# Literati: Towards Anytime Optimal Shape Generalized Trees via AO*

**Nakul Upadhya, Eldan Cohen**
Department of Mechanical and Industrial Engineering
University of Toronto, Toronto, Canada
nakul.upadhya@mail.utoronto.ca, eldan.cohen@utoronto.ca

## Abstract

Decision trees are prized for their interpretability and strong performance on tabular data, but popular greedy top-down induction algorithms can yield suboptimal and unnecessarily complex structures. Optimal decision tree methods address this through global optimization, yet remain restricted to axis-aligned threshold splits, which limit the expressivity of each node and often force deep, complex trees to capture non-linear feature effects. Shape Generalized Trees (SGTs) generalize threshold splits to learnable univariate shape functions, improving expressivity and enabling more compact trees. However, existing SGT induction algorithms are greedy and offer no optimality guarantees. In this work, we introduce Literati, the first algorithm for optimal SGT induction. We propose a novel AND/OR graph formulation of the problem that jointly optimizes tree structure and shape function complexity. To solve this AND/OR graph, we develop an AO*-based algorithm with two enhancements that improve anytime performance while preserving optimality: a secondary heuristic for OR-node selection and a round-robin policy for AND-node exploration. Across 24 real-world datasets, Literati achieves higher training and test accuracy than state-of-the-art tree approaches.

## 1 Introduction

Decision trees are one of the most widely used approaches for tabular datasets [1], largely due to their inherent interpretability [2] and robust performance [1]. Their hierarchical structure of simple threshold comparisons ($x_j \leq \theta$) can be easily visualized and audited, making them the model of choice when transparency is paramount. However, the common greedy top-down induction algorithms, such as CART [3], suffer from a fundamental limitation: by making locally optimal splits at each node, these methods often produce suboptimal and unnecessarily complex trees [4–7].

Recent years have witnessed significant progress in decision tree algorithms that overcome greedy suboptimality through global optimization [4–6, 8, 9]. These methods frame tree induction as an optimization problem over a regularized objective that balances accuracy and tree complexity, effectively optimizing both predictive performance and sparsity simultaneously [4, 6, 8]. Such non-greedy approaches have yielded trees that empirically generalize better to unseen data while reducing model complexity [4, 8].

While these algorithmic advances were underway, other work has focused on reimagining how tree branching is performed. Specifically, Upadhya and Cohen [10] proposed the Shape Generalized Tree (SGT), which replaces the axis-aligned threshold comparison with an axis-aligned *Shape Function*. By enabling non-linear decision boundaries at each node, SGTs have been empirically shown to achieve superior predictive performance with a more compact tree structure compared to threshold trees [10]. However, existing methods for learning SGTs rely on greedy induction, resulting in suboptimal trees and limited control over the learned shape functions.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

In this work, we introduce `Literati`[1], a novel approach to Optimal SGT induction. Our approach formulates SGT induction within an AND/OR graph search framework and leverages an AO*-inspired algorithm to efficiently search for an optimal SGT. Our contributions are as follows: *1) We propose a novel AND/OR graph formulation of the sparse SGT induction problem that jointly optimizes tree structure and shape function complexity, and prove that the resulting formulation yields optimal SGTs. 2) We introduce Literati, an AO*-based algorithm augmented with an inadmissible secondary heuristic for OR-node selection and round-robin AND-node exploration, yielding strong anytime performance while preserving optimality guarantees. 3) We evaluate Literati on a range of real-world datasets and demonstrate that our approach consistently outperforms state-of-the-art tree induction baselines in both train and test accuracy.*

## 2 Background

### 2.1 AND/OR Graph Search

An AND/OR graph $\mathcal{G}$ encodes problems whose solutions decompose into independent subproblems through a state-action model $(\mathcal{S}, \ell, \mathcal{A}, F, c)$ [11, 12]. Here, $\mathcal{S}$ is a set of states, $\ell \subseteq \mathcal{S}$ is the set of terminal states, $\mathcal{A}(s)$ is the set of actions available at a non-terminal state $s \in \mathcal{S} \setminus \ell$, $F : \mathcal{S} \times \mathcal{A} \to 2^{\mathcal{S}}$ maps each state-action pair to its non-empty set of successor states, and $c : \mathcal{S} \times \mathcal{A} \to \mathbb{R}_{\geq 0}$ assigns a non-negative cost to each action. Each action $a \in \mathcal{A}(s)$ has AND dynamics over its successors: $a$ solves $s$ only if every $s' \in F(s, a)$ is itself solved. Each non-terminal state $s$ has OR dynamics over its actions: $s$ is solved as soon as any single $a \in \mathcal{A}(s)$ solves it. For example, a state $s$ with two actions $a_1, a_2$ such that $F(s, a_1) = \{s_1, s_2\}$ and $F(s, a_2) = \{s_3, s_4\}$ is solved when $\{s_1$ AND $s_2\}$ are solved OR $\{s_3$ AND $s_4\}$ are solved. Terminal states are solved by definition.

Problems solvable by dynamic programming admit a natural formulation as AND/OR graph search, with subproblems mapped to states and their decompositions to actions [13, 14]. While bottom-up dynamic programming must evaluate all subproblems in the graph, a best-first search guided by admissible heuristic estimates can find a provably optimal solution while expanding only a relevant subset of the search space [5, 13, 15]. AO* realizes this approach for acyclic AND/OR graphs [11, 12]. It maintains a partial graph $G \subseteq \mathcal{G}$, initialized to the root node $\{s_0\}$, together with an admissible heuristic $V$ that lower-bounds the optimal cost at every state [12], and a *best partial policy* $\pi$ over $G$. A partial policy assigns an action $\pi(s) \in \mathcal{A}(s)$ to each non-terminal state in $G$ and is greedy with respect to $V$ at every such state, i.e. $\pi(s) = \arg\min_{a \in \mathcal{A}(s)} [c(s, a) + \sum_{s' \in F(s,a)} V(s')]$. Both $G$ and $\pi$ are refined iteratively through three phases: **selection** follows $\pi$ from $s_0$ until reaching a non-terminal tip state $s \in G$; **expansion** adds the successors of $s$ under each $a \in \mathcal{A}(s)$ to $G$ and initializes their heuristic values; and **backpropagation** revises $V$, and consequently $\pi$, bottom-up through $s$ and its ancestors. A state is marked COMPLETE once all its successors under $\pi$ are COMPLETE, with terminal states COMPLETE by definition, and the algorithm terminates when $s_0$ is COMPLETE, at which point $\pi$ is optimal [11, 12, 16]. The optimal solution graph $G^* \subseteq G$ is then collected by traversing $G$ from $s_0$ along $\pi$ until reaching terminal states. Pseudocode is provided in Algorithm 1.

**Algorithm 1:** AO$^*$ search.

**Input :** $V$, $\mathcal{G}$, $s_0$
**Output:** $\pi$
$G \leftarrow \{s_0\}$; $\pi \leftarrow \emptyset$
**while** $s_0$ *is not* COMPLETE **do**
  `// Selection`
  $s \leftarrow$ follow $\pi$ from $s_0$ to a non-terminal tip
  `// Expansion`
  $G \leftarrow G \cup$ Successors$(s)$
  Initialize $V(s')$ for each $s' \in$ Successors$(s)$
  `// Backpropagation`
  Update $V$, $\pi$, and COMPLETE status along $s \rightsquigarrow s_0$
**return** CollectSolutionGraph$(\pi, G)$

### 2.2 Optimal Decision Trees

While computationally efficient, greedy tree induction methods, such as CART [3] and C4.5 [17], often produce large trees with poor performance [4, 6]. To address this drawback, there exists a growing body of work on globally optimal tree induction. Although the problem is NP-complete [18],

[1] Named after the bonsai styling technique characterized by dramatic trunk shapes and minimal branches.

recent algorithmic advances have rendered it tractable for limited-depth trees on moderate-sized datasets. A prominent family employs branch-and-bound search with caching [8, 19, 20], with GOSDT [4] and STreeD [9] extending these methods to broader classes of objectives. All of these approaches require global pre-binarization of features. DPDT [6] instead uses a CART-based discretization that adaptively selects candidate thresholds at each node, while QuantBnB [21] and ConTree [22] operate directly on continuous features. There is also a growing line of work on *anytime optimal* tree methods that seek to identify high-quality solutions early in the search process [7, 23, 24]. Additionally, recent approaches, such as SPLIT and LicketySPLIT, trade optimality guarantees for computational efficiency by replacing parts of the search using greedy induction [25].

Particularly relevant to our work are AND/OR graph-based methods. Martelli and Montanari [15] employed AO* to seek optimal DTs in a cost-sensitivity context, while Branches [5] recently applied AO* to the sparse tree induction problem. While Literati also adopts the underlying AO* mechanism, it departs from prior work in two fundamental ways: (1) we propose a novel AND/OR graph formulation that enables optimal induction of SGTs (Section 3.1), and (2) we propose key enhancements to the AO* selection procedure that substantially improve anytime performance while preserving optimality (Section 3.2).

### 2.3 Shape Generalized Trees

Shape Generalized Trees (SGTs) [10] generalize decision trees by replacing threshold splits with learnable univariate shape functions $f(x_j)$ that map feature values to left or right branches. By capturing non-linear relations between features and the target within a single node, SGTs can reduce repeated branching on the same feature, resulting in more compact trees. As each split remains univariate, the learned shape functions can be easily visualized, preserving the inherent interpretability of decision trees [10].

However, due to their increased expressivity, inducing SGTs remains challenging. Current methods, primarily ShapeCART [10], are greedy and select shape functions by locally optimizing an impurity measure (e.g., Gini or entropy), yielding globally suboptimal trees. Post-hoc refinement methods like TAO [26] can improve these trees but are limited by the quality of the greedy initialization.

## 3 Literati

In this work, we consider a supervised classification setting over a dataset $\mathcal{D} = \{(\mathbf{x}_n, y_n)\}_{n=1}^{N}$, where each $\mathbf{x}_n \in \mathbb{R}^M$ is an $M$-dimensional feature vector and each $y_n \in \mathcal{C}$ is the corresponding class label. We seek a Shape Generalized Tree (SGT) that minimizes classification error while controlling model complexity along two axes: the number of internal nodes, which governs overall model size, and the number of discontinuities in the learned shape functions, which governs the complexity of individual branching rules. Specifically, we solve the following regularized optimization problem:

$$\begin{aligned} \min_{T \in \mathcal{T}} \quad & \text{Err}(T, \mathcal{D}) + \sum_{v \in \mathcal{I}(T)} [\lambda + \alpha \mathbf{K}(f_v)] \\ \text{s.t.} \quad & \mathbf{K}(f_v) \leq K \quad \forall\, v \in \mathcal{I}(T); \quad \mathbf{D}(T) \leq D \end{aligned} \tag{1}$$

where $\mathcal{T}$ denotes the set of all binary, axis-aligned SGTs, $\text{Err}(T, \mathcal{D})$ is the misclassification count of tree $T$ on dataset $\mathcal{D}$, and $\mathcal{I}(T)$ is the set of internal nodes of $T$. The hyperparameters $\lambda \geq 0$ and $\alpha \geq 0$ control the trade-off between error and model complexity. At each internal node $v$, a single feature $j_v$ is selected and a piecewise constant shape function $f_v \colon \mathbb{R} \to \{0, 1\}$ maps the value of feature $j_v$ to the left or right child. We define the complexity of $f_v$ as the number of discontinuities (change-points) in this function: $\mathbf{K}(f_v) = \sum_{i=1}^{u_v - 1} \mathbf{1}\big[f_v\big(x_{j_v}^{(i)}\big) \neq f_v\big(x_{j_v}^{(i+1)}\big)\big]$ where $x_{j_v}^{(1)} \leq \cdots \leq x_{j_v}^{(u_v)}$ are the sorted unique values of feature $j_v$ among instances reaching node $v$. Note that $\mathbf{K}(f_v) = 1$ recovers a standard threshold split, with higher values corresponding to more complex shape functions. The constraint $\mathbf{K}(f_v) \leq K$ with $K \geq 1$ bounds this complexity to ensure that individual branching rules remain amenable to human inspection. Similarly, the constraint on the maximum depth of the tree $\mathbf{D}(T) \leq D,\ D \geq 1$ limits the size of the tree and consequently improves interpretability [2]. Both constraints also help mitigate overfitting and can be tuned as hyperparameters.

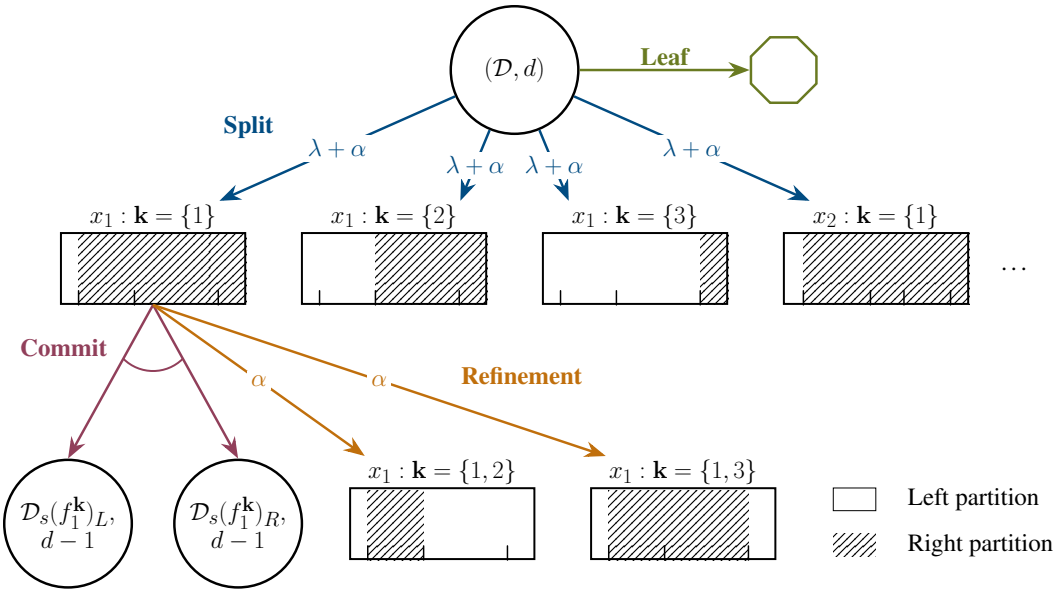



Figure 1: Visual representation of the AND/OR graph for Literati following the hypergraph notation from Nilsson [11]. Circle nodes correspond to data states and rectangle nodes correspond to Shape Function States. Each shape function state contains the candidate shape function being built. Edges connected by arcs are related via AND dynamics, edges without arcs are related via OR dynamics

Equation 1 is recursively decomposable across subtrees, and at internal node $v$, the optimal subtree cost is obtained by choosing the better of becoming a leaf or splitting with the best shape function:

$$\mathcal{L}_d^*(\mathcal{D}_v) = \min\left(\text{Err}(\mathcal{D}_v),\ \min_{f_v:\,\mathbf{K}(f_v)\le K} \lambda + \alpha\,\mathbf{K}(f_v) + \mathcal{L}_{d-1}^*(\mathcal{D}_v(f_v)_L) + \mathcal{L}_{d-1}^*(\mathcal{D}_v(f_v)_R)\right), \tag{2}$$

where $\mathcal{L}_d^*(\mathcal{D})$ denotes the cost of an optimal subtree for data $\mathcal{D}$ with depth budget $d$, computed recursively. The first argument, $\text{Err}(\mathcal{D}_v)$, is the cost of assigning the majority class at node $v$ without further splitting. $\mathcal{D}_v(f_v)_L$, $\mathcal{D}_v(f_v)_R$ are the left and right data partitions obtained by applying $f_v$. This recursion terminates when we reach the maximum depth, in which case $\mathcal{L}_0^*(\mathcal{D}_v) = \text{Err}(\mathcal{D}_v)$, or the node is pure with no error, i.e. $\text{Err}(\mathcal{D}_v) = 0$.

### 3.1 AND/OR Graph Representation

As Equation 2 decomposes separably over left and right subtrees, it can be solved via dynamic programming [9] and consequently admits a natural formulation as AND/OR graph search [13–15]. A direct translation of Equation 2 into this framework would require enumerating every feasible shape function at each internal node, leading to an explosion in the size of the partial search graph $G$. Instead, we further decompose Equation 2 by incrementally constructing shape functions one change-point at a time. This finer decomposition allows AO* to recognize suboptimal higher-order partial shape functions before they are admitted into $G$. More formally, let $\mathcal{U}_v^{(j)}$ be the set of unique values of feature $j$ from samples that reach internal node $v \in T$. The objective value at a node is defined as:

$$\mathcal{L}_d^*(\mathcal{D}_v) = \min\left(\text{Err}(\mathcal{D}_v),\ \min_{j\in\{1,\dots,M\}}\ \min_{u\in\mathcal{U}_v^{(j)}}\ \lambda + \alpha + \mathcal{L}_d^*(\mathcal{D}_v, j, \{u\})\right), \tag{3}$$

$$\mathcal{L}_d^*(\mathcal{D}_v, j, \mathbf{k}) = \min\left(\mathcal{L}_{d-1}^*(\mathcal{D}_v(f_j^{\mathbf{k}})_L) + \mathcal{L}_{d-1}^*(\mathcal{D}_v(f_j^{\mathbf{k}})_R),\ \alpha + \min_{\substack{u\in\mathcal{U}_v^{(j)},\,\max(\mathbf{k})<u,\\ |\mathbf{k}\cup\{u\}|\le K}} \mathcal{L}_d^*(\mathcal{D}_v, j, \mathbf{k}\cup\{u\})\right). \tag{4}$$

Here, $f_j^{\mathbf{k}}$ denotes the piecewise constant shape function on $j$ defined by the change-point set $\mathbf{k}$. At each stage of the recurrence (Equation 4), we can either recurse into the resulting left and right

subtrees, or introduce an additional change-point to the shape function at a cost of $\alpha$. For a state $s$ corresponding to a tree vertex $v$, we write $\mathcal{D}_s = \mathcal{D}_v$. Given this decomposition, we propose the following state-action model (depicted in Figure 1):

**Data States.** A data state $s = (\mathcal{D}_s, d)$, where $\mathcal{D}_s \subseteq \mathcal{D}$ is the subset of training instances reaching this node and $d \in \{0, 1, \ldots, D\}$ is the remaining depth budget. The root of the AND/OR graph is $s_0 = (\mathcal{D}, D)$. All children of a data state are related via OR dynamics. The available actions at a non-terminal data state are:

(i) **Split**: the split actions $a_{j,u}\ \forall j \in \{1, \ldots, M\},\ \forall u \in \mathcal{U}_s^{(j)}$. Each action transitions to the shape function state corresponding to the shape function on feature $j$ with an initial change-point on $u$. More formally, $F(s, a_{j,u}) = (\mathcal{D}_s, d, j, \{u\})$ with $c(s, a_{j,u}) = \lambda + \alpha$.

(ii) **Leaf**: The terminal action $\bar{a}$ transitions $s$ to its terminal state $\overline{(\mathcal{D}_s, d)} \in \ell$, with $c(s, \bar{a}) = 0$.

**Terminal States.** A terminal state $\overline{(\mathcal{D}_s, d)} \in \ell$ is a data state at which no further actions are available and which corresponds to a leaf node in the final tree. Data states are terminal by default when the depth budget is exhausted ($d = 0$) or all instances in $\mathcal{D}_s$ share the same class label.

**Shape Function States.** A shape function state $s_{\text{shape}} = (\mathcal{D}_s, d, j, \mathbf{k})$ represents a partially constructed branching rule on feature $j \in \{1, \ldots, M\}$, where $\mathcal{D}_s$ and $d$ are inherited from the parent data state. The set $\mathbf{k}$ contains indices of the ordered values in $\mathcal{U}_s^{(j)}$. These indices define the current piecewise constant shape function $f_j^{\mathbf{k}}$, such that $|\mathbf{k}| = \mathbf{K}(f_j^{\mathbf{k}}) \leq K$. The available actions are:

(i) **Commit**: apply the current shape function $f_j^{\mathbf{k}}$ to partition $\mathcal{D}_s$ into $\mathcal{D}_s(f_j^{\mathbf{k}})_L$ and $\mathcal{D}_s(f_j^{\mathbf{k}})_R$. The transition function returns a pair of successor data or terminal states, $F(s, a_{\text{commit}}) = \{(\mathcal{D}_s(f_j^{\mathbf{k}})_L, d-1), (\mathcal{D}_s(f_j^{\mathbf{k}})_R, d-1)\}$, which are related via AND dynamics: both must be solved, and their costs sum. The cost of this action is $c(s, a_{\text{commit}}) = 0$, as all regularization costs have already been incurred.

(ii) **Refine**: $a_{j,\mathbf{k},u}$ introduces an additional change-point $u \in \mathcal{U}_s^{(j)}$ satisfying $u > \max(\mathbf{k})$ and $|\mathbf{k} \cup \{u\}| \leq K$, transitioning to a new shape function state $F(s, a_{j,\mathbf{k},u}) = \{(\mathcal{D}_s, d, j, \mathbf{k} \cup \{u\})\}$ with $c(s, a_{j,\mathbf{k},u}) = \alpha$. The ordering constraint $u > \max(\mathbf{k})$ enforces a canonical construction of shape functions, avoiding redundant exploration of symmetric change-point sets.

The optimal cost function $V^*$ over this AND/OR graph is characterized by the Bellman equations:

$$V^*(s) = \begin{cases} \text{Err}(\mathcal{D}_s) & \text{if } s = \overline{(\mathcal{D}_s, d)} \in \ell \\ \min_{a \in \mathcal{A}(s)} Q^*(s, a) & \text{else} \end{cases}, \qquad Q^*(s, a) = c(s, a) + \sum_{s' \in F(s,a)} V^*(s'). \tag{5}$$

Our goal is to identify the optimal policy $\pi^*$ that prescribes an action at every non-terminal state reachable from the root $s_0 = (\mathcal{D}, D)$ under $\pi^*$ itself. The optimal policy is greedy with respect to $V^*$, i.e. $\pi^*(s) = \arg\min_{a \in \mathcal{A}(s)} Q^*(s, a)$, and we can use $\pi^*$ to traverse the AND/OR graph and obtain the optimal tree $T^*$.

**Proposition 1 (Validity of AND/OR Graph).** *The tree $T^*$ derived from the optimal policy $\pi^*$ is an optimal solution to the sparse SGT induction problem posed in Equation 1. Proof in Appendix A.1.*

### 3.1.1 Admissible Heuristic

To solve this AND/OR graph with AO*, we maintain a heuristic estimate $V(s)$ for each non-terminal state $s \in \mathcal{S} \setminus \ell$ which induces the current best partial policy $\pi$ that determines the next state to expand. For data states, we initialize $V(\mathcal{D}_s, d) = \min(\text{Err}(\mathcal{D}_s), \lambda + \alpha)$ [5]. For shape function states, we initialize $V(\mathcal{D}_s, d, j, \mathbf{k}) = \min\big(\alpha, \min(\text{Err}(\mathcal{D}_s(f_j^{\mathbf{k}})_L), \lambda + \alpha) + \min(\text{Err}(\mathcal{D}_s(f_j^{\mathbf{k}})_R), \lambda + \alpha)\big)$.

Whenever AO* expands a state $s$, its children are added to the search graph with their initial heuristic values, and $s$ and its ancestors are updated bottom-up as follows:

$$V(\widetilde{s}) = \min_{a \in \mathcal{A}(\widetilde{s})} Q(\widetilde{s}, a), \quad Q(\widetilde{s}, a) = c(\widetilde{s}, a) + \sum_{s' \in F(\widetilde{s}, a)} V(s'), \quad \forall\, \widetilde{s} \in \{s\} \cup \text{Ancestors}(s). \tag{6}$$

**Proposition 2 (Admissibility).** *The heuristic $V$ satisfies $V(s) \leq V^*(s)$ for every state $s$ throughout the execution of AO*. The proof is provided in Appendix A.2.*

By Proposition 2 and Proposition 1, the tree induced by $\pi^* = \pi$ at termination of AO* (Algorithm 1) is globally optimal. Furthermore, the heuristic enables early detection of terminal data states: if $\text{Err}(\mathcal{D}_s) \leq \lambda + \alpha$, no split can improve upon the leaf cost as per Corollary 1.

**Corollary 1 (Valid Inequality).** *If $\text{Err}(\mathcal{D}_s) \leq \lambda + \alpha$, then $(\mathcal{D}_s, d)$ can be marked terminal without expansion, without affecting optimality. Proof in Appendix A.2.*

### 3.1.2 Candidate Change Points

Searching for an optimal decision tree is computationally demanding, and the additional expressivity of Shape Generalized Trees makes the search space prohibitively large for deep trees or large datasets. Like prior optimal tree methods, Literati therefore restricts the search to a subset of candidate change-points. More formally, we define a candidate generation function $\Phi$, which maps each data state $(\mathcal{D}_s, d)$ to a collection of per-feature candidate sets,

$$\Phi(\mathcal{D}_s, d) = \big(U_s^{(1)}, \ldots, U_s^{(M)}\big), \tag{7}$$

where $U_s^{(j)}$ is a finite set of candidate change-points for feature $j$. Given $\Phi$, the actions available in our AND/OR graph become

$$\mathcal{A}_{\text{split}}(\mathcal{D}_s, d) = \big\{a_{j,u} : j \in \{1, \ldots, M\},\ u \in U_s^{(j)}\big\}, \tag{8}$$

$$\mathcal{A}_{\text{refine}}(\mathcal{D}_s, d, j, \mathbf{k}) = \big\{a_{j,\mathbf{k},u} : u \in U_s^{(j)},\ u > \max(\mathbf{k}),\ |\mathbf{k} \cup \{u\}| \leq K\big\}. \tag{9}$$

The generator $\Phi$ admits several instantiations. Setting $U_s^{(j)} = \mathcal{U}_s^{(j)}$ recovers the unrestricted search over all change-points. Pre-discretization, as used by most prior optimal tree methods, corresponds to a data-independent $\Phi$ that returns a fixed set of thresholds per feature regardless of the state, acting as a lookup table. These thresholds are typically feature quantiles [4, 5, 7–9] or ensemble-derived thresholds [27].

However, such data-independent discretization can miss locally informative change-points relevant to specific decision nodes. We therefore adopt the *adaptive* discretization strategy of DPDT [6]. At each data state $(\mathcal{D}_s, d)$, $\Phi$ fits a CART tree on $\mathcal{D}_s$ with at most $B_d + 1$ leaf nodes, yielding at most $B_d$ candidate thresholds in total, and returns its thresholds on each feature as $U_s^{(j)}$. The schedule $B_d$ is user-specified and typically increases with the remaining depth budget $d$, since states closer to the root have greater influence on tree structure; it provides a direct lever for trading solution quality against computational cost. In practice, this strategy yields substantial speedups in convergence while attaining higher training accuracy than quantile pre-discretization (Appendix G.2) and an objective close to that over the full candidate set (Section 4.1).

**Optimality Guarantees.** For any $\Phi$ that restricts the candidate set (i.e. $U_s^{(j)} \subsetneq \mathcal{U}_s^{(j)}$), Literati certifies optimality only over the SGTs reachable under $\Phi$, not over the original continuous feature space, a limitation shared by all discretization-based methods [4–6, 27]. The structure of this space, and hence the properties of the resultant optimality certificate, depends on $\Phi$. Under pre-discretization, Literati certifies optimality against all possible recompositions of the change-points, matching the guarantees provided by other methods requiring pre-discretization. Under adaptive discretization, Literati no longer provides closure under recomposition, but guarantees that its optimal regularized training objective is no greater than that of DPDT [6] and CART [3] under the matched conditions in Appendix A.3.

## 3.2 Anytime AO*

To support anytime behaviour, we maintain a primal bound $\overline{V}(s)$ and its induced policy $\overline{\pi}$, which tracks the best solution found so far from the subgraph rooted at $s$. Data states are initialized with $\overline{V}(\mathcal{D}_s, d) = \text{Err}(\mathcal{D}_s)$ (treating the data state as terminal). For shape function states, we eagerly compute the primal bounds of the two commit successors, and set $\overline{V}$ to their sum. $\overline{V}$ is updated on expansion using a bottom-up rule analogous to Equation 6. On early termination, we follow $\overline{\pi}$ from the root until we reach either a terminal state or an unexpanded state. If the reached state is a terminal or unexpanded data state, we treat it as a leaf of the returned tree. If it is an unexpanded shape function state, we apply the commit action and treat both resulting data states as leaves. Since

the admissible heuristic lower-bounds the value at a state, it serves as a dual bound and can be used to implement pruning strategies analogous to those used in AND/OR branch-and-bound [28]; If $Q(s,a) > \overline{V}(s)$, then the corresponding sub-graph is guaranteed to be sub-optimal, therefore we can prune $a$, i.e. $\mathcal{A}(s) := \mathcal{A}(s) \setminus \{a\}$.

While $\overline{\pi}$ provides a feasible tree at any point, we empirically observe that substantial search is required before a high-quality solution is found. This loose primal bound also limits the effectiveness of pruning, as $Q(s,a) > \overline{V}(s)$ is rarely satisfied during the early search iterations. We trace this to two distinct bottlenecks in the selection phase, one at OR dynamics and one at AND dynamics. Each bottleneck is addressed with a targeted modification that improves anytime performance while preserving the optimality of the policy returned at the termination of AO*. We describe each of them in the following sections. Detailed pseudocode and implementation details of our approach can be found in Appendix B.

### 3.2.1 Secondary Inadmissible Heuristic for OR Selection

The first issue is poor selection among sibling OR nodes, which AO* cycles through in a breadth-first-like pattern due to a weak admissible heuristic. We illustrate this for a data state $s$ in Figure 2, and a similar pattern arises for shape function states. For simplicity, we assume $\text{Err}(\mathcal{D}_s) \gg \alpha$ and $\text{Err}(\mathcal{D}_s) \gg \lambda$, and that these inequalities are preserved under splits of $\mathcal{D}_s$. Such a regime is representative of the upper levels of the tree, where partitions remain coarse, and leaf errors are substantially larger than the regularization parameters, so the regularization costs dominate. When state $s$ (root node in Figure 2) is expanded, every unexpanded shape function successor $s'$ enters $G$ with $V(s') = \min(2(\lambda+\alpha), \alpha) = \alpha$, yielding identical $Q$-values of $\lambda + 2\alpha$ across all split actions. Tie-breaking among these actions is therefore arbitrary: the algorithm expands one successor, lifting its $Q$-value above the tie, so $\pi(s)$ redirects to any remaining unexplored actions, and the pattern repeats. The result is poor anytime performance as AO* touches every split candidate once before deepening into any particular subtree.

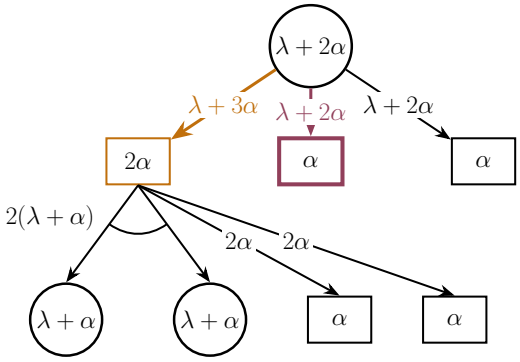


Figure 2: The blind heuristic problem. Nodes show $V$; edges show $Q$. Red: node selected this iteration; orange: node expanded in the prior iteration.

To address this problem, we introduce a secondary, more informative but potentially inadmissible, heuristic $\widetilde{V}$ to guide selection. Each state $s$ maintains both the admissible heuristic $V(s)$ and the informative heuristic $\widetilde{V}(s)$, which is updated on expansion using a bottom-up rule analogous to Equation 6. $\widetilde{V}$ induces a selection policy $\widetilde{\pi}$ that prescribes, at each state, the action with the lowest $\widetilde{Q}$-value. During selection, AO* follows $\widetilde{\pi}$ from the root to identify the next state to expand instead of $\pi$. Since $\widetilde{V}$ can potentially be inadmissible, when all successors in $F(s, \widetilde{\pi}(s))$ are COMPLETE, we cannot mark $s$ as COMPLETE. Instead, we revise $\widetilde{\pi}(s)$ to the action $a \in \mathcal{A}(s)$ with the lowest $\widetilde{Q}$-value among those whose successors are not all COMPLETE and continue the search to preserve optimality [29]. Pruning and termination remain governed exclusively by $\pi$: a state is marked COMPLETE only via its best partial policy under $V$, pruning conditions are unchanged, and AO* terminates when the root is COMPLETE under $\pi$. Since $\widetilde{V}$ and $\widetilde{\pi}$ only influence the order of expansion, the policy returned at termination is guaranteed to be optimal.

**CART-lookahead.** We propose a CART-based lookahead heuristic for $\widetilde{V}$. At a data state $(\mathcal{D}_s, d)$, we fit a CART tree to $\mathcal{D}_s$ with depth budget $d$ and set $\widetilde{V}(\mathcal{D}_s, d)$ to its misclassification count plus $(\lambda + \alpha)$ per internal node (Equation 1). For shape function states $(\mathcal{D}_s, d, j, \mathbf{k})$, we partition $\mathcal{D}_s$ via $f_j^{\mathbf{k}}$ and set $\widetilde{V}(\mathcal{D}_s, d, j, \mathbf{k}) = \widetilde{V}(\mathcal{D}_s(f_j^{\mathbf{k}})_L, d-1) + \widetilde{V}(\mathcal{D}_s(f_j^{\mathbf{k}})_R, d-1)$. Literati supports other informative heuristics, and alternatives are discussed in Appendix E.

