

#####  
#

Supplementary Material 2 - Electronic Appendix to the  
article "Evaluating Machine Learning Models in Non-Standard Settings: An Overview and  
New Findings" by Roman Hornung\*, 1,2, Malte Nalenz 3, Lennart Schneider 3,2, Andreas Bender 3,2,  
Ludwig Bothmann 3,2, Bernd Bischl 3,2, Thomas Augustin 3, Anne-Laure Boulesteix 1,2

1 Institute for Medical Information Processing, Biometry and Epidemiology,  
LMU Munich, Munich, Germany

2 Munich Center for Machine Learning (MCML), Munich, Germany

3 Department of Statistics, LMU Munich, Munich, Germany

\* For questions, please contact: [hornung@ibe.med.uni-muenchen.de](mailto:hornung@ibe.med.uni-muenchen.de)

#####  
#

Program and Platform:  
#####

- Program: R, versions 4.1.2 and 4.2.3.

- Used platforms: Linux (x86-64) (for the conduction of the analyses)  
Windows 7 and 10 64-bit (for the evaluation of the results)

- The following output from sessionInfo() describes which R packages and versions were used:

```
> sessionInfo()  
R version 4.2.3 (2023-03-15 ucrt)  
Platform: x86_64-w64-mingw32/x64 (64-bit)  
Running under: Windows 10 x64 (build 22621)
```

Matrix products: default

locale:  
[1] LC\_COLLATE=German\_Germany.utf8 LC\_CTYPE=German\_Germany.utf8 LC\_MONETARY=German\_Germany.utf8  
[4] LC\_NUMERIC=C LC\_TIME=German\_Germany.utf8

attached base packages:  
[1] stats graphics grDevices utils datasets methods base

other attached packages:  
[1] sf\_1.0-12 scales\_1.2.1 RColorBrewer\_1.1-3 ranger\_0.15.1 mlr3verse\_0.2.7  
[6] mlr3temporal\_0.1.0.9000 mlr3learners\_0.5.6 mlr3\_0.15.0 measures\_0.3 hierclass\_0.1.0  
[11] gridExtra\_2.3 ggpubr\_0.6.0 ggforce\_0.4.1 ggplot2\_3.4.1 dplyr\_1.1.1  
[16] data.table\_1.14.8

loaded via a namespace (and not attached):

[1] bbotk_0.7.2	prabclus_2.3-2	mlr3pipelines_0.4.3	tools_4.2.3	backports_1.4.1
[6] utf8_1.2.3	R6_2.5.1	KernSmooth_2.23-20	DBI_1.1.3	colorspace_2.1-0
[11] mlr3data_0.6.1	nnet_7.3-18	withr_2.5.0	mlr3viz_0.6.1	mlr3misc_0.11.0
[16] tidyselect_1.2.0	compiler_4.2.3	cli_3.6.0	lgr_0.4.4	diptest_0.76-0
[21] checkmate_2.1.0	DEoptimR_1.0-12	classInt_0.4-9	robustbase_0.95-1	mlr3tuningspaces_0.3.5
[26] palmerpenguins_0.1.1	mlr3tuning_0.18.0	digest_0.6.31	paradox_0.11.1	pkgconfig_2.0.3
[31] parallelly_1.36.0	rlang_1.1.0	rstudioapi_0.14	farver_2.1.1	generics_0.1.3
[36] mclust_6.0.0	car_3.1-1	magrittr_2.0.3	modeltools_0.2-23	Matrix_1.5-3
[41] Rcpp_1.0.10	mlr3fselect_0.11.0	munsell_0.5.0	fansi_1.0.4	abind_1.4-5
[46] lifecycle_1.0.3	carData_3.0-5	MASS_7.3-58.2	flexmix_2.3-19	grid_4.2.3
[51] parallel_4.2.3	listenv_0.9.0	crayon_1.5.2	lattice_0.20-45	mlr3cluster_0.1.8
[56] pillar_1.9.0	mlr3filters_0.7.1	uuid_1.1-0	fpc_2.2-10	ggsignif_0.6.4
[61] codetools_0.2-19	stats4_4.2.3	glue_1.6.2	vctrs_0.6.0	tweenr_2.0.2
[66] gtable_0.3.3	purrr_1.0.1	polyclip_1.10-4	tidyr_1.3.0	clue_0.3-64
[71] kernlab_0.9-32	future_1.32.0	broom_1.0.4	e1071_1.7-3	rstatix_0.7.2
[76] class_7.3-21	tibble_3.2.1	units_0.8-2	cluster_2.1.4	globals_0.16.2

General information and contents of this Electronic Appendix:

