Governance Review

Model

September, 2020

Table of Contents

Objective & Campaign Information

Smarter Marketing

Objective: To refit the model XXX by using XXX Credit and Census data sources plus the Infobase and house file data.

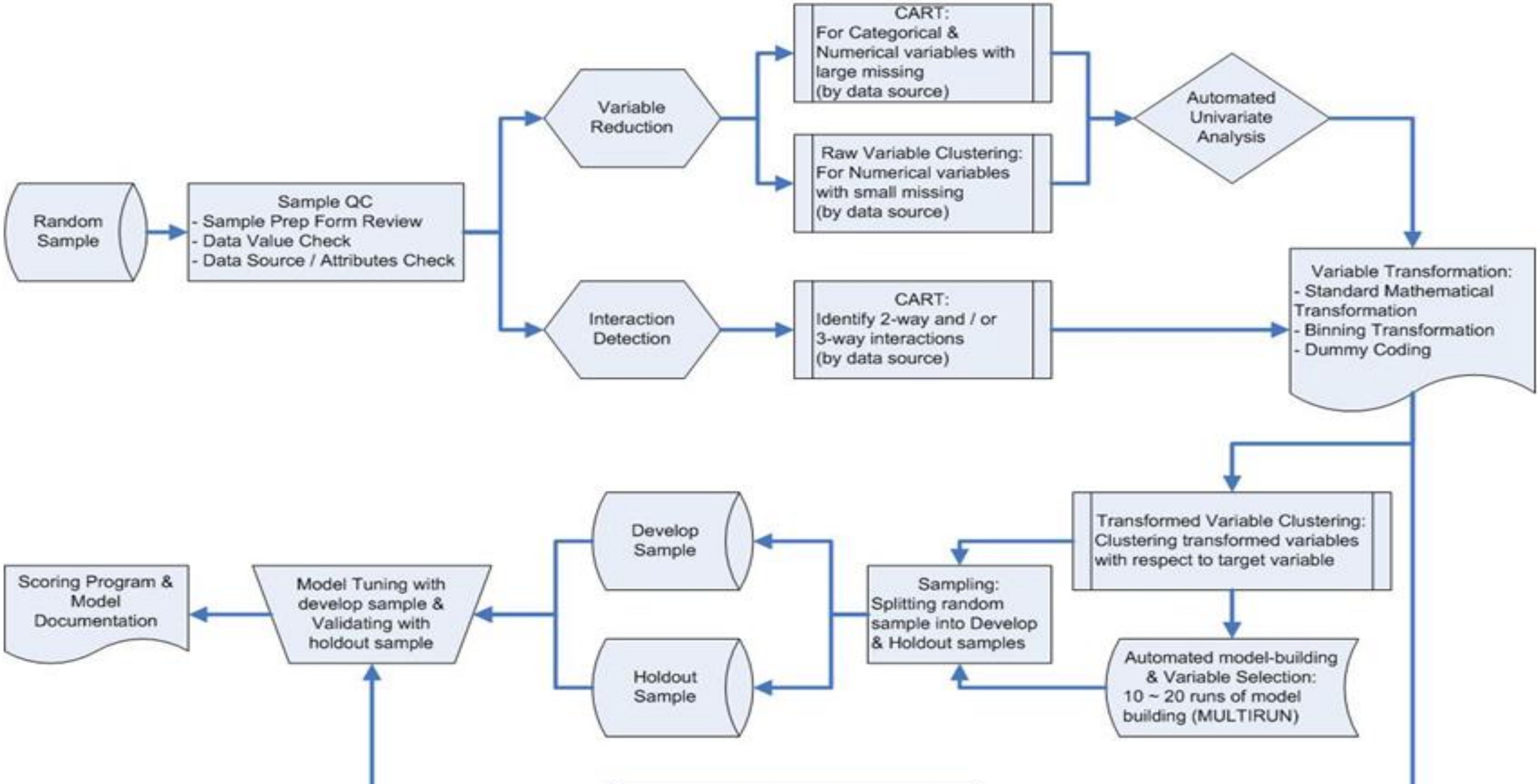
DSN: /u02/staging/temp/d042_spmj2736_mrgal_p1_prm (Dec'19 mailed) plus u02/staging/temp/c609_spmj2658_mrgal_p1_prm (Jun '19 mailed). u02/staging/temp/d296_spmj2743_mrgal_p1_prm (Jun' 20 mailed) was used for validation

Modeling Data Files	Total Observations	# of Responders	Response Rate	Total Premium Amount	PPPM
Whole Sample (Development)	209,583	733	.35%	$190,638	$0.91
Validation Sample	161,272	506	.31%	$137,010	$0.85

Data Sources Appended:

- House File & Promotion history
- Infobase
- Census
- Aggregate Credit
- Geographic

Fi


Random Sample
Sample QC
- Sample Prep Form Review
- Data Value Check
- Data Source / Attributes Check
Variable Reduction
Interaction Detection
CART:
For Categorical &
Numerical variables with
large missing
(by data source)
Raw Variable Clustering:
For Numerical variables
with small missing
(by data source)
CART:
Identify 2-way and / or
3-way interactions
(by data source)
Automated Univariate Analysis
Variable Transformation:
- Standard Mathematical Transformation
- Binning Transformation
- Dummy Coding
Transformed Variable Clustering:
Clustering transformed variables
with respect to target variable
Sampling:
Splitting random
sample into Develop
& Holdout samples
Automated model-building
& Variable Selection:
10 ~ 20 runs of model
building (MULTIRUN)
Develop Sample
Holdout Sample
Model Tuning with
develop sample &
Validating with
holdout sample
Scoring Program &
Model
Documentation

Final Model Variables

Smarter Marketing

MODEL	Housefile	InfoBase (XX Cluster)	Census	infobase (XX Cluster)	Client	Total
# of Variables	8	3	1	6	1	19
Total Weights	48.4%	9.1%	2.2%	24.3%	16%	100%

Model

Model: type II Tobit

$$y_{2i} = \begin{cases} y_{2i}^* & \text{if } y_{1i}^* > 0, \\ 0 & \text{if } y_{1i}^* \leq 0. \end{cases}$$

- Variable selection method:
- EDA (e.g. calculate correlation and examining data)
- Factor analysis: groups together predictors that are highly correlated with each other to create a composite variable as
- a weighted sum of those variables.
- l1 regularized regression (using response as target)
- Stepwise :selects the subset of the variables that maximizes adjusted R2.

Variable Screening with Clustering

For numerical variables with a small percent of missing, one may choose to use the clustering method to do the preliminary variable screening

- The end of the SAS output provides the list of variables which have the highest correlation with the Response variable from each cluster.

51 Clusters		R-squared with		1-R**2	Variable
Cluster	Variable	Own Cluster	Next Closest	Ratio	Label
Cluster 01	VN_HICDL	0.8751	0.5181	0.2591	High Credit for All Credit Type Accts >= 227296
	VN_HICML	0.81	0.4669	0.3564	High Credit for All Mortgage Type Accts >= 134632
	VN_HICNL	0.5748	0.2822	0.5923	High Credit for Non-Mortgage Loan Accts >= 84917
	ZINT3__L	0.5344	0.2014	0.583	
Cluster 02	VNHICRDH	0.788	0.2737	0.2919	High Credit for All Credit Type Accts <= 37899
	VNHICRMH	0.6092	0.3963	0.6474	High Credit for All Mortgage Type Accts <= 13589
	VNHICRNC	0.7656	0.4727	0.4445	LOG High Credit for Non-Mortgage Loan Accts
	VNHICRNH	0.7491	0.2756	0.3463	High Credit for Non-Mortgage Loan Accts <= 30894
	VN_HICNH	0.7434	0.3371	0.3871	High Credit for Non-Mortgage Loan Accts <= 42420

From SAS 9.4 documentation:
"The VARCLUS procedure divides a set of numeric variables into disjoint or hierarchical clusters. Associated with each cluster is a linear combination of the variables in the cluster. **This linear combination can be either the first principal component (the default)** or the centroid component (if you specify the CENTROID option). The first principal component is a weighted average of the variables that explains as much variance as possible. See Chapter 79: The PRINCOMP Procedure, for further details. Centroid components are unweighted averages of either the standardized variables (the default) or the raw variables (if you specify the COVARIANCE option). **PROC VARCLUS tries to maximize the variance that is explained by the cluster components, summed over all the clusters.**
The cluster components are oblique, not orthogonal, even when the cluster components are first principal components. In an ordinary principal component analysis, all components are computed from the same variables, and the first principal component is orthogonal to the second principal component and to every other principal component. In PROC VARCLUS, each cluster component is computed from a set of variables that is different from all the other cluster components. The first principal component of one cluster might be correlated with the first principal component of another cluster. Hence, the PROC VARCLUS algorithm is a type of oblique component analysis"

problem.

