Machine Learning Seminar

Prof. Dr. Markus Weinmann, Christopher Coors, Sercan Demir

University of Cologne

Agenda of the Course

- Kickoff April 08 - 14:00
- 1st check-in presentation (5 min) with tutor April 22 - 17:00
- 2nd check-in presentation (presentation: optional) with tutor May 03 - 16:00
- Submission of video, code, and report May 17 at 17:00

Course Setup

- 2 steps major steps in this seminar:
  - Students choose their topic and research for suitable data
  - Students show their understanding of the modelling techniques by writing up a paper and perform a final presentation in video format

Course Assessment

Grading	Description	Weight
Code	R or Python	25%
Report	Academic Paper	50%
Report	Video	25%

Course Assessment

- You write up your analysis in R or Python. Please comment your code, that it will be understandable. Also the quality of data analysis and scope of the chosen models will be graded.
- Teams will submit a 10-page written academic paper detailing their research context, research gap, data description, models, results, conclusion. Graded will be the effectiveness of communication and presentation of results in a scientific way.
- Teams will submit a video with a short 2 min. presentation, that is hold by everyone from the group. Graded will be the ability to work in a team and contribute to the project and also the adherence to project requirements and deadlines.

Data Sources

- Google dataset research: https://datasetsearch.research.google.com/
- UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets.php
- Kaggle: www.kaggle.com
- Microsoft Research Open Data: https://msropendata.com/

Example Projects

- Applying a FaceDetector to Images on Crowdfunding Platforms and choose a target variable that you want to analyze (e.g. average pledges).
- Tensor Decomposition/ Completion to predict solar activation on PV (https://www.nrel.gov/grid/solar-power-data.html)
- Predicting sales/ratings/conversions (depending on the context) from a set of features

Next Steps

- Explanatory data analysis (descriptives, best in visual and/or tabular format depending on your problem statement)
- Predictive modeling: A comparison of different machine learning methods in terms of performance. Elaborate on your choice of a final model: justify why your problem statement should be best by a interpretable or black-box model.
- Findings: Write up the report as proposed in previous slides
- Please list your participants clearly on the front