#####

- All R scripts use relative paths. Make sure that the R working directory is set to the directory of the 'PPerfEstComplex' folder where this README file is located. On Linux systems, R can also be run directly from the 'PPerfEstComplex' directory.
- 'PPerfEstComplex' contains the following subfolders: 'clustdata': This subfolder contains the code to reproduce and evaluate the clustered data simulation study; 'concdrift': This subfolder contains the code to reproduce and evaluate the concept drift simulation study; 'hierpr': This subfolder contains the code to reproduce and evaluate the simulation study on hierarchically structured outcomes; 'nsrs': This subfolder contains the code to reproduce and evaluate the unequal sampling probabilities simulation study; 'spatial': This subfolder contains the code and data (shapefiles of Bavaria in the subfolder 'data') to reproduce Figure 3. The raw and final results (in the form of figures) are also included.
- Except for 'spatial', all subfolders have very similar structures. They all contain three R scripts "simulation.R", "functions.R" and "evaluation.R". Here, "simulation.R" produces the raw results of the simulation studies, "functions.R" contains the functions used in the simulation studies, and "evaluation.R" evaluates the raw results to produce the figures and support the statements made in Supplementary Material 1. Furthermore, all subfolders (except 'spatial') contain a subfolder 'results', which contains two further subfolders 'intermediate\_results' and 'figures', where the former contains the raw results and the latter contains the figures as presented in the paper and in Supplementary Material 1. The subfolder 'hierpr' contains three further scripts "simulation\_subset\_test.R", "functions\_subset\_test.R" and "evaluation\_subset\_test.R" (and the corresponding results), which perform and evaluate the version of the simulation study on hierarchically structured outcomes, where only observations from classes that were included in the training data are included in the huge test data sets used to approximate the true evaluation metric values (cf. Statement "However, even when including in the test data only observations from classes that were contained in the training data, SCV still overestimated the true performance, albeit less strongly and not for all evaluation metrics (results not shown)." in Section C.2 of

Supplementary Material 1).

Some of the 'figures' subfolders also contain TeX files that generate the tikz figures shown in the manuscript. The script "obtain\_treestruc.R" in the subfolder 'hierpr' generates the tree structure used in the simulation of hierarchically structured results.

Evaluation of the results:

#####

- For the evaluation of the results it is not necessary to re-perform the analyses:

The R scripts "evaluation.R" contained in the subfolders (except 'spatial') produce all figures shown in the main paper and in Supplementary Material 1 as well as results that support statements made in Supplementary Material 1 without the need to re-run the analyses.

These R scripts load Rda files (stored in the subfolders "results/intermediateresults") that contain the raw results.

- Below is a detailed description of which scripts produce which results:

- "clustdata/evaluation.R": "Figure2.pdf", "FigureS1.pdf", "FigureS2.pdf", "FigureS3.pdf", "FigureS4.pdf", "FigureS5.pdf", "FigureS6.pdf"

- "clustdata/results/figures/Figure1.R": "Figure1.pdf"

- "concdrift/evaluation.R": "Figure7.pdf", "FigureS7.pdf", "FigureS8.pdf", "FigureS9.pdf", "FigureS10.pdf", "FigureS11.pdf", "FigureS12.pdf", "FigureS13.pdf", "FigureS14.pdf", "FigureS15.pdf", "FigureS16.pdf"

- "concdrift/results/figures/Figure5.R": "Figure5.pdf"

- "concdrift/results/figures/Figure6.R": "Figure6.pdf"

- "hierpr/evaluation.R": "Figure9.pdf", "FigureS17.pdf", "FigureS18.pdf", "FigureS19.pdf"

- "nsrs/evaluation.R": "Figure4.pdf"

- "spatial/spatial\_cv\_figure.R": "Figure3.pdf"

Full reproduction of the results:

#####

- All R code needed to fully reproduce the analyses is available in this electronic appendix.

- An MPI environment is required.

- The R scripts named "simulation.R" in the corresponding subfolders require the RMPISNOW shell script from the R package "snow".

Therefore, before executing these scripts you need to install the RMPISNOW shell script from the installed 'snow' R package or 'inst' directory of the package sources of the 'snow' R package in an appropriate location, preferably

on your path.

See <http://homepage.divms.uiowa.edu/~luke/R/cluster/cluster.html> (last accessed: 20th October 2023) for more details.

Subsequently, you need to create sh files, each for a different of the above R scripts. The following is the content of an example sh file "simulation\_clustdata.sh":

```
#!/bin/bash
#SBATCH -o /myoutfiledirectory/myjob.%j.%N.out
#SBATCH -D /myhomedirectory
#SBATCH -J LargeStudy
#SBATCH --get-user-env
#SBATCH --clusters=myclustername
#SBATCH --partition=mypartitionname
#SBATCH --qos=mypartitionname
#SBATCH --nodes=??
#SBATCH --tasks-per-node=??
#SBATCH --mail-type=end
#SBATCH --mail-user=my@mail.de
#SBATCH --time=??:??:??
```

```
mpirun RMPISNOW < ./PPerfEstComplex/clustdata/simulation.R
```

The above sh file of course has to be adjusted to be useable (e.g., the "?"s have to be replaced by actual numbers, the directories have to be adjusted and you need to specify your e-mail address; an e-mail will be sent to this address once the job is finished).