

IT in Finance Assignment 1b

Submission deadline and oral presentation: January 9, 2019

The purpose of this assignment is to understand the data mining process using classification and/or clustering algorithms on real-life data. As a data mining tool, the Orange package is suggested; other software packages such as STATISTICA, WEKA, SAS, R, ... may be also used.

The experiments may be carried out on the dataset Credit Rating Data available on my Website. The dataset contains information about the past customers of an institution and their applications for credit. These customers are classified as either “Bad” or “Good” credit risk based on their loan history with this institution.

In this project you should:

- 1) Analyze, review, and if needed preprocess the data,
- 2) Determine the attributes that are best predictors of credit rating,
- 3) Evaluate and find a high performance predictive model that classifies/clusters customers,
- 4) Discuss your results.

Students are free to use another datasets (for ex. German Credit Data, Credit Rating Data–US, Australian Credit Approval, Japanese Credit Scoring, private data, etc.). Some financial datasets can be downloaded from: <http://archive.ics.uci.edu/ml/machine-learning-databases/> .

Instructions and hints:

- A. Install and become familiar with the use of the Orange package (<http://orange.biolab.si/>)
- B. Preprocess and analyse the dataset
- C. You may use and test various classifiers or clustering algorithms available in Orange
- D. Submit a report-presentation as zipped file named **ITF-name.zip** by email to: jerzy.korczak@ue.wroc.pl

The report is a comprehensive presentation of your experiments using MS PowerPoint. You should include the following: experiment description, data analysis, key results (decision tree, rules, performance ratios, clusters) and conclusion of the experiments.

- E. Projects will be orally presented by the authors; approx. 10 min per project.

References:

- Lecture notes “IT in Finance” by J. Korczak (<http://www.korczak-leliwa.pl>)
- Doc Orange, <http://www.biolab.si>
- Lyn C. Thomas, *A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers*, <http://www.sciencedirect.com/science>
- Chuang C.L., Lin R.H., *Constructing a reassigning credit scoring model*, <http://www.sciencedirect.com/science>
- Hand D.J., Heney W.E., *Statistical Classification Methods in Consumer Client Scoring : A Review*, J. R. Statist. Soc., pp.523-541, 160, 1997.