Rashomon Ensembled Active Learning

Simon Dovan Nguyen [1] Kentaro Hoffman [1] Tyler H. McCormick [1 2]

[1]Department of Statistics [2]Department of Sociology

Abstract

Active learning's key task is selecting informative data points to enhance model predictions with a fixed labeling budget. However, when ensemble models such as random forests are used, there is a risk of the ensemble containing models with poor predictive accuracy or redundant trees with the same interpretation. To address this, we develop a novel approach called *UNique Rashomon Ensembled Active Learning (UNREAL)* to only ensemble the set of near-optimal models called the Rashomon set in order to guide the active learning process. We demonstrate how *UNREAL* can not only improve the accuracy and rate of convergence of the active learning procedure but also lead to improved interpretability compared to traditional approaches.

Contributions

1. We enhance active learning by exclusively aggregating the most plausible decision trees of the Rashomon set, ensuring efficiency and relevance.
2. We improve interpretability by pruning redundant trees, providing a more comprehensive ensemble of distinct yet near-optimal models.

Active Learning

Labeling training data for machine learning models is expensive and time-consuming. Active learning (AL) reduces this burden by strategically selecting the most informative data points for labeling.

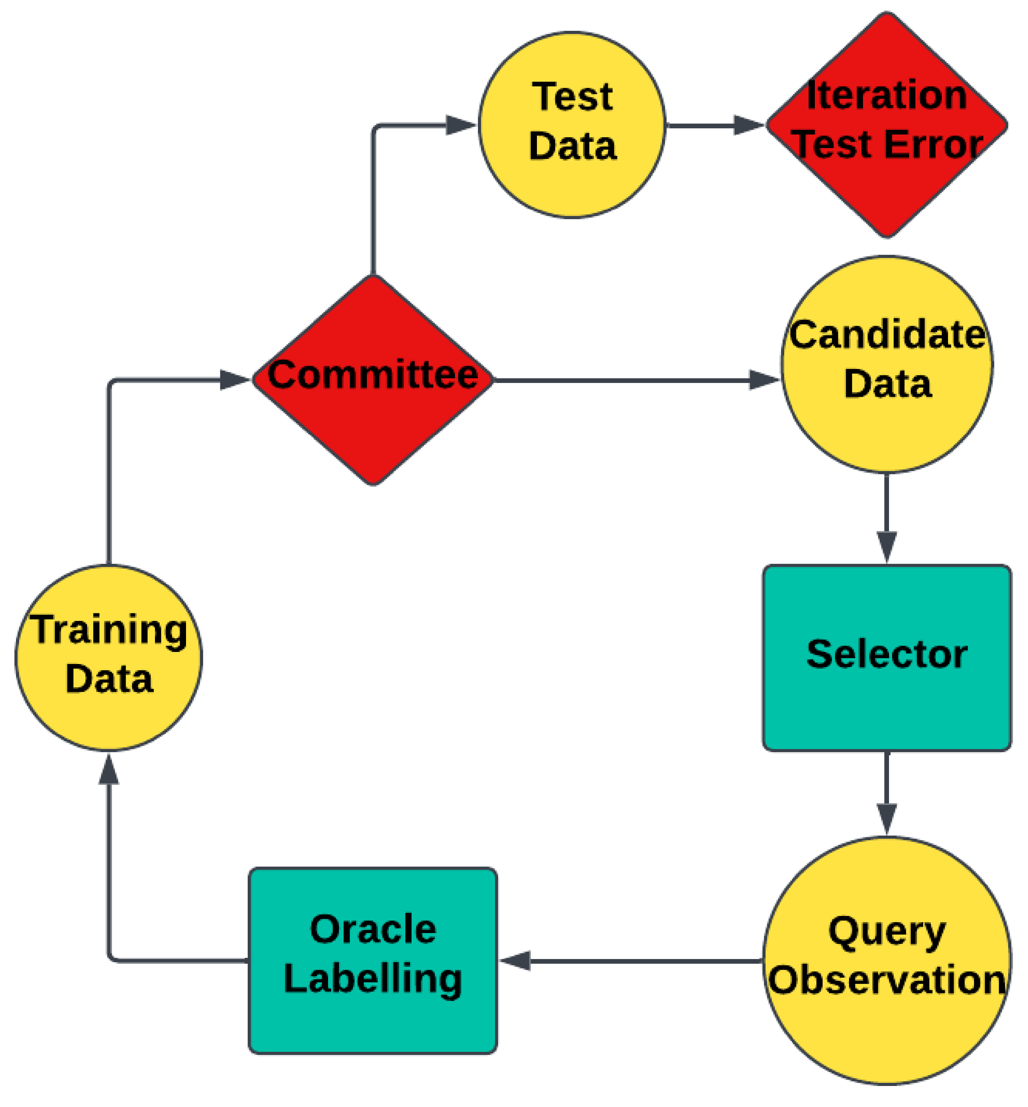


Figure 1. Active learning diagram.

Informativity is measured by vote entropy:

$$\delta(y, \mathbf{x}, \mathcal{C}) = \max_{\mathbf{x}} - \sum_{y \in \mathcal{Y}} \frac{\text{vote}_{\mathcal{C}}(y, \mathbf{x})}{|\mathcal{C}|} \log \frac{\text{vote}_{\mathcal{C}}(y, \mathbf{x})}{|\mathcal{C}|}$$

where

$$\text{vote}_{\mathcal{C}}(y, \mathbf{x}) = \sum_{c \in \mathcal{C}} \mathbb{I}\{c(\mathbf{x}) = y\}$$

is the number of "votes" that label y receives for covariate $\mathbf{x}$ from the members c of committee $\mathcal{C}$.

The base learners in ensemble methods naturally form a committee, with disagreement in "votes" serving as a common uncertainty metric. **However, redundant trees with duplicate explanations can artificially inflate agreement in the committee** [1].

Be Careful What You Average Over!

1. Incorporating implausible/poor trees

Most ensemble methods tend to aggregate over the space of all models, even if some of the models may have relatively poor accuracy.

