Paper 146-2008

Using SAS® Enterprise Miner™ to Prescribe a Pre-Screen Mailing

Rex Pruitt, PREMIER Bankcard, LLC, Sioux Falls, SD 57107

ABSTRACT

The Objective (Target) of this project was to prescribe a pre-screen test mailing program using a combination of a "Risk Score" and Specific Credit Bureau Attributes (CBA). CBA variables chosen were predictive of accounts that generate revenue in the Top 50% of the Company's customer portfolio, have acceptable delinquency levels and have a "Risk Score" less than what is normally acceptable.

A random sample of 20,000 records was selected from our total current portfolio of customers ranked by Revenue. The binary target value was determined using the objective statement noted above. The affirmative value for the target was represented by approximately 19% of the sample population.

Four modeling methods were used: Neural Network, Logistic Regression, Dmine Regression and Decision Tree. Based on model assessment, the Decision Tree is the strongest model based on misclassification rate, KS, Gini, etc. Since we will be establishing "Cut-Points" based on a decision process, the Decision Tree model will be used for identification of the selection parameters as opposed to deploying a record scoring methodology.

Tables that will be used to demonstrate the data preparation/selection, model selection scores, and chosen decision tree specifications are as follows:

TABLES:

1. Decision Tree Cut Point Table
2. Variable Definition Table
3. Sample Selection Statistics Tables (3.1 & 3.2)
4. Training Model Comparison Table
5. Validation Model Comparison Table
6. Decision Tree Variable Significance Table
7. Decision Tree Model Misclassification Plot
8. Decision Tree Model Average Square Error Plot.

INTRODUCTION

Predictive modeling using SAS Enterprise Miner is very powerful and capable of generating huge lifts in revenue for the organization. My experience, as demonstrated by the example illustrated in the context of this paper, is a clear depiction of the benefits resulting from predictive modeling. Using SAS tools in target marketing is having a significant impact on our Company's business. Venturing into the huge amounts of credit bureau information can be a daunting task. However, by employing SAS Enterprise Miner, coupled with some Base SAS techniques, gold nuggets can be identified.

The Objective (Target) of this project was to prescribe a pre-screen test mailing program using a combination of a "Risk Score" and Specific Credit Bureau Attributes (CBA). CBA variables to be chosen are predictive of accounts that generate revenue in the Top 50% of the Company's customer portfolio, have acceptable delinquency levels and have a "Risk Score" less than what is normally acceptable.

METHODS

A random sample of 20,000 records was selected from PREMIER's total current portfolio of customers ranked by Revenue. The binary target value was determined using the objective statement noted in the introduction.

Four modeling methods were used: Neural Network, Logistic Regression, Dmine Regression and Decision Tree. Due to the need to be able to express the resulting program in simple terms for use at the credit bureau, the recommendation will be in the form of establishing "Cut-Points" based on a decision process. As a result, the Decision Tree model will be used for identification of the selection parameters as opposed to deploying a record scoring methodology.

RESULTS

Sample Selection – The sample used for modeling was selected from the total current portfolio of customers ranked by Revenue divided into 5% segments. A random sample of 20,000 records was selected. The affirmative value for the target was represented by approximately 19% of the sample population (see Table #3.1 and 3.2).

Variable Selection Results – Four CBA variables (out of 18) were identified as predictive of the Objective (Target) in the modeling exercises. The Decision Tree results were used to establish the variable "Cut-Points." The Risk_Score_5 variable was used in the modeling exercise as well, but was forced below a normally acceptable value in the "Cut-Point" table. All significant variables are listed and defined in the "Variable Definition Table" (Table #2 omitted due to confidentiality).

Model Comparison Results – Four modeling methods were used: Neural Network, Logistic Regression, Dmine Regression & Decision Tree (see Table #4 & 5). Based on model assessment, the Decision Tree is the strongest model based on misclassification rate, KS, Gini, etc. This is especially good since, as noted earlier, there is a requirement that the final recommendation be based on establishing "Cut-Points" following the Decision Tree process. Therefore, the Decision Tree model will be used for identification of the selection parameters as opposed to deploying a record scoring methodology.

Actual Mailing Prospect Results – Using the cut-point table criteria at a single credit bureau, there were over 6.5 million new names identified that may potentially generate over $9 million in additional revenue each year. By adding the other two bureaus, there is a very good chance that the revenue number could be doubled to over $18 million per year.

