

The Comparisons of Data Mining Techniques for the Predictive Accuracy of
Probability of Default of Credit Card Clients

葉怡成, Che-hui Lien

Information Management

Computer Science and Informatics

icyeh@chu.edu.tw

Abstract

This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or not credible clients. Because the real probability of default is unknown, this study presented the novel "Sorting Smoothing Method" to estimate the real probability of default. With the real probability of default as the response variable (Y), and the predictive probability of default as the independent variable (X), the simple linear regression result ($Y = A + BX$) shows that the forecasting model produced by artificial neural network has the highest coefficient of determination; its regression intercept (A) is close to zero, and regression coefficient (B) to one. Therefore, among the six data mining techniques, artificial neural network is the only one that can accurately estimate the real probability of default.

Keyword : banking, neural network, probability, data mining.