SAS® GLOBAL FORUM 2015

The Journey Is Yours

Credit Scorecard Generation: An application of the Credit Scoring Node in SAS® Enterprise Miner

Ayush Priyadarshi[1] , Kushal Kathed[2] , Shilpi Prasad[3]
[1,2,3]Graduate Student, Oklahoma State University

SAS GLOBAL FORUM

Credit Scorecard Generation: An application of the Credit Scoring Node in SAS® Enterprise Miner

Ayush Priyadarshi[1], Kushal Kathed[2], Shilpi Prasad[3]
[1,2,3]SAS & OSU Data Mining Program, Oklahoma State University

Abstract

In today’s competitive world, acquiring new customers is crucial for businesses but what if most of the acquired customers turn out to be defaulters. This decision would backfire on the business and may lead to losses. The extant statistical methods have enabled businesses to identify good risk customers rather than intuitively judging them. The objective of this paper is to build a credit risk scorecard utilizing the Credit Risk Node inside SAS® Enterprise Miner 12.3, which maybe further used by a manager to make an instant decision on whether to accept or reject a customer’s credit application. To ensure generalization of the model, the dataset has been partitioned using the data partition node in two groups of 70:30 as training and validation respectively. The target is a binary variable which categorizes customers into good risk and bad risk group. After identifying the key variables required to generate the credit scorecard, a particular score was assigned to each of its sub groups. The final model generating the scorecard has a prediction accuracy of about 75%. A cumulative cut-off score of 120 was generated by SAS to make the demarcation between good and bad risk customers. Even in case of future variations in the data, model refinement is easy as the whole process is already defined and does not need to be rebuilt from scratch.

Objective

The objective of this project is to identify the key variables used to create the credit scorecard for managers. This score would be used to differentiate between good and bad risk customers. Also, we have tried to exemplify the ease of Credit Risk Node of SAS® Enterprise Miner to segregate the profiles of good and bad risk customers.

Methods

Data Preparation

The German Bank dataset used for credit scoring was extracted from UCI Machine Learning repository and consisted of 15 variables that capture details such as status of customer’s existing checking account, purpose of the credit, credit amount, employment status and property as well as other characteristics. The data was partitioned using the data partition node into two groups of 70:30 as training and validation respectively. The target variable is a binary variable which categorizes customers into good risk and bad risk.

Our next goal was to bin each input variable into significantly distinct sub groups. To accomplish this, we used the interactive grouping node inside Credit Risk tab of SAS® Enterprise Miner. Even after automatic binning of the data, some of the variable’s binning didn’t made complete business sense, so for these we regrouped these variables manually using domain expertise.

For instance: Attribute 10 (guarantors/ Other debtors), was grouped as one group but we separated it into three different sub-groups:
A101 : none (Group 1)
A102 : co-applicant (Group 2)
A103 : guarantor (Group 3)