Table 1: Training accuracy, runtime, and optimality proof rate across depth budgets averaged across all datasets evaluated, organized into Non-Optimal and Optimal approaches, with the latter grouped by discretization strategy. The dashed lines separate Literati from baselines. Bold indicates the best value per depth for training accuracy and runtime.

| | | ↑ Train Acc. (%) | | | | ↓ Runtime (s) | | | | ↑ Proof Obtained (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $K$ | $D$=3 | $D$=4 | $D$=5 | $D$=6 | $D$=3 | $D$=4 | $D$=5 | $D$=6 | $D$=3 | $D$=4 | $D$=5 | $D$=6 |
| **Non-Optimal** | | | | | | | | | | | | | |
| CART | 1 | 79.60 | 82.70 | 85.65 | 87.64 | **0.08** | **0.11** | **0.13** | **0.16** | — | — | — | — |
| AxTAO | 1 | 80.25 | 84.04 | 86.55 | 88.46 | 5.41 | 8.05 | 10.98 | 13.87 | — | — | — | — |
| ShapeCART | 3 | 80.43 | 83.64 | 86.34 | 88.52 | 1.82 | 1.96 | 3.03 | 3.84 | — | — | — | — |
| ShapeTAO | 3 | 81.28 | 85.07 | 87.33 | 89.49 | 3.85 | 8.07 | 17.26 | 29.84 | — | — | — | — |
| **Optimal (No Discretization)** | | | | | | | | | | | | | |
| ConTree | 1 | 84.12 | 86.68 | 86.13 | 84.35 | 388.18 | 1869.40 | 2542.97 | 2700.91 | 95.8 | 54.2 | 33.3 | 25.0 |
| **Optimal (Quantile Pre-Discretization)** | | | | | | | | | | | | | |
| Branches | 1 | 81.17 | 79.53 | 79.11 | 79.07 | 95.30 | 155.23 | 158.90 | 152.59 | 62.5 | 6.7 | 5.8 | 8.3 |
| STreeD | 1 | 83.50 | 86.57 | 88.44 | 88.26 | 169.90 | 413.52 | 1693.70 | 2445.12 | 99.2 | 91.7 | 65.0 | 37.5 |
| LDS-DL8.5 | 1 | 83.50 | 86.95 | 89.09 | 90.50 | 620.79 | 2459.50 | 3300.01 | 3212.81 | 87.5 | 37.5 | 8.3 | 12.5 |
| CADL8.5 | 1 | 83.50 | 86.14 | 87.29 | 84.12 | 315.62 | 1179.99 | 2481.14 | 3154.14 | 91.7 | 79.2 | 38.3 | 12.5 |
| **Optimal (Adaptive Discretization)** | | | | | | | | | | | | | |
| DPDT | 1 | 83.56 | 87.03 | 77.16 | 71.52 | 46.57 | 488.23 | 2387.41 | 3217.40 | 100.0 | 95.8 | 45.8 | 12.5 |
| Literati | 1 | 83.56 | 87.28 | 89.91 | 92.28 | 18.22 | 259.14 | 1156.29 | 2587.19 | 100.0 | 97.5 | 87.5 | 36.7 |
| | 2 | 84.03 | 87.75 | 90.36 | 92.58 | 60.82 | 851.76 | 2672.71 | 2817.87 | 100.0 | 87.5 | 33.3 | 22.5 |
| | 3 | **84.13** | **87.90** | **90.46** | **92.69** | 162.93 | 1393.71 | 2832.73 | 2782.09 | 100.0 | 82.5 | 25.0 | 24.2 |

### 3.2.2 Round Robin Selection for AND Nodes

The second issue occurs at AND dynamics. When the commit action is selected at a shape function state, the algorithm must choose which of the two resulting data states $(s_L, s_R)$ to explore first. Chaouki et al. [5] selects the child with the worst $\overline{V}$ to prioritize improving the subtree with the inferior incumbent solution. However, when primal values are significantly imbalanced - e.g. $\overline{V}(s_L) \gg \overline{V}(s_R)$ or vice versa - this rule keeps returning to $s_L$ for many iterations before $\overline{V}(s_L)$ drops enough for $s_R$ to be selected. As a result, early iterations repeatedly expand one subtree while the sibling remains unexplored, even when the larger $\overline{V}$ reflects an inherently harder subproblem. Substituting $\overline{V}$ for $V$ or $\widetilde{V}$ does not resolve this, as analogous imbalances may arise. To address this, we adopt a round-robin selection policy, and when both $s_L$ and $s_R$ are unsolved, we alternate between them on successive selection calls, ensuring balanced exploration of both subtrees. If one child is already solved, selection proceeds to the remaining unsolved child.

## 4 Experimental Evaluation

We evaluate Literati on 24 small-to-medium datasets (<500k samples) from Upadhya and Cohen [10], which includes the QuantBnB benchmark [21] used by Kohler et al. [6] and Brița et al. [22] (dataset details in Appendix C). Unless otherwise specified, results are averaged over five folds with a 70/10/20 train/val/test split. We compare Literati against SGT baselines ShapeCART and ShapeTAO [10], and non-SGT baselines CART [3], AxTAO [26], Branches [5], STreeD [9], LDS-DL8.5 [23], CADL8.5 [24], ConTree [22] (STreeD with continuous feature support), and DPDT [6]. Of these, CART, ShapeCART, AxTAO, ShapeTAO are non-optimal induction approaches. Among the optimal approaches, ConTree operates natively on continuous features, Branches, STreeD, LDS-DL8.5, and CADL8.5 require pre-discretization, while DPDT and Literati use adaptive discretization. Furthermore, LDS-DL8.5 [23] and CADL8.5 [24] both prioritize anytime performance. For runs that fail to produce any solution within the time limit (only DPDT), we impute the accuracy of a single-leaf tree. For runs that exhaust the 64 GB memory limit (only observed for Branches), we report the accuracy and runtime achieved at the time of the crash.

### 4.1 Optimization Performance

In our first experiment, we assess Literati's training performance across a range of depth and change-point budgets. We run Literati and our baselines on all datasets with a one-hour limit over depth budgets $D \in \{3, \ldots, 6\}$. For Literati, we sweep shape complexity $K \in \{1, 2, 3\}$ (where $K = 1$

recovers an axis-aligned threshold tree). ShapeTAO and ShapeCART are evaluated at $K = 3$, the most generous setting. For methods requiring pre-discretization, we preprocess features to 16 quantiles. For DPDT and Literati, we set the number of candidate thresholds to 16 for all depths ($B_d = 16\ \forall d = 1, \ldots, D$). Appendix C provides further details on the experimental setup. Table 1 reports average training accuracy and runtime. While proof rates are not directly comparable across different discretization methods, we report them where applicable for completeness. Per-dataset results and per-depth critical difference diagrams appear in Appendix G.

**Optimization Capabilities.** To isolate Literati's optimization capabilities from the expressiveness of shape functions, we first consider the $K = 1$ setting, where Literati produces axis-aligned threshold trees. Literati achieves the lowest runtime among optimal methods for $D \in \{3, 4, 5\}$, except for Branches, whose lower average runtimes stem from frequent early termination due to out-of-memory errors. Literati also achieves the highest training accuracy for $D \in \{4, 5, 6\}$ across all evaluated methods. While ConTree attains the highest training accuracy at the shallowest depth ($D = 3$), Literati significantly outperforms it for $D \in \{4, 5, 6\}$. Literati also substantially outperforms all prediscretized approaches (Branches, STreeD, LDS-DL8.5, and CADL8.5) in both training accuracy and runtime across all depths and $K$ values. In particular, Literati achieves significantly higher training accuracy than Branches, the other AO*-based approach, which frequently exhausts the 64 GB memory limit, highlighting the effectiveness of Literati's pruning strategy. Against DPDT, Literati matches its training accuracy exactly at $D = 3$ while achieving a lower runtime. Although both operate over the same search space when $K = 1$, Literati achieves a higher optimality proof rate, which we attribute to its AO*-based search reducing the number of subproblems that must be solved. Both methods also support depth-dependent $B_d$ schedules that trade solution quality for runtime, and Literati's advantage over DPDT persists under a lighter adaptive discretization regime (Appendix G).

**Effect of Shape Complexity.** Increasing $K$ substantially improves training accuracy: $K = 1$ to $K = 2$ yields $\approx 0.4\%$ across all depths, with $K = 2$ to $K = 3$ contributing a smaller but consistent gain. Consequently, $K = 3$ achieves the highest accuracy at every depth, exceeding the SGT baselines with similar $K$ (ShapeCART, ShapeTAO) and all optimal tree methods.

**Anytime Performance.** We observe that Literati's training accuracy continues to improve with depth at every $K$, even when proof rates fall, reflecting strong anytime behaviour. STreeD and LDS-DL8.5 exhibit a similar trend, but the gap between Literati and these two models widens with increasing depth. ConTree, CADL8.5, and Branches exhibit performance degradation as depth increases. DPDT does not return intermediate solutions upon timeout, resulting in poor performance at higher depths.

### 4.1.1 Selection Strategy Ablation

In this section, we provide an ablation of our selection strategies to isolate their contributions to anytime performance. Specifically, we ablate our OR selection method (best informative heuristic $\widetilde{V}$) by selecting using the admissible heuristic $V$. Additionally, we ablate our AND selection strategy (round-robin (RR)) and consider two alternatives: (1) selecting the child with the worst primal value [5] ($\uparrow \overline{V}$) and (2) selecting the child with the worst value under the OR heuristic in use ($\uparrow \widetilde{V}$ or $\uparrow V$). All configurations are evaluated on five representative datasets (Eye-Movements, Avila, Page, Fault, and Rice) across five folds with $D = 6$, $K = 3$ under a one-hour time limit. Figure 3 plots the distance to the optimal solution over time for Eye-Movements and Avila, and results for the other three datasets, plus distances across iterations, can be found in Appendix D.

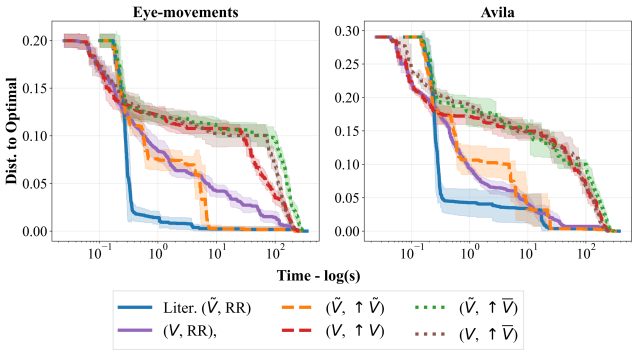



Figure 3: Dist. to Opt. Solution. Legend key: (OR Selection, AND Selection).

We observe that both the informative inadmissible heuristic and the round-robin AND selection meaningfully contribute to anytime performance, and that Literati achieves the best overall anytime performance when combining them.

### 4.2 Generalization

In addition to the optimization capabilities of Literati, we are interested in how well our approach performs on out-of-sample data points. For all methods, we tune hyperparameters using a 200-trial random search and select the trial with the best average validation accuracy across folds. To reflect practical training constraints, each model is given a 15-minute training budget per fold. Full details of each method's hyperparameter search space are provided in Appendix C.1. Across datasets and folds, Table 2 reports average test accuracy, and Figure 4 shows the corresponding critical difference diagram. Per dataset results are provided in Appendix H.

We observe that Literati achieves an average test accuracy 1.0% higher than the second-place approach, ShapeTAO, and a statistically significantly higher average rank than all evaluated baseline approaches.

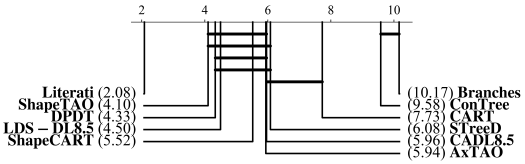



Figure 4: CDD of test accuracy. Axis shows each method's mean rank across datasets and folds; bars indicate statistical equivalence ($p \geq 0.05$, Wilcoxon).

Table 2: Average Test Accuracy Results.

| Model | Test Acc. (%) |
|---|---|
| CART | 87.6 |
| AxTAO | 88.5 |
| ShapeCART | 89.1 |
| ShapeTAO | <u>89.6</u> |
| Branches | 83.7 |
| STreeD | 88.5 |
| LDS-DL8.5 | <u>89.6</u> |
| CADL8.5 | 88.6 |
| ConTree | 85.5 |
| DPDT | 89.3 |
| Literati | **90.6** |

## 5 Conclusion

We introduced Literati, the first algorithm for optimal Shape Generalized Tree induction. By formulating SGT learning as an AND/OR graph search that jointly optimizes tree structure and shape function complexity, Literati provides optimality guarantees while retaining anytime performance through two modifications to AO* selection: an inadmissible secondary heuristic at OR nodes and round-robin exploration at AND nodes. Across 24 benchmark datasets, Literati achieves superior performance over all evaluated optimal-tree and greedy SGT baselines, establishing heuristic search as a strong foundation for SGT induction. Future work includes exploring additional admissible and informative heuristics within our AO* framework.

**Limitations.** As a tree-based model, Literati, like other SGTs, inherits intrinsic constraints of traditional decision trees, most notably that they are designed primarily for tabular data. Additionally, like other optimal tree induction approaches, Literati will struggle to prove optimality on datasets with a large number of samples within short time budgets. Nevertheless, the strong anytime performance of Literati makes it a valuable and practical tool for data science workflows.

**Broader Impact.** As ML increasingly governs high-stakes decisions in healthcare [30, 31], finance [32, 33], and manufacturing [34, 35], the need for models that balance accuracy and interpretability is pressing. Literati contributes to this need by producing inherently interpretable trees that outperform standard tree-based baselines in our experiments.

## 6 Acknowledgments and Disclosure of Funding

This research was supported by grant number DSI-CGY3R1P18 from the Data Sciences Institute at the University of Toronto. The authors also gratefully acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References


[1] Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? *Advances in neural information processing systems*, 35:507–520, 2022.

[2] Mitja Luštrek, Matjaž Gams, Sanda Martinčić-Ipšić, et al. What makes classification trees comprehensible? *Expert Systems with Applications*, 62:333–346, 2016.

[3] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[4] 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.

[5] Ayman Chaouki, Jesse Read, and Albert Bifet. Branches: Efficiently seeking optimal sparse decision trees via ao. In *Forty-second International Conference on Machine Learning*, 2025.

[6] Hector Kohler, Riad Akrour, and Philippe Preux. Breiman meets bellman: Non-greedy decision trees with mdps. In *KDD 2025-The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025.

[7] Emir Demirović, Emmanuel Hebrard, and Louis Jean. Blossom: an anytime algorithm for computing optimal decision trees. In *International Conference on Machine Learning*, pages 7533–7562. PMLR, 2023.

[8] Gaël Aglin, Siegfried Nijssen, and Pierre Schaus. Learning optimal decision trees using caching branch-and-bound search. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3146–3153, 2020.

[9] Jacobus van der Linden, Mathijs de Weerdt, and Emir Demirović. Necessary and sufficient conditions for optimal decision trees using dynamic programming. *Advances in Neural Information Processing Systems*, 36:9173–9212, 2023.

[10] Nakul Upadhya and Eldan Cohen. Empowering decision trees via shape function branching. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[11] Nils J Nilsson. *Principles of artificial intelligence*. Morgan Kaufmann, 2014.

[12] Blai Bonet and Héctor Geffner. An algorithm better than ao*? In *AAAI*, pages 1343–1348, 2005.

[13] Alberto Martelli, Ugo Montanari, et al. Additive and/or graphs. In *IJCAI*, volume 73, pages 1–11, 1973.

[14] Alberto Martelli, U Montanari, et al. From dynamic programming to search algorithms with functional costs. In *Proc. Fourth Int. Joint Conf. Artif. Intell. Tbilisi*, pages 345–350, 1975.

[15] Alberto Martelli and Ugo Montanari. Optimizing decision trees through heuristically guided search. *Communications of the ACM*, 21(12):1025–1039, 1978.

[16] Blai Bonet and Hector Geffner. Action selection for mdps: Anytime ao* versus uct. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pages 1749–1755, 2012.

[17] J Ross Quinlan. *C4.5: programs for machine learning*. Elsevier, 2014.

[18] Hyafil Laurent and Ronald L Rivest. Constructing optimal binary decision trees is np-complete. *Information processing letters*, 5(1):15–17, 1976.

[19] Xiyang Hu, Cynthia Rudin, and Margo Seltzer. Optimal sparse decision trees. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32, pages 7265–7273. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper_files/paper/2019/file/ac52c626afc10d4075708ac4c778ddfc-Paper.pdf`.

[20] Emir Demirović, Anna Lukina, Emmanuel Hebrard, Jeffrey Chan, James Bailey, Christopher Leckie, Kotagiri Ramamohanarao, and Peter J Stuckey. Murtree: Optimal decision trees via dynamic programming and search. *Journal of Machine Learning Research*, 23(26):1–47, 2022.

[21] Rahul Mazumder, Xiang Meng, and Haoyue Wang. Quant-BnB: A scalable branch-and-bound method for optimal decision trees with continuous features. In Kamalika Chaudhuri,

Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 15255–15277. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/mazumder22a.html.

[22] Cătălin E Briţa, Jacobus GM van der Linden, and Emir Demirović. Optimal classification trees for continuous feature data using dynamic programming with branch-and-bound. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11131–11139, 2025.

[23] Harold Kiossou, Pierre Schaus, Siegfried Nijssen, and Vinasetan Ratheil Houndji. Time constrained dl8.5 using limited discrepancy search. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 443–459. Springer, 2022.

[24] Harold Kiossou and Pierre Schaus. A generic complete anytime beam search for optimal decision tree. In *International Symposium on Intelligent Data Analysis*, pages 97–109. Springer, 2026.

[25] Varun Babbar, Hayden McTavish, Cynthia Rudin, and Margo Seltzer. Near optimal decision trees in a split second. *arXiv preprint arXiv:2502.15988*, 2025.

[26] Miguel A Carreira-Perpinán and Pooya Tavallali. Alternating optimization of decision trees, with application to learning sparse oblique trees. *Advances in neural information processing systems*, 31, 2018.

[27] Hayden McTavish, Chudi Zhong, Reto Achermann, Ilias Karimalis, Jacques Chen, Cynthia Rudin, and Margo Seltzer. Fast sparse decision tree optimization via reference ensembles. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 9604–9613, 2022.

[28] Radu Marinescu and Rina Dechter. And/or branch-and-bound search for combinatorial optimization in graphical models. *Artificial Intelligence*, 173(16-17):1457–1491, 2009.

[29] Eric A Hansen and Rong Zhou. Anytime heuristic search. *Journal of Artificial Intelligence Research*, 28:267–297, 2007.

[30] Gregor Stiglic, Primoz Kocbek, Nino Fijacko, Marinka Zitnik, Katrien Verbert, and Leona Cilar. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5):e1379, 2020.

[31] K. Srinivas, G. Raghavendra Rao, and A. Govardhan. Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. In *2010 5th International Conference on Computer Science & Education*, pages 1344–1349, 2010. doi: 10.1109/ICCSE.2010.5593711.

[32] Jozef Zurada. Could decision trees improve the classification accuracy and interpretability of loan granting decisions? In *2010 43rd Hawaii International Conference on System Sciences*, pages 1–9. IEEE, 2010.

[33] Pan Tang, Tiantian Tang, and Chennuo Lu. Predicting systemic financial risk with interpretable machine learning. *The North American Journal of Economics and Finance*, 71:102088, 2024. ISSN 1062-9408. doi: https://doi.org/10.1016/j.najef.2024.102088. URL https://www.sciencedirect.com/science/article/pii/S1062940824000123.

[34] Choi Sungsu, Battulga Lkhagvadorj, Aziz Nasridinov, Kwan-Hee Yoo, et al. A decision tree approach for identifying defective products in the manufacturing process. *International Journal of Contents*, 13(2):57–65, 2017.

[35] Sadeq Heydarbakian and Mehran Spehri. Interpretable machine learning to improve supply chain resilience, an industry 4.0 recipe. *IFAC-PapersOnLine*, 55(10):2834–2839, 2022. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2022.10.160. URL https://www.sciencedirect.com/science/article/pii/S2405896322021747. 10th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2022.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[37] Christoph Molnar. *Interpretable machine learning*. Lulu.com, 2020.

[38] Victor Feitosa Souza, Ferdinando Cicalese, Eduardo Laber, and Marco Molinaro. Decision trees with short explainable rules. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 12365–12379. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/500637d931d4feb99d5cce84af1f53ba-Paper-Conference.pdf.

[39] Dimitris Bertsimas and Jack Dunn. Optimal classification trees. *Machine Learning*, 106: 1039–1082, 2017.

# A Theoretical Results

## A.1 Validity of AND/OR formulation

Let $\mathcal{L}_D^*(\mathcal{D})$ denote the optimum of Equation 1. We first establish that the recurrences in Equations 3–4 compute $\mathcal{L}_D^*(\mathcal{D})$ (Lemma 1), then show that the Bellman equation on $(\mathcal{S}, \ell, \mathcal{A}, F, c)$ is the exact instantiation of these recurrences (Lemma 2). Composing the two yields $V^*(s_0) = \mathcal{L}_D^*(\mathcal{D})$, and the greedy policy $\pi^*$ traces out a corresponding optimal tree.

**Lemma 1 (Validity of the Recurrence).** *The function $\mathcal{L}^*$ defined by Equations 3–4 satisfies* $\mathcal{L}_D^*(\mathcal{D}) = \min_{T \in \mathcal{T}} \{\text{Err}(T, \mathcal{D}) + \sum_{v \in \mathcal{I}(T)} [\lambda + \alpha \mathbf{K}(f_v)] : \mathbf{K}(f_v) \leq K \; \forall v, \; \mathbf{D}(T) \leq D\}$.

*Proof.* The regularized loss is additive across the subtrees of an internal node, and both feasibility constraints are hereditary: every subtree of a feasible tree is itself feasible. Specifically, $\mathbf{K}(f_v) \leq K$ is local to node $v$, and the depth budget at depth $d$ is at most $D - d$, which is exactly what the recurrence's depth argument tracks. Since an optimal tree must be composed of optimal subtrees, the minimization over trees decomposes into nested minimizations over subtrees, reducing Equation 1 to Equation 2.

We now show that the incremental recurrence in Equations 3–4 enumerates the same feasible shape functions as the inner minimization of Equation 2 at identical cost. Every feasible shape function for node $v$ on feature $j$ is determined by its sorted change-point set $\mathbf{k} \subseteq \mathcal{U}_v^{(j)}, |\mathbf{k}| \leq K$ along with a choice of which child each interval routes to. Given $j$ and $\mathbf{k}$, there are exactly two such functions, $f_j^{\mathbf{k}}$ and $\bar{f}_j^{\mathbf{k}}$, which partition $\mathcal{D}_v$ into the same two groups but route them to opposite children. We adopt the convention $f_j^{\mathbf{k}}(x) = 0$ for $x < \min(\mathbf{k})$, and enumerate only $f_j^{\mathbf{k}}$. This loses no optimal trees: $f_j^{\mathbf{k}}$ and $\bar{f}_j^{\mathbf{k}}$ produce the same two data subsets up to a swap of $\mathcal{D}_v(f_j^{\mathbf{k}})_L$ and $\mathcal{D}_v(f_j^{\mathbf{k}})_R$, and the recurrence $\mathcal{L}_{d-1}^*(\mathcal{D}_v(f_j^{\mathbf{k}})_L) + \mathcal{L}_{d-1}^*(\mathcal{D}_v(f_j^{\mathbf{k}})_R)$ is symmetric in its two arguments, so both achieve identical cost. A set of change-points $\mathbf{k} = \{u_1, u_2, \ldots, u_m\}$ is constructed via the sequence $\text{split}(j, u_1) \to \text{refine}(u_2) \to \cdots \to \text{refine}(u_m) \to \text{commit}$, accumulating cost $(\lambda + \alpha) + (m-1)\alpha + 0 = \lambda + \alpha m$, which matches $\lambda + \alpha \mathbf{K}(f_v)$. Given $K \geq 1$, constant shape functions ($\mathbf{K}(f_j^{\mathbf{k}}) = 0$) are not enumerated. This is without loss of optimality, since a constant $f_v$ sends all of $\mathcal{D}_v$ to one child and yields a degenerate split with cost $\lambda + \mathcal{L}_{d-1}^*(\mathcal{D}_v) \geq \mathcal{L}_d^*(\mathcal{D}_v)$, which is dominated by either the leaf action or any nontrivial split. The ordering constraint $u > \max(\mathbf{k})$ on refine does not exclude any shape function; it only prevents symmetric shape functions. Hence the minimization in Equations 3–4 corresponds to the minimization over feasible $f_v$ in Equation 2, with identical costs. □

**Lemma 2 (AND/OR Graph Translation).** *For every data state $(\mathcal{D}_s, d) \in \mathcal{S}$ and shape function state $(\mathcal{D}_s, d, j, \mathbf{k}) \in \mathcal{S}$,*

$$V^*(\mathcal{D}_s, d) = \mathcal{L}_d^*(\mathcal{D}_s), \qquad V^*(\mathcal{D}_s, d, j, \mathbf{k}) = \mathcal{L}_d^*(\mathcal{D}_s, j, \mathbf{k}). \tag{10}$$

*Proof.* We show that $V^*$ satisfies the same recurrence as $\mathcal{L}^*$ over the same domain with the same base cases.

*Terminal states.* For $\overline{(\mathcal{D}_s, d)} \in \ell$, Equation 5 gives $V^*(s) = \text{Err}(\mathcal{D}_s)$, matching the base case of Equation 3. This covers both terminal conditions: depth exhaustion ($d = 0$) and label purity ($\text{Err}(\mathcal{D}_s) = 0$). For pure nodes the equality reads $V^*(s) = 0 = \mathcal{L}_d^*(\mathcal{D}_s)$, which agrees with the base case of Equation 3 since $\text{Err}(\mathcal{D}_s) = 0$ dominates any split cost $\lambda + \alpha \geq 0$.

*Data states.* For a non-terminal data state $s = (\mathcal{D}_s, d)$, the leaf action gives $Q^*(s, \bar{a}) = c(s, \bar{a}) + V^*(\overline{(\mathcal{D}_s, d)}) = 0 + \text{Err}(\mathcal{D}_s)$. Each split action gives $Q^*(s, a_{j,u}) = c(s, a_{j,u}) + V^*(\mathcal{D}_s, d, j, \{u\}) = \lambda + \alpha + V^*(\mathcal{D}_s, d, j, \{u\})$. Equation 5 then yields

$$V^*(\mathcal{D}_s, d) = \min\left(\text{Err}(\mathcal{D}_s), \min_{j \in \{1,\ldots,M\}} \min_{u \in \mathcal{U}_s^{(j)}} \lambda + \alpha + V^*(\mathcal{D}_s, d, j, \{u\})\right), \tag{11}$$

which is Equation 3.

*Shape function states.* For $s = (\mathcal{D}_s, d, j, \mathbf{k})$, the commit action transitions to the AND pair $((\mathcal{D}_s(f_j^{\mathbf{k}})_L, d-1), (\mathcal{D}_s(f_j^{\mathbf{k}})_R, d-1))$, and AND dynamics sum the children, giving $Q^*(s, a_{\text{commit}}) =$

$0 + V^*(\mathcal{D}_s(f_j^{\mathbf{k}})_L, d-1) + V^*(\mathcal{D}_s(f_j^{\mathbf{k}})_R, d-1)$. Each refine action gives $Q^*(s, a_{j,\mathbf{k},u}) = \alpha + V^*(\mathcal{D}_s, d, j, \mathbf{k} \cup \{u\})$. Equation 5 yields

$$V^*(\mathcal{D}_s, d, j, \mathbf{k}) = \min\left(V^*(\mathcal{D}_s(f_j^{\mathbf{k}})_L, d-1) + V^*(\mathcal{D}_s(f_j^{\mathbf{k}})_R, d-1),\ \alpha + \min_{\substack{u \in \mathcal{U}_s^{(j)},\, u > \max(\mathbf{k}),\\ |\mathbf{k} \cup \{u\}| \le K}} V^*(\mathcal{D}_s, d, j, \mathbf{k} \cup \{u\})\right), \quad (12)$$

which is Equation 4.

Hence $V^*$ and $\mathcal{L}^*$ satisfy identical recurrences with identical base cases, and they are well defined on the same finite domain (terminated by the depth budget and purity conditions). It follows that $V^* = \mathcal{L}^*$ pointwise on both state types. □

**Proof of Proposition 1.** At the root $s_0 = (\mathcal{D}, D)$, Lemma 2 gives $V^*(s_0) = \mathcal{L}_D^*(\mathcal{D})$, and Lemma 1 gives that $\mathcal{L}_D^*(\mathcal{D})$ equals the minimum of Equation 1. Hence $V^*(s_0)$ equals the optimum of the sparse SGT induction problem. As a result, the optimal tree $T^*$ can be obtained by following $\pi^*$ from the root node $s_0$ until terminal states are reached. More specifically, at each reachable data state $s$, if $\pi^*(s) = \bar{a}$, the corresponding node in $T^*$ is a leaf labeled by the majority class of $\mathcal{D}_s$; if $\pi^*(s) = a_{j,u}$, the node is internal with feature $j$, and $\pi^*$ is followed through the chain of shape function states until $\pi^*$ selects commit at some change-point set $\mathbf{k}^*(s)$; the branching rule at the node is then $f_j^{\mathbf{k}^*(s)}$, and the construction recurses on the two child data states. The extracted $T^*$ is feasible for Equation 1 by construction: the depth constraint $\mathbf{D}(T^*) \le D$ holds because $d$ strictly decreases by one on every commit transition and $d = D$ at $s_0$, so no path from the root to a leaf in $T^*$ exceeds depth $D$; the complexity constraint $\mathbf{K}(f_j^{\mathbf{k}^*(s)}) \le K$ holds because the refine action is guarded by $|\mathbf{k}| \le K$, so $|\mathbf{k}^*(s)| \le K$ at every internal node. Hence $T^*$ is optimal. □

### A.2 Admissibility of Heuristic

We prove by induction on the iterations of AO* that $V(s) \le V^*(s)$ for every state $s$ in the explicit search graph. We first establish that the initial heuristic values satisfy admissibility (Lemma 3), then show that the bottom-up updates in Equation 6 preserve admissibility (Lemma 4). The proposition then follows by induction, and the corollary on the valid inequality is immediate.

**Lemma 3 (Initial Admissibility).** *The initialization* $V(s) = V(\mathcal{D}_s, d) = \min(\mathit{Err}(\mathcal{D}_s),\ \lambda + \alpha)$ *at non-terminal data states and* $V(s) = V(\mathcal{D}_s, d, j, \mathbf{k}) = \min\big(\alpha,\ \min(\mathit{Err}(\mathcal{D}_s(f_j^{\mathbf{k}})_L),\ \lambda + \alpha) + \min(\mathit{Err}(\mathcal{D}_s(f_j^{\mathbf{k}})_R),\ \lambda + \alpha)\big)$ *at shape function states satisfies* $V(s) \le V^*(s)$.

*Proof.* All transition costs are non-negative ($\lambda, \alpha \ge 0$ and commit and leaf costs are 0), and the terminal values $\mathrm{Err}(\mathcal{D}_s) \ge 0$. Unrolling the Bellman recursion in Equation 5, $V^*(s)$ is a sum of non-negative quantities, so $V^*(s) \ge 0$ for every state $s$. For a terminal state, by definition $V^*(s) = V(s)$.

For a non-terminal data state $s = (\mathcal{D}_s, d)$, Equation 5 gives

$$V^*(s) \ = \ \min\left(Q^*(s, \bar{a}),\ \min_{j \in \{1,\dots,M\},\, u \in \mathcal{U}_s^{(j)}} Q^*(s, a_{j,u})\right). \quad (13)$$

The leaf action transitions to the terminal state $\overline{(\mathcal{D}_s, d)}$ at zero cost, so $Q^*(s, \bar{a}) = \mathrm{Err}(\mathcal{D}_s)$. Each split action incurs cost $\lambda + \alpha$ and transitions to a shape function state, so $Q^*(s, a_{j,u}) = \lambda + \alpha + V^*(\mathcal{D}_s, d, j, \{u\}) \ge \lambda + \alpha$. Therefore

$$V^*(s) \ \ge \ \min(\mathrm{Err}(\mathcal{D}_s),\ \lambda + \alpha) \ = \ V(s). \quad (14)$$

For a shape function state $s = (\mathcal{D}_s, d, j, \mathbf{k})$, we can either commit $f_j^{\mathbf{k}}$ to obtain the two data states $s_L = (\mathcal{D}_s(f_j^{\mathbf{k}})_L, d-1)$ and $s_R = (\mathcal{D}_s(f_j^{\mathbf{k}})_R, d-1)$, or refine and pay a cost of $\alpha$. By the data-state bound established above, the cost of committing satisfies

$$\begin{aligned} Q^*(s, a_{\mathrm{commit}}) &= V^*(s_L) + V^*(s_R) \\ &\ge \sum_{b \in \{L,R\}} \min\big(\mathrm{Err}(\mathcal{D}_s(f_j^{\mathbf{k}})_b), \lambda + \alpha\big). \end{aligned}$$

This bound also holds for terminal children, whose values equal their leaf errors. Each refine action incurs a cost of $\alpha$ and transitions to a state with non-negative optimal cost, so $Q^*(s, a_{j,\mathbf{k},u}) \geq \alpha$. Therefore,

$$V^*(s) \geq \min\Big(\alpha, \sum_{b\in\{L,R\}} \min\big(\mathrm{Err}(\mathcal{D}_s(f_j^{\mathbf{k}})_b), \lambda + \alpha\big)\Big) = V(s).$$

If no refine action is available, commit is the only action and the same bound holds. □

**Lemma 4 (Update Admissibility).** *Fix a non-terminal state $\widetilde{s}$ and suppose $V(s') \leq V^*(s')$ for every $s' \in F(\widetilde{s}, a)$ and every $a \in \mathcal{A}(\widetilde{s})$. Then the update in Equation 6 yields $V(\widetilde{s}) \leq V^*(\widetilde{s})$.*

*Proof.* By the update and the hypothesis,

$$\begin{aligned} V(\widetilde{s}) \;&=\; \min_{a\in\mathcal{A}(\widetilde{s})} \left[ c(\widetilde{s}, a) + \sum_{s'\in F(\widetilde{s},a)} V(s') \right] \\ &\leq\; \min_{a\in\mathcal{A}(\widetilde{s})} \left[ c(\widetilde{s}, a) + \sum_{s'\in F(\widetilde{s},a)} V^*(s') \right] \;=\; V^*(\widetilde{s}). \quad \square \end{aligned} \tag{15}$$

**Proof of Proposition 2.** We proceed by induction on the iteration count of AO*. At initialization, the search graph contains only the root $s_0 = (\mathcal{D}, D)$, which satisfies $V(s_0) \leq V^*(s_0)$ by Lemma 3.

Assume $V(s) \leq V^*(s)$ for every state $s$ in the partial search graph $G$ at the start of an iteration. During an iteration, AO* expands some non-terminal state $s$ and adds the successors of every action $a \in \mathcal{A}(s)$ to $G$ with their initial heuristic values; these satisfy admissibility by Lemma 3. AO* then updates $V$ for $s$ and its ancestors in bottom-up order via Equation 6. At the time each update is performed, every successor referenced in the update is admissible: either it was admissible before iteration $t$ (inductive hypothesis), it was just added with an admissible initial value (Lemma 3), or it was itself updated earlier in the bottom-up sweep and is admissible by Lemma 4. A further application of Lemma 4 then gives admissibility at $\widetilde{s}$. Iterating over all ancestors preserves admissibility throughout $G$. □

**Proof of Corollary 1.** Suppose $\mathrm{Err}(\mathcal{D}_s) \leq \lambda + \alpha$ at a data state $s = (\mathcal{D}_s, d)$. For every split action,

$$Q^*(s, a_{j,u}) \;=\; \lambda + \alpha + V^*(\mathcal{D}_s, d, j, \{u\}) \;\geq\; \lambda + \alpha \;\geq\; \mathrm{Err}(\mathcal{D}_s) \;=\; Q^*(s, \bar{a}), \tag{16}$$

so the leaf action is optimal and $V^*(s) = \mathrm{Err}(\mathcal{D}_s)$. □

### A.3 Upper Bound on Optimal Regularized Training Loss

**Proposition 3 (Dominance over DPDT).** *Suppose Literati and DPDT [6] are applied to the same dataset $\mathcal{D}$ under the following matched conditions: (i) identical depth budget $D$; (ii) DPDT uses the per-internal-node complexity penalty $C(T) = |\mathcal{I}(T)|$ with regularization weight $\alpha_{\mathrm{DPDT}} = \lambda + \alpha$; and (iii) both methods use identical CART-based adaptive change-point candidate sets at every data state [6] with identical $B_d$ schedules. Let $\mathcal{L}^*_{\mathrm{Liter.}}$ and $\mathcal{L}^*_{\mathrm{DPDT}}$ denote the optimal objective values attained by Literati and DPDT, respectively. Then $\mathcal{L}^*_{\mathrm{Liter.}} \leq \mathcal{L}^*_{\mathrm{DPDT}}$, with equality whenever $K = 1$.*

Throughout this proof, we denote $\mathcal{L}_{\mathrm{Liter}}(T)$ for the Literati objective in Equation 1 and $\mathcal{L}_{\mathrm{DPDT}}(T)$ for the DPDT objective under the node-count complexity penalty variant of Kohler et al. [6, Eq. 1]. We adopt two bookkeeping conventions to align the two formulations. First, DPDT is formulated as a maximization of expected return $J_\alpha(\pi)$, whereas Literati is formulated as a minimization of regularized loss; since $\mathcal{L}_{\mathrm{DPDT}}(T) = -J_\alpha(\pi)$ for the policy $\pi$ inducing $T$ [6, Proposition 4.1], we negate the DPDT return and work with the resulting loss throughout. Second, DPDT uses a normalized misclassification rate $\frac{1}{N}\sum_{i=1}^{N} \ell(y_i, T(x_i))$, whereas Literati uses the unnormalized count $\mathrm{Err}(T, \mathcal{D})$; since $1/N$ is a positive constant scaling factor that does not affect the argmin, we work with the unnormalized convention for both objectives, with the understanding that the DPDT regularization weight is rescaled accordingly. These conventions place the two objectives in a common form; the precise correspondence between their regularization terms is established below.

