Analysis of Credit Scoring using Machine Learning Techniques

Abstract

Credit Risk Assessment should be done by banks and financial institutions in order to make decisions like grant credit to new applicants, extend credit or increase credit limit for their customers existing customers using a technique called **Credit Scoring Model.** Credit scoring has become a very important task as the credit industry has been experiencing severe competition during the past few years. Credit scoring can be assumed as the assessment of risk associated with crediting to an organization or an individual entity. Thus, credit scoring is a very typical Data Mining (DM) classification problem. It uses numerical tools to rank order cases using data integrated into a single value that attempts to measure risk or credit worthiness. The decision-making   process for credit scoring can be either subjective or statistical.

Earlier, several statistical methods had been used such as **LDA** (Linear Discriminate Analysis) and **LR** (Logistic Regression) which only sets up linear relationship between the response and predictor variables; which are usually non-linear variables. The objective of the proposed study is to explore the performance of credit scoring using two commonly discussed data mining techniques—decision tree and random forest . To demonstrate the effectiveness of credit scoring, credit scoring tasks are performed on German Credit Dataset which is a real world dataset problem . As a result, the performance of the proposed models will be compared in terms of accuracy.