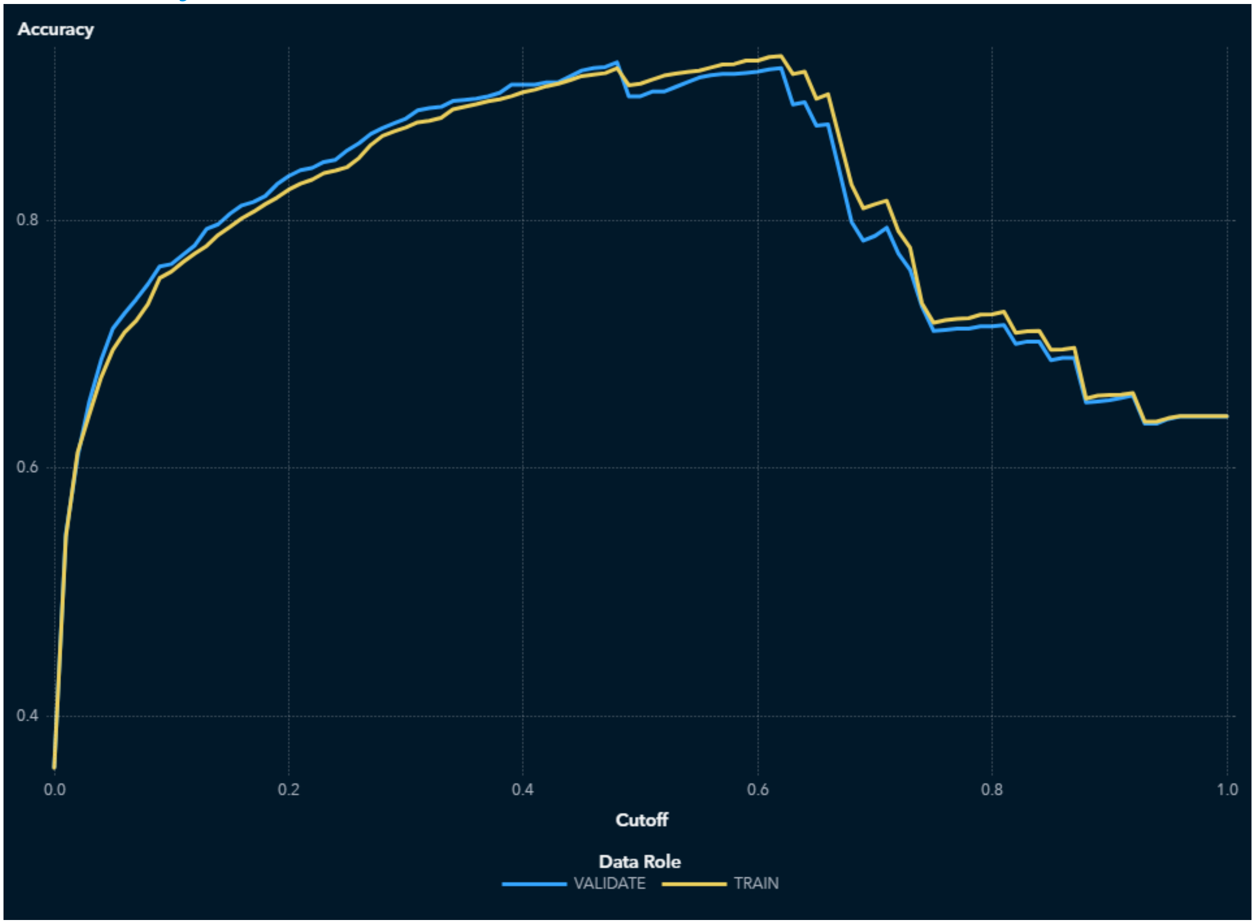
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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

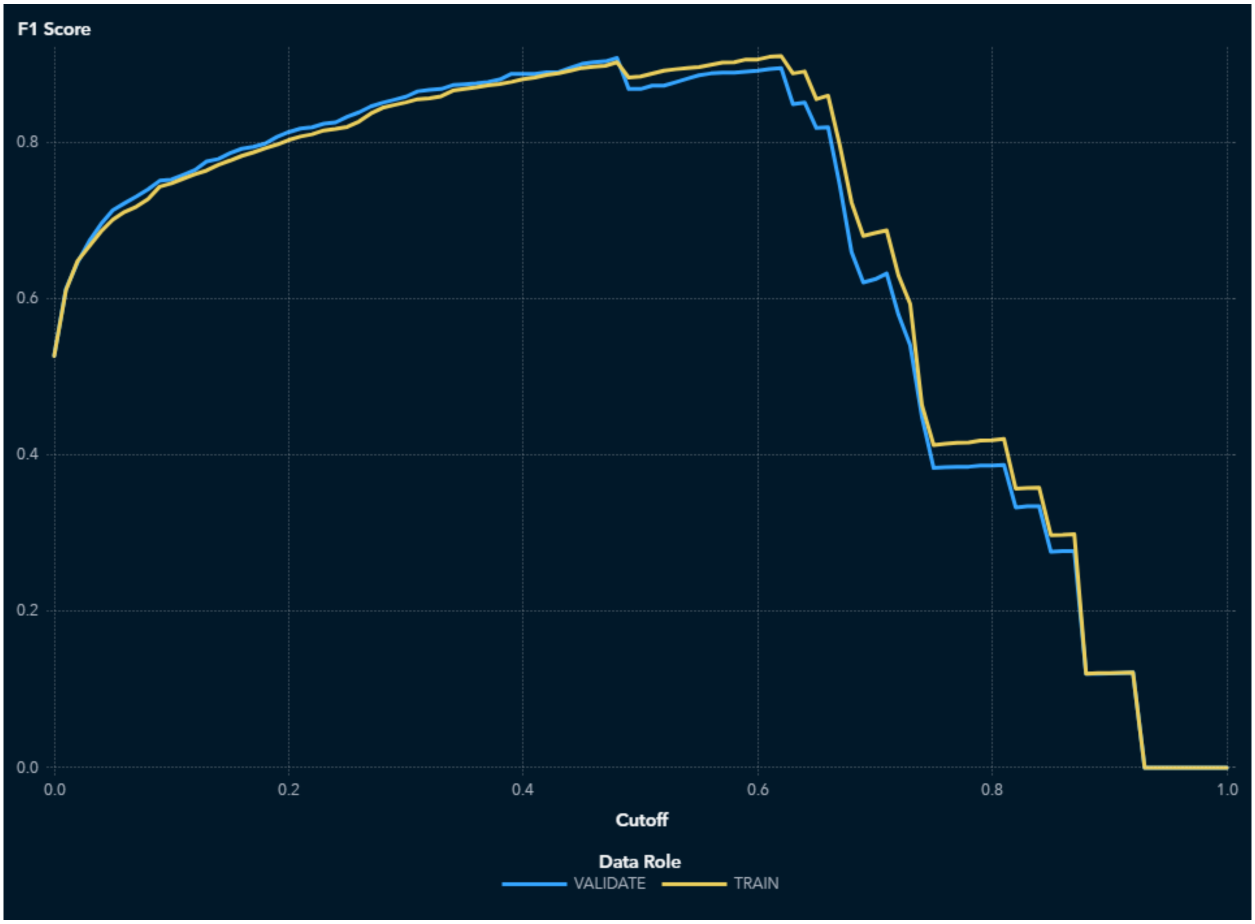


For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.911.

For this model, the accuracy in the VALIDATE partition at the cutoff of 0.5 is 0.9.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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 0.885.

For this model, the F1 score in the VALIDATE partition at the cutoff of 0.5 is 0.869.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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
acci_severity	TRAIN	1	1
acci_severity	VALIDATE	0	0

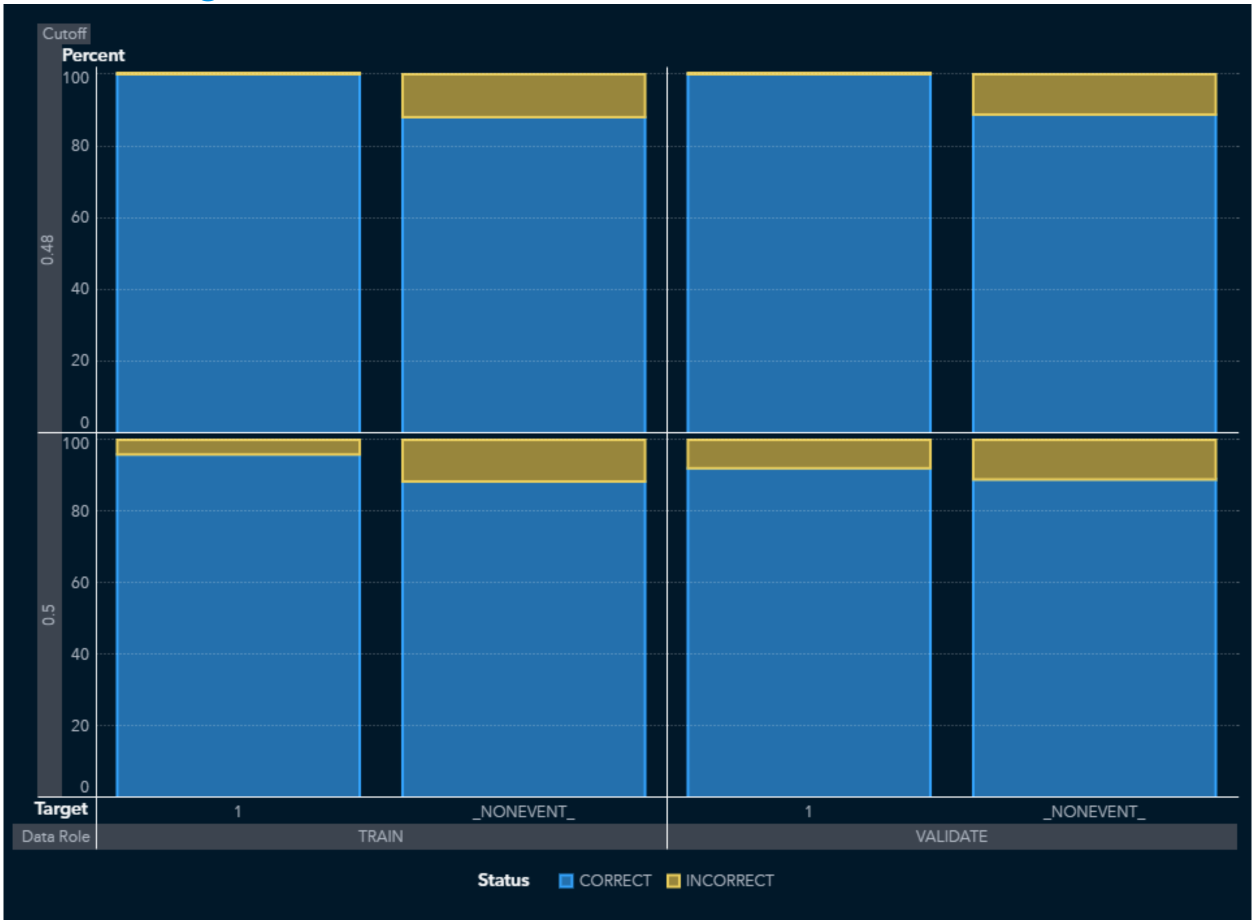
Number of Observations	Average Squared Error	Divisor for ASE	Root Average Squared Error
2,464	0.1409	2,464	0.3754
1,055	0.1450	1,055	0.3807

Misclassification Rate	Multi-Class Log Loss	KS (Youden)	Area Under ROC
0.3279	0.7021	0.8805	0.9538
0.3251	0.7140	0.8877	0.9504

Gini Coefficient	Gamma	Tau	KS Cutoff
0.9077	0.9098	0.4174	0.4800
0.9009	0.9027	0.4146	0.4800

KS at Default Cutoff	Misclassification Rate at KS Cutoff (Event)	Misclassification Rate (Event)
0.8429	0.0767	0.0893
0.8099	0.0720	0.0995

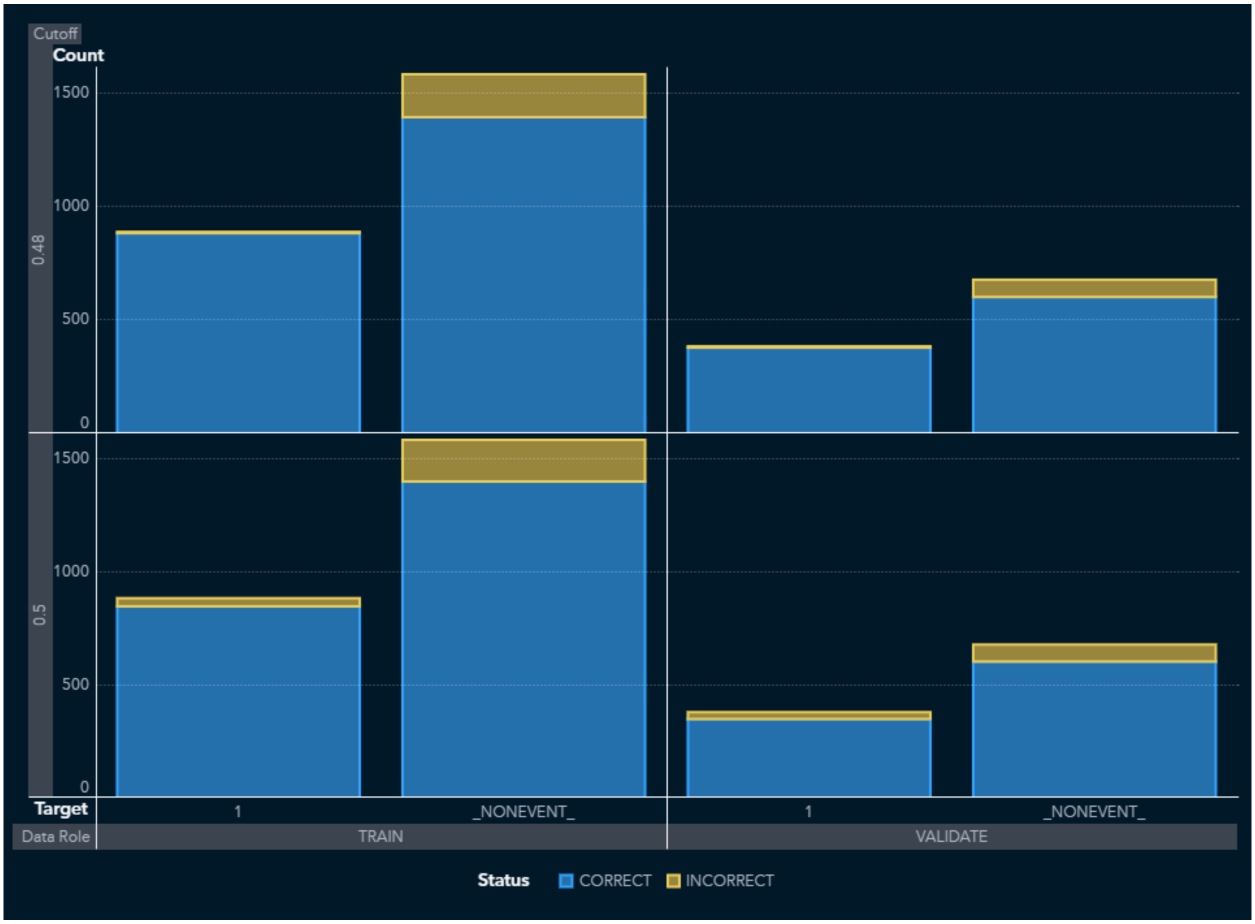
Percentage Plot



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: 0.48 (TRAIN), 0.48 (VALIDATE).

For this data, for the bar corresponding to the event level of acci_severity, "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 these KS cutoff values for existing partitions: 0.48 (TRAIN), 0.48 (VALIDATE).

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

Table

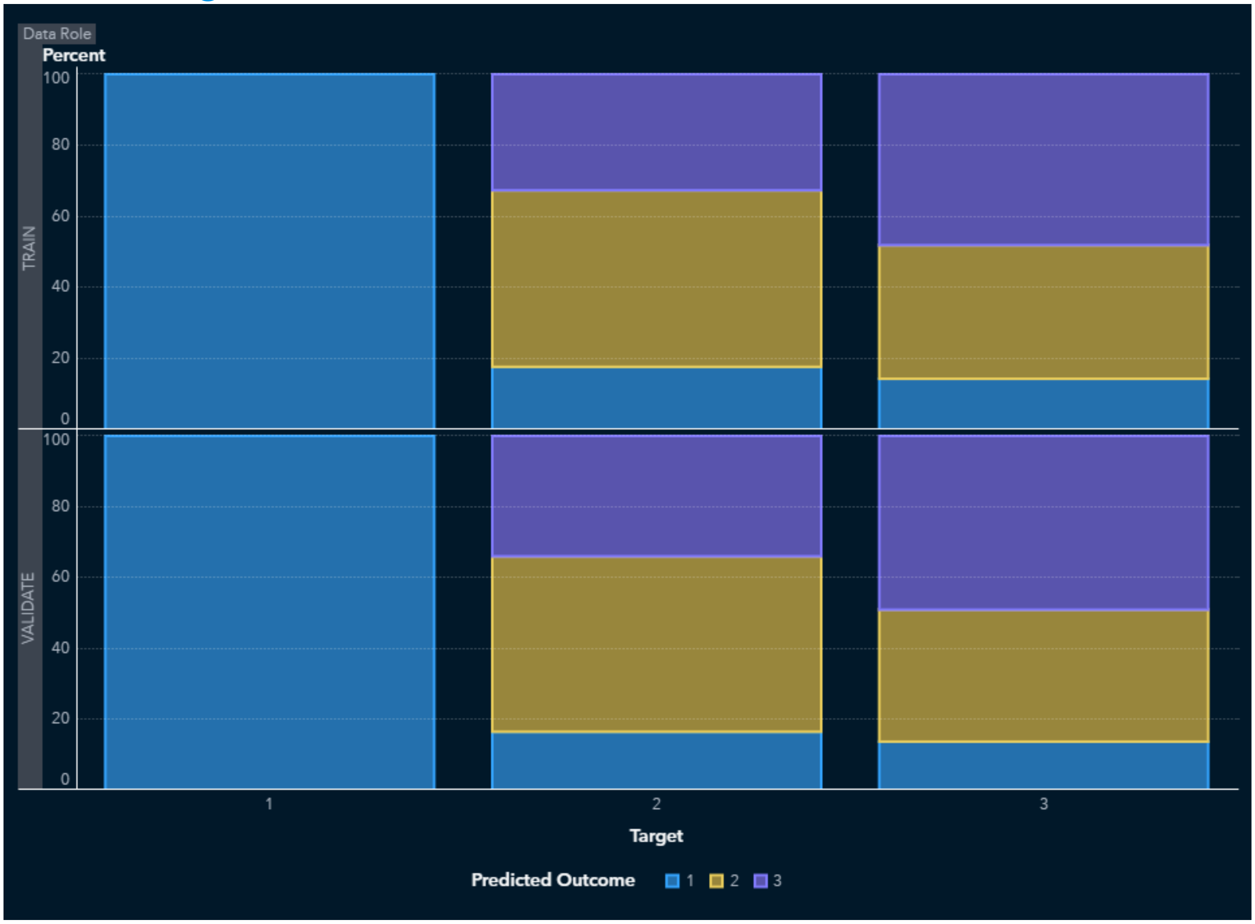
Cutoff	Cutoff Source	Target Name	Response
0.4800	KS	acci_severity	CORRECT
0.4800	KS	acci_severity	INCORRECT
0.4800	KS	acci_severity	CORRECT
0.4800	KS	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT

Event	Value	Training Frequency	Validation Frequency
1	True Positive	882	378
1	False Negative	0	0
NONEVENT	True Negative	1,393	601
NONEVENT	False Positive	189	76
1	True Positive	846	348
1	False Negative	36	30
NONEVENT	True Negative	1,398	602
NONEVENT	False Positive	184	75

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	100	100	
	0	0	
	88.0531	88.7740	
	11.9469	11.2260	
	95.9184	92.0635	
	4.0816	7.9365	
	88.3692	88.9217	

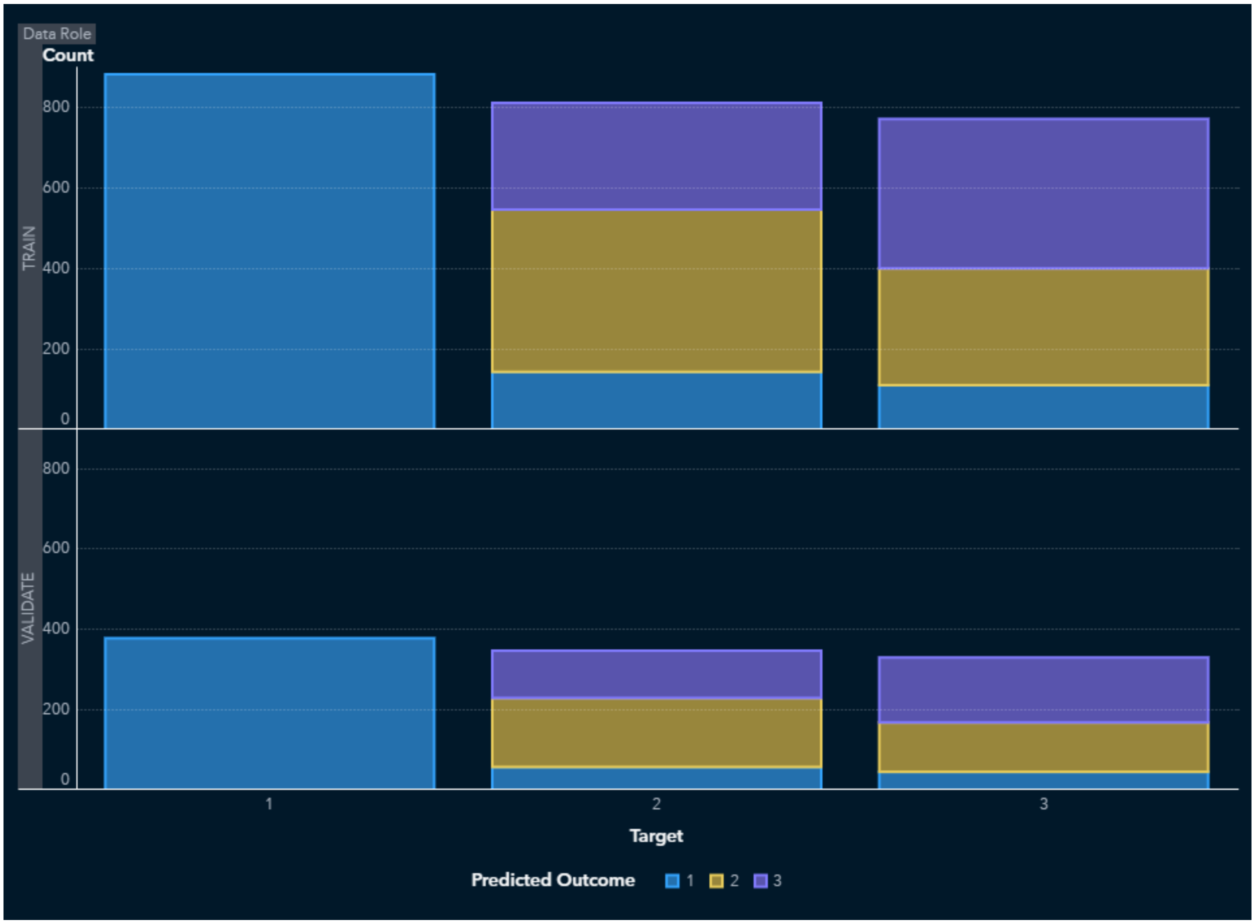
Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	11.6308	11.0783	

Percentage Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Count Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Table

Target Name	Data Role	Target	Unformatted Target
acci_severity	VALIDATE	1	1
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	3	3
acci_severity	VALIDATE	3	3
acci_severity	VALIDATE	3	3
acci_severity	TRAIN	1	1
acci_severity	TRAIN	2	2
acci_severity	TRAIN	2	2
acci_severity	TRAIN	2	2
acci_severity	TRAIN	3	3
acci_severity	TRAIN	3	3
acci_severity	TRAIN	3	3

