

## Abstract

- In any given machine learning problem, there may be many models that could explain the data almost equally well. Having a set of all almost-optimal models (the *Rashomon set*) allows practitioners to explore alternative models that might have desirable properties beyond what could be expressed within a loss function.
- We provide TreeFARMS (Trees FAST RASHOMON Sets), the first technique for completely enumerating the Rashomon set for sparse decision trees, a non-trivial problem with a highly nonlinear discrete function class. We represent the Rashomon set in a specialized data structure that supports efficient querying and sampling.
- We show three applications of the Rashomon set: 1) it can be used to study variable importance for the set of almost-optimal trees (as opposed to a single tree), 2) the Rashomon set for accuracy enables enumeration of the Rashomon sets for balanced accuracy and F1-score, and 3) the Rashomon set for a full dataset can be used to produce Rashomon sets constructed with only subsets of the data set.

## Objectives

### Harvest almost-optimal trees with various properties

- Practitioners may want to explore the Rashomon set to find models with important properties, such as interpretability, fairness, or use of specific variables, or they may want to choose a model that agrees with possible causal hypotheses, monotonicity trends, ease of calculation, or simply domain intuition.

### Explore several properties of the Rashomon set

- Size:** Does Rashomon set size vary between datasets?
- Variable importance:** How does the variable importance change among trees within the Rashomon set?
- Robustness:** How does the Rashomon set change if we remove a subset of data?
- Rashomon sets for other losses:** How can we construct Rashomon sets for balanced accuracy and F1-score?

## Methods

### Extraction Framework

- We implement TreeFARMS in GOSDT<sup>[1]</sup>, which uses a dynamic-programming-with-bounds formulation to find the optimal sparse decision trees.
- While searching the space, TreeFARMS prunes the search space by removing subproblems that lead to trees violating user's definition of *almost optimal*.

### Bounds for Reducing Search Space

- Rashomon Equivalent Points Bound for Subtrees*
  - Given an *objective threshold*  $\theta\epsilon$  such that all trees with objective below that value are considered *almost optimal*, and  $p_{\text{left}}\text{-lb}$ ,  $p_{\text{right}}\text{-lb}$  the lower bound of left and right subproblem, we have:
    - If  $p_{\text{left}}\text{-lb} + p_{\text{right}}\text{-lb} > \theta\epsilon$ , then we know all subtrees generated by  $p_{\text{left}}$  and  $p_{\text{right}}$  cannot be in the Rashomon Set.

## Scalability

- During merging process of the dynamic programming algorithm, we combine solved subtrees into larger trees.
- Instead of operating on individual trees, we operate on sets of trees grouped by their objective values.
- This dramatically boosts efficiency as the loss function for decision trees takes on a discrete number of values (approximately equal to the number of samples in the training data set), while the number of trees in the Rashomon set is frequently orders of magnitude larger.