The general SAS code for performing a cluster analysis is:

```
PROC CLUSTER <options>;
      VAR var1 var2 var3 ... var n;
```

Here the options control the printing, computational, and output of the procedures. Some examples are:

NOPRINT	- suppresses any printed output,
NOEIGEN	- suppresses printing of eigenvalues,
SIMPLE	- produces simple summary statistics for each variable,
METHOD =	- controls the clustering method used (required option),
STANDARD	- Uses the correlation matrix for computation, and
OUTTREE =	- create an output dataset for cluster diagrams.

The VAR statement, as before, lists the variables to be considered as responses.

For the flour example, the SAS program would be:

```
PROC CLUSTER METHOD = AVERAGE OUTTREE = TREE;
      VAR PEAK_VISC TROUGH_VISC FINAL_VISC BREAKDOWN
            TOTAL_SETBACK TIMEPEAK_VISC;
```

The method selected in this example is the AVERAGE which bases clustering decisions on the average distance (linkage) between points or clusters. Some other possibilities include CENTROID which uses the distance between the geometric centers of the clusters, MEDIAN which is similar to average, but based on median values, and SIMPLE which uses a nearest neighbor approach. The computed clusters will be saved in a dataset calledTREE for plotting purposes.

Smarter Marketing

Obs	VAR	NEWCLUST	ABSRESP	AMT_PRM
59	QN_PQ___	1	0.0066	0.0066
60	QB_PYD___	1	0.0076	0.0076
61	QD_PQ__H	1	0.0123	0.0123
62	QD_PQ__G	1	0.0126	0.0126
63	B_VINS09	2	0.001	-0.001
64	AX_AUUTIL	3	0.006	0.006
65	AX_PTRADESAT12	4	0.0131	-0.0131
66	AX_PTRADESAT03	4	0.0137	-0.0137
67	AX_PTRADESAT06	4	0.0145	-0.0145
68	AX_PTRADESAT24	4	0.0147	-0.0147
69	AC_PCT_2535K	5	0.0017	0.0017
70	AC_PCT_FAM_2535K	5	0.0018	0.0018
71	AC_PCT_2530K	5	0.0024	-0.0024
72	AC_PCT_FAM_2530K	5	0.0031	-0.0031
73	B_RHVALQ	6	0.0131	0.0131
74	AX_AUH59X	7	0.0094	0.0094
75	AX_AUH79X	7	0.0096	0.0096
76	AX_AUH39X	7	0.0096	0.0096
77	AX_AUC29X	7	0.0096	0.0096
78	AX_AU7924	7	0.0107	0.0107
79	AC_BOATRVVAN	8	0.0018	-0.0018
80	QN___D_4	9	0.0055	-0.0055
81	QN_____4	9	0.0055	-0.0055
82	QB___D___	10	0.0035	0.0035
83	QBS___D__	10	0.0035	0.0035
84	AMT_PRM	11	1	1

Univariate Index Calculation

Collect all variables from the screening process. Conduct the univariate index calculation by variable type

Nominal Categorical variables (%CATVAR)

- Geographic (STATE, ZIP1, ZIP2)
- Infobase

Ordinal Categorical variables (%SORTVAR)

- InfoBase variables (B_RHMVAL, B_FNEIHR etc.)

- **Numeric variables with large percent of missing (%NUMVAR)**

  membership variables (EN_, ED_, EC_, EA_, EH_)

  - promotion variables (QN_, QD_)
  - InfoBase variables (B_AGE2YR, etc.)

- **Numeric variables with small percent of missing (%TRANVAR)**
  - Census variables (Z_)
  - Credit variables (AF_)

Variable Transformation

- **Variable Lift Index is used as a guideline for variable transformation. Four types of variable transformations are performed in the indexing calculation:**

  - Standard mathematical transformations (Log, Square, Square Root, Reciprocal)*
  - Binning (ranking) transformation
  - Dummy coding
  - Interaction variable coding (Use JMP)

* To avoid mathematical errors, we define the continuous form of the variables as :

LOG = Log (1+ Raw variable)

Square = Raw variable * Raw variable

Square Root= SQRT(Raw variable)

Reciprocal= 1/(1+Raw variable)

Variable Transformation (cont.)

- **Continuous variable transformation**
  - Most appropriate when the lift index shows a monotonic trend from high to low or low to high. Candidates for continuous transformation are defined in the macro %TRANVAR. The choice for missing value imputation and replacement using the sample mean, median, min or max depends on the lift index trend from missing values.

- **Index or binning variable transformation**
  - Most appropriate when the lift index shows a general trend from high to low or low to high but the trend is not smooth. The missing values could be treated as a separate bin or combined with other values into one bin.

Variable Transformation (cont.)

- **Dummy variable transformation**
  - Most appropriate when the lift index shows a high or low response trend but neither a continuous nor binning transformation is suitable. The missing values could be included or excluded in the dummy transformation depending on the lift index of the missing value.
  - Dummy Coding Rule: A general rule of thumb in defining High or Low dummy variable:
    - Lift index >= 125 & Lower 95% index >= 100 for high dummy
    - Lift index <= 75 & Upper 95% index <= 100 for low dummy
  - For robustness, transformed dummy variables should contain at least 5% of the sample count
  - For a categorical variable, category values with at least 1% of the sample count could be combined with other categories with similar lift trend to form a dummy variable
  - In general, creating dummy variables in the middle interval values from ordinal or numeric variables are not recommended. There are some exceptions on age and income related variables, however.
- **The preference for variable transformation is**
  1. **Continuous transformation**
  2. **Binning transformation**
  3. **Dummy transformation**

VAR=B_INSU Obtain Insurance (Financial) Rank: Values of 01 (Most Likely) through 20 (Least Likely)