2. Explanations $\neq$ Trees

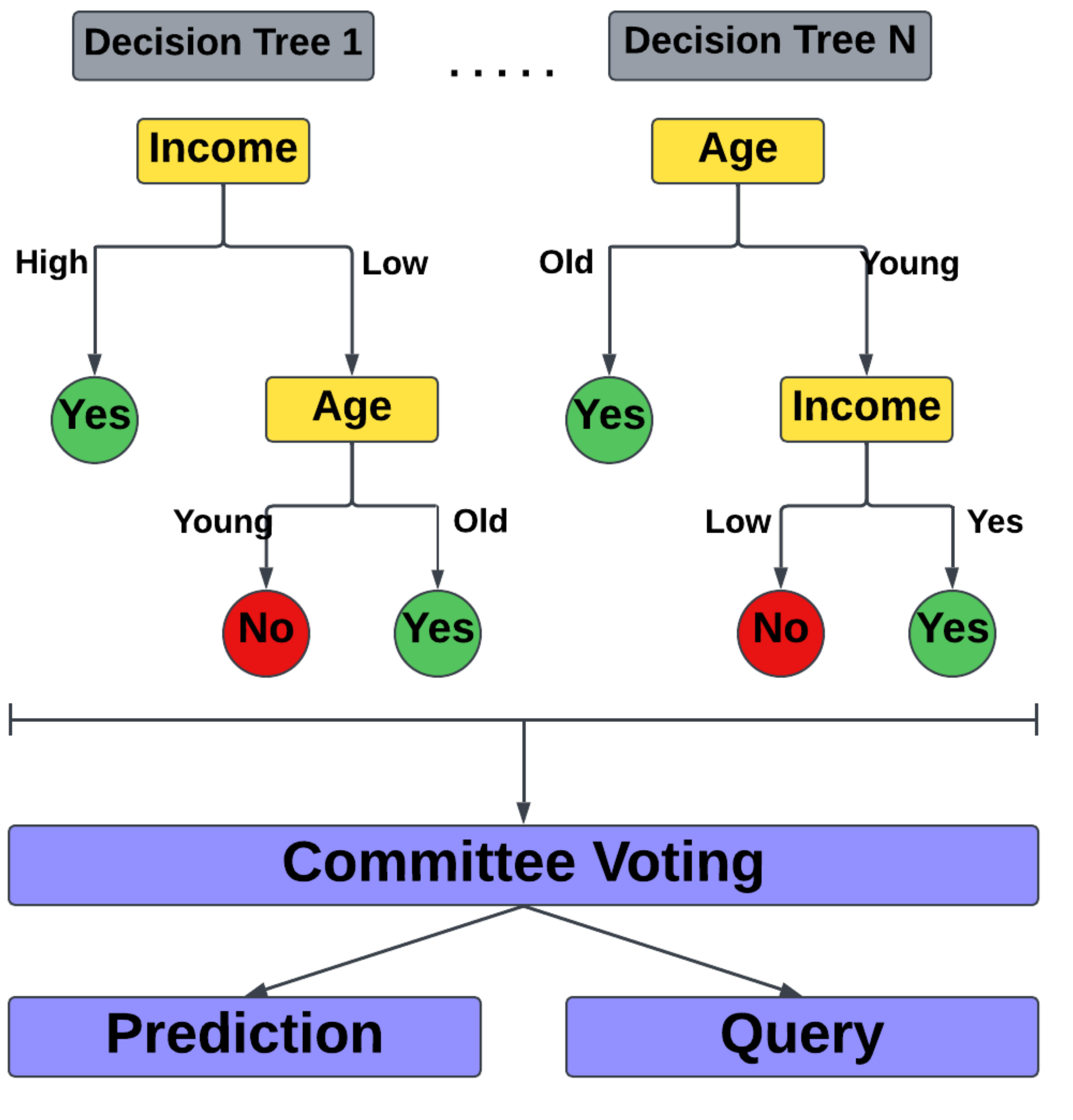


Figure 2. Multiplicity of explanations in decision trees.

Rashomon Decision Trees

The Rashomon set is a collection of models that are all near-optimal in predictive accuracy.

Definition 1 (Rashomon Set) Given a threshold ϵ, a hypothesis space $\mathcal{F}$, loss function $\mathcal{L}$, and reference model $\hat{f}$, a Rashomon set $\hat{\mathcal{R}}$ is defined as

$$\hat{\mathcal{R}} := \{f \in \mathcal{F} : \mathcal{L}(f) \leq (1+\epsilon) \cdot \mathcal{L}(\hat{f})\} \quad (1)$$

Xin et al. were the first to introduce an algorithm (`TreeFarms`) that fully enumerates the Rashomon set for sparse decision trees [3].

Adaptation: To address Issue 2, we restrict our aggregation to trees with unique explanations in the Rashomon set:

