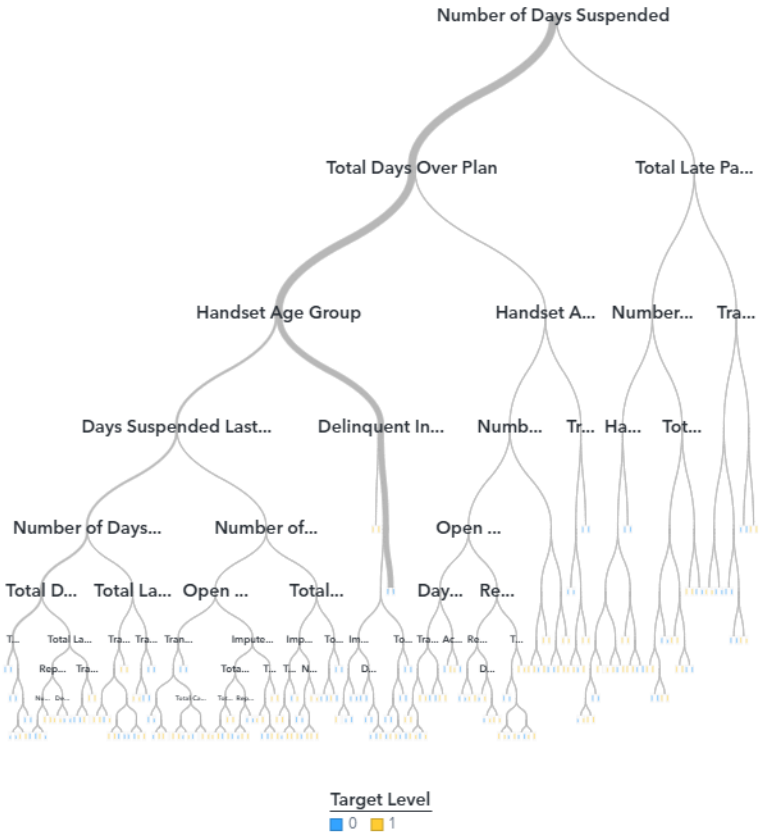
Demo

"Decision Tree" Results

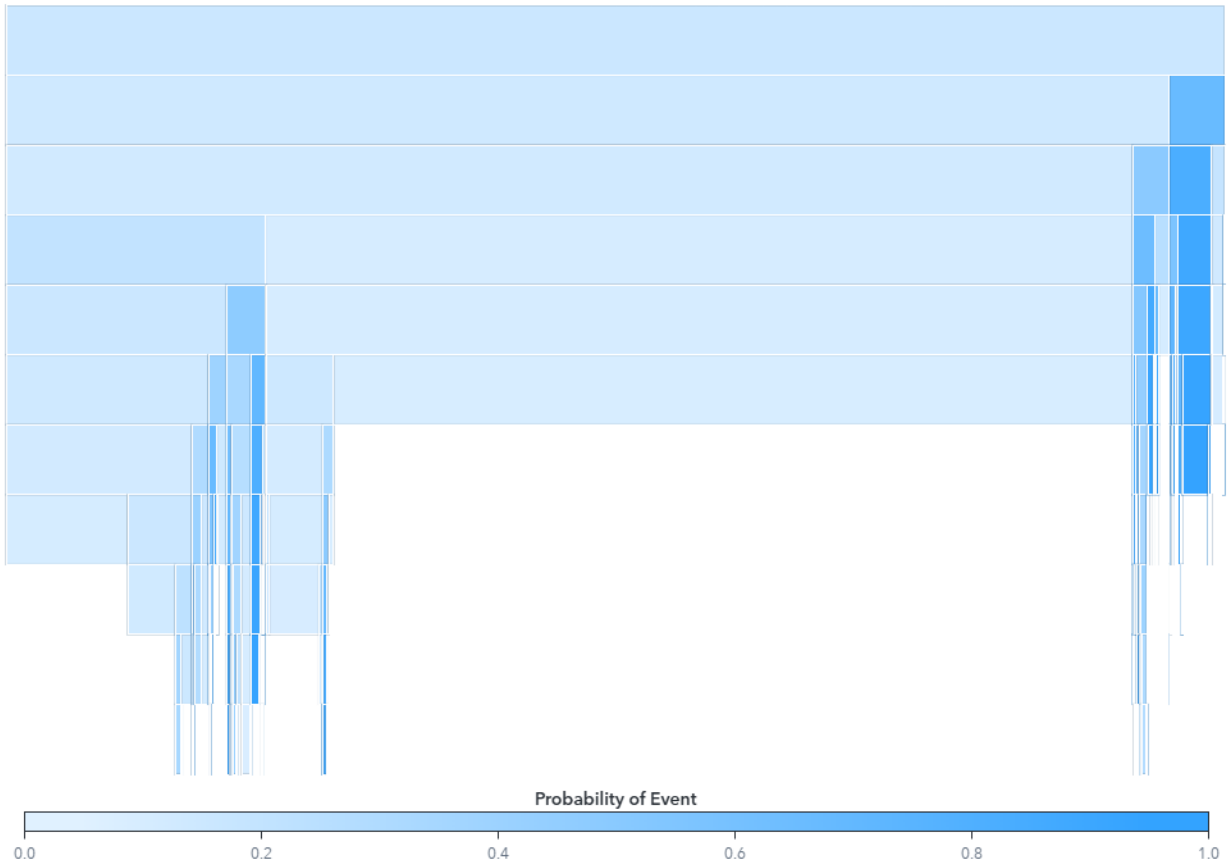
by: rapenaflor@mymail.mapua.edu.ph

Contents

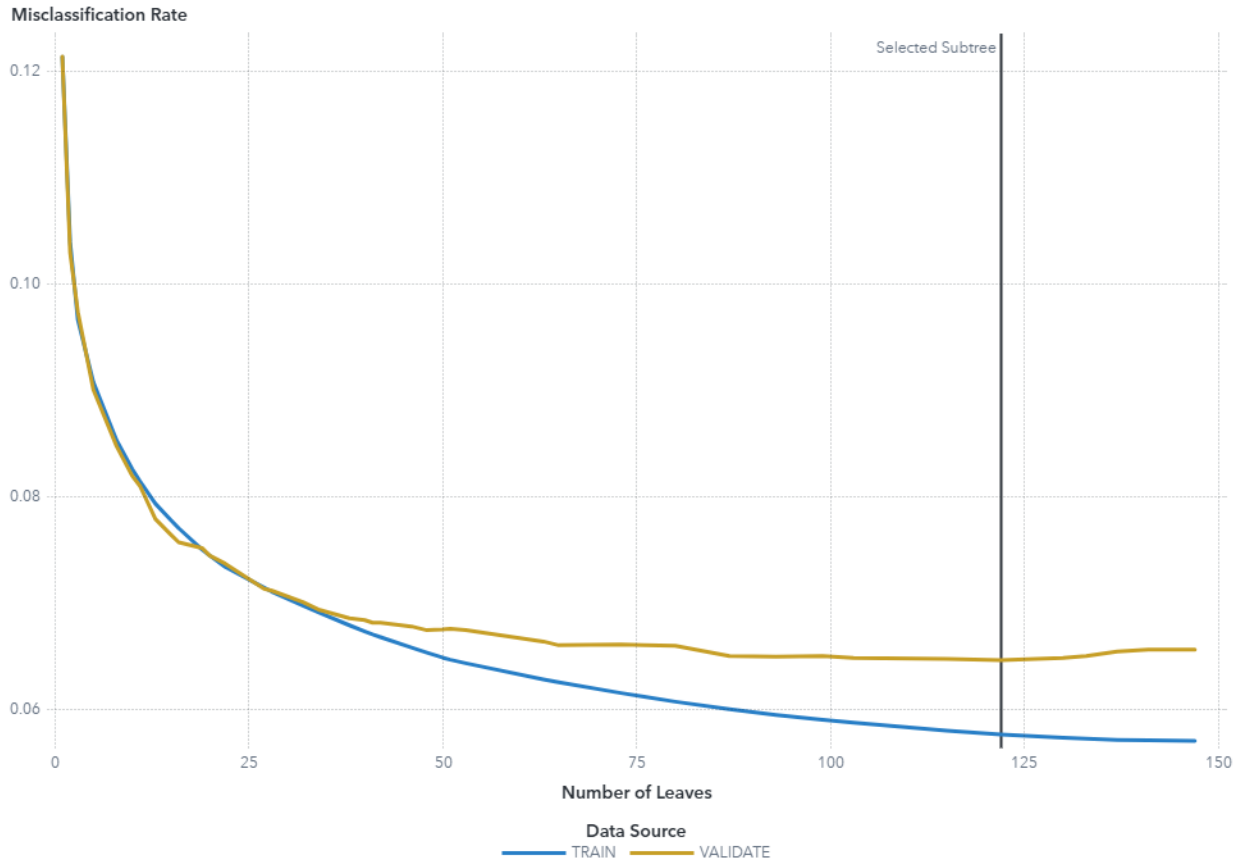
Tree Diagram



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 122 leaves with a misclassification rate of 0.065 for the VALIDATE partition.