Predicted Outcome	Count	Percent	Status
1	378	100	CORRECT
1	57	16.4265	INCORRECT
2	172	49.5677	CORRECT
3	118	34.0058	INCORRECT
1	45	13.6364	INCORRECT
2	123	37.2727	INCORRECT
3	162	49.0909	CORRECT
1	882	100	CORRECT
1	143	17.6326	INCORRECT
2	403	49.6917	CORRECT
3	265	32.6757	INCORRECT

Predicted Outcome	Count	Percent	Status
1	110	14.2672	INCORRECT
2	290	37.6135	INCORRECT
3	371	48.1193	CORRECT

Properties

Property Name	Property Value
actFunc1	TANH
actFunc10	TANH
actFunc2	TANH
actFunc3	TANH
actFunc4	TANH
actFunc5	TANH
actFunc6	TANH
actFunc7	TANH
actFunc8	TANH
actFunc9	TANH
actFuncAll	TANH
annealingRate	0.0000
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

Property Name	Property Value
atUpdateProperties	false
atUseLookup	false
atValidFold	5
atValidMethod	PARTITION
atValidProp	0.3000
atannealingRate	true
atannealingRateInit	0.0010
atannealingRateLB	0.0000
atannealingRateUB	0.1000
atlearningRate	true
atlearningRateInit	0.0010
atlearningRateLB	0
atlearningRateUB	0.1000
atnhidden	true
atnhiddenInit	1
atnhiddenLB	0
atnhiddenUB	2
atnunitsInit	1
atnunitsLB	1
atnunitsUB	100
atweightDecay1	true
atweightDecay1Init	0
atweightDecay1LB	0
atweightDecay1UB	10
atweightDecay2	true
atweightDecay2Init	0
atweightDecay2LB	0
atweightDecay2UB	10
autotune_enabled	false

Property Name	Property Value
binaryProbCutoff	0.5000
codeLocation	mlearning
dataMiningVersion	V2024.09
directConn	false
dnnAlg	ADAM
dnnBeta1	0.9000
dnnBeta2	0.9990
dnnGamma	0.1000
dnnLRPolicy	FIXED
dnnMaxEpochs	10
dnnMomentum	0.9000
dnnPower	0.7500
dnnStepSize	10
earlyStop	true
exactPctlLift	true
explainFidelity	false
explainInfo	false
fullDatasetReconstitution	false
goal	0
hidden1	50
hidden10	50
hidden2	50
hidden3	50
hidden4	50
hidden5	50
hidden6	50
hidden7	50
hidden8	50

Property Name	Property Value
hidden9	50
hiddenAll	true
hiddenAllNum	10
hiddenDropout	0
icePlots	false
inputDropout	0
inputStd	MIDRANGE
learningRate	0.0010
maxIter	300
maxNumShapVars	20
maxTime	0
miniBatchSize	50
missAsLevl	false
momentum	0
nBins	50
nHidden	1
numCorrections	6
numTries	1
optTech	AUTOMATIC
pdNumImportantInputs	5
pdObsSamples	1,000
pdPlots	false
performKernelShap	false
performLime	false
performVI	false
randomSeed	12,345
reportingOnly	false
seedId	12,345

Property Name	Property Value
sgdSeed	12,345
specifyRows	RANDOM
stagnation	5
targetAct	IDENTITY
targetError	NORMAL
targetStd	MIDRANGE
templateRevision	5
train	true
truncateLl	5
truncateUl	95
useLocking	false
userProbCutoff	false
weightDecay	0.1000
weightDecay1	0

Output

The SAS System

The NNET Procedure

Model Information	
Model	Neural Net
Number of Observations Used	2464
Number of Observations Read	2464
Target/Response Variable	acci_severity
Number of Nodes	79
Number of Input Nodes	66
Number of Output Nodes	3
Number of Hidden Nodes	10
Number of Hidden Layers	1
Number of Weight Parameters	690
Number of Bias Parameters	13
Architecture	MLP
Seed for Initial Weight	12345
Optimization Technique	LBFGS
Number of Neural Nets	1
Objective Value	3.0848500886
Misclassification Rate for Validation	0.3251

Iteration History

Iteration Number	Objective Function	Norm of Gradient	Loss	Validate Error	Step Size	Norm L1	Norm L2	Norm Maximum	Fit Error
1	3.859734	0.165198	3.820517	0.741232	0	6.685947	0.392173	0.094916	0.726867
2	3.833423	0.134721	3.803013	0.641706	1.443735	5.431683	0.304109	0.061618	0.642045
3	3.824056	0.155049	3.793688	0.641706	1	4.911209	0.303678	0.079760	0.642045
4	3.756622	0.760358	3.620708	0.562085	2.943247	22.08051	1.359142	0.518357	0.587662
5	3.719942	0.843697	3.537627	0.567773	0.371182	30.48902	1.823146	0.669670	0.598214
6	3.633714	0.397437	3.406606	0.536493	0.478453	39.12346	2.271082	0.786585	0.548295
7	3.595825	0.370001	3.397017	0.491943	1	33.86417	1.988080	0.701726	0.483360
8	3.442952	0.641536	3.198941	0.471090	1	40.12429	2.440119	0.903929	0.467532
9	3.384144	0.508896	3.059261	0.398104	1	52.58942	3.248831	1.214443	0.396916
10	3.315500	0.332750	2.984968	0.445498	1	54.14533	3.305326	1.208043	0.409091
11	3.277881	0.321454	2.929809	0.382938	1	56.93601	3.480723	1.258138	0.368506
12	3.255640	0.316106	2.898702	0.421801	1	57.61898	3.569374	1.280117	0.383523
13	3.236558	0.229659	2.867795	0.373460	1	59.14146	3.687637	1.327787	0.362419
14	3.212485	0.225033	2.824301	0.358294	1	61.49322	3.881836	1.401572	0.353490
15	3.190892	0.222562	2.787914	0.360190	1	63.22318	4.029780	1.447378	0.348214
16	3.164949	0.164473	2.740698	0.342180	1	65.98564	4.242511	1.499981	0.330763
17	3.153177	0.147048	2.720087	0.349763	1	67.30262	4.330896	1.502084	0.330357
18	3.148723	0.295471	2.717260	0.345024	0.337585	67.15731	4.314628	1.491850	0.327110
19	3.142228	0.240129	2.712678	0.348815	1	66.98546	4.295501	1.480458	0.331169
20	3.130338	0.155950	2.701591	0.344076	1	67.07759	4.287462	1.465718	0.340909
21	3.122033	0.175956	2.695203	0.345972	1	66.83323	4.268301	1.452396	0.343344
22	3.111480	0.307427	2.668613	0.345972	1	69.92950	4.428666	1.475602	0.342532
23	3.099581	0.179451	2.651050	0.341232	1	70.86115	4.485304	1.497365	0.335227
24	3.091632	0.179567	2.632138	0.331754	1	72.64216	4.594937	1.538901	0.335227
25	3.084850	0.245187	2.612357	0.325118	1	74.97540	4.724931	1.578779	0.327922
26	3.075781	0.268333	2.565805	0.326066	1	81.62096	5.099765	1.702086	0.322240
27	3.067376	0.364593	2.537649	0.327014	1	85.52735	5.297269	1.750871	0.312500
28	3.058743	0.263180	2.538908	0.327962	1	83.56876	5.198349	1.733863	0.320211
29	3.043042	0.158489	2.512139	0.328910	1	85.30854	5.309034	1.792678	0.317776
30	3.032750	0.221518	2.480542	0.332701	1	88.83244	5.522071	1.879425	0.311282

The optimizationexited on validationcriteria.

Predicted Probability Variables	
LevName	Variable
1	P_acci_severity1
3	P_acci_severity3
2	P_acci_severity2

Predicted Target Variable
Variable
I_acci_severity

Score Information for Training	
Number of Observations Read	2464
Number of Observations Used	2464
Misclassification Rate	0.3279

Score Information for Validation	
Number of Observations Read	1055
Number of Observations Used	1055
Misclassification Rate	0.3251

Predicting_Accidents_D

"Variable Selection" Results

by: di00222@surrey.ac.uk

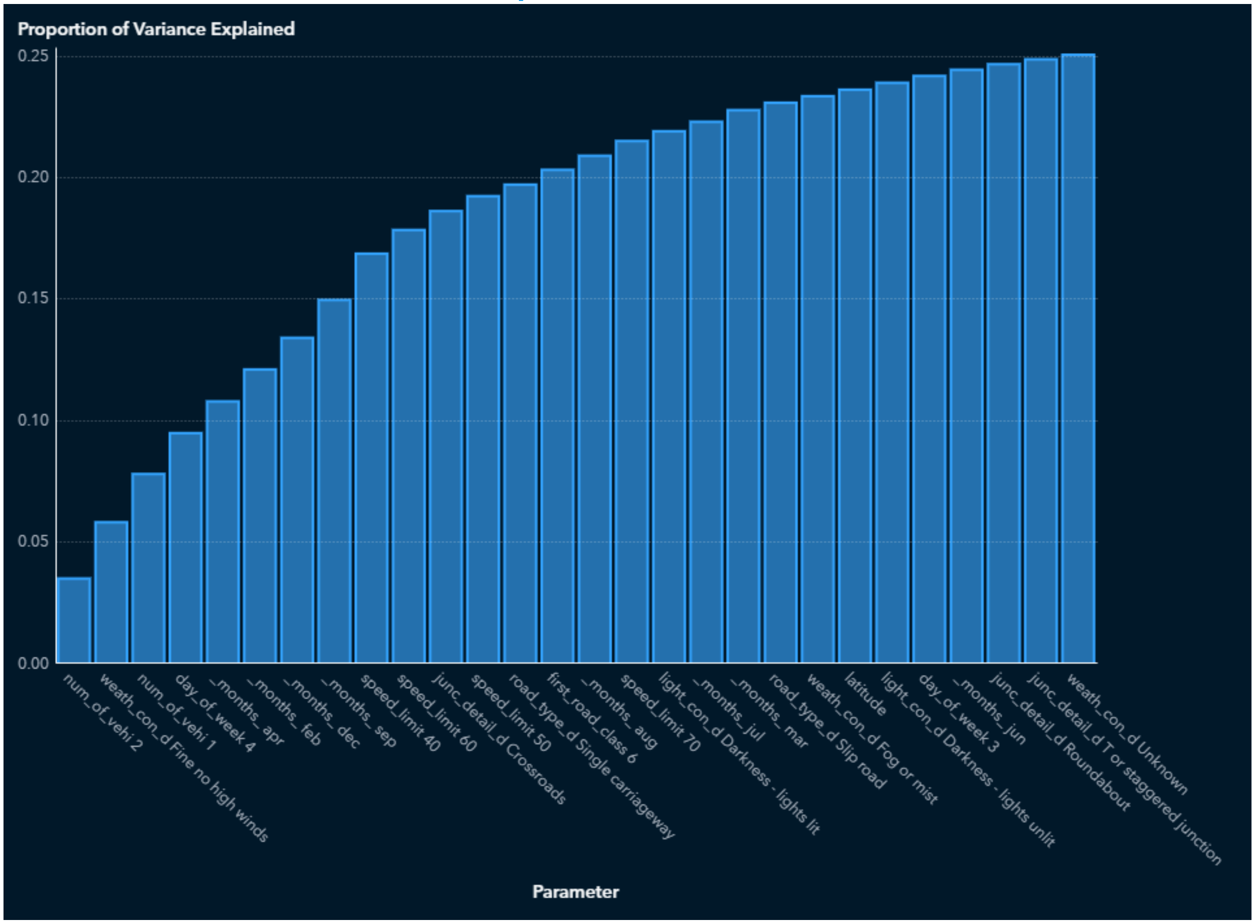
Contents

Variable Selection

Name	Variable Label	Variable Level	Role
ACCI_SEVERITY		NOMINAL	TARGET
DAY_OF_WEEK		NOMINAL	INPUT
FIRST_ROAD_CLASS		NOMINAL	INPUT
JUNC_DETAIL_D		NOMINAL	INPUT
LATITUDE		INTERVAL	INPUT
LIGHT_CON_D		NOMINAL	INPUT
NUM_OF_VEHI		NOMINAL	INPUT
ROAD_TYPE_D		NOMINAL	INPUT
SPEED_LIMIT		NOMINAL	INPUT
WEATH_CON_D		NOMINAL	INPUT
MONTHS		NOMINAL	INPUT
ACCI_REF		INTERVAL	ID
DMINDEX		NOMINAL	KEY
PARTIND	Partition Indicator	NOMINAL	PARTITION
LONGITUDE		INTERVAL	REJECTED

Reason
Variance Explained (Supervised)

Cumulative Variance Explained - Fast Selection



Variance Explained Values - Fast Selection

Iteration	Parameter	Variable Label	Proportion of Variance Explained
1	num_of_vehi 2	num_of_vehi	0.0350
2	weath_con_d Fine no high winds	weath_con_d	0.0582
3	num_of_vehi 1	num_of_vehi	0.0781
4	day_of_week 4	day_of_week	0.0949
5	_months_ apr	_months_	0.1079
6	_months_ feb	_months_	0.1211
7	_months_ dec	_months_	0.1341
8	_months_ sep	_months_	0.1497
9	speed_limit 40	speed_limit	0.1687
10	speed_limit 60	speed_limit	0.1785
11	junc_detail_d Crossroads	junc_detail_d	0.1863
12	speed_limit 50	speed_limit	0.1924
13	road_type_d Single carriageway	road_type_d	0.1972
14	first_road_class 6	first_road_class	0.2033
15	_months_ aug	_months_	0.2090
16	speed_limit 70	speed_limit	0.2151
17	light_con_d Darkness - lights lit	light_con_d	0.2191
18	_months_ jul	_months_	0.2230
19	_months_ mar	_months_	0.2278
20	road_type_d Slip road	road_type_d	0.2309
21	weath_con_d Fog or mist	weath_con_d	0.2336
22	latitude	latitude	0.2362

Iteration	Parameter	Variable Label	Proportion of Variance Explained
23	light_con_d Darkness - lights unlit	light_con_d	0.2391
24	day_of_week 3	day_of_week	0.2419
25	_months_ jun	_months_	0.2444
26	junc_detail_d Roundabout	junc_detail_d	0.2467
27	junc_detail_d T or staggered junction	junc_detail_d	0.2487
28	weath_con_d Unknown	weath_con_d	0.2506

Incremental Variance Explained
0
0.0232
0.0199
0.0168
0.0130
0.0132
0.0130
0.0156
0.0190
0.0097
0.0078
0.0062
0.0047
0.0061

Incremental Variance Explained
0.0058
0.0061
0.0040
0.0039
0.0048
0.0030
0.0027
0.0027
0.0028
0.0028
0.0026
0.0023
0.0020
0.0019

Properties

Property Name	Property Value
alpha	0.2000
bonferroni	false
chooseCriterion	SBC
chooseFcp	SBC
codeLocation	mlearning
confidence	0.2500
criterionMethod	ENTROPY
criterionMethod_f	IGR
cumulativeRedu	1
cumulative_Super	1
cvccFolds	10
dataMiningVersion	V2024.09
defaultVarsPerTree_f	true
defaultVarsPerTree_g	true
distribution	GAUSSIAN
earlyStop	true
earlyStopMethod	STAGNATION
ensembleCrit	ANY
esMetric	MCR
esMinimum	false
esThreshold	0
esThresholdIter	0
fast_Enabled	true
forest_Enabled	false
fullDatasetReconstitution	false

Property Name	Property Value
gradient_Enabled	false
hLeafSize	5
iCriterionMethod	VARIANCE
iCriterionMethod_f	VARIANCE
incrementRedu	0.0010
increment_Super	0.0010
intBinMethod_f	QUANTILE
intBinMethod_g	QUANTILE
intBinMethod_t	QUANTILE
intervalBins_f	50
intervalBins_g	50
intervalBins_t	50
lasso	0
leafSize	5
learningRate	0.1000
loh	0
maxBranch_f	2
maxBranch_g	2
maxBranch_t	2
maxCategories_f	128
maxCategories_g	128
maxCategories_t	128
maxDepth_f	20
maxDepth_g	4
maxDepth_t	10
maxEffects	0
maxL2	1
maxLevel	50
maxSteps	0

Property Name	Property Value
maxStepsRedu	200
maxStepsSuper	200
maxTime	600
maxTrees	100
maxeffects_Redu	200
maxeffects_Super	200
minEffects	0
minL2	0
minLeafSize	5
minUseInSearch_f	1
minUseInSearch_g	1
minUseinsearch_t	1
missingPercent	50
missingValue	USEINSEARCH
nPLeaves	1
ntrees	100
numAlpha	4
numL2	50
numLambda	10
optTech	MILP
parmValueAlpha	SEARCH
parmValueL2	SEARCH
parmValueLambda	SEARCH
partitionSeed	12,345
pearson_Redu	CORR
pearson_Redu_Super	CORR
power	1.5000
prescreen_Enabled	false

Property Name	Property Value
pruningMethod	COSTCOMPLEXITY
reg_Enabled	false
relativeImportance_f	0.2500
relativeImportance_g	0.2500
relativeImportance_t	0.2500
ridge	1
seRule	false
seed_f	12,345
seed_g	12,345
seed_t	12,345
selMethod	AUTOMATIC
selectCriterion	SBC
selectMethod	STEPWISE
selectProcess	SEQUENTIAL
slEntry	0.0500
slStay	0.0500
stagnation	5
stopCriterion	SBC
stopSuper	BIC
subsampleRate	0.5000
suppressIntercept	false
templateRevision	5
tolerance	0
topx	10
topx_Enabled	false
trainFraction	0.6000

Property Name	Property Value
tree_Enabled	false
unsuper_enabled	false
useVarOnce	false
validation	true
validationProb	0.3000
varsToTry_f	100
varsToTry_g	100

Output

Fast Variable Selection

The VARREDUCE Procedure

Number of Observations Read	2464
Number of Observations Used	2464

Selection Summary

Iteration	Parameter	Proportion of Variance Explained	SSE	MSE	AIC	AICC	BIC
1	num_of_vehi 2	0.035041	1.929919	0.00078356	0.664783	3.664798	0.660647
2	weath_con_d Fine no high winds	0.058196	1.883608	0.00076507	0.642929	3.642953	0.639528
3	num_of_vehi 1	0.078058	1.843885	0.00074924	0.624050	3.624085	0.621383
4	day_of_week 4	0.094888	1.810224	0.00073586	0.608061	3.608108	0.606128
5	_months_ apr	0.107922	1.784155	0.00072556	0.595991	3.596053	0.594792
6	_months_ feb	0.121099	1.757802	0.00071513	0.583545	3.583624	0.583081
7	_months_ dec	0.134099	1.731802	0.00070484	0.571078	3.571176	0.571349
8	_months_ sep	0.149696	1.700609	0.00069243	0.555337	3.555456	0.556342
9	speed_limit 40	0.168743	1.662514	0.00067719	0.535116	3.535258	0.536856
10	speed_limit 60	0.178487	1.643027	0.00066953	0.525761	3.525928	0.528235
11	junc_detail_d Crossroads	0.186255	1.627490	0.00066347	0.518695	3.518888	0.521903
12	speed_limit 50	0.192414	1.615171	0.00065872	0.513532	3.513755	0.517475
13	road_type_d Single carriageway	0.197158	1.605683	0.00065511	0.510075	3.510329	0.514752
14	first_road_class 6	0.203259	1.593481	0.00065040	0.504882	3.505169	0.510294
15	_months_ aug	0.209040	1.581919	0.00064594	0.500035	3.500357	0.506181
16	speed_limit 70	0.215103	1.569794	0.00064126	0.494776	3.495134	0.501656
17	light_con_d Darkness - lights lit	0.219125	1.561751	0.00063823	0.492074	3.492471	0.499688
18	_months_ jul	0.223024	1.553952	0.00063530	0.489502	3.489941	0.497851
19	_months_ mar	0.227821	1.544358	0.00063164	0.485745	3.486227	0.494828
20	road_type_d Slip road	0.230858	1.538284	0.00062941	0.484239	3.484766	0.494057
21	weath_con_d Fog or mist	0.233558	1.532885	0.00062746	0.483158	3.483732	0.493710
22	latitude	0.236230	1.527540	0.00062553	0.482100	3.482723	0.493387
23	light_con_d Darkness - lights unlit	0.239062	1.521875	0.00062346	0.480820	3.481494	0.492841
24	day_of_week 3	0.241873	1.516253	0.00062142	0.479554	3.480282	0.492309
25	_months_ jun	0.244431	1.511137	0.00061957	0.478609	3.479392	0.492099
26	junc_detail_d Roundabout	0.246710	1.506580	0.00061796	0.478024	3.478864	0.492248
27	junc_detail_d T or staggered junction	0.248735	1.502530	0.00061655	0.477767	3.478667	0.492726
28	weath_con_d Unknown	0.250608	1.498784	0.00061526	0.477706	3.478667	0.493399

Selected Effects

Number	Selected Variable	Variable Type
1	num_of_vehi	CLASS
2	weath_con_d	CLASS
3	day_of_week	CLASS
4	_months_	CLASS
5	speed_limit	CLASS
6	junc_detail_d	CLASS
7	road_type_d	CLASS
8	first_road_class	CLASS
9	light_con_d	CLASS
10	latitude	INTERVAL