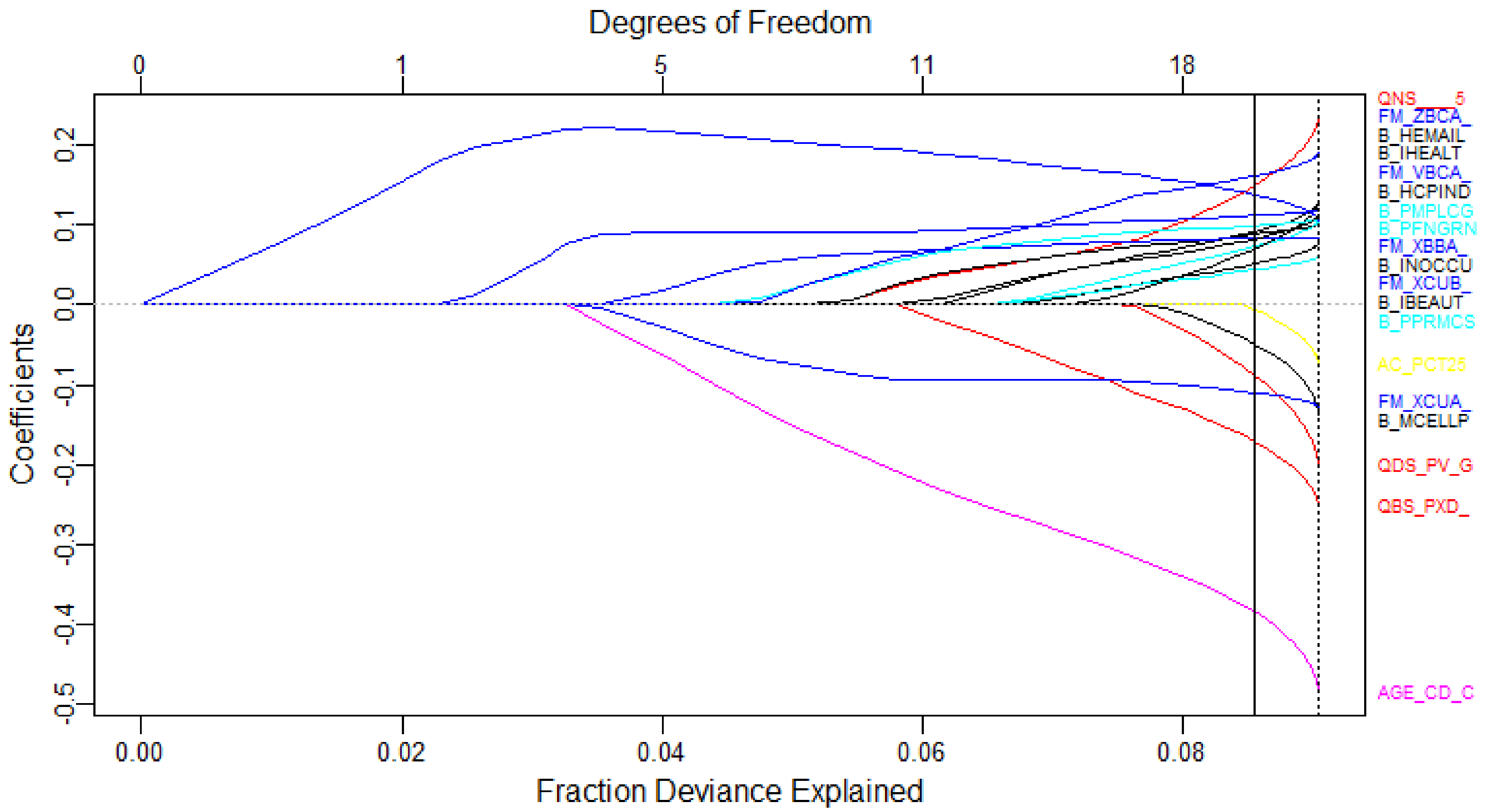
VALUE	NUMBER OF OBS	% OBS	TOTAL OF PRM_AMT	% PRM_AMT	PRM_AMT MEAN	UPPER 95% INDEX	LOWER 95% INDEX	INDEX OF PRM_AMT MEAN	NUMBER OF RESPONSE	% RESPONSE	RESPONSE RATE X100	UPPER 95% INDEX	LOWER 95% INDEX	INDEX OF RESPONSE RATE
	15823	12.57	16830	9.76	1.06	102	53	78	53	10.08	0.33	102	59	80
1	272	0.22	594	0.34	2.18	472	-153	159	1	0.19	0.37	260	-84	88
2	1000	0.79	1386	0.8	1.39	211	-8	101	4	0.76	0.4	189	2	96
3	1907	1.51	1848	1.07	0.97	139	3	71	5	0.95	0.26	118	8	63
4	3176	2.52	4422	2.56	1.39	161	42	102	14	2.66	0.44	161	50	106
5	4189	3.33	3399	1.97	0.81	99	19	59	11	2.09	0.26	100	26	63
6	5241	4.16	6831	3.96	1.3	142	48	95	19	3.61	0.36	126	48	87
7	6198	4.92	7656	4.44	1.24	134	47	90	22	4.18	0.35	120	50	85
8	7214	5.73	6633	3.85	0.92	100	34	67	21	3.99	0.29	99	40	70
9	7520	5.97	8943	5.19	1.19	124	50	87	31	5.89	0.41	133	64	99
10	7530	5.98	9009	5.22	1.2	125	49	87	30	5.7	0.4	129	61	95
11	8136	6.46	14784	8.57	1.82	176	90	133	49	9.32	0.6	184	104	144
12	7409	5.89	9669	5.61	1.31	135	56	95	30	5.7	0.4	132	62	97
13	6340	5.04	10527	6.1	1.66	172	70	121	27	5.13	0.43	140	64	102
14	6555	5.21	7379	4.28	1.13	122	42	82	24	4.56	0.37	123	53	88
15	6284	4.99	11583	6.72	1.84	188	81	135	31	5.89	0.49	160	77	118
16	5688	4.52	7161	4.15	1.26	136	48	92	23	4.37	0.4	136	57	97
17	5361	4.26	10131	5.87	1.89	195	81	138	32	6.08	0.6	192	94	143
18	6483	5.15	11121	6.45	1.72	174	76	125	32	6.08	0.49	159	77	118
19	7608	6.04	13926	8.08	1.83	181	86	134	39	7.41	0.51	161	84	123
20	5957	4.73	8613	4.99	1.45	152	59	106	28	5.32	0.47	154	71	112
VAR	125891	100	172445	100					526	100				

Transformation of Numeric Variables

VARNAME: D_AGE												
Rank	Min	Max	# of Obs	% Obs	# of Response	% Response	Cum # of Obs	Cum # of Response	Response Rate	Upper 95% Index	Lower 95% Index	Index of Response Rate
.	.	.	54656	15.73	120	17.47	54656	120	0.22	131	91	111
0	10	37	30436	8.76	22	3.20	85092	142	0.07	52	21	37
1	38	41	27737	7.98	25	3.64	112829	167	0.09	63	28	46
2	42	44	26190	7.54	21	3.06	139019	188	0.08	58	23	41
3	45	47	28255	8.13	45	6.55	167274	233	0.16	104	57	81
4	48	50	29034	8.35	57	8.30	196308	290	0.20	125	74	99
5	51	54	37677	10.84	83	12.08	233985	373	0.22	135	87	111
6	55	57	25671	7.39	68	9.90	259656	441	0.26	166	102	134
7	58	61	28386	8.17	75	10.92	288042	516	0.26	164	103	134
8	62	67	30779	8.86	81	11.79	318821	597	0.26	162	104	133
9	68	99	28721	8.26	90	13.10	347542	687	0.31	191	126	159
Total			347542	100.00	687	100.00						

- Variable Type: A numeric variable with high percent of missing values
- Continuous Transformation? NO
  - Lift index distribution is not very smooth and high percent of missing values
- Binning Transformation? YES
  - The index distribution has a good increasing trend. Suggested bins are color coded.
- Dummy Transformation? YES
  - ZD_AGE_L = (0 <= D_AGE <= 44); Note: Ranks 0-2 meet the low dummy coding rule
    Note: Even though the variable starts with a value 10, we will use 0 as the lower bound
  - ZD_AGE_H = (D_AGE >= 55); Note: Ranks 6-9 meet the high dummy coding rule
  - Even though 68+ group is more responsive in the sample, we may not want to code it as a separate dummy variable since we may not have enough names to select from in future campaigns.

Variable reduction



2.5 Interaction Variable Detection

CART provides an easy way to derive interaction dummy variables. Interaction variables capture multi-dimension information from the raw variables which could help to improve model performance. Here are the basic guidelines:

