4.1.2 Comparison Methods

# Original

Current state-of-the-art approaches to model extraction were trialled. These were Bayesian rule lists from [35] (as pysbrl on pip), an h2o decision tree ([31]), a simplified scikit-learn [25] decision tree [22], and logistic regression with ℓ1-regularization (from scikit-learn). For the scikit-learn methods, unfortunately, these do not natively support categorical features, so a one-hot-encoding is needed to be applied prior. For Bayesian rule lists, they currently only support discrete features, so as is commonly done, multi-interval discretization as proposed in [16] was first applied. No pre-processing was required for the h2o trees. Unfortunately, to our knowledge there are no current approaches in literature to multi-objective XAI, so we only compare the best resulting solutions rather than comparing frontiers as is more commonly done with multi-objective optimisation algorithms.

# Condensed

Current state-of-the-art approaches to model extraction were trialled:

* Bayesian rule lists [35] (as pysbrl on pip)
* A (h2o) decision tree [31]
* A simplified (scikit-learn) [25] decision tree [22]
* Logistic regression with ℓ1-regularization (from scikit-learn)

Pre-Processing: One-hot-encoding was required for the scikit-learn methods to support categorical features. Multi-interval discretization is applied for the Bayesian rule lists to support continuous features.

Unfortunately, to our knowledge there are no current approaches in literature to multi-objective XAI, so we only compare the best resulting solutions rather than comparing frontiers as is more commonly done with multi-objective optimisation algorithms. – ADD AS FOOTNOTE???