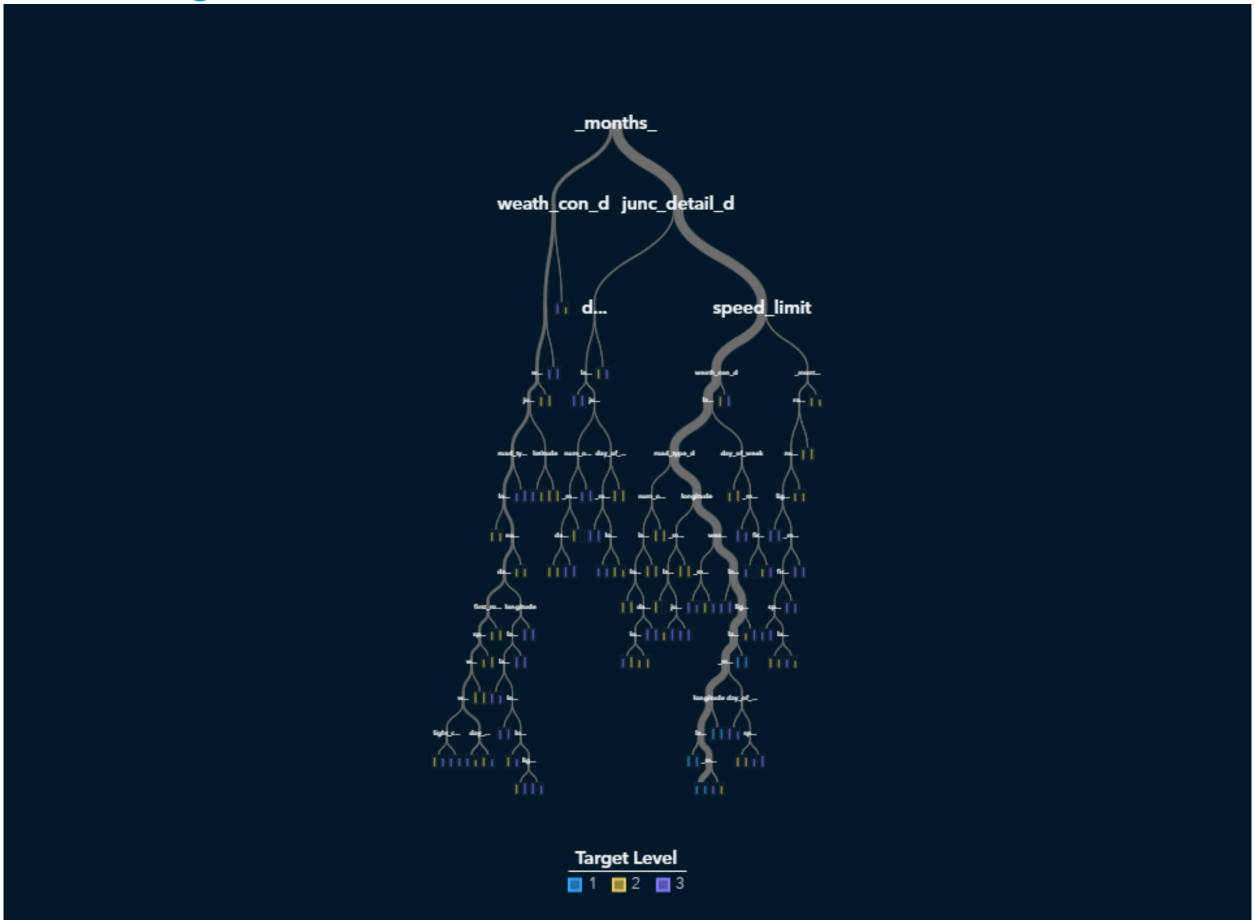
Predicting_Accidents_D

"Decision Tree" Results

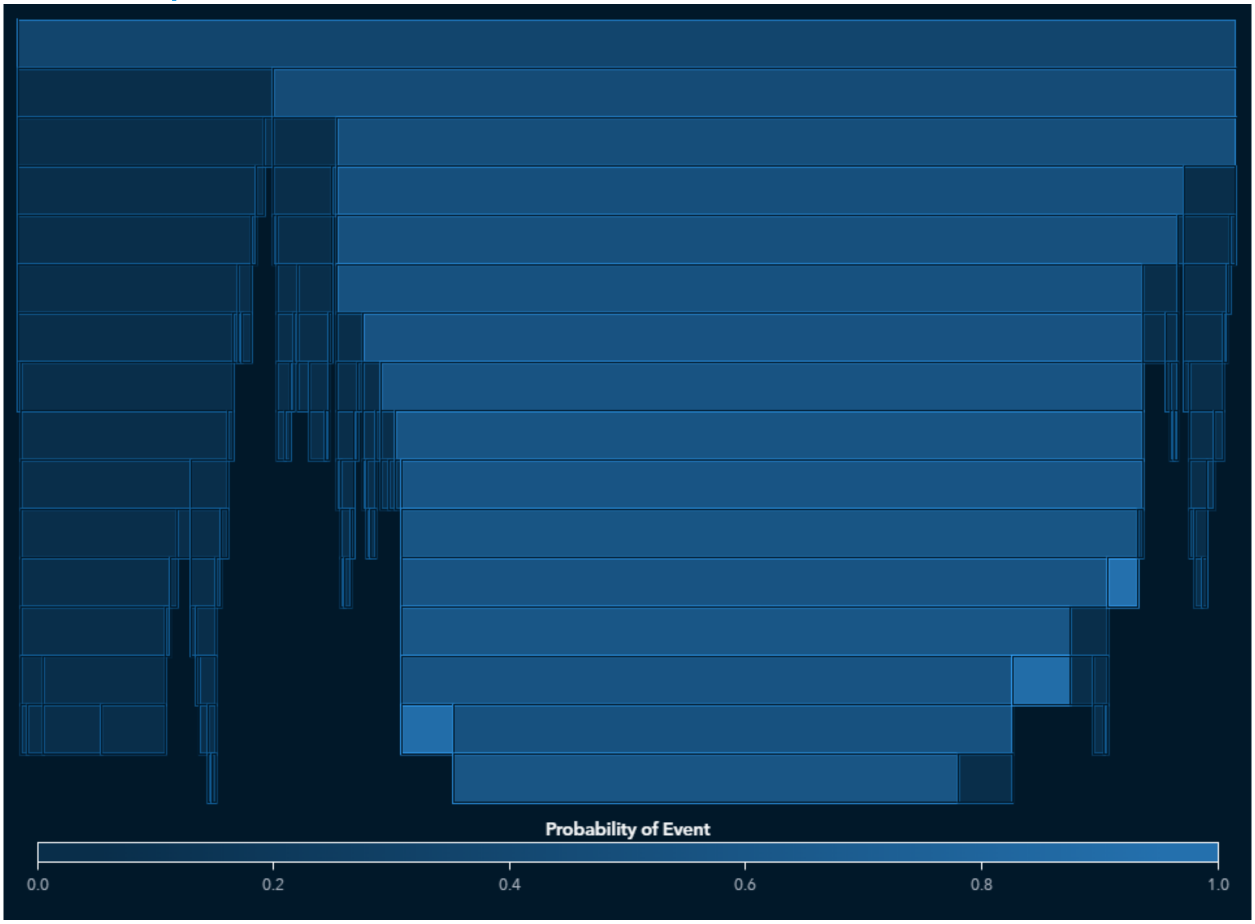
by: di00222@surrey.ac.uk

Contents

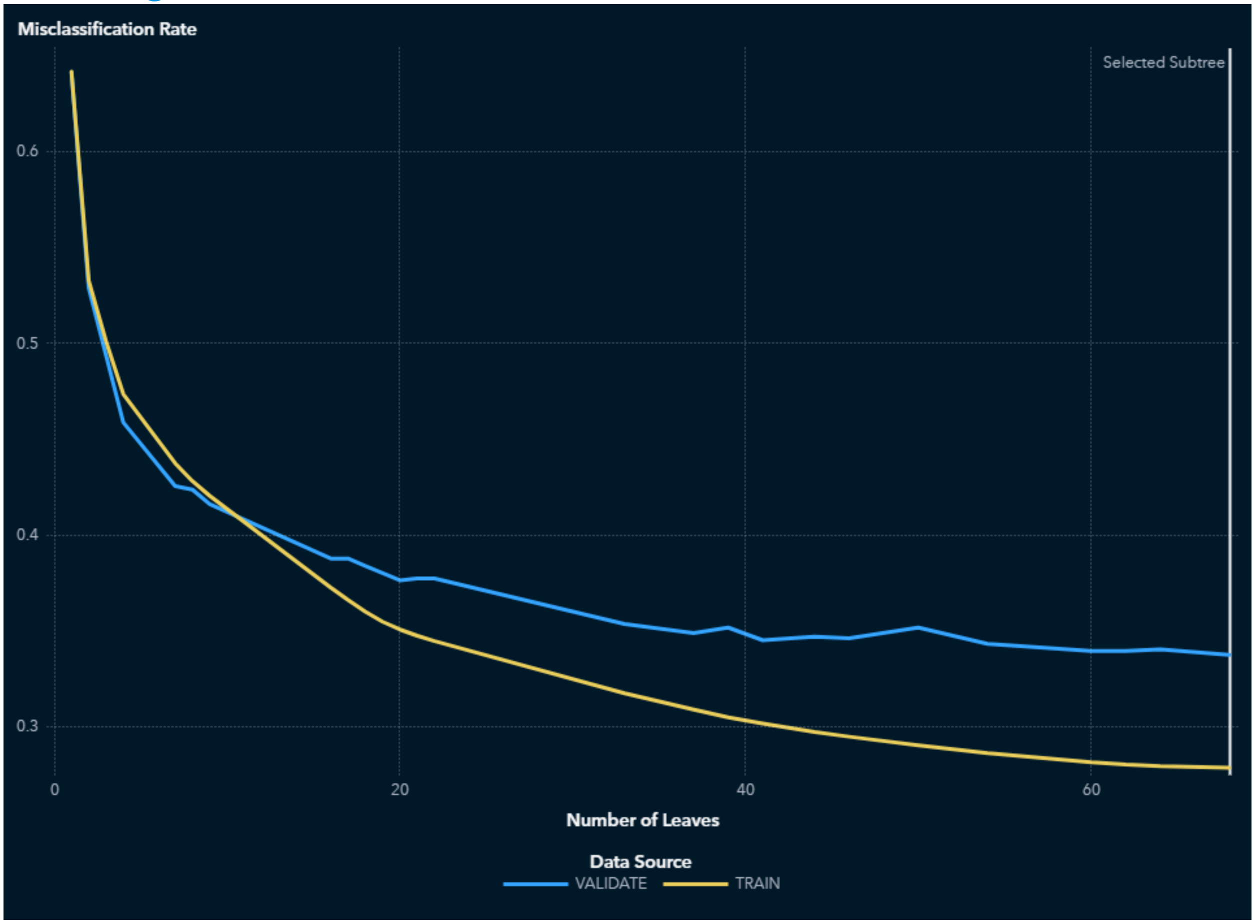
Tree Diagram



Treemap



Pruning Error Plot



This plot shows how the misclassification rate changes for subtrees, which are created by pruning the full decision tree to various numbers of leaves. The training error decreases as the number of leaves increases, so the VALIDATE partition can be used to prune the tree to prevent overfitting. For this decision tree model, the selected subtree based on the pruning options has 68 leaves with a misclassification rate of 0.337 for the VALIDATE partition.

Variable Importance

Variable Name	Training Importance	Training Relative Importance	Validation Relative Importance
months	243.5914	1	1
speed_limit	49.9083	0.2049	0.3764
latitude	111.2987	0.4569	0.3100
junc_detail_d	57.1999	0.2348	0.2453
road_type_d	30.3278	0.1245	0.1776
day_of_week	55.9951	0.2299	0.1437
longitude	62.0510	0.2547	0.1243
weath_con_d	43.4392	0.1783	0.0799
num_of_vehi	12.0205	0.0493	0.0477
first_road_class	7.1313	0.0293	0.0369
light_con_d	19.6281	0.0806	0.0355

Count	Validation Importance
10	86.2291
4	32.4535
11	26.7304
4	21.1547
3	15.3166
8	12.3884
9	10.7155
6	6.8865
4	4.1103
3	3.1844
5	3.0603

Score Inputs

Name	Role	Variable Level	Type
day_of_week	INPUT	NOMINAL	N
first_road_class	INPUT	NOMINAL	N
junc_detail_d	INPUT	NOMINAL	C
latitude	INPUT	INTERVAL	N
light_con_d	INPUT	NOMINAL	C
longitude	INPUT	INTERVAL	N
num_of_vehi	INPUT	NOMINAL	N
road_type_d	INPUT	NOMINAL	C
speed_limit	INPUT	NOMINAL	N
weath_con_d	INPUT	NOMINAL	C
months	INPUT	NOMINAL	C

Variable Type	Variable Label	Variable Format	Variable Length
double			8
double			8
varchar			35
double			8
varchar			23
double			8
double			8
varchar			18
double			8
varchar			21
varchar			3

Score Outputs

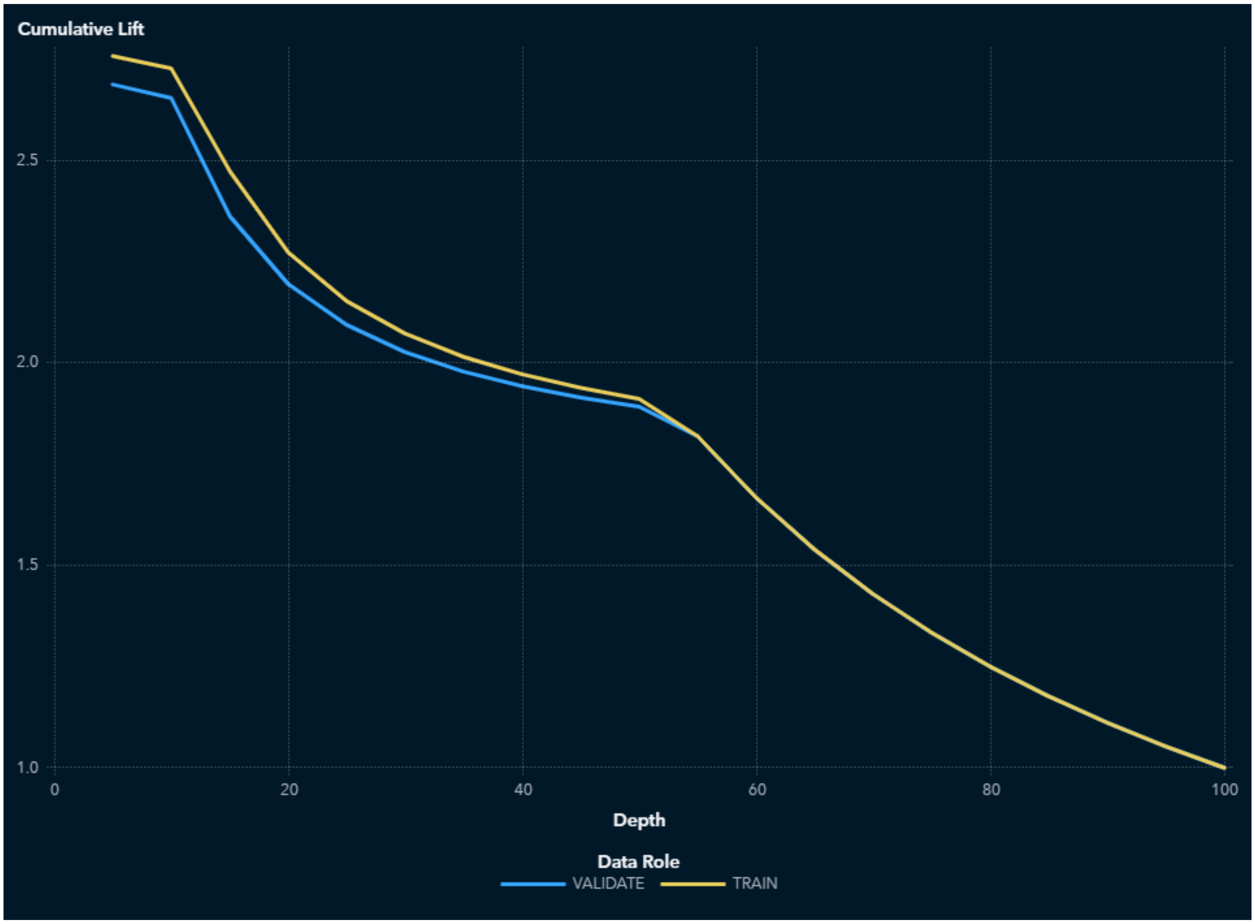
Name	Role	Type	Variable Type
EM_CLASSIFICATION	CLASSIFICATION	C	char
EM_EVENTPROBABILITY	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
I_acci_severity	CLASSIFICATION	C	char
P_acci_severity1	PREDICT	N	double
P_acci_severity2	PREDICT	N	double
P_acci_severity3	PREDICT	N	double
WARN	ASSESS	C	char
_leaf_id_	SEGMENT	N	double

Variable Label	Variable Format	Variable Length	Creator
Predicted for acci_severity		12	tree
Probability for acci_severity=1		8	tree
Probability of Classification		8	tree
Into: acci_severity		32	tree
Predicted: acci_severity=1		8	tree
Predicted: acci_severity=2		8	tree
Predicted: acci_severity=3		8	tree
Warnings		4	tree
_leaf_id_		8	tree

Function	Creator GUID

Function	Creator GUID
CLASSIFICATION	5badf9c0-c1af-4288-ba1e-745de840de8a
PREDICT	5badf9c0-c1af-4288-ba1e-745de840de8a
PREDICT	5badf9c0-c1af-4288-ba1e-745de840de8a
CLASSIFICATION	5badf9c0-c1af-4288-ba1e-745de840de8a
PREDICT	5badf9c0-c1af-4288-ba1e-745de840de8a
PREDICT	5badf9c0-c1af-4288-ba1e-745de840de8a
PREDICT	5badf9c0-c1af-4288-ba1e-745de840de8a
ASSESS	5badf9c0-c1af-4288-ba1e-745de840de8a
TRANSFORM	5badf9c0-c1af-4288-ba1e-745de840de8a

Cumulative Lift



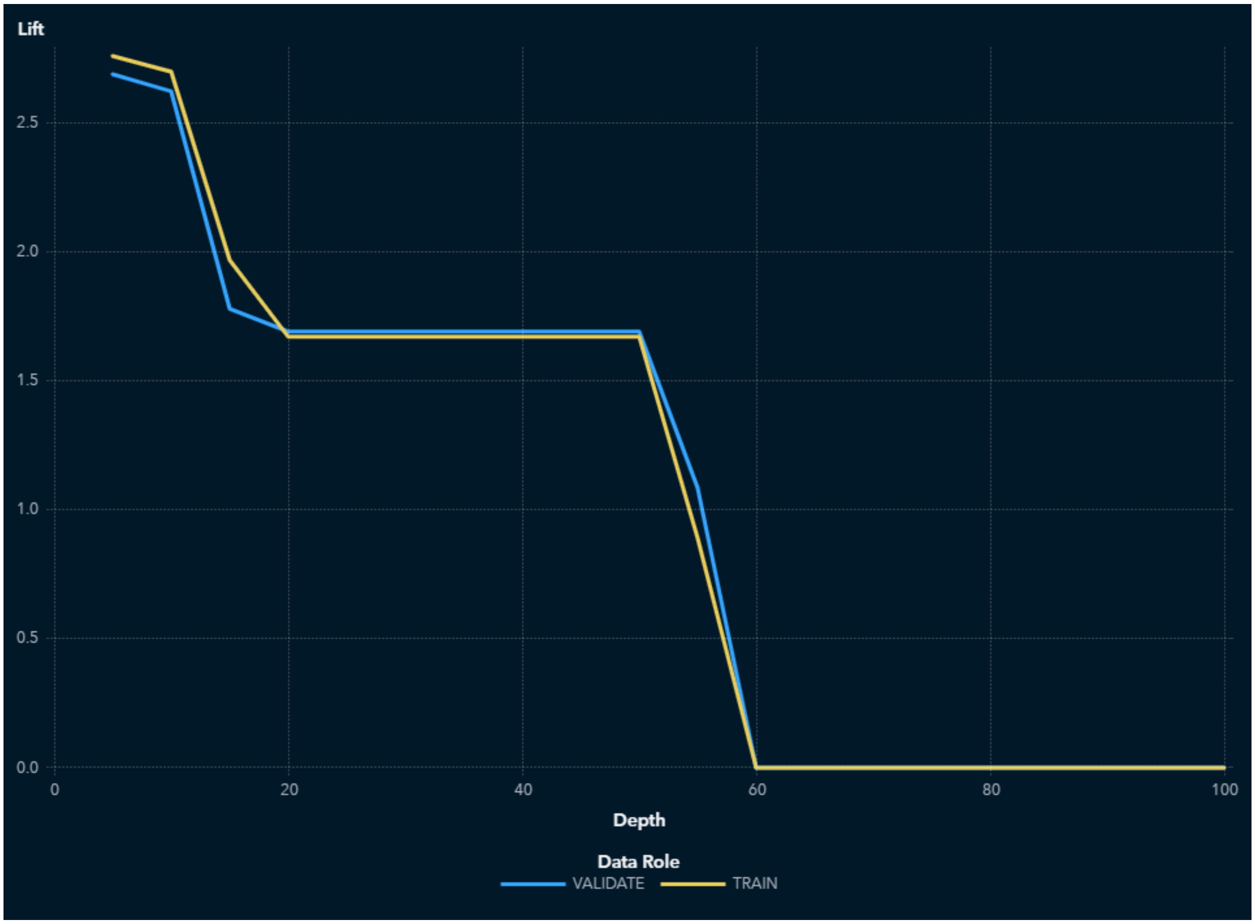
The VALIDATE partition has a Cumulative Lift of 2.65 in the 10% quantile (depth of 10) meaning there are 2.65 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Cumulative Lift of 2.73 in the 10% quantile (depth of 10) meaning there are 2.73 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



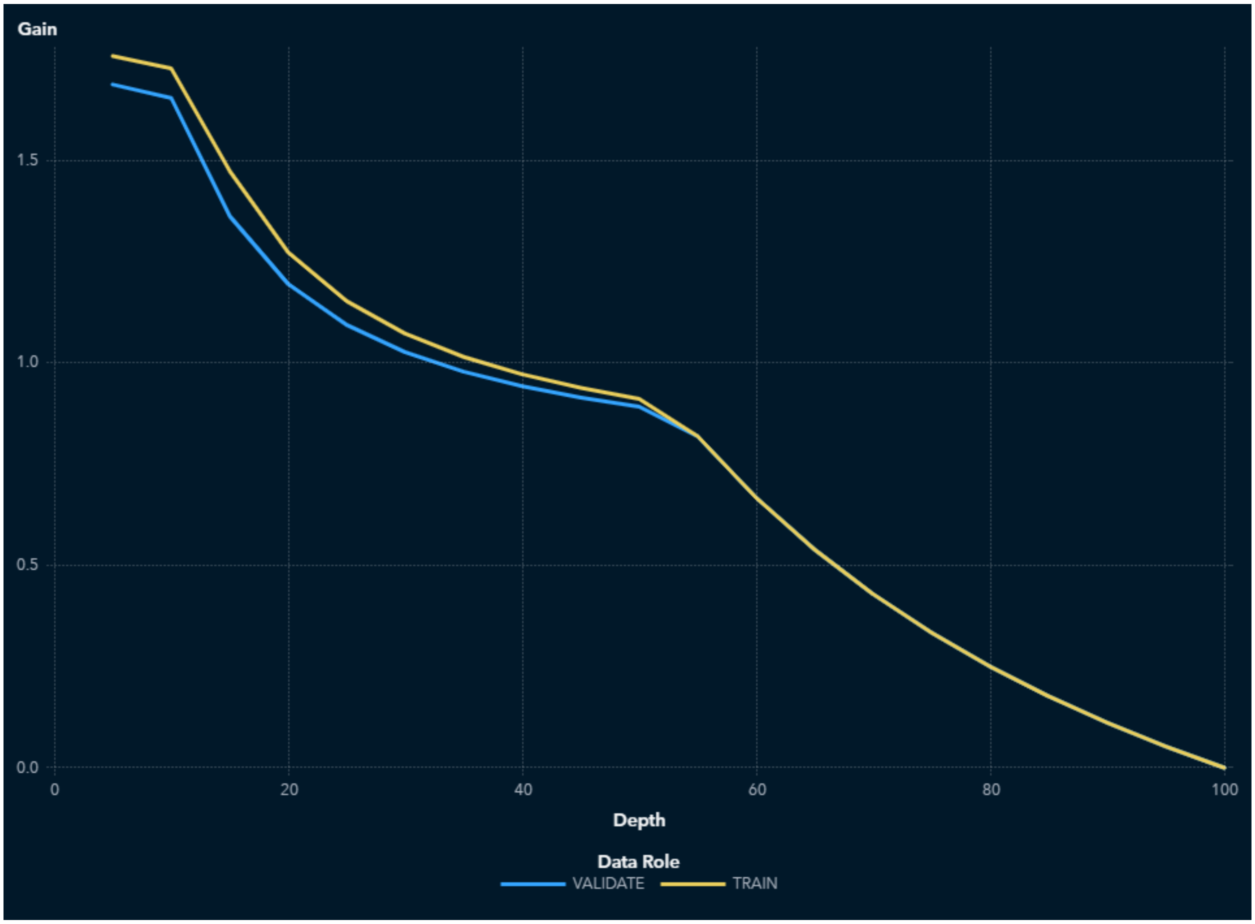
The VALIDATE partition has a Lift of 2.69 in the 5% quantile (depth of 5) meaning there are 2.69 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Lift of 2.76 in the 5% quantile (depth of 5) meaning there are 2.76 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



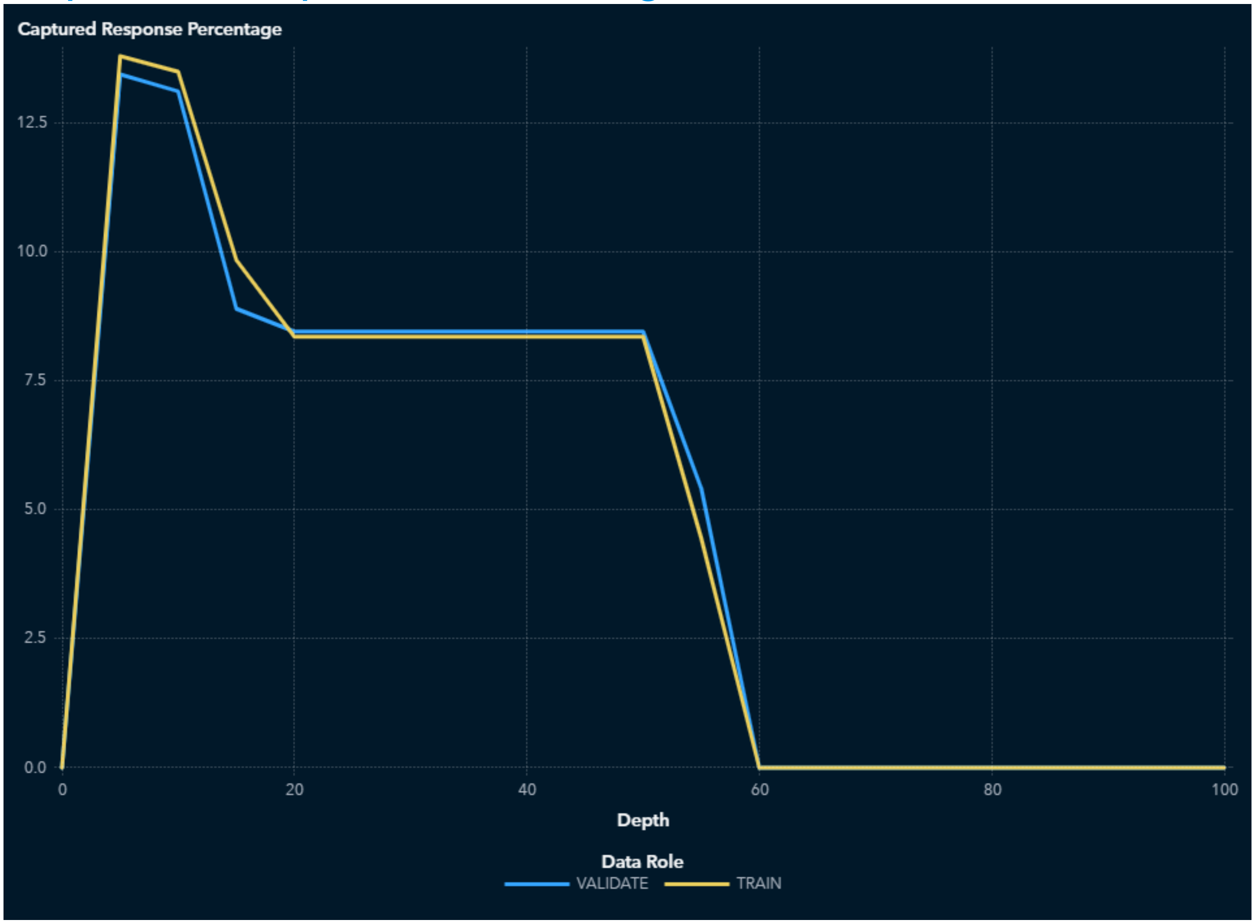
The VALIDATE partition has a Gain of 1.7 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 1.8.