We first show that for any axis-aligned threshold tree $T$, the Literati objective with $K = 1$ and the DPDT objective with matched regularization weight $\alpha_{\mathrm{DPDT}} = \lambda + \alpha$ coincide. Writing both objectives as a sum of an error term and a per-internal-node complexity, we obtain the following:

$$\mathcal{L}_{\mathrm{Liter}}(T) = \mathrm{Err}(T, \mathcal{D}) + \sum_{v \in \mathcal{I}(T)} \left[\lambda + \alpha\, \mathbf{K}(f_v)\right], \tag{17}$$

$$\mathcal{L}_{\mathrm{DPDT}}(T) = \mathrm{Err}(T, \mathcal{D}) + \sum_{v \in \mathcal{I}(T)} \alpha_{\mathrm{DPDT}}. \tag{18}$$

Under $K = 1$, the constraint $\mathbf{K}(f_v) \leq K$ combined with the requirement that $f_v$ be non-constant at any internal node (else $v$ would not induce a split) pins $\mathbf{K}(f_v) = 1$ at every $v \in \mathcal{I}(T)$. The Literati per-node penalty thus reduces to $\lambda + \alpha \cdot 1 = \lambda + \alpha$, and applying the matching condition $\alpha_{\mathrm{DPDT}} = \lambda + \alpha$ yields

$$\sum_{v \in \mathcal{I}(T)} \left[\lambda + \alpha\, \mathbf{K}(f_v)\right] = \sum_{v \in \mathcal{I}(T)} (\lambda + \alpha) = \sum_{v \in \mathcal{I}(T)} \alpha_{\mathrm{DPDT}}. \tag{19}$$

The error terms agree by construction: a Literati node with change-point set $\mathbf{k} = \{u\}$ on feature $j$ routes samples by $\mathbf{1}[x_j > u]$, which is the DPDT threshold split $x_j \leq u$ up to a left-right relabeling. Routing-equivalent trees have identical $\mathrm{Err}(\cdot, \mathcal{D})$. Combining the matched error and complexity terms gives

$$\mathcal{L}^{K=1}_{\mathrm{Liter}}(T) = \mathcal{L}_{\mathrm{DPDT}}(T) \qquad \forall\, T \in \mathcal{T}_{\mathrm{DPDT}}. \tag{20}$$

By assumption (iii) of Proposition 3, Literati and DPDT employ the same CART-based discretization function $\Phi$. Provided the candidate schedule $B_d$ is the same for all depths, DPDT and Literati utilize identical collections of per-feature candidate sets $\{U_s^{(j)}\}_{j=1}^{M}$ for any dataset $\mathcal{D}_s \subseteq \mathcal{D}$. Under $K = 1$, the constraint $|\mathbf{k}| \leq K$ forces the Literati refine action set $\mathcal{A}_{\mathrm{refine}}$ to be empty at every shape function state, leaving the commit action $a_{\mathrm{commit}}$ as the only available action. Each shape function state $(\mathcal{D}_s, d, j, \{u\})$ thus deterministically transitions to the data state pair $\big((\mathcal{D}_s(f_j^{\{u\}})_L, d-1), (\mathcal{D}_s(f_j^{\{u\}})_R, d-1)\big)$ with cost zero, exactly mirroring the partition induced by the DPDT split $a_{j,u}$ [6]. The two methods therefore induce identical search spaces, yielding the same discretized feasible set $\mathcal{T}_{\mathrm{DPDT}} = \mathcal{T}^{K=1}_{\mathrm{Liter}}$. Combined with Equation 20, this gives $\mathcal{L}^*_{\mathrm{Liter}}\big|_{K=1} = \mathcal{L}^*_{\mathrm{DPDT}}$. For $K \geq 2$, the constraint $|\mathbf{k}| \leq K$ relaxes that of $K = 1$, so $\mathcal{T}^{K=1}_{\mathrm{Liter}} \subseteq \mathcal{T}^{K}_{\mathrm{Liter}}$ and minimizing the Literati objective over the larger set gives $\mathcal{L}^*_{\mathrm{Liter}}\big|_{K\geq 2} \leq \mathcal{L}^*_{\mathrm{Liter}}\big|_{K=1} = \mathcal{L}^*_{\mathrm{DPDT}}$. □

Under the corresponding matched conditions, CART's regularized loss upper-bounds the optimal regularized loss of DPDT [6] and, consequently, that of Literati.

### A.4 Growth Bound of our AND/OR Graph

We analyze the size of the AND/OR graph defined in Section 3.1 in terms of the number of training instances $N$, features $M$, maximum tree depth $D$, and shape function complexity budget $K$, by examining the branching factor at each state type.

Each non-terminal data state $(\mathcal{D}_s, d)$ has at most $\sum_{j=1}^{M} |\mathcal{U}_s^{(j)}| \leq MN$ split successors (one shape function state per feature–change-point pair) and one terminal successor. Each shape function state $(\mathcal{D}_s, d, j, \mathbf{k})$ has at most $|\mathcal{U}_s^{(j)}| - |\mathbf{k}| \leq N - 1$ refine successors (since the ordering constraint $u > \max(\mathbf{k})$ excludes $\max(\mathbf{k})$ and all values below it) and one commit action producing two AND-successors at depth budget $d - 1$, for at most $N + 1$ successor states. Since $|\mathbf{k}|$ ranges from 1 to $K$, there are up to $K$ successive layers of shape function states between consecutive data state levels. The effective branching factor from a data state at depth budget $d$ to child data states at depth budget $d - 1$ is therefore $O(MN^K)$, and applying this recursively across $D$ depth levels yields

$$|\mathcal{S}_{\mathrm{data}}| = O\big((MN^K)^D\big). \tag{21}$$

Each data state gives rise to $O(MN^K)$ shape function states, so the total state space is

$$|\mathcal{S}| = O\big((MN^K)^{D+1}\big), \tag{22}$$

dominated by shape function states. This exponential growth motivates the anytime property of Literati: a procedure that achieves the optimal or near-optimal solution early, without exhaustively expanding the graph, is critical for practical scalability.

### A.4.1 Adaptive Discretization

Here, we analyze the growth rate of the AND/OR graph under adaptive discretization. Given a discretization schedule $B_d$, $d = 1, \ldots, D$, fitting a CART tree with at most $B_d + 1$ leaves at each data state with depth budget $d$ yields at most $B_d$ candidate thresholds distributed across features, so $\sum_{j=1}^{M} |U_s^{(j)}| \leq B_d$. The branching factor at a data state reduces from $MN$ to at most $B_d$ split successors, and the refine branching at each shape function layer reduces from $N - 1$ to at most $B_d - 1$. With up to $K$ shape function layers, the effective branching from data states at depth budget $d$ to child data states at depth budget $d - 1$ is $O(B_d^K)$. Since this branching factor varies with $d$, the total state space is bounded by the product across depth levels:

$$|\mathcal{S}| = O\left(D \cdot \prod_{d=1}^{D} B_d^K\right). \tag{23}$$

As $B_d \ll M \cdot N$, adaptive discretization significantly reduces the size of the AND/OR graph.

## B Literati Algorithmic Details

In this section, we provide more details on our algorithm, including notable design decisions that may be useful for re-implementing our algorithm, as well as pseudocode in Algorithms 2–5. Source code for Literati can be found at `https://github.com/upadhyan/literati`.

**Architecture.** Literati is implemented in Rust with Python bindings. Each state $s \in \mathcal{S}$ is represented as a struct storing the index vector identifying samples in $\mathcal{D}_s$, the heuristic values $V(s)$, $\widetilde{V}(s)$, and $\overline{V}(s)$, a COMPLETE flag, and additional fields described in the paragraphs below. The AND dynamics induced by a commit action are encoded in a separate struct that tracks the joint solved status of $(s_L, s_R) = F(s_{\text{shape}}, a_{\text{commit}})$ and supports round-robin selection (Section 3.2.2).

**Efficient Data Partitioning.** Applying a shape function $f_j^{\mathbf{k}}$ requires partitioning $\mathcal{D}_s$ according to $\mathbf{k}$. To avoid repeatedly scanning $\mathcal{D}_s$, when we expand a data state $(\mathcal{D}_s, d)$ we precompute, for every feature $j$, the sample indices falling in each interval between consecutive candidates in $U_s^{(j)}$, and store these bin partitions on the data state's struct. Descendant shape function states sharing the same ancestor data state hold pointers to these bins rather than copies. A shape function on feature $j$ can then be represented compactly as a boolean array over bins indicating left or right assignment, and partitioning $\mathcal{D}_s$ reduces to a single linear scan of this array followed by concatenation of the corresponding index vectors.

**Policy Queues.** Each partial policy ($\pi$, $\widetilde{\pi}$, and $\overline{\pi}$) is represented as a min-queue over $\mathcal{A}(s)$ keyed on the corresponding $Q$-values, with the queues stored on the state's struct. The queues are re-sorted whenever a state is touched during backpropagation. For $\pi$ and $\overline{\pi}$, the prescribed action is simply the queue head. For $\widetilde{\pi}$, since $\widetilde{V}$ is potentially inadmissible, we instead traverse the queue until we find an action whose successors are not all COMPLETE (Section 3.2.1).

**Eager Initialization of Commit Successors.** The heuristic values $V$, $\widetilde{V}$, and $\overline{V}$ at a shape function state $s_{\text{shape}} = (\mathcal{D}_s, d, j, \mathbf{k})$ are functions of the two data states resulting from its commit action. We therefore eagerly construct and populate the structs for $(\mathcal{D}_s(f_j^{\mathbf{k}})_L, d-1)$ and $(\mathcal{D}_s(f_j^{\mathbf{k}})_R, d-1)$ (using the bin pointers inherited from the ancestor data state) when $s_{\text{shape}}$ is added to the graph, and retain them until $s_{\text{shape}}$ is itself expanded. This avoids recomputing partition statistics that are reused throughout backpropagation, yielding substantial runtime savings in practice.

**Lazy Backpropagation.** To avoid unnecessary updates and priority queue reorderings, backpropagation along an ancestor chain is terminated as soon as the quantities $V(s)$, $\widetilde{V}(s)$, $\overline{V}(s)$, and the COMPLETE flag are all unchanged after updating state $s$. Propagation continues to the parent of $s$ only if at least one of these quantities is modified.

**Regularization Reparameterization.** For analytical clarity, the objective in Equation 1 charges $\lambda + \alpha\,\mathbf{K}(f_v)$ at every internal node. However, to make the two regularization terms easier for users to interpret independently, our implementation instead exposes parameters $\lambda_r \geq 0$ and $\alpha_r \geq 0$ and charges $\lambda_r + \alpha_r\,(\mathbf{K}(f_v) - 1)$ per node, so that $\alpha_r$ penalizes only the structure beyond a standard threshold split. This is an affine reparameterization of the original objective, recovered exactly by setting $\lambda = \lambda_r - \alpha_r$ and $\alpha = \alpha_r$. In the AND/OR graph, the split transition cost decomposes as $c(s, a_{j,u}) = (\lambda_r - \alpha_r) + \alpha_r = \lambda_r$ and the refine cost remains $c(s, a_{j,\mathbf{k},u}) = \alpha_r$, both non-negative, so the correctness of the AND/OR formulation, the AO* search procedure, and all theoretical results are preserved.

**Caching** Unlike Branches [5], we do not cache states keyed on their reaching subset $\mathcal{D}_s$. Empirically, the memory cost of maintaining such a cache, together with the lookup overhead at every state construction, outweighs any savings from reuse. Two factors compound this: (i) our adaptive candidate generation produces per-node threshold sets $U_s^{(j)}$ that yield largely distinct partitions across the search tree, reducing cache hit rates, and (ii) caching invalidates the memory advantage Literati gains from pruning states, as we would still need to retain pruned states for future reuse. Instead, we maintain a limited cache of $V$, $\bar{V}$, and $\widetilde{V}$ values for the 10,000 most recently explored data states. This is substantially more memory efficient because it avoids retaining pointers to the underlying states, allowing Literati to free pruned states from memory while still benefiting from prior exploration through tighter bounds.

### B.1 Pseudocode

**Algorithm 2:** Literati: main loop

**Input:** data $\mathcal{D}$, depth budget $D$; complexity budget $K$; regularizers $\lambda, \alpha$; time limit $\tau$
**Output:** Shape Generalized Tree

$s_0 \leftarrow (\mathcal{D}, D)$ // root data state
`InitDataState`$(s_0)$
**while** $\neg\,\mathrm{COMPLETE}(s_0)\ \wedge\ \neg$ `TimedOut`$()$ **do**
  $s \leftarrow$ `Select`$(s_0)$
  `Expand`$(s)$
  `Backpropagate`$(s)$
**return** `ExtractTree`$(s_0)$ // unroll $\bar{\pi}$ from the root

**Algorithm 3:** Selection

**Function** `Select`$(s)$**:**
  **if** $\neg\,\mathrm{expanded}(s)$ **then return** $s$
  $a \leftarrow \tilde{\pi}(s)$
  **if** $|F(s,a)| > 1$ **then**
    **return** `Select`(`ANDSelection`$(F(s,a),\ s)$)
  **else**
    $\{s'\} \leftarrow F(s,a)$
    **return** `Select`$(s')$

**Function** `ANDSelection`$(\{s_L, s_R\},\ s)$**:**
  **if** $\mathrm{COMPLETE}(s_L)$ **then return** $s_R$
  **if** $\mathrm{COMPLETE}(s_R)$ **then return** $s_L$
  **if** $\mathrm{chooseLeft}(s)$ **then**
    $\mathrm{chooseLeft}(s) \leftarrow$ false
    **return** $s_L$
  **else**
    $\mathrm{chooseLeft}(s) \leftarrow$ true
    **return** $s_R$

**Algorithm 4:** Expansion

**Function** `Expand(`$s$`):`
- **if** $s$ *is Data State* **then** // data state
  - $\mathcal{D}_s, d \leftarrow s$
  - $t \leftarrow \overline{(\mathcal{D}_s, d)}$
  - $V(t) \leftarrow \mathrm{Err}(\mathcal{D}_s)$; $\quad \overline{V}(t) \leftarrow V(t)$; $\quad \widetilde{V}(t) \leftarrow V(t)$
  - $\mathrm{COMPLETE}(t) \leftarrow$ true
  - add $\bar{a}$ with cost 0 and $F(s, \bar{a}) = \{t\}$ to $\mathcal{A}(s)$
  - $U_s^{(1)}, \ldots, U_s^{(M)} \leftarrow$ `Discretize`$(\mathcal{D}_s, d)$
  - **foreach** $j = 1, \ldots, M$ **do**
    - `// add Initial Shape Function states`
    - **foreach** $u \in U_s^{(j)}$ **do**
      - $s' \leftarrow (\mathcal{D}_s, d, j, \{u\})$
      - `InitSFState`$(s')$
      - add $a_{j,u}$ with cost $\lambda + \alpha$, $F = \{s'\}$ to $\mathcal{A}(s)$
- **else** // shape function state
  - **foreach** $s' \in \{s_L, s_R\}$ **do**
    - `InitDataState`$(s')$
  - add $a_{\text{commit}}$ with cost 0, $F = \{s_L, s_R\}$ to $\mathcal{A}(s)$
  - **if** $|\mathbf{k}| < K$ **then**
    - $U_s^{(j)} \leftarrow$ `RetrieveChangepoints`$(\mathcal{D}_s, j)$
    - **foreach** $u \in U_s^{(j)}$ **with** $u > \max(\mathbf{k})$ **do**
      - $s' \leftarrow (\mathcal{D}_s, d, j, \mathbf{k} \cup \{u\})$
      - `InitSFState`$(s')$
      - add $a_{j,\mathbf{k},u}$ with cost $\alpha$, $F = \{s'\}$ to $\mathcal{A}(s)$
  - $\mathrm{expanded}(s) \leftarrow$ true

**Function** `InitSFState(`$s$`):`
- $(\mathcal{D}_s, d, j, \mathbf{k}) \leftarrow s$
- $(\mathcal{D}_L, \mathcal{D}_R) \leftarrow$ `Partition`$(\mathcal{D}_s, f_j^{\mathbf{k}})$
- $s_L \leftarrow (\mathcal{D}_L, d-1)$; $\quad s_R \leftarrow (\mathcal{D}_R, d-1)$
- `// Precompute left and right Data State statistics.`
- $V(s_L) \leftarrow \min(\mathrm{Err}(\mathcal{D}_L), \lambda + \alpha)$; $V(s_R) \leftarrow \min(\mathrm{Err}(\mathcal{D}_R), \lambda + \alpha)$
- $\overline{V}(s_L) \leftarrow \mathrm{Err}(\mathcal{D}_L)$; $\overline{V}(s_R) \leftarrow \mathrm{Err}(\mathcal{D}_R)$
- **if** $d > 1$ **then**
  - $T_L \leftarrow$ `FitCART`$(\mathcal{D}_L, d-1)$; $\quad T_R \leftarrow$ `FitCART`$(\mathcal{D}_R, d-1)$
  - $\widetilde{V}(s_L) \leftarrow \mathcal{L}(T_L, \mathcal{D}_L)$; $\quad \widetilde{V}(s_R) \leftarrow \mathcal{L}(T_R, \mathcal{D}_R)$
- **else**
  - $\widetilde{V}(s_L) \leftarrow \mathrm{Err}(\mathcal{D}_L)$; $\quad \widetilde{V}(s_R) \leftarrow \mathrm{Err}(\mathcal{D}_R)$
- `//` $\mathcal{L}(T, \mathcal{D})$ `is the objective value of Equation 1`
- **foreach** $s' \in \{s_L, s_R\}$ **do**
  - `InitDataState`$(s')$
- `// Assign Shape Function State statistics`
- $V(s) \leftarrow \min(\alpha, V(s_L) + V(s_R))$
- $\overline{V}(s) \leftarrow \overline{V}(s_L) + \overline{V}(s_R)$
- $\widetilde{V}(s) \leftarrow \widetilde{V}(s_L) + \widetilde{V}(s_R)$
- $\mathrm{COMPLETE}(s) \leftarrow$ FALSE

**Function** `InitDataState(`$s$`):`
- $(\mathcal{D}_s, d) \leftarrow s$
- $\mathrm{COMPLETE}(s) \leftarrow (d = 0) \vee (\mathrm{Err}(\mathcal{D}_s) \leq \lambda + \alpha)$
- **if** $\mathrm{COMPLETE}(s)$ **then**
  - $V(s) \leftarrow \mathrm{Err}(\mathcal{D}_s)$
  - $\overline{V}(s) \leftarrow V(s)$; $\quad \widetilde{V}(s) \leftarrow V(s)$
  - **return**
- **if** $V(s), \overline{V}(s), \widetilde{V}(s)$ do not exist **then**
  - $V(s) \leftarrow \min(\mathrm{Err}(\mathcal{D}_s), \lambda + \alpha)$
  - $\overline{V}(s) \leftarrow \mathrm{Err}(\mathcal{D}_s)$
  - $T \leftarrow$ `FitCART`$(\mathcal{D}_s, d)$
  - $\widetilde{V}(s) \leftarrow \mathcal{L}(T, \mathcal{D}_s)$

**Algorithm 5:** Backpropagation

**Function** `Backpropagate(`$s$`)`**:**

OLD $\leftarrow (V(s), \bar{V}(s), \tilde{V}(s),$ COMPLETE$(s))$

`// primal`

$\bar{\pi}(s) \leftarrow \arg\min_{a \in \mathcal{A}(s)} \bar{Q}(s,a)$; $\quad \bar{V}(s) \leftarrow \bar{Q}(s, \bar{\pi}(s))$

`// prune`

$\mathcal{A}(s) \leftarrow \{\, a \in \mathcal{A}(s) \,:\, Q(s,a) \leq \bar{V}(s) \,\}$

Retain the incumbent policy and its solution graph before freeing pruned subgraphs

Remove subgraphs of any pruned actions

`// Policy Updates`

$\pi(s) \leftarrow \arg\min_{a \in \mathcal{A}(s)} Q(s,a)$; $\quad V(s) \leftarrow Q(s, \pi(s))$

`// COMPLETE check`

**if** $\forall\, s' \in F(s, \pi(s)) \,:\,$ COMPLETE$(s')$ **then**

  COMPLETE$(s) \leftarrow$ TRUE

  $\mathcal{A}(s) \leftarrow \{\pi(s)\}$

  `// Search complete; Fix policy`

`// Select an action with unsolved successors`

**if** COMPLETE$(s)$ **then**

  $\tilde{\pi}(s) \leftarrow \pi(s)$; $\quad \tilde{V}(s) \leftarrow V(s)$

**else**

  $\mathcal{A}_{\text{open}}(s) \leftarrow \{a \in \mathcal{A}(s) : \exists s' \in F(s,a),\ \neg\text{COMPLETE}(s')\}$

  $\tilde{\pi}(s) \leftarrow \arg\min_{a \in \mathcal{A}_{\text{open}}(s)} \tilde{Q}(s,a)$

  $\tilde{V}(s) \leftarrow \tilde{Q}(s, \tilde{\pi}(s))$

`// Propagate Upward`

**if** $(V(s), \bar{V}(s), \tilde{V}(s),$ COMPLETE$(s)) \neq$ OLD $\wedge$ parent$(s)$ exists **then**

  `Backpropagate(`parent$(s)$`)`

**Algorithm 6:** Discretization with caching

**Function** `Discretize(`$\mathcal{D}, d$`)`**:**

**if** $(\mathcal{D}, d) \in$ cache **then return** cache$[(\mathcal{D}, d)]$

$T \leftarrow$ `FitCART`$(\mathcal{D},$ maxLeaves $= B_d + 1)$

$U_{\mathcal{D}}^{(j)} \leftarrow \emptyset \quad \forall\, j \in \{1, \ldots, M\}$

**foreach** *internal node* $v \in \mathcal{I}(T)$ **do**

  let $(j_v, u_v)$ be the feature and threshold at node $v$

  $U_{\mathcal{D}}^{(j_v)} \leftarrow U_{\mathcal{D}}^{(j_v)} \cup \{u_v\}$

cache$[(\mathcal{D}, d)] \leftarrow \{U_{\mathcal{D}}^{(j)}\}_{j=1}^{M}$

**return** $\{U_{\mathcal{D}}^{(j)}\}_{j=1}^{M}$

# C Additional Experimental Details

All experiments were run on Intel Granite Forest compute nodes (Xeon 6972P @ 2.4 GHz, 12 vCPUs, 64 GB RAM). For the Optimization Performance experiments (Section 4.1), all regularization is turned off for all approaches.

**Baseline Implementation Details:** For CART, we utilize the implementation found in Scikit-Learn [36]. For DPDT, we utilize the official implementation provided by the original authors [6] found at `https://github.com/KohlerHECTOR/DPDTreeEstimator`. For STreeD [9], we utilize the official implementation provided by the original authors found at `https://github.com/AlgTUDelft/pystreed`. For ConTree [22], we utilize the official implementation provided by the original authors found at `https://github.com/ConSol-Lab/contree/tree/main`. For Branches [5], we utilize the official implementation provided by the original authors found at `https://github.com/Chaoukia/branches`. NOTE: when faced with categorical variables, Branches supports multi-way branching, routing each category to a separate sub-tree. For fair comparison with binary trees, we turn this feature off in our experiments by setting the `encoding` parameter to `"multi"` or `"binary"`[2]. For ShapeCART and ShapeTAO [10], we utilize the official implementation provided by the original authors found at `https://github.com/optimal-uoft/Empowering-DTs-via-Shape-Functions`.

## C.1 Generalization Experiment Details

Below are the hyperparameters tuned for each evaluated approach. Depth is treated as a hyperparameter for all models and tuned over $D \in \{2, \dots, 6\}$ following Upadhya and Cohen [10].

- **Literati**: `impurity` $\in$ {gini, entropy}; `lambda_`, `alpha_` $\in \{1, 2, 4, 8, 16, 32, 64\}$ with `alpha_` $\leq$ `lambda_`; `max_order` $\in \{2, 3\}$;
  `cart_nodes_list` $\in \{(9, 9), (5, 5), (17, 17), (17, 9), (9, 5)\}$[3]. `lambda_`, `alpha_` divided by `n_samples`.
- **ShapeCART**: `criterion` $\in$ {gini, entropy};
  `min_samples_split` $\in \{2, 4, 8, 16, 32, 64\}$; `min_samples_leaf` $\in \{1, 2, 4, 8, 16, 32\}$;
  `min_impurity_decrease` $\in \{0, 10^{-4}, 5{\cdot}10^{-4}, 10^{-3}, 5{\cdot}10^{-3}, 10^{-2}\}$;
  `inner_max_leaf_nodes` $\in \{3, 4\}$. Note that this parameter controls the max shape function complexity (e.g. $K =$ `inner_max_leaf_nodes` $- 1$);
  `inner_min_samples_leaf` $\in \{1, 10^{-4}, 5{\cdot}10^{-4}, 10^{-3}, 5{\cdot}10^{-3}, 10^{-2}\}$.
- **ShapeTAO**: Same as ShapeCART **plus** `tao_reg` $\in \{1, 2, 4, 8, 16, 32, 64, 128\}$.
  `lambda_` $=$ `tao_reg` / `n_samples`.
- **CART**: `criterion` $\in$ {gini, entropy}; `min_samples_split` $= 2^k$, $k \in \{1, \dots, 5\}$;
  `min_samples_leaf` $\in \{1, \dots, 32\}$;
  `min_impurity_decrease` $\in \{0, 10^{-4}, 5{\cdot}10^{-4}, 10^{-3}, 5{\cdot}10^{-3}, 10^{-2}\}$;
  `ccp_alpha` $\in \{0, 10^{-4}, 5{\cdot}10^{-4}, 10^{-3}, 5{\cdot}10^{-3}, 10^{-2}\}$.
- **AxTAO**: `criterion` $\in$ {gini, entropy}; `min_samples_split` $= 2^k$, $k \in \{1, \dots, 5\}$;
  `min_samples_leaf` $\in \{1, \dots, 32\}$;
  `min_impurity_decrease` $\in \{0, 10^{-4}, 5{\cdot}10^{-4}, 10^{-3}, 5{\cdot}10^{-3}, 10^{-2}\}$;
  `lambda_` $\in \{0, 10^{-4}, 5{\cdot}10^{-4}, 10^{-3}, 5{\cdot}10^{-3}, 10^{-2}\}$. `lambda_` used as raw fraction (not scaled by n).
- **DPDT**: `criterion` $\in$ {gini, entropy}; `min_samples_split` $= 2^k$, $k \in \{1, \dots, 5\}$;
  `min_samples_leaf` $\in \{1, \dots, 32\}$;
  `min_impurity_decrease` $\in \{0, 10^{-4}, 5{\cdot}10^{-4}, 10^{-3}, 5{\cdot}10^{-3}, 10^{-2}\}$;
  `cart_nodes_list` $\in \{(9, 9), (5, 5), (17, 17), (17, 9), (9, 5)\}$.
- **Branches**: `lambda_` $\in \{1, 2, 4, 8, 16, 32, 64\}$; `n_quantiles` $\in [4, 16]$. QuantileBinarizer preprocessing; `lambda_` scaled by `n_samples`.

[2] See Line 27 of `https://github.com/Chaoukia/branches/blob/main/src/branches.py`

[3] $B_d =$ `cart_nodes_list`$[D - d] - 1$. For both Literati and DPDT, tuple entries specify CART leaf budgets from the root downward. If $D - d$ is out of index for the tuple, we set $B_d = K$ to allow for shape functions at all depths

- **ConTree**: `lambda_` $\in \{1, 2, 4, 8, 16, 32, 64\}$. `lambda_` mapped to `complexity_cost` $=$ `lambda_` $/$ `n_samples`.
- **STreeD**: `lambda_` $\in \{1, 2, 4, 8, 16, 32, 64\}$; `n_quantiles` $\in [4, 16]$. QuantileBinarizer preprocessing; `lambda_` mapped to `cost_complexity` $=$ `lambda_` $/$ `n_samples`.
- **LDS-DL8.5**: `lambda_` $\in \{1, 2, 4, 8, 16, 32, 64\}$; `n_quantiles` $\in [4, 16]$. QuantileBinarizer preprocessing; `lambda_` mapped to `min_support`
- **CADL8.5**: `lambda_` $\in \{1, 2, 4, 8, 16, 32, 64\}$; `n_quantiles` $\in [4, 16]$. QuantileBinarizer preprocessing; `lambda_` mapped to `min_support`

Table 3: Dataset statistics and source information.

| Dataset | $N$ | $M$ | $\lvert\mathcal{C}\rvert$ | Source | Source ID |
|---|---|---|---|---|---|
| adult | 45222 | 108 | 2 | OpenML | 179 |
| avila | 20867 | 10 | 12 | OpenML | 42932 |
| bank | 1372 | 4 | 2 | OpenML | 1462 |
| bean | 13611 | 16 | 7 | UCI | 602 |
| bidding | 6321 | 9 | 2 | OpenML | 42889 |
| electricity | 45312 | 13 | 2 | OpenML | 151 |
| eucalyptus | 736 | 1487 | 5 | OpenML | 43924 |
| eye-movements | 10936 | 27 | 3 | OpenML | 1044 |
| eye-state | 14980 | 14 | 2 | OpenML | 1471 |
| fault | 1941 | 27 | 7 | OpenML | 40982 |
| gas-drift | 13910 | 128 | 6 | OpenML | 1476 |
| htru | 17898 | 8 | 2 | OpenML | 45558 |
| magic | 13376 | 10 | 2 | OpenML | 44125 |
| mini-boone | 130064 | 50 | 2 | OpenML | 41150 |
| mushroom | 8124 | 94 | 2 | UCI | 73 |
| occupancy | 20560 | 5 | 2 | UCI | 357 |
| page | 5473 | 10 | 5 | OpenML | 30 |
| pendigits | 10992 | 16 | 10 | OpenML | 32 |
| raisin | 900 | 7 | 2 | UCI | 850 |
| rice | 3810 | 7 | 2 | UCI | 545 |
| room | 10129 | 16 | 4 | UCI | 864 |
| segment | 2310 | 18 | 7 | OpenML | 40984 |
| skin | 245057 | 3 | 2 | OpenML | 1502 |
| wilt | 4839 | 5 | 2 | OpenML | 40983 |

# D Selection Strategy Ablation Continued

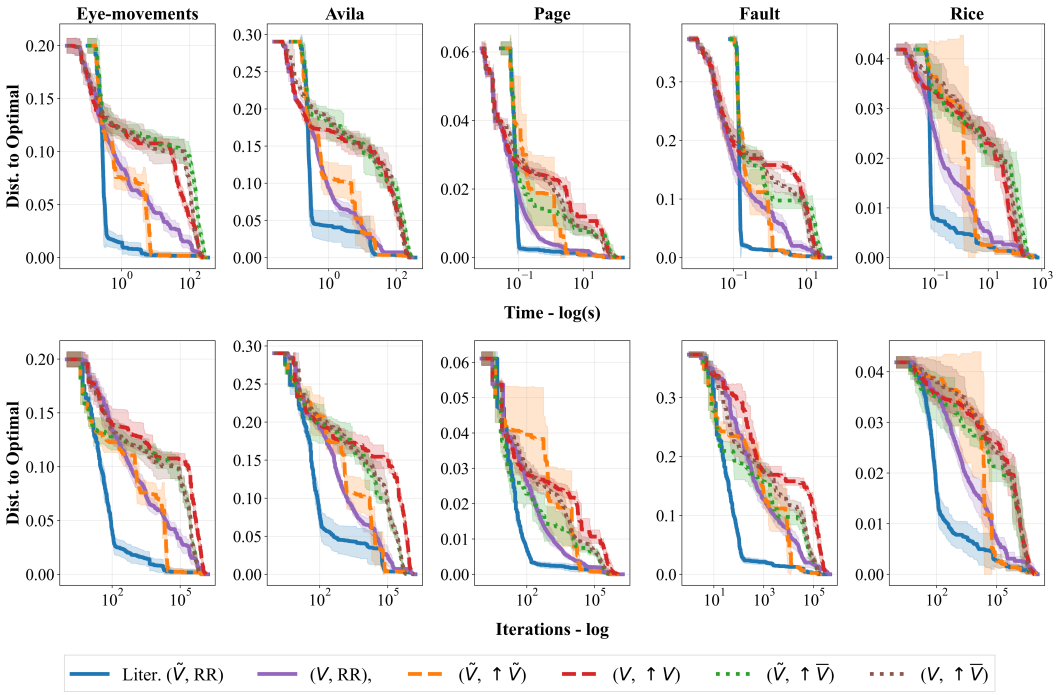



Figure D.5: Distance to Optimal Solution across time (top) and iterations (bottom) for Literati (Full) and its ablations on Eye-Movements, Avila, Page, Fault, and Rice.

Figure D.5 shows results for Eye-Movements, Avila, Page, Fault, and Rice for both time and iterations. We observe that, across five datasets, using an informative heuristic for OR selection and round-robin AND selection configuration consistently achieves the best anytime performance among all configurations when looking both at time and iterations.