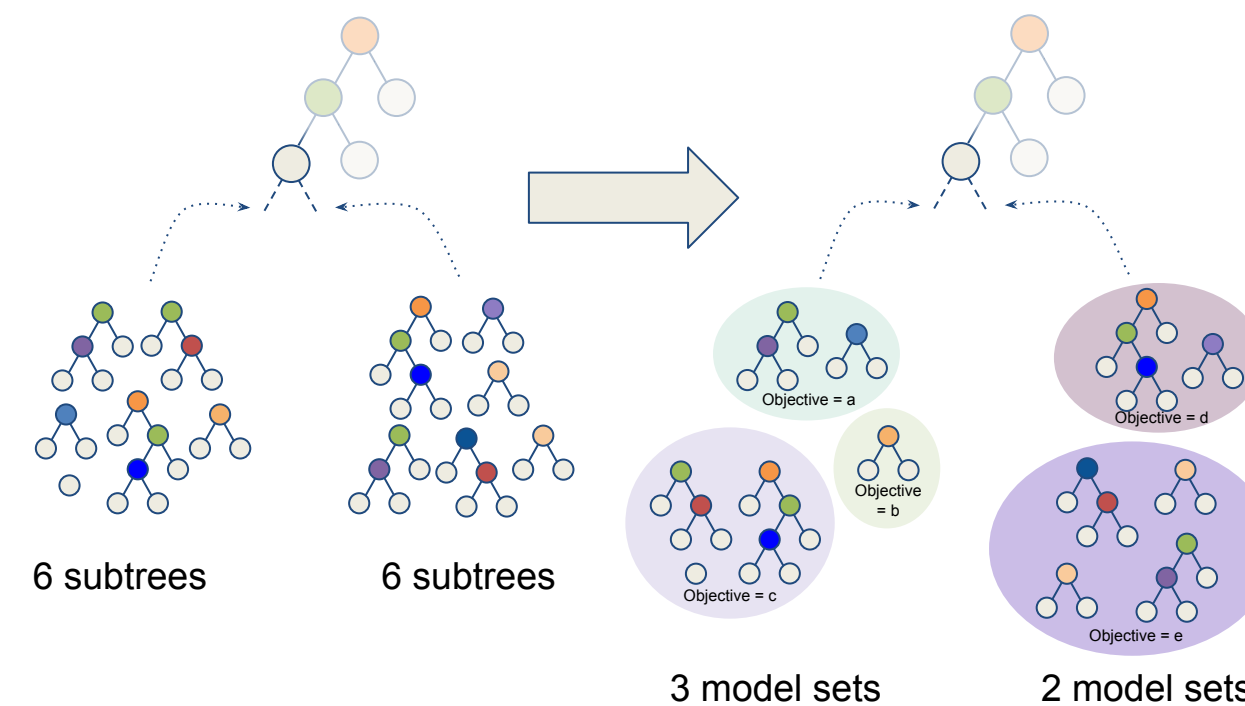


Figure 1: Illustration of Model Set representation performance boost. During merging, we need to compare every pair of subtrees. Prior to using model set, we require  $6 \times 6 = 36$  operations; whereas after combining subtrees using model set, we only need  $2 \times 3 = 6$  operations. Doing this on each level leads to exponential speedup.

## Results

We compare TreeFARMS against four baselines on the Monk2<sup>[2]</sup> dataset.

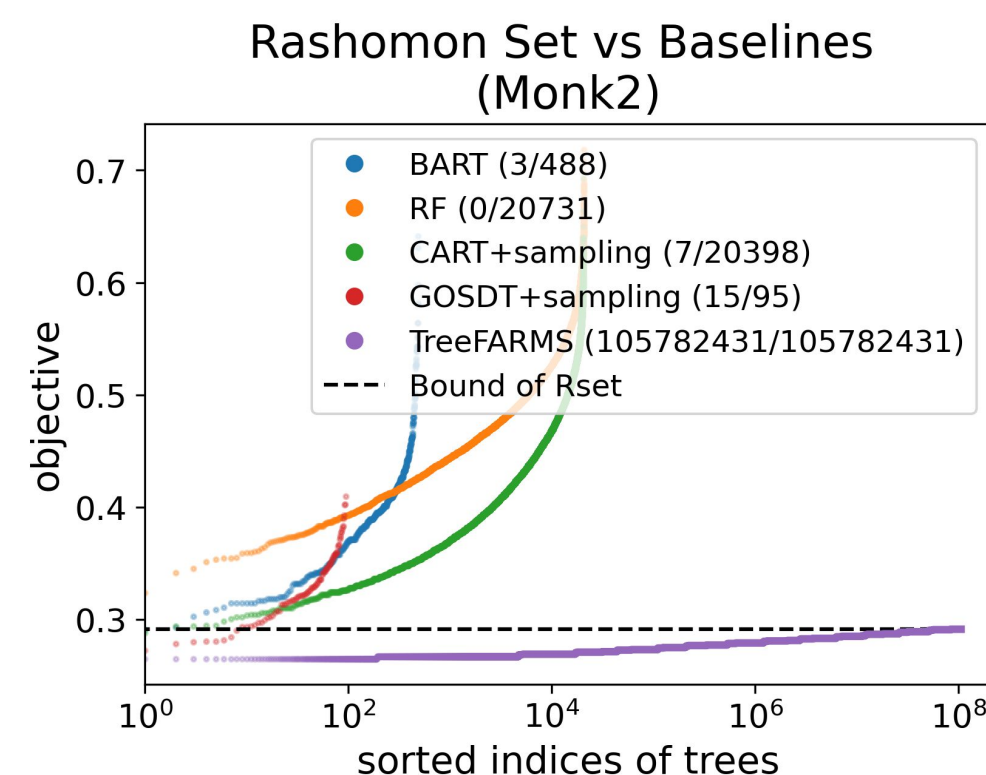


Figure 2: Comparison of trees found in the Rashomon set and trees generated by baselines in about the same amount of time. Trees in the Rashomon set have objective below the dashed line. (A/B) in legend represents that A trees among B trees trained by the baseline are in the Rashomon set.

## Applications

### Variable importance

- We consider model class reliance (MCR)<sup>[3]</sup>, as a variable is important to one model does not mean that it is important in general. MCR provides the range of variable importance values across the set of all well-performing models.