The TRAIN partition has a Gain of 1.7 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 1.81.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

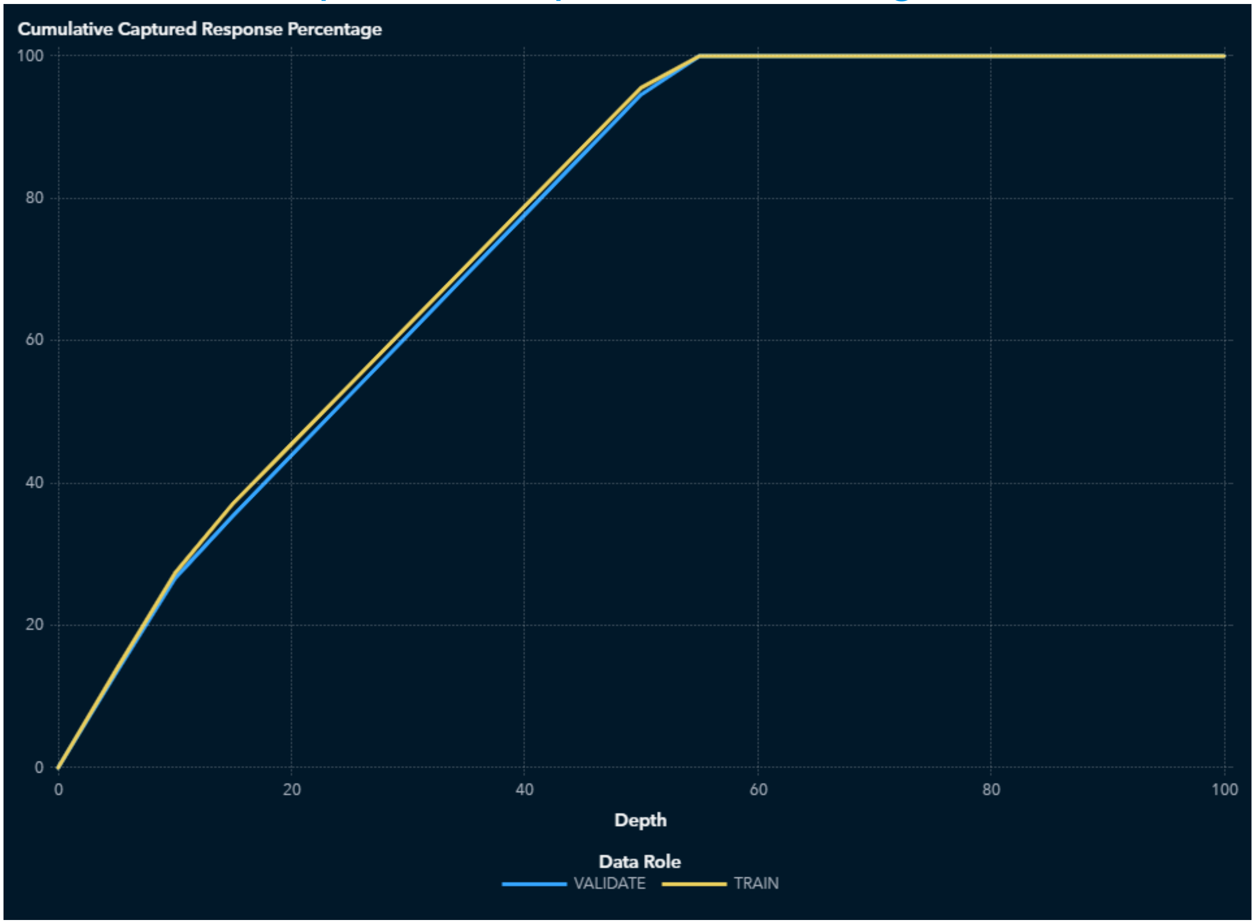


At the 5% quantile (depth of 5), the VALIDATE partition has a Captured response percentage of 13.4 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 14.02.

At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 13.8 (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 14.06.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



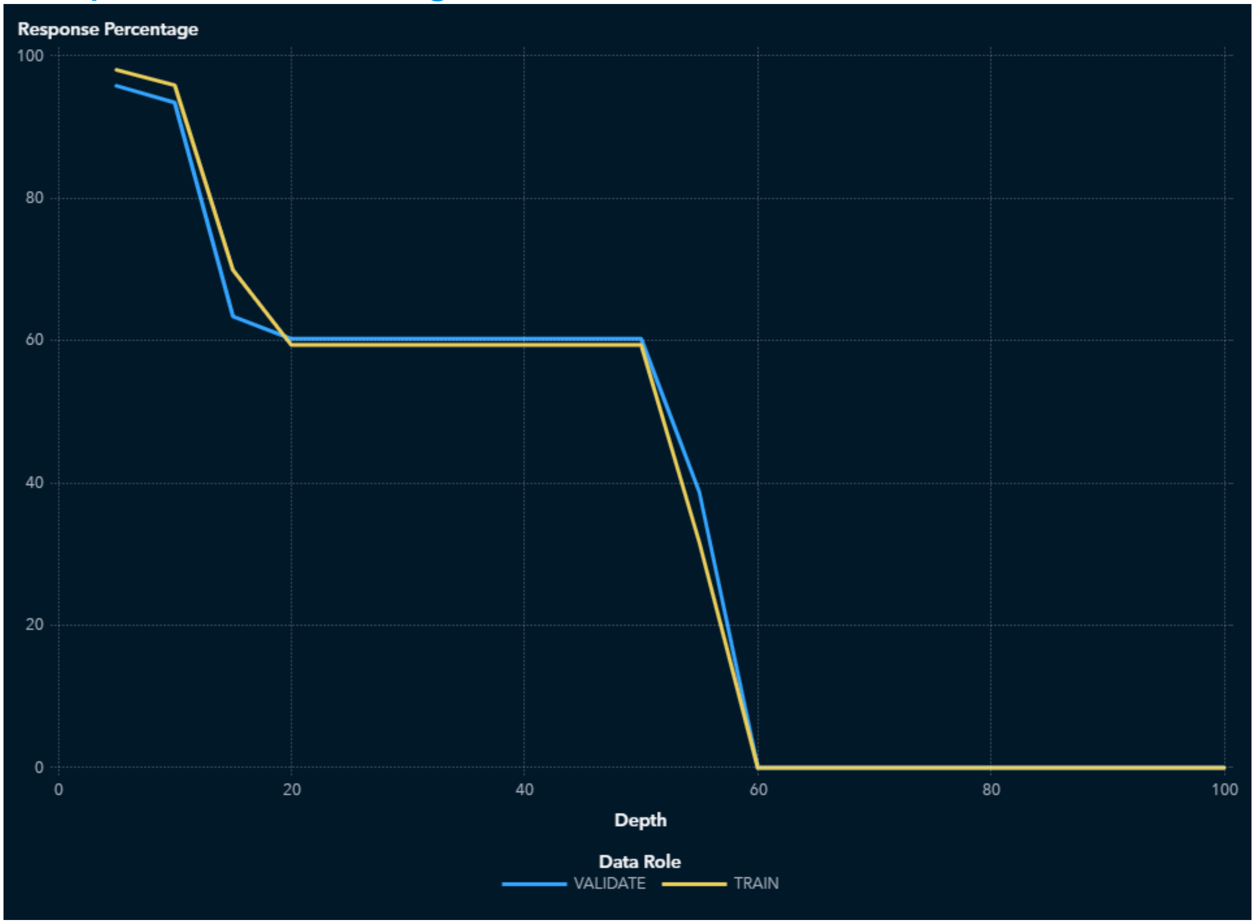
In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative captured response percentage of 26.5 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 28.04.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 27.3 (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 28.12.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

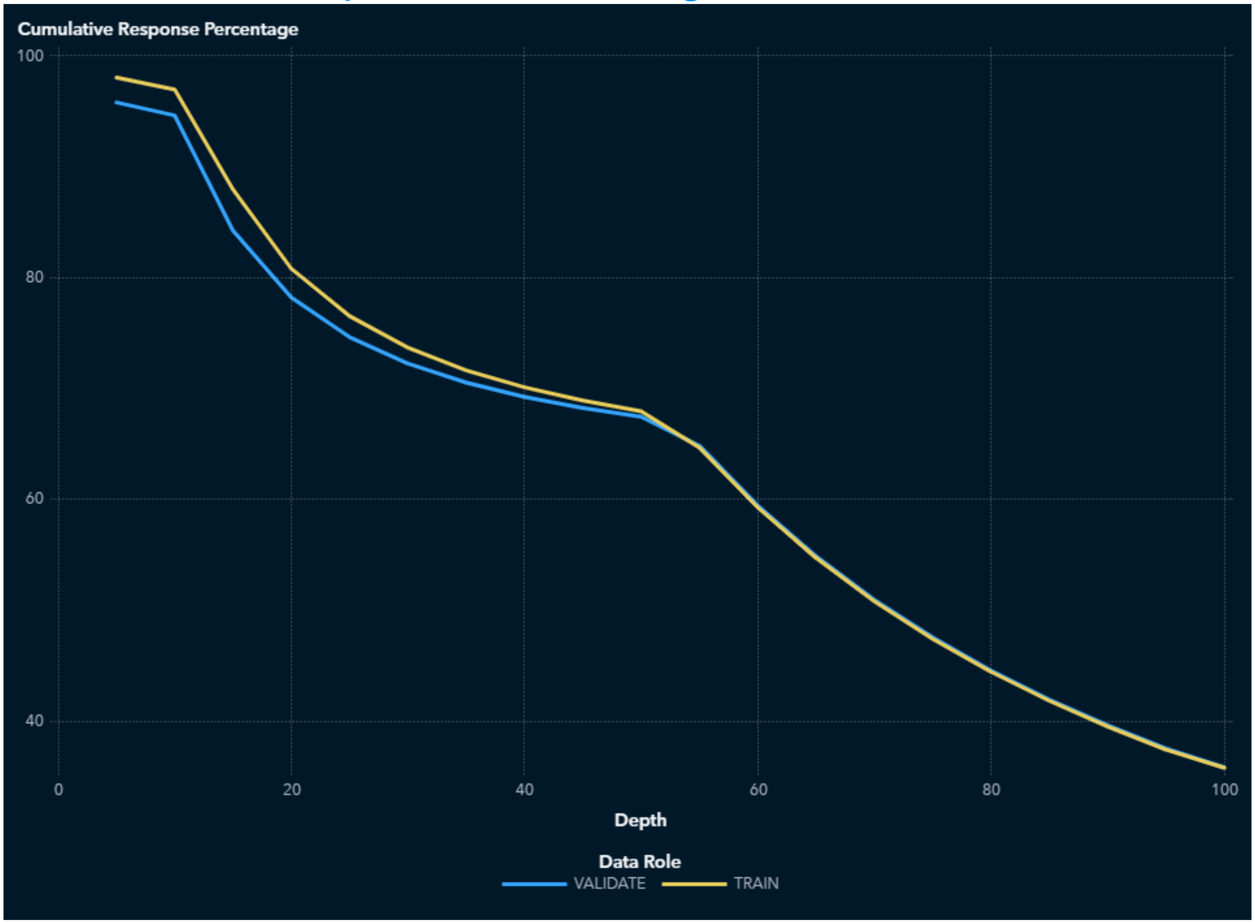


At the 5% quantile (depth of 5), the VALIDATE partition has a Response percentage of 95.8. The best possible value of Response percentage for this partition at depth 5 is 100.

At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 98.1. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

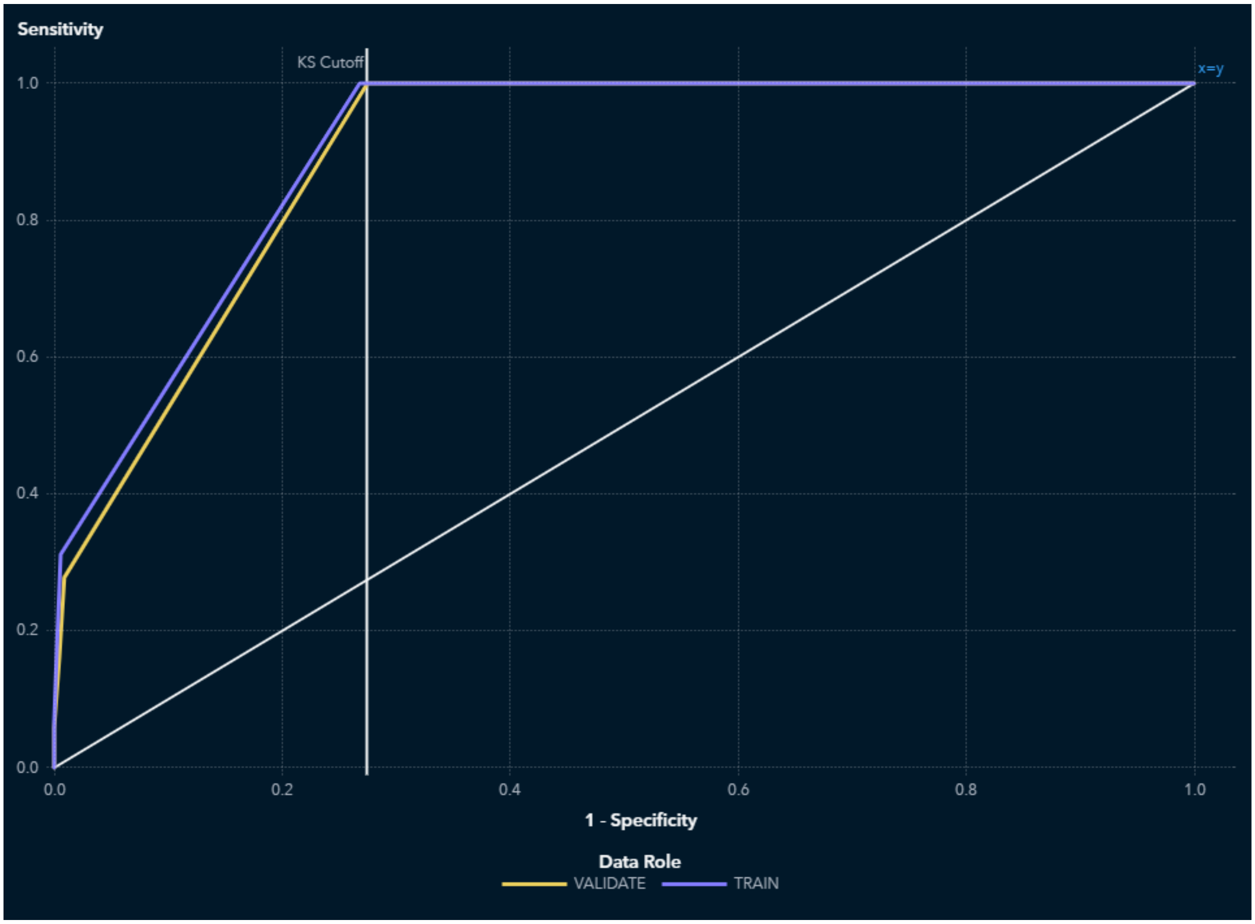


In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative response percentage of 94.6. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 97. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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 VALIDATE partition. The KS Cutoff line is drawn at the cutoff value 0.01, where the 1-specificity value is 0.275 and the sensitivity value is 1.

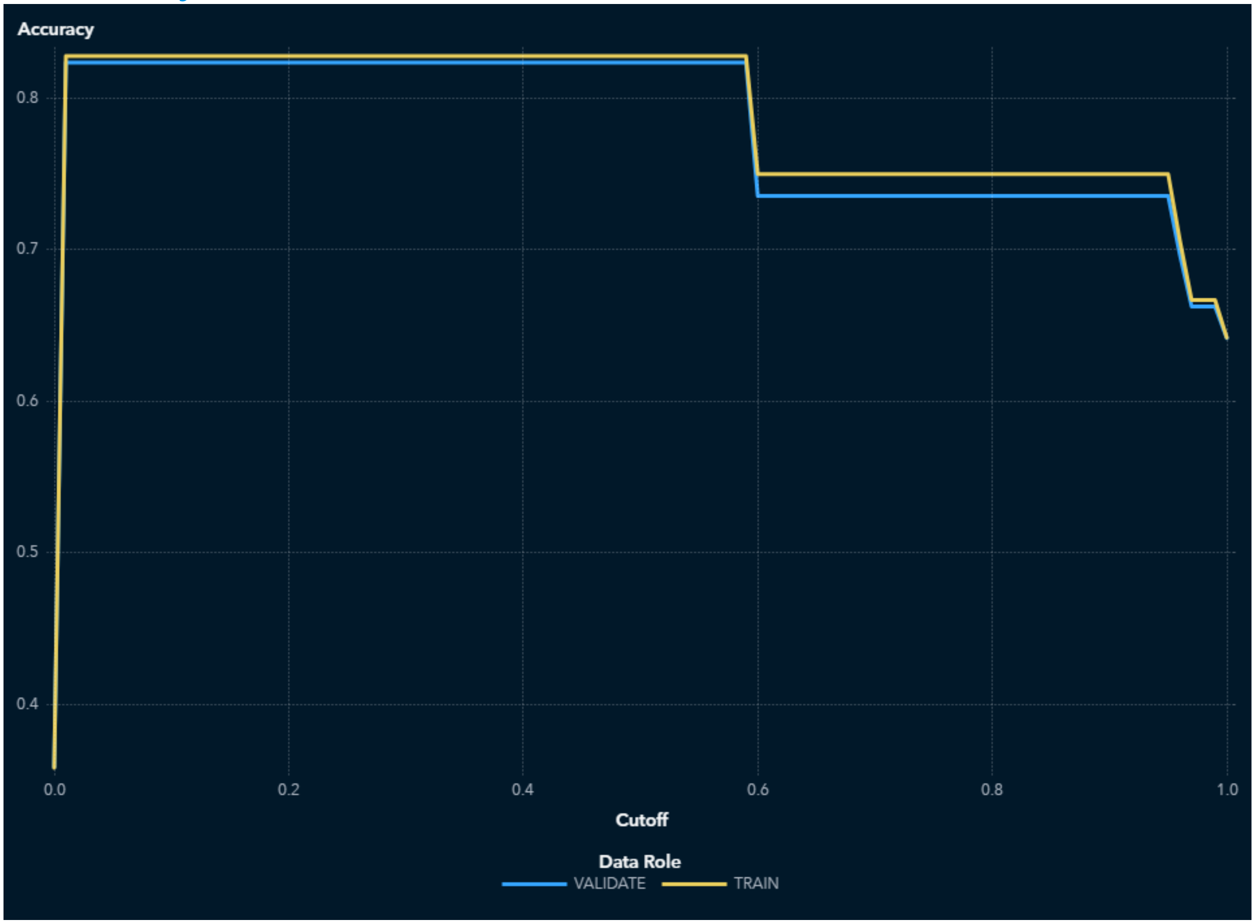
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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

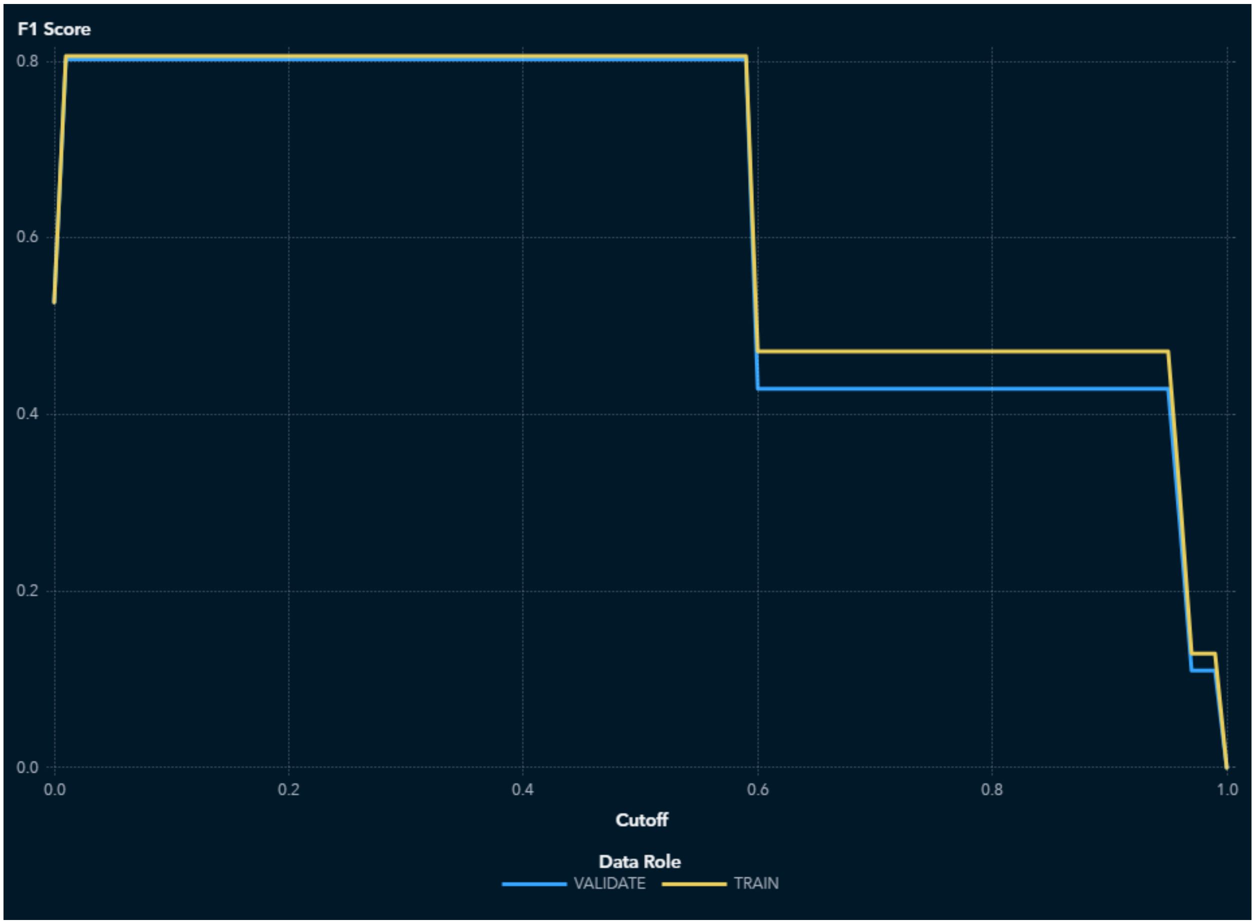


For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.828.