# E Other Inadmissible Heuristics

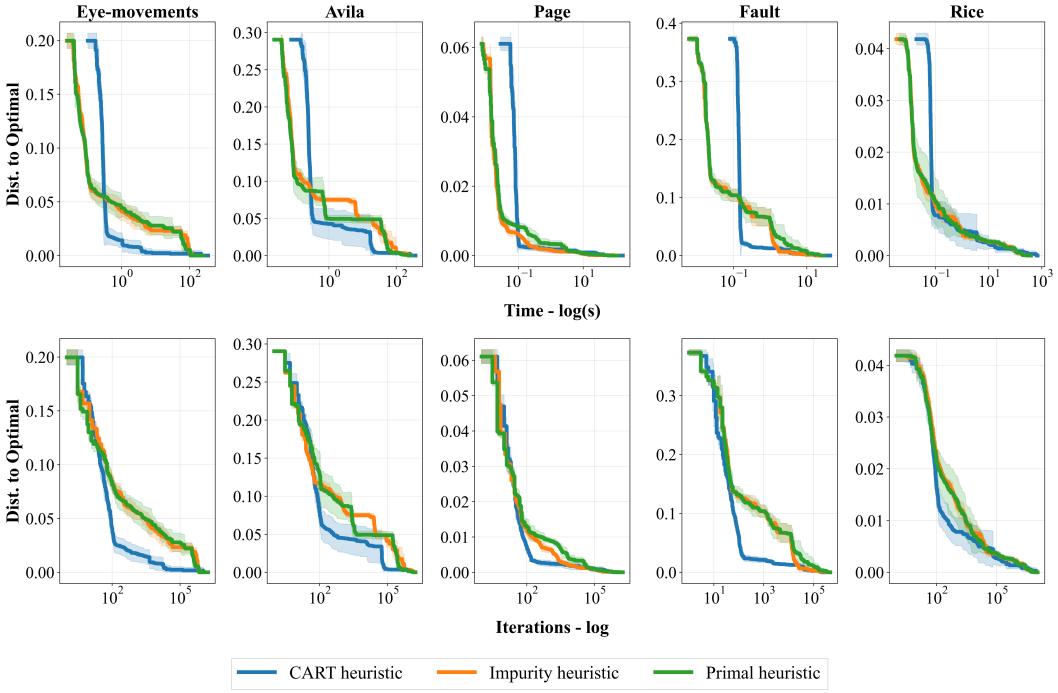



Figure E.6: Distance to Optimal Solution across time (top) and iterations (bottom) for Literati under different $\widetilde{V}$ candidates.

In this section, we evaluate two other candidates for inadmissible heuristics:

- **Primal:** Set $\widetilde{V} = \overline{V}$. This requires no additional computation, as the primal bound is already maintained during search.
- **Impurity:** Set $\widetilde{V}(\mathcal{D}_v, d) = I(\mathcal{D}_v) \cdot |\mathcal{D}_v|$, where $I(\cdot)$ denotes a weighted impurity measure (Gini or entropy). This is cheap to compute and better accounts for the class distribution.

We follow the experimental setting used in our selection strategy analysis (Section 4.1.1 and Appendix D), replacing only the inadmissible heuristic driving the search. Figure E.6 reports the distance-to-optimal against both wall-clock time and iteration count on Eye-Movements, Avila, Page, Fault, and Rice. Although the CART heuristic incurs minor additional overhead (in the order of $10^{-2}$–$10^{-1}$ seconds) during the initial iterations, it yields substantial runtime improvement on most datasets. This advantage is even more pronounced in the iteration view, where CART consistently approaches the optimum in fewer expansions than either the primal or impurity heuristics. The primal and impurity heuristics exhibit nearly similar behaviour across all five datasets, which we attribute to their shared myopic structure: both score a candidate split using only information local to the node. The CART heuristic instead performs an implicit lookahead by fitting a surrogate subtree, yielding tighter estimates.

# F Impact of Shapes

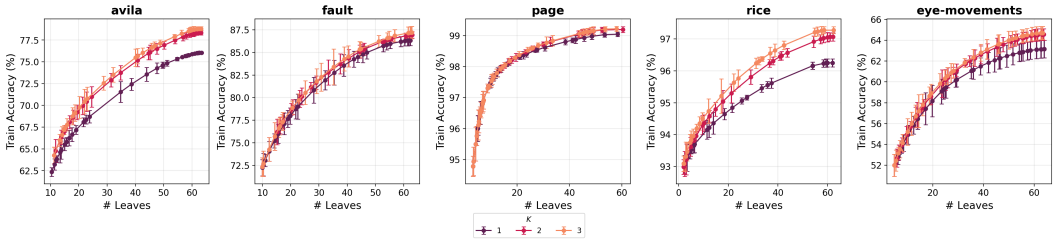


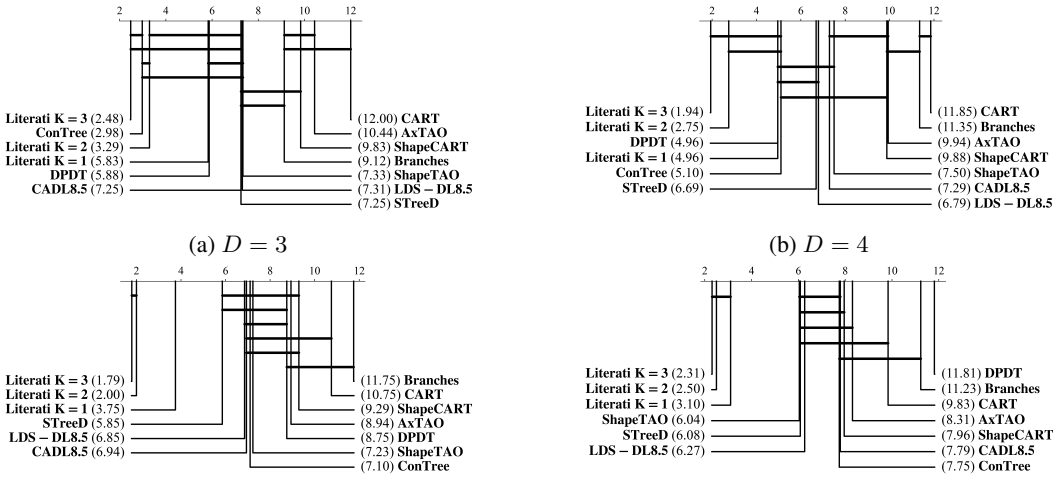

Figure F.7: Train Accuracy across Number of Leaves. The horizontal bars describe the standard deviation across folds.

Decision trees are prized for their inherent interpretability. However, this interpretability is highly sensitive to the structural complexity of the tree, particularly the depth of decision paths and the number of leaves [2, 37–39]. As an SGT, Literati leverages shape function branching, which yields higher-performing trees than axis-aligned threshold trees when controlling for the number of leaves. We empirically evaluate this benefit by comparing the train accuracy vs. model size tradeoff across shape complexity budgets.

**Setup.** We sweep 100 values of $\lambda$ log-uniformly spaced over $[10^{-8}, 1]$ and train Literati with a maximum depth of 6 using the light configuration from Appendix G.1, imposing a one-hour time limit per run. We compare $K \in \{1, 2, 3\}$; under this configuration, where $K = 1$ yields a threshold tree equivalent to DPDT [6] (Proposition 3). Figure F.7 reports training accuracy against the number of leaves on five representative datasets: *Avila*, *Fault*, *Page*, *Rice*, and *Eye-Movements*.

**Results.** Across all five datasets, increasing $K$ results in higher training accuracy consistently across tree size. The effect is most pronounced on *Avila* and *Rice*. On *Avila*, $K = 2$ matches the training accuracy that $K = 1$ reaches with 35 leaves using only 25; on *Rice*, $K = 2$ matches the 40-leaf $K = 1$ model with just 20 leaves. We further observe diminishing returns in $K$: the gap between $K = 2$ and $K = 3$ is consistently smaller than the gap between $K = 1$ and $K = 2$.

# G Optimality Experiment Results Continued




(a) $D = 3$ (b) $D = 4$

(c) $D = 5$ (d) $D = 6$

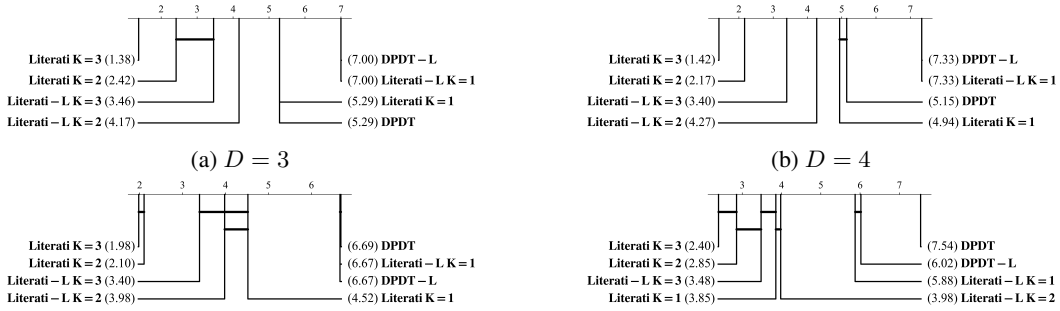
Figure G.8: Critical Difference Diagram of Train Accuracy for the Optimization capability experiments across datasets and folds

Table 4: Training accuracy, runtime, and optimality proof rate across depth budgets averaged across all datasets evaluated for DPDT and Literati, organized by Heavy and Light (-L) Discretization. The dashed lines separate Literati from baselines. Bold indicates the best value per depth for training accuracy and runtime.

| | | ↑ Train Acc. (%) | | | | ↓ Runtime (s) | | | | ↑ Proof Obtained (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $K$ | $D=3$ | $D=4$ | $D=5$ | $D=6$ | $D=3$ | $D=4$ | $D=5$ | $D=6$ | $D=3$ | $D=4$ | $D=5$ | $D=6$ |
| **Heavy Discretization** ($B_d = 16\ \forall d$) | | | | | | | | | | | | | |
| DPDT | 1 | 83.56 | 87.03 | 77.16 | 71.52 | 46.57 | 488.23 | 2387.41 | 3217.40 | 100.0 | 95.8 | 45.8 | 12.5 |
| Literati | 1 | 83.56 | 87.28 | 89.91 | 92.28 | 18.22 | 259.14 | 1156.29 | 2587.19 | 100.0 | 97.5 | 87.5 | 36.7 |
| | 2 | 84.03 | 87.75 | 90.36 | 92.58 | 60.82 | 851.76 | 2672.71 | 2817.87 | 100.0 | 87.5 | 33.3 | 22.5 |
| | 3 | **84.13** | **87.90** | **90.46** | **92.69** | 162.93 | 1393.71 | 2832.73 | 2782.09 | 100.0 | 82.5 | 25.0 | 24.2 |
| **Light Discretization** ($B_D = 16, B_{D-1} = 8, B_d = 4\ \forall\, 1 \leq d < D-1$) | | | | | | | | | | | | | |
| DPDT-L | 1 | 83.48 | 87.11 | 89.64 | 91.5 | 15.09 | 65.21 | 260.43 | 624.28 | 100.0 | 100.0 | 100.0 | 95.8 |
| Literati-L | 1 | 83.48 | 87.11 | 89.64 | 91.83 | **7.68** | **33.00** | **131.70** | **404.74** | 100.0 | 100.0 | 100.0 | 95.8 |
| | 2 | 83.85 | 87.46 | 89.99 | 92.26 | 17.00 | 87.67 | 377.58 | 981.87 | 100.0 | 100.0 | 95.8 | 84.2 |
| | 3 | 83.91 | 87.53 | 90.05 | 92.33 | 24.66 | 135.06 | 564.28 | 1124.52 | 100.0 | 100.0 | 92.5 | 82.5 |

Figure G.8 contains critical difference diagrams for each depth on train accuracy. We observe that above $D = 3$, Literati across values of $K$ achieves statistically significantly higher train accuracy over all other compared approaches, with Literati $K = 2$ and $K = 3$ achieving statistically significantly higher train accuracy for $D \in \{4, 5\}$ over $K = 1$.

### G.1 Light and Depth-Adaptive Configurations




(a) $D = 3$ (b) $D = 4$

(c) $D = 5$ (d) $D = 6$

Figure G.9: Critical Difference Diagram of Train Accuracy for the Optimization capability experiments across datasets and folds between the heavy and light (-L) variants of DPDT and Literati.

In this section, we evaluate a light discretization scheme for DPDT and Literati. More formally, we set $B_D = 16$, $B_{D-1} = 8$, and $B_d = 4\ \forall\, 1 \leq d < D-1$ for both DPDT and Literati and compare it to the heavier discretization scheme ($B_d = 16\ \forall d$) used in Section 4.1. Table 4 contains training accuracy, runtime, and proof rate for both the heavy and light configurations while Figure G.9 contains critical difference diagrams for each depth on training accuracy. Per dataset results can be found in Appendix G.3.

We observe that switching from a heavy to a light discretization configuration reduces runtime by orders of magnitude, though this, as expected, comes at the cost of training accuracy. Furthermore, we observe that we obtain proofs on a majority of datasets across all depths. We observe that under the light configuration, Literati ($K = 1$) and DPDT ($K = 1$) achieve identical training accuracy across all depths except $D = 6$, where neither converges on Mini-Boone. However, because of its anytime nature, Literati can return an intermediate solution, whereas DPDT cannot. When looking at runtime, we observe that the Literati-L $K = 1$ achieves lower average runtime than DPDT-L, demonstrating the quality of our search strategy.

## G.2 Literati Pre-Discretized

In addition to adaptive discretization, Literati can operate on pre-discretized candidate sets. In this section, we evaluate Literati using quantile-based pre-discretized features (Literati-PreQuant) and compare its performance with STreeD. We focus on the $K = 1$ setting, for which Literati and STreeD operate over equivalent search spaces. To ensure a fair comparison, we also equip Literati-PreQuant with the D2-solver subroutine used by STreeD [9, 20].

Table 5: Comparison of STreeD and Literati under quantile pre-discretization (Literati-PreQuant), averaged over all datasets at each depth budget $D$. Bold marks the best value at each depth between the two pre-discretized methods. Literati with its default adaptive discretization is shown below the solid line for reference.

| | | ↑ Train Acc. (%) | | | | ↓ Runtime (s) | | | | ↑ Proof Obtained (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $K$ | $D=3$ | $D=4$ | $D=5$ | $D=6$ | $D=3$ | $D=4$ | $D=5$ | $D=6$ | $D=3$ | $D=4$ | $D=5$ | $D=6$ |
| STreeD | 1 | **83.50** | 86.57 | 88.44 | 88.26 | 169.90 | 413.52 | 1693.70 | **2445.12** | 99.2 | **91.7** | 65.0 | **37.5** |
| Literati-PreQuant | 1 | **83.50** | **87.03** | **89.50** | **91.74** | **58.93** | **350.05** | **1623.85** | 2488.06 | **100.0** | **91.7** | **70.8** | 35.8 |
| Literati | 1 | 83.56 | 87.28 | 89.91 | 92.28 | 18.22 | 259.14 | 1156.29 | 2587.19 | 100.0 | 97.5 | 87.5 | 36.7 |

We observe that Literati-PreQuant consistently achieves higher training accuracy across all depths except for $D = 3$ (where it ties STreeD), highlighting the strong anytime performance of our approach. Furthermore, Literati-PreQuant maintains competitive proof rates across all depths, obtaining higher proof rates than STreeD at $D = 3$ and $D = 5$, whereas STreeD obtains a higher proof rate at $D = 6$.

Comparing Literati-PreQuant with Literati using adaptive discretization, we observe that using adaptive discretization results in higher training accuracy at all depths and lower runtime at all depths except $D = 6$.

## G.3 Per Dataset Results

Table 6: Training accuracy, runtime, and optimality proof rate across depth budgets across all datasets evaluated for Depths 3 and 4. The three blocks for each dataset report the normal, light (-L), and Literati-PreQuant configurations, respectively. The dashed lines separate Literati from baselines. Bold indicates the best value per dataset and depth within the normal and light blocks for training accuracy and runtime.

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D=3$ | $D=4$ | $D=3$ | $D=4$ | $D=3$ | $D=4$ |
| adult | CART | 1 | $82.32 \pm 0.16$ | $82.81 \pm 0.16$ | $\mathbf{0.09 \pm 0.01}$ | $\mathbf{0.10 \pm 0.02}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $82.34 \pm 0.18$ | $83.31 \pm 0.14$ | $6.24 \pm 2.59$ | $15.88 \pm 4.16$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $82.32 \pm 0.16$ | $82.81 \pm 0.16$ | $2.11 \pm 0.39$ | $2.99 \pm 1.06$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $82.32 \pm 0.16$ | $83.31 \pm 0.14$ | $8.51 \pm 11.19$ | $28.80 \pm 26.94$ | 0.0 | 0.0 |
| | Branches | 1 | $83.25 \pm 0.13$ | $82.88 \pm 0.33$ | $50.45 \pm 6.22$ | $68.76 \pm 6.36$ | 100.0 | 0.0 |
| | STreeD | 1 | $83.25 \pm 0.13$ | $83.84 \pm 0.12$ | $0.37 \pm 0.05$ | $9.18 \pm 0.57$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $83.25 \pm 0.13$ | $83.84 \pm 0.12$ | $38.67 \pm 0.53$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $83.25 \pm 0.13$ | $83.84 \pm 0.12$ | $1.35 \pm 0.01$ | $82.72 \pm 2.82$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{83.27 \pm 0.14}$ | $\mathbf{83.85 \pm 0.14}$ | $181.77 \pm 105.87$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $83.25 \pm 0.19$ | $83.83 \pm 0.16$ | $34.54 \pm 2.01$ | $581.03 \pm 28.52$ | 100.0 | 100.0 |
| | Literati | 1 | $83.25 \pm 0.19$ | $83.83 \pm 0.16$ | $28.04 \pm 2.52$ | $408.74 \pm 61.86$ | 100.0 | 100.0 |
| | | 2 | $83.25 \pm 0.19$ | $83.83 \pm 0.16$ | $49.04 \pm 9.17$ | $623.54 \pm 156.00$ | 100.0 | 100.0 |
| | | 3 | $83.25 \pm 0.19$ | $83.83 \pm 0.16$ | $56.84 \pm 10.28$ | $831.52 \pm 213.82$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $\mathbf{83.18 \pm 0.15}$ | $83.69 \pm 0.15$ | $\mathbf{10.26 \pm 0.55}$ | $46.55 \pm 3.61$ | 100.0 | 100.0 |
| | Literati-L | 1 | $\mathbf{83.18 \pm 0.15}$ | $83.69 \pm 0.15$ | $14.39 \pm 1.72$ | $\mathbf{42.38 \pm 10.08}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{83.18 \pm 0.15}$ | $\mathbf{83.70 \pm 0.15}$ | $20.58 \pm 3.08$ | $62.92 \pm 12.68$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{83.18 \pm 0.15}$ | $\mathbf{83.70 \pm 0.15}$ | $24.39 \pm 7.23$ | $83.15 \pm 24.75$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $83.25 \pm 0.13$ | $83.84 \pm 0.12$ | $0.27 \pm 0.00$ | $14.08 \pm 0.46$ | 100.0 | 100.0 |

**Table 6 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ |
| avila | CART | 1 | $52.80 \pm 0.40$ | $55.43 \pm 0.98$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.03 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $53.70 \pm 0.16$ | $56.71 \pm 1.08$ | $0.37 \pm 0.08$ | $0.82 \pm 0.19$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $55.00 \pm 0.45$ | $57.54 \pm 0.46$ | $0.74 \pm 0.36$ | $0.97 \pm 0.48$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $55.71 \pm 0.13$ | $59.29 \pm 0.93$ | $1.82 \pm 0.34$ | $3.49 \pm 0.48$ | 0.0 | 0.0 |
| | Branches | 1 | $56.97 \pm 0.19$ | $57.16 \pm 0.26$ | $139.81 \pm 14.55$ | $120.49 \pm 14.37$ | 100.0 | 0.0 |
| | STreeD | 1 | $56.97 \pm 0.19$ | $62.30 \pm 0.28$ | $5.00 \pm 0.22$ | $307.22 \pm 9.96$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $56.97 \pm 0.19$ | $62.30 \pm 0.28$ | $243.71 \pm 19.46$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $56.97 \pm 0.19$ | $62.30 \pm 0.28$ | $11.54 \pm 0.37$ | $1204.33 \pm 95.23$ | 100.0 | 100.0 |
| | ConTree | 1 | $58.45 \pm 0.10$ | $63.82 \pm 0.22$ | $19.95 \pm 1.47$ | $2940.90 \pm 156.55$ | 100.0 | 100.0 |
| | DPDT | 1 | $58.23 \pm 0.13$ | $63.33 \pm 0.40$ | $29.40 \pm 0.44$ | $508.28 \pm 14.41$ | 100.0 | 100.0 |
| | Literati | 1 | $58.23 \pm 0.13$ | $63.33 \pm 0.40$ | $6.14 \pm 0.23$ | $93.26 \pm 2.13$ | 100.0 | 100.0 |
| | | 2 | $58.99 \pm 0.14$ | $64.64 \pm 0.18$ | $32.45 \pm 3.48$ | $1096.71 \pm 80.51$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{59.19 \pm 0.22}$ | $\mathbf{65.10 \pm 0.27}$ | $55.39 \pm 8.16$ | $2747.93 \pm 386.57$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $57.84 \pm 0.33$ | $62.93 \pm 0.26$ | $6.59 \pm 0.12$ | $30.56 \pm 1.25$ | 100.0 | 100.0 |
| | Literati-L | 1 | $57.84 \pm 0.33$ | $62.93 \pm 0.26$ | $\mathbf{2.20 \pm 0.09}$ | $\mathbf{9.02 \pm 0.35}$ | 100.0 | 100.0 |
| | | 2 | $58.78 \pm 0.18$ | $64.01 \pm 0.38$ | $6.13 \pm 0.69$ | $27.67 \pm 1.93$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{58.82 \pm 0.19}$ | $\mathbf{64.03 \pm 0.40}$ | $7.45 \pm 0.90$ | $34.34 \pm 4.40$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $56.97 \pm 0.19$ | $62.30 \pm 0.28$ | $0.65 \pm 0.04$ | $57.43 \pm 1.72$ | 100.0 | 100.0 |
| bank | CART | 1 | $93.89 \pm 1.23$ | $96.33 \pm 1.94$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $95.41 \pm 0.34$ | $98.08 \pm 1.03$ | $0.28 \pm 0.08$ | $0.49 \pm 0.08$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $94.66 \pm 0.86$ | $97.41 \pm 1.23$ | $0.47 \pm 0.23$ | $0.71 \pm 0.34$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $97.00 \pm 1.09$ | $99.46 \pm 0.19$ | $2.08 \pm 0.65$ | $7.14 \pm 5.71$ | 0.0 | 0.0 |
| | Branches | 1 | $97.66 \pm 0.17$ | $99.65 \pm 0.06$ | $2.51 \pm 0.13$ | $56.99 \pm 6.05$ | 100.0 | 100.0 |
| | STreeD | 1 | $97.66 \pm 0.17$ | $99.65 \pm 0.06$ | $0.03 \pm 0.03$ | $0.12 \pm 0.03$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $97.66 \pm 0.17$ | $99.65 \pm 0.06$ | $0.98 \pm 0.08$ | $29.88 \pm 1.24$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $97.66 \pm 0.17$ | $99.65 \pm 0.06$ | $0.04 \pm 0.00$ | $1.08 \pm 0.05$ | 100.0 | 100.0 |
| | ConTree | 1 | $98.35 \pm 0.09$ | $\mathbf{100.00 \pm 0.00}$ | $0.02 \pm 0.00$ | $0.01 \pm 0.00$ | 100.0 | 100.0 |
| | DPDT | 1 | $97.94 \pm 0.29$ | $99.83 \pm 0.09$ | $0.91 \pm 0.10$ | $8.72 \pm 1.35$ | 100.0 | 100.0 |
| | Literati | 1 | $97.94 \pm 0.29$ | $99.81 \pm 0.09$ | $0.14 \pm 0.01$ | $1.06 \pm 0.20$ | 100.0 | 100.0 |
| | | 2 | $98.58 \pm 0.24$ | $99.98 \pm 0.05$ | $1.24 \pm 0.19$ | $12.73 \pm 11.97$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.67 \pm 0.23}$ | $\mathbf{100.00 \pm 0.00}$ | $3.82 \pm 0.89$ | $49.10 \pm 72.70$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $97.94 \pm 0.29$ | $99.75 \pm 0.12$ | $0.41 \pm 0.01$ | $1.78 \pm 0.07$ | 100.0 | 100.0 |
| | Literati-L | 1 | $97.94 \pm 0.29$ | $99.75 \pm 0.12$ | $\mathbf{0.07 \pm 0.00}$ | $\mathbf{0.24 \pm 0.01}$ | 100.0 | 100.0 |
| | | 2 | $98.58 \pm 0.24$ | $99.94 \pm 0.09$ | $0.37 \pm 0.03$ | $0.89 \pm 0.52$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.60 \pm 0.22}$ | $\mathbf{99.98 \pm 0.05}$ | $0.73 \pm 0.09$ | $1.25 \pm 1.19$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $97.66 \pm 0.17$ | $99.65 \pm 0.06$ | $0.01 \pm 0.00$ | $0.14 \pm 0.00$ | 100.0 | 100.0 |
| bean | CART | 1 | $77.84 \pm 0.19$ | $82.79 \pm 0.19$ | $\mathbf{0.04 \pm 0.00}$ | $\mathbf{0.05 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $78.78 \pm 0.16$ | $83.78 \pm 0.18$ | $1.10 \pm 0.05$ | $2.20 \pm 0.38$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $77.84 \pm 0.19$ | $82.62 \pm 0.18$ | $1.03 \pm 0.32$ | $0.99 \pm 0.30$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $78.79 \pm 0.17$ | $84.46 \pm 1.89$ | $2.42 \pm 0.31$ | $6.16 \pm 2.46$ | 0.0 | 0.0 |
| | Branches | 1 | $81.18 \pm 0.77$ | $83.98 \pm 0.21$ | $221.53 \pm 13.49$ | $206.76 \pm 4.31$ | 0.0 | 0.0 |
| | STreeD | 1 | $85.79 \pm 0.15$ | $90.05 \pm 0.10$ | $2.05 \pm 0.07$ | $194.48 \pm 2.72$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $85.79 \pm 0.15$ | $90.04 \pm 0.12$ | $463.23 \pm 33.11$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $85.79 \pm 0.15$ | $90.05 \pm 0.10$ | $18.81 \pm 0.95$ | $3067.72 \pm 283.09$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{87.26 \pm 0.12}$ | $90.89 \pm 0.12$ | $23.65 \pm 1.62$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $85.82 \pm 0.27$ | $90.86 \pm 0.20$ | $25.53 \pm 0.22$ | $429.52 \pm 6.90$ | 100.0 | 100.0 |
| | Literati | 1 | $85.82 \pm 0.27$ | $90.87 \pm 0.20$ | $8.91 \pm 0.22$ | $130.95 \pm 3.72$ | 100.0 | 100.0 |
| | | 2 | $86.31 \pm 0.18$ | $91.14 \pm 0.34$ | $39.93 \pm 4.79$ | $997.31 \pm 43.47$ | 100.0 | 100.0 |
| | | 3 | $86.31 \pm 0.18$ | $\mathbf{91.15 \pm 0.34}$ | $60.69 \pm 5.92$ | $1884.36 \pm 235.77$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $85.81 \pm 0.28$ | $90.82 \pm 0.20$ | $7.66 \pm 0.40$ | $33.43 \pm 0.53$ | 100.0 | 100.0 |
| | Literati-L | 1 | $85.81 \pm 0.28$ | $90.83 \pm 0.21$ | $\mathbf{3.47 \pm 0.04}$ | $\mathbf{14.59 \pm 0.92}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{86.06 \pm 0.42}$ | $\mathbf{91.00 \pm 0.26}$ | $9.92 \pm 0.45$ | $44.87 \pm 2.76$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{86.06 \pm 0.42}$ | $\mathbf{91.00 \pm 0.26}$ | $13.01 \pm 1.19$ | $58.62 \pm 5.03$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $85.79 \pm 0.15$ | $90.05 \pm 0.10$ | $0.95 \pm 0.10$ | $94.18 \pm 21.24$ | 100.0 | 100.0 |
| bidding | CART | 1 | $98.21 \pm 0.09$ | $98.61 \pm 0.10$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $98.21 \pm 0.09$ | $98.62 \pm 0.10$ | $0.09 \pm 0.01$ | $0.19 \pm 0.06$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $98.24 \pm 0.10$ | $99.55 \pm 0.06$ | $0.47 \pm 0.27$ | $1.27 \pm 1.05$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $98.24 \pm 0.10$ | $99.55 \pm 0.06$ | $1.21 \pm 0.32$ | $1.28 \pm 0.30$ | 0.0 | 0.0 |
| | Branches | 1 | $99.22 \pm 0.05$ | $99.42 \pm 0.04$ | $11.15 \pm 0.81$ | $184.72 \pm 17.58$ | 100.0 | 0.0 |
| | STreeD | 1 | $99.22 \pm 0.05$ | $99.67 \pm 0.01$ | $0.09 \pm 0.03$ | $1.68 \pm 0.07$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $99.22 \pm 0.05$ | $99.67 \pm 0.01$ | $3.96 \pm 0.04$ | $349.83 \pm 17.08$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $99.22 \pm 0.05$ | $99.67 \pm 0.01$ | $0.33 \pm 0.01$ | $13.53 \pm 0.51$ | 100.0 | 100.0 |
| | ConTree | 1 | $99.22 \pm 0.05$ | $99.66 \pm 0.07$ | $0.73 \pm 0.16$ | $5.22 \pm 1.55$ | 100.0 | 100.0 |
| | DPDT | 1 | $99.22 \pm 0.05$ | $99.68 \pm 0.02$ | $1.98 \pm 0.25$ | $18.93 \pm 3.17$ | 100.0 | 100.0 |
| | Literati | 1 | $99.22 \pm 0.05$ | $99.68 \pm 0.02$ | $0.46 \pm 0.08$ | $2.64 \pm 0.36$ | 100.0 | 100.0 |
| | | 2 | $99.42 \pm 0.05$ | $99.89 \pm 0.04$ | $2.89 \pm 0.41$ | $27.13 \pm 11.41$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.55 \pm 0.03}$ | $\mathbf{99.95 \pm 0.02}$ | $5.19 \pm 1.52$ | $59.58 \pm 22.86$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $99.22 \pm 0.05$ | $99.67 \pm 0.03$ | $0.86 \pm 0.04$ | $3.40 \pm 0.25$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.22 \pm 0.05$ | $99.67 \pm 0.03$ | $\mathbf{0.26 \pm 0.02}$ | $\mathbf{0.80 \pm 0.14}$ | 100.0 | 100.0 |
| | | 2 | $99.42 \pm 0.06$ | $99.88 \pm 0.04$ | $1.08 \pm 0.18$ | $3.21 \pm 0.51$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.55 \pm 0.03}$ | $\mathbf{99.95 \pm 0.02}$ | $1.80 \pm 0.30$ | $5.30 \pm 1.22$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $99.22 \pm 0.05$ | $99.67 \pm 0.01$ | $0.05 \pm 0.01$ | $1.82 \pm 0.11$ | 100.0 | 100.0 |