Variable Importance

Variable Name	Training Importance	Training Relative Importance	Validation Relative Importance
curr_days_susp	1,396.4214	1	1
pymts_late_ltd	534.3120	0.3826	0.3339
ever_days_over_plan	423.6394	0.3034	0.3303
handset_age_grp	383.2681	0.2745	0.2813
avg_days_susp	300.8500	0.2154	0.1745
times_susp	226.5809	0.1623	0.1646
calls_care_ltd	178.3559	0.1277	0.0895
LOG_MB_Data_Usg_M04	115.6250	0.0828	0.0868
LOG_MB_Data_Usg_M05	80.5528	0.0577	0.0569
LOG_MB_Data_Usg_M07	84.3520	0.0604	0.0496
wrk_orders	79.1102	0.0567	0.0407
REP_seconds_of_data_norm	64.0916	0.0459	0.0287
delinq_indicator	87.4760	0.0626	0.0229
LOG_MB_Data_Usg_M08	24.9799	0.0179	0.0195
IMP_seconds_of_data_log	31.3355	0.0224	0.0167
ever_times_over_plan	51.8382	0.0371	0.0135
IMP_LOG_MB_Data_Usg_M09	20.9723	0.0150	0.0124
LOG_MB_Data_Usg_M06	9.4500	0.0068	0.0080

Variable Name	Training Importance	Training Relative Importance	Validation Relative Importance
REP_tot_mb_data_curr	5.7356	0.0041	0.0065
days_openwrkorders	11.4754	0.0082	0.0052
call_category_1	15.2126	0.0109	0.0029
REP_mb_data_usg_roamm02	12.1477	0.0087	0.0028
IMP_REP_mou_onnet_pct_MOM	35.4546	0.0254	0.0025
REP_mb_data_usg_m03	4.7619	0.0034	0.0021
unsolv_tsupcomplnt	9.5466	0.0068	0.0012
bill_data_usg_m09	8.8088	0.0063	0.0012
REP_mb_data_usg_roamm03	1.6563	0.0012	0.0006
REP_bill_data_usg_m06	4.7376	0.0034	0.0005
IMP_REP_mb_data_ndist_mo6m	3.7556	0.0027	-0.0004
REP_calls_in_offpk	6.7493	0.0048	-0.0008
num_tsupcomplnts	3.2000	0.0023	-0.0016
IMP_REP_data_device_age	4.7647	0.0034	-0.0027
IMP_cs_ttl_rural	5.6343	0.0040	-0.0036
billing_cycle	10.3700	0.0074	-0.0046
region	3.2000	0.0023	-0.0053
IMP_nbr_data_cdrs	4.7417	0.0034	-0.0054
IMP_cs_afr_amer	9.7074	0.0070	-0.0061
REP_mb_data_usg_m02	10.0204	0.0072	-0.0150

Count	Validation Importance
5	603.5223
8	201.5150
2	199.3302
5	169.7938
6	105.3371
5	99.3158
14	54.0095
6	52.3845
7	34.3625
6	29.9358
4	24.5786
5	17.3260
7	13.8462
2	11.7489
2	10.0639
7	8.1237
1	7.4606
1	4.8488
1	3.8970
2	3.1507
3	1.7638
1	1.6990
3	1.4837
1	1.2698
2	0.7425
1	0.7246
1	0.3623
1	0.2936

Count	Validation Importance
1	-0.2664
1	-0.4932
1	-0.9600
1	-1.6129
1	-2.1618
2	-2.7995
1	-3.2000
1	-3.2482
2	-3.6656
1	-9.0654

Score Inputs

Name	Role	Variable Level	Type
avg_arpu_3m	INPUT	INTERVAL	N
avg_data_chrgs_3m	INPUT	INTERVAL	N
avg_data_prem_chrgs_3m	INPUT	INTERVAL	N
avg_days_susp	INPUT	INTERVAL	N
avg_overage_chrgs_3m	INPUT	INTERVAL	N
billing_cycle	INPUT	NOMINAL	N
bill_data_usg_m03	INPUT	INTERVAL	N
bill_data_usg_m06	INPUT	INTERVAL	N
bill_data_usg_m09	INPUT	INTERVAL	N
calls_care_ltd	INPUT	INTERVAL	N
calls_in_offpk	INPUT	INTERVAL	N
calls_in_pk	INPUT	INTERVAL	N
calls_out_offpk	INPUT	INTERVAL	N
calls_out_pk	INPUT	INTERVAL	N
calls_total	INPUT	INTERVAL	N
call_category_1	INPUT	NOMINAL	C
cs_afr_amer	INPUT	INTERVAL	N
cs_caucasian	INPUT	INTERVAL	N
cs_hispanic	INPUT	INTERVAL	N
cs_med_home_value	INPUT	INTERVAL	N
cs_other	INPUT	INTERVAL	N
cs_pct_home_owner	INPUT	INTERVAL	N
cs_ttl_female	INPUT	INTERVAL	N
cs_ttl_hhlds	INPUT	INTERVAL	N
cs_ttl_male	INPUT	INTERVAL	N

Name	Role	Variable Level	Type
cs_ttl_mdage	INPUT	INTERVAL	N
cs_ttl_pop	INPUT	INTERVAL	N
cs_ttl_rural	INPUT	INTERVAL	N
cs_ttl_urban	INPUT	INTERVAL	N
curr_days_susp	INPUT	INTERVAL	N
data_device_age	INPUT	INTERVAL	N
data_prem_chrgs_curr	INPUT	INTERVAL	N
days_openwrkorders	INPUT	INTERVAL	N
delinq_indicator	INPUT	NOMINAL	N
ever_days_over_plan	INPUT	INTERVAL	N
ever_times_over_plan	INPUT	INTERVAL	N
handset_age_grp	INPUT	NOMINAL	C
lifetime_value	INPUT	INTERVAL	N
mb_data_ndist_mo 6m	INPUT	INTERVAL	N
mb_data_usg_m01	INPUT	INTERVAL	N
mb_data_usg_m02	INPUT	INTERVAL	N
mb_data_usg_m03	INPUT	INTERVAL	N
MB_Data_Usg_M04	INPUT	INTERVAL	N
MB_Data_Usg_M05	INPUT	INTERVAL	N
MB_Data_Usg_M06	INPUT	INTERVAL	N
MB_Data_Usg_M07	INPUT	INTERVAL	N
MB_Data_Usg_M08	INPUT	INTERVAL	N
MB_Data_Usg_M09	INPUT	INTERVAL	N
mb_data_usg_roamm01	INPUT	INTERVAL	N
mb_data_usg_roa	INPUT	INTERVAL	N

Name	Role	Variable Level	Type
mm02			
mb_data_usg_roamm03	INPUT	INTERVAL	N
mou_onnet_pct_MOM	INPUT	INTERVAL	N
mou_total_pct_MOM	INPUT	INTERVAL	N
nbr_data_cdrs	INPUT	INTERVAL	N
num_tsupcomplnts	INPUT	NOMINAL	N
pymts_late_ltd	INPUT	NOMINAL	N
region	INPUT	NOMINAL	C
seconds_of_data_log	INPUT	INTERVAL	N
seconds_of_data_norm	INPUT	INTERVAL	N
times_susp	INPUT	NOMINAL	N
tot_drpd_pr1	INPUT	INTERVAL	N
tot_mb_data_curr	INPUT	INTERVAL	N
tot_mb_data_roam_curr	INPUT	INTERVAL	N
tot_overage_chgs	INPUT	INTERVAL	N
tot_voice_chrgs_curr	INPUT	INTERVAL	N
unsolv_tsupcomplnt	INPUT	NOMINAL	N
voice_tot_bill_mou_curr	INPUT	INTERVAL	N
wrk_orders	INPUT	NOMINAL	N

Variable Type	Variable Label	Variable Format	Variable Length
double	3M Avg Revenue per User	DOLLAR8.2	8
double	3M Avg Data	DOLLAR8.2	8

Variable Type	Variable Label	Variable Format	Variable Length
	Charges		
double	3M Avg Premium Data Charges	DOLLAR8.2	8
double	Days Suspended Last 6M	BEST2.0	8
double	3M Avg Overage Charges	DOLLAR8.2	8
double	Billing Cycle	BESTD2.0	8
double	3M Avg Billed Data Usage	COMMA8.0	8
double	6M Avg Billed Data Usage	COMMA8.0	8
double	9M Avg Billed Data Usage	COMMA8.0	8
double	Total Calls to Care Lifetime	BEST12.0	8
double	Calls Incoming Off-Peak	COMMA8.0	8
double	Calls Incoming Peak	COMMA8.0	8
double	Calls Outgoing Off-Peak	COMMA8.0	8
double	Calls Outgoing Peak	COMMA8.0	8
double	Total Calls Curr	COMMA8.0	8
char	Call Center Category 1	$CHAR28.	28
double	Census Area African-American	BEST8.3	8
double	Census Area Caucasian	BEST8.3	8
double	Census Area Hispanic	BEST8.3	8

Variable Type	Variable Label	Variable Format	Variable Length
double	Census Area Median Home Value Index	BEST4.2	8
double	Census Area Other	BEST8.3	8
double	Census Area Percent Home Owner	BEST8.3	8
double	Census Area Total Female	BEST8.3	8
double	Census Area Total Households	COMMA12.0	8
double	Census Area Total Males	BEST8.3	8
double	Census Area Median Age	BEST3.0	8
double	Census Area Total Population	COMMA12.0	8
double	Census Area Total Rural	BEST8.3	8
double	Census Area Total Urban	BEST8.3	8
double	Number of Days Suspended	BEST4.0	8
double	Avg Age of Devices on Plan	COMMA10.0	8
double	Premium Data Charges	DOLLAR8.2	8
double	Days of Open Work Orders	BEST2.0	8
double	Delinquent Indicator	BEST2.0	8
double	Total Days Over Plan	BEST2.0	8
double	Total Times Over	BEST2.0	8

Variable Type	Variable Label	Variable Format	Variable Length
	Plan		
char	Handset Age Group	$CHAR12.	12
double	Lifetime Value	DOLLAR8.2	8
double	6M Avg Billed Data Usage Normally Distributed	BEST12.0	8
double	MB Data Usage 1 Mth Prior	COMMA8.0	8
double	MB Data Usage 2 Mths Prior	COMMA8.0	8
double	MB Data Usage 3 Mths Prior	COMMA8.0	8
double	MB of Data Usage Month 4	BEST12.0	8
double	MB of Data Usage Month 5	BEST12.0	8
double	MB of Data Usage Month 6	BEST12.0	8
double	MB of Data Usage Month 7	BEST12.0	8
double	MB of Data Usage Month 8	BEST12.0	8
double	MB of Data Usage Month 9	BEST12.0	8
double	MB Data Usage Roam 1 Mth Prior	COMMA8.0	8
double	MB Data Usage Roam 2 Mths Prior	COMMA8.0	8
double	MB Data Usage Roam 3 Mths Prior	COMMA8.0	8
double	Minutes On Network Pct Change Month over	PERCENT8.2	8

Variable Type	Variable Label	Variable Format	Variable Length
	Month		
double	Minutes Total Pct Change Month over Month	PERCENT8.2	8
double	Number of Data Records	COMMA10.0	8
double	Tech Support Complaints - LTD	BEST2.0	8
double	Total Late Payments Lifetime	BEST2.0	8
char	Account Region	$CHAR13.	13
double	Seconds of Data - Natural Log		8
double	Seconds of Data - Normalized		8
double	Number of Times Suspended	BEST4.0	8
double	Number of Dropped Calls 1 Mth Prior	COMMA8.0	8
double	Total MB of Data Usage	COMMA8.0	8
double	Total MB of Roam Data Usage	COMMA8.0	8
double	Total Overage Charges	DOLLAR8.2	8
double	Total Voice Charges	DOLLAR8.2	8
double	Unresolved Tech Support Complaint - LTD	BEST2.0	8
double	Total Voice Billed Minutes of Use	COMMA8.0	8
double	Open Work Orders	BEST8.0	8