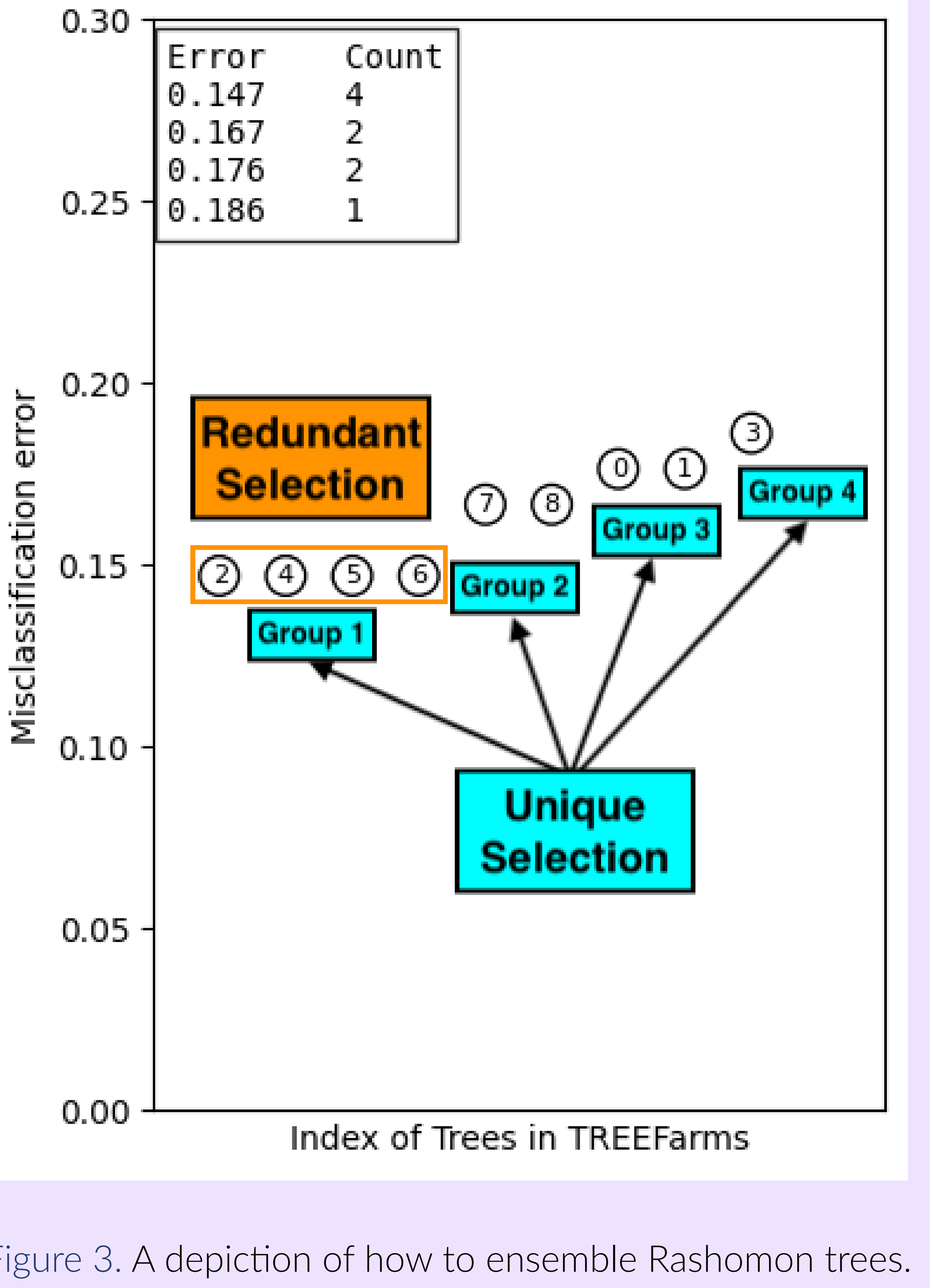


Figure 3. A depiction of how to ensemble Rashomon trees.

Methods/Algorithm

