

# Data Analysis 3

## A course introduction

---

Jan Platoš

September 12, 2020

Department of Computer Science  
Faculty of Electrical Engineering and Computer Science  
VŠB - Technical University of Ostrava

1. Associative Pattern and Rule Mining
2. Explorative analysis
3. Clustering algorithms and principles
  - 3.1 Representative based clustering
  - 3.2 Hierarchical clustering
  - 3.3 Density based clustering
  - 3.4 Clustering validation
4. Anomaly detection
5. Classification
  - 5.1 Feature selection
  - 5.2 Probabilistic Classification
  - 5.3 Decision trees
  - 5.4 Rule-Based Classification
  - 5.5 Linear discriminant analysis
  - 5.6 Support Vector Machine
6. Regression

- Aggarwal C.C. (2015), Data Mining: The Textbook, Springer.
- Bramer, M. (2013). Principles of data mining. Springer.
- Leskovec, J., Rajaraman, A., Ullman, J. D. (2014). Mining of massive datasets. Cambridge University Press.
- Witten, I. H., Frank, E. (2011). Data Mining: Practical machine learning tools and techniques [3rd Ed.]. Morgan Kaufmann.
- Zaki, M. J., Meira Jr, W. (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press.

# Course Evaluation - Full-time students

- Exercise tasks finishing (5 to 10 points):
  - up to 2 points for finishing task (small implementation/analytical tasks) at the exercise or at home (1-2 hours of work).
- Exploratory analysis and clustering (15 to 30 points):
  - Selection of the dataset, exploratory analysis and clustering analysis on it. (2-4h)
  - A PDF report + source codes or a Jupyter Notebook with description.
- Data classification analysis (15 to 30 points):
  - Analysis of the selected dataset with standard algorithms (2-4h).
  - A PDF report + source codes or a Jupyter Notebook with description.
- A method implementation (15 to 30 points):
  - Implementation of a selected method we have mentioned/used during the course or any other (2-4h).

Questions?