Machine Learning in R

Michel Lang
LMU München

Jakob Richter
TU Dortmund
Why a Machine Learning Framework?

- New algorithms are mostly implemented in C(++) with bindings to R/python/julia
- Frameworks unify the API, crucial for comparison of many algorithms
- Popular Frameworks:
  - Python: sk-learn
  - Julia: mlj
  - R: caret (no longer maintained), tidymodels, mldr
  - Rapidminer
  - Weka
  - ...

MLR
mlr

- Abstraction for machine learning in R:
  - > 100 Learners, > 30 performance measures, > 30 variable selection filters
  - classification, regression, clustering, survival, multiclass, anomaly detection, functional data analysis
  - tuning, visualization, pipelines, parallelization
  - Addons for Bayesian Optimization, AutoML, XAI,
- Popular and stable, recommended by NIST
- > 20 Contributors, 8 GSOC projects
- Reboot started by MCML (mlr3) and ATI (mlj)
- Object orientation
- Modularization into multiple packages
- Native support for big data
- Full parallelization support (from threading to distributed computing on HPCs)
- First released version includes:
  - objects for all building blocks: tasks, learners, measures, resamplings, ...
  - parallel resampling and benchmarking
  - hyperparameter tuning in mlr3tuning
  - nested resampling
  - data flow programming as DAGs in mlr3pipelines
mlr3 is for scientists

- Completely reproducible results
- Compare many algorithms with only a few lines of code
- Interpretable machine learning (XAI)
- Publication-ready yet customizable plots
- Construct custom pipelines
- Extend objects for your domain as needed
mldr3 is for scientists

- Completely reproducible results
- Compare many algorithms with only a few lines of code
- Interpretable machine learning (XAI)
- Publication-ready yet customizable plots
- Construct custom pipelines
- Extend objects for your domain as needed
mlr3 is for scientists

- Completely reproducible results
- Compare many algorithms with only a few lines of code
- Interpretable machine learning (XAI)
- Publication-ready yet customizable plots
- Construct custom pipelines
- Extend objects for your domain as needed