Score Outputs

Name	Role	Type	Variable Type
EM_CLASSIFICATION	CLASSIFICATION	C	char
EM_EVENTPROBABILITY	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
IMP_LOG_MB_Data_Usg_M09	INPUT	N	double
IMP_REP_data_device_age	INPUT	N	double
IMP_REP_mb_data_ndist_mo6m	INPUT	N	double
IMP_REP_mou_onnet_pct_MOM	INPUT	N	double
IMP_REP_mou_total_pct_MOM	INPUT	N	double
IMP_avg_arpu_3m	INPUT	N	double
IMP_avg_data_chrgs_3m	INPUT	N	double
IMP_avg_data_prem_chrgs_3m	INPUT	N	double
IMP_avg_overage_chrgs_3m	INPUT	N	double
IMP_cs_afr_amer	INPUT	N	double
IMP_cs_caucasian	INPUT	N	double
IMP_cs_hispanic	INPUT	N	double
IMP_cs_med_home_value	INPUT	N	double
IMP_cs_other	INPUT	N	double
IMP_cs_pct_home_owner	INPUT	N	double
IMP_cs_ttl_female	INPUT	N	double

Name	Role	Type	Variable Type
IMP_cs_ttl_hhlds	INPUT	N	double
IMP_cs_ttl_male	INPUT	N	double
IMP_cs_ttl_mdage	INPUT	N	double
IMP_cs_ttl_pop	INPUT	N	double
IMP_cs_ttl_rural	INPUT	N	double
IMP_cs_ttl_urban	INPUT	N	double
IMP_data_prem_chrgs_curr	INPUT	N	double
IMP_nbr_data_cdrs	INPUT	N	double
IMP_seconds_of_data_log	INPUT	N	double
IMP_tot_drpd_pr1	INPUT	N	double
IMP_tot_overage_chgs	INPUT	N	double
IMP_tot_voice_chrgs_curr	INPUT	N	double
I_churn	CLASSIFICATION	C	char
LOG_MB_Data_Usg_M04	INPUT	N	double
LOG_MB_Data_Usg_M05	INPUT	N	double
LOG_MB_Data_Usg_M06	INPUT	N	double
LOG_MB_Data_Usg_M07	INPUT	N	double
LOG_MB_Data_Usg_M08	INPUT	N	double
LOG_MB_Data_Usg_M09	REJECTED	N	double
P_churn0	PREDICT	N	double
P_churn1	PREDICT	N	double
REP_bill_data_usg	INPUT	N	double

Name	Role	Type	Variable Type
_m03			
REP_bill_data_usg_m06	INPUT	N	double
REP_calls_in_offpk	INPUT	N	double
REP_calls_in_pk	INPUT	N	double
REP_calls_out_offpk	INPUT	N	double
REP_calls_out_pk	INPUT	N	double
REP_calls_total	INPUT	N	double
REP_data_device_age	REJECTED	N	double
REP_lifetime_value	INPUT	N	double
REP_mb_data_ndist_mo6m	REJECTED	N	double
REP_mb_data_usg_m01	INPUT	N	double
REP_mb_data_usg_m02	INPUT	N	double
REP_mb_data_usg_m03	INPUT	N	double
REP_mb_data_usg_roamm01	INPUT	N	double
REP_mb_data_usg_roamm02	INPUT	N	double
REP_mb_data_usg_roamm03	INPUT	N	double
REP_mou_onnet_pct_MOM	REJECTED	N	double
REP_mou_total_pct_MOM	REJECTED	N	double
REP_seconds_of_data_norm	INPUT	N	double

Name	Role	Type	Variable Type
REP_tot_mb_data_curr	INPUT	N	double
REP_tot_mb_data_roam_curr	INPUT	N	double
REP_voice_tot_bill_mou_curr	INPUT	N	double
WARN	ASSESS	C	char

Variable Label	Variable Format	Variable Length	Creator
Predicted for churn		2	tree
Probability for churn=1		8	tree
Probability of Classification		8	tree
Imputed Transformed MB of Data Usage Month 9		8	impute
Imputed Replacement: Avg Age of Devices on Plan	COMMA10.0	8	impute
Imputed Replacement: 6M Avg Billed Data Usage Normally Distributed	BEST12.0	8	impute
Imputed Replacement: Minutes On Network Pct Change Month over Month	PERCENT8.2	8	impute
Imputed Replacement:	PERCENT8.2	8	impute

Variable Label	Variable Format	Variable Length	Creator
Minutes Total Pct Change Month over Month			
Imputed 3M Avg Revenue per User	DOLLAR8.2	8	impute
Imputed 3M Avg Data Charges	DOLLAR8.2	8	impute
Imputed 3M Avg Premium Data Charges	DOLLAR8.2	8	impute
Imputed 3M Avg Overage Charges	DOLLAR8.2	8	impute
Imputed Census Area African-American	BEST8.3	8	impute
Imputed Census Area Caucasian	BEST8.3	8	impute
Imputed Census Area Hispanic	BEST8.3	8	impute
Imputed Census Area Median Home Value Index	BEST4.2	8	impute
Imputed Census Area Other	BEST8.3	8	impute
Imputed Census Area Percent Home Owner	BEST8.3	8	impute
Imputed Census Area Total Female	BEST8.3	8	impute
Imputed Census Area Total Households	COMMA12.0	8	impute
Imputed Census Area Total Males	BEST8.3	8	impute

Variable Label	Variable Format	Variable Length	Creator
Imputed Census Area Median Age	BEST3.0	8	impute
Imputed Census Area Total Population	COMMA12.0	8	impute
Imputed Census Area Total Rural	BEST8.3	8	impute
Imputed Census Area Total Urban	BEST8.3	8	impute
Imputed Premium Data Charges	DOLLAR8.2	8	impute
Imputed Number of Data Records	COMMA10.0	8	impute
Imputed Seconds of Data - Natural Log		8	impute
Imputed Number of Dropped Calls 1 Mth Prior	COMMA8.0	8	impute
Imputed Total Overage Charges	DOLLAR8.2	8	impute
Imputed Total Voice Charges	DOLLAR8.2	8	impute
Into: churn		32	tree
Transformed MB of Data Usage Month 4		8	transform
Transformed MB of Data Usage Month 5		8	transform
Transformed MB of Data Usage Month 6		8	transform
Transformed MB of Data Usage Month		8	transform

Variable Label	Variable Format	Variable Length	Creator
7			
Transformed MB of Data Usage Month 8		8	transform
Transformed MB of Data Usage Month 9		8	transform
Predicted: churn=0		8	tree
Predicted: churn=1		8	tree
Replacement: 3M Avg Billed Data Usage	COMMA8.0	8	replacement
Replacement: 6M Avg Billed Data Usage	COMMA8.0	8	replacement
Replacement: Calls Incoming Off-Peak	COMMA8.0	8	replacement
Replacement: Calls Incoming Peak	COMMA8.0	8	replacement
Replacement: Calls Outgoing Off-Peak	COMMA8.0	8	replacement
Replacement: Calls Outgoing Peak	COMMA8.0	8	replacement
Replacement: Total Calls Curr	COMMA8.0	8	replacement
Replacement: Avg Age of Devices on Plan	COMMA10.0	8	replacement
Replacement: Lifetime Value	DOLLAR8.2	8	replacement
Replacement: 6M Avg Billed Data Usage Normally Distributed	BEST12.0	8	replacement

Variable Label	Variable Format	Variable Length	Creator
Replacement: MB Data Usage 1 Mth Prior	COMMA8.0	8	replacement
Replacement: MB Data Usage 2 Mths Prior	COMMA8.0	8	replacement
Replacement: MB Data Usage 3 Mths Prior	COMMA8.0	8	replacement
Replacement: MB Data Usage Roam 1 Mth Prior	COMMA8.0	8	replacement
Replacement: MB Data Usage Roam 2 Mths Prior	COMMA8.0	8	replacement
Replacement: MB Data Usage Roam 3 Mths Prior	COMMA8.0	8	replacement
Replacement: Minutes On Network Pct Change Month over Month	PERCENT8.2	8	replacement
Replacement: Minutes Total Pct Change Month over Month	PERCENT8.2	8	replacement
Replacement: Seconds of Data - Normalized		8	replacement
Replacement: Total MB of Data Usage	COMMA8.0	8	replacement
Replacement: Total MB of Roam Data Usage	COMMA8.0	8	replacement

Variable Label	Variable Format	Variable Length	Creator
Replacement: Total Voice Billed Minutes of Use	COMMA8.0	8	replacement
Warnings		4	tree

Function	Creator GUID
CLASSIFICATION	5ba566be-ee0a-4065-8502-fa90299311fc
PREDICT	5ba566be-ee0a-4065-8502-fa90299311fc
PREDICT	5ba566be-ee0a-4065-8502-fa90299311fc
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-

