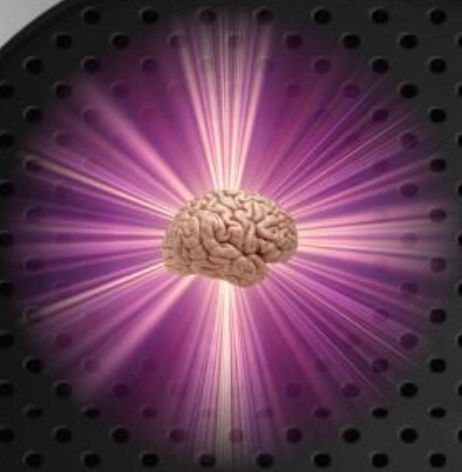
AGH University of Science and Technology
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
Department of Biocybernetics and Biomedical Engineering



Knowledge-based CI and ML in Biomedicine

Data Mining Laboratory Classes

Adrian Horzyk
horzyk@agh.edu.pl

Google: Adrian Horzyk

Data Mining Assignments

- ✓ **Download any dataset of biomedical data (vectorized or sequential) where you can search for frequent patterns or association rules.**
- ✓ **Make yourself familiar with the data mining library SPMF containing more than two hundred data mining and knowledge exploration methods:**
  - **association rule mining,**
  - **itemset mining,**
  - **sequential pattern**
  - **sequential rule mining,**
  - **sequence prediction,**
  - **periodic pattern mining,**
  - **episode mining,**
  - **high-utility pattern mining,**
  - **time-series mining,**
  - **clustering and classification.**

Data Mining Assignments

- ✓ Implement one of the methods (Apriori, Eclat, FP-tree) presented during the lecture or reuse (using wrappers for Python) a selected method from the data mining library.
  You can also use any other Python library or methods built in Python after your choice!
- ✓ Use the implemented or adapted method(s) from the library to find:
  - ➢ all frequent patterns in the dataset meeting the defined criterium of the minimum support;
  - ➢ all closed and maximum patterns in the dataset meeting the defined criterium of the minimum support;
  - ➢ all associative rules meeting the defined criteria of the minimum support and the minimum confidence;
- ✓ Present your final results (found patterns and associative rules) and check whether the computed results are correct.
- ✓ Send the code and the results as an answer to the assignment in MS Teams.

Literature and Bibliography

1. Robert Layton, Learning Data Mining with Python, Second Edition, O'Reilly.
2. Robert Layton, Learning Data Mining with Python (on-line), Second Edition, O'Reilly.
3. Useful Python libraries for CI and DM.
4. Python libraries and modules for data mining:
    1. Numpy
    2. Scipy
    3. Pandas – Python Data Analysis Library
    4. Scikit-learn