**Table 6 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ |
| electricity | CART | 1 | $76.05 \pm 0.17$ | $76.29 \pm 0.10$ | $\mathbf{0.04 \pm 0.00}$ | $\mathbf{0.05 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $76.18 \pm 0.25$ | $77.07 \pm 0.89$ | $1.31 \pm 0.74$ | $4.87 \pm 4.62$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $77.70 \pm 1.66$ | $79.22 \pm 0.91$ | $1.27 \pm 0.73$ | $1.96 \pm 1.25$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $77.78 \pm 1.75$ | $79.94 \pm 0.39$ | $3.57 \pm 0.69$ | $6.24 \pm 0.80$ | 0.0 | 0.0 |
| | Branches | 1 | $79.06 \pm 0.09$ | $78.11 \pm 0.12$ | $40.08 \pm 3.00$ | $73.23 \pm 10.42$ | 100.0 | 0.0 |
| | STreeD | 1 | $79.06 \pm 0.09$ | $80.06 \pm 0.07$ | $0.87 \pm 0.08$ | $34.74 \pm 5.86$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $79.06 \pm 0.09$ | $79.92 \pm 0.12$ | $54.80 \pm 0.99$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $79.06 \pm 0.09$ | $80.06 \pm 0.07$ | $2.85 \pm 0.17$ | $186.96 \pm 13.91$ | 100.0 | 100.0 |
| | ConTree | 1 | $79.65 \pm 0.04$ | $80.84 \pm 0.07$ | $106.50 \pm 4.25$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $79.48 \pm 0.05$ | $81.00 \pm 0.13$ | $33.96 \pm 1.05$ | $609.02 \pm 24.94$ | 100.0 | 100.0 |
| | Literati | 1 | $79.48 \pm 0.05$ | $81.00 \pm 0.13$ | $11.15 \pm 0.55$ | $167.38 \pm 7.98$ | 100.0 | 100.0 |
| | | 2 | $80.32 \pm 0.17$ | $81.83 \pm 0.13$ | $181.25 \pm 18.23$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $\mathbf{80.57 \pm 0.19}$ | $\mathbf{82.06 \pm 0.18}$ | $1135.26 \pm 180.24$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT-L | 1 | $79.44 \pm 0.06$ | $80.77 \pm 0.17$ | $9.57 \pm 0.60$ | $42.52 \pm 1.57$ | 100.0 | 100.0 |
| | Literati-L | 1 | $79.44 \pm 0.06$ | $80.77 \pm 0.17$ | $\mathbf{4.38 \pm 0.11}$ | $\mathbf{17.93 \pm 0.46}$ | 100.0 | 100.0 |
| | | 2 | $79.92 \pm 0.16$ | $81.48 \pm 0.17$ | $30.92 \pm 2.68$ | $198.13 \pm 11.92$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{80.04 \pm 0.19}$ | $\mathbf{81.65 \pm 0.15}$ | $91.82 \pm 10.63$ | $617.17 \pm 68.52$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $79.06 \pm 0.09$ | $80.06 \pm 0.07$ | $0.62 \pm 0.06$ | $26.12 \pm 2.86$ | 100.0 | 100.0 |
| eucalyptus | CART | 1 | $48.60 \pm 0.84$ | $51.87 \pm 0.65$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.02 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $48.60 \pm 0.84$ | $51.91 \pm 0.57$ | $5.69 \pm 0.37$ | $7.87 \pm 2.89$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $48.60 \pm 0.84$ | $51.87 \pm 0.65$ | $2.29 \pm 0.32$ | $3.55 \pm 0.08$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $48.60 \pm 0.84$ | $51.91 \pm 0.57$ | $4.69 \pm 0.56$ | $8.00 \pm 1.10$ | 0.0 | 0.0 |
| | Branches | 1 | $49.30 \pm 0.85$ | $52.22 \pm 1.36$ | $164.44 \pm 9.60$ | $172.80 \pm 24.99$ | 0.0 | 0.0 |
| | STreeD | 1 | $\mathbf{51.67 \pm 1.19}$ | $\mathbf{58.91 \pm 1.15}$ | $1.67 \pm 0.13$ | $61.27 \pm 6.25$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $\mathbf{51.67 \pm 1.19}$ | $\mathbf{58.91 \pm 1.15}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $\mathbf{51.67 \pm 1.19}$ | $58.56 \pm 1.35$ | $254.35 \pm 40.65$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $\mathbf{51.67 \pm 1.19}$ | $57.04 \pm 1.10$ | $1621.80 \pm 58.82$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $50.16 \pm 0.90$ | $57.74 \pm 1.19$ | $8.24 \pm 0.75$ | $135.45 \pm 10.09$ | 100.0 | 100.0 |
| | Literati | 1 | $50.16 \pm 0.90$ | $57.78 \pm 1.17$ | $6.99 \pm 0.46$ | $61.36 \pm 10.87$ | 100.0 | 100.0 |
| | | 2 | $50.35 \pm 0.84$ | $58.02 \pm 1.15$ | $7.84 \pm 0.86$ | $70.56 \pm 9.57$ | 100.0 | 100.0 |
| | | 3 | $50.35 \pm 0.84$ | $58.02 \pm 1.15$ | $7.83 \pm 0.59$ | $68.30 \pm 11.41$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $50.16 \pm 0.90$ | $57.59 \pm 1.17$ | $\mathbf{2.20 \pm 0.22}$ | $11.02 \pm 3.22$ | 100.0 | 100.0 |
| | Literati-L | 1 | $50.16 \pm 0.90$ | $57.63 \pm 1.15$ | $3.15 \pm 0.29$ | $\mathbf{8.93 \pm 0.51}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{50.35 \pm 0.84}$ | $\mathbf{57.78 \pm 1.17}$ | $3.37 \pm 0.21$ | $10.01 \pm 1.28$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{50.35 \pm 0.84}$ | $\mathbf{57.78 \pm 1.17}$ | $3.23 \pm 0.25$ | $9.71 \pm 1.03$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $51.67 \pm 1.19$ | $58.91 \pm 1.15$ | $0.56 \pm 0.06$ | $45.78 \pm 4.52$ | 100.0 | 100.0 |
| eye-movements | CART | 1 | $48.45 \pm 2.15$ | $51.75 \pm 0.47$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.03 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $51.41 \pm 0.34$ | $53.31 \pm 0.42$ | $3.96 \pm 0.90$ | $4.84 \pm 2.06$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $50.92 \pm 0.43$ | $52.72 \pm 0.34$ | $0.85 \pm 0.09$ | $0.93 \pm 0.12$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $52.42 \pm 0.54$ | $53.64 \pm 0.25$ | $4.51 \pm 0.61$ | $5.65 \pm 1.93$ | 0.0 | 0.0 |
| | Branches | 1 | $49.25 \pm 0.21$ | $45.10 \pm 0.24$ | $175.39 \pm 13.10$ | $135.56 \pm 14.18$ | 0.0 | 0.0 |
| | STreeD | 1 | $52.23 \pm 0.34$ | $55.52 \pm 0.25$ | $3.79 \pm 0.16$ | $484.98 \pm 16.52$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $52.23 \pm 0.34$ | $55.28 \pm 0.12$ | $737.06 \pm 35.46$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $52.23 \pm 0.34$ | $55.36 \pm 0.13$ | $22.80 \pm 0.56$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $52.60 \pm 0.27$ | $55.28 \pm 0.79$ | $195.72 \pm 27.05$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $52.29 \pm 0.16$ | $55.54 \pm 0.21$ | $18.31 \pm 1.45$ | $322.09 \pm 11.12$ | 100.0 | 100.0 |
| | Literati | 1 | $52.29 \pm 0.16$ | $55.49 \pm 0.11$ | $5.45 \pm 0.73$ | $82.20 \pm 7.91$ | 100.0 | 100.0 |
| | | 2 | $52.68 \pm 0.30$ | $56.51 \pm 0.25$ | $17.92 \pm 3.43$ | $446.34 \pm 83.24$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{52.82 \pm 0.48}$ | $\mathbf{57.04 \pm 0.18}$ | $24.10 \pm 6.23$ | $741.29 \pm 188.44$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $52.21 \pm 0.22$ | $55.12 \pm 0.39$ | $5.10 \pm 0.02$ | $25.11 \pm 0.46$ | 100.0 | 100.0 |
| | Literati-L | 1 | $52.21 \pm 0.22$ | $55.03 \pm 0.29$ | $\mathbf{2.11 \pm 0.29}$ | $\mathbf{8.47 \pm 0.64}$ | 100.0 | 100.0 |
| | | 2 | $52.62 \pm 0.28$ | $55.80 \pm 0.32$ | $4.57 \pm 0.92$ | $18.74 \pm 3.54$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{52.63 \pm 0.29}$ | $\mathbf{55.96 \pm 0.45}$ | $5.32 \pm 1.43$ | $21.06 \pm 5.10$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $52.23 \pm 0.34$ | $55.52 \pm 0.25$ | $2.04 \pm 0.22$ | $310.78 \pm 30.30$ | 100.0 | 100.0 |
| eye-state | CART | 1 | $67.37 \pm 1.21$ | $71.25 \pm 0.66$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.03 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $68.56 \pm 0.73$ | $72.38 \pm 0.54$ | $1.80 \pm 0.86$ | $2.80 \pm 0.56$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $70.86 \pm 0.58$ | $73.55 \pm 0.59$ | $1.31 \pm 0.91$ | $1.44 \pm 0.92$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $\mathbf{71.59 \pm 0.59}$ | $74.36 \pm 0.56$ | $3.06 \pm 0.82$ | $5.84 \pm 2.35$ | 0.0 | 0.0 |
| | Branches | 1 | $68.74 \pm 0.39$ | $60.62 \pm 0.21$ | $113.45 \pm 5.07$ | $106.93 \pm 4.73$ | 0.0 | 0.0 |
| | STreeD | 1 | $70.40 \pm 0.16$ | $74.40 \pm 0.31$ | $2.30 \pm 0.19$ | $214.82 \pm 27.26$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $70.40 \pm 0.16$ | $74.01 \pm 0.19$ | $415.63 \pm 38.01$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $70.40 \pm 0.16$ | $74.40 \pm 0.31$ | $12.77 \pm 0.83$ | $2056.99 \pm 83.49$ | 100.0 | 100.0 |
| | ConTree | 1 | $70.90 \pm 0.20$ | $75.08 \pm 0.38$ | $64.30 \pm 3.97$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $70.40 \pm 0.48$ | $74.79 \pm 0.37$ | $15.11 \pm 0.65$ | $287.16 \pm 12.56$ | 100.0 | 100.0 |
| | Literati | 1 | $70.42 \pm 0.49$ | $74.79 \pm 0.37$ | $5.40 \pm 0.12$ | $87.80 \pm 2.27$ | 100.0 | 100.0 |
| | | 2 | $71.58 \pm 0.27$ | $75.74 \pm 0.34$ | $21.69 \pm 2.37$ | $594.56 \pm 67.56$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{71.59 \pm 0.26}$ | $\mathbf{75.90 \pm 0.47}$ | $36.69 \pm 8.62$ | $1103.18 \pm 237.60$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $70.31 \pm 0.58$ | $74.56 \pm 0.41$ | $4.19 \pm 0.24$ | $20.87 \pm 0.96$ | 100.0 | 100.0 |
| | Literati-L | 1 | $70.31 \pm 0.58$ | $74.56 \pm 0.41$ | $\mathbf{1.84 \pm 0.06}$ | $\mathbf{8.37 \pm 0.13}$ | 100.0 | 100.0 |
| | | 2 | $70.94 \pm 0.51$ | $\mathbf{75.11 \pm 0.36}$ | $4.78 \pm 0.44$ | $22.55 \pm 1.97$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{71.11 \pm 0.42}$ | $\mathbf{75.11 \pm 0.36}$ | $6.46 \pm 1.01$ | $31.38 \pm 4.38$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $70.40 \pm 0.16$ | $74.40 \pm 0.31$ | $1.63 \pm 0.58$ | $137.57 \pm 29.05$ | 100.0 | 100.0 |

**Table 6 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D=3$ | $D=4$ | $D=3$ | $D=4$ | $D=3$ | $D=4$ |
| fault | CART | 1 | $55.41 \pm 0.49$ | $62.42 \pm 0.92$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $56.82 \pm 0.56$ | $64.27 \pm 1.40$ | $0.73 \pm 0.27$ | $0.98 \pm 0.35$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $58.54 \pm 1.48$ | $65.18 \pm 1.72$ | $0.58 \pm 0.11$ | $0.71 \pm 0.26$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $62.92 \pm 1.12$ | $68.03 \pm 1.58$ | $4.53 \pm 3.92$ | $6.52 \pm 5.12$ | 0.0 | 0.0 |
| | Branches | 1 | $61.19 \pm 0.71$ | $51.74 \pm 0.70$ | $207.75 \pm 15.75$ | $214.70 \pm 26.02$ | 0.0 | 0.0 |
| | STreeD | 1 | $67.28 \pm 0.48$ | $73.99 \pm 0.19$ | $5.28 \pm 0.15$ | $853.06 \pm 39.97$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $67.28 \pm 0.48$ | $73.99 \pm 0.19$ | $1340.98 \pm 77.95$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $67.28 \pm 0.48$ | $73.99 \pm 0.19$ | $18.86 \pm 1.00$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $\mathbf{67.75 \pm 0.50}$ | $70.13 \pm 3.36$ | $24.21 \pm 2.50$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $67.17 \pm 0.61$ | $74.34 \pm 0.73$ | $5.43 \pm 0.39$ | $99.43 \pm 3.74$ | 100.0 | 100.0 |
| | Literati | 1 | $67.17 \pm 0.61$ | $74.33 \pm 0.74$ | $1.65 \pm 0.09$ | $23.13 \pm 1.89$ | 100.0 | 100.0 |
| | | 2 | $67.57 \pm 0.48$ | $75.61 \pm 0.86$ | $3.22 \pm 0.23$ | $64.83 \pm 5.12$ | 100.0 | 100.0 |
| | | 3 | $67.72 \pm 0.64$ | $\mathbf{75.68 \pm 0.88}$ | $3.55 \pm 0.54$ | $78.26 \pm 12.97$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $67.03 \pm 0.89$ | $73.74 \pm 0.93$ | $1.52 \pm 0.08$ | $7.46 \pm 0.15$ | 100.0 | 100.0 |
| | Literati-L | 1 | $67.03 \pm 0.89$ | $73.74 \pm 0.93$ | $\mathbf{0.62 \pm 0.04}$ | $\mathbf{2.36 \pm 0.10}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{67.28 \pm 0.76}$ | $74.33 \pm 1.20$ | $0.94 \pm 0.11$ | $3.81 \pm 0.48$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{67.28 \pm 0.76}$ | $\mathbf{74.37 \pm 1.12}$ | $1.02 \pm 0.10$ | $4.01 \pm 0.62$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $67.28 \pm 0.48$ | $73.99 \pm 0.19$ | $1.21 \pm 0.21$ | $220.48 \pm 29.40$ | 100.0 | 100.0 |
| gas-drift | CART | 1 | $66.33 \pm 0.43$ | $73.39 \pm 0.49$ | $\mathbf{0.31 \pm 0.00}$ | $\mathbf{0.40 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $67.51 \pm 0.28$ | $79.72 \pm 2.49$ | $23.10 \pm 3.42$ | $43.28 \pm 8.19$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $66.85 \pm 0.45$ | $74.20 \pm 0.39$ | $4.08 \pm 1.26$ | $4.70 \pm 0.63$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $69.84 \pm 1.08$ | $78.95 \pm 1.24$ | $9.60 \pm 2.10$ | $17.66 \pm 2.30$ | 0.0 | 0.0 |
| | Branches | 1 | $42.91 \pm 0.11$ | $42.88 \pm 0.12$ | $170.19 \pm 39.91$ | $155.10 \pm 25.60$ | 0.0 | 0.0 |
| | STreeD | 1 | $73.60 \pm 0.27$ | $73.60 \pm 0.27$ | $3274.76 \pm 184.10$ | $3600.00 \pm 0.00$ | 80.0 | 0.0 |
| | LDS-DL8.5 | 1 | $73.60 \pm 0.27$ | $83.81 \pm 0.28$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $73.60 \pm 0.27$ | $69.66 \pm 0.10$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $\mathbf{75.19 \pm 0.16}$ | $\mathbf{84.61 \pm 0.23}$ | $3324.21 \pm 143.92$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $70.67 \pm 0.41$ | $83.66 \pm 0.53$ | $140.61 \pm 0.86$ | $2266.18 \pm 157.51$ | 100.0 | 100.0 |
| | Literati | 1 | $70.67 \pm 0.41$ | $83.66 \pm 0.53$ | $71.94 \pm 1.04$ | $1092.49 \pm 20.12$ | 100.0 | 100.0 |
| | | 2 | $70.67 \pm 0.41$ | $83.70 \pm 0.57$ | $90.07 \pm 6.37$ | $1504.90 \pm 61.89$ | 100.0 | 100.0 |
| | | 3 | $70.67 \pm 0.41$ | $83.70 \pm 0.57$ | $89.62 \pm 3.11$ | $1530.03 \pm 58.41$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $\mathbf{70.66 \pm 0.40}$ | $83.14 \pm 0.47$ | $50.50 \pm 0.61$ | $210.94 \pm 1.24$ | 100.0 | 100.0 |
| | Literati-L | 1 | $\mathbf{70.66 \pm 0.40}$ | $\mathbf{83.15 \pm 0.46}$ | $\mathbf{26.91 \pm 0.54}$ | $\mathbf{113.33 \pm 2.62}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{70.66 \pm 0.40}$ | $\mathbf{83.15 \pm 0.46}$ | $31.28 \pm 1.00$ | $130.68 \pm 3.97$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{70.66 \pm 0.40}$ | $\mathbf{83.15 \pm 0.46}$ | $30.92 \pm 0.93$ | $130.94 \pm 2.98$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $73.60 \pm 0.27$ | $84.54 \pm 0.15$ | $1038.14 \pm 52.79$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| htru | CART | 1 | $97.96 \pm 0.04$ | $98.11 \pm 0.03$ | $\mathbf{0.03 \pm 0.00}$ | $\mathbf{0.04 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $98.01 \pm 0.06$ | $98.14 \pm 0.02$ | $0.54 \pm 0.22$ | $0.47 \pm 0.09$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $98.10 \pm 0.04$ | $98.29 \pm 0.03$ | $1.29 \pm 0.15$ | $1.44 \pm 0.73$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $98.15 \pm 0.03$ | $98.34 \pm 0.04$ | $2.01 \pm 0.50$ | $16.87 \pm 14.98$ | 0.0 | 0.0 |
| | Branches | 1 | $97.73 \pm 0.05$ | $97.62 \pm 0.04$ | $45.50 \pm 3.23$ | $120.46 \pm 19.08$ | 100.0 | 0.0 |
| | STreeD | 1 | $97.73 \pm 0.05$ | $97.89 \pm 0.04$ | $0.46 \pm 0.04$ | $18.49 \pm 0.78$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $97.73 \pm 0.05$ | $97.89 \pm 0.04$ | $49.05 \pm 4.38$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $97.73 \pm 0.05$ | $97.89 \pm 0.04$ | $2.00 \pm 0.10$ | $146.00 \pm 2.34$ | 100.0 | 100.0 |
| | ConTree | 1 | $98.12 \pm 0.04$ | $98.33 \pm 0.03$ | $75.33 \pm 17.25$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $98.09 \pm 0.04$ | $98.30 \pm 0.05$ | $14.79 \pm 0.41$ | $275.32 \pm 12.95$ | 100.0 | 100.0 |
| | Literati | 1 | $98.09 \pm 0.04$ | $98.30 \pm 0.05$ | $4.43 \pm 0.26$ | $67.89 \pm 3.96$ | 100.0 | 100.0 |
| | | 2 | $98.15 \pm 0.04$ | $98.40 \pm 0.04$ | $28.89 \pm 4.45$ | $1045.71 \pm 156.16$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.19 \pm 0.06}$ | $\mathbf{98.46 \pm 0.04}$ | $73.58 \pm 21.87$ | $3346.46 \pm 322.06$ | 100.0 | 60.0 |
| | DPDT-L | 1 | $98.09 \pm 0.04$ | $98.28 \pm 0.05$ | $4.65 \pm 0.07$ | $22.55 \pm 0.82$ | 100.0 | 100.0 |
| | Literati-L | 1 | $98.08 \pm 0.05$ | $98.28 \pm 0.05$ | $\mathbf{1.73 \pm 0.04}$ | $\mathbf{7.22 \pm 0.41}$ | 100.0 | 100.0 |
| | | 2 | $98.12 \pm 0.06$ | $98.34 \pm 0.04$ | $6.79 \pm 0.85$ | $31.90 \pm 3.61$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.14 \pm 0.04}$ | $\mathbf{98.36 \pm 0.04}$ | $10.57 \pm 2.55$ | $54.50 \pm 12.67$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $97.73 \pm 0.05$ | $97.89 \pm 0.04$ | $0.31 \pm 0.04$ | $15.20 \pm 0.81$ | 100.0 | 100.0 |
| magic | CART | 1 | $77.76 \pm 0.49$ | $79.30 \pm 0.43$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.03 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $78.15 \pm 0.35$ | $80.16 \pm 0.45$ | $0.82 \pm 0.24$ | $1.81 \pm 0.50$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $78.15 \pm 0.22$ | $80.23 \pm 0.45$ | $1.22 \pm 0.71$ | $1.09 \pm 0.92$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $78.49 \pm 0.28$ | $81.01 \pm 0.47$ | $2.27 \pm 1.08$ | $4.65 \pm 0.90$ | 0.0 | 0.0 |
| | Branches | 1 | $79.84 \pm 0.43$ | $77.50 \pm 0.42$ | $111.41 \pm 8.96$ | $119.51 \pm 9.36$ | 100.0 | 0.0 |
| | STreeD | 1 | $79.84 \pm 0.43$ | $81.69 \pm 0.39$ | $0.90 \pm 0.33$ | $44.43 \pm 1.53$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $79.84 \pm 0.43$ | $81.69 \pm 0.39$ | $98.00 \pm 9.54$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $79.84 \pm 0.43$ | $81.69 \pm 0.39$ | $3.68 \pm 0.23$ | $414.25 \pm 15.68$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{80.35 \pm 0.34}$ | $82.59 \pm 0.34$ | $32.01 \pm 3.93$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $80.08 \pm 0.33$ | $82.17 \pm 0.36$ | $16.51 \pm 0.12$ | $299.17 \pm 6.33$ | 100.0 | 100.0 |
| | Literati | 1 | $80.07 \pm 0.32$ | $82.17 \pm 0.36$ | $4.93 \pm 0.09$ | $76.79 \pm 2.40$ | 100.0 | 100.0 |
| | | 2 | $80.33 \pm 0.46$ | $82.57 \pm 0.27$ | $24.47 \pm 0.76$ | $786.11 \pm 54.23$ | 100.0 | 100.0 |
| | | 3 | $80.33 \pm 0.46$ | $\mathbf{82.60 \pm 0.29}$ | $41.83 \pm 4.36$ | $1727.63 \pm 211.33$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $79.60 \pm 0.31$ | $82.09 \pm 0.38$ | $4.87 \pm 0.03$ | $23.26 \pm 0.67$ | 100.0 | 100.0 |
| | Literati-L | 1 | $79.60 \pm 0.31$ | $82.09 \pm 0.38$ | $\mathbf{1.79 \pm 0.03}$ | $\mathbf{7.95 \pm 0.16}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{80.08 \pm 0.53}$ | $\mathbf{82.18 \pm 0.33}$ | $5.30 \pm 0.15$ | $26.42 \pm 0.52$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{80.08 \pm 0.53}$ | $\mathbf{82.18 \pm 0.34}$ | $6.90 \pm 0.23$ | $33.51 \pm 1.28$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $79.84 \pm 0.43$ | $81.69 \pm 0.39$ | $0.62 \pm 0.08$ | $34.62 \pm 1.75$ | 100.0 | 100.0 |

**Table 6 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ |
| mini-boone | CART | 1 | $86.75 \pm 0.10$ | $87.94 \pm 0.11$ | $\mathbf{1.30 \pm 0.01}$ | $\mathbf{1.70 \pm 0.03}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $87.25 \pm 0.09$ | $88.51 \pm 0.21$ | $77.32 \pm 52.94$ | $92.00 \pm 16.49$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $87.00 \pm 0.23$ | $88.50 \pm 0.08$ | $16.64 \pm 11.14$ | $12.83 \pm 5.12$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $87.80 \pm 0.05$ | $88.77 \pm 0.03$ | $15.22 \pm 2.87$ | $19.30 \pm 4.26$ | 0.0 | 0.0 |
| | Branches | 1 | $83.63 \pm 0.07$ | $83.56 \pm 0.14$ | $61.27 \pm 6.32$ | $63.24 \pm 4.39$ | 0.0 | 0.0 |
| | STreeD | 1 | $88.36 \pm 0.02$ | $89.14 \pm 0.04$ | $770.61 \pm 20.44$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | LDS-DL8.5 | 1 | $88.35 \pm 0.02$ | $88.87 \pm 0.04$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $88.36 \pm 0.02$ | $83.27 \pm 0.06$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $87.66 \pm 0.05$ | $83.67 \pm 0.20$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $88.29 \pm 0.10$ | $83.28 \pm 0.08$ | $631.02 \pm 39.27$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $88.29 \pm 0.10$ | $89.49 \pm 0.02$ | $253.02 \pm 15.49$ | $3564.98 \pm 51.70$ | 100.0 | 40.0 |
| | | 2 | $\mathbf{88.45 \pm 0.12}$ | $\mathbf{89.61 \pm 0.10}$ | $646.62 \pm 41.13$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $\mathbf{88.45 \pm 0.12}$ | $\mathbf{89.61 \pm 0.10}$ | $899.90 \pm 79.36$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT-L | 1 | $88.27 \pm 0.09$ | $89.42 \pm 0.03$ | $218.99 \pm 10.07$ | $927.56 \pm 33.39$ | 100.0 | 100.0 |
| | Literati-L | 1 | $88.27 \pm 0.09$ | $89.42 \pm 0.03$ | $\mathbf{91.88 \pm 3.86}$ | $\mathbf{412.30 \pm 30.85}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{88.37 \pm 0.11}$ | $\mathbf{89.48 \pm 0.09}$ | $205.09 \pm 20.65$ | $949.58 \pm 74.73$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{88.37 \pm 0.11}$ | $\mathbf{89.48 \pm 0.09}$ | $242.83 \pm 40.51$ | $1201.26 \pm 146.04$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $88.36 \pm 0.02$ | $89.35 \pm 0.09$ | $363.70 \pm 124.40$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| mushroom | CART | 1 | $98.50 \pm 0.09$ | $99.40 \pm 0.05$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $98.53 \pm 0.06$ | $99.40 \pm 0.05$ | $1.70 \pm 0.63$ | $1.31 \pm 0.08$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $98.50 \pm 0.09$ | $99.40 \pm 0.05$ | $0.30 \pm 0.03$ | $0.40 \pm 0.01$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $98.53 \pm 0.06$ | $99.40 \pm 0.05$ | $1.23 \pm 0.28$ | $1.31 \pm 0.11$ | 0.0 | 0.0 |
| | Branches | 1 | $\mathbf{99.13 \pm 0.03}$ | $100.00 \pm 0.00$ | $13.38 \pm 0.38$ | $84.33 \pm 110.55$ | 100.0 | 60.0 |
| | STreeD | 1 | $\mathbf{99.13 \pm 0.03}$ | $\mathbf{100.00 \pm 0.00}$ | $0.08 \pm 0.05$ | $0.05 \pm 0.00$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $\mathbf{99.13 \pm 0.03}$ | $\mathbf{100.00 \pm 0.00}$ | $4.35 \pm 0.06$ | $0.06 \pm 0.00$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $\mathbf{99.13 \pm 0.03}$ | $\mathbf{100.00 \pm 0.00}$ | $0.23 \pm 0.00$ | $0.02 \pm 0.00$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{99.13 \pm 0.03}$ | $\mathbf{100.00 \pm 0.00}$ | $10.05 \pm 2.15$ | $0.07 \pm 0.01$ | 100.0 | 100.0 |
| | DPDT | 1 | $99.01 \pm 0.11$ | $\mathbf{100.00 \pm 0.00}$ | $1.76 \pm 0.16$ | $12.37 \pm 1.41$ | 100.0 | 100.0 |
| | Literati | 1 | $99.09 \pm 0.03$ | $\mathbf{100.00 \pm 0.00}$ | $1.67 \pm 0.18$ | $0.15 \pm 0.01$ | 100.0 | 100.0 |
| | | 2 | $99.09 \pm 0.03$ | $\mathbf{100.00 \pm 0.00}$ | $1.68 \pm 0.14$ | $0.17 \pm 0.04$ | 100.0 | 100.0 |
| | | 3 | $99.09 \pm 0.03$ | $\mathbf{100.00 \pm 0.00}$ | $1.70 \pm 0.17$ | $0.14 \pm 0.01$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $99.01 \pm 0.11$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{1.03 \pm 0.08}$ | $3.83 \pm 0.23$ | 100.0 | 100.0 |
| | Literati-L | 1 | $\mathbf{99.09 \pm 0.03}$ | $\mathbf{100.00 \pm 0.00}$ | $1.34 \pm 0.03$ | $0.21 \pm 0.08$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{99.09 \pm 0.03}$ | $\mathbf{100.00 \pm 0.00}$ | $1.35 \pm 0.14$ | $0.17 \pm 0.03$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.09 \pm 0.03}$ | $\mathbf{100.00 \pm 0.00}$ | $1.33 \pm 0.10$ | $\mathbf{0.16 \pm 0.02}$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $99.13 \pm 0.03$ | $100.00 \pm 0.00$ | $0.04 \pm 0.00$ | $0.20 \pm 0.04$ | 100.0 | 100.0 |
| occupancy | CART | 1 | $99.00 \pm 0.04$ | $99.03 \pm 0.04$ | $\mathbf{0.02 \pm 0.01}$ | $\mathbf{0.02 \pm 0.01}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.01 \pm 0.04$ | $99.04 \pm 0.04$ | $0.52 \pm 0.28$ | $0.60 \pm 0.20$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $99.08 \pm 0.10$ | $99.28 \pm 0.03$ | $0.90 \pm 0.26$ | $0.96 \pm 0.66$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.09 \pm 0.12$ | $99.31 \pm 0.04$ | $2.27 \pm 1.20$ | $2.82 \pm 1.00$ | 0.0 | 0.0 |
| | Branches | 1 | $98.81 \pm 0.03$ | $98.86 \pm 0.04$ | $8.27 \pm 1.10$ | $112.32 \pm 8.44$ | 100.0 | 0.0 |
| | STreeD | 1 | $98.81 \pm 0.03$ | $98.97 \pm 0.02$ | $0.17 \pm 0.01$ | $3.36 \pm 0.05$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $98.81 \pm 0.03$ | $98.97 \pm 0.02$ | $6.05 \pm 0.43$ | $474.01 \pm 43.74$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $98.81 \pm 0.03$ | $98.97 \pm 0.02$ | $0.40 \pm 0.02$ | $14.36 \pm 0.29$ | 100.0 | 100.0 |
| | ConTree | 1 | $99.12 \pm 0.04$ | $99.19 \pm 0.32$ | $11.05 \pm 3.68$ | $323.50 \pm 35.47$ | 100.0 | 100.0 |
| | DPDT | 1 | $99.09 \pm 0.04$ | $99.31 \pm 0.02$ | $10.07 \pm 0.20$ | $168.64 \pm 4.78$ | 100.0 | 100.0 |
| | Literati | 1 | $99.09 \pm 0.04$ | $99.31 \pm 0.02$ | $2.61 \pm 0.07$ | $34.19 \pm 2.33$ | 100.0 | 100.0 |
| | | 2 | $99.24 \pm 0.03$ | $99.42 \pm 0.01$ | $27.00 \pm 4.82$ | $1072.56 \pm 122.05$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.27 \pm 0.03}$ | $\mathbf{99.48 \pm 0.02}$ | $79.63 \pm 18.00$ | $3578.83 \pm 47.34$ | 100.0 | 20.0 |
| | DPDT-L | 1 | $99.09 \pm 0.04$ | $99.30 \pm 0.01$ | $2.86 \pm 0.24$ | $13.54 \pm 1.61$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.09 \pm 0.04$ | $99.30 \pm 0.02$ | $\mathbf{0.98 \pm 0.06}$ | $\mathbf{3.97 \pm 0.18}$ | 100.0 | 100.0 |
| | | 2 | $99.22 \pm 0.04$ | $99.38 \pm 0.02$ | $4.43 \pm 0.60$ | $22.99 \pm 4.02$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.24 \pm 0.05}$ | $\mathbf{99.40 \pm 0.02}$ | $7.06 \pm 1.43$ | $37.20 \pm 6.77$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $98.81 \pm 0.03$ | $98.97 \pm 0.02$ | $0.12 \pm 0.01$ | $2.27 \pm 0.09$ | 100.0 | 100.0 |
| page | CART | 1 | $96.14 \pm 0.42$ | $96.90 \pm 0.26$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $96.26 \pm 0.34$ | $96.93 \pm 0.24$ | $0.38 \pm 0.09$ | $0.62 \pm 0.35$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $95.95 \pm 0.19$ | $97.09 \pm 0.30$ | $1.58 \pm 2.41$ | $1.81 \pm 1.33$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $96.32 \pm 0.37$ | $97.39 \pm 0.27$ | $2.30 \pm 0.93$ | $4.18 \pm 0.58$ | 0.0 | 0.0 |
| | Branches | 1 | $96.08 \pm 0.12$ | $95.35 \pm 0.17$ | $97.28 \pm 14.01$ | $206.99 \pm 9.52$ | 100.0 | 0.0 |
| | STreeD | 1 | $96.08 \pm 0.12$ | $97.30 \pm 0.13$ | $0.63 \pm 0.02$ | $36.19 \pm 1.25$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $96.08 \pm 0.12$ | $97.30 \pm 0.13$ | $56.48 \pm 3.35$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $96.08 \pm 0.12$ | $97.30 \pm 0.13$ | $2.03 \pm 0.05$ | $203.10 \pm 8.56$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{96.97 \pm 0.16}$ | $97.07 \pm 1.79$ | $2.41 \pm 0.27$ | $230.90 \pm 23.38$ | 100.0 | 100.0 |
| | DPDT | 1 | $96.85 \pm 0.14$ | $97.87 \pm 0.08$ | $6.04 \pm 0.07$ | $107.10 \pm 1.85$ | 100.0 | 100.0 |
| | Literati | 1 | $96.85 \pm 0.14$ | $97.87 \pm 0.07$ | $1.70 \pm 0.06$ | $23.56 \pm 2.26$ | 100.0 | 100.0 |
| | | 2 | $96.94 \pm 0.12$ | $97.97 \pm 0.11$ | $9.03 \pm 1.44$ | $256.02 \pm 45.39$ | 100.0 | 100.0 |
| | | 3 | $96.94 \pm 0.12$ | $\mathbf{98.02 \pm 0.12}$ | $15.44 \pm 4.96$ | $546.27 \pm 180.00$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $96.84 \pm 0.15$ | $97.84 \pm 0.11$ | $1.71 \pm 0.05$ | $8.55 \pm 0.16$ | 100.0 | 100.0 |
| | Literati-L | 1 | $96.83 \pm 0.15$ | $97.84 \pm 0.11$ | $\mathbf{0.65 \pm 0.07}$ | $\mathbf{2.66 \pm 0.22}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{96.91 \pm 0.13}$ | $\mathbf{97.95 \pm 0.10}$ | $1.91 \pm 0.34$ | $8.41 \pm 1.60$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{96.91 \pm 0.13}$ | $\mathbf{97.95 \pm 0.10}$ | $2.77 \pm 1.21$ | $11.47 \pm 4.78$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $96.08 \pm 0.12$ | $97.30 \pm 0.13$ | $0.19 \pm 0.03$ | $13.02 \pm 1.18$ | 100.0 | 100.0 |

