

# Data Analytics Course 2014

## Katharina Morik (TU Dortmund University)

6 sessions of in-depth data analysis methods and tools

18.3. - 25.3.2014

First meeting, 18.3.2014 at 14 o'clock

Last meeting, 25.3.2014 ends before 14 o'clock

The availability of data from scientific experiments, from computer services, online sales, machine measurements in production and logistics allows us to obtain insight into complex process and make good use of it. In almost every subject, analyzing data has become a key issue. Under the heading of Big Data, the importance of handling data efficiently and effectively been stressed even more. Theories and methods from statistics, machine learning, data mining, and databases in concert are needed to master the challenges of data analytics.

The course introduces Data Analytics to students from all faculties. It explains the process of Data Analytics and illustrates each step by examples from real-world applications. Using RapidMiner, the world's most popular open source tool for Data Analytics, hands-on experience is offered. Methods and models are taught, thus providing participants with some theoretical background. Hence, participants will become able to form their own first processes of Data Analytics and understand more complex analytics.

Students will receive a script.

Hands-on experience with RapidMiner is easy. The community edition is free:  
<http://rapidminer.com/download-rapidminer/>

It is recommended that students download the RapidMiner tool and bring their laptops to the sessions.

### Content

#### Data Mining Basics

- The data mining process
- Learning Tasks
- RapidMiner introduction

#### Data Mining Models and Methods

- Additive Models: Decision Trees, Random Forest
- Local Models: k-Nearest Neighbors
- Linear Models: Regression, Support Vector Machine
- Clustering

#### Getting Data in

- Import
- Types and Roles
- Transformations and Normalizations
- Handling Missing Data

- Feature Selection
  - Inspecting Data
- Validating Results
- Cross Validation
  - Accuracy, precision, recall

### **Further Reading**

Matthew North (2012) *Data Mining for the Masses*. Global Text Project Book  
Markus Hofmann, Ralf Klinkenberg (eds) (2014) *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. Chapman&Hall/CRC Press  
Trevor Hastie, Robert Tibshirani, Jerome Friedman (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Verlag