Function	Creator GUID
	a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f

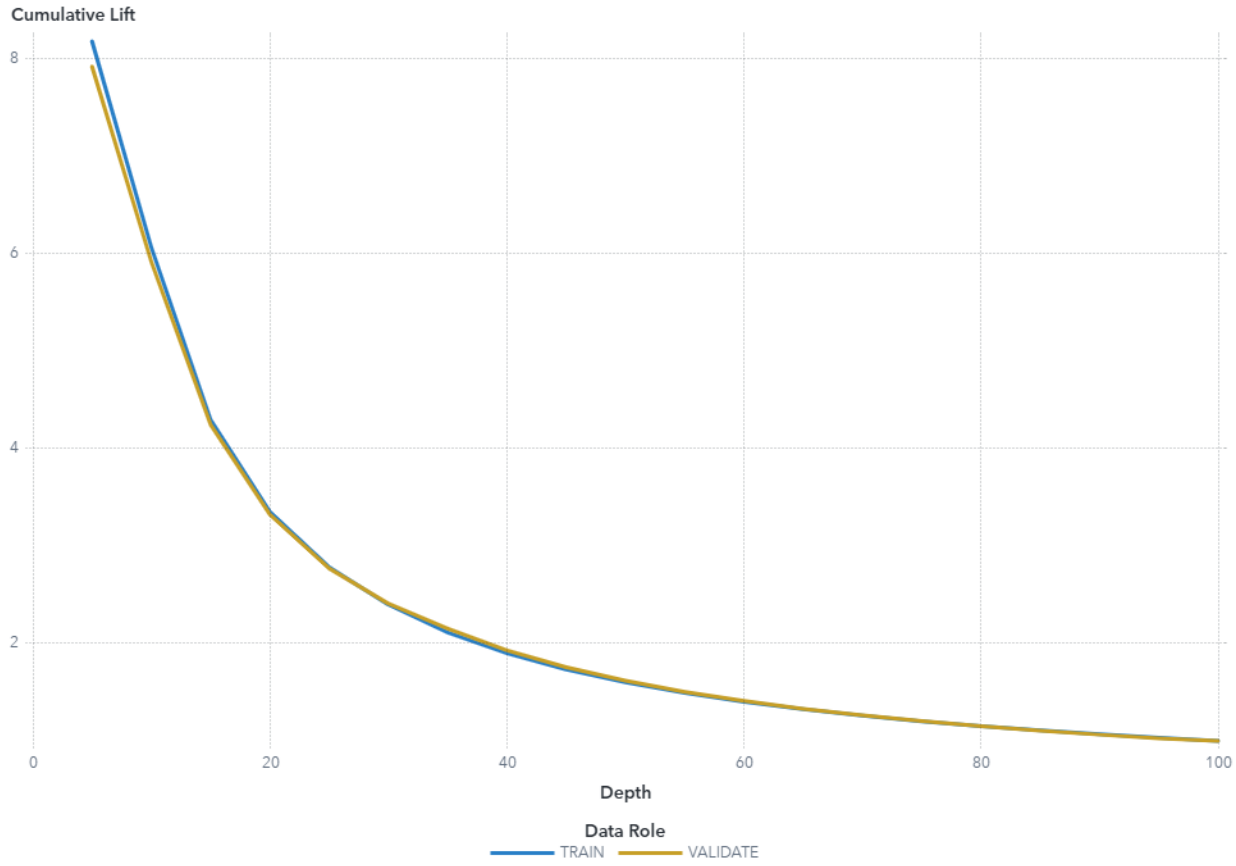
Function	Creator GUID
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
TRANSFORM	907a829f-c39c-4442-9da4-a51e2549996f
CLASSIFICATION	5ba566be-ee0a-4065-8502-fa90299311fc
TRANSFORM	7cd5f313-cfef-45df-abd8-746010ad74d

Function	Creator GUID
	4
TRANSFORM	7cd5f313-cfef-45df-abd8-746010ad74d4
TRANSFORM	7cd5f313-cfef-45df-abd8-746010ad74d4
TRANSFORM	7cd5f313-cfef-45df-abd8-746010ad74d4
TRANSFORM	7cd5f313-cfef-45df-abd8-746010ad74d4
TRANSFORM	7cd5f313-cfef-45df-abd8-746010ad74d4
PREDICT	5ba566be-ee0a-4065-8502-fa90299311fc
PREDICT	5ba566be-ee0a-4065-8502-fa90299311fc
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950

Function	Creator GUID
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a9

Function	Creator GUID
	50
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
TRANSFORM	2c293f10-115e-4978-9f1d-0a2aa860a950
ASSESS	5ba566be-ee0a-4065-8502-fa90299311fc

Cumulative Lift



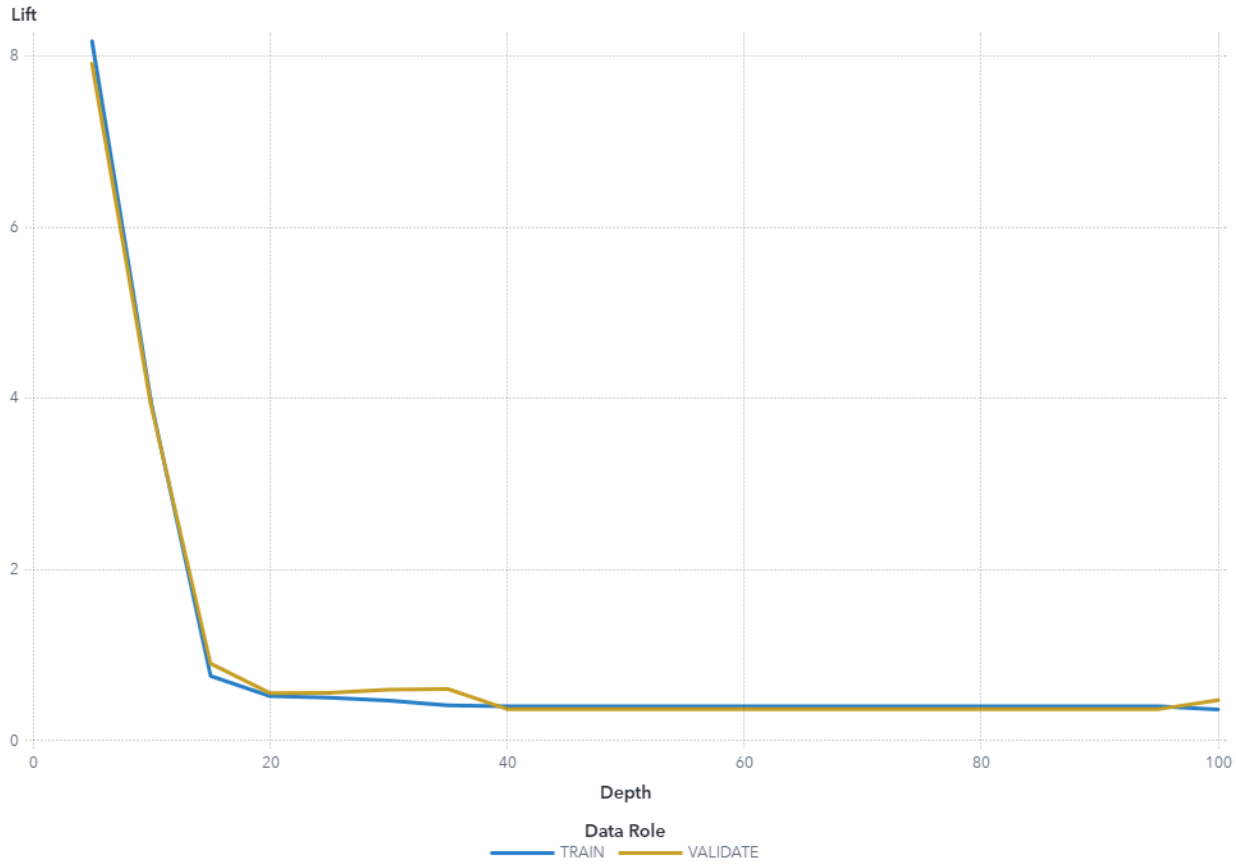
The VALIDATE partition has a Cumulative Lift of 5.91 in the 10% quantile (depth of 10) meaning there are 5.91 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Cumulative Lift of 6.06 in the 10% quantile (depth of 10) meaning there are 6.06 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_churn1, which represents the predicted probability of the event "1" for the target churn. 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



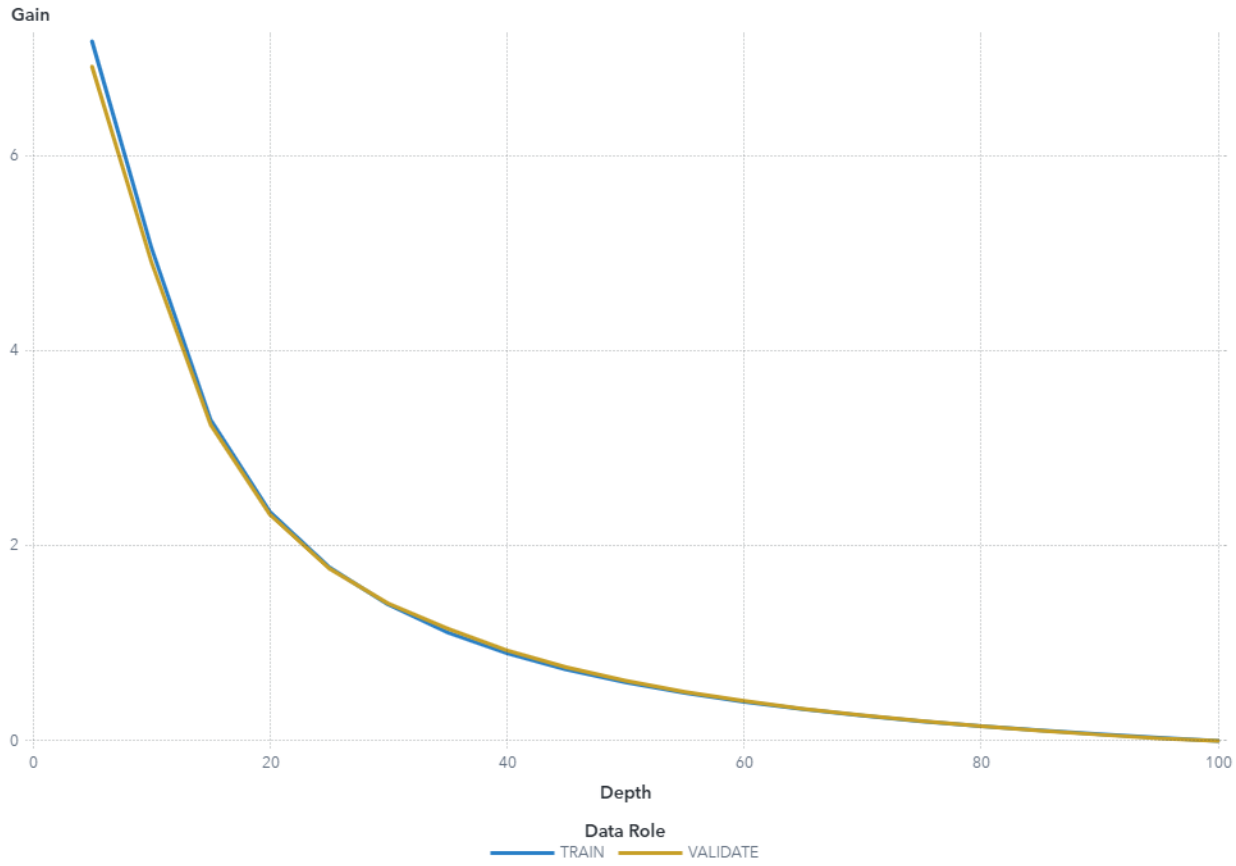
The VALIDATE partition has a Lift of 7.92 in the 5% quantile (depth of 5) meaning there are 7.92 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

The TRAIN partition has a Lift of 8.18 in the 5% quantile (depth of 5) meaning there are 8.18 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_churn1, which represents the predicted probability of the event "1" for the target churn. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Lift is the ratio of the number of events in that quantile to the number of events that would be there at random, or equivalently, the ratio of the response percentage to the baseline response percentage. With 20 quantiles, it is expected that 5% of the events

occur in each quantile. Thus, Lift measures how much more likely it is to observe an event in each quantile than by selecting observations at random.

