ABSTRACT

RAKHMAWATI. Credit Risk Modeling Using Logistic Ridge Regression. Supervised by AAM ALAMUDI and DIAN KUSUMANINGRUM.

The growth of credit of national banking may cause a greater risk faced by banks. One thing we must highlight is a way to determine whether the new applicant will be good in loan repayments. A well known and widely used method for classifying the new applicant of credit is Logistic Regression. Multicollinearity is a problem that is frequently encountered in model building. Usually, variable selection method is used for handling this problem. But sometimes it creates a new problem when the important variable does not enter to the model. Logistic Ridge Regression could be an alternative in logistic regression when multicollinearity exists. The advantage of this method is that it can handle multicollinearity without deleting any predictor variables. This research compared the performance of logistic ridge regression and logistic regression with variable selection to predict the collectability status of new applicants of credit. There were 1000 observations of German Credit data set. The 740 observations were used for modeling and 260 observations were used for validation. Backward was the best among other selection variable methods which had the highest c statistic and the model was fit by Hosmer and Lemeshow Goodness-of-Fit Test. By using backward logistic regression, it showed that among 17 variables there were eight variables which were significant in the wald test. There were many significant correlations among the predictors but the highest correlation coefficient was 0.628 which exist between duration of credit (V1) and credit amount (V2). The ridge parameter or $\lambda$ was 0.001. The optimal cut point of backward logistic regression was 0.680, while for logistic ridge regression was 0.677. By comparing the c statistic and the total correctly predicted cases, we can see that the logistic ridge regression was better than backward logistic regression in training data. However, with testing data (validation), backward logistic regression was better. To have a better understanding of the model with higher correlation values between V1 and V2, V2* was generated to replace V2 and logistic regression with variable selection and ridge were also built. The result pointed out that logistic ridge regression has a little higher capability to predict the new applicant’s collectability status than logistic regression with variable selection.

Key words: credit risk modeling, logistic ridge regression, multicollinearity