

Abstract

The aim of this study is to model the probability of a customer to attrite/defect from a bank where, for example, the bank is not their preferred/primary bank for salary deposits. The termination of deposit inflow serves as the outcome parameter and the random forest modelling technique was used to predict the outcome, in which new data sources (transactional data) were explored to add predictive power. The conventional logistic regression modelling technique was used to benchmark the random forest's results.

It was found that the random forest model slightly overfit during the training process and loses predictive power during validation and out of training period data. The random forest model, however, remains predictive and performs better than logistic regression at a cut-off probability of 20%.