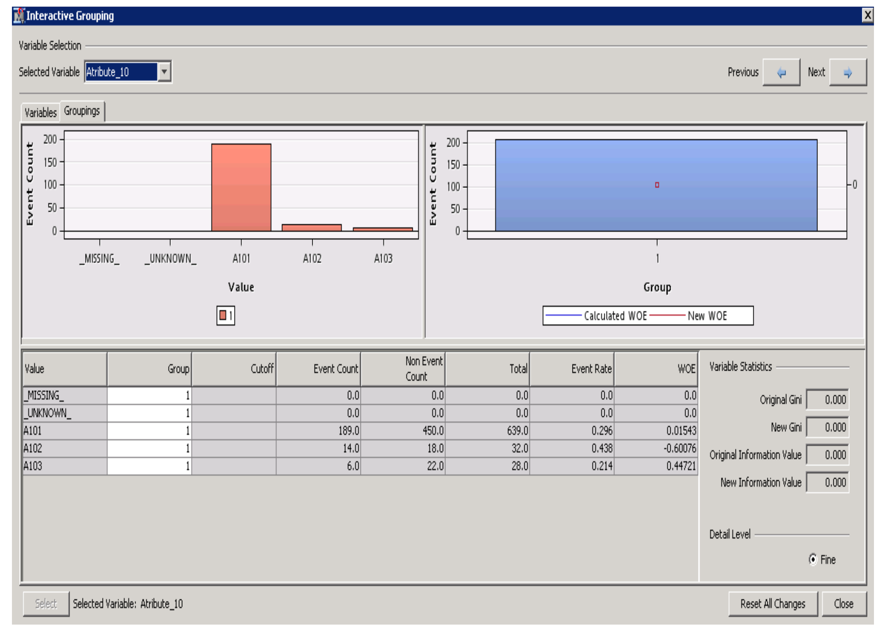

Fig. 1 Interactive binning of variables

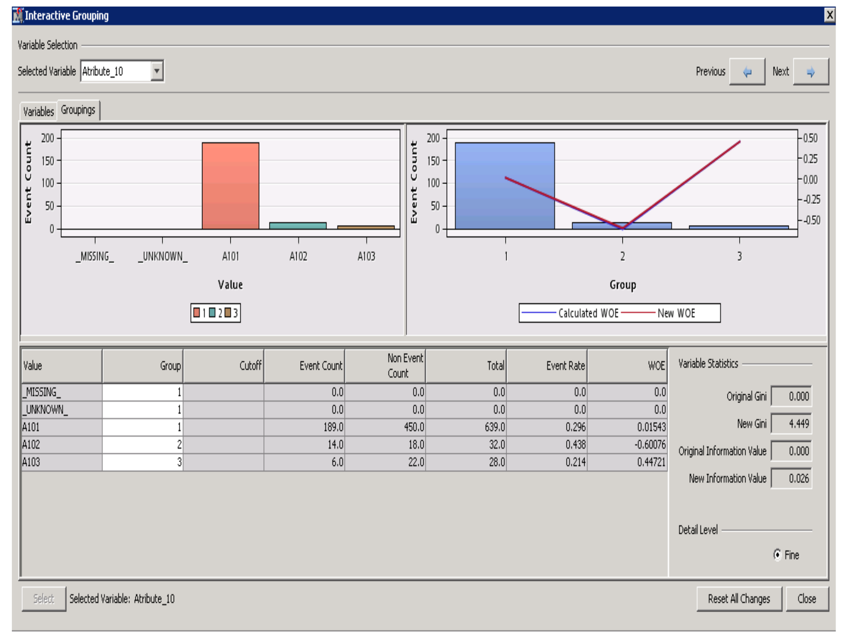

Fig. 2 Manual separation of bins

After separation, Weight Of Evidence (WOE) plot made more sense because the presence of a guarantor increases the chances of the customer being a better risks.

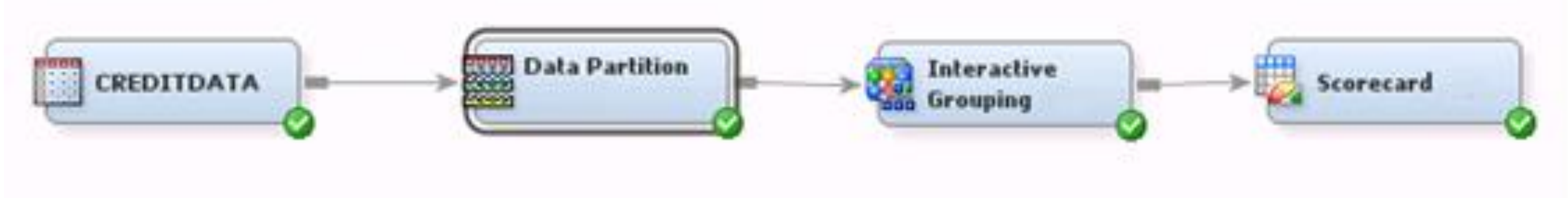

Fig. 3 SAS Enterprise Miner diagram

Model Assessment

Misclassification rate for the validation dataset is 25 % which means that we are able to predict about 75 % of the cases correctly which is quite reasonable.

Area under ROC (AUR) curve is 81.4% for training and 78% for the validation datasets. This plot is generally regarded as providing a good measure of the scorecard strength. A scorecard that is no better than random selection has an AUR value equal to 0.50.

KS Statistics for validation is 47.6 %. According to the KS plot, the best cutoff score to distinguish between good and bad risk customers is 120 for the validation dataset.

