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

Ayush Priyadarshi¹, Kushal Kathed², Shilpi Prasad³

¹,²,³Graduate Student, Oklahoma State University
Credit Scorecard Generation: An application of the Credit Scoring Node in SAS® Enterprise Miner
Ayush Priyadarshi¹, Kushal Kathed², Shilpi Prasad³
1,2,3SAS & 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.

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 (AUC) 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 AUC 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.

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 make 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)

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

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.

References

- https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29

Fig. 1 Interactive binning of variables

Fig. 2 Manual separation of bins

Fig. 3 SAS Enterprise Miner diagram

Fig. 4 Fit statistics of the credit risk model

Fig. 5 ROC plot

Fig. 6 Cutoff value to segregate good and bad customers