The tables used to demonstrate the data preparation/selection, model selection scores, and chosen decision tree specifications are as follows:

RESULTS TABLES:

1. Decision Tree Cut Point Table
2. Variable Definition Table (omitted due to confidentiality)
3. Sample Selection Statistics Tables (3.1 & 3.2)
4. Training Model Comparison Table
5. Validation Model Comparison Table
6. Decision Tree Variable Significance Table
7. Decision Tree Model Misclassification Plot
8. Decision Tree Model Average Square Error Plot

Attribute	Tree Model #3 Decision/Cut-Point
Risk_Score_5	<XX
CBA_14	No
CBA_9	Yes
CBA_10	No
CBA_1	No

Decision Tree "Cut-Point" (Table

Variable	N	N Miss	Mean	Minimum	Maximum	t Value	Pr > |t|	Variance	Std Dev
Target	20001	0	0.19	0.00	1.00	67.68	<.0001	0.15	0.39
Risk_Score_5	19858	143	77.26	46.00	102.00	1052.07	<.0001	107.09	10.35
CBA_1	17352	2649	0.42	0.00	1.00	111.75	<.0001	0.24	0.49
CBA_9	17352	2649	0.61	0.00	1.00	165.57	<.0001	0.24	0.49
CBA_10	17352	2649	0.17	0.00	1.00	60.02	<.0001	0.14	0.38
CBA_14	17352	2649	0.00	0.00	1.00	7.50	<.0001	0.00	0.06

Sample Selection Statistics (Table #3.1):

Target	N Obs	Variable	N	N Miss	Mean	Minimum	Maximum	t Value	Pr > |t|	Variance	Std Dev
0	16274	Risk_Score_5	16207	67	79.55	46.00	102.00	1021.18	<.0001	98.34	9.92
		CBA_1	14192	2082	0.43	0.00	1.00	102.51	<.0001	0.24	0.49
		CBA_9	14192	2082	0.62	0.00	1.00	152.78	<.0001	0.24	0.48
		CBA_10	14192	2082	0.20	0.00	1.00	59.11	<.0001	0.16	0.40
		CBA_14	14192	2082	0.00	0.00	1.00	6.72	<.0001	0.00	0.06
1	3727	Risk_Score_5	3651	76	67.12	49.00	74.00	909.32	<.0001	19.89	4.46
		CBA_1	3160	567	0.39	0.00	1.00	44.69	<.0001	0.24	0.49
		CBA_9	3160	567	0.57	0.00	1.00	64.66	<.0001	0.25	0.50
		CBA_10	3160	567	0.06	0.00	1.00	13.77	<.0001	0.05	0.23
		CBA_14	3160	567	0.00	0.00	1.00	3.32	0.0009	0.00	0.06

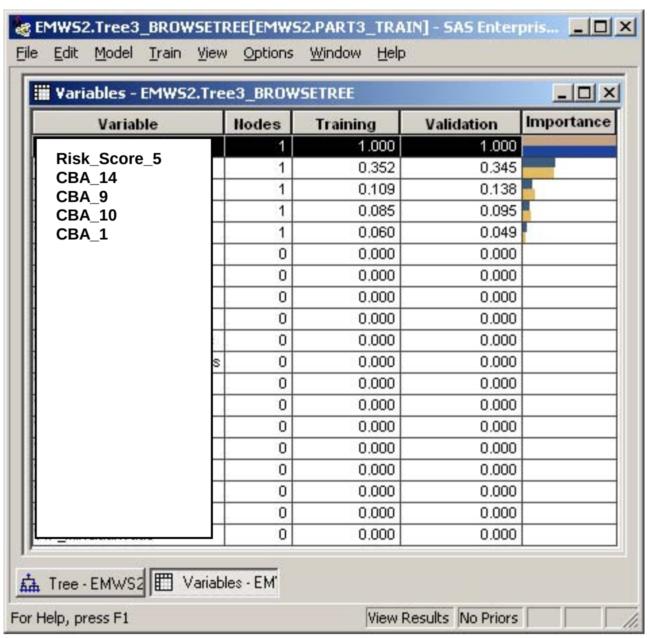
Sample Selection Statistics (Table #3.2):

Test Type	Fit Statistic	DmineReg3	Neural3	Reg3	Tree3
0-Use Indicator	Model Selected (1-Yes, 0=No)	0.00	0.00	0.00	1.00
1-KS	Bin-Based Two-Way Kolmogorov-Smirnov Statisti	0.67	0.58	0.56	0.68
1-KS	Kolmogorov-Smirnov Statistic	0.67	0.59	0.56	0.69
2-GINI	Gini Coefficient	0.71	0.67	0.63	0.75
4-Classification	Frequency of Classified Cases	10000.00	.	.	.
4-Classification	Misclassification Rate	0.19	0.19	0.20	0.16
4-Classification	Number of Wrong Classifications	1903.00	1861.00	.	.

