Abstract

Data mining is an emerging area of research that aims at extracting meaningful patterns from available data. This paper highlights the significance of classification in predicting new trends from voluminous data. Performance analysis of various data mining algorithm viz. BayesNet, Meta-Stacking, Naïve Bayes, Random Forest, SMO and ZeroR in predicting credit-card defaulters is discussed in this paper. Dataset from the UCI machine learning repository comprising of 25 attributes and 30000 instances have been employed to analyze the performance of algorithms. Moreover, the effect of feature selection has also been identified with respect to each classification algorithm. It has been concluded from the experimental results that both Correlation Feature Subset and Information Gain feature selection methods yield the most useful features for prediction and the accuracy of Random Forest Ensemble method is highest in predicting credit card defaulters.
1. Han, Jiawei, and Micheline Kamber. "Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems)." (2000).
3. Dataset from University of California, Irvine available in their online repository http://archive.ics.uci.edu/ml/index.html.
7. Investopedia is an online encyclopedia relating to investments. http://www.investopedia.com/terms/c/credit.asp

Index Terms

Computer Science Information Sciences
Keywords

Credit card defaulter, datamining, patterns, Knowledge patterns.