Fit Statistics

Target	Fit Statistics	Statistics Label	Train	Validation
Target	_AIC_	Akaike's Information Criterion	674.5298	.
Target	_ASE_	Average Squared Error	0.155806	0.164778
Target	_AVERR_	Average Error Function	0.46819	0.500795
Target	_DFE_	Degrees of Freedom for Error	689	.
Target	_DFM_	Model Degrees of Freedom	10	.
Target	_DFT_	Total Degrees of Freedom	699	.
Target	_DIV_	Divisor for ASE	1398	602
Target	_ERR_	Error Function	654.5298	301.4788
Target	_FPE_	Final Prediction Error	0.160329	.
Target	_MAX_	Maximum Absolute Error	0.966561	0.984461
Target	_MSE_	Mean Square Error	0.158068	0.164778
Target	_NOBS_	Sum of Frequencies	699	301
Target	_NW_	Number of Estimate Weights	10	.
Target	_RASE_	Root Average Sum of Squares	0.394723	0.405929
Target	_RFPE_	Root Final Prediction Error	0.400411	.
Target	_RMSE_	Root Mean Squared Error	0.397577	0.405929
Target	_SBC_	Schwarz's Bayesian Criterion	720.0263	.
Target	_SSE_	Sum of Squared Errors	217.8172	99.19658
Target	_SUMW_	Sum of Case Weights Times Freq	1398	602
Target	_MISC_	Misclassification Rate	0.240343	0.252492
Target	_KS_	Kolmogorov-Smirnov Statistic	0.492784	0.476923
Target	_AUR_	Area Under ROC	0.814281	0.789011
Target	_Gini_	Gini Coefficient	0.628562	0.578022
Target	_ARATIO_	Accuracy Ratio	0.628562	0.578022

Fig. 4 Fit statistics of the credit risk model

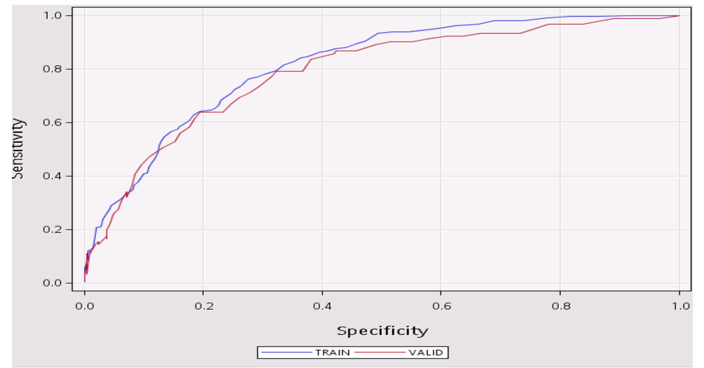

Fig. 5 ROC plot

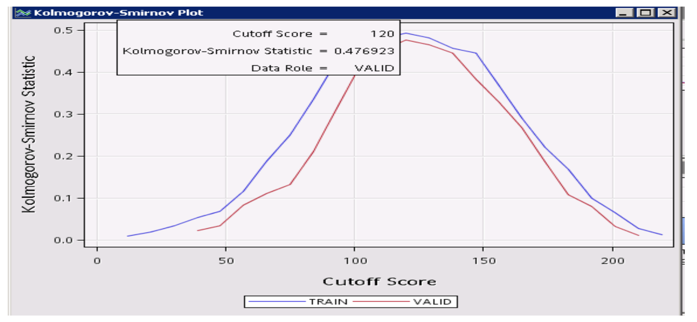

Fig. 6 Cutoff value to segregate good and bad customers

Results

Customer Profiles based on the credit scorecard

Attributes of Good Risk Customers:
Age group 44 to 53
Customers who own a house
Have a saving account
Outstanding credit between € 1372 to € 1503
Credits paid back duly

Attributes of Bad Risk Customers:
Age group 0 to 30 and 53+
Customers who live in a rented house
Saving account has debit memo (DM) greater than 1000
Outstanding credit greater than € 9572
Delaying in paying off in the past

References

- https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29
- http://www.sas.com/resources/whitepaper/wp_10961.pdf

Acknowledgements

We thank Dr. Goutam Chakraborty, Professor of Marketing & Founder of SAS and OSU Data Mining Certificate Program at Oklahoma State University for all his support and guidance throughout the project.

SAS® GLOBAL FORUM

April 26-29
Dallas, TX