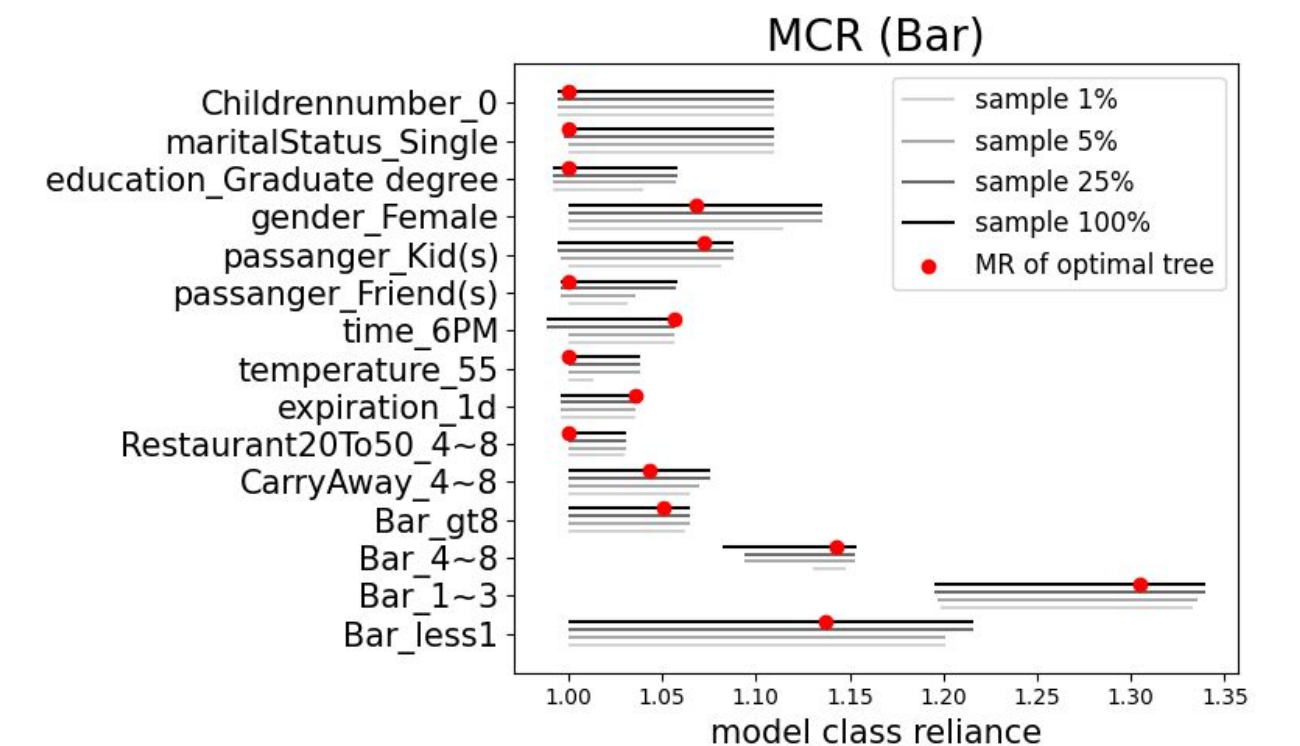


Figure 3: Variable Importance: Model class reliance on the Bar Dataset<sup>[4]</sup>. Red dots indicate the model reliance (variable importance) calculated from the optimal tree. Each line connects MCR- and MCR+ showing the range of variable importance among all good models.

### Sensitivity to Missing Groups of Samples

- We can measure how a sample or a group of samples influences all well-performing models.

### Constructing the Rashomon Set for Metrics other than Accuracy

- By default, TreeFARMS constructs Rashomon Set using accuracy metric.
- We show that, given the Rashomon set constructed using accuracy, we can directly find the Rashomon sets for balanced accuracy and F1-score.

## Conclusion

This work opens the door to interesting discussions on variable importance, distributional shift, and user options. By efficiently representing all optimal and slightly suboptimal models for complex nonlinear function classes with interactions between variables, we provide a range of new user-centered capabilities for machine learning systems, and a new understanding of the importance of variables. Importantly, our method allows users a choice rather than handing them a single model. Thus, we are able to examine Rashomon sets across problems with a new lens, enabling users to choose models rather than be at the mercy of an algorithm that produces only a single model.

## References

- [1] Jimmy Lin, Chudi Zhong, Diane Hu, Cynthia Rudin, and Margo Seltzer. Generalized and scalable optimal sparse decision trees. In International Conference on Machine Learning, pages 6150–6160. PMLR, 2020.
- [2] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.
- [3] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research, 20(177):1–81, 2019.
- [4] Tong Wang, Cynthia Rudin, Finale Doshi-Velez, Yimin Liu, Erica Klampfl, and Perry MacNeille. A bayesian framework for learning rule sets for interpretable classification. The Journal of Machine Learning Research, 18(1):2357–2393, 2017.