Modeling Small Business Credit Scoring Using Logistic Regression, Neural Networks, and Decision Trees

Previous research on credit scoring that used statistical and intelligent methods was mostly focused on commercial and consumer lending. The main purpose of this paper is to extract important features for credit scoring in small business lending on a dataset with specific transitional economic conditions using relatively small data set. To do this we compare the accuracy of best models extracted by different methodologies, such as logistic regression, neural networks, and CART decision trees. Four different neural network algorithms are tested, including backpropagation, radial basis function network, probabilistic and learning vector quantization, using forward nonlinear variable selection strategy. Although the test of differences in proportion and McNemar’s test do not show statistically significant difference in the tested models, the probabilistic NN model produces the highest hit rate and the lowest type I error. According to the measures of association the best NN model also shows the highest degree of association with the data, and it yields the lowest total relative cost of misclassification for all examined scenarios. The best model extracts a set of important features for small business credit scoring for the observed sample, emphasizing credit program characteristics, as well as entrepreneur’s personal and business characteristics as the most important.