For this model, the accuracy in the VALIDATE partition at the cutoff of 0.5 is 0.824.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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 0.806.

For this model, the F1 score in the VALIDATE partition at the cutoff of 0.5 is 0.803.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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
acci_severity	TRAIN	1	1
acci_severity	VALIDATE	0	0

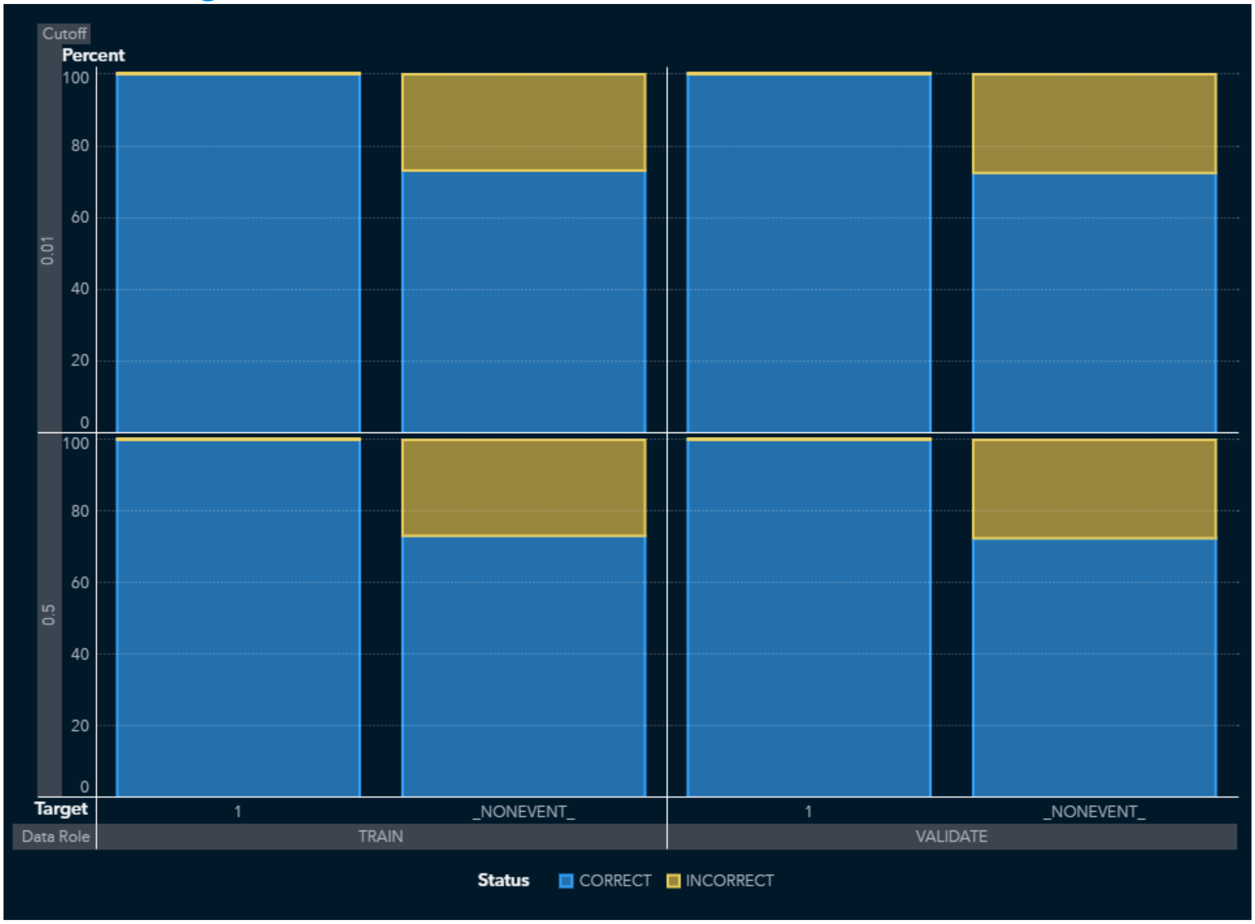
Number of Observations	Average Squared Error	Divisor for ASE	Root Average Squared Error
2,464	0.1282	2,464	0.3580
1,055	0.1517	1,055	0.3894

Misclassification Rate	Multi-Class Log Loss	KS (Youden)	Area Under ROC
0.2784	0.6285	0.7320	0.9051
0.3355	0.8236	0.7253	0.8966

Gini Coefficient	Gamma	Tau	KS Cutoff
0.8103	0.9896	0.3726	0.0100
0.7932	0.9829	0.3651	0.0100

KS at Default Cutoff	Misclassification Rate at KS Cutoff (Event)	Misclassification Rate (Event)
0.7320	0.1721	0.1721
0.7253	0.1763	0.1763

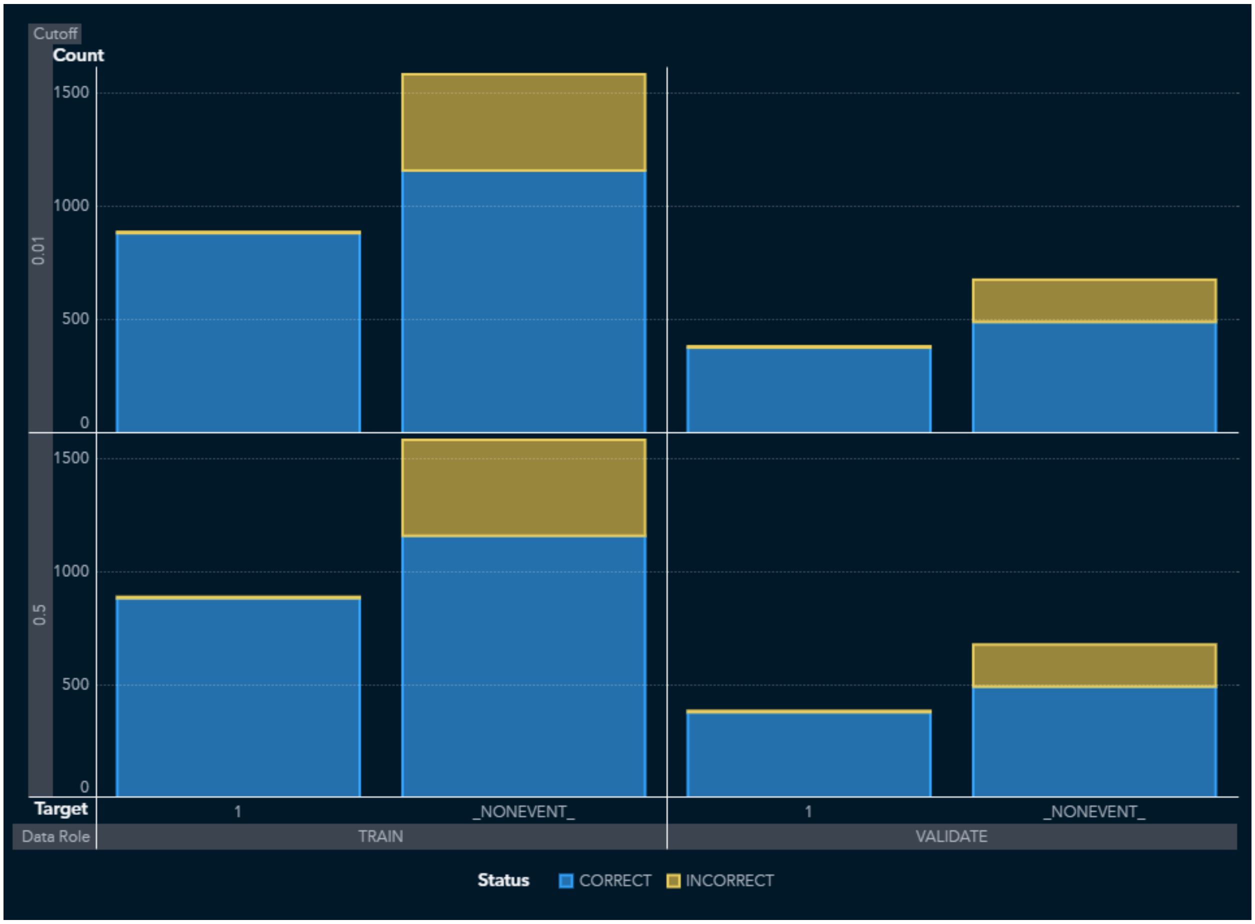
Percentage Plot



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: 0.01 (TRAIN), 0.01 (VALIDATE).

For this data, for the bar corresponding to the event level of acci_severity, "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 these KS cutoff values for existing partitions: 0.01 (TRAIN), 0.01 (VALIDATE).

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

Table

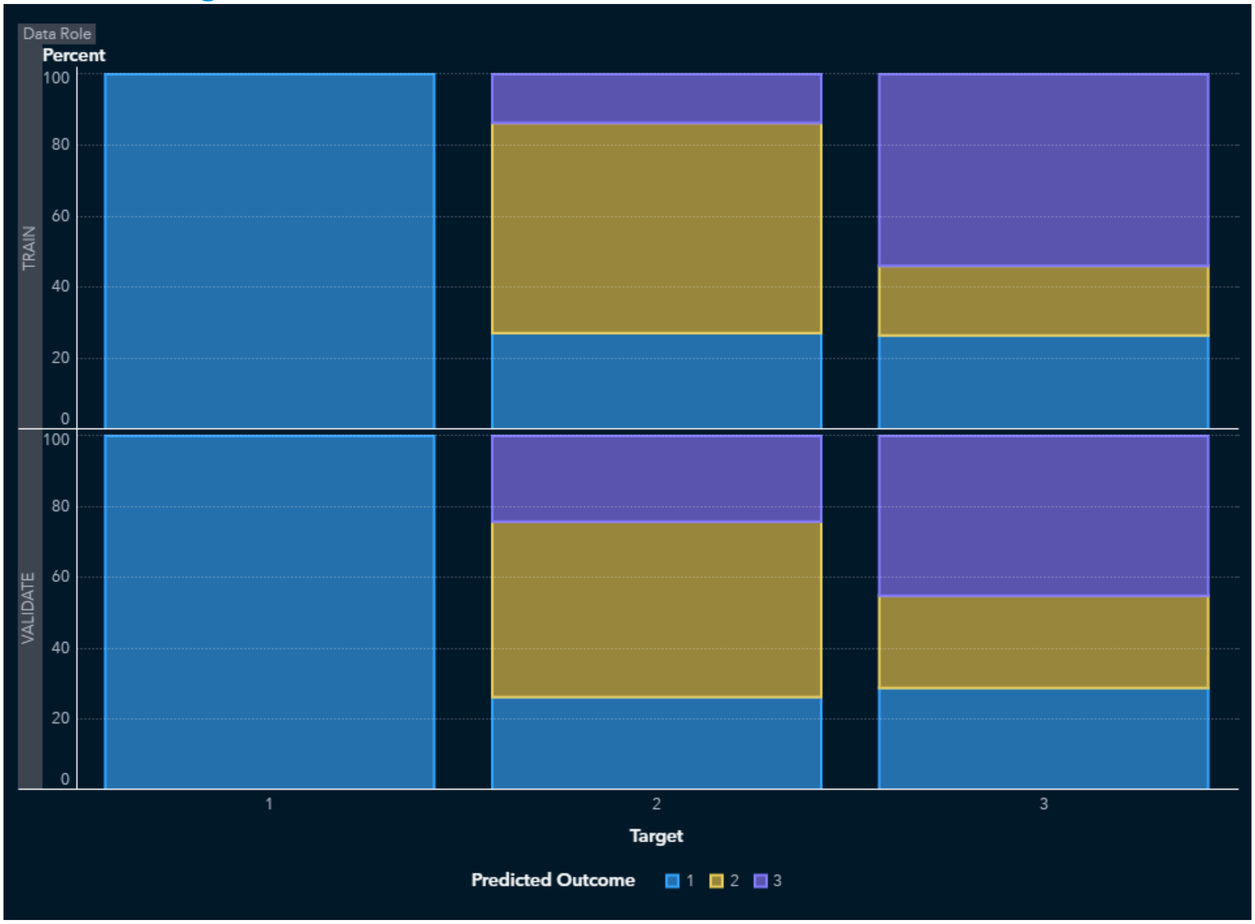
Cutoff	Cutoff Source	Target Name	Response
0.0100	KS	acci_severity	CORRECT
0.0100	KS	acci_severity	INCORRECT
0.0100	KS	acci_severity	CORRECT
0.0100	KS	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT

Event	Value	Training Frequency	Validation Frequency
1	True Positive	882	378
1	False Negative	0	0
NONEVENT	True Negative	1,158	491
NONEVENT	False Positive	424	186
1	True Positive	882	378
1	False Negative	0	0
NONEVENT	True Negative	1,158	491
NONEVENT	False Positive	424	186

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	100	100	
	0	0	
	73.1985	72.5258	
	26.8015	27.4742	
	100	100	
	0	0	
	73.1985	72.5258	

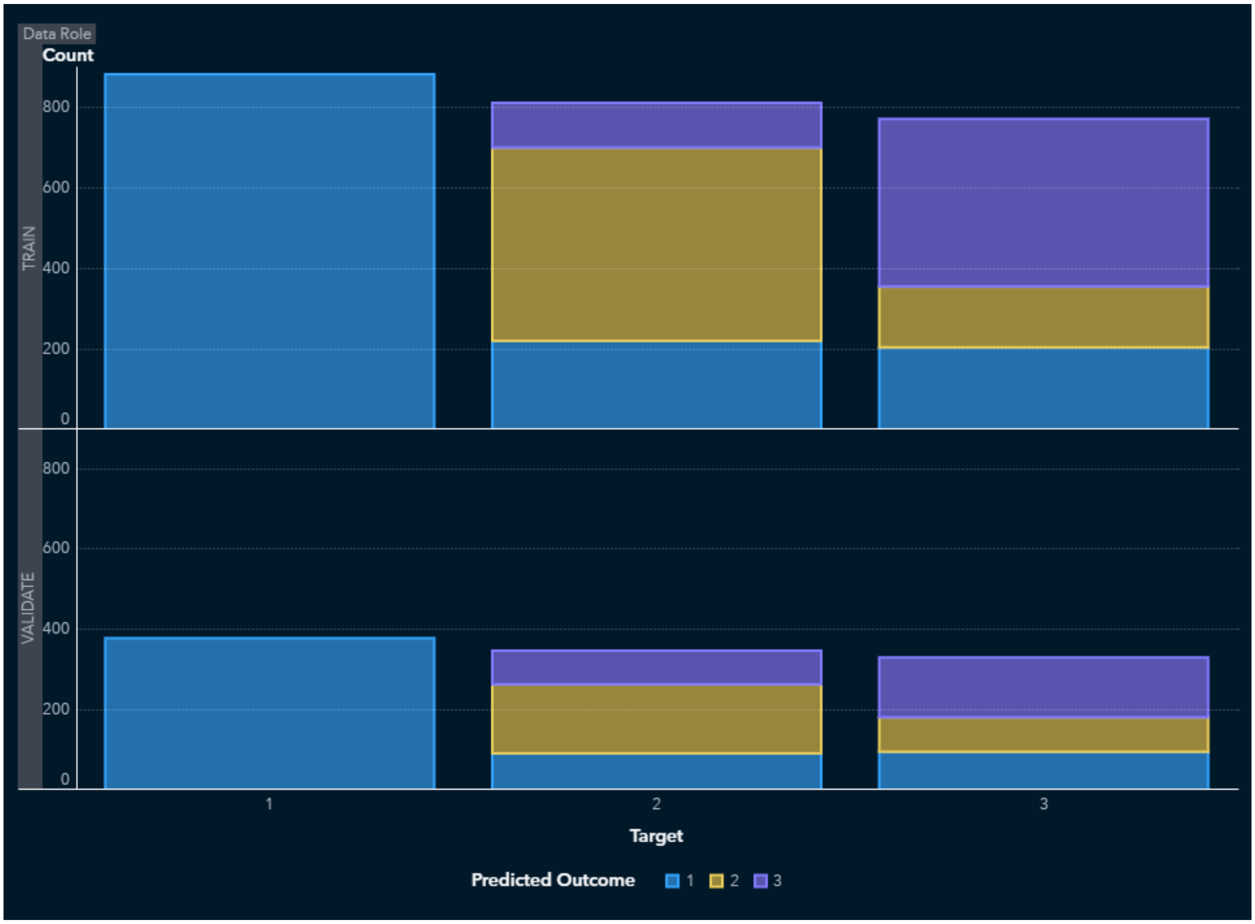
Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	26.8015	27.4742	

Percentage Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Count Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Table

Target Name	Data Role	Target	Unformatted Target
acci_severity	VALIDATE	1	1
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	3	3
acci_severity	VALIDATE	3	3
acci_severity	VALIDATE	3	3
acci_severity	TRAIN	1	1
acci_severity	TRAIN	2	2
acci_severity	TRAIN	2	2
acci_severity	TRAIN	2	2
acci_severity	TRAIN	3	3
acci_severity	TRAIN	3	3
acci_severity	TRAIN	3	3

Predicted Outcome	Count	Percent	Status
1	378	100	CORRECT
1	91	26.2248	INCORRECT
2	172	49.5677	CORRECT
3	84	24.2075	INCORRECT
1	95	28.7879	INCORRECT
2	86	26.0606	INCORRECT
3	149	45.1515	CORRECT
1	882	100	CORRECT
1	220	27.1270	INCORRECT
2	480	59.1862	CORRECT
3	111	13.6868	INCORRECT

Predicted Outcome	Count	Percent	Status
1	204	26.4591	INCORRECT
2	151	19.5850	INCORRECT
3	416	53.9559	CORRECT

Properties

Property Name	Property Value
alpha	0.2000
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
atgrowcrit	true
atgrowcritValsi	VARIANCE FTEST CHAID
atgrowcritValsn	ENTROPY CHAID IGR GINI CHISQUARE

Property Name	Property Value
atleafSize	false
atleafSizeInit	5
atleafSizeLB	1
atleafSizeUB	100
atmaxdepth	true
atmaxdepthInit	10
atmaxdepthLB	1
atmaxdepthUB	19
atnumbin	true
atnumbinInit	50
atnumbinLB	20
atnumbinUB	200
autotune_enabled	false
binaryProbCutoff	0.5000
bonferroni	false
ccAlpha	0
codeLocation	mlearning
confidence	0.2500
criterionMethod	IGR
cvccFolds	10
dataMiningVersion	V2024.09
embeddedBarChart	true
exactPctlLift	true
explainFidelity	false
explainInfo	false
fullDatasetReconstitution	false
hLeafSize	7
iCriterionMethod	VARIANCE

Property Name	Property Value
icePlots	false
inodeColor	AVERAGE
intBinMethod	QUANTILE
intervalBins	50
maxBranch	2
maxCategories	128
maxDepth	15
maxNumShapVars	20
minUseinsearch	1
missingValue	USEINSEARCH
nBins	50
nPLeaves	1
nodeColor	PROBEVENT
pdNumImportantInputs	5
pdObsSamples	1,000
pdPlots	false
performKernelShap	false
performLime	false
performVI	false
pruningMethod	COSTCOMPLEXITY
reportingOnly	false
seRule	false
seed	12,345
seedId	12,345
selMethod	AUTOMATIC
specifyRows	RANDOM
templateRevision	4
train	true

Property Name	Property Value
truncateLI	5
truncateUI	95
useVarOnce	false
userProbCutoff	false

Output

The SAS System

The TREESPLIT Procedure

Model Information	
Split Criterion	IGR
Pruning Method	Cost Complexity
Max Branches per Node	2
Max Tree Depth	15
Tree Depth Before Pruning	15
Tree Depth After Pruning	15
Number of Leaves Before Pruning	91
Number of Leaves After Pruning	68

	Training	Validation	Total
Number of Observations Read	2464	1055	3519
Number of Observations Used	2464	1055	3519

The SAS System

The TREESPLIT Procedure

Fit Statistics for Selected Tree		
	Number of Leaves	Misclassification Rate
Training	68	0.2784
Validation	68	0.3374

Variable Importance					
	Training		Validation		
Variable	Importance	Relative Importance	Importance	Relative Importance	Count
months	243.59	1.0000	86.2291	1.0000	10
speed_limit	49.9083	0.2049	32.4535	0.3764	4
latitude	111.30	0.4569	26.7304	0.3100	11
junc_detail_d	57.1999	0.2348	21.1547	0.2453	4
road_type_d	30.3278	0.1245	15.3166	0.1776	3
day_of_week	55.9951	0.2299	12.3884	0.1437	8
longitude	62.0510	0.2547	10.7155	0.1243	9
weath_con_d	43.4392	0.1783	6.8865	0.0799	6
num_of_vehi	12.0205	0.0493	4.1103	0.0477	4
first_road_class	7.1313	0.0293	3.1844	0.0369	3
light_con_d	19.6281	0.0806	3.0603	0.0355	5

Cost Complexity Pruning			
		Misclassification Rate	
Alpha	Number of Leaves	Training	Validation
36E-20	68	0.2784	0.3374
0.00020	64	0.2792	0.3403
0.00041	62	0.2800	0.3393
0.00061	60	0.2813	0.3393
0.00081	54	0.2861	0.3431
0.00101	50	0.2902	0.3517
0.00112	46	0.2946	0.3460
0.00122	44	0.2971	0.3469
0.00149	41	0.3015	0.3450
0.00162	39	0.3048	0.3517
0.00203	37	0.3088	0.3488
0.00213	33	0.3174	0.3536
0.00247	22	0.3446	0.3773
0.00284	21	0.3474	0.3773
0.00325	20	0.3506	0.3763
0.00406	19	0.3547	0.3801
0.00528	18	0.3600	0.3839
0.00609	17	0.3661	0.3877
0.00649	16	0.3726	0.3877
0.00684	9	0.4205	0.4161
0.00771	8	0.4282	0.4237
0.00933	7	0.4375	0.4256
0.01204	4	0.4736	0.4588
0.02760	3	0.5012	0.4938
0.03166	2	0.5329	0.5289
0.1092	1	0.6420	0.6417

The SAS System

The TREESPLIT Procedure

Predicted Probability Variables	
acci_severity	Variable
1	P_acci_severity1
3	P_acci_severity3
2	P_acci_severity2

