# GLANCE: GLOBAL ACTIONS IN A NUTSHELL FOR COUNTERFACTUAL EXPLAINABILITY

Anonymous authors

Paper under double-blind review

### Abstract

The widespread deployment of machine learning systems in critical real-world decision-making applications has highlighted the urgent need for counterfactual explainability methods that operate effectively. Global counterfactual explanations, expressed as actions to offer recourse, aim to provide succinct explanations and insights applicable to large population subgroups. Effectiveness is measured by the fraction of the population that is provided recourse, ensuring that the actions benefit as many individuals as possible. Keeping the cost of actions low ensures the proposed recourse actions remain practical and actionable. Limiting the number of actions that provide global counterfactuals is essential to maximize interpretability. The primary challenge, therefore, is balancing these tradeoffs—maximizing effectiveness, minimizing cost, while maintaining a small number of actions. We introduce GLANCE, a versatile and adaptive framework, comprising two algorithms, that allows the careful balancing of the trade-offs among the three key objectives, with the size objective functioning as a tunable parameter to keep the actions few and easy to interpret. C-GLANCE employs a clustering approach that considers both the feature space and the space of counterfactual actions, thereby accounting for the distribution of points in a way that aligns with the structure of the model. T-GLANCE provides additional features to enhance flexibility. It employs a tree-based approach, that allows users to specify split features, to build a decision tree with a single counterfactual action at each node that can be used as a subgroup policy. Our extensive experimental evaluation demonstrates that our method consistently shows greater robustness and performance compared to existing methods across various datasets and models.

033

004

010 011

012

013

014

015

016

017

018

019

021

024

025

026

027

028

029

031

034 035

## 1 INTRODUCTION

037

Machine learning models are increasingly deployed in critical domains such as loan approvals, hiring, and healthcare. This widespread adoption has intensified the need for transparency and interpretability in model decisions, requiring users to understand how their input features influence the outcomes and how they might change them to achieve favorable outcomes, known as recourse (Miller, 2019). Counterfactual explanations have gathered extensive attention for their suitability for achieving algorithmic recourse (Karimi et al., 2020), their interpretability (Wachter et al., 2017), actionability (Ustun et al., 2019), utility in fairness audits (Sharma et al., 2019; Kavouras et al., 2023), etc. A counterfactual action, or simply an *action*, defines the specific feature changes needed to convert an unfavorable decision into a favorable one.

Traditionally, counterfactual explanations refer to *local explainability*, being tied to a particular negatively affected instance. However, in many real-world scenarios, *global counterfactual explain-ability* is more useful, offering generalized explanations that apply across the affected population.
 While a collection of all local counterfactuals could technically cover all affected individuals, this approach sacrifices interpretability, which is central to global explainability.

 Similar to past work (Rawal & Lakkaraju, 2020; Kanamori et al., 2022), we define Global Counterfactual Explanations (GCE) as a small set of *global actions* that provide effective recourse for the affected population. Any global counterfactual solution must meet three objectives as identified from previous research: (1) be composed of a small number of actions to ensure interpretability (size), (2) minimize the cost of implementing those actions (cost), and (3) offer recourse to as many affected individuals as possible (effectiveness).

As noted by Branke (2008), the relationships 058 between multiple optimization objectives are often complex, and aggregating them into a 060 single objective, which is common in practice 061 (Rawal & Lakkaraju, 2020), can be problematic 062 since they are typically non-commensurable. 063 Framing GCEs as multi-objective optimization, 064 allows us to explore the inherent trade-offs between effectiveness and cost, especially when 065 the solution size is constrained. 066

067To understand this optimization problem, as-068sume a 2-d numerical feature space, and con-069sider the two affected instances  $x_1, x_2$  depicted070in Figure 1a. Assume that recourse cost equals071the distance to the decision boundary, drawn as072a line in the figure. Observe that  $a_1$  (resp.  $a_2$ ) is



Figure 1: A toy example depicting two negative instances  $x_1, x_2$ , and five actions:  $a_1, a_2$  are the optimal local actions for  $x_1, x_2$ ;  $a_3$  is the optimal GCE maximizing effectiveness;  $a_4$  and  $a_5$  are actions along the optimal direction d as defined in Ley et al. (2023) (a) The feature space, where the line is the decision boundary. (b) The action space, where  $l_1, l_2$  depict the decision boundary from the perspective of  $x_1, x_2$ , respectively.

the local action that provides recourse for  $x_1$  (resp.  $x_2$ ) at minimum cost.

Further, consider the *action space* depicted in Figure 1b, where every action is represented as a point (or equivalently a vector relative to the center o of the coordinate system). The blue  $l_1$  and red  $l_2$ lines represent the decision boundary seen from the perspectives of  $x_1$  and  $x_2$ , respectively. The blue line  $l_1$  separates the actions that provide recourse for  $x_1$  (any action on the outside, away from o) from those that do not. Action  $a_1$  lies on  $l_1$ , and is the closest point to o, and thus the min-cost local action for  $x_1$ . Similarly, the red line  $l_2$  concerns  $x_2$  and contains its min-cost action  $a_2$ .

Consider now the problem of finding a single action GCE. To provide recourse for both  $x_1$  and  $x_2$ , we look for an action that lies outside both lines in Figure 1b. Among all such actions,  $a_3$  has the minimum cost and thus is the optimal global action that maximizes effectiveness. If we trade off effectiveness for cost,  $a_2$  is the optimal global action that minimizes cost, but has 50% effectiveness (brings recourse to  $x_2$  but not to  $x_1$ ).

