Institut zur Qualitätsentwicklung im Bildungswesen
Forschungsdatenzentrum

WISSENSCHAFTLICHE EINRICHTUNG DER LÄNDER
AN DER HUMBOLDT-UNIVERSITÄT ZU BERLIN E.V.

Workshop:	Machine Learning and Interpretable Machine Learning with R
Instructor:	Prof. Dr. Mirka Henninger (Universität Basel / University of Basel)
Date:	Tuesday, 27.02.2024, 9:00 – 17:00 h

Abstract

The aim of this course is to provide the audience with 1) an overview of the variety of machine learning (ML) methods as well as the general principles of ML, and what distinguishes it from classical statistics, 2) a deeper understanding of two of the most widely used families of ML methods: ensemble methods (bagging, random forests, boosting) and neural networks, 3) an overview of graphical and numeric means for enhancing the interpretability of ML results, as well as caveats in their interpretation, and 4) how to do all this in the free, open source software R.

At the end of the course participants will understand key principles of ML and interpretable ML, be able to apply several widely used ML methods in R, know where to be careful not to mis- or overinterpret ML results, and be able to judge if and how ML methods could contribute to their own research.

Inhalte /Course Content.

- General introduction to machine learning and general principles of machine learning
- Detailed introduction to two of the most widely used families of machine learning methods:
  - ensemble methods (bagging, random forests, and boosting), including their construction principles and properties
  - neural networks, with a focus on single hidden-layer, feed-forward networks, the role of activation functions, and parameter tuning
- Presentation of graphical and numeric interpretation techniques from the field of Interpretable Machine Learning (IML) to assess importance, shape, and interaction effects of predictor variables, including a discussion of potential risks of misinterpretation
- Practical exercises to apply the presented methods in the free-open-source software R.

Voraussetzungen / Intended Audience.

Everyone interested in machine learning and interpretable machine learning is welcome to this course, from student to professor. Previous experience with R is a plus, but the course materials and presenters are prepared in a way that makes it possible to follow even for R novices. The workshop will be held in English, unless everyone on site agrees that the workshop should take place in German (but the materials will still be in English).

Literatur / Literature.

General text books:

James, G., Witten, D., Hastie, T., & Tibshirani, R. (Hrsg.). (2021). *An introduction to statistical learning with applications in R*. Springer.
https://link.springer.com/content/pdf/10.1007/978-1-0716-1418-1.pdf

Chollet, F. & Allaire J.J. (2018). *Deep Learning with R*. Manning.
https://www.manning.com/books/deep-learning-with-r

Journal (tutorial-style) articles:

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*, 323–348. https://doi.org/10.1037/a0016973

Urban, C. J., & Gates, K. M. (2021). Deep learning: A primer for psychologists. *Psychological Methods, 26*, 743–773. https://doi.org/10.1037/met0000374

Henninger, M., Debelak, R., Rothacher, Y., & Strobl, C. (2023). Interpretable machine learning for psychological research: Opportunities and pitfalls. *Psychological Methods.* https://doi.org/10.1037/met0000560

Software.

The participants bring their own laptop with R and RStudio installed. They will receive detailed installation instructions before the course.