Predicted Target Variable	
Level Index	Variable
	I_acci_severity

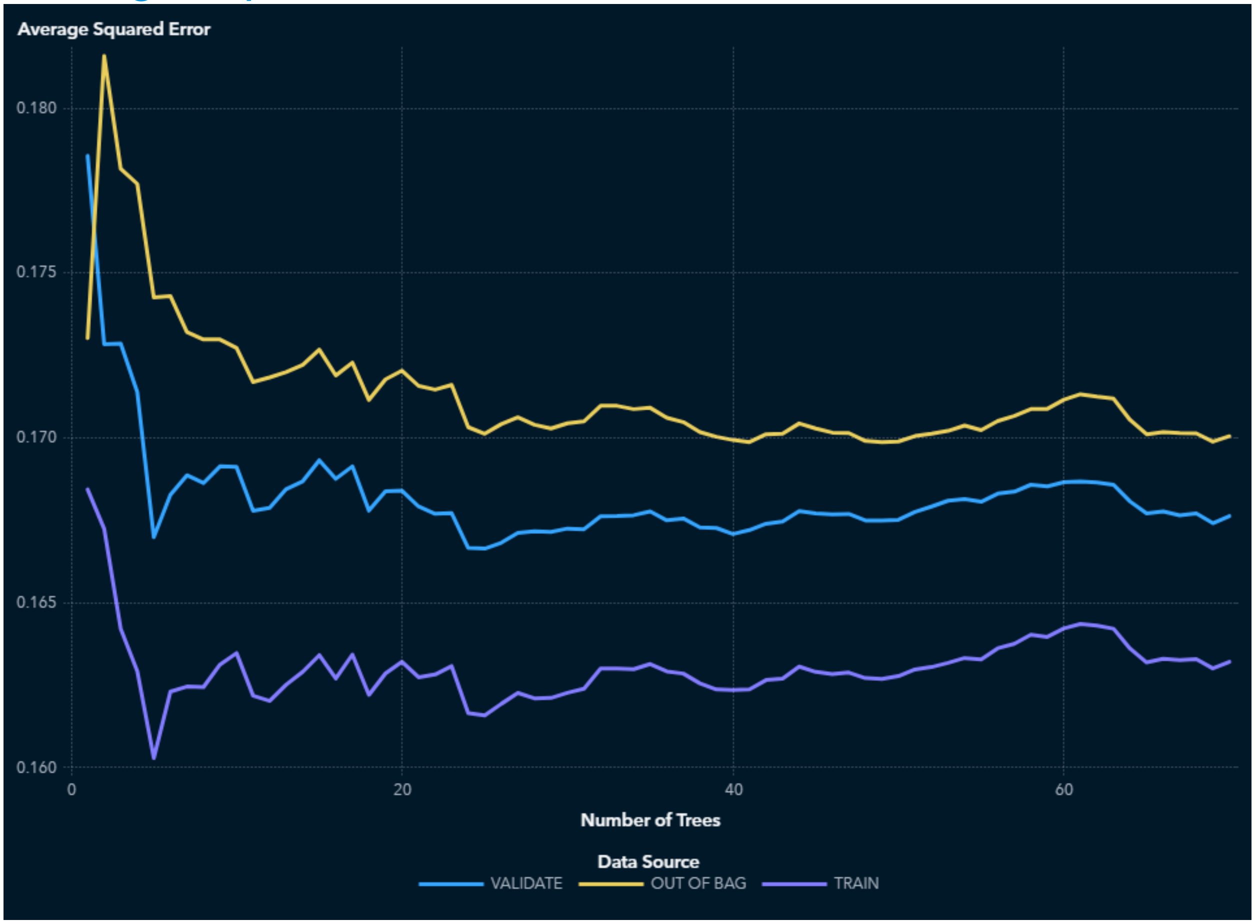
Predicting_Accidents_D

"Forest" Results

by: di00222@surrey.ac.uk

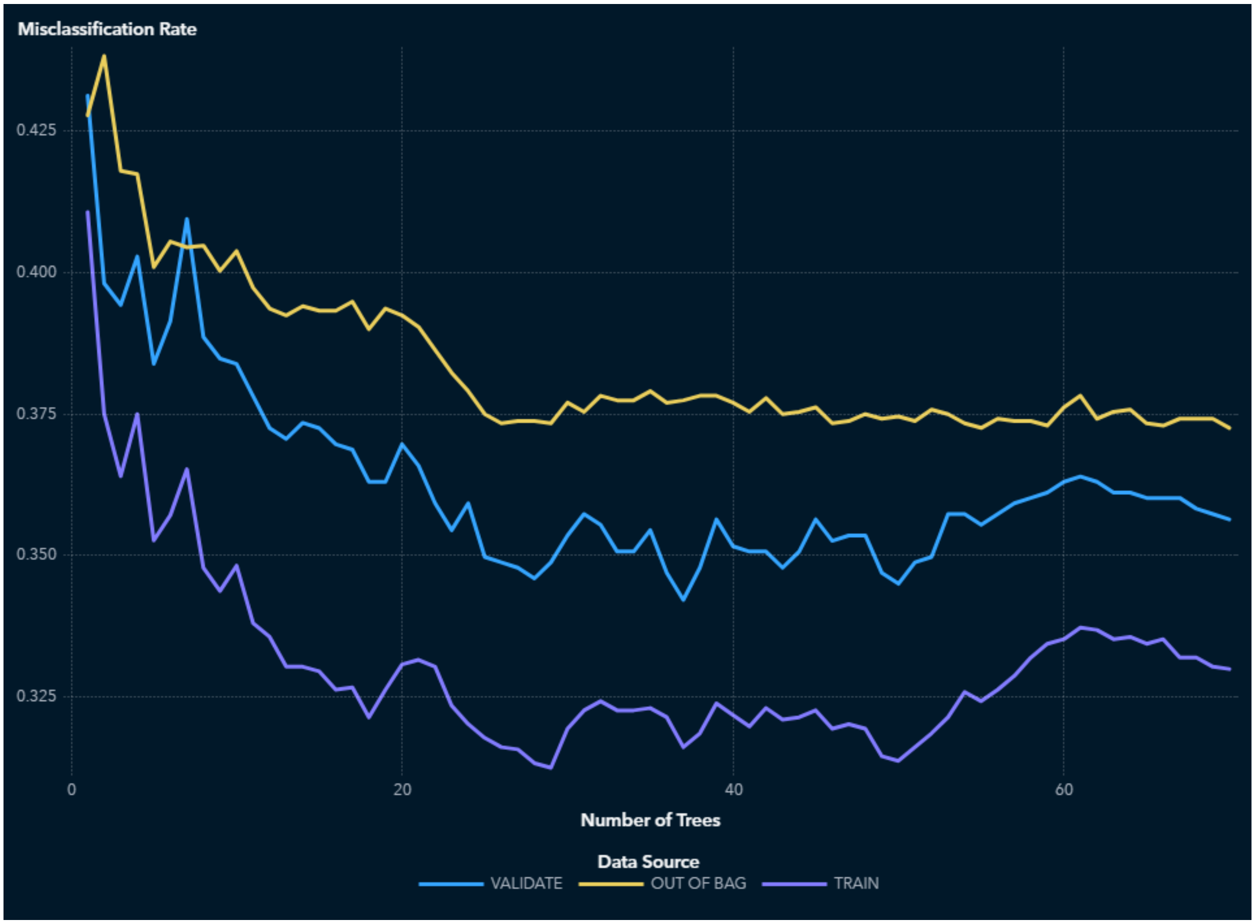
Contents

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 0.167 and occurs for 25 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 0.342 and occurs for 37 trees.

Variable Importance

Variable Label	Role	Variable Name	Training Importance
	INPUT	_months_	78.0266
	INPUT	latitude	29.7745
	INPUT	longitude	23.8729
	INPUT	junc_detail_d	20.9440
	INPUT	speed_limit	15.2127
	INPUT	weath_con_d	12.2947
	INPUT	road_type_d	8.2363
	INPUT	num_of_vehi	7.2818
	INPUT	day_of_week	7.2308
	INPUT	light_con_d	4.4609
	INPUT	first_road_class	1.7597

Importance Standard Deviation	Relative Importance
29.4937	1
23.6086	0.3816
15.8589	0.3060
13.3732	0.2684
12.6443	0.1950
6.0333	0.1576
8.0025	0.1056
13.0698	0.0933
6.9295	0.0927
3.8183	0.0572
2.8610	0.0226

Score Inputs

Name	Role	Variable Level	Type
acci_ref	ID	INTERVAL	N
day_of_week	INPUT	NOMINAL	N
first_road_class	INPUT	NOMINAL	N
junc_detail_d	INPUT	NOMINAL	C
latitude	INPUT	INTERVAL	N
light_con_d	INPUT	NOMINAL	C
longitude	INPUT	INTERVAL	N
num_of_vehi	INPUT	NOMINAL	N
road_type_d	INPUT	NOMINAL	C
speed_limit	INPUT	NOMINAL	N
weath_con_d	INPUT	NOMINAL	C
months	INPUT	NOMINAL	C

Variable Type	Variable Label	Variable Format	Variable Length
double			8
double			8
double			8
varchar			35
double			8
varchar			23
double			8
double			8
varchar			18
double			8
varchar			21
varchar			3

Score Outputs

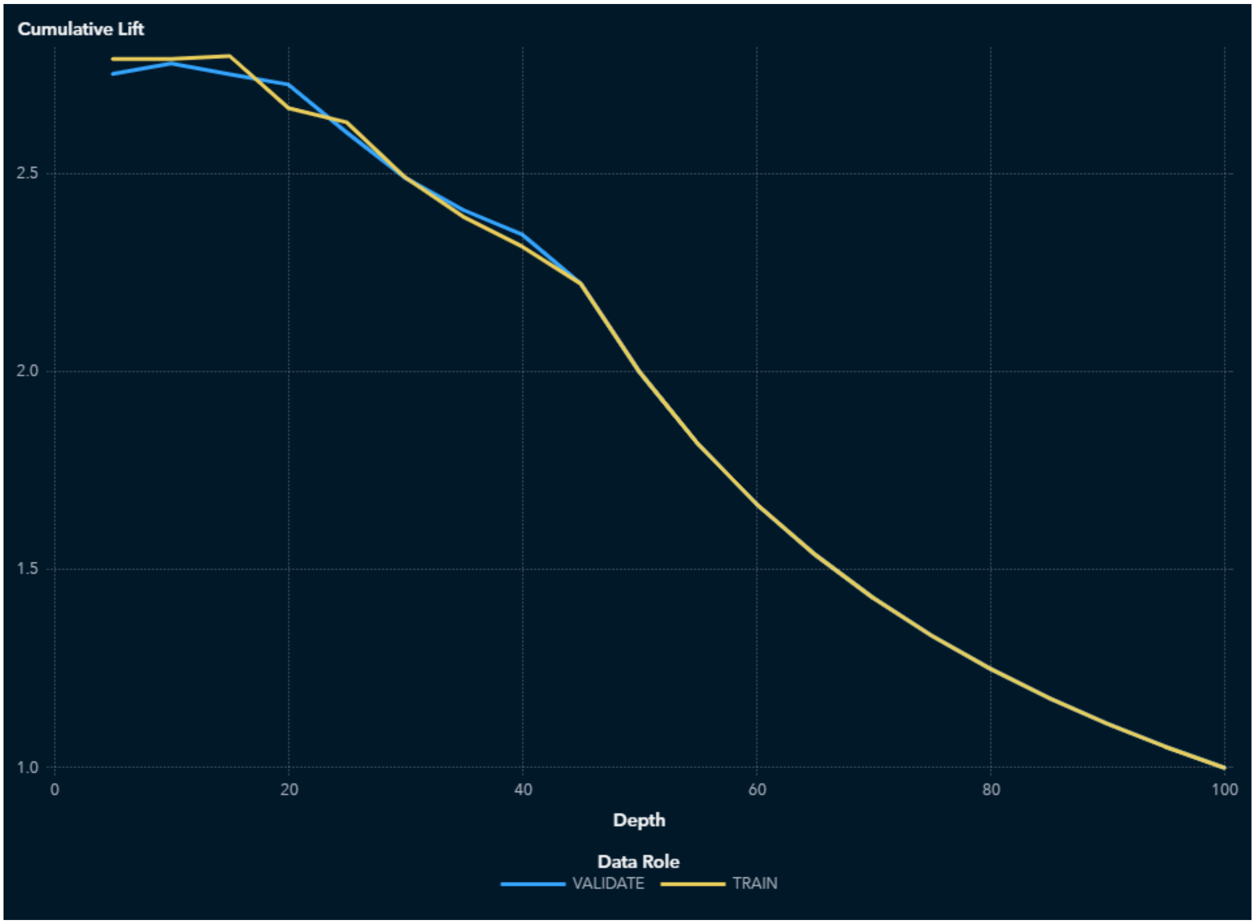
Name	Role	Type	Variable Type
EM_CLASSIFICATION	CLASSIFICATION	C	char
EM_EVENTPROBABILITY	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
I_acci_severity	CLASSIFICATION	C	char
P_acci_severity1	PREDICT	N	double
P_acci_severity2	PREDICT	N	double
P_acci_severity3	PREDICT	N	double
WARN	ASSESS	C	char

Variable Label	Variable Format	Variable Length	Creator
Predicted for acci_severity		12	forest
Probability for acci_severity=1		8	forest
Probability of Classification		8	forest
Into: acci_severity		32	forest
Predicted: acci_severity=1		8	forest
Predicted: acci_severity=2		8	forest
Predicted: acci_severity=3		8	forest
Warnings		4	forest

Function	Creator GUID
CLASSIFICATION	529fe103-ac27-45c9-b30c-c9a1db6310be

Function	Creator GUID
PREDICT	529fe103-ac27-45c9-b30c-c9a1db6310be
PREDICT	529fe103-ac27-45c9-b30c-c9a1db6310be
CLASSIFICATION	529fe103-ac27-45c9-b30c-c9a1db6310be
PREDICT	529fe103-ac27-45c9-b30c-c9a1db6310be
PREDICT	529fe103-ac27-45c9-b30c-c9a1db6310be
PREDICT	529fe103-ac27-45c9-b30c-c9a1db6310be
ASSESS	529fe103-ac27-45c9-b30c-c9a1db6310be

Cumulative Lift



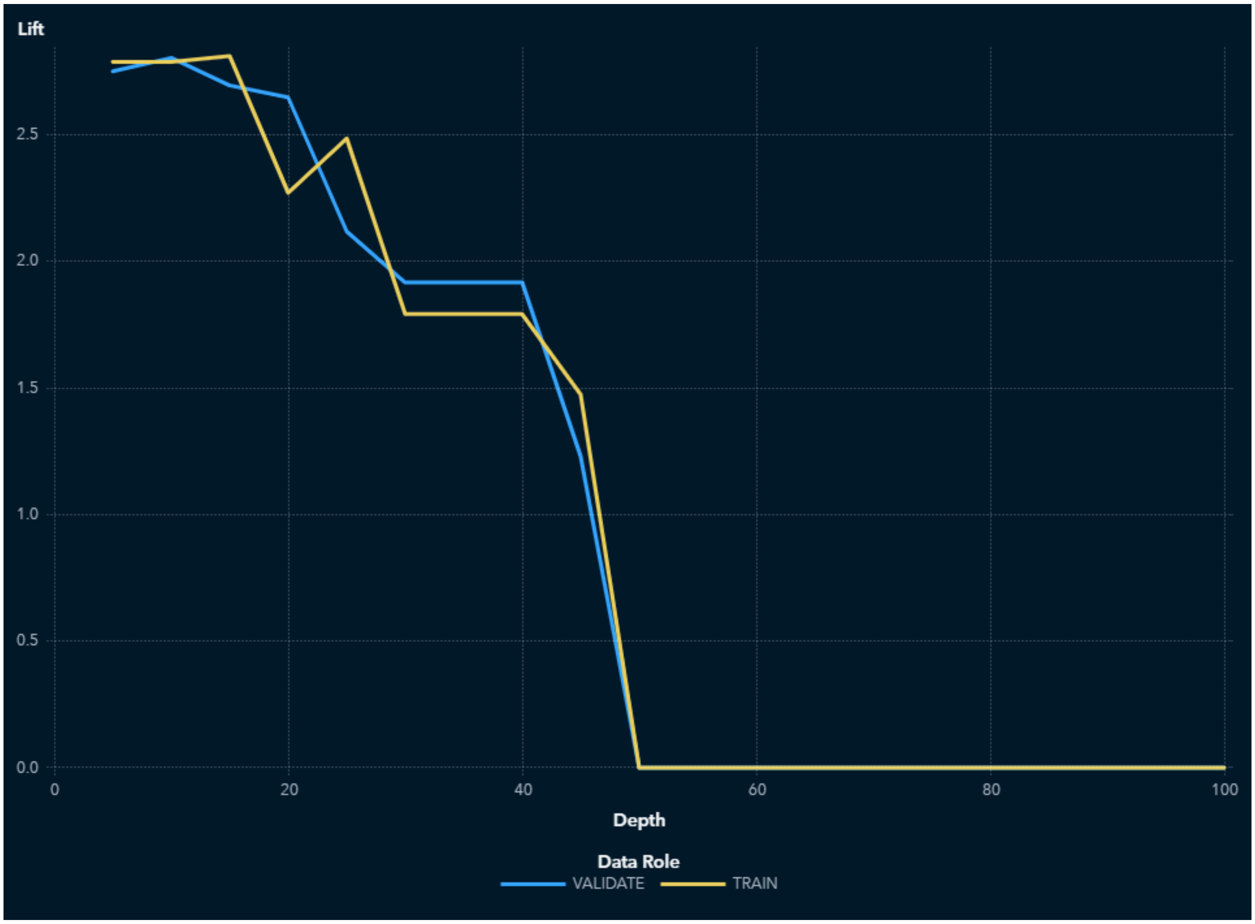
The VALIDATE partition has a Cumulative Lift of 2.78 in the 10% quantile (depth of 10) meaning there are 2.78 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Cumulative Lift of 2.79 in the 10% quantile (depth of 10) meaning there are 2.79 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



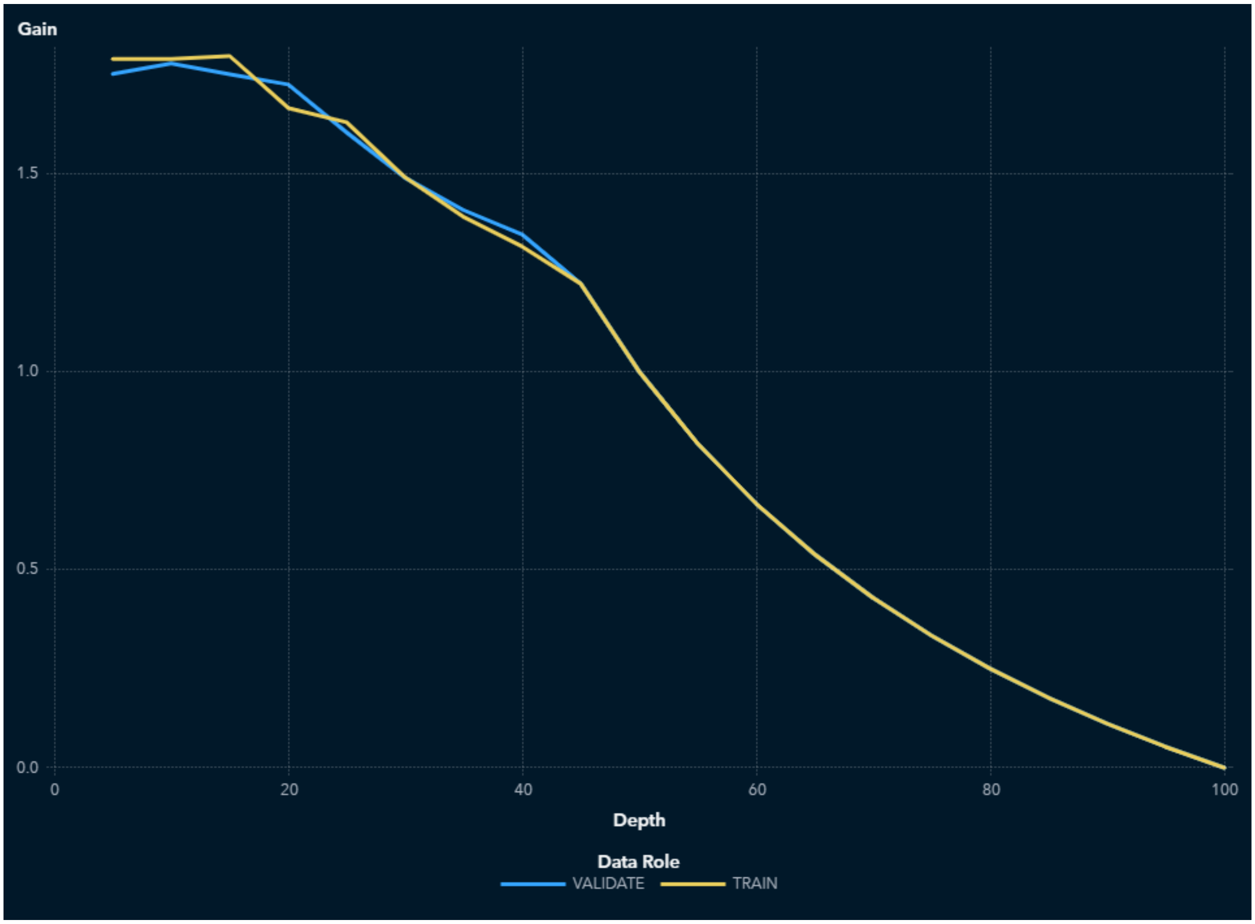
The VALIDATE partition has a Lift of 2.75 in the 5% quantile (depth of 5) meaning there are 2.75 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Lift of 2.79 in the 5% quantile (depth of 5) meaning there are 2.79 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



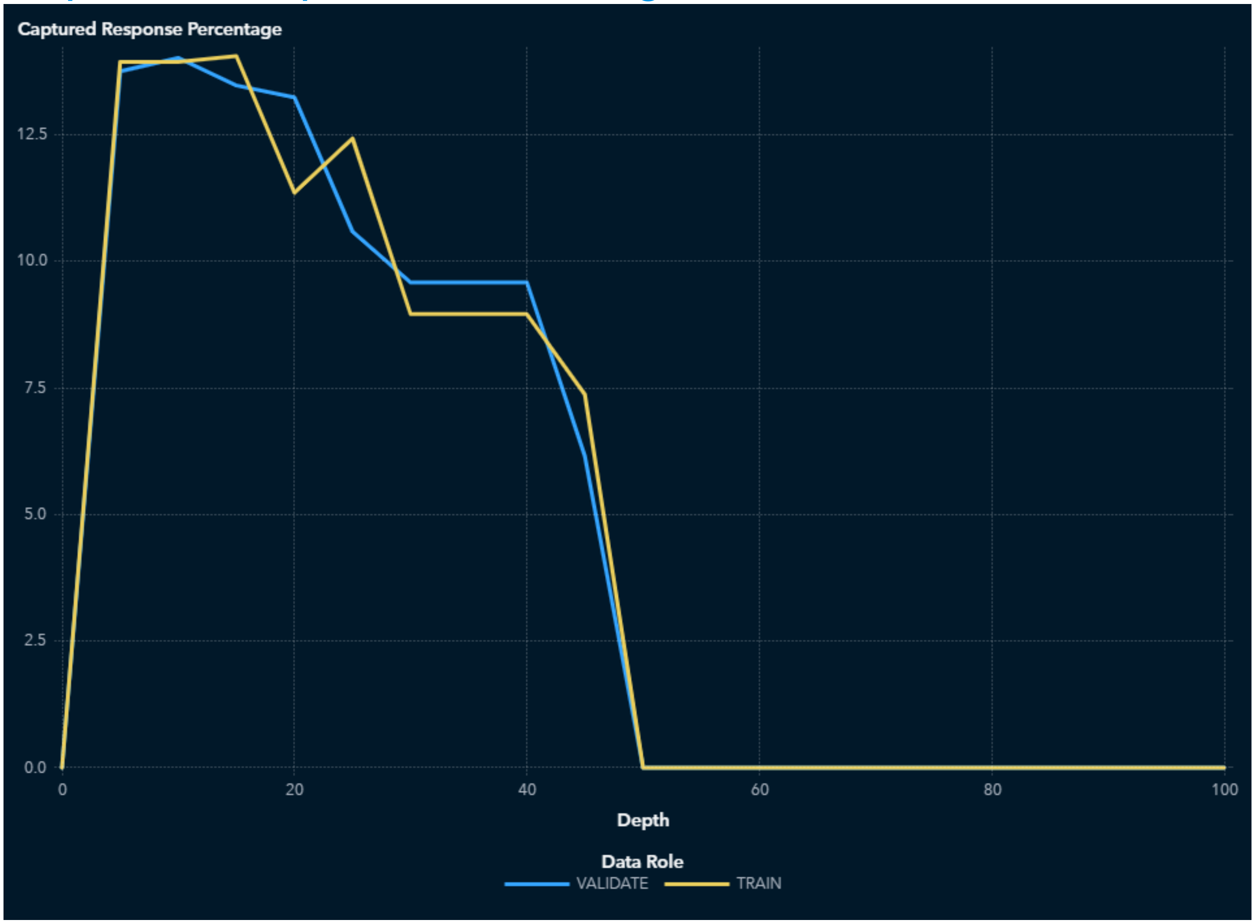
The VALIDATE partition has a Gain of 1.8 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 1.8.

