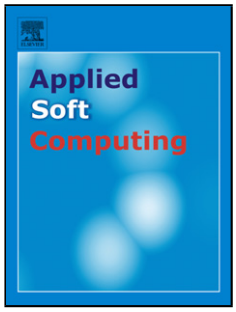

Title: Information Gain Directed Genetic Algorithm Wrapper Feature selection for Credit Rating

Authors: Swati Jadhav, Hongmei He, Karl Jenkins

PII: S1568-4946(18)30224-2
DOI: https://doi.org/10.1016/j.asoc.2018.04.033
Reference: ASOC 4837

To appear in: *Applied Soft Computing*

Received date: 22-7-2017
Revised date: 29-1-2018
Accepted date: 13-4-2018

Please cite this article as: Swati Jadhav, Hongmei He, Karl Jenkins, Information Gain Directed Genetic Algorithm Wrapper Feature selection for Credit Rating, Applied Soft Computing Journal https://doi.org/10.1016/j.asoc.2018.04.033

Information Gain Directed Genetic Algorithm Wrapper Feature selection for Credit Rating

Swati Jadhav, Hongmei He and Karl Jenkins

School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK

{s.jadhav,h.he, k.w.jenkins@cranfield.ac.uk}

Highlights

- The problem of credit scoring is addressed as a classification and feature subset selection problem.
- A novel algorithm "Information Gain Directed Feature selection algorithm (IGDFS)" is proposed, which performs the ranking of features based on information gain and uses three classical machine learning algorithms of SVM, KNN and Naïve Bayes for credit scoring.
- The average prediction results by IGDFS, Genetic Algorithm Wrapper and Baseline classification models are compared for three public credit datasets.
- The classification accuracy achieved with different feature selection strategies is highly sensitive to the type of data, total number of samples and the number of positive and negative samples in the dataset.
- There is a potential for improvement in the models' performances if the feature selection method is chosen carefully.

A B S T R A C T

Financial credit scoring is one of the most crucial processes in the finance industry sector to be able to assess the credit-worthiness of individuals and enterprises. Various statistics-based machine learning techniques have been employed for this task. "Curse of Dimensionality" is still a significant challenge in machine learning techniques. Some research has been carried out on Feature Selection (FS) using genetic algorithm as wrapper to improve the performance of credit scoring models.