Chapter 12
Credit Risk Analysis and Prediction

Akhil Raman Sarepalli
Vellore Institue of Technology, India

Maheshwari S
Vellore Institue of Technology, India

ABSTRACT

Credit scoring is used to divide applicants into two groups: those with good credit and those with bad credit. When a bank gets a loan request, borrowers with strong credit have a high likelihood of repaying debt. The likelihood of default is higher for applicants with bad credit. The profitability of financial organisations depends on the accuracy of credit scoring. Financial institutions will experience less of a loss if their credit scoring of applicants with poor credit is even 1 percent more accurate. This study seeks to solve this categorization issue by examining the risk of granting a loan to the applicant using the applicant's socioeconomic and demographic attributes from German credit data. In terms of overall accuracy, the authors evaluated the efficiency of several ML techniques like decision tree, logistic regression model, neural network, SVM, as well as random forest. The authors compared and evaluated several models for the model optimization process, integrating the impacts of balancing the AUC (area under the ROC curve)and accuracy values.

INTRODUCTION

Motivation

Credit granting to consumers is an important operation of the banking sector. Credit risk management assesses the information that is available and determines a customer's reliability to protect the financial institution from fraud.

DOI: 10.4018/978-1-6684-9804-0.ch012

Datasets

A collection of datasets helpful for assessing machine learning algorithms may be found at the UC Irvine Machine Learning Repository.The German credit data and another dataset from UCI - ML 1 repository were both used for this study. The Australian dataset comprises 14 characteristics and 690 instances, compared to 20 features and 1,000 examples in the German sample. A customer's level of credibility is the response variable, which is a binary decision. The loan's purpose, the credit score, and client information are some of the shared attributes between the two datasets -age, occupation, account duration, and salary.

Credit Risk

A critical phase of knowledge discovery is data mining, which looks for patterns in data using theories, processes, and tools. In order to ensure that tools and procedures are correctly corresponded to the data and the purpose of pattern recognition, it is essential to understand the fundamental principles of the methodology. There may be a number of tool choices for a certain data collection. When a bank receives a loan application, the bank must decide based on the applicant's profile whether to authorize the loan or not. The bank's decision includes two different categories of risks:

- If the borrower has good credit risk or is likely to pay back the loan, then the bank loses business by rejecting the loan application.
- The bank can experience a loss by making the loan if the customer has a poor credit risk, which means they are unable to pay back the debt.

OBJECTIVE AND SCOPE OF PROJECT

Objective

- The bank needs a decision rule that specifies who gets to approve loans and who doesn't to reduce loss from the bank's standpoint. Before deciding on a borrower's loan application, loan managers take into account the applicant's socioeconomic and demographic data.