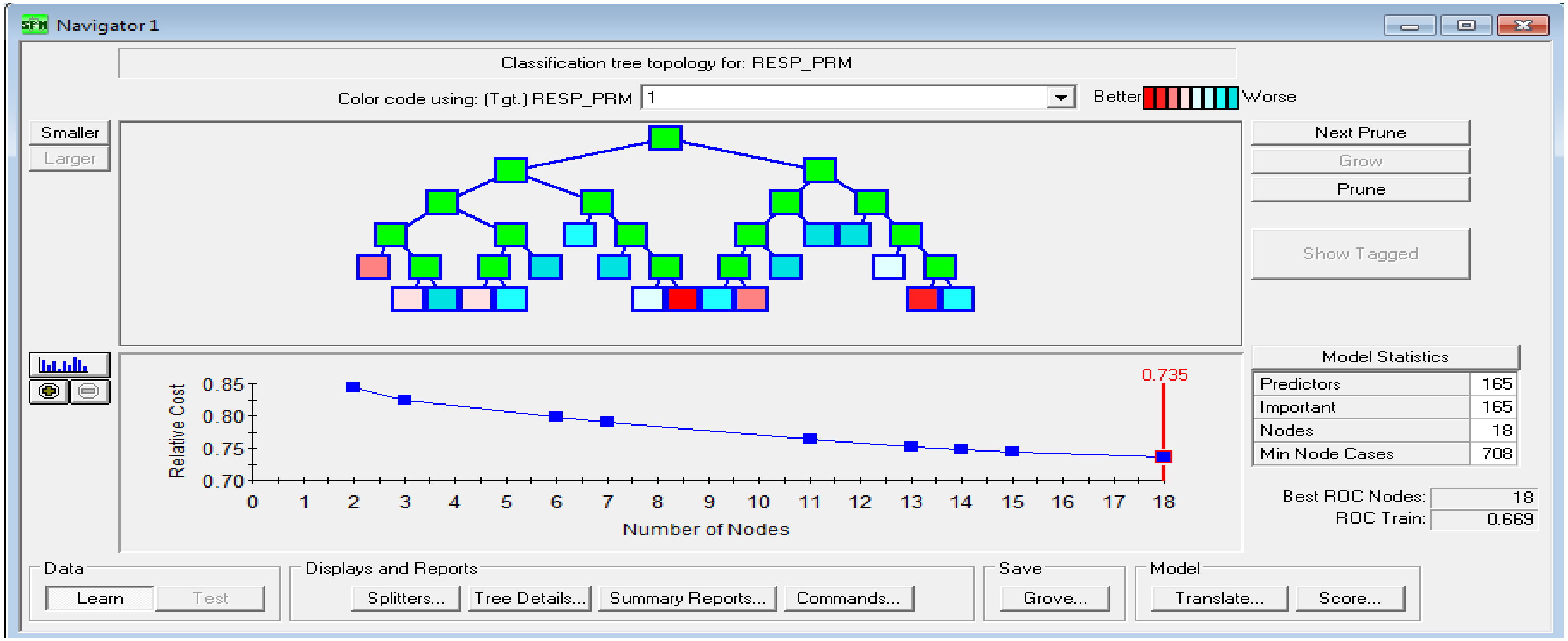
- Derive two-way interaction dummy variables (3-way interaction is acceptable but could be less robust and more difficult to interpret)
- The interaction dummy should contains greater than 10% of the sample count (combining interaction terms if necessary)
- Lift Index <= 60 for low interaction dummy and Lift Index > 140 for high interaction dummy
- Derive interaction variables from same data source
  - We recommend to derive interaction variables from the same data source to avoid difficulties in variable interpretation in the model.
  - Exceptions apply when interacting with demographic variables such as age, income, etc.
- Define the lower boundary/value for the interaction variables - unless the missing values have similar lift index with the defined data range

2.5 Interaction Variable Detection (cont.)

To detect two-way interaction variables:

- Set up the CART the same way as we did using CART for variable screening

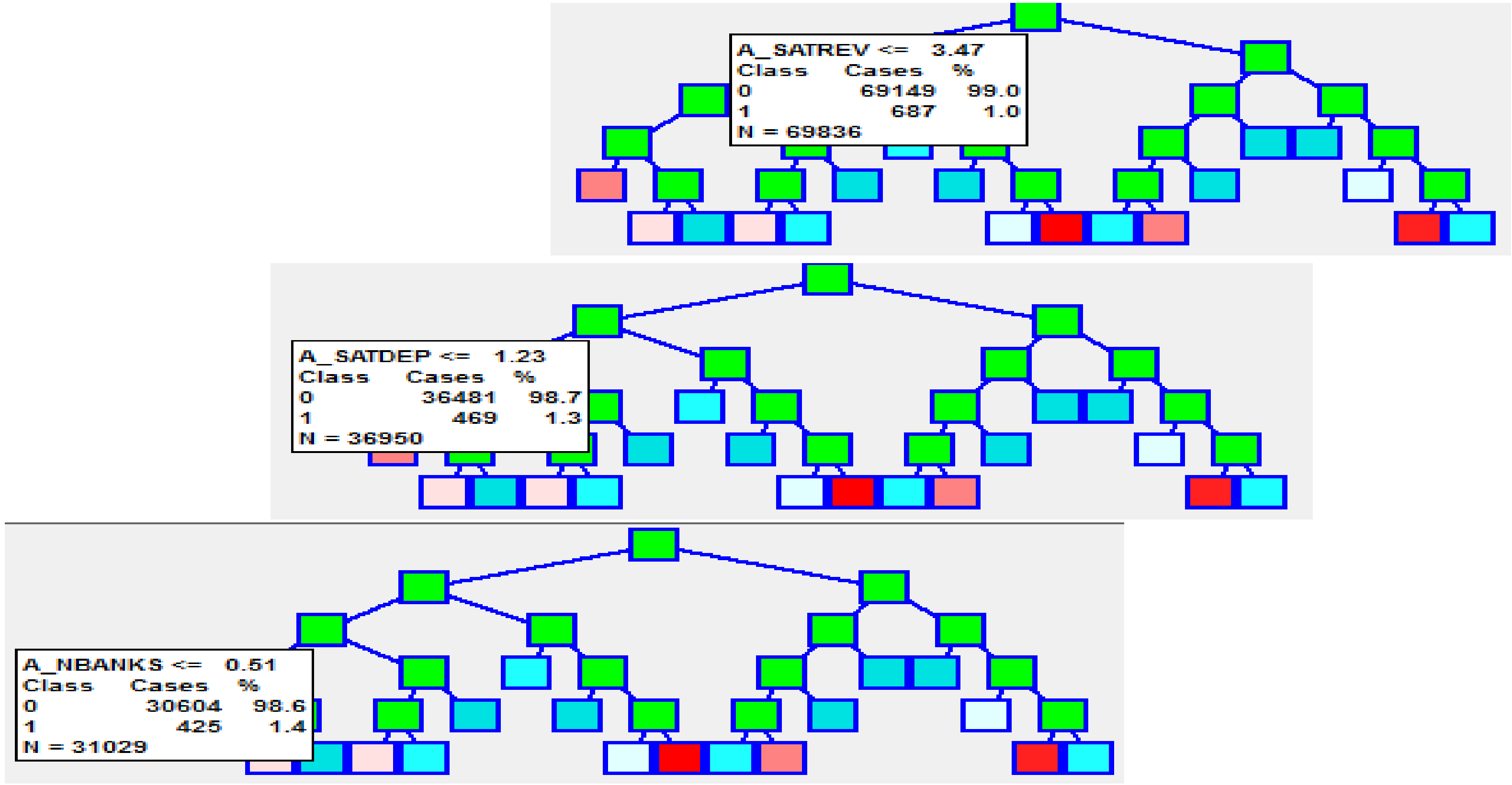
  Note: set up the tree depth to 4 to 5 levels in “Advance” tab



- The initial navigator window displays the overview of the tree on the top panel. In this example, we have a CART tree with 18 terminal nodes (the end nodes of the tree). The color-coded nodes help us locate the terminal nodes of interest. Red terminal nodes indicate the interaction variables with high lift index while blue terminal nodes indicate interaction with low lift index. Other colors indicate mixed results and should not be used as interaction dummy variables.

