A Study on Loan Performance Using Data Mining Techniques

E.B.Thisara[1], M.C.Wijegunasekara[2]

Most of the modern financial companies offer loans to customers in order to build up their own business. Such companies have a major problem when they recover the loan as the customers do not pay the installments according to the signed contract. It is crucial to determine/create the appropriate strategies and to identify the risk free customers as there is high potential of non-performing loans. In order to predict the risk factors that affect to non-performing loan, Data Mining techniques were considered. This research discovered the factors/reasons for non-performing loan using the data from a reputed Finance Company.

This research focused on eighteen attributes which were referred to as factors affecting a non-performing loan state and the dataset contained with 30% of test data and 70% of training data from 750 records. Among those attributes eleven key attributes namely: Age, Area, Branch Name, Customer Job, Income, Loan State, Mortgage, Number of Terms, Overdue days, Product Type and Interest Rate were selected to create the data mining models. The considered mining models were namely: Neural Networks (NN), Decision Trees (DT) and Clustering (CL). These models were created using the Business Intelligence tool and the database was created in SQL Server Management Studio 2008R2.

The predicted probabilities (as a percentage) of Neural Networks, Decision Trees and Clustering models were 1.57%, 0.44% and 10.46% for non-performing loan state respectively. As the Clustering Model had the highest value it was chosen as the best algorithm to evaluate loan state by using Microsoft clustering method. The Clustering model was given ten clusters numbered from 1 to 10 and five clusters namely: 3, 6, 8, 9 and 10 were identified as the most inclined towards the non-performing loan state by comparative analysis. The predicted probabilities of selected clusters were 23%, 41%, 32%, 23% and 35% respectively and cluster number 6 showed a highest value and cluster number 10 showed the next highest value. Based on cluster performance, numbers 1, 2, 4, 5, 7 had a high probability of becoming performing loan and thus were not included in the analysis. According to the states of attributes within each cluster profiles Product Type, Customer Job, Mortgage, Income, Number of Terms and Interest Rate were identified and shortlisted as the factors affecting the non-performing loan state most.

The research identified that if the customer is self-employed or individual, a small property owner, or having a low income and depending on the type of mortgage (building, vehicle or non-mortgage) the loan tend to be non-performing. The longer duration for loan repayment or higher interest rates will also cause a loan to be non-performing. According to the above results it can be concluded that the high interest loans provided for the unemployed customers or customers with low income have a higher potential to be non-performing and hence resulting in a monetary loss for the financial company. Therefore a financial company will be able to improve its profits if they are more concerned about such customers and undertake suitable decisions. The model will support the financial sector in identifying the amount of loans that could be transformed into the non-performing state. Therefore the findings of this research will benefit the financial industry to reduce the risk of granting loans when providing loans in future.

Key words: Business Intelligence tool, Data Mining, SQL Server Management Studio 2008R2

[1] Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka. buddhisha.et@gmail.com

[2] Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka. carmel@kln.ac.lk