PREDICTIVE MODELING IN CREDIT RISK: SOME COMMON PRACTICES

S.Sarkar*

Department of Mathematics, School of Science and Technology, Nazarbayev University, Astana, Kazakhstan;
*shuchismita.sarkar@nu.edu.kz

INTRODUCTION.

A country's economy is highly dependent on its banking industry. Increased competition, continuous pressure to generate revenue while maintaining a low loss rate has encouraged the financial institutions to venture on different avenues of attracting creditworthy customers and maintaining a healthy portfolio. A credit score is a numerical expression that summarizes the financial behavior exhibited by a person in the past. It is a common practice in most of the leading financial institutions to make use of predictive scorecards (developed in-house or borrowed from external source e.g. credit reporting agency) to assess loan risk. A large range of predictive models have been used for credit scoring. Logistic model is one of them which is widely preferred because it does not make any assumption of normality, linearity, and homogeneity of variance for the independent variables. It also produces real probability which is easy to interpret.

MATERIALS AND METHODS.

In the current study, some common practices for predictive modeling used in the banking industry are discussed. The entire process flow is presented in three sections – data mining, predictive modeling and validation. In each section the industry practice is described using flow charts, inherent mathematical formulae and examples.

RESULTS AND DISCUSSION.

Data mining techniques are used in collecting, cleaning and the initial processing of the data. 80% of the data is randomly selected for building the model and 20% is reserved for validating it. The selected data is then classified into different segments. Segmentation is driven by preliminary analysis and business need. The next step is to study the relationship between the independent and dependent variables for each predictor. Weaker predictors may be discarded at this stage. Transformations of variables are done if needed. After that, stepwise logistic regression is applied on the clean data which eventually produces the model. Model fit statistics are observed. A model that rank orders and displays the best separation from good to bad is considered as the best.

ACKNOWLEDGMENTS.

This study uses some common techniques of Predictive Modeling that I learned while working in the banking industry. I would also like to thank SST for their kind cooperation.

REFERENCES.

1. N. Siddiqi. (2006). Credit Risk Scorecards, John Wiley & Sons.