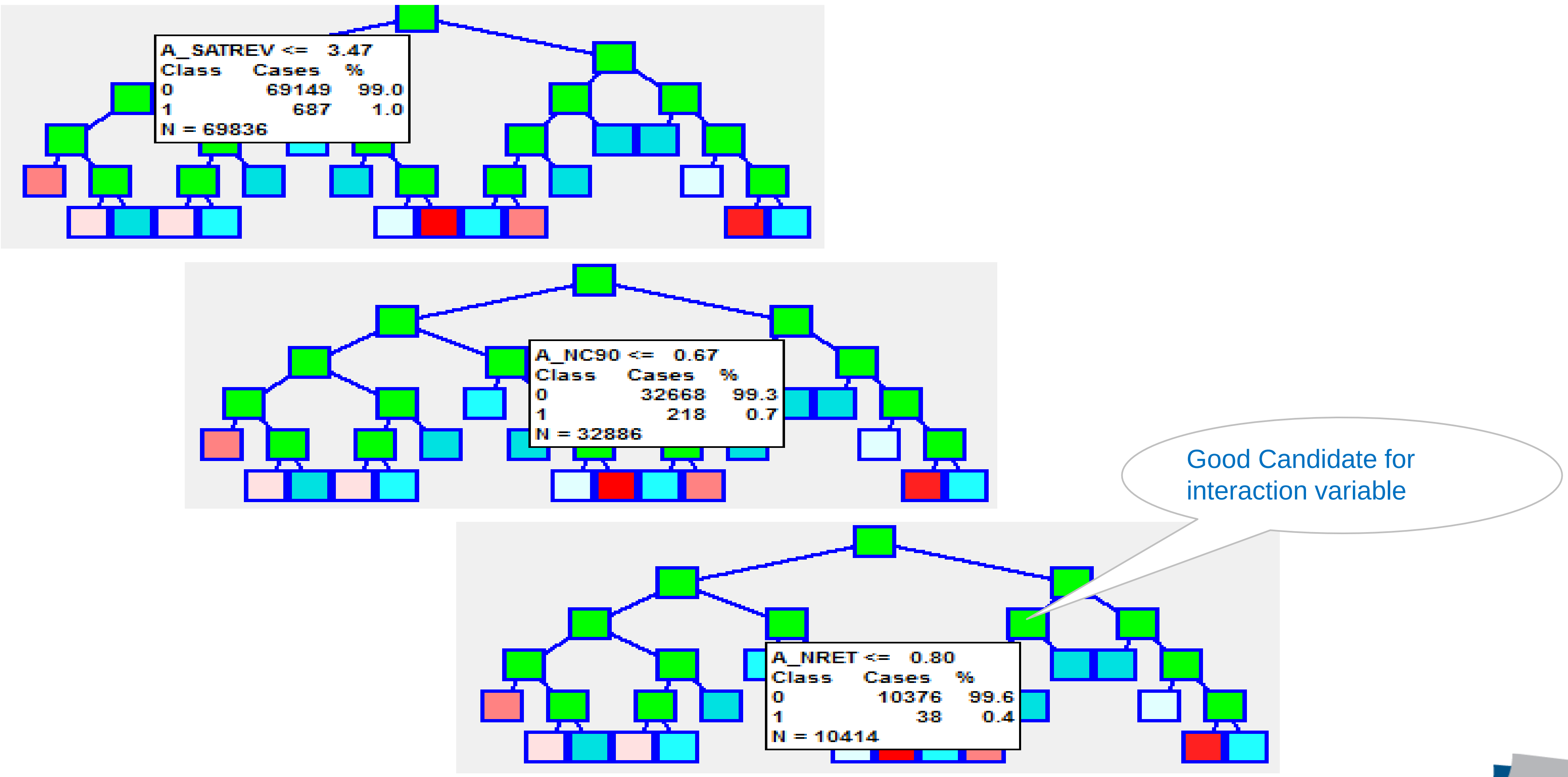
2.5 Interaction Variable Detection (cont.)

- We read the node information by putting the pointer on any node. The first panel below shows the root node information. The total analysis sample size is 69,836 with 687 responders. The response rate for the analysis file is 1.0%.
- The next two panels show the first left splitting node and second level left splitting node. The second level left splitting node forms the interaction of primary parent node (A_SATREV<= 3.47) and secondary parent node (A_SATDEP <= 1.23). The size of this node is 31,029 and the response rate is 1.4%. Compared with the root node, the response rate lift is 140 and coverage rate is 44.4%. This could be a good candidate for interaction variables.



2.5 Interaction Variable Detection

- The panels below show the information of root node, first level right splitting node, and second level left splitting node. The second level left splitting node forms the interaction of root node (A_SATREV > 3.47) and the secondary parent node (A_NC90 <= 0.67). The size of this node is 10,414 and the response rate is 0.4%. Compared with the root node, the response rate lift is 40 and the coverage rate is 14.9%. Therefore, it is a good candidate for an interaction dummy variable.



Model Fitting and Tuning

After the variable reduction step, we would usually reduce the model variable list to 20~ 40 coded variables.

Majority of modelers include 40 or fewer variables

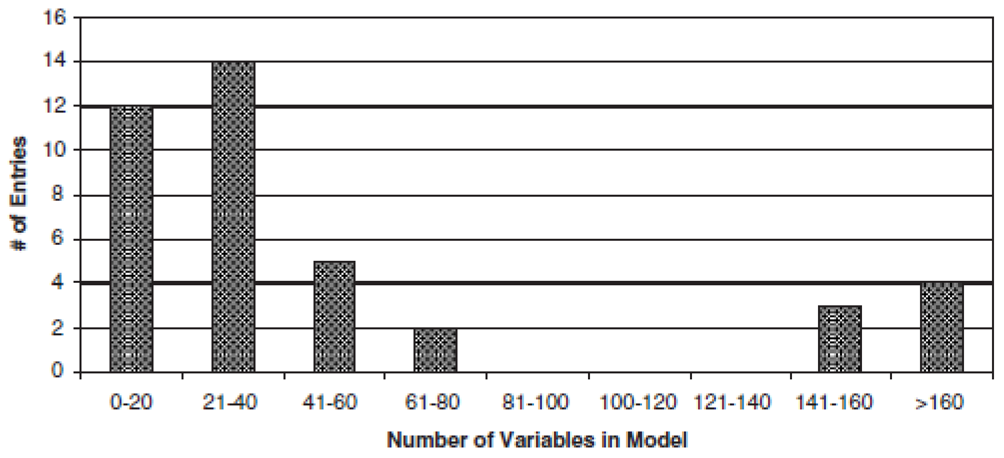
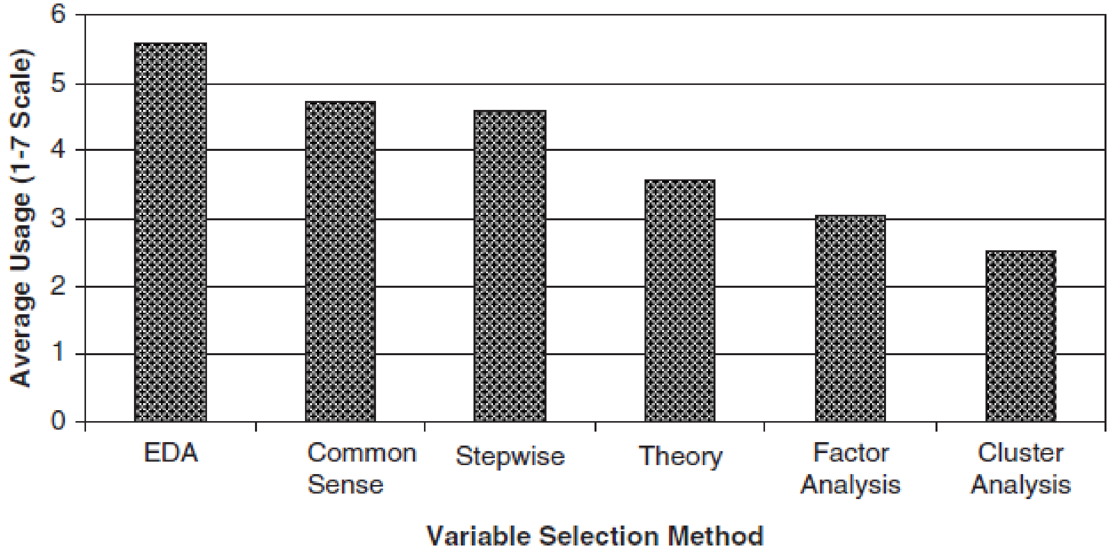


Fig. 10.4 Number of predictors used in predictive models for churn modeling tournament (Statistics provided by Neslin et al. 2006a).

