Data Analytics and Machine Learning with Applications in KNIME
(taught by Prof. Martin Huber)

This course provides an introduction to data analytics and machine learning, including applications in the easy-to-use software "KNIME".

Data analytics aims at analyzing data to detect informative patterns and draw conclusions about specific processes (like customer behavior, production, turnover…). One subfield is (supervised) machine learning, which aims at predicting an outcome in the future, e.g. sales, based on analyzing patterns of potentially relevant factors (or "predictors") like price, quality, etc. in past data. This course discusses important tools for prediction (such as regression and decision trees), focusing on the intuition of the various methods rather than technical details.

The tools are applied to the analysis of economic and business-related data in "KNIME", which is an open source software that is free of charge. "KNIME" is based on a graphical interface that represents all analytical steps by an intuitive flow chart and does not require programming skills. "KNIME" is therefore ideal for those interested in applying data analytics and machine learning without learning a programming language.

Objectives
- To understand the idea and goals of data analytics and machine learning
- To understand the intuition, advantages, and disadvantages of alternative methods
- To be able to apply the methods to real-world data using the software "KNIME"

Content
- Introduction to the concept and purpose of data analytics and machine learning
- Linear and non-linear regression (OLS, logit regression)
- Penalized regression for variable selection and shrinking (lasso and ridge regression)
- Tree-based approaches (trees, bagging, random forests)
- Model tuning (cross-validation)
- Performance evaluation (out-of-sample testing)
- Application of all methods to commercial data using the statistical software "KNIME"; course participants may bring along their own data sets

Preconditions
- None (introductory course). Basic knowledge of statistics is an advantage, but not required.

Duration
- 1 day on Feb 11th (roughly 7*45 minutes)

Evaluation
- take home exam: project work to be solved in KNIME

ECTS
- 0.5