**Table 6 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ |
| pendigits | CART | 1 | $57.37 \pm 0.37$ | $73.04 \pm 0.41$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.02 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $58.78 \pm 0.61$ | $79.70 \pm 0.17$ | $0.91 \pm 0.42$ | $2.04 \pm 0.29$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $57.29 \pm 0.32$ | $72.73 \pm 0.22$ | $1.88 \pm 0.40$ | $2.26 \pm 1.15$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $58.88 \pm 0.66$ | $79.70 \pm 0.16$ | $2.52 \pm 0.86$ | $5.71 \pm 2.28$ | 0.0 | 0.0 |
| | Branches | 1 | $63.85 \pm 0.19$ | $38.89 \pm 0.41$ | $205.99 \pm 37.07$ | $148.12 \pm 4.84$ | 0.0 | 0.0 |
| | STreeD | 1 | $66.62 \pm 0.19$ | $84.51 \pm 0.36$ | $5.91 \pm 0.24$ | $366.36 \pm 23.96$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $66.62 \pm 0.19$ | $84.51 \pm 0.36$ | $376.35 \pm 21.50$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $66.62 \pm 0.19$ | $84.51 \pm 0.36$ | $15.26 \pm 0.69$ | $2076.27 \pm 136.63$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{66.71 \pm 0.28}$ | $\mathbf{85.10 \pm 0.31}$ | $7.84 \pm 0.52$ | $642.14 \pm 38.64$ | 100.0 | 100.0 |
| | DPDT | 1 | $63.95 \pm 0.45$ | $83.96 \pm 0.35$ | $17.18 \pm 0.16$ | $303.99 \pm 5.28$ | 100.0 | 100.0 |
| | Literati | 1 | $63.95 \pm 0.45$ | $83.96 \pm 0.35$ | $4.05 \pm 0.34$ | $62.29 \pm 1.22$ | 100.0 | 100.0 |
| | | 2 | $64.73 \pm 1.73$ | $84.06 \pm 0.24$ | $8.82 \pm 1.22$ | $203.38 \pm 16.25$ | 100.0 | 100.0 |
| | | 3 | $64.73 \pm 1.73$ | $84.06 \pm 0.24$ | $9.24 \pm 0.85$ | $215.23 \pm 16.68$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $63.52 \pm 0.33$ | $83.28 \pm 0.43$ | $4.08 \pm 0.19$ | $18.87 \pm 0.21$ | 100.0 | 100.0 |
| | Literati-L | 1 | $63.52 \pm 0.33$ | $83.36 \pm 0.40$ | $\mathbf{1.39 \pm 0.10}$ | $\mathbf{6.46 \pm 0.27}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{64.40 \pm 1.77}$ | $\mathbf{83.38 \pm 0.41}$ | $2.31 \pm 0.25$ | $11.66 \pm 1.64$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{64.40 \pm 1.77}$ | $\mathbf{83.38 \pm 0.41}$ | $2.37 \pm 0.43$ | $11.30 \pm 1.15$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $66.62 \pm 0.19$ | $84.51 \pm 0.36$ | $1.71 \pm 0.19$ | $158.40 \pm 8.65$ | 100.0 | 100.0 |
| raisin | CART | 1 | $88.38 \pm 0.78$ | $89.97 \pm 0.85$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $88.67 \pm 1.06$ | $90.51 \pm 0.74$ | $0.19 \pm 0.10$ | $0.41 \pm 0.13$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $90.89 \pm 0.37$ | $93.43 \pm 0.18$ | $0.58 \pm 0.34$ | $1.46 \pm 1.64$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $91.49 \pm 0.60$ | $94.03 \pm 0.58$ | $2.26 \pm 1.26$ | $3.54 \pm 1.12$ | 0.0 | 0.0 |
| | Branches | 1 | $89.21 \pm 0.40$ | $90.19 \pm 0.34$ | $18.13 \pm 1.10$ | $343.55 \pm 34.20$ | 100.0 | 0.0 |
| | STreeD | 1 | $89.21 \pm 0.40$ | $91.33 \pm 0.37$ | $0.05 \pm 0.03$ | $1.04 \pm 0.07$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $89.21 \pm 0.40$ | $91.33 \pm 0.37$ | $9.93 \pm 0.77$ | $940.65 \pm 68.05$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $89.21 \pm 0.40$ | $91.33 \pm 0.37$ | $0.19 \pm 0.01$ | $11.67 \pm 0.76$ | 100.0 | 100.0 |
| | ConTree | 1 | $90.32 \pm 0.34$ | $92.57 \pm 1.10$ | $0.36 \pm 0.09$ | $18.87 \pm 3.08$ | 100.0 | 100.0 |
| | DPDT | 1 | $90.00 \pm 0.39$ | $92.51 \pm 0.31$ | $1.98 \pm 0.13$ | $34.02 \pm 2.23$ | 100.0 | 100.0 |
| | Literati | 1 | $90.00 \pm 0.39$ | $92.51 \pm 0.26$ | $0.25 \pm 0.03$ | $3.21 \pm 0.19$ | 100.0 | 100.0 |
| | | 2 | $90.83 \pm 0.31$ | $93.78 \pm 0.17$ | $1.41 \pm 0.29$ | $46.07 \pm 8.80$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{91.62 \pm 0.45}$ | $\mathbf{95.02 \pm 0.21}$ | $3.81 \pm 1.95$ | $178.40 \pm 97.39$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $89.84 \pm 0.46$ | $92.35 \pm 0.41$ | $0.57 \pm 0.02$ | $3.33 \pm 0.24$ | 100.0 | 100.0 |
| | Literati-L | 1 | $89.84 \pm 0.46$ | $92.35 \pm 0.41$ | $\mathbf{0.09 \pm 0.01}$ | $\mathbf{0.38 \pm 0.03}$ | 100.0 | 100.0 |
| | | 2 | $90.76 \pm 0.33$ | $93.30 \pm 0.21$ | $0.31 \pm 0.07$ | $1.52 \pm 0.32$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{91.30 \pm 0.60}$ | $\mathbf{93.94 \pm 0.44}$ | $0.45 \pm 0.15$ | $2.37 \pm 1.03$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $89.21 \pm 0.40$ | $91.33 \pm 0.37$ | $0.03 \pm 0.00$ | $1.05 \pm 0.06$ | 100.0 | 100.0 |
| rice | CART | 1 | $93.15 \pm 0.27$ | $93.55 \pm 0.34$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $93.15 \pm 0.27$ | $93.80 \pm 0.29$ | $0.08 \pm 0.02$ | $0.63 \pm 0.19$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $93.59 \pm 0.39$ | $94.53 \pm 0.26$ | $0.47 \pm 0.34$ | $0.56 \pm 0.18$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $93.80 \pm 0.31$ | $94.74 \pm 0.26$ | $2.44 \pm 0.79$ | $3.35 \pm 0.37$ | 0.0 | 0.0 |
| | Branches | 1 | $93.47 \pm 0.23$ | $93.60 \pm 0.18$ | $23.43 \pm 5.40$ | $239.21 \pm 21.92$ | 100.0 | 0.0 |
| | STreeD | 1 | $93.47 \pm 0.23$ | $93.99 \pm 0.20$ | $0.07 \pm 0.01$ | $2.95 \pm 0.34$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $93.47 \pm 0.23$ | $93.99 \pm 0.20$ | $13.44 \pm 0.62$ | $1274.31 \pm 25.89$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $93.47 \pm 0.23$ | $93.99 \pm 0.20$ | $0.41 \pm 0.03$ | $23.68 \pm 1.02$ | 100.0 | 100.0 |
| | ConTree | 1 | $93.96 \pm 0.27$ | $91.94 \pm 6.58$ | $6.91 \pm 2.01$ | $881.23 \pm 97.97$ | 100.0 | 100.0 |
| | DPDT | 1 | $93.81 \pm 0.26$ | $94.56 \pm 0.28$ | $4.16 \pm 0.15$ | $82.30 \pm 5.25$ | 100.0 | 100.0 |
| | Literati | 1 | $93.81 \pm 0.26$ | $94.51 \pm 0.28$ | $0.84 \pm 0.06$ | $11.72 \pm 0.90$ | 100.0 | 100.0 |
| | | 2 | $94.14 \pm 0.28$ | $95.04 \pm 0.30$ | $7.57 \pm 1.09$ | $300.65 \pm 34.59$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{94.42 \pm 0.27}$ | $\mathbf{95.52 \pm 0.22}$ | $24.22 \pm 5.85$ | $1722.30 \pm 433.68$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $93.81 \pm 0.25$ | $94.40 \pm 0.26$ | $1.23 \pm 0.05$ | $6.60 \pm 0.64$ | 100.0 | 100.0 |
| | Literati-L | 1 | $93.81 \pm 0.25$ | $94.38 \pm 0.24$ | $\mathbf{0.33 \pm 0.02}$ | $\mathbf{1.23 \pm 0.13}$ | 100.0 | 100.0 |
| | | 2 | $94.08 \pm 0.30$ | $94.82 \pm 0.32$ | $1.85 \pm 0.15$ | $8.54 \pm 0.43$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{94.26 \pm 0.28}$ | $\mathbf{95.02 \pm 0.28}$ | $3.98 \pm 0.84$ | $18.16 \pm 3.61$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $93.47 \pm 0.23$ | $93.99 \pm 0.20$ | $0.07 \pm 0.01$ | $2.61 \pm 0.08$ | 100.0 | 100.0 |
| room | CART | 1 | $96.92 \pm 0.09$ | $99.06 \pm 0.04$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.14 \pm 0.04$ | $99.33 \pm 0.12$ | $0.85 \pm 0.18$ | $1.03 \pm 0.33$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $98.27 \pm 0.07$ | $99.39 \pm 0.27$ | $0.98 \pm 0.74$ | $0.84 \pm 0.53$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $\mathbf{99.59 \pm 0.16}$ | $99.59 \pm 0.08$ | $4.58 \pm 4.62$ | $14.54 \pm 7.55$ | 0.0 | 0.0 |
| | Branches | 1 | $97.77 \pm 0.25$ | $96.96 \pm 0.49$ | $58.41 \pm 4.46$ | $194.74 \pm 11.29$ | 100.0 | 0.0 |
| | STreeD | 1 | $97.77 \pm 0.25$ | $99.70 \pm 0.05$ | $0.43 \pm 0.04$ | $13.05 \pm 0.93$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $97.77 \pm 0.25$ | $99.70 \pm 0.05$ | $15.79 \pm 1.34$ | $1081.59 \pm 62.75$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $97.77 \pm 0.25$ | $99.70 \pm 0.05$ | $1.19 \pm 0.02$ | $58.85 \pm 2.11$ | 100.0 | 100.0 |
| | ConTree | 1 | $99.22 \pm 0.06$ | $\mathbf{99.90 \pm 0.03}$ | $0.65 \pm 0.06$ | $14.62 \pm 2.36$ | 100.0 | 100.0 |
| | DPDT | 1 | $99.19 \pm 0.06$ | $99.78 \pm 0.04$ | $5.04 \pm 0.20$ | $64.28 \pm 5.67$ | 100.0 | 100.0 |
| | Literati | 1 | $99.18 \pm 0.05$ | $99.78 \pm 0.04$ | $1.89 \pm 0.15$ | $18.13 \pm 1.40$ | 100.0 | 100.0 |
| | | 2 | $99.18 \pm 0.05$ | $99.81 \pm 0.04$ | $5.62 \pm 0.79$ | $83.22 \pm 9.89$ | 100.0 | 100.0 |
| | | 3 | $99.21 \pm 0.01$ | $99.82 \pm 0.03$ | $6.51 \pm 0.72$ | $107.91 \pm 18.87$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $\mathbf{99.19 \pm 0.05}$ | $99.72 \pm 0.04$ | $1.60 \pm 0.02$ | $7.56 \pm 0.41$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.18 \pm 0.05$ | $99.74 \pm 0.05$ | $\mathbf{0.77 \pm 0.06}$ | $\mathbf{2.74 \pm 0.15}$ | 100.0 | 100.0 |
| | | 2 | $99.18 \pm 0.05$ | $\mathbf{99.77 \pm 0.02}$ | $1.50 \pm 0.14$ | $5.54 \pm 0.72$ | 100.0 | 100.0 |
| | | 3 | $99.18 \pm 0.05$ | $\mathbf{99.77 \pm 0.03}$ | $1.64 \pm 0.26$ | $5.94 \pm 0.85$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $97.77 \pm 0.25$ | $99.70 \pm 0.05$ | $0.24 \pm 0.05$ | $9.16 \pm 0.81$ | 100.0 | 100.0 |

**Table 6 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ | $D = 3$ | $D = 4$ |
| segment | CART | 1 | $56.99 \pm 0.06$ | $69.76 \pm 0.91$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $57.12 \pm 0.03$ | $74.99 \pm 6.46$ | $0.13 \pm 0.00$ | $0.33 \pm 0.13$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $56.99 \pm 0.06$ | $70.92 \pm 0.22$ | $0.68 \pm 0.33$ | $0.77 \pm 0.47$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $57.12 \pm 0.03$ | $78.63 \pm 2.39$ | $1.67 \pm 0.76$ | $3.40 \pm 1.49$ | 0.0 | 0.0 |
| | Branches | 1 | $85.68 \pm 0.26$ | $86.68 \pm 1.13$ | $324.12 \pm 22.55$ | $299.77 \pm 57.33$ | 0.0 | 0.0 |
| | STreeD | 1 | $85.68 \pm 0.26$ | $93.96 \pm 0.25$ | $0.76 \pm 0.06$ | $62.32 \pm 2.41$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $85.68 \pm 0.26$ | $93.96 \pm 0.25$ | $149.64 \pm 3.95$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | CADL8.5 | 1 | $85.68 \pm 0.26$ | $93.96 \pm 0.25$ | $4.27 \pm 0.16$ | $724.23 \pm 28.87$ | 100.0 | 100.0 |
| | ConTree | 1 | $88.02 \pm 0.46$ | $90.83 \pm 6.79$ | $0.95 \pm 0.13$ | $104.89 \pm 14.16$ | 100.0 | 100.0 |
| | DPDT | 1 | $87.66 \pm 0.36$ | $94.76 \pm 0.47$ | $3.34 \pm 0.10$ | $51.14 \pm 1.92$ | 100.0 | 100.0 |
| | Literati | 1 | $87.69 \pm 0.35$ | $94.66 \pm 0.37$ | $1.03 \pm 0.02$ | $11.24 \pm 0.36$ | 100.0 | 100.0 |
| | | 2 | $89.05 \pm 1.13$ | $95.75 \pm 0.24$ | $5.08 \pm 0.58$ | $123.24 \pm 15.63$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{89.07 \pm 1.11}$ | $\mathbf{95.96 \pm 0.33}$ | $9.70 \pm 2.16$ | $360.57 \pm 88.96$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $87.66 \pm 0.36$ | $94.53 \pm 0.44$ | $1.08 \pm 0.02$ | $5.09 \pm 0.70$ | 100.0 | 100.0 |
| | Literati-L | 1 | $87.69 \pm 0.35$ | $94.51 \pm 0.44$ | $\mathbf{0.47 \pm 0.01}$ | $\mathbf{1.61 \pm 0.03}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{87.85 \pm 0.41}$ | $95.52 \pm 0.19$ | $1.43 \pm 0.17$ | $5.24 \pm 0.59$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{87.85 \pm 0.41}$ | $\mathbf{95.65 \pm 0.20}$ | $2.05 \pm 0.42$ | $7.29 \pm 1.45$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $85.68 \pm 0.26$ | $93.96 \pm 0.25$ | $0.38 \pm 0.01$ | $44.95 \pm 2.06$ | 100.0 | 100.0 |
| skin | CART | 1 | $96.58 \pm 0.02$ | $97.48 \pm 0.01$ | $\mathbf{0.06 \pm 0.00}$ | $\mathbf{0.07 \pm 0.01}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $96.61 \pm 0.03$ | $98.71 \pm 0.01$ | $1.51 \pm 0.88$ | $6.99 \pm 1.06$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $97.34 \pm 0.04$ | $98.47 \pm 0.02$ | $1.29 \pm 0.29$ | $1.74 \pm 0.12$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $98.15 \pm 0.03$ | $99.23 \pm 0.05$ | $5.24 \pm 1.09$ | $14.15 \pm 2.56$ | 0.0 | 0.0 |
| | Branches | 1 | $96.33 \pm 0.02$ | $97.14 \pm 0.10$ | $14.99 \pm 1.50$ | $63.65 \pm 3.98$ | 100.0 | 0.0 |
| | STreeD | 1 | $96.33 \pm 0.02$ | $98.53 \pm 0.02$ | $1.23 \pm 0.10$ | $13.71 \pm 0.67$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $96.33 \pm 0.02$ | $98.53 \pm 0.02$ | $16.97 \pm 0.07$ | $477.27 \pm 13.42$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $96.33 \pm 0.02$ | $98.53 \pm 0.02$ | $1.37 \pm 0.06$ | $25.92 \pm 0.38$ | 100.0 | 100.0 |
| | ConTree | 1 | $96.72 \pm 0.10$ | $98.73 \pm 0.07$ | $5.59 \pm 0.22$ | $97.07 \pm 9.76$ | 100.0 | 100.0 |
| | DPDT | 1 | $96.59 \pm 0.02$ | $98.71 \pm 0.05$ | $88.07 \pm 3.24$ | $1396.45 \pm 104.36$ | 100.0 | 100.0 |
| | Literati | 1 | $96.59 \pm 0.02$ | $98.71 \pm 0.05$ | $13.82 \pm 0.86$ | $185.07 \pm 22.46$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{98.59 \pm 0.01}$ | $\mathbf{99.58 \pm 0.01}$ | $236.92 \pm 6.23$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $\mathbf{98.59 \pm 0.01}$ | $99.37 \pm 0.16$ | $1233.84 \pm 81.81$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT-L | 1 | $96.59 \pm 0.02$ | $98.63 \pm 0.05$ | $19.37 \pm 1.22$ | $84.91 \pm 6.11$ | 100.0 | 100.0 |
| | Literati-L | 1 | $96.59 \pm 0.02$ | $98.64 \pm 0.05$ | $\mathbf{5.34 \pm 0.06}$ | $\mathbf{23.02 \pm 0.69}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{98.33 \pm 0.03}$ | $99.48 \pm 0.01$ | $36.19 \pm 1.05$ | $208.37 \pm 16.08$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.33 \pm 0.03}$ | $\mathbf{99.57 \pm 0.08}$ | $79.08 \pm 4.84$ | $432.89 \pm 21.30$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $96.33 \pm 0.02$ | $98.53 \pm 0.02$ | $0.88 \pm 0.09$ | $10.52 \pm 1.31$ | 100.0 | 100.0 |
| wilt | CART | 1 | $97.66 \pm 0.19$ | $98.39 \pm 0.16$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $97.76 \pm 0.21$ | $98.64 \pm 0.16$ | $0.22 \pm 0.05$ | $0.65 \pm 0.05$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $97.77 \pm 0.28$ | $98.49 \pm 0.17$ | $0.74 \pm 0.30$ | $0.61 \pm 0.43$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $97.98 \pm 0.30$ | $98.67 \pm 0.21$ | $2.44 \pm 0.78$ | $3.18 \pm 1.28$ | 0.0 | 0.0 |
| | Branches | 1 | $97.93 \pm 0.07$ | $98.55 \pm 0.06$ | $8.28 \pm 0.73$ | $233.58 \pm 8.50$ | 100.0 | 0.0 |
| | STreeD | 1 | $97.93 \pm 0.07$ | $98.59 \pm 0.08$ | $0.07 \pm 0.03$ | $1.08 \pm 0.10$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $97.93 \pm 0.07$ | $98.59 \pm 0.08$ | $3.89 \pm 0.46$ | $400.43 \pm 47.26$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $97.93 \pm 0.07$ | $98.59 \pm 0.08$ | $0.18 \pm 0.01$ | $8.17 \pm 0.23$ | 100.0 | 100.0 |
| | ConTree | 1 | $98.28 \pm 0.10$ | $99.08 \pm 0.02$ | $0.23 \pm 0.04$ | $6.24 \pm 1.21$ | 100.0 | 100.0 |
| | DPDT | 1 | $98.19 \pm 0.10$ | $98.97 \pm 0.02$ | $3.69 \pm 0.34$ | $56.86 \pm 5.62$ | 100.0 | 100.0 |
| | Literati | 1 | $98.19 \pm 0.10$ | $98.96 \pm 0.02$ | $0.71 \pm 0.03$ | $9.03 \pm 0.62$ | 100.0 | 100.0 |
| | | 2 | $98.30 \pm 0.10$ | $99.19 \pm 0.07$ | $8.98 \pm 2.17$ | $286.56 \pm 57.16$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.41 \pm 0.08}$ | $\mathbf{99.34 \pm 0.09}$ | $31.83 \pm 13.72$ | $1771.64 \pm 739.89$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $98.19 \pm 0.10$ | $98.93 \pm 0.03$ | $1.16 \pm 0.06$ | $5.64 \pm 0.13$ | 100.0 | 100.0 |
| | Literati-L | 1 | $98.19 \pm 0.10$ | $98.93 \pm 0.03$ | $\mathbf{0.30 \pm 0.02}$ | $\mathbf{1.13 \pm 0.07}$ | 100.0 | 100.0 |
| | | 2 | $98.28 \pm 0.10$ | $99.14 \pm 0.07$ | $1.77 \pm 0.26$ | $8.49 \pm 1.05$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.32 \pm 0.09}$ | $\mathbf{99.24 \pm 0.06}$ | $3.75 \pm 1.18$ | $18.89 \pm 6.04$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $97.93 \pm 0.07$ | $98.59 \pm 0.08$ | $0.03 \pm 0.00$ | $0.81 \pm 0.05$ | 100.0 | 100.0 |

Table 7: Training accuracy, runtime, and optimality proof rate across depth budgets across all datasets evaluated for Depths 5 and 6. The three blocks for each dataset report the normal, light (-L), and Literati-PreQuant configurations, respectively. The dashed lines separate Literati from baselines. Bold indicates the best value per dataset and depth within the normal and light blocks for training accuracy and runtime.

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s ± std.) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ |
| adult | CART | 1 | $83.28 \pm 0.16$ | $83.72 \pm 0.15$ | $\mathbf{0.10 \pm 0.01}$ | $\mathbf{0.11 \pm 0.01}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $83.75 \pm 0.21$ | $83.98 \pm 0.18$ | $22.57 \pm 4.53$ | $24.69 \pm 5.62$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $83.28 \pm 0.16$ | $83.72 \pm 0.14$ | $3.16 \pm 0.64$ | $4.73 \pm 0.88$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $83.71 \pm 0.23$ | $83.98 \pm 0.18$ | $58.31 \pm 32.75$ | $79.33 \pm 52.67$ | 0.0 | 0.0 |
| | Branches | 1 | $82.92 \pm 0.32$ | $82.88 \pm 0.45$ | $70.52 \pm 9.93$ | $68.82 \pm 9.11$ | 0.0 | 0.0 |
| | STreeD | 1 | $84.34 \pm 0.10$ | $\mathbf{84.85 \pm 0.08}$ | $337.08 \pm 7.08$ | $3599.66 \pm 0.34$ | 100.0 | 0.0 |
| | LDS-DL8.5 | 1 | $84.34 \pm 0.10$ | $84.60 \pm 0.08$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $84.34 \pm 0.10$ | $84.80 \pm 0.08$ | $3587.47 \pm 28.02$ | $3600.00 \pm 0.00$ | 20.0 | 0.0 |
| | ConTree | 1 | $83.86 \pm 0.16$ | $76.28 \pm 0.07$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $75.22 \pm 0.00$ | $75.22 \pm 0.00$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $83.79 \pm 1.22$ | $84.72 \pm 0.13$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 2 | $\mathbf{84.37 \pm 0.13}$ | $84.75 \pm 0.14$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $84.36 \pm 0.13$ | $84.71 \pm 0.14$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $84.21 \pm 0.16$ | $\mathbf{84.56 \pm 0.13}$ | $202.25 \pm 12.82$ | $914.85 \pm 35.05$ | 100.0 | 100.0 |
| | Literati-L | 1 | $84.21 \pm 0.16$ | $\mathbf{84.56 \pm 0.13}$ | $\mathbf{91.37 \pm 16.62}$ | $\mathbf{294.38 \pm 82.81}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{84.22 \pm 0.16}$ | $\mathbf{84.56 \pm 0.13}$ | $164.78 \pm 49.19$ | $499.07 \pm 144.96$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{84.22 \pm 0.16}$ | $\mathbf{84.56 \pm 0.13}$ | $187.37 \pm 82.99$ | $585.48 \pm 199.47$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $84.34 \pm 0.10$ | $84.80 \pm 0.09$ | $573.57 \pm 52.73$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| avila | CART | 1 | $61.11 \pm 0.60$ | $64.26 \pm 1.27$ | $\mathbf{0.03 \pm 0.01}$ | $\mathbf{0.03 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $61.70 \pm 0.89$ | $65.48 \pm 0.96$ | $1.27 \pm 0.31$ | $2.67 \pm 0.65$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $61.94 \pm 1.18$ | $66.26 \pm 1.54$ | $1.95 \pm 2.38$ | $2.19 \pm 1.59$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $64.24 \pm 1.93$ | $69.31 \pm 1.60$ | $6.28 \pm 1.95$ | $9.50 \pm 2.03$ | 0.0 | 0.0 |
| | Branches | 1 | $55.91 \pm 2.17$ | $54.09 \pm 2.30$ | $109.60 \pm 8.67$ | $108.42 \pm 13.41$ | 0.0 | 0.0 |
| | STreeD | 1 | $67.37 \pm 0.15$ | $67.37 \pm 0.15$ | $3599.75 \pm 0.34$ | $3599.67 \pm 0.32$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $67.07 \pm 0.14$ | $73.06 \pm 0.16$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $67.35 \pm 0.14$ | $71.10 \pm 0.09$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $66.73 \pm 0.29$ | $66.29 \pm 1.36$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $47.94 \pm 0.05$ | $47.94 \pm 0.05$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $69.10 \pm 0.39$ | $77.14 \pm 0.18$ | $1376.47 \pm 58.47$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $\mathbf{71.61 \pm 0.50}$ | $76.58 \pm 3.02$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $69.75 \pm 1.96$ | $\mathbf{77.50 \pm 2.93}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $68.57 \pm 0.41$ | $76.06 \pm 0.35$ | $132.82 \pm 2.48$ | $601.06 \pm 12.80$ | 100.0 | 100.0 |
| | Literati-L | 1 | $68.57 \pm 0.41$ | $76.02 \pm 0.46$ | $\mathbf{32.97 \pm 1.02}$ | $\mathbf{137.44 \pm 4.68}$ | 100.0 | 100.0 |
| | | 2 | $70.36 \pm 0.32$ | $78.35 \pm 0.23$ | $121.20 \pm 10.41$ | $532.36 \pm 50.02$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{70.59 \pm 0.37}$ | $\mathbf{78.76 \pm 0.19}$ | $151.86 \pm 16.59$ | $653.07 \pm 90.81$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $67.86 \pm 0.21$ | $74.10 \pm 1.62$ | $3244.04 \pm 112.75$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| bank | CART | 1 | $99.12 \pm 0.26$ | $99.62 \pm 0.20$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.67 \pm 0.29$ | $99.90 \pm 0.15$ | $0.64 \pm 0.17$ | $0.67 \pm 0.22$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $99.40 \pm 0.14$ | $99.90 \pm 0.10$ | $1.22 \pm 1.86$ | $0.47 \pm 0.08$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.94 \pm 0.09$ | $\mathbf{100.00 \pm 0.00}$ | $12.45 \pm 8.98$ | $10.73 \pm 10.95$ | 0.0 | 0.0 |
| | Branches | 1 | $99.92 \pm 0.09$ | $\mathbf{100.00 \pm 0.00}$ | $218.87 \pm 184.02$ | $78.00 \pm 40.59$ | 40.0 | 100.0 |
| | STreeD | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.11 \pm 0.01$ | $0.11 \pm 0.02$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.09 \pm 0.02$ | $0.01 \pm 0.00$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.03 \pm 0.02$ | $0.01 \pm 0.01$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 100.0 | 100.0 |
| | DPDT | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $65.62 \pm 14.28$ | $389.23 \pm 105.42$ | 100.0 | 100.0 |
| | Literati | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.02 \pm 0.00$ | $0.03 \pm 0.01$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.02 \pm 0.00$ | $0.04 \pm 0.01$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.02 \pm 0.00$ | $0.04 \pm 0.01$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $6.67 \pm 0.34$ | $19.55 \pm 1.82$ | 100.0 | 100.0 |
| | Literati-L | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.04 \pm 0.05$ | $\mathbf{0.02 \pm 0.00}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.03 \pm 0.02$ | $0.03 \pm 0.00$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{0.02 \pm 0.01}$ | $0.03 \pm 0.00$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $100.00 \pm 0.00$ | $100.00 \pm 0.00$ | $0.10 \pm 0.03$ | $0.12 \pm 0.02$ | 100.0 | 100.0 |

**Table 7 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s ± std.) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ |
| bean | CART | 1 | $89.04 \pm 0.21$ | $91.33 \pm 0.21$ | $\mathbf{0.06 \pm 0.00}$ | $\mathbf{0.07 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $90.13 \pm 0.14$ | $92.34 \pm 0.33$ | $3.14 \pm 0.71$ | $5.37 \pm 0.59$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $89.46 \pm 0.19$ | $92.03 \pm 0.23$ | $1.19 \pm 0.15$ | $4.90 \pm 7.33$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $90.37 \pm 0.60$ | $93.07 \pm 0.31$ | $16.15 \pm 12.50$ | $58.25 \pm 60.21$ | 0.0 | 0.0 |
| | Branches | 1 | $82.59 \pm 0.96$ | $82.35 \pm 0.96$ | $221.30 \pm 17.01$ | $206.18 \pm 17.53$ | 0.0 | 0.0 |
| | STreeD | 1 | $91.58 \pm 0.10$ | $91.58 \pm 0.10$ | $3599.69 \pm 0.25$ | $3599.41 \pm 0.08$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $91.36 \pm 0.32$ | $92.78 \pm 0.17$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $91.46 \pm 0.09$ | $61.09 \pm 0.03$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $88.61 \pm 8.75$ | $73.55 \pm 0.65$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $40.82 \pm 0.09$ | $40.82 \pm 0.09$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $92.39 \pm 0.19$ | $93.79 \pm 0.10$ | $2112.90 \pm 164.69$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $92.57 \pm 0.22$ | $93.95 \pm 0.13$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{92.59 \pm 0.20}$ | $\mathbf{94.03 \pm 0.13}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $92.24 \pm 0.16$ | $93.49 \pm 0.13$ | $144.61 \pm 2.23$ | $650.35 \pm 34.17$ | 100.0 | 100.0 |
| | Literati-L | 1 | $92.24 \pm 0.16$ | $93.49 \pm 0.13$ | $\mathbf{50.95 \pm 1.03}$ | $\mathbf{211.91 \pm 5.98}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{92.38 \pm 0.21}$ | $93.76 \pm 0.20$ | $201.46 \pm 14.29$ | $867.59 \pm 67.38$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{92.38 \pm 0.21}$ | $\mathbf{93.78 \pm 0.20}$ | $267.56 \pm 23.43$ | $1130.77 \pm 92.55$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $91.71 \pm 0.10$ | $92.93 \pm 0.18$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| bidding | CART | 1 | $99.62 \pm 0.06$ | $99.66 \pm 0.06$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.63 \pm 0.05$ | $99.70 \pm 0.07$ | $0.26 \pm 0.09$ | $0.37 \pm 0.14$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $99.75 \pm 0.08$ | $99.95 \pm 0.06$ | $0.82 \pm 0.46$ | $0.66 \pm 0.13$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.77 \pm 0.09$ | $99.96 \pm 0.06$ | $1.71 \pm 0.47$ | $2.70 \pm 0.61$ | 0.0 | 0.0 |
| | Branches | 1 | $99.44 \pm 0.06$ | $99.45 \pm 0.06$ | $170.13 \pm 7.27$ | $168.86 \pm 4.35$ | 0.0 | 0.0 |
| | STreeD | 1 | $99.95 \pm 0.02$ | $\mathbf{100.00 \pm 0.00}$ | $23.78 \pm 1.71$ | $27.65 \pm 3.56$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $99.95 \pm 0.01$ | $\mathbf{100.00 \pm 0.00}$ | $3600.00 \pm 0.00$ | $1507.31 \pm 804.29$ | 0.0 | 100.0 |
| | CADL8.5 | 1 | $99.95 \pm 0.02$ | $\mathbf{100.00 \pm 0.00}$ | $346.17 \pm 24.55$ | $97.31 \pm 62.55$ | 100.0 | 100.0 |
| | ConTree | 1 | $99.99 \pm 0.01$ | $\mathbf{100.00 \pm 0.00}$ | $9.60 \pm 6.58$ | $0.01 \pm 0.00$ | 100.0 | 100.0 |
| | DPDT | 1 | $99.99 \pm 0.01$ | $\mathbf{100.00 \pm 0.00}$ | $142.87 \pm 32.55$ | $874.19 \pm 274.92$ | 100.0 | 100.0 |
| | Literati | 1 | $99.98 \pm 0.02$ | $\mathbf{100.00 \pm 0.00}$ | $9.74 \pm 4.04$ | $0.85 \pm 1.77$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $1.12 \pm 2.28$ | $0.07 \pm 0.03$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $1.27 \pm 2.74$ | $0.07 \pm 0.05$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $99.97 \pm 0.02$ | $\mathbf{100.00 \pm 0.00}$ | $11.88 \pm 1.21$ | $37.73 \pm 4.52$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.97 \pm 0.02$ | $\mathbf{100.00 \pm 0.00}$ | $1.20 \pm 0.35$ | $\mathbf{0.08 \pm 0.08}$ | 100.0 | 100.0 |
| | | 2 | $99.99 \pm 0.01$ | $\mathbf{100.00 \pm 0.00}$ | $2.81 \pm 3.69$ | $0.09 \pm 0.09$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{0.10 \pm 0.13}$ | $\mathbf{0.08 \pm 0.07}$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $99.95 \pm 0.02$ | $100.00 \pm 0.00$ | $41.37 \pm 2.54$ | $280.74 \pm 45.97$ | 100.0 | 100.0 |
| electricity | CART | 1 | $77.68 \pm 1.09$ | $79.33 \pm 0.48$ | $\mathbf{0.06 \pm 0.00}$ | $\mathbf{0.07 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $79.23 \pm 0.74$ | $80.94 \pm 0.22$ | $4.31 \pm 2.55$ | $10.50 \pm 4.19$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $80.09 \pm 0.94$ | $82.21 \pm 1.19$ | $3.11 \pm 1.44$ | $6.26 \pm 8.70$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $80.96 \pm 0.37$ | $83.85 \pm 0.34$ | $19.92 \pm 16.73$ | $46.72 \pm 36.37$ | 0.0 | 0.0 |
| | Branches | 1 | $76.03 \pm 0.15$ | $75.98 \pm 0.15$ | $80.53 \pm 11.76$ | $81.48 \pm 6.00$ | 0.0 | 0.0 |
| | STreeD | 1 | $81.20 \pm 0.06$ | $82.11 \pm 0.04$ | $856.64 \pm 37.81$ | $3599.60 \pm 0.29$ | 100.0 | 0.0 |
| | LDS-DL8.5 | 1 | $80.74 \pm 0.07$ | $81.45 \pm 0.11$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $81.04 \pm 0.03$ | $81.87 \pm 0.05$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $81.00 \pm 0.09$ | $77.30 \pm 0.41$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $75.81 \pm 0.12$ | $75.81 \pm 0.12$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $82.39 \pm 0.11$ | $84.21 \pm 0.08$ | $2668.13 \pm 281.03$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $\mathbf{83.47 \pm 0.11}$ | $85.26 \pm 0.39$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $83.44 \pm 1.28$ | $\mathbf{85.46 \pm 0.90}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $81.97 \pm 0.14$ | $83.52 \pm 0.06$ | $200.34 \pm 9.31$ | $970.37 \pm 19.42$ | 100.0 | 100.0 |
| | Literati-L | 1 | $81.97 \pm 0.14$ | $83.52 \pm 0.06$ | $\mathbf{69.98 \pm 3.32}$ | $\mathbf{289.38 \pm 8.36}$ | 100.0 | 100.0 |
| | | 2 | $83.01 \pm 0.16$ | $85.01 \pm 0.10$ | $1089.81 \pm 50.06$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $\mathbf{83.37 \pm 0.05}$ | $\mathbf{85.42 \pm 0.13}$ | $3218.44 \pm 243.81$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati-PreQuant | 1 | $81.20 \pm 0.06$ | $82.29 \pm 0.08$ | $798.37 \pm 52.61$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| eucalyptus | CART | 1 | $55.53 \pm 0.99$ | $59.03 \pm 1.43$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.02 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $55.64 \pm 1.02$ | $59.18 \pm 1.39$ | $14.35 \pm 7.98$ | $16.44 \pm 5.62$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $55.33 \pm 1.67$ | $59.61 \pm 1.12$ | $5.95 \pm 1.05$ | $7.92 \pm 0.60$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $55.56 \pm 1.32$ | $59.69 \pm 1.06$ | $12.95 \pm 2.65$ | $57.22 \pm 84.23$ | 0.0 | 0.0 |
| | Branches | 1 | $50.70 \pm 1.79$ | $52.02 \pm 2.55$ | $166.08 \pm 21.92$ | $186.35 \pm 23.95$ | 0.0 | 0.0 |
| | STreeD | 1 | $\mathbf{66.85 \pm 1.00}$ | $66.85 \pm 1.00$ | $3326.32 \pm 366.39$ | $3599.65 \pm 0.31$ | 60.0 | 0.0 |
| | LDS-DL8.5 | 1 | $66.19 \pm 1.23$ | $73.66 \pm 1.09$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $60.74 \pm 0.72$ | $61.75 \pm 1.65$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $49.84 \pm 0.65$ | $43.50 \pm 2.63$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $38.68 \pm 0.35$ | $38.68 \pm 0.35$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $65.64 \pm 0.79$ | $73.62 \pm 0.70$ | $661.13 \pm 195.64$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $66.26 \pm 1.22$ | $\mathbf{74.20 \pm 0.99}$ | $635.89 \pm 173.16$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $66.26 \pm 1.22$ | $\mathbf{74.20 \pm 0.99}$ | $701.66 \pm 237.98$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT-L | 1 | $65.06 \pm 0.76$ | $\mathbf{72.37 \pm 0.28}$ | $42.35 \pm 0.74$ | $193.81 \pm 15.62$ | 100.0 | 100.0 |
| | Literati-L | 1 | $65.06 \pm 0.76$ | $72.22 \pm 0.16$ | $\mathbf{19.11 \pm 2.78}$ | $\mathbf{33.70 \pm 3.38}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{65.10 \pm 0.68}$ | $72.33 \pm 0.32$ | $20.85 \pm 3.42$ | $37.26 \pm 6.26$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{65.10 \pm 0.68}$ | $72.33 \pm 0.32$ | $20.57 \pm 2.25$ | $37.42 \pm 3.57$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $66.85 \pm 1.00$ | $76.38 \pm 0.88$ | $2923.47 \pm 223.20$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |

**Table 7 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s ± std.) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ |
| eye-movements | CART | 1 | $54.78 \pm 0.84$ | $56.80 \pm 0.49$ | $\mathbf{0.04 \pm 0.00}$ | $\mathbf{0.05 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $55.69 \pm 0.61$ | $57.53 \pm 0.42$ | $5.19 \pm 0.60$ | $8.32 \pm 2.38$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $54.71 \pm 1.13$ | $58.57 \pm 1.58$ | $3.58 \pm 4.50$ | $1.97 \pm 0.35$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $56.28 \pm 1.27$ | $60.23 \pm 1.64$ | $11.26 \pm 1.25$ | $45.05 \pm 42.97$ | 0.0 | 0.0 |
| | Branches | 1 | $44.82 \pm 0.31$ | $44.73 \pm 0.17$ | $161.28 \pm 38.19$ | $152.91 \pm 42.96$ | 0.0 | 0.0 |
| | STreeD | 1 | $58.64 \pm 0.30$ | $58.64 \pm 0.30$ | $3599.51 \pm 0.20$ | $3599.49 \pm 0.18$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $58.08 \pm 0.31$ | $60.88 \pm 0.48$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $50.02 \pm 0.21$ | $49.82 \pm 0.26$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $54.92 \pm 0.42$ | $53.38 \pm 2.48$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $43.85 \pm 0.70$ | $43.85 \pm 0.70$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $59.71 \pm 0.30$ | $64.94 \pm 0.76$ | $1301.77 \pm 118.27$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $59.60 \pm 4.29$ | $67.12 \pm 0.50$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{62.15 \pm 0.71}$ | $\mathbf{67.88 \pm 0.43}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $59.24 \pm 0.28$ | $63.54 \pm 0.66$ | $115.66 \pm 3.58$ | $559.45 \pm 8.25$ | 100.0 | 100.0 |
| | Literati-L | 1 | $59.25 \pm 0.27$ | $63.54 \pm 0.66$ | $\mathbf{29.99 \pm 3.78}$ | $\mathbf{125.12 \pm 17.16}$ | 100.0 | 100.0 |
| | | 2 | $60.02 \pm 0.39$ | $65.05 \pm 0.43$ | $74.12 \pm 15.20$ | $338.83 \pm 68.35$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{60.14 \pm 0.36}$ | $\mathbf{65.19 \pm 0.50}$ | $87.93 \pm 25.30$ | $414.05 \pm 113.81$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $59.56 \pm 0.26$ | $63.89 \pm 0.48$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| eye-state | CART | 1 | $72.70 \pm 0.43$ | $75.60 \pm 0.67$ | $\mathbf{0.03 \pm 0.00}$ | $\mathbf{0.03 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $74.73 \pm 0.78$ | $77.94 \pm 0.92$ | $3.32 \pm 0.66$ | $10.30 \pm 5.67$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $75.02 \pm 0.63$ | $78.21 \pm 0.62$ | $1.31 \pm 0.34$ | $1.48 \pm 0.31$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $77.26 \pm 0.45$ | $79.96 \pm 0.67$ | $14.34 \pm 2.96$ | $23.08 \pm 3.73$ | 0.0 | 0.0 |
| | Branches | 1 | $60.64 \pm 0.25$ | $60.65 \pm 0.22$ | $110.11 \pm 4.79$ | $113.89 \pm 8.92$ | 0.0 | 0.0 |
| | STreeD | 1 | $78.33 \pm 0.09$ | $78.33 \pm 0.09$ | $3599.64 \pm 0.25$ | $3599.65 \pm 0.20$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $77.06 \pm 0.17$ | $80.09 \pm 0.46$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $77.87 \pm 0.22$ | $64.31 \pm 3.92$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $75.40 \pm 0.50$ | $73.48 \pm 0.46$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $59.53 \pm 0.21$ | $59.53 \pm 0.21$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $79.07 \pm 0.26$ | $83.12 \pm 0.33$ | $1515.20 \pm 72.66$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $79.60 \pm 0.22$ | $\mathbf{83.62 \pm 0.56}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{79.62 \pm 0.29}$ | $83.54 \pm 0.46$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $78.37 \pm 0.27$ | $82.22 \pm 0.29$ | $101.39 \pm 5.13$ | $494.10 \pm 23.02$ | 100.0 | 100.0 |
| | Literati-L | 1 | $78.37 \pm 0.27$ | $82.21 \pm 0.29$ | $\mathbf{35.22 \pm 1.50}$ | $\mathbf{157.65 \pm 21.07}$ | 100.0 | 100.0 |
| | | 2 | $78.88 \pm 0.49$ | $82.69 \pm 0.32$ | $106.28 \pm 10.31$ | $526.22 \pm 54.18$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{78.93 \pm 0.44}$ | $\mathbf{82.70 \pm 0.31}$ | $149.27 \pm 22.56$ | $682.40 \pm 119.61$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $78.66 \pm 0.08$ | $81.60 \pm 0.78$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| fault | CART | 1 | $67.39 \pm 1.47$ | $72.53 \pm 0.69$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $69.28 \pm 0.86$ | $74.40 \pm 0.77$ | $1.35 \pm 0.37$ | $2.34 \pm 0.47$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $69.91 \pm 1.87$ | $74.09 \pm 2.08$ | $5.51 \pm 3.85$ | $1.94 \pm 0.96$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $72.37 \pm 1.94$ | $75.96 \pm 1.74$ | $10.08 \pm 7.57$ | $7.61 \pm 1.49$ | 0.0 | 0.0 |
| | Branches | 1 | $51.94 \pm 0.66$ | $52.02 \pm 0.67$ | $214.86 \pm 27.94$ | $229.61 \pm 22.65$ | 0.0 | 0.0 |
| | STreeD | 1 | $77.11 \pm 0.89$ | $77.11 \pm 0.89$ | $3599.63 \pm 0.18$ | $3599.64 \pm 0.43$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $79.97 \pm 0.66$ | $85.08 \pm 0.47$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $76.22 \pm 0.38$ | $76.24 \pm 0.38$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $72.46 \pm 0.44$ | $74.52 \pm 1.87$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $81.61 \pm 0.60$ | $49.71 \pm 0.52$ | $1856.06 \pm 42.10$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $81.69 \pm 0.57$ | $88.63 \pm 0.21$ | $347.23 \pm 8.93$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $82.64 \pm 0.70$ | $89.57 \pm 0.66$ | $1353.60 \pm 175.17$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $\mathbf{82.78 \pm 0.69}$ | $\mathbf{90.04 \pm 0.48}$ | $1947.47 \pm 323.11$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT-L | 1 | $80.82 \pm 0.77$ | $86.95 \pm 0.30$ | $37.21 \pm 0.71$ | $182.75 \pm 3.02$ | 100.0 | 100.0 |
| | Literati-L | 1 | $80.81 \pm 0.62$ | $86.94 \pm 0.38$ | $\mathbf{8.06 \pm 0.90}$ | $\mathbf{29.08 \pm 4.05}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{81.33 \pm 0.74}$ | $87.66 \pm 0.51$ | $13.06 \pm 0.82$ | $51.08 \pm 4.55$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{81.33 \pm 0.74}$ | $\mathbf{87.75 \pm 0.49}$ | $13.59 \pm 2.05$ | $51.48 \pm 8.45$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $81.15 \pm 0.34$ | $88.03 \pm 0.55$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| gas-drift | CART | 1 | $81.10 \pm 0.44$ | $86.74 \pm 0.54$ | $\mathbf{0.48 \pm 0.01}$ | $\mathbf{0.58 \pm 0.05}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $86.21 \pm 1.41$ | $90.97 \pm 0.82$ | $71.72 \pm 19.33$ | $89.20 \pm 7.29$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $83.60 \pm 0.44$ | $88.22 \pm 0.66$ | $5.56 \pm 0.34$ | $11.54 \pm 4.78$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $88.05 \pm 1.44$ | $92.31 \pm 0.63$ | $55.34 \pm 32.77$ | $79.12 \pm 48.13$ | 0.0 | 0.0 |
| | Branches | 1 | $42.73 \pm 0.16$ | $42.87 \pm 0.21$ | $154.02 \pm 10.86$ | $146.57 \pm 5.81$ | 0.0 | 0.0 |
| | STreeD | 1 | $73.60 \pm 0.27$ | $73.60 \pm 0.27$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $90.25 \pm 0.23$ | $94.02 \pm 0.17$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $69.06 \pm 0.14$ | $69.10 \pm 0.20$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $65.48 \pm 0.98$ | $57.75 \pm 2.65$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $38.95 \pm 0.12$ | $38.95 \pm 0.12$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $92.34 \pm 0.36$ | $\mathbf{97.05 \pm 0.26}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 2 | $\mathbf{92.56 \pm 0.49}$ | $\mathbf{97.05 \pm 0.30}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $92.54 \pm 0.53$ | $\mathbf{97.05 \pm 0.30}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $91.26 \pm 0.36$ | $\mathbf{96.85 \pm 0.23}$ | $826.71 \pm 18.28$ | $3090.66 \pm 74.23$ | 100.0 | 100.0 |
| | Literati-L | 1 | $91.26 \pm 0.36$ | $96.83 \pm 0.25$ | $\mathbf{443.99 \pm 32.17}$ | $\mathbf{1762.02 \pm 130.74}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{91.28 \pm 0.37}$ | $96.84 \pm 0.25$ | $500.36 \pm 16.89$ | $1916.86 \pm 62.08$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{91.28 \pm 0.37}$ | $96.84 \pm 0.25$ | $504.35 \pm 17.54$ | $1960.49 \pm 102.92$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $91.73 \pm 0.23$ | $97.18 \pm 0.45$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |

**Table 7 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s ± std.) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ |
| htru | CART | 1 | $98.28 \pm 0.03$ | $98.42 \pm 0.05$ | $\mathbf{0.04 \pm 0.00}$ | $\mathbf{0.05 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $98.33 \pm 0.05$ | $98.47 \pm 0.05$ | $1.95 \pm 0.47$ | $4.53 \pm 1.63$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $98.51 \pm 0.06$ | $98.68 \pm 0.06$ | $1.16 \pm 0.78$ | $1.08 \pm 0.52$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $98.59 \pm 0.04$ | $98.72 \pm 0.05$ | $16.92 \pm 15.32$ | $12.91 \pm 15.01$ | 0.0 | 0.0 |
| | Branches | 1 | $97.64 \pm 0.05$ | $97.41 \pm 0.31$ | $112.01 \pm 23.86$ | $107.80 \pm 7.20$ | 0.0 | 0.0 |
| | STreeD | 1 | $98.07 \pm 0.04$ | $98.33 \pm 0.03$ | $518.06 \pm 57.71$ | $3599.45 \pm 0.34$ | 100.0 | 0.0 |
| | LDS-DL8.5 | 1 | $98.02 \pm 0.05$ | $98.19 \pm 0.07$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $98.07 \pm 0.04$ | $98.32 \pm 0.03$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $98.43 \pm 0.03$ | $98.52 \pm 0.24$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $97.76 \pm 0.05$ | $97.76 \pm 0.05$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $98.56 \pm 0.04$ | $98.86 \pm 0.04$ | $1107.77 \pm 49.05$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $98.70 \pm 0.05$ | $99.00 \pm 0.08$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{98.77 \pm 0.04}$ | $\mathbf{99.07 \pm 0.10}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $98.51 \pm 0.04$ | $98.80 \pm 0.04$ | $103.29 \pm 4.24$ | $484.63 \pm 57.95$ | 100.0 | 100.0 |
| | Literati-L | 1 | $98.51 \pm 0.04$ | $98.79 \pm 0.04$ | $\mathbf{26.48 \pm 1.33}$ | $\mathbf{109.88 \pm 7.62}$ | 100.0 | 100.0 |
| | | 2 | $98.62 \pm 0.04$ | $98.96 \pm 0.05$ | $170.19 \pm 22.62$ | $832.25 \pm 98.08$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.65 \pm 0.05}$ | $\mathbf{98.98 \pm 0.04}$ | $283.01 \pm 62.38$ | $1345.54 \pm 264.28$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $98.07 \pm 0.04$ | $98.33 \pm 0.03$ | $460.36 \pm 63.56$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| magic | CART | 1 | $81.40 \pm 0.38$ | $83.18 \pm 0.75$ | $\mathbf{0.04 \pm 0.00}$ | $\mathbf{0.05 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $82.89 \pm 0.43$ | $84.11 \pm 0.71$ | $5.76 \pm 1.23$ | $6.27 \pm 1.66$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $82.10 \pm 0.92$ | $84.60 \pm 0.42$ | $1.55 \pm 1.13$ | $2.22 \pm 1.91$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $83.31 \pm 0.40$ | $85.42 \pm 0.41$ | $13.90 \pm 12.34$ | $15.62 \pm 4.24$ | 0.0 | 0.0 |
| | Branches | 1 | $74.35 \pm 0.23$ | $74.38 \pm 0.25$ | $123.10 \pm 4.70$ | $137.29 \pm 8.60$ | 0.0 | 0.0 |
| | STreeD | 1 | $83.46 \pm 0.31$ | $84.91 \pm 0.60$ | $2124.74 \pm 177.97$ | $3599.74 \pm 0.30$ | 100.0 | 0.0 |
| | LDS-DL8.5 | 1 | $83.10 \pm 0.39$ | $84.59 \pm 0.35$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $83.44 \pm 0.33$ | $79.11 \pm 0.84$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $79.79 \pm 1.74$ | $79.33 \pm 0.88$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $72.84 \pm 0.19$ | $72.84 \pm 0.19$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $84.43 \pm 0.29$ | $86.41 \pm 0.24$ | $1235.00 \pm 9.83$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $\mathbf{84.73 \pm 0.24}$ | $87.04 \pm 0.27$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{84.73 \pm 0.35}$ | $\mathbf{87.14 \pm 0.33}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $84.21 \pm 0.30$ | $86.08 \pm 0.22$ | $110.06 \pm 2.82$ | $524.25 \pm 6.57$ | 100.0 | 100.0 |
| | Literati-L | 1 | $84.22 \pm 0.30$ | $86.08 \pm 0.22$ | $\mathbf{31.79 \pm 0.93}$ | $\mathbf{133.02 \pm 4.06}$ | 100.0 | 100.0 |
| | | 2 | $84.44 \pm 0.31$ | $86.48 \pm 0.25$ | $129.51 \pm 5.84$ | $583.04 \pm 38.31$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{84.46 \pm 0.32}$ | $\mathbf{86.53 \pm 0.23}$ | $164.23 \pm 6.96$ | $786.54 \pm 66.41$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $83.46 \pm 0.31$ | $85.30 \pm 0.25$ | $1764.00 \pm 191.96$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| mini-boone | CART | 1 | $89.32 \pm 0.27$ | $90.29 \pm 0.17$ | $\mathbf{2.09 \pm 0.01}$ | $\mathbf{2.48 \pm 0.01}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $89.90 \pm 0.26$ | $90.70 \pm 0.07$ | $105.16 \pm 22.30$ | $119.96 \pm 38.37$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $89.50 \pm 0.12$ | $90.53 \pm 0.10$ | $23.01 \pm 19.74$ | $30.59 \pm 27.06$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $89.85 \pm 0.11$ | $90.88 \pm 0.09$ | $37.72 \pm 8.50$ | $61.94 \pm 20.62$ | 0.0 | 0.0 |
| | Branches | 1 | $83.56 \pm 0.14$ | $83.57 \pm 0.15$ | $70.50 \pm 21.46$ | $62.66 \pm 4.79$ | 0.0 | 0.0 |
| | STreeD | 1 | $89.10 \pm 0.09$ | $89.10 \pm 0.12$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $89.82 \pm 0.05$ | $90.52 \pm 0.07$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $83.27 \pm 0.06$ | $83.27 \pm 0.06$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $81.82 \pm 0.14$ | $80.92 \pm 0.13$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $83.28 \pm 0.08$ | $83.28 \pm 0.08$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $90.31 \pm 0.09$ | $\mathbf{91.23 \pm 0.24}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 2 | $\mathbf{90.35 \pm 0.10}$ | $\mathbf{91.23 \pm 0.25}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{90.35 \pm 0.10}$ | $\mathbf{91.23 \pm 0.23}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $90.34 \pm 0.07$ | $83.28 \pm 0.08$ | $3555.90 \pm 26.97$ | $\mathbf{3600.00 \pm 0.00}$ | 100.0 | 0.0 |
| | Literati-L | 1 | $90.34 \pm 0.07$ | $91.24 \pm 0.05$ | $\mathbf{1570.92 \pm 79.85}$ | $\mathbf{3600.00 \pm 0.00}$ | 100.0 | 0.0 |
| | | 2 | $\mathbf{90.41 \pm 0.05}$ | $\mathbf{91.30 \pm 0.06}$ | $3600.00 \pm 0.00$ | $\mathbf{3600.00 \pm 0.00}$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{90.41 \pm 0.05}$ | $91.26 \pm 0.11$ | $3600.00 \pm 0.00$ | $\mathbf{3600.00 \pm 0.00}$ | 0.0 | 0.0 |
| | Literati-PreQuant | 1 | $89.98 \pm 0.06$ | $90.92 \pm 0.08$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| mushroom | CART | 1 | $99.69 \pm 0.02$ | $99.70 \pm 0.04$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.69 \pm 0.02$ | $99.70 \pm 0.04$ | $1.60 \pm 0.10$ | $1.46 \pm 0.40$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $99.69 \pm 0.02$ | $99.70 \pm 0.04$ | $0.48 \pm 0.01$ | $0.54 \pm 0.02$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.69 \pm 0.02$ | $99.70 \pm 0.04$ | $1.54 \pm 0.06$ | $6.07 \pm 9.82$ | 0.0 | 0.0 |
| | Branches | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $18.69 \pm 20.92$ | $8.90 \pm 4.90$ | 100.0 | 100.0 |
| | STreeD | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.05 \pm 0.00$ | $0.05 \pm 0.00$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.05 \pm 0.00$ | $0.05 \pm 0.00$ | 100.0 | 100.0 |
| | CADL8.5 | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.09 \pm 0.01$ | $1.98 \pm 0.42$ | 100.0 | 100.0 |
| | ConTree | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.07 \pm 0.00$ | $0.07 \pm 0.01$ | 100.0 | 100.0 |
| | DPDT | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $72.14 \pm 13.21$ | $354.14 \pm 59.77$ | 100.0 | 100.0 |
| | Literati | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.31 \pm 0.03$ | $0.36 \pm 0.11$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.33 \pm 0.02$ | $0.36 \pm 0.03$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.38 \pm 0.11$ | $0.34 \pm 0.06$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $12.62 \pm 0.67$ | $35.67 \pm 3.83$ | 100.0 | 100.0 |
| | Literati-L | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.34 \pm 0.03$ | $0.34 \pm 0.05$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{0.31 \pm 0.03}$ | $0.40 \pm 0.13$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $0.33 \pm 0.04$ | $\mathbf{0.33 \pm 0.02}$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $100.00 \pm 0.00$ | $100.00 \pm 0.00$ | $0.36 \pm 0.06$ | $0.48 \pm 0.03$ | 100.0 | 100.0 |

**Table 7 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s ± std.) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ |
| occupancy | CART | 1 | $99.06 \pm 0.04$ | $99.20 \pm 0.06$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.02 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.07 \pm 0.05$ | $99.22 \pm 0.06$ | $0.75 \pm 0.69$ | $1.51 \pm 0.71$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $99.40 \pm 0.09$ | $99.51 \pm 0.12$ | $0.55 \pm 0.20$ | $1.71 \pm 1.25$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.43 \pm 0.08$ | $99.53 \pm 0.10$ | $8.06 \pm 8.63$ | $4.17 \pm 0.86$ | 0.0 | 0.0 |
| | Branches | 1 | $98.86 \pm 0.09$ | $98.86 \pm 0.05$ | $104.12 \pm 8.42$ | $111.87 \pm 27.54$ | 0.0 | 0.0 |
| | STreeD | 1 | $99.14 \pm 0.04$ | $99.29 \pm 0.03$ | $48.59 \pm 3.14$ | $400.41 \pm 8.77$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $99.14 \pm 0.04$ | $99.23 \pm 0.06$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $99.14 \pm 0.04$ | $99.28 \pm 0.03$ | $320.78 \pm 19.63$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $99.39 \pm 0.20$ | $99.64 \pm 0.07$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $99.52 \pm 0.03$ | $98.94 \pm 0.05$ | $2587.46 \pm 131.84$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $99.52 \pm 0.03$ | $99.77 \pm 0.02$ | $450.90 \pm 45.60$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $99.63 \pm 0.02$ | $\mathbf{99.84 \pm 0.02}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{99.67 \pm 0.03}$ | $\mathbf{99.84 \pm 0.08}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $99.48 \pm 0.03$ | $99.70 \pm 0.02$ | $58.40 \pm 0.71$ | $246.52 \pm 11.94$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.48 \pm 0.03$ | $99.70 \pm 0.01$ | $\mathbf{13.74 \pm 0.75}$ | $\mathbf{45.85 \pm 3.05}$ | 100.0 | 100.0 |
| | | 2 | $99.58 \pm 0.03$ | $99.80 \pm 0.02$ | $108.43 \pm 23.15$ | $411.63 \pm 70.03$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.60 \pm 0.03}$ | $\mathbf{99.82 \pm 0.02}$ | $171.87 \pm 40.61$ | $644.73 \pm 163.89$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $99.14 \pm 0.04$ | $99.29 \pm 0.03$ | $37.81 \pm 1.91$ | $356.10 \pm 22.72$ | 100.0 | 100.0 |
| page | CART | 1 | $97.50 \pm 0.24$ | $97.95 \pm 0.25$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $97.67 \pm 0.22$ | $98.04 \pm 0.26$ | $1.40 \pm 0.51$ | $1.47 \pm 0.27$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $97.91 \pm 0.22$ | $98.40 \pm 0.23$ | $2.01 \pm 1.62$ | $0.90 \pm 0.25$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $98.12 \pm 0.21$ | $98.68 \pm 0.12$ | $5.62 \pm 1.63$ | $27.30 \pm 17.73$ | 0.0 | 0.0 |
| | Branches | 1 | $94.67 \pm 0.06$ | $94.64 \pm 0.09$ | $209.00 \pm 11.15$ | $212.80 \pm 14.49$ | 0.0 | 0.0 |
| | STreeD | 1 | $98.21 \pm 0.11$ | $98.89 \pm 0.11$ | $1181.46 \pm 56.33$ | $3599.62 \pm 0.33$ | 100.0 | 0.0 |
| | LDS-DL8.5 | 1 | $98.08 \pm 0.10$ | $98.77 \pm 0.15$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $98.17 \pm 0.10$ | $96.94 \pm 0.10$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $97.68 \pm 0.51$ | $96.69 \pm 0.69$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $98.62 \pm 0.08$ | $93.17 \pm 0.16$ | $1657.04 \pm 71.62$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $98.62 \pm 0.09$ | $99.32 \pm 0.07$ | $307.96 \pm 28.83$ | $3343.90 \pm 230.98$ | 100.0 | 80.0 |
| | | 2 | $98.76 \pm 0.08$ | $99.44 \pm 0.03$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{98.80 \pm 0.06}$ | $\mathbf{99.46 \pm 0.06}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $98.54 \pm 0.10$ | $99.16 \pm 0.07$ | $39.54 \pm 1.48$ | $173.07 \pm 6.07$ | 100.0 | 100.0 |
| | Literati-L | 1 | $98.54 \pm 0.09$ | $99.16 \pm 0.06$ | $\mathbf{8.77 \pm 0.33}$ | $\mathbf{27.73 \pm 2.86}$ | 100.0 | 100.0 |
| | | 2 | $98.65 \pm 0.10$ | $99.33 \pm 0.07$ | $34.38 \pm 10.76$ | $126.62 \pm 24.53$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.67 \pm 0.08}$ | $\mathbf{99.35 \pm 0.07}$ | $43.63 \pm 16.82$ | $167.29 \pm 57.26$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $98.21 \pm 0.11$ | $99.06 \pm 0.13$ | $584.21 \pm 83.57$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| pendigits | CART | 1 | $82.85 \pm 0.27$ | $88.27 \pm 0.15$ | $\mathbf{0.02 \pm 0.00}$ | $\mathbf{0.02 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $84.77 \pm 0.39$ | $90.50 \pm 0.47$ | $3.41 \pm 0.91$ | $6.44 \pm 2.99$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $82.40 \pm 0.33$ | $88.12 \pm 0.27$ | $2.29 \pm 0.89$ | $2.83 \pm 0.83$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $84.76 \pm 0.50$ | $90.99 \pm 0.37$ | $24.99 \pm 21.69$ | $43.90 \pm 33.22$ | 0.0 | 0.0 |
| | Branches | 1 | $37.42 \pm 1.51$ | $37.00 \pm 1.13$ | $146.59 \pm 4.12$ | $147.95 \pm 11.59$ | 0.0 | 0.0 |
| | STreeD | 1 | $91.11 \pm 0.20$ | $91.11 \pm 0.20$ | $3599.74 \pm 0.21$ | $3599.59 \pm 0.26$ | 0.0 | 0.0 |
| | LDS-DL8.5 | 1 | $91.43 \pm 0.18$ | $95.63 \pm 0.26$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $90.67 \pm 0.29$ | $51.72 \pm 2.51$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $88.02 \pm 5.89$ | $79.10 \pm 1.42$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $20.24 \pm 0.54$ | $20.24 \pm 0.54$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $91.59 \pm 0.33$ | $\mathbf{96.15 \pm 0.14}$ | $1061.35 \pm 31.53$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $\mathbf{91.63 \pm 0.36}$ | $96.14 \pm 0.13$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $91.61 \pm 0.34$ | $96.14 \pm 0.13$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $91.08 \pm 0.19$ | $95.75 \pm 0.17$ | $92.34 \pm 3.05$ | $452.73 \pm 14.02$ | 100.0 | 100.0 |
| | Literati-L | 1 | $91.09 \pm 0.18$ | $95.74 \pm 0.17$ | $\mathbf{28.50 \pm 1.92}$ | $\mathbf{118.73 \pm 7.38}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{91.11 \pm 0.18}$ | $\mathbf{95.82 \pm 0.21}$ | $53.77 \pm 7.57$ | $221.15 \pm 17.55$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{91.11 \pm 0.18}$ | $\mathbf{95.82 \pm 0.21}$ | $51.67 \pm 7.89$ | $233.49 \pm 46.52$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $91.60 \pm 0.15$ | $95.89 \pm 0.39$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| raisin | CART | 1 | $91.90 \pm 0.59$ | $93.71 \pm 0.37$ | $\mathbf{0.00 \pm 0.00}$ | $\mathbf{0.00 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $92.29 \pm 0.52$ | $93.81 \pm 0.45$ | $0.86 \pm 0.23$ | $0.75 \pm 0.46$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $94.92 \pm 0.75$ | $96.13 \pm 0.72$ | $1.27 \pm 1.13$ | $0.92 \pm 0.92$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $95.33 \pm 0.80$ | $96.38 \pm 0.85$ | $5.32 \pm 1.23$ | $3.95 \pm 0.80$ | 0.0 | 0.0 |
| | Branches | 1 | $90.03 \pm 0.41$ | $90.06 \pm 1.01$ | $394.75 \pm 91.07$ | $361.59 \pm 19.48$ | 0.0 | 0.0 |
| | STreeD | 1 | $94.03 \pm 0.21$ | $97.56 \pm 0.24$ | $17.42 \pm 1.20$ | $132.43 \pm 9.22$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $93.62 \pm 0.34$ | $86.63 \pm 0.63$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $94.03 \pm 0.21$ | $97.43 \pm 0.31$ | $592.07 \pm 58.43$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $95.65 \pm 1.17$ | $\mathbf{100.00 \pm 0.00}$ | $304.54 \pm 53.24$ | $20.45 \pm 18.12$ | 100.0 | 100.0 |
| | DPDT | 1 | $95.71 \pm 0.16$ | $86.63 \pm 0.63$ | $524.64 \pm 26.73$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $95.68 \pm 0.13$ | $99.17 \pm 0.21$ | $36.19 \pm 1.57$ | $366.46 \pm 40.00$ | 100.0 | 100.0 |
| | | 2 | $97.56 \pm 0.18$ | $99.94 \pm 0.09$ | $1361.21 \pm 321.54$ | $1755.44 \pm 1701.05$ | 100.0 | 60.0 |
| | | 3 | $\mathbf{98.54 \pm 0.21}$ | $99.90 \pm 0.21$ | $3600.00 \pm 0.00$ | $1319.54 \pm 1283.09$ | 0.0 | 80.0 |
| | DPDT-L | 1 | $95.21 \pm 0.26$ | $98.16 \pm 0.27$ | $16.45 \pm 1.47$ | $74.25 \pm 9.62$ | 100.0 | 100.0 |
| | Literati-L | 1 | $95.21 \pm 0.34$ | $98.16 \pm 0.14$ | $\mathbf{1.18 \pm 0.05}$ | $\mathbf{3.25 \pm 0.37}$ | 100.0 | 100.0 |
| | | 2 | $96.57 \pm 0.37$ | $99.49 \pm 0.21$ | $5.76 \pm 1.34$ | $20.34 \pm 5.03$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{97.08 \pm 0.37}$ | $\mathbf{99.78 \pm 0.18}$ | $8.88 \pm 3.75$ | $28.73 \pm 12.43$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $94.03 \pm 0.21$ | $97.56 \pm 0.24$ | $25.48 \pm 1.37$ | $334.45 \pm 19.04$ | 100.0 | 100.0 |