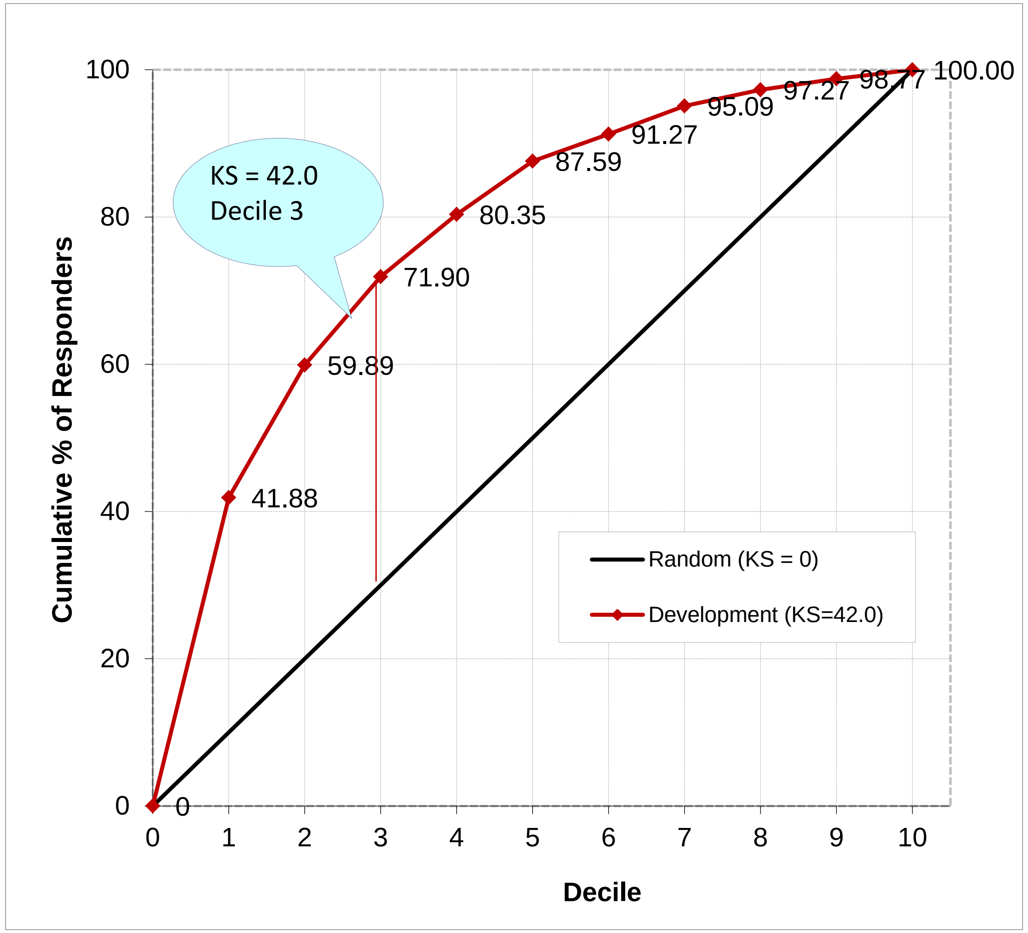


Model Fitting and Tuning (cont.)

- **Proc Logistic**
  - In general, we look at the Logistic Regression results on all 3 samples to make sure that the variable performance is consistent across samples. For example, is the sign of the variable coefficients the same across samples? Are all variables statistically significant across samples?
  - Check the Chi-Square value which is the indicator of variable explanatory power for develop sample, holdout sample and entire samples.
  - Check the p-value for variable statistical significance in all 3 samples.
  - If the number of responders is small, it is hard to tune the model using the develop sample only. In this case, we could use the entire sample to tune the model.
  - Option “OUTEST=OVERDEST” provides the estimates from the entire/develop sample so we could use them for the gains table.
  - %VARRANK ranks variables by their importance in the model using the standardized coefficients.

- **Gains Table**
  - Gains table gives us a better view of the model performance by showing the distribution of the responders, response rate, and cumulative % of responders.
  - Usually, our goal is to have as many responders as possible in the top decile/pentile to meet the campaign requirement. A good model should show a smooth top heavy distribution of responders in the gains table.
  - For xx products, a deeper penetration above hurdle is another important goal.

Model Statistics



Analysis of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
InteXept	1	-2139.01	87.7614	594.05	<.0001
XQNS__5_H	1	139.9459	29.2497	22.89	<.0001
XQDSPVG_L	1	-100.157	24.5885	16.59	<.0001
XQBSPXD_L	1	-182.316	29.4822	38.24	<.0001
XAGE_CD_L	1	-242.518	32.4796	55.75	<.0001
XBIHEAL_H	1	78.3517	25.4435	9.48	0.0021
XBIBEAU_H	1	51.5696	27.4619	3.53	0.0604
XBPMPLC_H	1	50.6529	21.8229	5.39	0.0203
XBPPRMC_H	1	42.0778	23.7526	3.14	0.0765
XBINOCC_H	1	84.0744	26.4135	10.13	0.0015
XBPFNGR_H	1	76.5219	25.1971	9.22	0.0024
XBHEMAI_H	1	64.864	21.6705	8.96	0.0028
XFVBCAN_C	1	2717.209	498.3764	29.73	<.0001
XFXBBAN_C	1	36.4133	9.5888	14.42	0.0001
XFXCUBN_C	1	25.9485	9.878	6.9	0.0086
XBMCELL_L	1	-90.3109	27.2049	11.02	0.0009
XBHCPIN_H	1	56.8443	24.3303	5.46	0.0195
XFZBCAN_H	1	86.1774	27.9998	9.47	0.0021
XFXCUAN_L	1	-87.3424	26.1246	11.18	0.0008
XACPBLR_L	1	-36.7705	25.1031	2.15	0.143
Scale	1	768.711	26.5318		

Fit Statistics

-2 Log Likelihood	18459.79
AIC (smaller is better)	18501.8
AICC (smaller is better)	18501.8
BIC (smaller is better)	18717.1

CALCULATION OF AUC AND GINI INDEX

Obs	AUC	Gini
1	0.78446	0.56893

CUTOFFS FOR SCORE

DECILE	MINIMUM ESTIMATE	MAXIMUM ESTIMATE
0	2.048752901	108.0117663
1	1.219306293	2.048733136
2	0.830278766	1.219301069
3	0.6004892	0.830273012
4	0.446513319	0.60048438
5	0.326086142	0.446507014
6	0.222714586	0.326085834
7	0.13781964	0.222703442
8	0.075148564	0.137819478
9	0.018632214	0.075146276

Model Fitting and Tuning (cont.)

Cutoff		
Decile	Minimum Estimate	Maximum Estimate
[illegible]	0.004282151	0.074203222
1	0.002657780	0.004282037
2	0.001928607	0.002657704
3	0.001457147	0.001928524
4	0.001184807	0.001457099

Decile Minimum Estimates give out the cutoff for each decile.

Predicted Response Rate,should be close to the list level response rate

Gains Table															
Decile	# of Obs	% of Obs	# of Coverage Names	% of Coverage Names	Cum % of Coverage Names	Predicted Coverage Prob	Coverage Rate	Cum Coverage Rate	Coverage Rate Lift	Total Coverage Amount ($)	% of Coverage	Cum % of Coverage	Coverage Amount Per Prospect (Cents)	Cum Coverage Amount Per Prospect (Cents)	Cum Coverage Amount Lift
0	34755	10.00	211	30.71	30.71	0.62	0.61	0.61	307	47168	30.46	30.46	135.70	135.70	305
1	34754	10.00	109	15.87	46.58	0.32	0.31	0.46	233	24497	15.82	46.28	70.49	103.10	231
2	34754	10.00	83	12.08	58.66	0.25	0.24	0.39	196	19061	12.31	58.59	54.85	87.02	195
3	34754	10.00	71	10.33	69.00	0.20	0.20	0.34	172	15853	10.24	68.82	45.61	76.67	172
4	34754	10.00	62	9.02	78.02	0.17	0.18	0.31	156	14058	9.08	77.90	40.45	69.42	156
5	34755	10.00	44	6.40	84.43	0.13	0.13	0.28	141	9498	6.13	84.03	27.33	62.41	140
6	34754	10.00	42	6.11	90.54	0.11	0.12	0.26	129	9612	6.21	90.24	27.66	57.44	129
7	34754	10.00	37	5.39	95.92	0.08	0.11	0.24	120	8437	5.45	95.69	24.28	53.30	120
8	34754	10.00	16	2.33	98.25	0.06	0.05	0.22	109	3804	2.46	98.15	10.95	48.59	109
9	34754	10.00	12	1.75	100.00	0.03	0.03	0.20	100	2871	1.85	100.00	8.26	44.56	100
	347542	100.00	687							154859	100.00				

It measures the responders capture rate of the model. A good model should have a high percentage on the top 50% of the file.