Gain



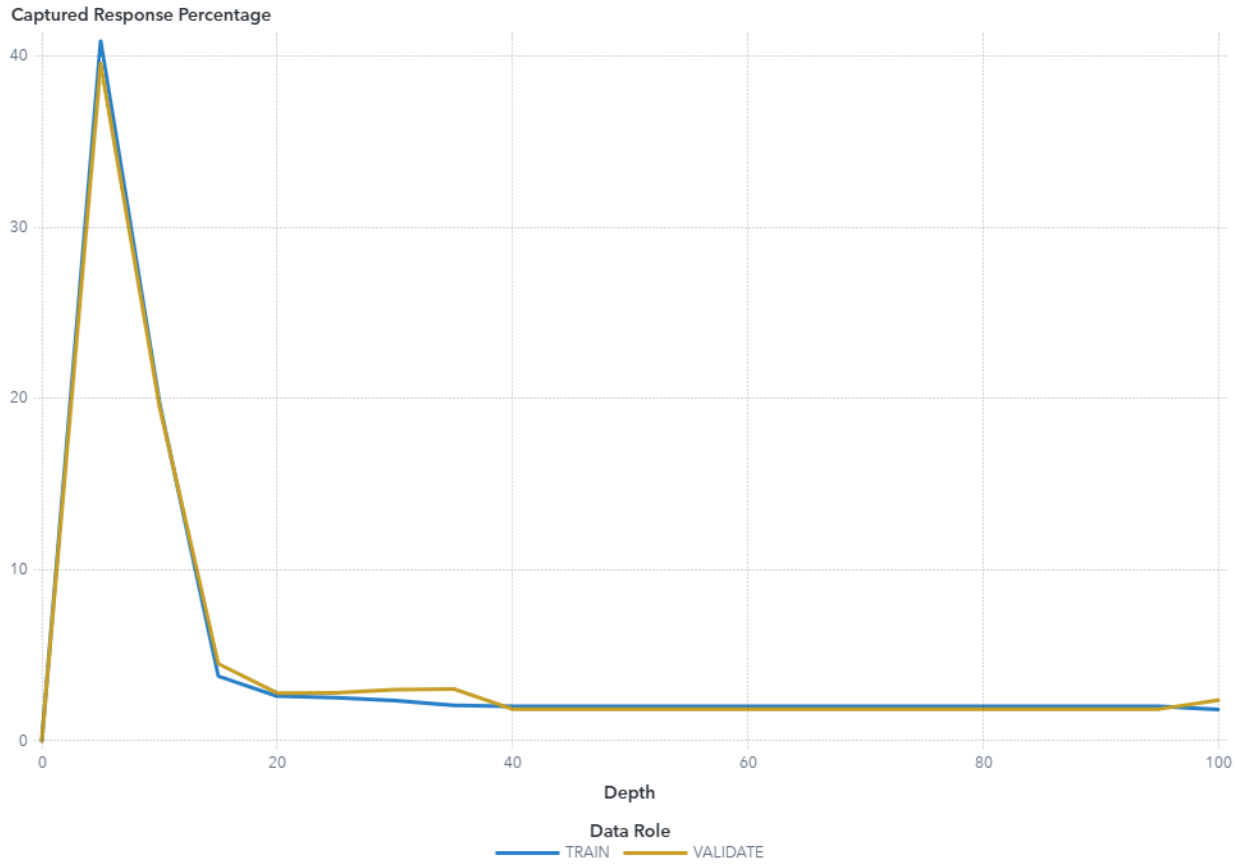
The VALIDATE partition has a Gain of 4.9 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 7.25.

The TRAIN partition has a Gain of 5.1 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 7.24.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_churn1, which represents the predicted probability of the event "1" for the target churn. 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

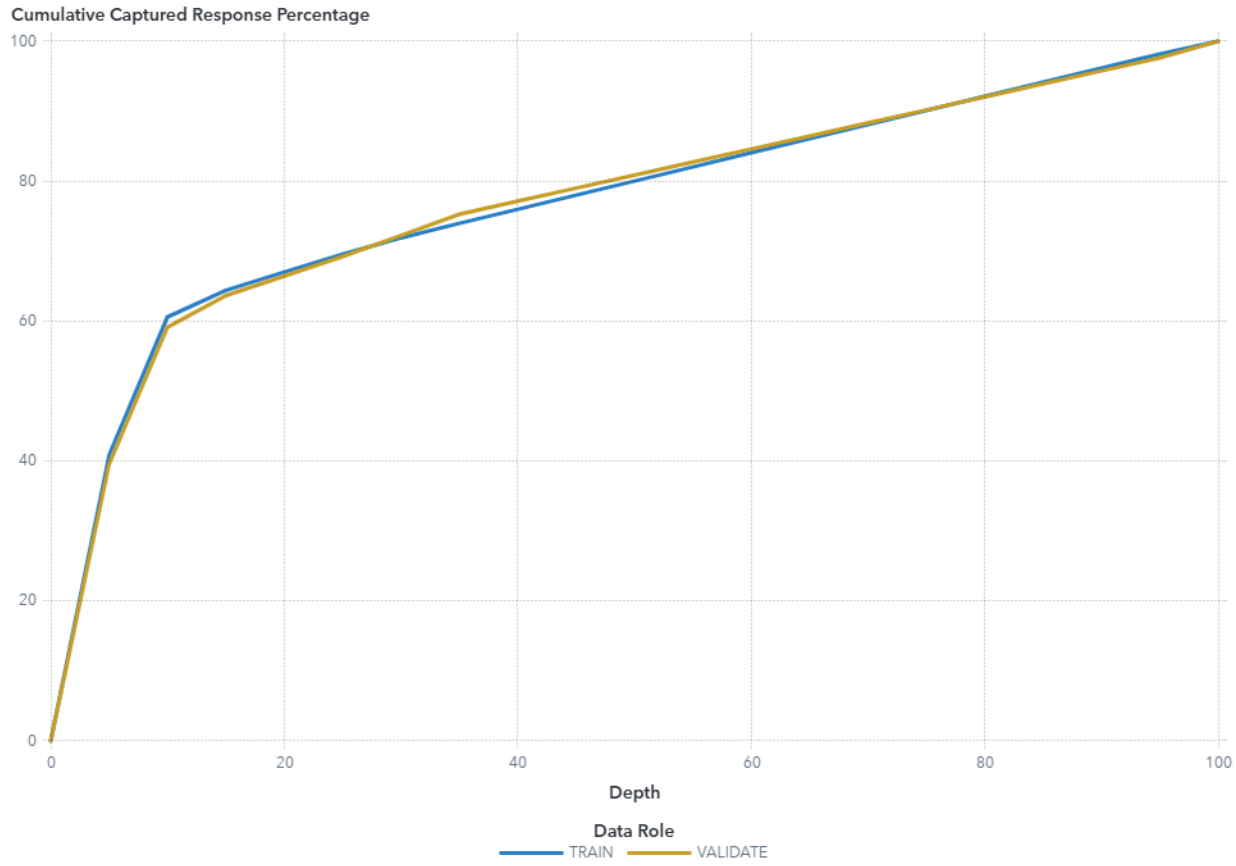


At the 5% quantile (depth of 5), the VALIDATE partition has a Captured response percentage of 39.6 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 41.23.

At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 40.9 (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 41.22.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_churn1, which represents the predicted probability of the event "1" for the target churn. 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



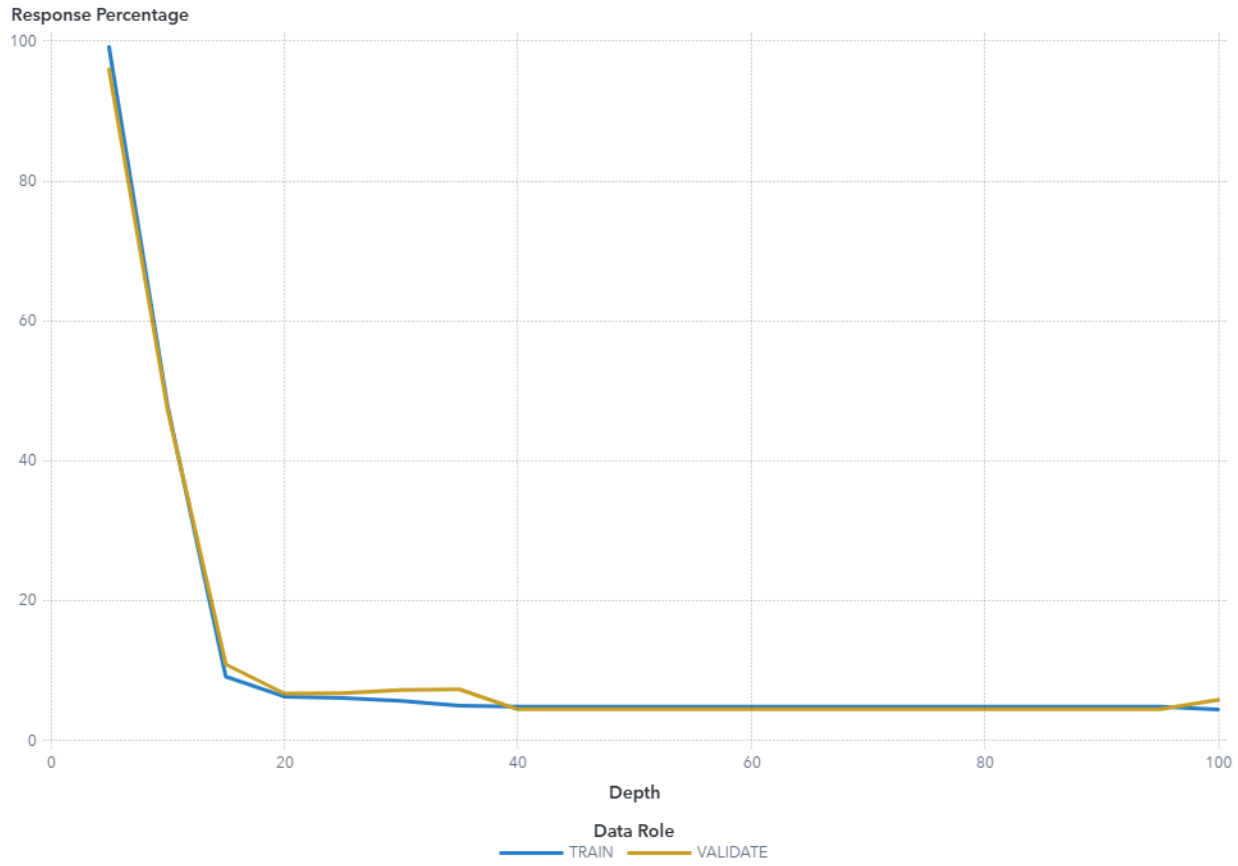
In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative captured response percentage of 59.1 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 82.47.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 60.6 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 82.45.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_churn1, which represents the predicted probability of the event "1" for the target churn. 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

