Call for Contributions

Multi-Analyst-Study and Hackathon on

Interpretable Machine Learning for Longitudinal and Clustered Data

Data with stochastic dependencies is common in many fields, including education, psychology, social sciences, and health sciences. Two common sources of dependencies are repeated measurements of the same individuals in a **longitudinal design** (e.g., panel data) and data with noticeable **cluster-level effects**, like cross-sectional data with students nested in schools. Both designs are important to understand associations and causal mechanisms. Data analysis is changing, since current practices in data collection yield many variables. This creates **opportunities** like the study of dynamic behavior or analyses comparing many competing theories, but poses **challenges** in variable selection, feature estimation, and causal effect estimation.

Supervised machine learning (ML) includes powerful prediction models. While their predictive performance is sometimes astounding, many ML approaches are opaque and hard to interpret. In addition, common ML models like the random forest are not designed for dependent observations, as all observations are weighted equally by default. The big interest in ML in the social and behavioral sciences calls for adaptations of existing approaches and innovative techniques, overcoming inherent limitations with respect to dependent data.

At the same time, research and applications need to go beyond mere pragmatic predictive performance. **Interpretability is a key requirement for deepened understanding**. Interpretability might be either in-built into a method (including regularized regression approaches or CART) or obtained post-hoc with techniques like feature importance measures or partial dependency plots. Interestingly, many examples show that interpretability and performance are not necessarily at odds, and especially in high-stakes applications, it will often be beneficial to employ methods with in-built interpretability.

We cordially invite all interested academic researchers to contribute to our multi-analyst study and the related in-person Hackathon in Dortmund, Germany, hosted by the interdisciplinary research center Agile Prevention and Intervention Research (Agile PAIR) (agile-pair@tu-dortmund.de).

Aims
- Discuss **recent developments in prediction modelling of dependent outcomes** and trajectories.
- Compare approaches on two datasets with **a multi-analyst** approach.
- Derive and publish **authoritative guidelines** for research with supervised machine learning with stochastic dependencies

Timeline
- March 23rd, 2026: Publication of the Call for Contribution
- April 20th, 2026: Deadline for brief outlines (max. 300 words)
- April, 27th, 2026: Notice of acceptance; data sharing will happen after completed pre-registrations
- June, 18th, 2026: Deadline for multi-analyst study pre-registration and brief analysis plans (via osf.io)
- July, 2nd, 2026: In-person workshop in Dortmund, Germany
- September 30th, 2026: Results finalized (end of analysis phase); start of guideline phase
- December 20th, 2026: Submission of multi-analyst study and guideline paper

Examples of methods within scope
- classification and regression trees and related methods
- regularized regression models, regularized structural equation models, and other statistical learning models
- well-explained black box models, including (tree-)ensembles and (deep) neural networks
- functional/additive models
- surrogate modelling and explainability techniques

Datasets
Organizers provide datasets with documentation. Participants are required to apply at least one method of their choice to at least one of the datasets and share their findings. All contributions provide **an analysis plan to be pre-registered** with the outline of the data-analytic approach.

Dataset 1: National Educational Panel Study (NEPS)
NEPS provides longitudinal data on educational trajectories in a broad sense. Participant age at recruitment ranges from newborns to older adults. It comprises several panel studies investigating educational processes in Germany. More than 70,000 participants have been surveyed about their educational activities, and their competence development has been measured since 2009. We thank Dr. Elisabeth Graf for communicating the cleaned NEPS dataset.
- **Goal:** Can trajectories of motivation and interest predict the uptake of a university degree or professional degree? This includes two supervised classification problems with binary target variables: (1) Indicator of transition to a university degree, (2) Indicator of transition to a professional degree.
- **Data** are from NEPS' third starting cohort (SC3) that sampled 5th graders in 2009 and followed many of them until 2022. Dependencies are on the individual level, and there are also cluster effects. Among others, variables include scales for motivation, interest, personality, and student background variables. Imputations of missing data will be provided.
- **Evaluation of generalizability** is on a holdout sample of 20% of the data. The AUC is the main metric for both target variables.

Dataset 2: Progress in International Reading Literacy Study (PIRLS)
PIRLS is an international large-scale assessment in fourth graders designed to compare reading literacy between countries and over time. It has been conducted every five years since 2001. Next to reading literacy, the study gathers contextual information about children's home and school experiences from children, parents, teachers, and school principles. The PIRLS dataset is kindly provided by the PIRLS team at TU Dortmund University.
- **Goal:** Find weights for PIRLS 2016 and 2021 that balance the covariate distribution between cohorts. It is of great interest to attribute reading literacy trends to demographic composition of cohorts or more closely to school-related learning. Since demographic composition varies between years, differences in the covariate distributions could bias the interpretation of trends results. The original survey weights are provided.
- **Data** are from PIRLS 2016 and PIRLS 2021 from the Italian subsample (low number of missings, listwise deletion) with pronounced cluster-level effects. We provide mostly student-level variables.
- **Evaluation of generalizability** is on parent-level variables. Metrics include the sum of standardized mean differences and squared differences in the covariance matrices (Frobenius norm).

Travel Costs
Participants generally cover their travel costs. Hackathon participation is otherwise free of charge. Two travel grants for non-tenured participants can be applied for as part of an outline submission.

Outline Details
Brief Outlines with max. 300 words contain information on your preferred methodological approach(es). Please include details on pre-processing and tuning. Mention your target dataset(s) (Dataset 1 and/or Dataset 2) and add citations of key papers. Explain in what sense your approach is interpretable. All outlines need to be sent via email to agile-pair@tu-dortmund.de until April 20th, 2026, 23:59 (CET — Central European Time).

Outline Checklist
- Title
- Authors with institutions and email
- Max. 300 words in addition to the two above
- In your email, please specify who the corresponding author is (if not identical to the author submitting the abstract) and who will participate in the in-person workshop.
- In case you want to apply for a travel grant and do not have tenure, please mention this in the email, too.

Analysis Plan Details
To prevent data peeking, analysis plans are required before data are shared with participants. The plans provide further details of the methods from the abstracts. Researchers are free to refine the approaches from the outlines. Beyond the outlines, more details on implementation, validation, cross-validation, and interpretability are expected. Code snippets can be part of an analysis plan if they help to clarify implementation details. There is no word limit for the analysis plans.

All analysis plans will be part of the pre-registration. For this, all researchers with accepted outlines will be able to upload their analysis plans to the hackathon's osf.io repository until June 18th, 23:59 (CET — Central European Time), thereby pre-registering the analysis.

Analysis Phase Details
Analysis can begin after the pre-registration on osf.io is complete. Participants are assumed to follow their analysis plans, and deviations from the analysis plans are to be documented. All participants participating in the in-person workshop are requested to provide preliminary results as a basis for the discussions.

Organizers
- Philipp Doebler, Department of Statistics & Center for Agile PAIR , TU Dortmund University
- Jörg-Tobias Kuhn, Faculty of Rehabilitation Sciences, Methods of Empirical Educational Research & Center for Agile PAIR, TU Dortmund University
- Annette Lohbeck, Faculty of Rehabilitation Sciences & Center for Agile PAIR, TU Dortmund University
- Katja Ickstadt, Department of Statistics & TU Dortmund Center for Data Science and Simulation
- Claus Weihs, Department of Statistics
- Jakob Schwerter, Hector Research Institute of Education Sciences and Psychology, University of Tübingen
- Ulrich Ludewig, Department of Educational Sciences and Psychology, TU Dortmund University

Please send all questions to agile-pair@tu-dortmund.de.