

## Mining in Financial Databases

### Assignment 2

**Submission deadline and oral presentation: March 27-28, 2017**

The aim of the project is to master the methods of knowledge discovery from financial databases using neural networks and data mining. Your task is to develop a model of classifying clients to appropriate risk groups on the basis of historical data and their financial situation. This model will be applied in assessing the credit worthiness of the customer while granting credit, buying insurance and determining the strategy of debt collection.

The model is to classify customers into one of the following groups of risk assessment of customer, namely: Low (L) - is a low risk, medium (M) - the risk Medium, High (H) - high risk, high Plus (HP) - risk very high. On the basis of the experiment you should propose risk classification rules and the most appropriate strategy for debt collection (procedure) for a given debtor. Model assessment of the financial situation can also be used for interpretation / explanation of recovery decisions based on risk assessment of the financial situation of the debtor.

The analysis can also use other data mining platforms, such as Orange, R, Weka, Statistica, SAS, ... In order to classify the data you may take advantage of the available algorithms (eg. in the tab Classify Orange), such as decision trees, MLP, naive Bayes, CN2, ... You may also use the platform Pathfinder's VSoft.

The presentation of the project should comment on classification accuracy of decision rules (or neural network), and visualize the data mining process and the results. It should also assess the quality of the applied classifier.

Due to the constraint of the VSoft license, the platform and data will be delivered personally.

#### Hints:

- A. *Become familiar with the use of Pathfinder (see Manual)*
- B. *Use Excel to transform raw data*
- C. *Install and become familiar with the use Orange (<http://orange.biolab.si/>).*
- D. *Edit and present the results in Power Point, and submit your compressed presentation as file **DMF-nazwisko.zip** to: [jerzy.korczak@ue.wroc.pl](mailto:jerzy.korczak@ue.wroc.pl)*  
The presentation should include the following information and data: experiment description, description of features, data sets used, data mining schema, visual representations of the results, conclusion of the experiments.
- E. *Oral presentation has to be given by all authors of the assignment; 10min per project.*
- F. *The assignment can be done by max two students*

#### References:

- Lecture notes (<http://www.korczak-leliwa.pl>)
- Pathfinder Manual (<http://www.korczak-leliwa.pl>)
- Orange Tutorial and documentations