Selection of machine learning models from Rashomon set gives a broader perspective on explaining the data

Katarzyna Kobylińska[1][0000−0002−0292−4982], Rafał Machowicz[2], Mariusz Adamek[3,4][0000−0002−1885−9257], and Przemysław Biecek[1,5][0000−0001−8423−1823]

[1] University of Warsaw, Faculty of Mathematics, Informatics and Mechanics, Poland
[2] Department of Hematology, Transplantation and Internal Medicine, Medical University of Warsaw
[3] Faculty of Medicine and Dentistry, Medical University of Silesia
[4] Faculty of Health Sciences, Medical University of Gdańsk, 80-210 Gdańsk, Poland
[5] Faculty of Mathematics and Information Science, Warsaw University of Technology

Abstract. The classical machine learning modeling process ends with the selection of an accurate model that minimizes loss function. The process favors this model and abandons a more profound analysis of slightly worse models. In the case of complex relationships, such tunnel vision can lead to misleading or incomplete conclusions. The set of almost equally good models is called the Rashomon set, can be numerous, and may contain models that differ significantly in explaining data. This study proposes a novel method to explore models from the Rashomon set. We demonstrate that it is better to consider many good models than only one. Such a broader perspective allows for finding models with other features, e.g., compatible with domain knowledge or different explanations of the data. The process contains a choice of k the most diverse models from that set. We propose a measure to compare the model's explanations. The process and the measure would work on any Explainable Artificial Intelligence method that presents the dependency between the variable and the prediction. We strongly believe that the introduced method will help enlarge the perspective on explaining the data. For the researcher, our approach is a way to find, analyze and compare models that could have features that were unnoticed by a selection of models based on the loss function. For the physician, the process helps to trust the model's results. We present the use case on models trained to predict the survival of patients with Hemophagocytic lymphohistiocytosis (HLH).

Keywords: Rashomon set · Explainable Artificial Intelligence · Partial Dependence Plot · Similarity measures

A Appendix

In the Appendix, we attach the detailed results of the study. The exact results of the simulation study from Section **??** are presented in Table A.1 and Figure A.1.

Table A.1. Similarity measures for synthetic data set

Models	Similarity measuremnet	X1	X2	X4
$\text{GBM}_{dataset_0}$	**Frechet**	0.59	0.08	0.04
$\text{GBM}_{dataset_1}$	**Measure**	0.19	0.63	0.55
$\text{RF}_{dataset_0}$	**Frechet**	0.61	0.07	0.06
$\text{GBM}_{dataset_1}$	**Measure**	0.17	0.61	0.63
$\text{RF}_{dataset_1}$	**Frechet**	0.09	0.09	0.05
$\text{GBM}_{dataset_1}$	**Measure**	0.87	0.55	0.63

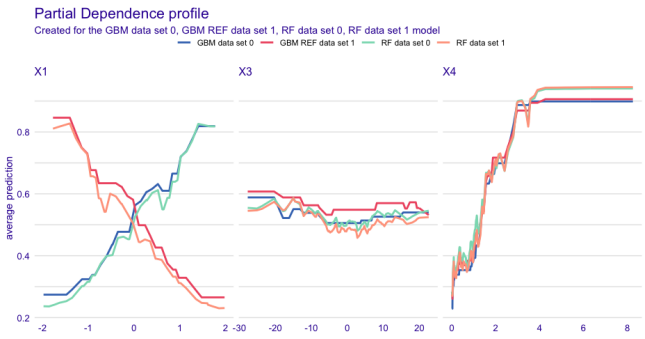

Fig. A.1. Partial Dependence Plots for RF and GBM models. The red curves represent the winner model.

Figure A.2 presents the values of measure computed for each pair of models.

Figure A.3 presents Partial Dependence Profiles calculated on three the most different models for each continuous variable.

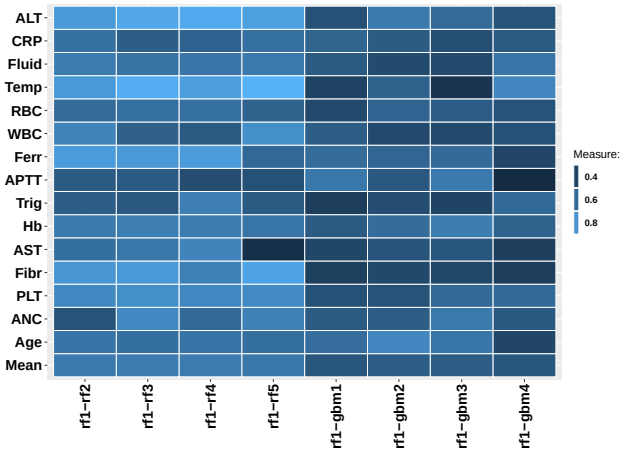

Fig. A.2. Heatmap of measure values computed for pairs of models in step 1

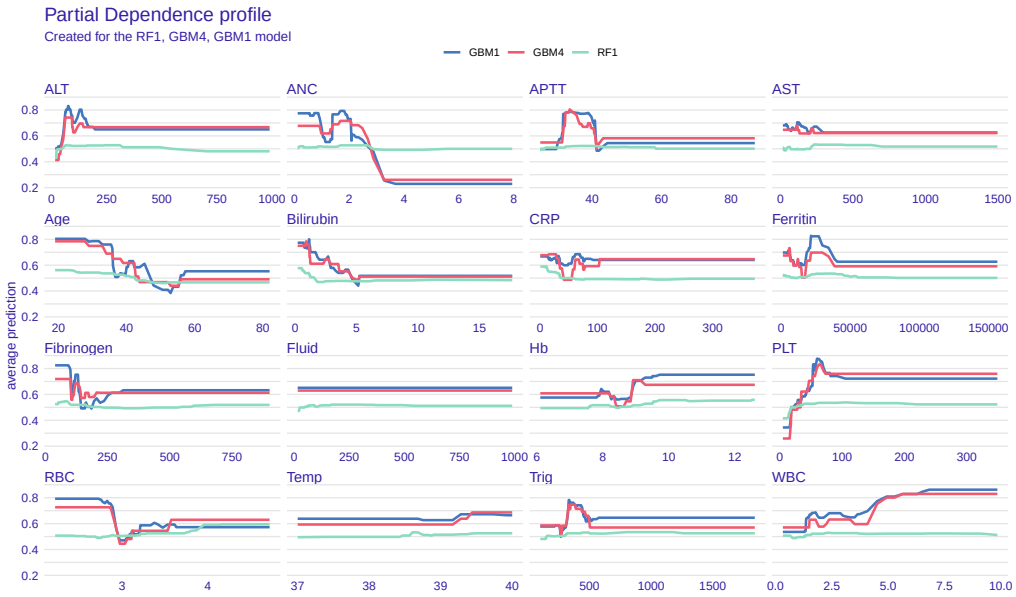

Fig. A.3. Partial Dependence Profiles for three the most different models from Rashomon set and following variables: ALT (Alanine aminotransferase), ANC (Absolute Neutrophil Count), APTT (Activated Partial Thrombin Time), AST (Aspartate aminotransferase), Age, Bilirubin, CRP (C-Reactive Protein), Ferritin, Fibronectin, Fluid, Hb (Hemoglobin), PLT (Platelet Count), RBC (Red Blood Cell Count), Temperature, Trig (Triglycerides), WBC (White Blood Cell Count)