

IT in Finance Assignment 3 (option A)

Submission deadline and oral presentation: **January 18, 2017**

The aim of the Assignment 3A 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 EU countries, based on selected macroeconomic indicators, and comment the position of Polish economy in 2014-2016. To select and extract the data you might be inspired by an example of the dataset available on my Web page. However in case of reuse, the dataset should be extended in terms of the number of indicators.

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

To download the statistics you may also use the EU Open Data Portal (<http://open-data.europa.eu/en/apps>).

The second part of the work should be done using the Orange data mining platform. To process the extracted data you should first transform it and store into the native Orange's data format ".tab". The 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 apply one of the available algorithms such as Hierarchical Clustering", k-Means, SOM. Visualize and comment your results.

Instructions and hints:

- A. *Become familiar with the use of the Passport database*
- B. *Use Excel to transform the raw data extracted from the Passport database*
- C. *Use of the Orange software (<http://orange.biolab.si/>).*
- D. *Submit a report-presentation as zipped file named **ITF-your-name.zip** by email to: jerzy.korczak@ue.wroc.pl*

This report should be a comprehensive presentation of your work using MS PowerPoint. You should provide the following information and data: experiment description, description of indicators (categories), datasets used, data mining schema, visual representations of the results, conclusions of the experiments.

- E. *Oral presentation has to be given by all authors of the project; 10 min per project.*

References:

- Lecture notes "Mining of Financial Databases" (<http://www.korczak-leliwa.pl>)
- Passport User Guide, docs and tutorials
- Orange Tutorial and documentations