Algorithm 1: *Unique Rashomon Ensembled AL*

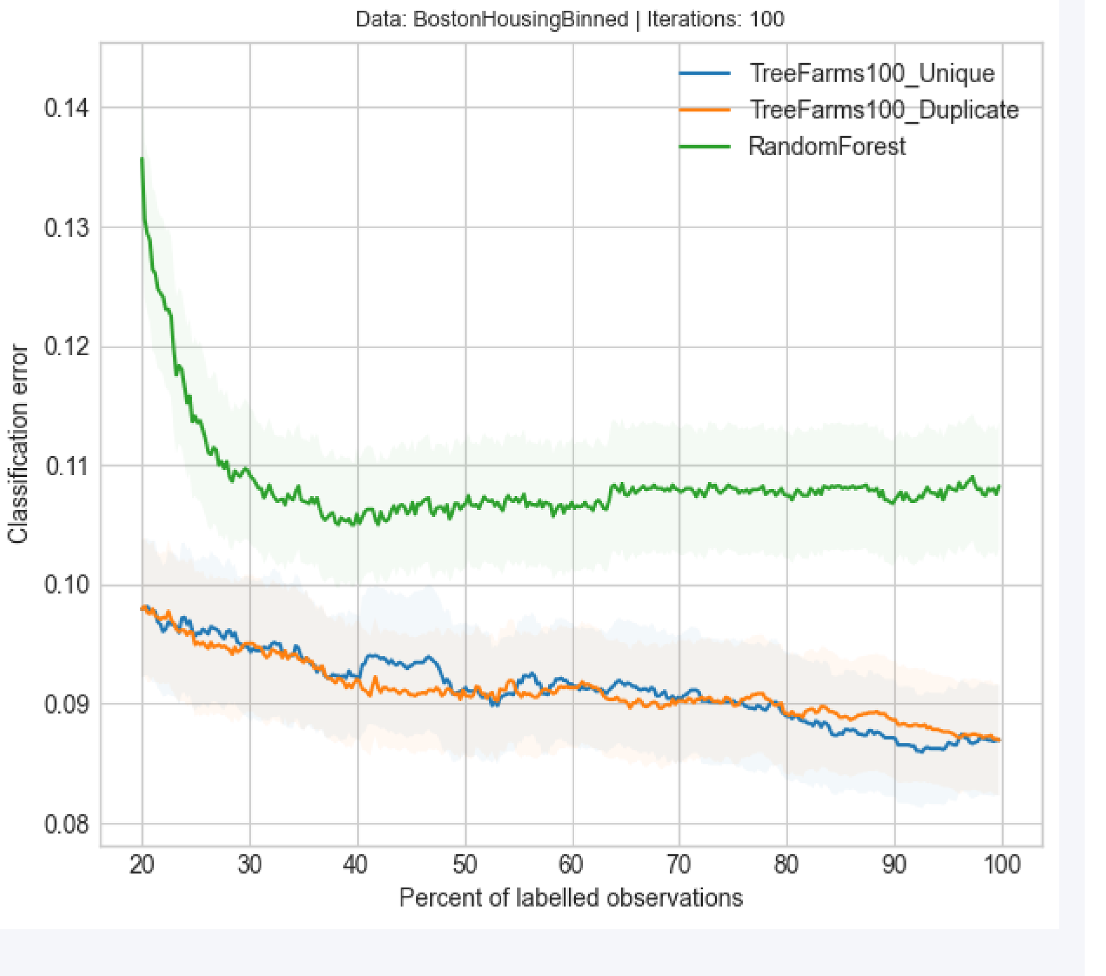
Input: $D^{(0)}_{trn}$; D_{tst}; $D^{(0)}_{cdd}$; ϵ;