Response Rate in list level. A good model should have a smooth trend from top to bottom

The premium section will be used as a reference in model tuning stage of HAP response model building

Final Model Variables

Smarter Marketing

Variable	Variable Description	Transformation	Missing Imputation	Variable Weight	Correlation	Data Source
XAGE_CD_L	AGE_CD_CLIENT_IB in ('B','C')	Dummy	-	16	Negative	Client
XQBSPXD_L	PROPS:PRODUCT FLAG X(1,0) DM = 1	Dummy	-	9.6	Negative	Housefile
XFZBCAN_H	DP05:GENERIC DDA X XX MODEL >= 0.015806	Dummy	-	5.6	Positive	Housefile
XQNS__5_H	PROPS:TOT PROMOTIONS 12-24M ANY MEDIA = 0	Dummy	-	7.8	Positive	Housefile
XFXBBAN_C	BB03: GENERIC X Bank	Continuous	-	4	Positive	Housefile
XFVBCAN_C	DP05:GENERIC Regional Banks and CU X	Continuous	-	5.4	Positive	Housefile
XFXCUAN_L	CU04:GENERIC CU DDA <= 3.0349	Dummy	-	5.7	Negative	Housefile
XBPMPLC_H	XX -Place Groups in ('12C','07C','05C','04B','02A')	Dummy	-	3.1	Positive	InfoBase
XBIHEAL_H	Health / Medical - General - Interest = 1	Dummy	-	4.6	Positive	InfoBase
XBINOCC_H	Occupation - Input Individual in ('8','2','7','5')	Dummy	-	3.7	Positive	InfoBase
XFXCUBN_C	CU02: GENERIC X Insured Basics CUs	Continuous	-	3.6	Positive	Housefile
XQDSPVG_L	PROPS:DUR SNCE LAST PROMO PROD X = 0	Dummy	-	6.7	Negative	Housefile
XBHCPIN_H	Consumer Prominence Indicator >=8	Dummy	-	3.7	Positive	InfoBase
XBPFNGR_H	XX Financial Group Code in ('04H','03M','02U')	Dummy	-	3.9	Positive	InfoBase
XBHEMAI_H	eMail Append Available Indicator in ('Y')	Dummy	-	4.3	Positive	InfoBase
XBIBEAU_H	Beauty and Cosmetics - Interest = 1	Dummy	-	2.5	Positive	InfoBase
XBMCELL_L	Media Channel Usage Cell Phone <=2	Dummy	-	5.5	Negative	InfoBase
XBPPRMC_H	XX Cluster in ('84','41','87','61','70','35','69','68')	Dummy	-	2.1	Positive	InfoBase
XACPBLR_L	Population 25+ - % Bachelors Degree >= 26	Dummy	-	2.2	Negative	Census

Final Model Variables

Smarter Marketing

VAR=XARN
EW_H

RANK	MIN	MAX	NUMBER OF OBS	% OBS	TOTAL OF PRM_AMT	% PRM_AMT	PRM_AMT MEAN	UPPER 95% INDEX	LOWER 95% INDEX	INDEX OF PRM_AMT MEAN	NUMBER OF RESPONSE	% RESPONSE	RESPONSE RATE X100	UPPER 95% INDEX	LOWER 95% INDEX	INDEX OF RESPONSE RATE
4	0	0	139269	82.13	101933	77.34	0.73	104	84	94	472	77.12	0.34	102	85	94
9	1	1	30304	17.87	29866	22.66	0.99	151	102	127	140	22.88	0.46	149	107	128
VAR			169573	100	131798	100					612	100				

Responder vs Non-responder Characteristics

Smarter Marketing

Most likely to be responders	Least likely to be responders
DP05:GENERIC DDA X XX MODEL >= 0.015806	Client age in 18-39
TOT PROMOTIONS 12-24M ANY MEDIA = 0	PRODUCT FLAG X(1,0) DM = 1
Score higher in BB03: GENERIC X Bank	CU04:GENERIC CU DDA <= 3.0349
Score higher in DP05:GENERIC Regional Banks and CU X	DUR SNCE LAST PROMO PROD X = 0
XX -Place Groups in ('12C','07C','05C','04B','02A')	BB02: GENERIC X Bank -21 clients FF <= 1.2535
Interest in Health / Medical - General	Population 25+ - % Bachelors Degree >= 26
Occupation - Input Individual in ('8','2','7','5')	
Score higher in CU02: GENERIC X Insured Basics Cus	
Consumer Prominence Indicator >=8	
XX Financial Group Code in ('04H','03M','02U')	
eMail Append Indicator Available (‘Y”)	
XX Cluster in ('84','41','87','61','70','35','69','68')	
Interest in Beauty and Cosmetics	

Model Statistics

Smarter Marketing

variable	rank 1 (|AUC-0.5|)	rank 2 (partial correlation)	rank avg.	estimate (standardized)	final estimate (standardized)
XAGE_CD_L	5 (0.136)	1 (0.0067)	3.0	1 (0.2730)	**1 (0.2078)**
XFVBCAN_C	3 (0.143)	3 (0.0032)	3.0	10 (0.0662)	**6 (0.0614)**
XFXBBAN_C	1 (0.222)	6 (0.0017)	3.5	15 (0.0565)	**5 (0.0777)**
XQBSPXD_L	8 (0.084)	2 (0.0042)	5.0	2 (0.1433)	**2 (0.0906)**
XFZBCAN_H	6 (0.131)	7 (0.0014)	6.5	5 (0.1062)	**3 (0.0878)**
XQNS__5_H	10 (0.066)	4 (0.0031)	7.0	3 (0.1337)	**4 (0.0789)**
XFXCUAN_L	2 (0.144)	14 (0.0007)	8.0	9 (0.0710)	**7 (0.0597)**
XQDSPVG_L	13 (0.059)	5 (0.0021)	9.0	4 (0.1169)	**12 (0.0445)**
XBPMPLC_H	7 (0.098)	12 (0.0008)	9.5	13 (0.0598)	**8 (0.0541)**
XFXCUBN_C	4 (0.138)	16 (0.0006)	10.0	16 (0.0460)	**11 (0.0459)**
XBINOCC_H	14 (0.051)	9 (0.0011)	11.5	14 (0.0567)	**10 (0.0481)**
XBHEMAI_H	17 (0.036)	8 (0.0011)	12.5	7 (0.0741)	**15 (0.0360)**
XBMCELL_L	12 (0.063)	13 (0.0007)	12.5	6 (0.0812)	**17 (0.0236)**
XBIHEAL_H	16 (0.039)	10 (0.0010)	13.0	8 (0.0723)	**9 (0.0492)**
XBHCPIN_H	11 (0.065)	15 (0.0007)	13.0	11 (0.0637)	**13 (0.0438)**
XBPPRMC_H	9 (0.074)	18 (0.0003)	13.5	19 (0.0343)	**18 (0.0234)**
XBPFNGR_H	18 (0.035)	11 (0.0009)	14.5	12 (0.0604)	**14 (0.0387)**
XACPBLR_L	15 (0.048)	19 (0.0003)	17.0	17 (0.0445)	**19 (8.8112E-05)**
XBIBEAU_H	19 (0.034)	17 (0.0004)	18.0	18 (0.0437)	**16 (0.0269)**

- Weights more evenly distributed than the other model (response model).
- The other model shifts weights towards predictors that are likely to be more useful (using l1 regularization).

