

Data Mining Cup

Data Mining Cup

General information

Teacher	Michel Lang and Markus Pauly
Module requirements	<i>For Data Science students: If there are requirements in case of conditional admission (any of modules MD Req1 to MD Req8), these have to be fulfilled by the start of the course "Case Studies" or the external internship.</i>
Place and Time	<i>TBA, probably Thursdays, 10am – 2pm, SkyLounge@LogistikCampus</i>
Language	<i>English</i>
First date	<i>April, 16</i>

Binding registration by March 15 via mail to sia.statistik@tu-dortmund.de

Data Mining Cup

- This is an on-site course
 - 10 participants from Statistics Department and 6 from CS
 - with active participation
- Equivalent to
 - Advanced Case Studies (Econometrics)
 - Case Studies (MD4, Data Science)
 - Fallstudien 2 (M.Sc. Statistics)
- Competition based on a current data analytical problem from REMONDIS
 - complex data from recycling plants
 - a very good solution may get a prize 😊

Data Mining Cup

Potential 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
- Time Series Models
- Optimization
- Feature Selection
- Feature Generation
- Ensembles and Stacking

Potential Software

- Python or R
- For ML:
 - General ML frameworks like mlr3 (R) or scikit-learn (Python)
 - Modern packages, e.g. ranger, xgboost, fable, glmnet, sklearn, sktime, prophet etc.

Data Mining Cup

Examination

- Active participation in discussions
- Final Presentation – Date: TBA
- Final Report (~20-30 pages)
We will announce specific formalia for this report at the end of the competition) Deadline: August 31
– no extension!