Reject inference in credit scoring using semi-supervised machine learning algorithms

Credit scoring models are commonly built on a sample of accepted applicants whose repayment and behaviour information is observable once the loan has been issued. However in practice these models are regularly applied to new applicants, which may cause sample bias. This bias is even more pronounced in online lending, where over 90% of total loan requests are rejected. Reject inference is a technique to infer the outcomes for rejected applicants and incorporate them in the scoring system, with the expectation that predictive accuracy is improved. This paper extends previous studies in two main ways: firstly, we propose new methods involving machine learning to solve the reject inference problem; secondly, the Semi-supervised Support Vector Machines and k-prototype fuzzy c-means models are found to improve the performance of scoring models compared to the industrial benchmark of logistic regression based on real unsecured consumer loan data.