Model Comparison "Training" (Table

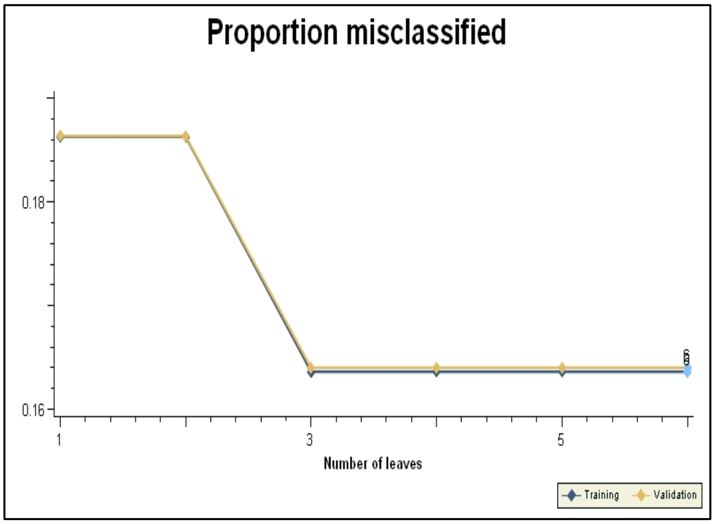
Test Type	Fit Statistic	DmineReg3	Neural3	Reg3	Tree3
1-KS	Bin-Based Two-Way Kolmogorov-Smirnov Statisti	0.66	0.58	0.56	0.70
1-KS	Kolmogorov-Smirnov Statistic	0.67	0.58	0.56	0.70
2-GINI	te: Gini Coefficient	0.72	0.68	0.64	0.77
4-Classification	Frequency of Classified Cases	10001.00	.	.	.
4-Classification	Misclassification Rate	0.19	0.18	0.20	0.16
4-Classification	Number of Wrong Classifications	1898.00	1838.00	.	.

Model Comparison "Validation" (Table

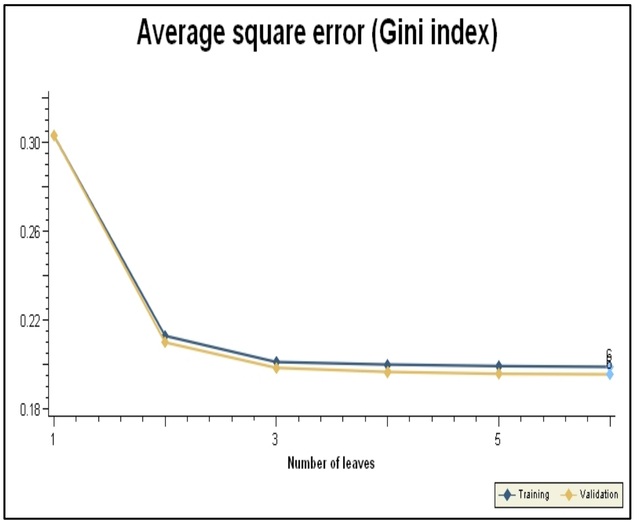


Variable	Nodes	Training	Validation	Importance
Risk_Score_5	1	1.000	1.000	
CBA_14	1	0.352	0.345	
CBA_9	1	0.109	0.138	
CBA_10	1	0.085	0.095	
CBA_1	1	0.060	0.049	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	
	0	0.000	0.000	

Decision Tree Variable Significance (Table



Decision Tree Model Misclassification Plot (Table #7):
Lower misclassification ratios represent a good model. Also, the closeness of the Decision Tree Model results for the Training and Validation datasets, the better the model performance.



Decision Tree Model Average Square Error Plot (Table #8):
The lower error demonstrates minimal target selection bias by the model.

CONCLUSION

By using predictive modeling with SAS Enterprise Miner is my company is able to generate huge lifts in revenue for the organization. My experience, as demonstrated by the example illustrated in the context of this paper, is a clear depiction of the benefits resulting from predictive modeling. Using SAS tools in target marketing is having a significant impact on our Company's business. Venturing into the huge amounts of credit bureau information, or any other huge set of data, can be a daunting task. However, by employing SAS Enterprise Miner, coupled with some Base SAS techniques, gold nuggets can be identified.

Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Rex Pruitt
PREMIER Bankcard
3820 N. Louise Ave.
Sioux Falls, SD 57107
(605) 575-9810
(605) 575-9866:
rpruitt@premierbankcard.com