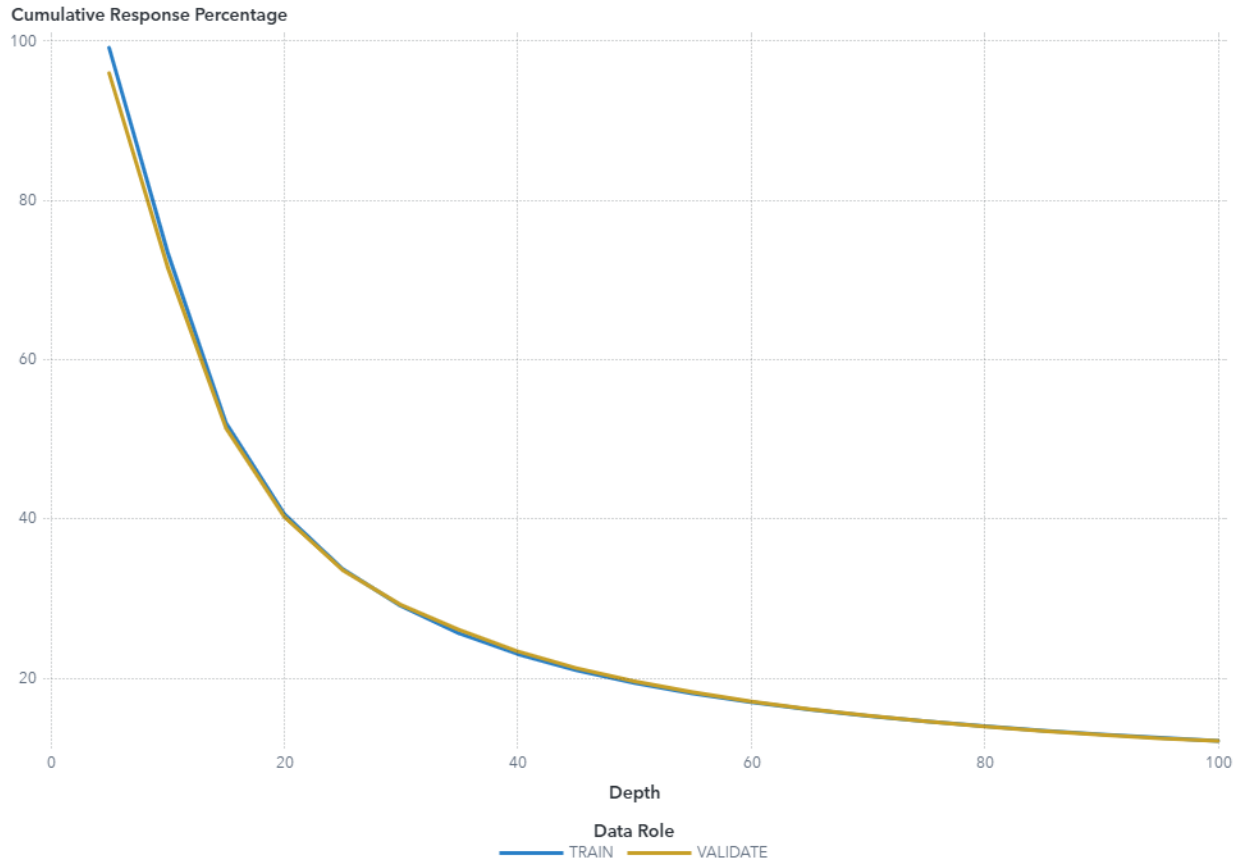


At the 5% quantile (depth of 5), the VALIDATE partition has a Response percentage of 96. The best possible value of Response percentage for this partition at depth 5 is 100.

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_churn1, which represents the predicted probability of the event "1" for the target churn. 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

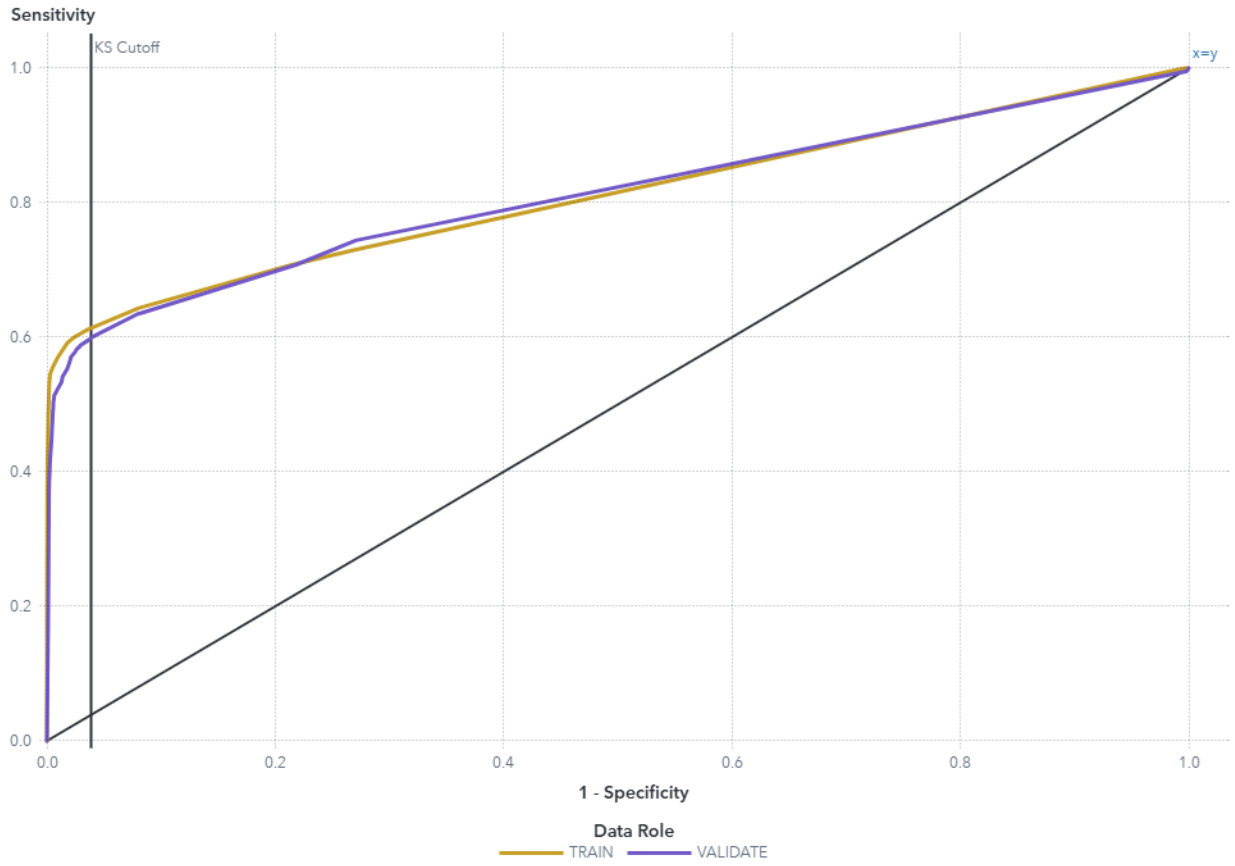


In the top 10% of the data (depth 10), the VALIDATE partition has a Cumulative response percentage of 71.7. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 73.5. 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_churn1, which represents the predicted probability of the event "1" for the target churn. 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 0.09, where the 1-specificity value is 0.039 and the sensitivity value is 0.6.

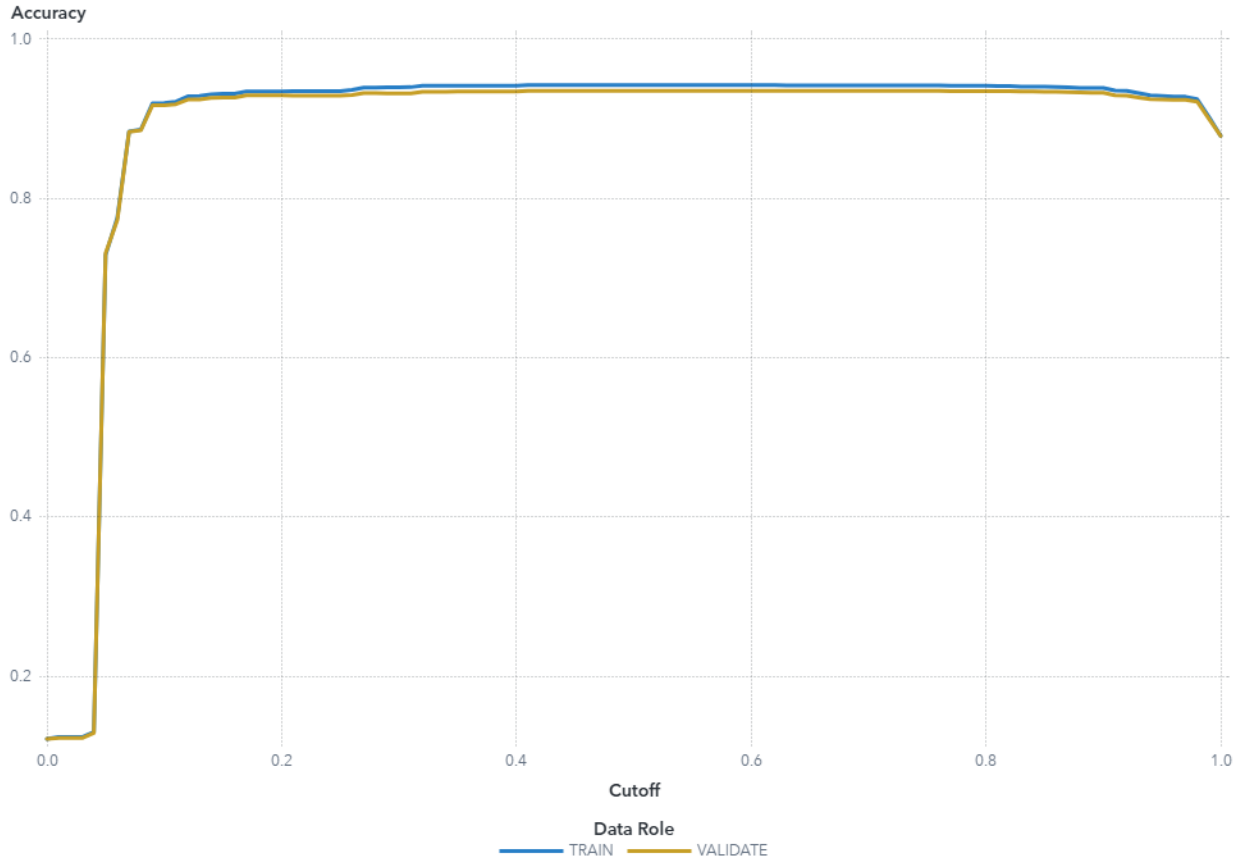
Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_churn1, which is the predicted probability of the event "1" for the target churn, is greater than or equal to the cutoff value. When P_churn1 is greater 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