The TRAIN partition has a Gain of 1.8 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 1.81.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

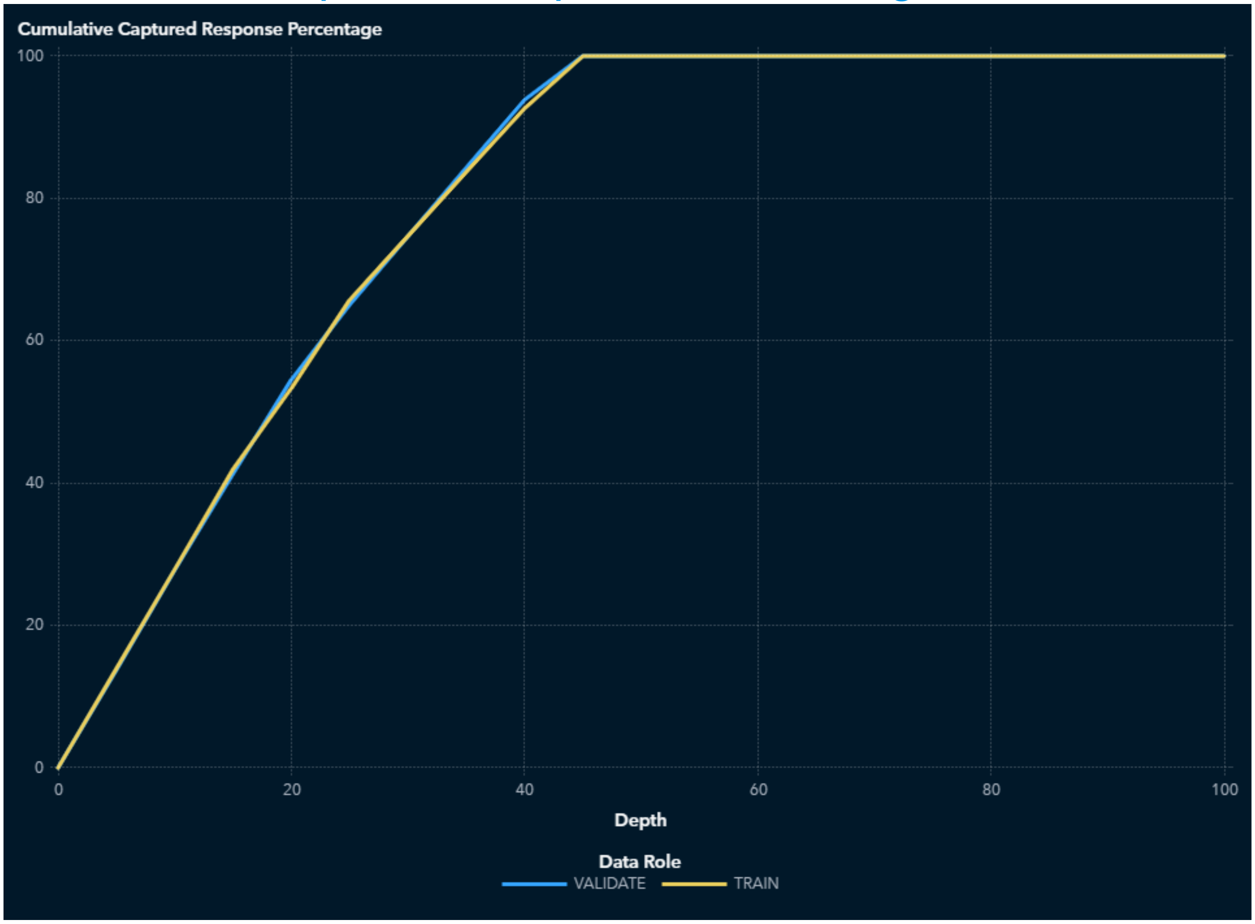


At the 5% quantile (depth of 5), the VALIDATE partition has a Captured response percentage of 13.8 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 14.02.

At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 13.9 (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 14.06.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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



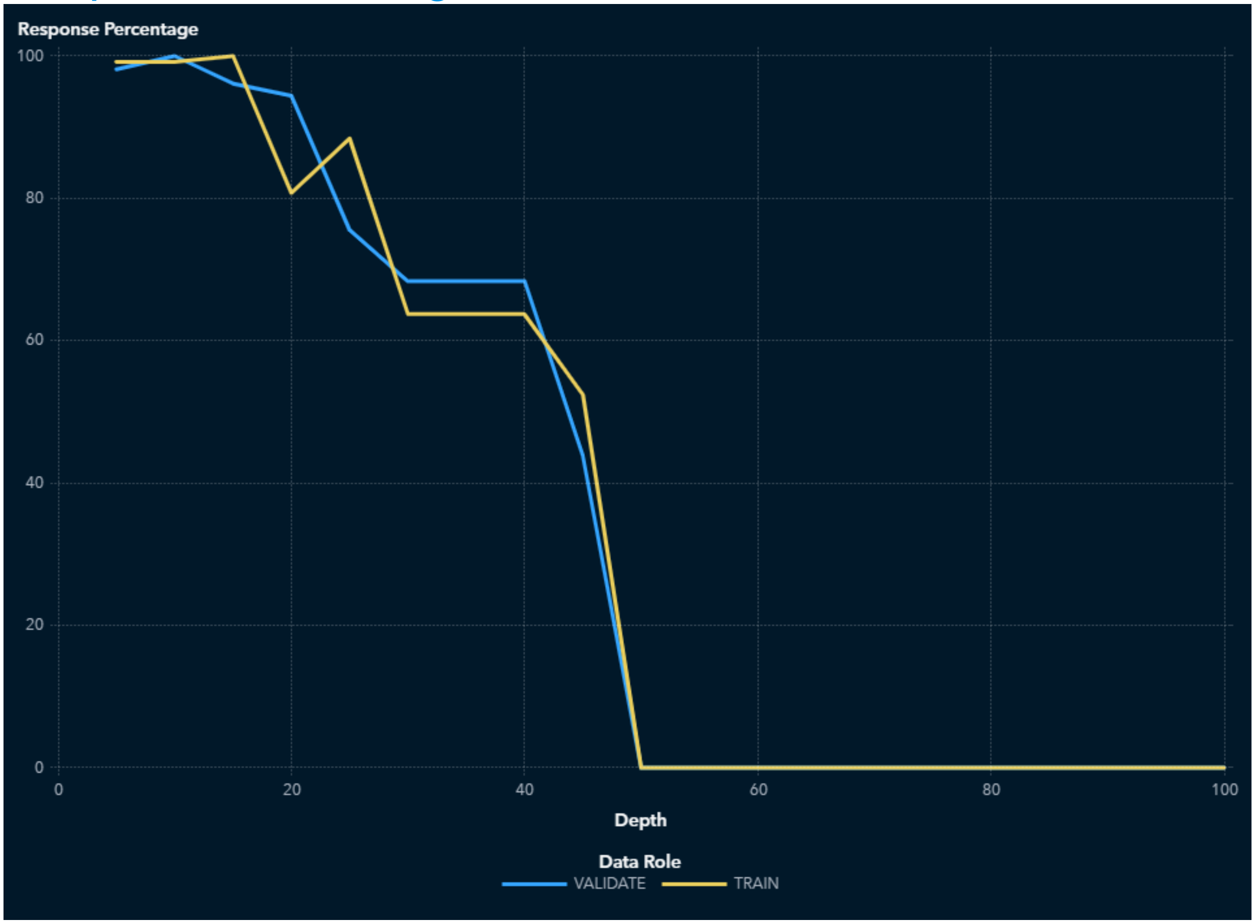
In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative captured response percentage of 27.8 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 28.04.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 27.9 (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 28.12.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

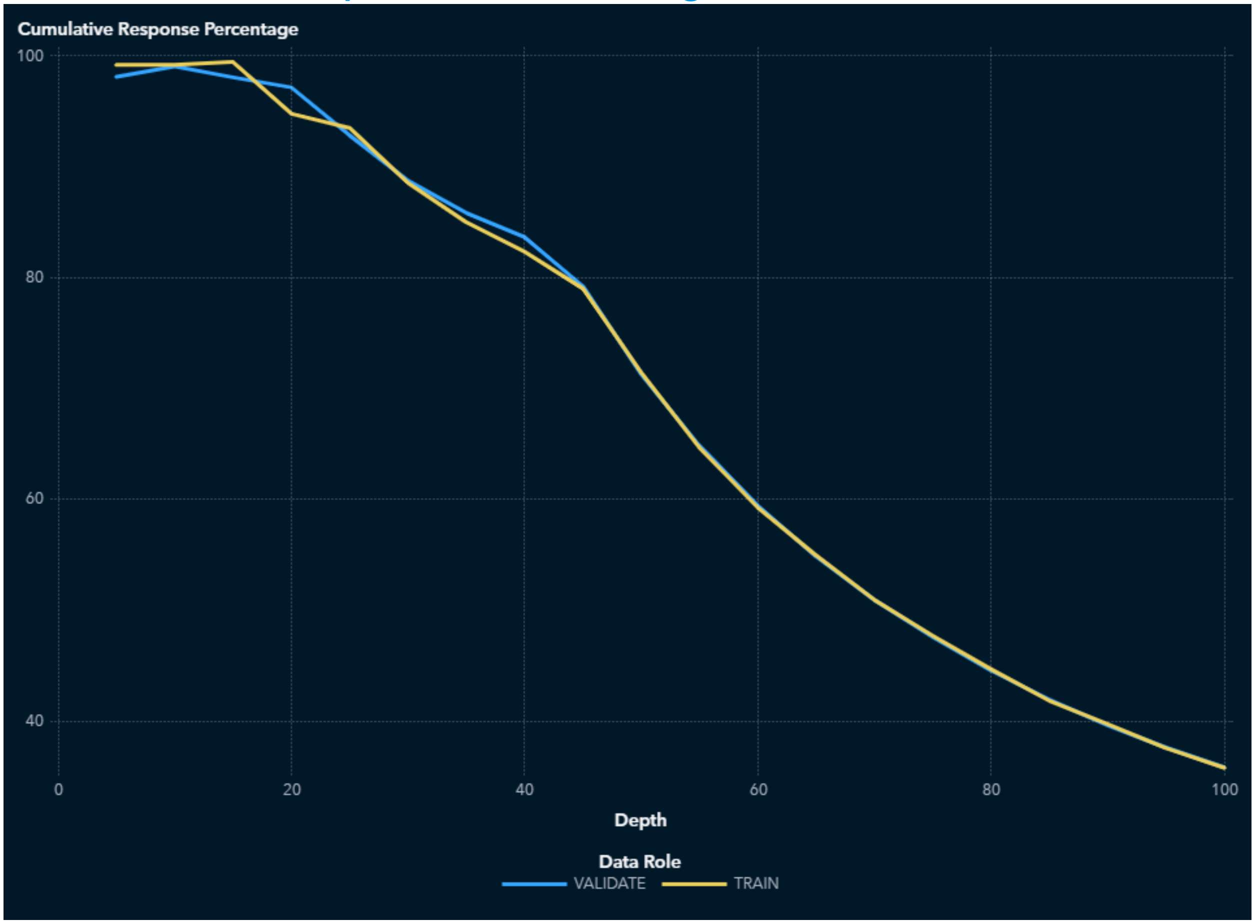


At the 5% quantile (depth of 5), the VALIDATE partition has a Response percentage of 98.1. The best possible value of Response percentage for this partition at depth 5 is 100.

At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 99.2. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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

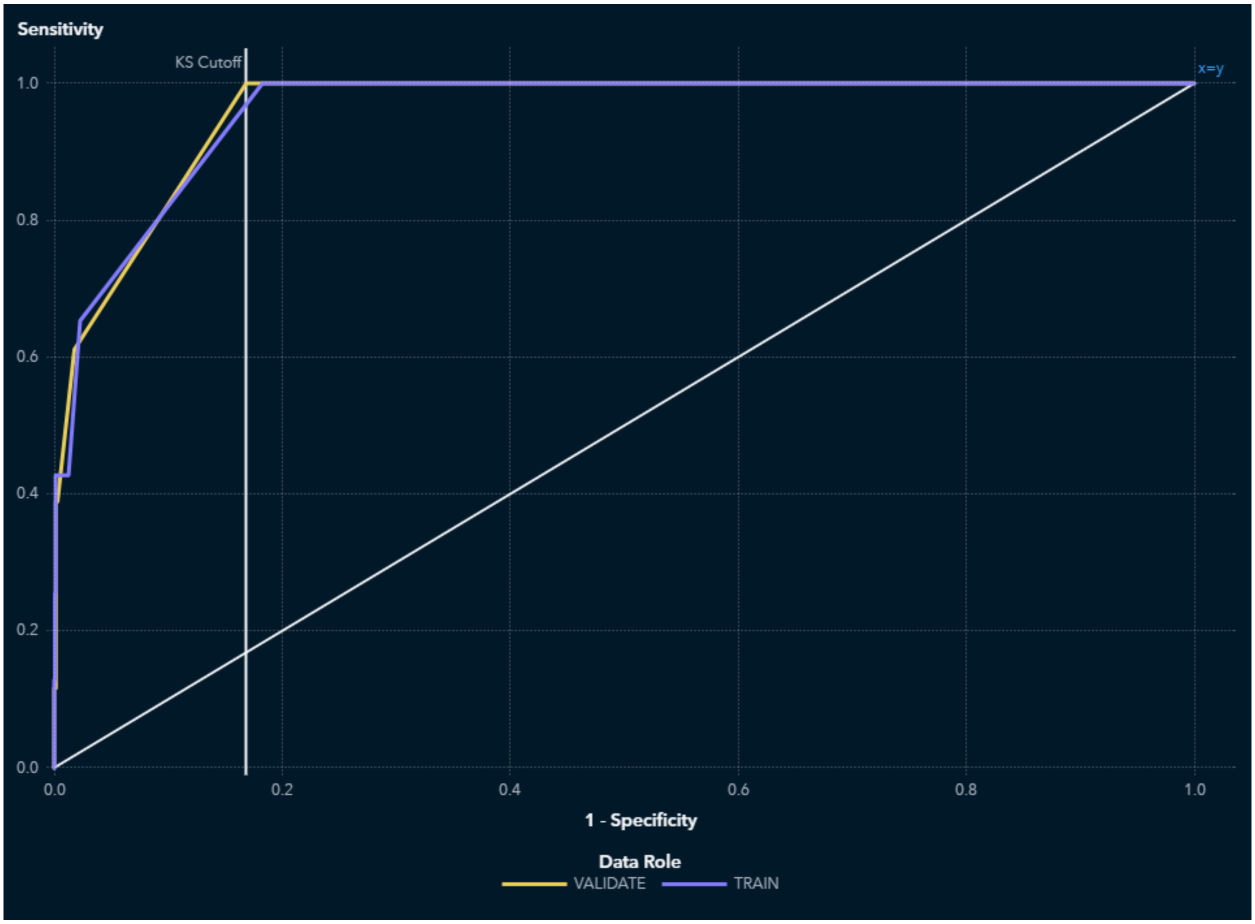


In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative response percentage of 99.1. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 99.2. 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_acci_severity1, which represents the predicted probability of the event "1" for the target acci_severity. 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 VALIDATE partition. The KS Cutoff line is drawn at the cutoff value 0.49, where the 1-specificity value is 0.168 and the sensitivity value is 1.

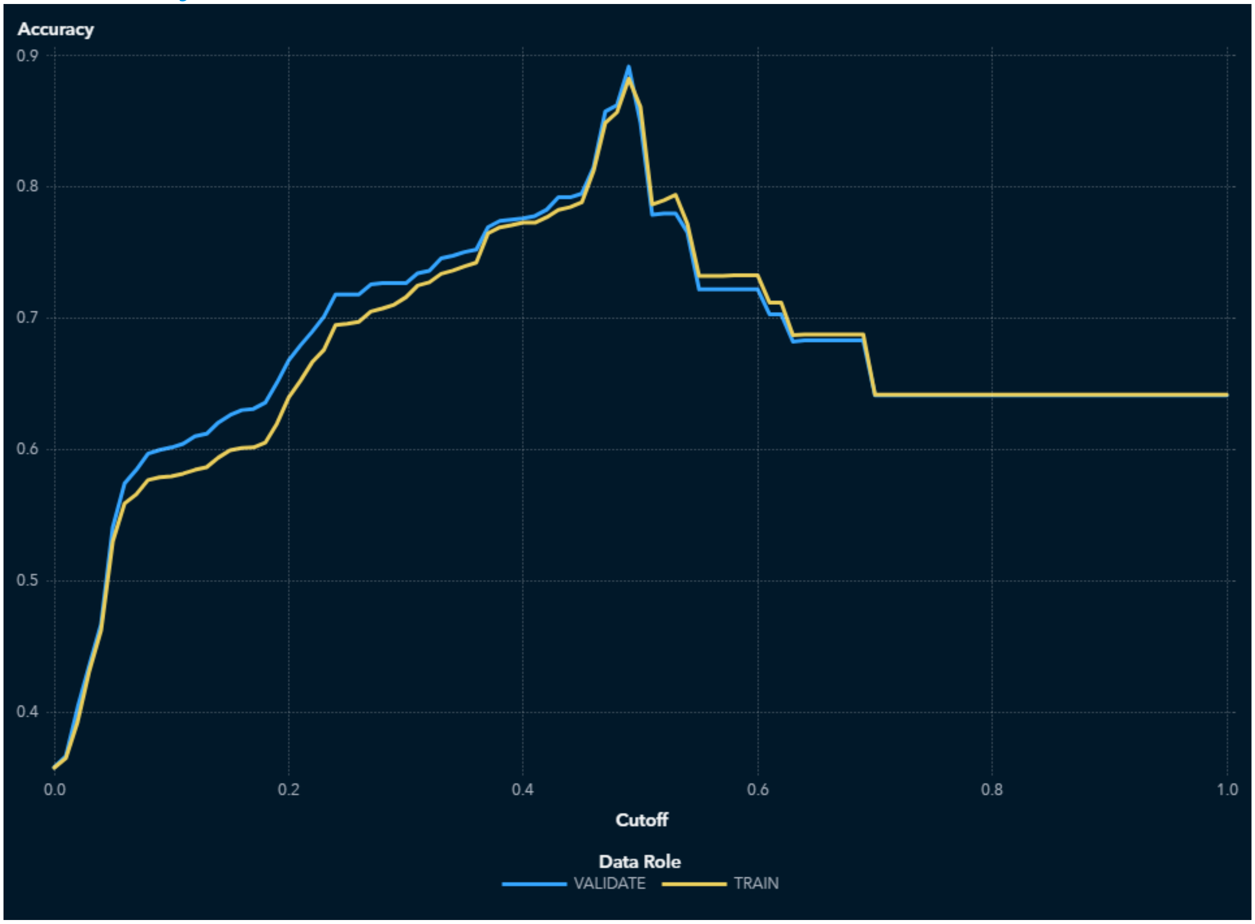
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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

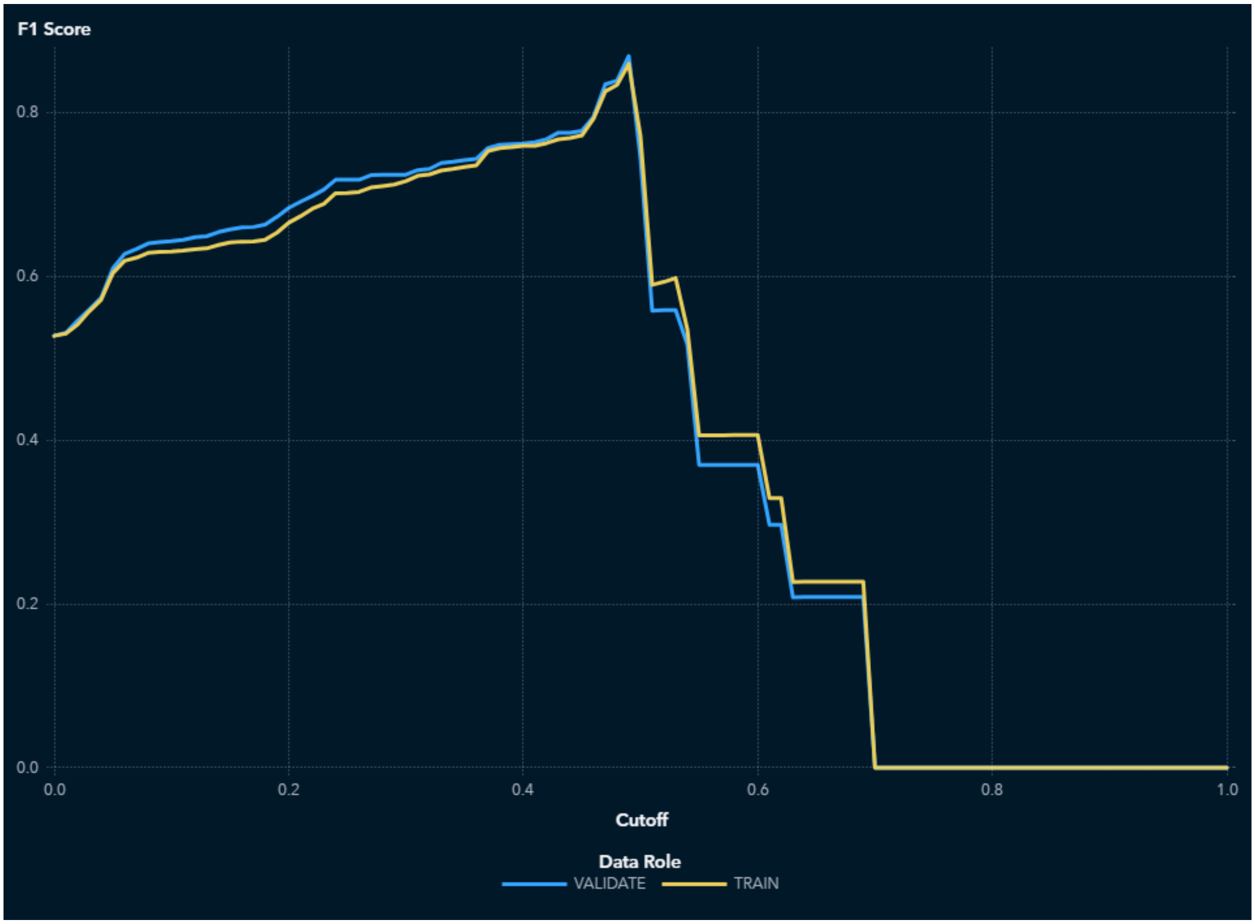


For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.861.

For this model, the accuracy in the VALIDATE partition at the cutoff of 0.5 is 0.849.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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 0.771.

For this model, the F1 score in the VALIDATE partition at the cutoff of 0.5 is 0.744.

