Data Mining Cup 2024

Statistics: Jonas Rieger, Steffen Maletz

Computer Science: Emmanuel Müller, Michel Lang, Simon Klütermann
DMC 2024

- This is an on-site course
  - Max. 12 participants from Statistics department and max. 12 from Computer Science
- Predictive modeling competition
  - Training dataset + unlabeled test data for prediction.
  - Optimize against specified quality measure
Statistical Methods

- EDA (Explorative Data Analysis)
- Preprocessing (Imputation, ...)
- Resampling and Evaluation
- Discriminant Analysis
- Nearest Neighbours
- Trees and Forests
- Support Vector Machines
- Regularized Linear Models
- Gradient Boosting
- Neural Networks
- Hyperparameter optimization
- Feature Selection
- Feature Generation
- Ensembles and Stacking
- [...]
Course Plan

- TBA: Start of competition, release of data and task
- During lecture period: Regular meetings (2 per week), active participation
  - Tuesday and Thursday, each 10:15 - 11:45, CDI 121
- TBA: End of competition, upload of predictions for test data
- August 31: Final Report (~ 25 pages)
Requirements

- Familiarity with data analysis tools like Python/sklearn, R or Julia

- Master Statistik: Fallstudien I (recommended)
- Master Econometrics: Minor Introductory Case Studies
- Master Data Science:
  - All requirement courses (Introductory Case Studies, ...) must have been passed
  - Advanced Statistical Learning is recommended to be passed

- Computer Science: Big Data Analytics (recommended), Mathematics Courses
Examination **Statistics/Data Science**

- active participation in competition and discussions
- poster session at the end of the competition
- final report (~ 25 pages, we will announce specific formalia for this report at the end of the competition) - deadline: August 31 – no extension!
Examination Computer Science

- **active participation in competition and discussions**
  - initiative for open tasks
  - imagination for what could be useful tasks
  - take and fill necessary roles in team
  - think both in and beyond your team

- **poster session**
  - explanation of task, teams and your role in the DMC
  - outline how your team’s process going from early to later solutions
  - explain team’s contributions to the final solution