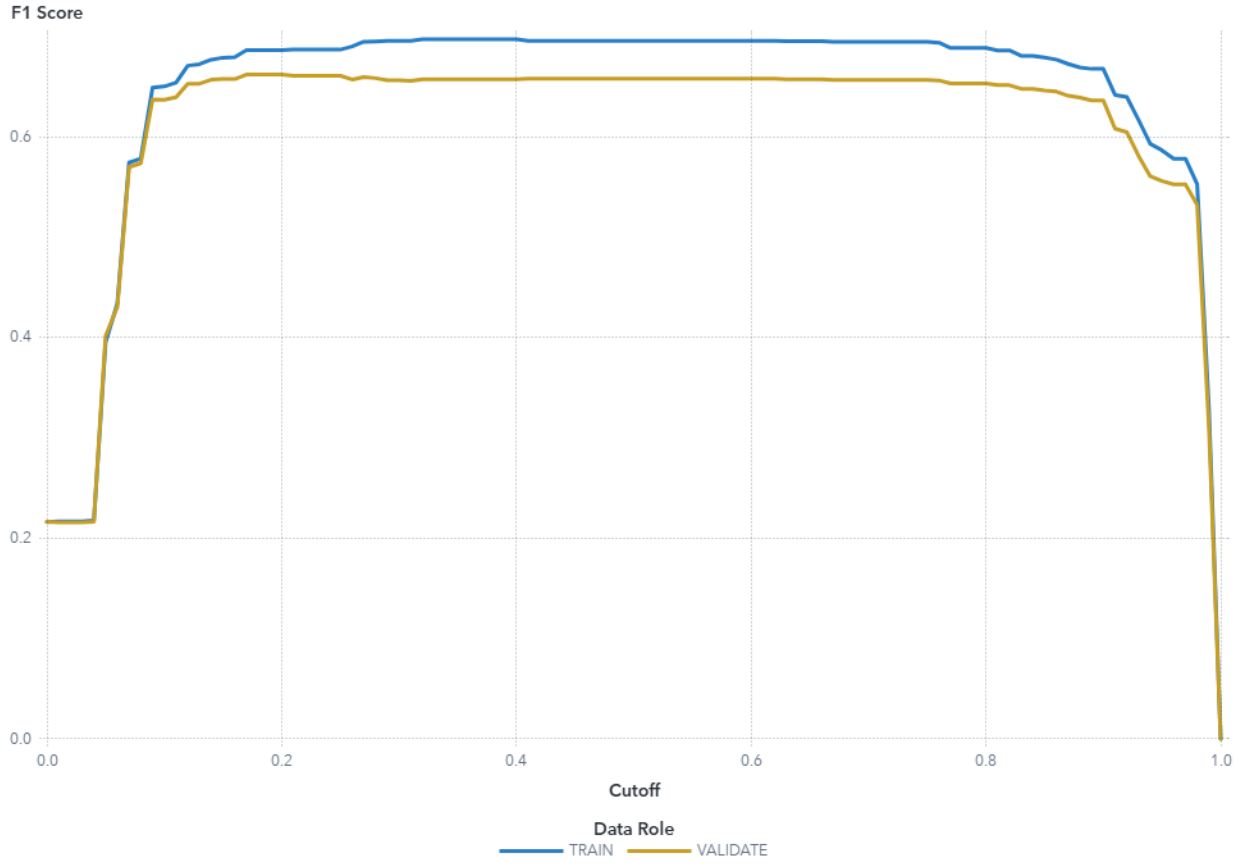


For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.942.

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

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_churn1, which is the predicted probability of the event "1" for the target churn, is greater than or equal to the cutoff value. When P_churn1 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 0.696.

For this model, the F1 score in the VALIDATE partition at the cutoff of 0.5 is 0.658.

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_churn1, which is the predicted probability of the event "1" for the target churn, is greater than or equal to the cutoff value. When P_churn1 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
churn	TRAIN	1	1
churn	VALIDATE	0	0

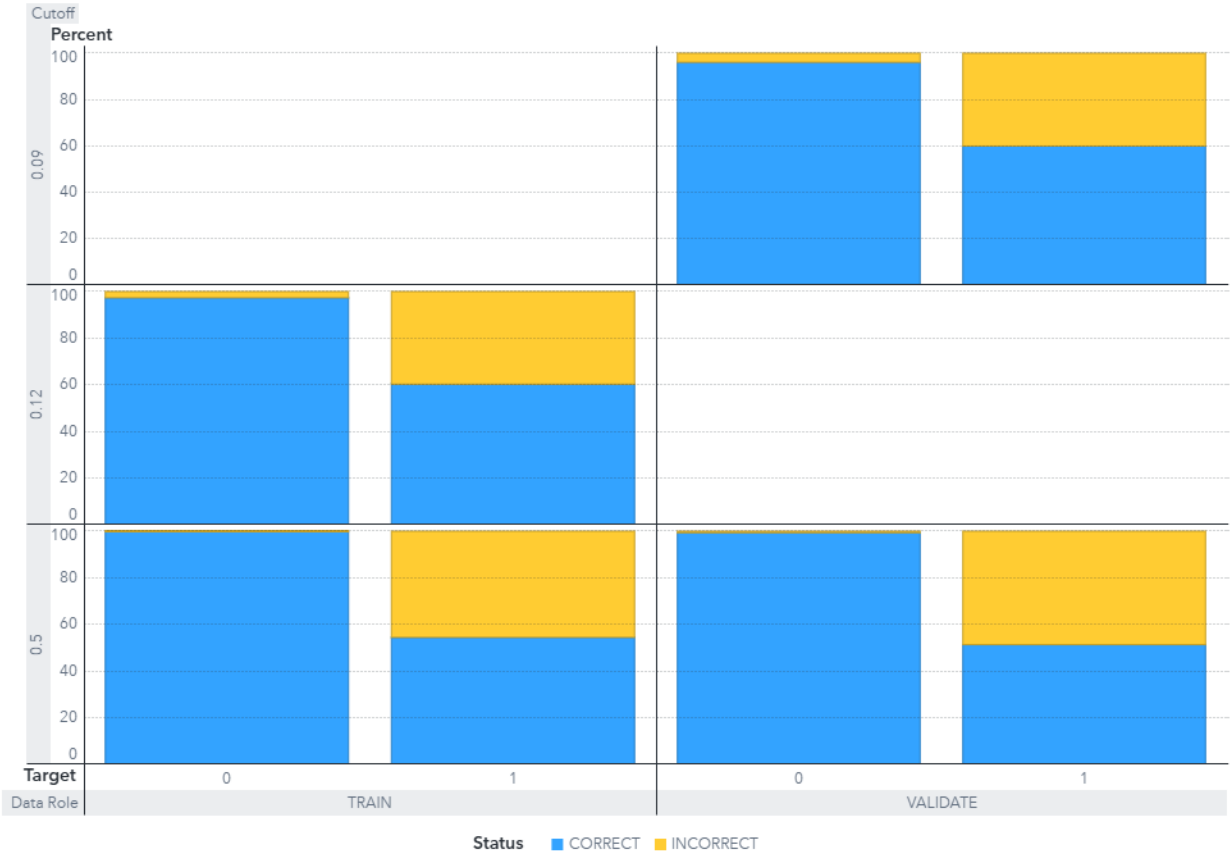
Number of Observations	Average Squared Error	Divisor for ASE	Root Average Squared Error
39,590	0.0527	39,590	0.2296
16,967	0.0584	16,967	0.2417

Misclassification Rate	Multi-Class Log Loss	KS (Youden)	Area Under ROC
0.0577	0.2107	0.5762	0.8112
0.0647	0.2516	0.5608	0.8115

Gini Coefficient	Gamma	Tau	KS Cutoff
0.6223	0.7811	0.1327	0.1200
0.6229	0.7718	0.1328	0.0900

KS at User-Specified Cutoff	Misclassification Rate at KS Cutoff (Event)	Misclassification Rate (Event)
0.5419	0.0717	0.0577
0.5070	0.0828	0.0647