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_acci_severity1, which is the predicted probability of the event "1" for the target acci_severity, is greater than or equal to the cutoff value. When P_acci_severity1 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
acci_severity	TRAIN	1	1
acci_severity	VALIDATE	0	0

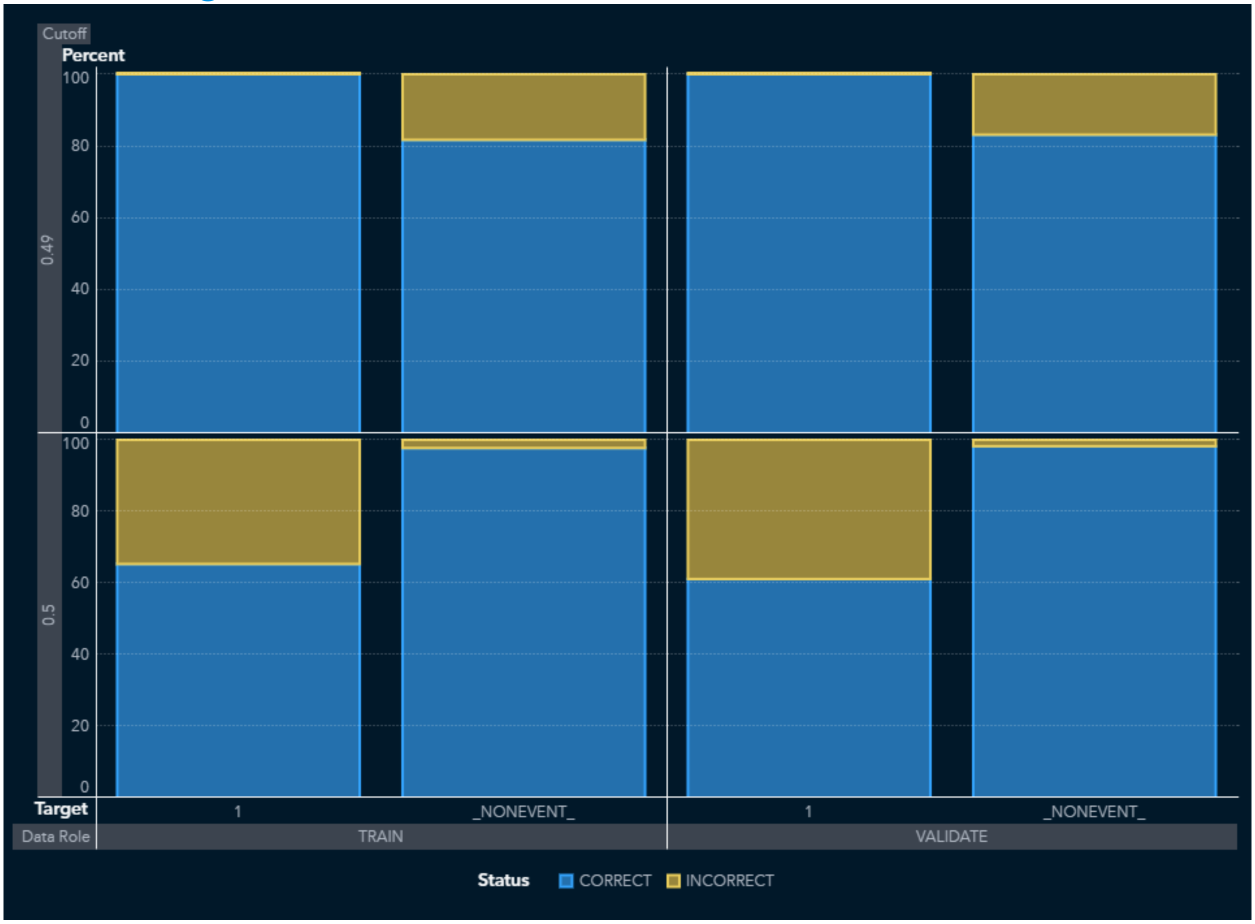
Number of Observations	Average Squared Error	Divisor for ASE	Root Average Squared Error
2,464	0.1632	2,464	0.4040
1,055	0.1676	1,055	0.4094

Misclassification Rate	Multi-Class Log Loss	KS (Youden)	Area Under ROC
0.3300	0.8264	0.8173	0.9601
0.3564	0.8397	0.8316	0.9611

Gini Coefficient	Gamma	Tau	KS Cutoff
0.9201	0.9766	0.4231	0.4900
0.9222	0.9830	0.4245	0.4900

KS at Default Cutoff	Misclassification Rate at KS Cutoff (Event)	Misclassification Rate (Event)
0.6303	0.1173	0.1388
0.5934	0.1081	0.1507

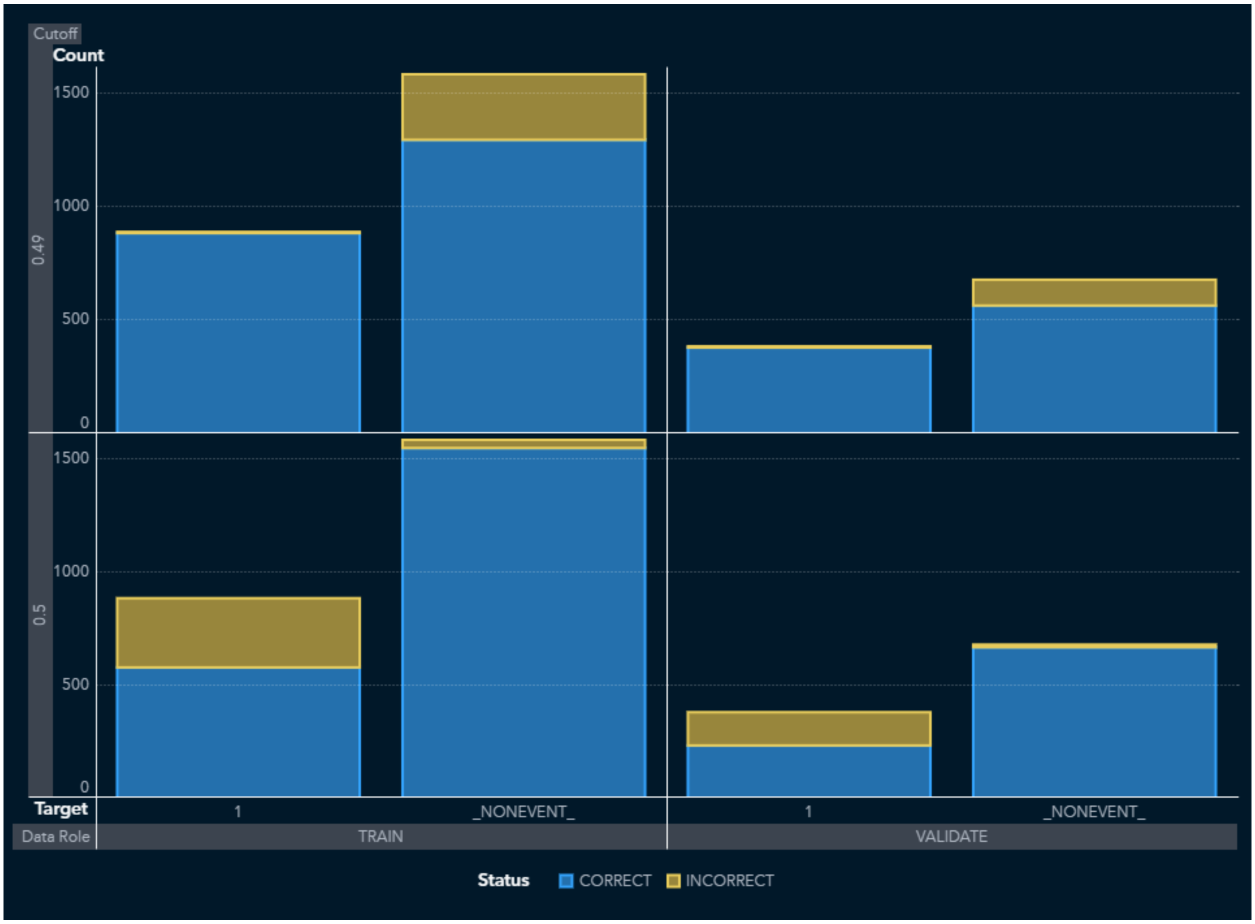
Percentage Plot



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: 0.49 (TRAIN), 0.49 (VALIDATE).

For this data, for the bar corresponding to the event level of acci_severity, "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 these KS cutoff values for existing partitions: 0.49 (TRAIN), 0.49 (VALIDATE).

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

Table

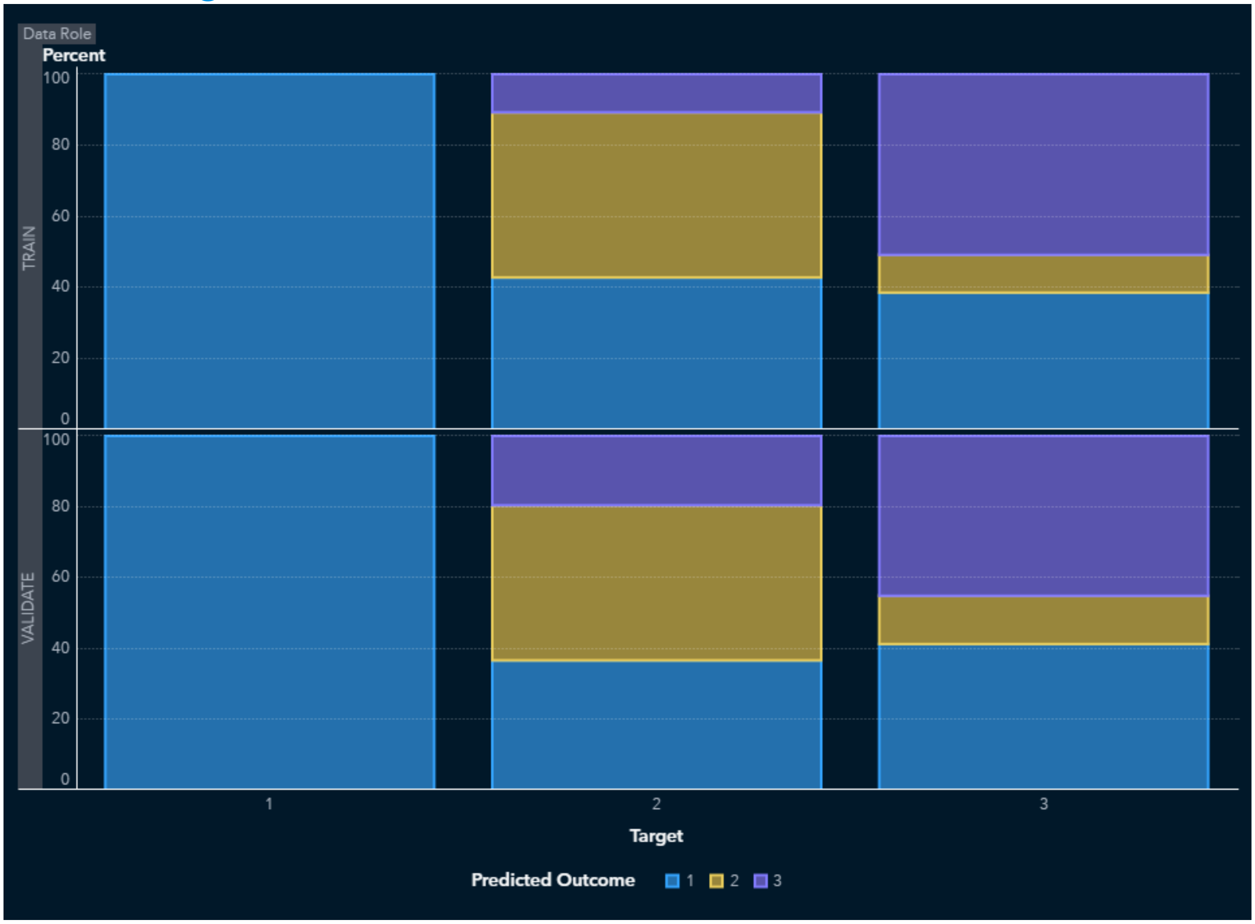
Cutoff	Cutoff Source	Target Name	Response
0.4900	KS	acci_severity	CORRECT
0.4900	KS	acci_severity	INCORRECT
0.4900	KS	acci_severity	CORRECT
0.4900	KS	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT
0.5000	Default	acci_severity	CORRECT
0.5000	Default	acci_severity	INCORRECT

Event	Value	Training Frequency	Validation Frequency
1	True Positive	882	378
1	False Negative	0	0
NONEVENT	True Negative	1,293	563
NONEVENT	False Positive	289	114
1	True Positive	576	231
1	False Negative	306	147
NONEVENT	True Negative	1,546	665
NONEVENT	False Positive	36	12

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	100	100	
	0	0	
	81.7320	83.1610	
	18.2680	16.8390	
	65.3061	61.1111	
	34.6939	38.8889	
	97.7244	98.2275	

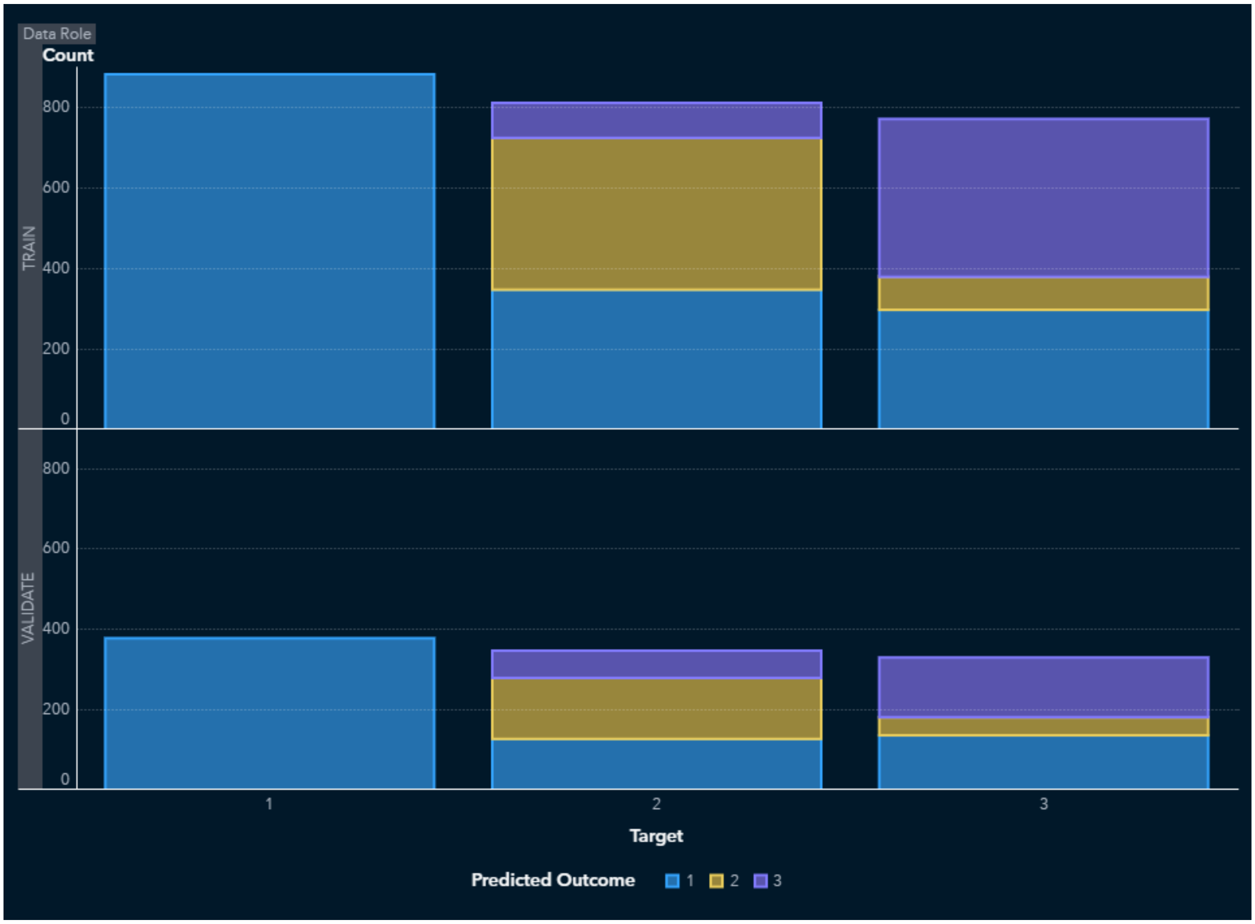
Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	2.2756	1.7725	

Percentage Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Count Plot



The Nominal Classification report displays either the percentage of or the number of observations predicting each target level. The plot is segmented by target level and partition level. The target level with the greatest predicted probability is the predicted outcome. A greater number of observations where the target and predicted outcome are the same indicates a better model.

Table

Target Name	Data Role	Target	Unformatted Target
acci_severity	VALIDATE	1	1
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	2	2
acci_severity	VALIDATE	3	3
acci_severity	VALIDATE	3	3
acci_severity	VALIDATE	3	3
acci_severity	TRAIN	1	1
acci_severity	TRAIN	2	2
acci_severity	TRAIN	2	2
acci_severity	TRAIN	2	2
acci_severity	TRAIN	3	3
acci_severity	TRAIN	3	3
acci_severity	TRAIN	3	3

Predicted Outcome	Count	Percent	Status
1	378	100	CORRECT
1	127	36.5994	INCORRECT
2	152	43.8040	CORRECT
3	68	19.5965	INCORRECT
1	136	41.2121	INCORRECT
2	45	13.6364	INCORRECT
3	149	45.1515	CORRECT
1	882	100	CORRECT
1	347	42.7867	INCORRECT
2	377	46.4858	CORRECT
3	87	10.7275	INCORRECT

Predicted Outcome	Count	Percent	Status
1	297	38.5214	INCORRECT
2	82	10.6355	INCORRECT
3	392	50.8431	CORRECT

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	V2024.09
defaultVarsPerTree	true
exactPctlLift	true
explainFidelity	false
explainInfo	false
fullDatasetReconstitution	false
iCriterionMethod	VARIANCE

Property Name	Property Value
icePlots	false
intBinMethod	QUANTILE
intervalBins	50
leafProp	0.0001
leafSize	5
leafSpec	COUNT
loh	0
maxBranch	3
maxCategories	128
maxDepth	6
maxNumShapVars	20
maxTrees	70
minUseInSearch	1
missingValue	USEINSEARCH
nBins	50
pdNumImportantInputs	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
truncateLI	5
truncateUI	95
userProbCutoff	false
varsToTry	100
voteMethod	PROBABILITY

Output

The SAS System

The FOREST Procedure

Model Information	
Number of Trees	70
Number of Variables Per Split	4
Seed	12345
Bootstrap Percentage	60
Number of Bins	50
Number of Input Variables	11
Maximum Number of Tree Nodes	59
Minimum Number of Tree Nodes	17
Maximum Number of Branches	3
Minimum Number of Branches	2
Maximum Depth	6
Minimum Depth	6
Maximum Number of Leaves	32
Minimum Number of Leaves	10
Maximum Leaf Size	1246
Minimum Leaf Size	5
OOB Misclassification Rate	0.37256494
Average Number of Leaves	21.4142857

	Training	Validation	Total
Number of Observations Read	2464	1055	3519
Number of Observations Used	2464	1055	3519

Variable Importance			
Variable	Importance	Std Dev Importance	Relative Importance
months	78.0266	29.4937	1.0000
latitude	29.7745	23.6086	0.3816
longitude	23.8729	15.8589	0.3060
junc_detail_d	20.9440	13.3732	0.2684
speed_limit	15.2127	12.6443	0.1950
weath_con_d	12.2947	6.0333	0.1576
road_type_d	8.2363	8.0025	0.1056
num_of_vehi	7.2818	13.0698	0.0933
day_of_week	7.2308	6.9295	0.0927
light_con_d	4.4609	3.8183	0.0572
first_road_class	1.7597	2.8610	0.0226

Fit Statistics

Number of Trees	OOB Average Square Error	Training Average Square Error	Validation Average Square Error	OOB Misclassification Rate	Training Misclassification Rate	Validation Misclassification Rate	OOB Log Loss	Training Log Loss	Validation Log Loss
1	0.173	0.168	0.179	0.428	0.411	0.431	0.953	0.872	1.089
2	0.182	0.167	0.173	0.438	0.375	0.398	0.986	0.828	0.846
3	0.178	0.164	0.173	0.418	0.364	0.394	0.954	0.815	0.846
4	0.178	0.163	0.171	0.417	0.375	0.403	0.938	0.808	0.838
5	0.174	0.160	0.167	0.401	0.353	0.384	0.909	0.801	0.823
6	0.174	0.162	0.168	0.405	0.357	0.391	0.868	0.809	0.829
7	0.173	0.162	0.169	0.404	0.365	0.409	0.856	0.808	0.830
8	0.173	0.162	0.169	0.405	0.348	0.389	0.865	0.810	0.831
9	0.173	0.163	0.169	0.400	0.344	0.385	0.858	0.813	0.834
10	0.173	0.163	0.169	0.404	0.348	0.384	0.848	0.815	0.833
11	0.172	0.162	0.168	0.397	0.338	0.378	0.846	0.812	0.830
12	0.172	0.162	0.168	0.394	0.336	0.373	0.848	0.813	0.832
13	0.172	0.163	0.168	0.392	0.330	0.371	0.850	0.815	0.834
14	0.172	0.163	0.169	0.394	0.330	0.373	0.851	0.817	0.835
15	0.173	0.163	0.169	0.393	0.330	0.373	0.853	0.819	0.838
16	0.172	0.163	0.169	0.393	0.326	0.370	0.850	0.817	0.836
17	0.172	0.163	0.169	0.395	0.327	0.369	0.853	0.821	0.839
18	0.171	0.162	0.168	0.390	0.321	0.363	0.848	0.815	0.833
19	0.172	0.163	0.168	0.394	0.326	0.363	0.851	0.818	0.835
20	0.172	0.163	0.168	0.392	0.331	0.370	0.852	0.820	0.836
21	0.172	0.163	0.168	0.390	0.332	0.366	0.850	0.818	0.834
22	0.171	0.163	0.168	0.386	0.330	0.359	0.850	0.818	0.834
23	0.172	0.163	0.168	0.382	0.323	0.355	0.851	0.820	0.834
24	0.170	0.162	0.167	0.379	0.320	0.359	0.847	0.815	0.831
25	0.170	0.162	0.167	0.375	0.318	0.350	0.846	0.815	0.831
26	0.170	0.162	0.167	0.373	0.316	0.349	0.848	0.817	0.832
27	0.171	0.162	0.167	0.374	0.316	0.348	0.851	0.820	0.835
28	0.170	0.162	0.167	0.374	0.313	0.346	0.850	0.819	0.835
29	0.170	0.162	0.167	0.373	0.313	0.349	0.850	0.819	0.835
30	0.170	0.162	0.167	0.377	0.319	0.354	0.850	0.820	0.835
31	0.171	0.162	0.167	0.375	0.323	0.357	0.851	0.820	0.835
32	0.171	0.163	0.168	0.378	0.324	0.355	0.853	0.823	0.837
33	0.171	0.163	0.168	0.377	0.323	0.351	0.853	0.824	0.838
34	0.171	0.163	0.168	0.377	0.323	0.351	0.854	0.825	0.839
35	0.171	0.163	0.168	0.379	0.323	0.355	0.855	0.825	0.839
36	0.171	0.163	0.168	0.377	0.321	0.347	0.853	0.824	0.838
37	0.170	0.163	0.168	0.377	0.316	0.342	0.852	0.824	0.838
38	0.170	0.163	0.167	0.378	0.319	0.348	0.852	0.823	0.838
39	0.170	0.162	0.167	0.378	0.324	0.356	0.851	0.822	0.837
40	0.170	0.162	0.167	0.377	0.322	0.352	0.850	0.822	0.837
41	0.170	0.162	0.167	0.375	0.320	0.351	0.851	0.823	0.838
42	0.170	0.163	0.167	0.378	0.323	0.351	0.852	0.824	0.839
43	0.170	0.163	0.167	0.375	0.321	0.348	0.852	0.824	0.839
44	0.170	0.163	0.168	0.375	0.321	0.351	0.853	0.826	0.840
45	0.170	0.163	0.168	0.376	0.323	0.356	0.852	0.825	0.839
46	0.170	0.163	0.168	0.373	0.319	0.353	0.852	0.824	0.839
47	0.170	0.163	0.168	0.374	0.320	0.354	0.851	0.824	0.839
48	0.170	0.163	0.168	0.375	0.319	0.354	0.851	0.824	0.839
49	0.170	0.163	0.167	0.374	0.315	0.347	0.851	0.824	0.839
50	0.170	0.163	0.168	0.375	0.314	0.345	0.851	0.825	0.839
51	0.170	0.163	0.168	0.374	0.316	0.349	0.852	0.826	0.840
52	0.170	0.163	0.168	0.376	0.319	0.350	0.852	0.826	0.841
53	0.170	0.163	0.168	0.375	0.321	0.357	0.853	0.826	0.842
54	0.170	0.163	0.168	0.373	0.326	0.357	0.853	0.827	0.842
55	0.170	0.163	0.168	0.373	0.324	0.355	0.853	0.828	0.842
56	0.171	0.164	0.168	0.374	0.326	0.357	0.855	0.829	0.843
57	0.171	0.164	0.168	0.374	0.329	0.359	0.855	0.829	0.843
58	0.171	0.164	0.169	0.374	0.332	0.360	0.856	0.830	0.844
59	0.171	0.164	0.169	0.373	0.334	0.361	0.856	0.830	0.844
60	0.171	0.164	0.169	0.376	0.335	0.363	0.857	0.831	0.845
61	0.171	0.164	0.169	0.378	0.337	0.364	0.858	0.832	0.845
62	0.171	0.164	0.169	0.374	0.337	0.363	0.858	0.832	0.845
63	0.171	0.164	0.169	0.375	0.335	0.361	0.857	0.831	0.844
64	0.171	0.164	0.168	0.376	0.336	0.361	0.855	0.829	0.843
65	0.170	0.163	0.168	0.373	0.334	0.360	0.853	0.827	0.841
66	0.170	0.163	0.168	0.373	0.335	0.360	0.853	0.827	0.841
67	0.170	0.163	0.168	0.374	0.332	0.360	0.852	0.827	0.840
68	0.170	0.163	0.168	0.374	0.332	0.358	0.852	0.827	0.840
69	0.170	0.163	0.167	0.374	0.330	0.357	0.851	0.826	0.839
70	0.170	0.163	0.168	0.373	0.330	0.356	0.852	0.826	0.840

Output CAS Tables

CAS Library	Name	Number of Rows	Number of Columns
CASUSER (di00222@surrey.ac.uk)	_model_5db0ca64-0393-f64a-8c8c-654d27f491a	2798	56

Predicted Probability Variables	
acci_severity	Variable
1	P_acci_severity1
3	P_acci_severity3
2	P_acci_severity2

Predicted Target Variable	
Level Index	Variable
	I_acci_severity