repeat

- Train $\mathbf{F}$ on $D^{(n)}_{trn}$;
- Test $\mathbf{F}$ on $D^{(n)}_{tst}$;
- Enumerate $\mathcal{R}$ with `TreeFarms`;
- *(Optionally)* Reduce the Rashomon set $\mathcal{R}$ to the top k models in $\mathcal{R}$;
- Predict labels $\hat{y}^{(n)}_{tst,m}$ from D_{tst} and calculate the classification error for each tree f_m in $\mathcal{R}$;
- Define the the smallest classification error from amongst the trees in $\mathcal{R}$ as the current iteration error;
- Compute vote-entropy $\delta^{(n)}(y, x, \mathcal{C})$ on $D^{(n)}_{cdd}$;
- Resample $B^{(n)}$ from $D^{(n)}_{cdd}$ based on the observation with the highest vote entropy: $B^{(n)} := \arg\max_x \delta^{(n)}(y, x, \mathcal{C})$;
- Query $B^{(n)}$ for oracle labeling;
- Update training set $D^{(n+1)}_{trn} = D^{(n)}_{trn} \cup B^{(n)}$;
- Update candidate set: $D^{(n+1)}_{cdd} = D^{(n)}_{cdd} \setminus B^{(n)}$

until *labelling budget is depleted or test error is sufficiently small*;

Simulation Results



Result 1: Aggregating over the Rashomon sets outperform random forest by a significant margin.

Result 2: Pruning to only retain trees with unique explanations improves interpretability while maintaining predictive accuracy.

Future Work

Unlike `TreeFarms`, Rashomon Partition Sets [2] exhaustively enumerate the Rashomon set without relying on a predefined geometry. This geometry-free approach *plants the seeds* for potential active learning improvements in prediction and interpretability.

References

[1] Prem Melville and Raymond J. Mooney. Diverse ensembles for active learning. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, New York, NY, USA, 2004. Association for Computing Machinery.

[2] Aparajithan Venkateswaran, Anirudh Sankar, Arun G. Chandrasekhar, and Tyler H. McCormick. Robustly estimating heterogeneity in factorial data using rashomon partitions. *ArXiv*, 2024.

[3] Rui Xin, Chudi Zhong, Zhi Chen, Takuya Takagi, Margo Seltzer, and Cynthia Rudin. Exploring the whole rashomon set of sparse decision trees. In *Advances in Neural Information Processing Systems*, 2022.