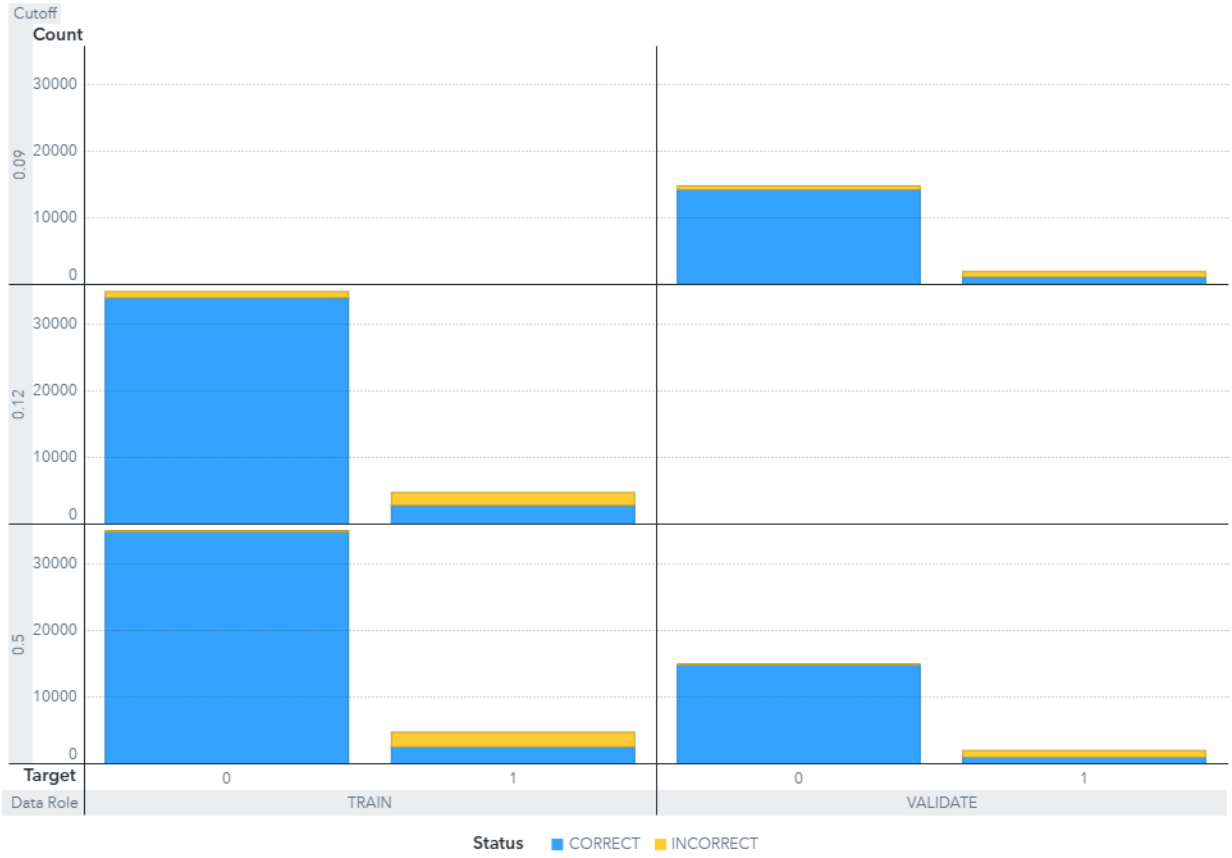
Percentage Plot



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: 0.12 (TRAIN), 0.09 (VALIDATE).

For this data, for the bar corresponding to the event level of churn, " 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: 0.12 (TRAIN), 0.09 (VALIDATE).

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

Table

Cutoff	Cutoff Source	Target Name	Response
0.0900	KS	churn	CORRECT
0.0900	KS	churn	INCORRECT
0.0900	KS	churn	CORRECT
0.0900	KS	churn	INCORRECT
0.1200	KS	churn	CORRECT
0.1200	KS	churn	INCORRECT
0.1200	KS	churn	CORRECT
0.1200	KS	churn	INCORRECT
0.5000	Default	churn	CORRECT
0.5000	Default	churn	INCORRECT
0.5000	Default	churn	CORRECT
0.5000	Default	churn	INCORRECT

Event	Value	Training Frequency	Validation Frequency
1	True Positive		1,235
1	False Negative		824
0	True Negative		14,327
0	False Positive		581
1	True Positive	2,896	
1	False Negative	1,907	
0	True Negative	33,857	
0	False Positive	930	
1	True Positive	2,616	1,057
1	False Negative	2,187	1,002
0	True Negative	34,690	14,813
0	False Positive	97	95

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
		59.9806	
		40.0194	
		96.1028	
		3.8972	
	60.2956		
	39.7044		
	97.3266		
	2.6734		
	54.4660	51.3356	
	45.5340	48.6644	
	99.7212	99.3628	
	0.2788	0.6372	

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	GINI
cvccFolds	10
dataMiningVersion	V2024.03
embeddedBarChart	true
exactPctlLift	true
explainFidelity	false
explainInfo	false
fullDatasetReconstitution	false
hLeafSize	5
iCriterionMethod	VARIANCE

Property Name	Property Value
icePlots	false
inodeColor	AVERAGE
intBinMethod	QUANTILE
intervalBins	50
maxBranch	2
maxCategories	128
maxDepth	10
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
rapidGrowth	false
reportingOnly	false
seRule	false
seed	12,345
seedId	12,345
selMethod	AUTOMATIC
specifyRows	RANDOM
templateRevision	4

Property Name	Property Value
train	true
truncateLl	5
truncateUl	95
useVarOnce	false
userProbCutoff	false

Output

The SAS System

The TREESPLIT Procedure

Model Information	
Split Criterion	Gini
Pruning Method	Cost Complexity
Max Branches per Node	2
Max Tree Depth	10
Tree Depth Before Pruning	10
Tree Depth After Pruning	10
Number of Leaves Before Pruning	243
Number of Leaves After Pruning	122

	Training	Validation	Total
Number of Observations Read	39590	16967	56557
Number of Observations Used	39590	16967	56557

The SAS System

The TREESPLIT Procedure

Fit Statistics for Selected Tree

	Number of Leaves	Misclassification Rate
Training	122	0.0577
Validation	122	0.0647

Variable Importance

	Training		Validation		
Variable	Importance	Relative Importance	Importance	Relative Importance	Count
curr_days_susp	1396.42	1.0000	603.52	1.0000	5
pymts_late_ltd	534.31	0.3826	201.52	0.3339	8
ever_days_over_plan	423.64	0.3034	199.33	0.3303	2
handset_age_grp	383.27	0.2745	169.79	0.2813	5
avg_days_susp	300.85	0.2154	105.34	0.1745	6
times_susp	226.58	0.1623	99.3158	0.1646	5
calls_care_ltd	178.36	0.1277	54.0095	0.0895	14
LOG_MB_Data_Usg_M04	115.63	0.0828	52.3845	0.0868	6
LOG_MB_Data_Usg_M05	80.5528	0.0577	34.3625	0.0569	7
LOG_MB_Data_Usg_M07	84.3520	0.0604	29.9358	0.0496	6
wrk_orders	79.1102	0.0567	24.5786	0.0407	4
REP_seconds_of_data_norm	64.0916	0.0459	17.3260	0.0287	5
delinq_indicator	87.4760	0.0626	13.8462	0.0229	7
LOG_MB_Data_Usg_M08	24.9799	0.0179	11.7489	0.0195	2
IMP_seconds_of_data_log	31.3355	0.0224	10.0639	0.0167	2
ever_times_over_plan	51.8382	0.0371	8.1237	0.0135	7
IMP_LOG_MB_Data_Usg_M09	20.9723	0.0150	7.4606	0.0124	1
LOG_MB_Data_Usg_M06	9.4500	0.0068	4.8488	0.0080	1
REP_tot_mb_data_curr	5.7356	0.0041	3.8970	0.0065	1
days_openwrkorders	11.4754	0.0082	3.1507	0.0052	2
call_category_1	15.2126	0.0109	1.7638	0.0029	3
REP_mb_data_usg_roamm02	12.1477	0.0087	1.6990	0.0028	1
IMP_REP_mou_onnet_pct_MOM	35.4546	0.0254	1.4837	0.0025	3
REP_mb_data_usg_m03	4.7619	0.0034	1.2698	0.0021	1
unsolv_tsupcomplnt	9.5466	0.0068	0.7425	0.0012	2
bill_data_usg_m09	8.8088	0.0063	0.7246	0.0012	1
REP_mb_data_usg_roamm03	1.6563	0.0012	0.3623	0.0006	1
REP_bill_data_usg_m06	4.7376	0.0034	0.2936	0.0005	1
IMP_REP_mb_data_ndist_mo6m	3.7556	0.0027	-0.2664	-44E-5	1
REP_calls_in_offpk	6.7493	0.0048	-0.4932	-82E-5	1
num_tsupcomplnts	3.2000	0.0023	-0.9600	-0.002	1
IMP_REP_data_device_age	4.7647	0.0034	-1.6129	-0.003	1
IMP_cs_ttl_rural	5.6343	0.0040	-2.1618	-0.004	1
billing_cycle	10.3700	0.0074	-2.7995	-0.005	2
region	3.2000	0.0023	-3.2000	-0.005	1
IMP_nbr_data_cdrs	4.7417	0.0034	-3.2482	-0.005	1
IMP_cs_afr_amer	9.7074	0.0070	-3.6656	-0.006	2
REP_mb_data_usg_m02	10.0204	0.0072	-9.0654	-0.015	1

Cost Complexity Pruning

	Number	Misclassification Rate	
Alpha	of Leaves	Training	Validation
359E-21	147	0.05709	0.06566
8.42E-6	141	0.05714	0.06566
0.00001	137	0.05719	0.06548
0.00003	133	0.05729	0.06507
0.00003	130	0.05739	0.06489
0.00004	122	0.05769	0.06465
0.00005	115	0.05804	0.06477
0.00006	103	0.05880	0.06489
0.00007	99	0.05908	0.06507
0.00008	93	0.05954	0.06501
0.00009	87	0.06007	0.06507
0.00010	80	0.06077	0.06601
0.00011	76	0.06123	0.06607
0.00012	73	0.06158	0.06613
0.00013	65	0.06259	0.06607
0.00014	63	0.06287	0.06642
0.00015	53	0.06438	0.06748
0.00016	51	0.06471	0.06760
0.00020	50	0.06492	0.06754
0.00023	48	0.06537	0.06748
0.00024	46	0.06585	0.06784
0.00025	42	0.06684	0.06819
0.00025	41	0.06709	0.06819
0.00028	40	0.06737	0.06843
0.00029	38	0.06795	0.06860
0.00030	34	0.06916	0.06943
0.00032	32	0.06979	0.07014
0.00033	28	0.07110	0.07120
0.00038	27	0.07148	0.07137
0.00039	22	0.07343	0.07373
0.00053	20	0.07449	0.07450
0.00058	19	0.07507	0.07520
0.00067	16	0.07706	0.07574
0.00076	15	0.07782	0.07644
0.00077	13	0.07936	0.07792
0.00106	11	0.08149	0.08104
0.00114	10	0.08262	0.08198
0.00134	8	0.08530	0.08475
0.00183	5	0.09078	0.09006
0.00293	3	0.09664	0.09737
0.00730	2	0.1039	0.1030
0.01738	1	0.1213	0.1214

The SAS System

The TREESPLIT Procedure

Predicted Probability Variables	
churn	Variable
1	P_churn1
0	P_churn0

Predicted Target Variable	
Level Index	Variable
	I_churn