

Mining of Financial Database Assignment 1b

Submission deadline and oral presentation: March 18, 2018

The aim of the Assignment is to master the skills of the use of financial and economic databases in data analysis projects. Your task is to discover the clusters of the IT companies in Lower Silesia based on selected financial indicators and evaluate/interpret the clusters.

The data can be found in the EMIS database (<http://www.bg.ue.wroc.pl/resources>). The selection of the indicators should be argued. Please note that to do this part of the work, you should use the computers or your laptops ONLY in the UE library.

The second part of the work should be done using the Orange data mining platform. To process the extracted data you should first transform them and store into native Orange's data format ".tab". Native format starts with feature (attribute) names, their type (continuous, discrete, string). Note that Orange can read files in native and other data formats (.txt, .csv,...).

To cluster the data use one of the available algorithms such as Hierarchical clustering, k-Means, SOM. Visualize and comment your results of clustering.

Instructions and hints:

A. *Become familiar with the use of the EMIS database. The language changes in "quick links", and the "account settings". Indicate "Company", click "Quick Screener" - in "Industries" choose industries (there you should probably look at range 517-519, but this is just a suggestion), enter region "Lower Silesia" (dolnośląskie). In the preview you'd check the criteria - then just click on the "Search" and you'll get your table - then you'll personalize it ("Customize Columns") to have all the data you want. At the end, you should click "Export" and choose export options*

B. *Use Excel or Orange to transform the raw data extracted from the EMIS database*

C. *Install and become familiar with the use of the Orange software (<http://orange.biolab.si/>).*

D. *Submit a report-presentation as zipped file named **MFD-your-name.zip** by email to: jerzy.korczak@ue.wroc.pl*

This is a comprehensive presentation of your work using MS PowerPoint. You should provide the following information and data: experiment description, description of indicators (attributes), data sets used, data mining schema, visual representations of the results, conclusion of the experiments.

E. *Oral presentation has to be given by selected students; ca 10 min per project.*

References:

- Lecture notes "Mining of Financial Databases" (<http://www.korczak-leliwa.pl>)
- User guide (<https://static-emis.emis.com/docs/EMISPRO%20User%20Guide.pdf>).
- Orange Tutorial and documentations (<http://orange.biolab.si/>)