Gains Table

Smarter Marketing

Modeling sample

Revenue DECILE Gains Table

DECILE	Number Of Obs	Percent Of Obs	Premium Amount (Dollars)	Percent Of Premium	Cum Percent Of Premium	Predicted PPPM (Cents)	PPPM (Cents)	Cum PPPM (Cents)	Revenue Lift
0	20959	10	$74,910	39.29	39.29	400.99	357.41	357.41	393
1	20958	10	$34,218	17.95	57.24	156.96	163.27	260.34	286
2	20958	10	$24,510	12.86	70.1	100.54	116.95	212.55	234
3	20959	10	$17,880	9.38	79.48	70.65	85.31	180.74	199
4	20958	10	$15,210	7.98	87.46	51.88	72.57	159.1	175
5	20958	10	$8,070	4.23	91.69	38.46	38.51	139	153
6	20959	10	$8,520	4.47	96.16	27.35	40.65	124.95	137
7	20958	10	$3,780	1.98	98.14	17.87	18.04	111.59	123
8	20958	10	$1,320	0.69	98.84	10.44	6.3	99.89	110
9	20958	10	$2,220	1.16	100	5.11	10.59	90.96	100
	209583	*100*	*$190,638*	*100*					

Supplemental Response DECILE Gains Table

DECILE	Number Of Obs	Percent Of Obs	Number Of Supp Responses	Cum Number Of Supp Responses	Percent Of Supp Responses	Cum Percent Of Supp Responses	Supp Response Rate	Cumulative Supp Response Rate	Supp Response Lift
0	20959	10	306	306	41.75	41.75	1.46	1.46	417
1	20958	10	134	440	18.28	60.03	0.64	1.05	300
2	20958	10	87	527	11.87	71.9	0.42	0.84	240
3	20959	10	64	591	8.73	80.63	0.31	0.7	202
4	20958	10	50	641	6.82	87.45	0.24	0.61	175
5	20958	10	27	668	3.68	91.13	0.13	0.53	152
6	20959	10	28	696	3.82	94.95	0.13	0.47	136
7	20958	10	17	713	2.32	97.27	0.08	0.43	122
8	20958	10	11	724	1.5	98.77	0.05	0.38	110
9	20958	10	9	733	1.23	100	0.04	0.35	100
	209583	*100*	*733*						

Gains Table

Smarter Marketing

Validation sample

Revenue DECILE Gains Table

DECILE	Number Of Obs	Percent Of Obs	Premium Amount (Dollars)	Percent Of Premium	Cum Percent Of Premium	Predicted PPPM (Cents)	PPPM (Cents)	Cum PPPM (Cents)	Revenue Lift
0	16128	10	$51,156	37.34	37.34	472.12	317.19	317.19	373
1	16127	10	$16,842	12.29	49.63	195.22	104.43	210.81	248
2	16127	10	$13,422	9.8	59.43	130.97	83.23	168.29	198
3	16127	10	$13,170	9.61	69.04	94.98	81.66	146.63	173
4	16127	10	$17,190	12.55	81.59	70.62	106.59	138.62	163
5	16128	10	$7,440	5.43	87.02	53.11	46.13	123.21	145
6	16127	10	$5,400	3.94	90.96	39.43	33.48	110.39	130
7	16127	10	$5,070	3.7	94.66	27.88	31.44	100.52	118
8	16127	10	$5,190	3.79	98.45	17.51	32.18	92.93	109
9	16127	10	$2,130	1.55	100	7.97	13.21	84.96	100
	161272	*100*	*$137,010*	*100*					

Supplemental Response DECILE Gains Table

DECILE	Number Of Obs	Percent Of Obs	Number Of Supp Responses	Cum Number Of Supp Responses	Percent Of Supp Responses	Cum Percent Of Supp Responses	Supp Response Rate	Cumulative Supp Response Rate	Supp Response Lift
0	16128	10	184	184	36.36	36.36	1.14	1.14	364
1	16127	10	67	251	13.24	49.6	0.42	0.78	248
2	16127	10	45	296	8.89	58.5	0.28	0.61	195
3	16127	10	49	345	9.68	68.18	0.3	0.53	170
4	16127	10	70	415	13.83	82.02	0.43	0.51	164
5	16128	10	25	440	4.94	86.96	0.16	0.45	145
6	16127	10	17	457	3.36	90.32	0.11	0.4	129
7	16127	10	20	477	3.95	94.27	0.12	0.37	118
8	16127	10	20	497	3.95	98.22	0.12	0.34	109
9	16127	10	9	506	1.78	100	0.06	0.31	100
	161272	*100*	*506*						

Gains Table

Smarter Marketing

Validation sample

DECILE	Number Of Obs	Percent Of Obs	Premium Amount (Dollars)	Percent Of Premium	Cum Percent Of Premium	Predicted PPPM (Cents)	PPPM (Cents)	Cum PPPM (Cents)	Revenue Lift
0	16128	10.00	$52,236	38.13	38.13	1.27	323.88	323.88	381
1	16127	10.00	$16,332	11.92	50.05	0.65	101.27	212.58	250
2	16127	10.00	$13,152	9.60	59.65	0.48	81.55	168.91	199
3	16127	10.00	$13,710	10.01	69.65	0.38	85.01	147.93	174
4	16127	10.00	$14,460	10.55	80.21	0.31	89.66	136.28	160
5	16128	10.00	$9,540	6.96	87.17	0.26	59.15	123.42	145
6	16127	10.00	$4,350	3.17	90.34	0.22	26.97	109.65	129
7	16127	10.00	$5,250	3.83	94.18	0.18	32.55	100.01	118
8	16127	10.00	$5,550	4.05	98.23	0.13	34.41	92.72	109
9	16127	10.00	$2,430	1.77	100.00	0.08	15.07	84.96	100
	161272	*100.00*	*$137,010*	*100.00*					

Modeling sample

DECILE	Number Of Obs	Percent Of Obs	Premium Amount (Dollars)	Percent Of Premium	Cum Percent Of Premium	Predicted PPPM (Cents)	PPPM (Cents)	Cum PPPM (Cents)	Revenue Lift
0	20959	10.00	$77,394	40.60	40.60	1.20	369.26	369.26	406
1	20958	10.00	$30,744	16.13	56.72	0.58	146.69	257.98	284
2	20958	10.00	$24,810	13.01	69.74	0.41	118.38	211.45	232
3	20959	10.00	$16,290	8.54	78.28	0.32	77.72	178.02	196
4	20958	10.00	$15,240	7.99	86.28	0.26	72.72	156.96	173
5	20958	10.00	$11,850	6.22	92.49	0.22	56.54	140.22	154
6	20959	10.00	$7,350	3.86	96.35	0.18	35.07	125.20	138
7	20958	10.00	$3,720	1.95	98.30	0.14	17.75	111.77	123
8	20958	10.00	$1,080	0.57	98.87	0.09	5.15	99.92	110
9	20958	10.00	$2,160	1.13	100.00	0.07	10.31	90.96	100
	209583	*100.00*	*$190,638*	*100.00*					

Model Assessment

Smarter Marketing

ON MODELING SAMPLE												
Model	Cum PPNM (Cents)				Lift				Premium Amount Capture Rate			
	Top 10%	Top 30%	Top 50%	Top 80%	Top 10%	Top 30%	Top 50%	Top 80%	Top 10%	Top 30%	Top 50%	Top 80%
X	357.80	212.16	159.45	111.59	393	233	175	123	39.33%	69.97%	87.65%	98.14%
Y	272.96	177.91	146.71	107.1	300	196	161	118	30.01%	58.68%	80.64%	94.19%
GENERIC	258.25	174.59	141.05	107.99	284	192	155	119	28.39%	57.58%	77.53%	94.98%

ON VALIDATION SAMPLE (Jun '20 mailed)												
Model	Cum PPNM (Cents)				Lift				Premium Amount Capture Rate			
	Top 10%	Top 30%	Top 50%	Top 80%	Top 10%	Top 30%	Top 50%	Top 80%	Top 10%	Top 30%	Top 50%	Top 80%
X	317.19	168.29	138.62	100.52	373	198	163	118	37.34%	59.43%	81.59%	94.66%
Y	202.71	141.23	118.47	96.38	244	170	142	116	24.37%	50.93%	71.20%	92.69%
GENERIC	254.66	163.2	122.11	94.66	306	196	147	114	30.61%	58.85%	73.39%	91.03%

ON VALIDATION SAMPLE (Dec '19 mailed)												
Model	Cum PPNM (Cents)				Lift				Premium Amount Capture Rate			
	Top 10%	Top 30%	Top 50%	Top 80%	Top 10%	Top 30%	Top 50%	Top 80%	Top 10%	Top 30%	Top 50%	Top 80%
X	227.56	128.54	100.85	77.55	350	197	155	119	34.96%	59.25%	77.48%	95.32%
Y	180.7	111.77	95.31	74.82	278	172	146	115	27.76%	51.52%	73.22%	91.96%
GENERIC	215.34	122.22	93.12	74.63	331	188	143	115	33.09%	56.33%	71.54%	91.74%

- The new model outperforms the previous version and the generic model in the modeling sample and validation sample everywhere