**Table 7 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s ± std.) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ |
| rice | CART | 1 | $94.05 \pm 0.21$ | $94.52 \pm 0.22$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $94.17 \pm 0.18$ | $94.68 \pm 0.23$ | $0.96 \pm 0.41$ | $1.59 \pm 0.38$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $95.49 \pm 0.37$ | $96.29 \pm 0.23$ | $1.19 \pm 0.98$ | $0.89 \pm 0.34$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $95.73 \pm 0.30$ | $96.48 \pm 0.21$ | $6.64 \pm 2.00$ | $8.09 \pm 2.64$ | 0.0 | 0.0 |
| | Branches | 1 | $93.68 \pm 0.25$ | $93.69 \pm 0.23$ | $248.94 \pm 25.49$ | $236.81 \pm 13.01$ | 0.0 | 0.0 |
| | STreeD | 1 | $94.77 \pm 0.15$ | $95.89 \pm 0.15$ | $65.07 \pm 3.29$ | $899.23 \pm 43.55$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $94.65 \pm 0.16$ | $95.49 \pm 0.10$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $94.77 \pm 0.15$ | $95.83 \pm 0.15$ | $1263.87 \pm 229.75$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $95.53 \pm 0.26$ | $95.86 \pm 1.11$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT | 1 | $95.60 \pm 0.26$ | $93.09 \pm 0.26$ | $1446.62 \pm 54.64$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $95.58 \pm 0.21$ | $97.04 \pm 0.18$ | $181.77 \pm 16.45$ | $2391.14 \pm 310.24$ | 100.0 | 100.0 |
| | | 2 | $96.39 \pm 0.22$ | $97.79 \pm 0.51$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $\mathbf{96.80 \pm 0.30}$ | $\mathbf{98.32 \pm 0.48}$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $95.34 \pm 0.24$ | $96.56 \pm 0.29$ | $32.43 \pm 1.56$ | $155.29 \pm 8.91$ | 100.0 | 100.0 |
| | Literati-L | 1 | $95.32 \pm 0.26$ | $96.54 \pm 0.27$ | $\mathbf{4.13 \pm 0.54}$ | $\mathbf{12.93 \pm 3.66}$ | 100.0 | 100.0 |
| | | 2 | $95.92 \pm 0.28$ | $97.35 \pm 0.21$ | $38.40 \pm 5.15$ | $205.07 \pm 39.12$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{96.14 \pm 0.35}$ | $\mathbf{97.59 \pm 0.30}$ | $101.11 \pm 26.31$ | $465.94 \pm 136.32$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $94.77 \pm 0.15$ | $95.89 \pm 0.15$ | $73.27 \pm 4.68$ | $1221.82 \pm 102.57$ | 100.0 | 100.0 |
| room | CART | 1 | $99.49 \pm 0.04$ | $99.62 \pm 0.06$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.57 \pm 0.04$ | $99.69 \pm 0.09$ | $1.25 \pm 0.40$ | $1.27 \pm 0.27$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $99.61 \pm 0.12$ | $99.86 \pm 0.05$ | $2.15 \pm 2.21$ | $1.35 \pm 1.47$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.72 \pm 0.07$ | $99.85 \pm 0.12$ | $18.74 \pm 9.30$ | $18.94 \pm 14.33$ | 0.0 | 0.0 |
| | Branches | 1 | $97.79 \pm 0.78$ | $97.87 \pm 0.64$ | $176.54 \pm 14.90$ | $181.78 \pm 13.93$ | 0.0 | 0.0 |
| | STreeD | 1 | $99.89 \pm 0.04$ | $99.98 \pm 0.02$ | $239.49 \pm 5.72$ | $2098.40 \pm 262.28$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $99.86 \pm 0.07$ | $99.93 \pm 0.04$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $99.89 \pm 0.04$ | $99.91 \pm 0.03$ | $2354.10 \pm 241.32$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $3.61 \pm 2.46$ | $0.42 \pm 0.56$ | 100.0 | 100.0 |
| | DPDT | 1 | $99.98 \pm 0.02$ | $88.62 \pm 0.00$ | $681.19 \pm 58.11$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $99.98 \pm 0.01$ | $\mathbf{100.00 \pm 0.00}$ | $106.13 \pm 55.95$ | $0.59 \pm 0.75$ | 100.0 | 100.0 |
| | | 2 | $99.99 \pm 0.01$ | $\mathbf{100.00 \pm 0.00}$ | $430.38 \pm 496.87$ | $0.30 \pm 0.26$ | 100.0 | 100.0 |
| | | 3 | $99.99 \pm 0.01$ | $\mathbf{100.00 \pm 0.00}$ | $534.69 \pm 643.44$ | $0.32 \pm 0.27$ | 100.0 | 100.0 |
| | DPDT-L | 1 | $99.97 \pm 0.03$ | $\mathbf{100.00 \pm 0.00}$ | $29.53 \pm 1.52$ | $102.64 \pm 10.87$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.97 \pm 0.03$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{5.91 \pm 3.07}$ | $\mathbf{0.43 \pm 0.51}$ | 100.0 | 100.0 |
| | | 2 | $\mathbf{99.98 \pm 0.01}$ | $\mathbf{100.00 \pm 0.00}$ | $11.64 \pm 6.36$ | $1.04 \pm 1.85$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.98 \pm 0.01}$ | $\mathbf{100.00 \pm 0.00}$ | $12.20 \pm 6.54$ | $1.11 \pm 2.03$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $99.89 \pm 0.04$ | $99.98 \pm 0.02$ | $285.07 \pm 15.33$ | $2650.37 \pm 1019.45$ | 100.0 | 60.0 |
| segment | CART | 1 | $83.13 \pm 0.50$ | $91.37 \pm 0.36$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $84.72 \pm 0.90$ | $92.50 \pm 0.65$ | $0.46 \pm 0.11$ | $0.70 \pm 0.11$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $82.42 \pm 0.30$ | $91.53 \pm 0.50$ | $0.63 \pm 0.25$ | $1.22 \pm 1.20$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $84.22 \pm 0.93$ | $93.62 \pm 0.89$ | $2.76 \pm 0.65$ | $9.38 \pm 5.75$ | 0.0 | 0.0 |
| | Branches | 1 | $87.10 \pm 1.04$ | $87.26 \pm 3.61$ | $276.55 \pm 18.77$ | $301.41 \pm 55.45$ | 0.0 | 0.0 |
| | STreeD | 1 | $97.72 \pm 0.15$ | $83.87 \pm 30.91$ | $2936.98 \pm 127.74$ | $3599.54 \pm 0.37$ | 100.0 | 0.0 |
| | LDS-DL8.5 | 1 | $97.18 \pm 0.27$ | $98.71 \pm 0.22$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $97.25 \pm 0.24$ | $98.10 \pm 0.12$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | ConTree | 1 | $93.05 \pm 4.11$ | $\mathbf{99.72 \pm 0.22}$ | $1915.42 \pm 442.89$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $98.21 \pm 0.12$ | $28.57 \pm 0.00$ | $699.94 \pm 12.05$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $98.14 \pm 0.09$ | $99.63 \pm 0.10$ | $122.28 \pm 6.07$ | $1131.65 \pm 72.35$ | 100.0 | 100.0 |
| | | 2 | $98.55 \pm 0.11$ | $99.51 \pm 0.21$ | $2762.48 \pm 416.38$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $\mathbf{98.59 \pm 0.05}$ | $99.58 \pm 0.15$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $97.88 \pm 0.18$ | $99.28 \pm 0.19$ | $20.89 \pm 1.00$ | $83.20 \pm 2.41$ | 100.0 | 100.0 |
| | Literati-L | 1 | $97.86 \pm 0.19$ | $99.26 \pm 0.22$ | $\mathbf{4.32 \pm 0.22}$ | $\mathbf{9.64 \pm 0.93}$ | 100.0 | 100.0 |
| | | 2 | $98.32 \pm 0.09$ | $99.62 \pm 0.17$ | $14.38 \pm 1.72$ | $44.10 \pm 6.36$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{98.33 \pm 0.08}$ | $\mathbf{99.68 \pm 0.18}$ | $20.28 \pm 4.08$ | $68.05 \pm 16.28$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $97.72 \pm 0.15$ | $99.46 \pm 0.15$ | $2834.81 \pm 123.77$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| skin | CART | 1 | $98.49 \pm 0.02$ | $99.00 \pm 0.01$ | $\mathbf{0.09 \pm 0.01}$ | $\mathbf{0.09 \pm 0.01}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.38 \pm 0.03$ | $99.69 \pm 0.02$ | $10.54 \pm 1.98$ | $14.54 \pm 2.25$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $98.90 \pm 0.08$ | $99.25 \pm 0.18$ | $2.30 \pm 0.22$ | $2.97 \pm 0.48$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.67 \pm 0.05$ | $99.77 \pm 0.06$ | $49.56 \pm 12.14$ | $74.77 \pm 13.03$ | 0.0 | 0.0 |
| | Branches | 1 | $97.60 \pm 0.38$ | $97.58 \pm 0.61$ | $50.94 \pm 5.62$ | $49.39 \pm 5.48$ | 0.0 | 0.0 |
| | STreeD | 1 | $99.08 \pm 0.02$ | $99.25 \pm 0.02$ | $156.68 \pm 11.49$ | $930.82 \pm 42.42$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $99.08 \pm 0.02$ | $99.23 \pm 0.04$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $99.08 \pm 0.02$ | $99.25 \pm 0.02$ | $373.67 \pm 12.83$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $99.71 \pm 0.04$ | $98.61 \pm 0.61$ | $1144.45 \pm 154.16$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | DPDT | 1 | $88.23 \pm 0.06$ | $88.23 \pm 0.06$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | Literati | 1 | $99.73 \pm 0.01$ | $\mathbf{99.87 \pm 0.03}$ | $2245.96 \pm 83.00$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 2 | $\mathbf{99.75 \pm 0.02}$ | $99.80 \pm 0.12$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | | 3 | $99.72 \pm 0.01$ | $99.50 \pm 0.82$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | DPDT-L | 1 | $99.64 \pm 0.04$ | $99.86 \pm 0.01$ | $331.00 \pm 15.23$ | $1238.05 \pm 81.39$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.60 \pm 0.01$ | $99.86 \pm 0.01$ | $\mathbf{76.89 \pm 2.84}$ | $\mathbf{239.18 \pm 11.71}$ | 100.0 | 100.0 |
| | | 2 | $99.82 \pm 0.01$ | $\mathbf{99.92 \pm 0.00}$ | $942.92 \pm 34.04$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | | 3 | $\mathbf{99.83 \pm 0.00}$ | $\mathbf{99.92 \pm 0.00}$ | $2053.03 \pm 105.39$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati-PreQuant | 1 | $99.08 \pm 0.02$ | $99.25 \pm 0.02$ | $108.54 \pm 7.74$ | $680.27 \pm 71.46$ | 100.0 | 100.0 |

**Table 7 continued from previous page**

| | | | ↑ Train Acc. (%) | | ↓ Runtime (s ± std.) | | ↑ Proof Obtained (%) | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Model | K | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ | $D = 5$ | $D = 6$ |
| wilt | CART | 1 | $99.06 \pm 0.10$ | $99.44 \pm 0.11$ | $\mathbf{0.01 \pm 0.00}$ | $\mathbf{0.01 \pm 0.00}$ | 0.0 | 0.0 |
| | AxTAO | 1 | $99.19 \pm 0.16$ | $99.50 \pm 0.07$ | $1.41 \pm 0.66$ | $1.43 \pm 0.49$ | 0.0 | 0.0 |
| | ShapeCART | 3 | $98.90 \pm 0.25$ | $99.12 \pm 0.31$ | $0.74 \pm 0.29$ | $0.97 \pm 0.72$ | 0.0 | 0.0 |
| | ShapeTAO | 3 | $99.06 \pm 0.23$ | $99.29 \pm 0.28$ | $3.73 \pm 1.11$ | $9.82 \pm 10.78$ | 0.0 | 0.0 |
| | Branches | 1 | $98.38 \pm 0.09$ | $98.37 \pm 0.11$ | $204.56 \pm 17.85$ | $200.73 \pm 12.57$ | 0.0 | 0.0 |
| | STreeD | 1 | $99.08 \pm 0.06$ | $99.58 \pm 0.05$ | $18.46 \pm 0.78$ | $199.03 \pm 18.79$ | 100.0 | 100.0 |
| | LDS-DL8.5 | 1 | $99.08 \pm 0.06$ | $99.43 \pm 0.12$ | $3600.00 \pm 0.00$ | $3600.00 \pm 0.00$ | 0.0 | 0.0 |
| | CADL8.5 | 1 | $99.08 \pm 0.06$ | $99.55 \pm 0.08$ | $309.01 \pm 16.84$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | ConTree | 1 | $99.70 \pm 0.05$ | $\mathbf{100.00 \pm 0.00}$ | $53.55 \pm 24.04$ | $0.99 \pm 0.67$ | 100.0 | 100.0 |
| | DPDT | 1 | $99.53 \pm 0.01$ | $94.60 \pm 0.00$ | $764.34 \pm 41.68$ | $3600.00 \pm 0.00$ | 100.0 | 0.0 |
| | Literati | 1 | $99.53 \pm 0.02$ | $99.94 \pm 0.02$ | $102.79 \pm 3.95$ | $857.49 \pm 38.62$ | 100.0 | 100.0 |
| | | 2 | $99.79 \pm 0.03$ | $99.99 \pm 0.01$ | $3600.00 \pm 0.00$ | $1072.62 \pm 1470.26$ | 0.0 | 80.0 |
| | | 3 | $\mathbf{99.88 \pm 0.02}$ | $\mathbf{100.00 \pm 0.00}$ | $3600.00 \pm 0.00$ | $649.84 \pm 665.19$ | 0.0 | 100.0 |
| | DPDT-L | 1 | $99.49 \pm 0.04$ | $99.89 \pm 0.01$ | $26.09 \pm 2.26$ | $97.67 \pm 4.53$ | 100.0 | 100.0 |
| | Literati-L | 1 | $99.49 \pm 0.04$ | $99.90 \pm 0.02$ | $\mathbf{3.79 \pm 0.31}$ | $8.87 \pm 0.47$ | 100.0 | 100.0 |
| | | 2 | $99.70 \pm 0.06$ | $\mathbf{100.00 \pm 0.00}$ | $35.65 \pm 5.97$ | $12.29 \pm 19.35$ | 100.0 | 100.0 |
| | | 3 | $\mathbf{99.78 \pm 0.06}$ | $\mathbf{100.00 \pm 0.00}$ | $80.66 \pm 28.11$ | $\mathbf{1.03 \pm 1.32}$ | 100.0 | 100.0 |
| | Literati-PreQuant | 1 | $99.08 \pm 0.06$ | $99.58 \pm 0.05$ | $17.45 \pm 1.30$ | $189.04 \pm 13.39$ | 100.0 | 100.0 |

# H Per Dataset Generalization Results

Table 8: Test Accuracy (%) results for our generalization experiments. Highest average test accuracy across folds per dataset is **bolded**.

| Dataset | CART | AxTAO | ShapeCART | ShapeTAO | Branches | STreeD | LDS-DL8.5 | CADL8.5 | ConTree | DPDT | Literati |
|---|---|---|---|---|---|---|---|---|---|---|---|
| adult | $85.40 \pm 0.20$ | $85.90 \pm 0.30$ | $87.10 \pm 0.30$ | $87.30 \pm 0.30$ | $83.17 \pm 0.29$ | $86.60 \pm 0.30$ | $86.90 \pm 0.40$ | $86.10 \pm 0.20$ | $84.70 \pm 0.30$ | $86.00 \pm 0.30$ | $\mathbf{87.40 \pm 0.20}$ |
| avila | $68.10 \pm 0.90$ | $69.90 \pm 0.40$ | $71.50 \pm 1.70$ | $75.10 \pm 2.30$ | $63.48 \pm 0.78$ | $73.00 \pm 0.70$ | $74.20 \pm 0.50$ | $72.50 \pm 0.60$ | $63.90 \pm 0.80$ | $73.70 \pm 0.70$ | $\mathbf{79.40 \pm 1.10}$ |
| bank | $97.90 \pm 0.50$ | $98.80 \pm 0.60$ | $97.50 \pm 1.00$ | $98.40 \pm 0.60$ | $\mathbf{99.27 \pm 0.51}$ | $99.20 \pm 0.70$ | $98.90 \pm 0.40$ | $99.20 \pm 0.50$ | $96.20 \pm 2.50$ | $98.30 \pm 0.30$ | $98.70 \pm 0.30$ |
| bean | $91.80 \pm 0.10$ | $92.50 \pm 0.20$ | $93.50 \pm 0.20$ | $93.70 \pm 0.30$ | $85.41 \pm 0.56$ | $92.10 \pm 0.40$ | $93.40 \pm 0.20$ | $92.20 \pm 0.40$ | $90.40 \pm 0.70$ | $93.00 \pm 0.20$ | $\mathbf{94.00 \pm 0.40}$ |
| bidding | $99.50 \pm 0.30$ | $99.50 \pm 0.30$ | $99.70 \pm 0.10$ | $99.70 \pm 0.10$ | $99.49 \pm 0.24$ | $99.50 \pm 0.20$ | $99.70 \pm 0.20$ | $99.70 \pm 0.10$ | $99.00 \pm 0.90$ | $99.70 \pm 0.10$ | $\mathbf{99.80 \pm 0.10}$ |
| electricity | $80.10 \pm 0.60$ | $82.50 \pm 0.50$ | $85.70 \pm 0.80$ | $86.10 \pm 0.50$ | $78.56 \pm 0.23$ | $82.60 \pm 0.40$ | $83.20 \pm 0.30$ | $82.30 \pm 0.30$ | $81.40 \pm 0.30$ | $83.80 \pm 0.30$ | $\mathbf{86.90 \pm 0.20}$ |
| eucalyptus | $57.10 \pm 3.70$ | $57.80 \pm 4.10$ | $58.70 \pm 3.10$ | $\mathbf{60.50 \pm 4.60}$ | $43.51 \pm 2.32$ | $53.50 \pm 4.80$ | $56.40 \pm 1.90$ | $59.80 \pm 4.20$ | $48.30 \pm 5.20$ | $60.00 \pm 3.20$ | $57.60 \pm 3.40$ |
| eye-movements | $55.90 \pm 0.70$ | $57.30 \pm 1.30$ | $58.50 \pm 0.90$ | $59.30 \pm 1.80$ | $52.29 \pm 1.16$ | $59.50 \pm 0.80$ | $59.70 \pm 0.90$ | $59.10 \pm 0.50$ | $54.50 \pm 1.20$ | $59.70 \pm 0.70$ | $\mathbf{63.50 \pm 1.00}$ |
| eye-state | $75.00 \pm 0.50$ | $77.50 \pm 1.20$ | $79.40 \pm 1.10$ | $79.90 \pm 1.10$ | $72.71 \pm 1.26$ | $\mathbf{81.10 \pm 0.70}$ | $81.00 \pm 0.60$ | $79.80 \pm 0.80$ | $74.60 \pm 1.40$ | $79.00 \pm 0.70$ | $\mathbf{81.10 \pm 1.70}$ |
| fault | $69.70 \pm 1.10$ | $72.80 \pm 1.50$ | $73.60 \pm 2.90$ | $74.50 \pm 2.80$ | $68.53 \pm 2.88$ | $74.50 \pm 2.70$ | $\mathbf{77.00 \pm 1.70}$ | $73.70 \pm 2.40$ | $66.90 \pm 3.70$ | $71.60 \pm 1.20$ | $76.50 \pm 2.80$ |
| gas-drift | $89.90 \pm 0.90$ | $92.50 \pm 1.40$ | $93.70 \pm 1.60$ | $95.10 \pm 0.90$ | $58.05 \pm 0.40$ | $84.60 \pm 0.70$ | $96.30 \pm 0.70$ | $83.30 \pm 0.90$ | $74.20 \pm 1.10$ | $96.00 \pm 0.60$ | $\mathbf{99.00 \pm 0.60}$ |
| htru | $\mathbf{97.90 \pm 0.10}$ | $\mathbf{97.90 \pm 0.10}$ | $97.80 \pm 0.20$ | $\mathbf{97.90 \pm 0.20}$ | $97.81 \pm 0.18$ | $97.80 \pm 0.10$ | $97.80 \pm 0.20$ | $97.80 \pm 0.10$ | $97.80 \pm 0.10$ | $\mathbf{97.90 \pm 0.10}$ | $97.80 \pm 0.10$ |
| magic | $82.40 \pm 0.80$ | $83.10 \pm 1.00$ | $84.20 \pm 1.00$ | $84.40 \pm 0.90$ | $79.80 \pm 0.75$ | $84.10 \pm 0.90$ | $84.90 \pm 1.10$ | $82.60 \pm 1.10$ | $81.80 \pm 1.10$ | $84.20 \pm 1.00$ | $\mathbf{85.40 \pm 1.10}$ |
| mini-boone | $91.80 \pm 0.20$ | $92.40 \pm 0.10$ | $93.50 \pm 0.20$ | $93.30 \pm 0.20$ | $87.03 \pm 0.20$ | $91.10 \pm 0.20$ | $93.10 \pm 0.10$ | $90.80 \pm 0.10$ | $88.90 \pm 0.20$ | $92.70 \pm 0.10$ | $\mathbf{93.80 \pm 0.20}$ |
| mushroom | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.10}$ | $\mathbf{100.00 \pm 0.00}$ | $98.92 \pm 0.09$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $\mathbf{100.00 \pm 0.00}$ | $99.60 \pm 0.60$ | $\mathbf{100.00 \pm 0.10}$ | $\mathbf{100.00 \pm 0.00}$ |
| occupancy | $98.80 \pm 0.10$ | $98.90 \pm 0.10$ | $99.00 \pm 0.10$ | $98.90 \pm 0.10$ | $98.83 \pm 0.10$ | $98.80 \pm 0.10$ | $98.70 \pm 0.00$ | $98.80 \pm 0.10$ | $98.90 \pm 0.10$ | $99.00 \pm 0.10$ | $\mathbf{99.10 \pm 0.10}$ |
| page | $96.60 \pm 0.50$ | $\mathbf{96.90 \pm 0.40}$ | $96.30 \pm 0.50$ | $96.40 \pm 0.50$ | $95.87 \pm 0.59$ | $96.10 \pm 0.30$ | $96.70 \pm 0.40$ | $96.70 \pm 0.50$ | $96.00 \pm 0.90$ | $96.70 \pm 0.40$ | $96.80 \pm 0.40$ |
| pendigits | $92.20 \pm 0.70$ | $93.90 \pm 0.60$ | $93.80 \pm 0.80$ | $94.90 \pm 1.10$ | $81.06 \pm 0.77$ | $95.60 \pm 0.50$ | $97.20 \pm 0.50$ | $94.90 \pm 0.50$ | $85.00 \pm 0.60$ | $95.60 \pm 0.60$ | $\mathbf{97.70 \pm 0.60}$ |
| raisin | $88.50 \pm 2.10$ | $88.30 \pm 0.70$ | $\mathbf{89.40 \pm 2.00}$ | $88.10 \pm 1.00$ | $85.22 \pm 0.84$ | $87.70 \pm 1.30$ | $87.80 \pm 1.40$ | $88.40 \pm 2.10$ | $87.30 \pm 2.00$ | $89.10 \pm 2.60$ | $89.30 \pm 2.40$ |
| rice | $94.30 \pm 0.60$ | $94.40 \pm 0.70$ | $\mathbf{96.00 \pm 0.80}$ | $95.40 \pm 1.20$ | $92.68 \pm 0.80$ | $94.90 \pm 0.80$ | $94.80 \pm 1.10$ | $94.60 \pm 0.80$ | $93.70 \pm 0.80$ | $94.60 \pm 0.80$ | $\mathbf{96.00 \pm 0.80}$ |
| room | $99.50 \pm 0.10$ | $99.60 \pm 0.10$ | $99.50 \pm 0.10$ | $99.60 \pm 0.10$ | $98.70 \pm 0.17$ | $99.60 \pm 0.00$ | $99.50 \pm 0.10$ | $99.50 \pm 0.10$ | $99.30 \pm 0.50$ | $99.60 \pm 0.20$ | $\mathbf{99.70 \pm 0.10}$ |
| segment | $93.60 \pm 1.60$ | $94.40 \pm 1.40$ | $93.90 \pm 1.20$ | $94.30 \pm 1.30$ | $90.48 \pm 1.00$ | $94.70 \pm 0.90$ | $95.80 \pm 1.00$ | $95.80 \pm 0.70$ | $92.90 \pm 3.20$ | $96.00 \pm 1.40$ | $\mathbf{96.80 \pm 0.80}$ |
| skin | $98.90 \pm 0.10$ | $99.50 \pm 0.00$ | $99.10 \pm 0.10$ | $99.70 \pm 0.10$ | $98.59 \pm 0.05$ | $99.30 \pm 0.00$ | $99.30 \pm 0.00$ | $99.30 \pm 0.00$ | $99.70 \pm 0.00$ | $99.70 \pm 0.00$ | $\mathbf{99.90 \pm 0.00}$ |
| wilt | $98.30 \pm 0.50$ | $98.30 \pm 0.30$ | $98.10 \pm 0.30$ | $98.20 \pm 0.20$ | $98.16 \pm 0.42$ | $98.10 \pm 0.20$ | $\mathbf{98.50 \pm 0.30}$ | $98.30 \pm 0.50$ | $97.30 \pm 0.70$ | $98.30 \pm 0.30$ | $98.40 \pm 0.20$ |

Table 9: Runtime (s) of the selected model for our generalization experiments. Lowest average across-folds is **bolded**.

| Dataset | CART | AxTAO | ShapeCART | ShapeTAO | Branches | STreeD | LDS-DL8.5 | CADL8.5 | ConTree | DPDT | Literati |
|---|---|---|---|---|---|---|---|---|---|---|---|
| adult | $\mathbf{0.12 \pm 0.01}$ | $21.68 \pm 6.44$ | $2.37 \pm 1.26$ | $47.51 \pm 5.84$ | $47.96 \pm 4.25$ | $899.68 \pm 0.31$ | $901.72 \pm 0.04$ | $900.15 \pm 0.04$ | $900.09 \pm 0.01$ | $12.73 \pm 1.23$ | $838.64 \pm 58.03$ |
| avila | $\mathbf{0.06 \pm 0.01}$ | $2.88 \pm 0.29$ | $2.50 \pm 1.73$ | $13.90 \pm 5.06$ | $143.81 \pm 20.48$ | $899.41 \pm 0.30$ | $901.66 \pm 0.03$ | $900.11 \pm 0.04$ | $900.02 \pm 0.03$ | $30.62 \pm 1.29$ | $39.33 \pm 4.47$ |
| bank | $\mathbf{0.00 \pm 0.00}$ | $0.62 \pm 0.14$ | $0.67 \pm 0.39$ | $15.46 \pm 2.09$ | $55.37 \pm 6.80$ | $0.04 \pm 0.00$ | $0.08 \pm 0.04$ | $0.01 \pm 0.00$ | $0.02 \pm 0.00$ | $0.43 \pm 0.02$ | $0.26 \pm 0.03$ |
| bean | $\mathbf{0.10 \pm 0.00}$ | $4.53 \pm 1.25$ | $1.97 \pm 0.42$ | $14.47 \pm 3.34$ | $74.16 \pm 2.99$ | $899.55 \pm 0.28$ | $903.94 \pm 0.06$ | $900.06 \pm 0.00$ | $900.01 \pm 0.01$ | $40.25 \pm 0.67$ | $105.04 \pm 7.68$ |
| bidding | $\mathbf{0.00 \pm 0.00}$ | $0.42 \pm 0.17$ | $0.89 \pm 0.68$ | $17.89 \pm 3.16$ | $48.17 \pm 4.32$ | $9.43 \pm 1.48$ | $903.50 \pm 0.35$ | $142.83 \pm 10.97$ | $103.92 \pm 63.81$ | $1.59 \pm 0.32$ | $12.30 \pm 1.65$ |
| electricity | $\mathbf{0.08 \pm 0.00}$ | $10.78 \pm 6.30$ | $1.56 \pm 0.30$ | $83.11 \pm 3.26$ | $26.78 \pm 3.97$ | $899.65 \pm 0.32$ | $902.86 \pm 0.29$ | $900.13 \pm 0.01$ | $900.03 \pm 0.02$ | $22.51 \pm 1.16$ | $320.99 \pm 51.27$ |
| eucalyptus | $\mathbf{0.02 \pm 0.00}$ | $10.24 \pm 0.28$ | $1.27 \pm 0.36$ | $4.08 \pm 0.51$ | $133.85 \pm 9.25$ | $899.57 \pm 0.42$ | $901.86 \pm 0.12$ | $900.31 \pm 0.10$ | $900.04 \pm 0.02$ | $0.87 \pm 0.05$ | $5.10 \pm 0.26$ |
| eye-movements | $\mathbf{0.04 \pm 0.00}$ | $5.71 \pm 0.95$ | $2.11 \pm 0.33$ | $59.75 \pm 10.71$ | $116.58 \pm 13.00$ | $899.60 \pm 0.34$ | $905.50 \pm 1.42$ | $900.07 \pm 0.01$ | $900.02 \pm 0.02$ | $29.78 \pm 1.71$ | $231.22 \pm 28.95$ |
| eye-state | $\mathbf{0.04 \pm 0.00}$ | $8.78 \pm 5.29$ | $2.23 \pm 1.74$ | $93.67 \pm 37.02$ | $42.28 \pm 1.06$ | $899.45 \pm 0.38$ | $904.34 \pm 0.77$ | $900.22 \pm 0.05$ | $900.02 \pm 0.02$ | $21.80 \pm 1.20$ | $41.03 \pm 9.58$ |
| fault | $\mathbf{0.02 \pm 0.00}$ | $1.78 \pm 0.41$ | $3.97 \pm 4.78$ | $31.90 \pm 12.55$ | $426.34 \pm 46.49$ | $899.70 \pm 0.17$ | $907.39 \pm 0.38$ | $900.06 \pm 0.03$ | $24.05 \pm 3.21$ | $8.57 \pm 0.51$ | $3.50 \pm 0.38$ |
| gas-drift | $\mathbf{0.88 \pm 0.03}$ | $48.25 \pm 8.70$ | $10.71 \pm 1.04$ | $123.43 \pm 81.11$ | $179.61 \pm 7.67$ | $901.78 \pm 0.14$ | $902.50 \pm 0.11$ | $900.36 \pm 0.12$ | $900.02 \pm 0.01$ | $288.21 \pm 2.06$ | $810.08 \pm 43.74$ |
| htru | $\mathbf{0.05 \pm 0.00}$ | $1.40 \pm 0.42$ | $1.10 \pm 0.56$ | $2.74 \pm 0.91$ | $900.00 \pm 0.00$ | $161.84 \pm 13.11$ | $903.49 \pm 0.46$ | $900.10 \pm 0.02$ | $70.08 \pm 32.57$ | $12.23 \pm 0.41$ | $172.15 \pm 51.95$ |
| magic | $\mathbf{0.05 \pm 0.00}$ | $5.24 \pm 1.48$ | $1.66 \pm 0.28$ | $60.95 \pm 4.75$ | $76.43 \pm 3.72$ | $899.47 \pm 0.22$ | $905.26 \pm 0.74$ | $900.50 \pm 0.03$ | $900.02 \pm 0.03$ | $11.49 \pm 1.54$ | $100.29 \pm 9.58$ |
| mini-boone | $\mathbf{2.51 \pm 0.04}$ | $119.82 \pm 38.29$ | $14.00 \pm 5.87$ | $161.83 \pm 18.41$ | $76.93 \pm 4.48$ | $901.17 \pm 0.43$ | $906.98 \pm 0.60$ | $901.05 \pm 0.24$ | $900.04 \pm 0.02$ | $855.88 \pm 22.08$ | $901.36 \pm 0.22$ |
| mushroom | $\mathbf{0.01 \pm 0.00}$ | $0.40 \pm 0.10$ | $0.87 \pm 0.52$ | $1.50 \pm 0.23$ | $12.38 \pm 0.45$ | $0.05 \pm 0.00$ | $0.08 \pm 0.02$ | $0.02 \pm 0.00$ | $900.02 \pm 0.01$ | $0.18 \pm 0.01$ | $4.20 \pm 1.23$ |
| occupancy | $\mathbf{0.01 \pm 0.00}$ | $0.78 \pm 0.35$ | $1.28 \pm 0.76$ | $21.91 \pm 3.23$ | $900.00 \pm 0.00$ | $28.61 \pm 2.41$ | $901.98 \pm 0.10$ | $145.09 \pm 7.66$ | $673.46 \pm 373.26$ | $6.50 \pm 0.28$ | $88.37 \pm 15.35$ |
| page | $\mathbf{0.01 \pm 0.00}$ | $1.25 \pm 0.46$ | $1.01 \pm 0.82$ | $27.12 \pm 4.90$ | $69.47 \pm 7.27$ | $882.00 \pm 17.98$ | $906.30 \pm 0.25$ | $900.13 \pm 0.04$ | $2.52 \pm 0.60$ | $5.96 \pm 0.19$ | $10.56 \pm 0.72$ |
| pendigits | $\mathbf{0.02 \pm 0.00}$ | $6.39 \pm 1.72$ | $1.70 \pm 0.30$ | $84.35 \pm 13.79$ | $65.24 \pm 9.26$ | $899.50 \pm 0.30$ | $905.39 \pm 0.60$ | $900.42 \pm 0.04$ | $742.05 \pm 100.46$ | $21.27 \pm 1.15$ | $97.81 \pm 5.28$ |
| raisin | $\mathbf{0.00 \pm 0.00}$ | $0.33 \pm 0.11$ | $0.50 \pm 0.13$ | $10.50 \pm 6.14$ | $1.59 \pm 0.17$ | $0.01 \pm 0.00$ | $4.66 \pm 0.54$ | $0.03 \pm 0.00$ | $0.31 \pm 0.22$ | $0.72 \pm 0.06$ | $0.46 \pm 0.15$ |
| rice | $\mathbf{0.01 \pm 0.00}$ | $0.72 \pm 0.44$ | $1.47 \pm 0.88$ | $2.91 \pm 1.49$ | $2.89 \pm 0.17$ | $0.02 \pm 0.00$ | $78.81 \pm 4.25$ | $0.07 \pm 0.00$ | $0.02 \pm 0.01$ | $1.18 \pm 0.03$ | $14.64 \pm 2.64$ |
| room | $\mathbf{0.01 \pm 0.00}$ | $1.05 \pm 0.32$ | $0.94 \pm 0.16$ | $25.45 \pm 4.33$ | $127.34 \pm 18.92$ | $162.68 \pm 12.03$ | $903.04 \pm 0.30$ | $900.37 \pm 0.08$ | $900.02 \pm 0.01$ | $5.49 \pm 0.62$ | $12.17 \pm 3.63$ |
| segment | $\mathbf{0.01 \pm 0.00}$ | $0.97 \pm 0.34$ | $1.23 \pm 0.42$ | $9.27 \pm 6.66$ | $380.31 \pm 43.75$ | $899.62 \pm 0.32$ | $909.96 \pm 0.63$ | $900.08 \pm 0.03$ | $900.02 \pm 0.02$ | $2.80 \pm 0.12$ | $15.31 \pm 1.09$ |
| skin | $\mathbf{0.08 \pm 0.01}$ | $13.52 \pm 2.07$ | $2.87 \pm 2.64$ | $41.44 \pm 22.96$ | $17.03 \pm 0.50$ | $584.25 \pm 39.38$ | $902.06 \pm 0.04$ | $900.19 \pm 0.02$ | $899.55 \pm 3.68$ | $57.78 \pm 4.44$ | $530.38 \pm 27.13$ |
| wilt | $\mathbf{0.01 \pm 0.00}$ | $1.61 \pm 0.29$ | $1.30 \pm 0.77$ | $45.72 \pm 26.21$ | $2.83 \pm 0.12$ | $4.88 \pm 0.28$ | $65.86 \pm 2.85$ | $23.30 \pm 0.43$ | $0.27 \pm 0.09$ | $5.21 \pm 0.26$ | $55.29 \pm 20.59$ |

# I Interpretability

SGTs inherit many structural qualities of decision trees that make decision trees interpretable [10]. As a tree variant, SGTs are modular: like standard decision trees, they decompose into a hierarchical set of rules whose subtrees can be interpreted independently. Because each node's shape function operates on only one feature, it can be directly visualized and queried, which preserves simulability, the ability of a user to trace and reproduce the model's root-to-leaf decision process by hand (See Appendix I.1 below). Finally, SGTs improve sparsity: the added expressiveness of each node compresses repeated splits into more compact trees, and sparsity is what most directly aids simulability, since a smaller model gives a practitioner a feasible chance of reconstructing its predictions.

## I.1 Visualized Trees

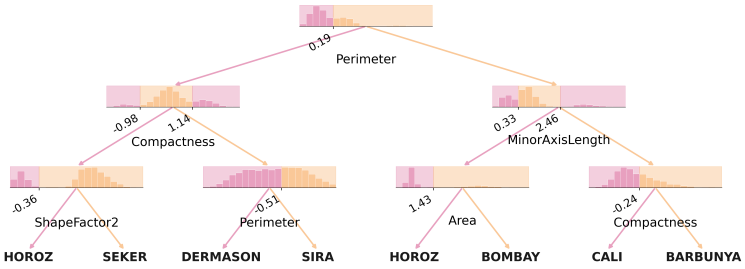



Figure I.10: $D = 3, K = 2$ SGT for Bean

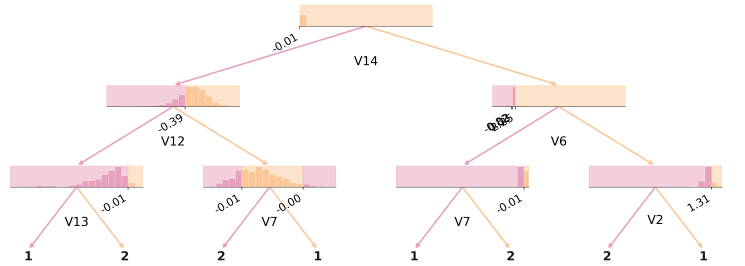



Figure I.11: $D = 3, K = 2$ SGT for Eye-State

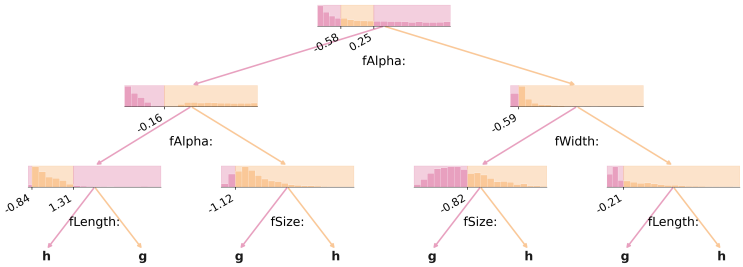



Figure I.12: $D = 3, K = 2$ SGT for Magic

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We describe all main claims in our abstract and provide a list of contributions in our introduction.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Conclusion.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: See Appendix for proofs of all theoretical results along with relevant assumptions.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide detailed instructions on our experimental configuration, allowing others to reproduce our experiments if needed. Additionally, we provide detailed pseudocode on how to implement our algorithm.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We utilize open source datasets in our experiments and provide code to replicate our experiments.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: See Experimental Evaluation section as well as Appendix.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We provide statistical significance results along with std. deviation across folds.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

Justification: See experimental evaluation section as well as appendix.

Guidelines:

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We agree and confirm our work follows the NeurIPS code of ethics.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See Conclusion. The goal of our work is to provide an improved decision tree algorithm that addresses many of the interpretability concerns with binary axis-aligned trees. As discussed in the conclusion, increased interpretability is important in high-stakes, sensitive fields, and improving interpretability has a positive societal impact.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Our approach has minor risk of misuse and we only consider open source datasets.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Original producers of relevant code are cited.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: No assets introduced.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: No human subjects used.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No study participants.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not a core part of our work.