GCE gives rise to a different optimization problem than its local counterpart. Even if optimal local actions are generated for each instance and a small subset is chosen as the global actions, this may still result in a suboptimal GCE. In our example, the optimal global action  $a_3$  is not an optimal local action for either  $x_1$  or  $x_2$ ; in fact,  $a_3$  can be viewed as a compromise between  $a_1$  and  $a_2$ , the locally optimal actions. Finding good GCEs requires a nuanced exploration of the action space.

In this work, we distinguish between two variants of GCE that differ in how they assign actions to individuals. In *implicit* GCE, each individual chooses the best (i.e., min-cost) global action in GCE to achieve (if possible) recourse (as in e.g., Ley et al., 2023). Implicit GCE is helpful when one wishes to obtain a model-level understanding of recourse, or to select a global (horizontal) policy to apply to the entire affected population. In *explicit* GCE, individuals are partitioned according to their features and a single action applies to each partition (as in e.g., Rawal & Lakkaraju, 2020; Kanamori et al., 2022). Explicit GCE provides a more interpretable summary as it associates actions to partitions of the feature space, and can also be used to implement targeted policies to bring recourse to subpopulations.

099

## 100 1.1 KEY CONTRIBUTIONS

Recognizing that both the effectiveness and the cost of actions depend on the individuals' features and the model's decision boundary, we propose a novel clustering approach for GCEs that considers both feature and action spaces. Unlike previous claims that clustering is ineffective for global counterfactuals (see e.g., Kanamori et al. (2022)), we are the first to propose the joint clustering in feature and action spaces and demonstrate its effectiveness. We introduce GLANCE, a novel and highly adaptable framework, that consists of two algorithms based on hierarchical clustering techniques, C-GLANCE and T-GLANCE. GLANCE bridges local and global explanations by combining



Figure 2: Intuition behind clustering approaches. (a) First, C-GLANCE generates diverse candidate action from the centroids of feature-based clusters. (b) Then, C-GLANCE merges clusters based on similarity in either the feature space or the possible recourse actions, grouping instances that may be further apart but can be explained by similar actions. (c) An alternative where clusters are formed solely based on the feature space may lead to suboptimal global actions—either high effectiveness at a very high cost or low cost but very low effectiveness.

123

clustering with local counterfactual generation, offering a scalable and effective methodology for
 GCEs. GLANCE balances actionable, minimal changes with interpretability through a tunable size
 objective, ensuring practical and interpretable solutions.

127 C-GLANCE. We present C-GLANCE, a clustering-based approach for the implicit GCE problem
 128 that returns s global actions. At a high level, C-GLANCE consists of two phases: candidate action
 129 generation, and action selection.

To generate candidates, C-GLANCE first performs a fine-grained clustering of the individuals in the feature space. For each cluster, C-GLANCE then discovers a few diverse actions that bring recourse to the cluster centroid. The intuition here is that highly similar individuals are expected to have a similar perspective of the decision boundary, and can thus achieve recourse through similar cost-efficient actions. Figure 2a shows an example with four fine-grained clusters, where three diverse actions are discovered from each centroid. Each action represents an alternate way to bring recourse to cluster members, giving C-GLANCE the flexibility to later select among them.

137 At the end of the first phase, C-GLANCE has associated each cluster of individuals with a set of 138 actions. In the second phase, C-GLANCE iteratively merges such clusters until exactly s clusters 139 remain. Two clusters are merged if they either concern similar individuals or contain similar actions. 140 The intuition behind the former is similar to that of the candidate generation phase. The intuition 141 for the latter is that if similar actions bring recourse to them it should be easier to identify a common low-cost, highly effective action, even though they are dissimilar feature-wise. Figure 2b shows an 142 example where the red individuals are grouped together either through affinity in the feature or in 143 the action space. 144

As discussed in (Kanamori et al., 2022), performing a coarse-grained clustering on feature space and then assigning a single action per cluster leads to inadequate global actions, that suffer either in cost or effectiveness. For example, in Figure 2c, individuals are clustered based on feature similarity and a single action is selected per cluster. Because the blue cluster is associated with rather dissimilar actions, choosing a single actions leads to a compromise in either effectiveness (as depicted) or cost.

150 T-GLANCE. Recognizing the need for explicit GCE, we present T-GLANCE, an algorithm em-151 ploying a tree-based approach, similar to explainable clustering techniques (e.g., Moshkovitz et al., 152 2020). This allows assigning actions to individuals while satisfying the transparency and consistency properties required by Kanamori et al. (2022). T-GLANCE also allows users to select features 153 that define specific groups, to produce targeted GCEs. For the user's facilitation, we also provide 154 suggestions based on the importance feature permutation score for selecting features to be used 155 in splits. T-GLANCE directly supports policymaking and audit scenarios—an application not ad-156 dressed by existing methods. As shown in our experiments, T-GLANCE delivers competitive results, 157 outperforming previous tree-based methods while maintaining efficiency. 158

Experimental Evaluation. We perform a comprehensive evaluation across various datasets and
 models, demonstrating that GLANCE consistently outperforms or matches state-of-the-art methods
 in both cost and effectiveness while remaining computational efficiency. To facilitate comparison
 in global counterfactual methods, we invoke Pareto dominance criteria from multiobjective opti-

mization, further showcasing GLANCE's stable, superior performance across diverse datasets and models.

164 165 166

## 1.2 Related Work

167 Counterfactual Explanations. There has been a plethora of work with an emphasis on counter-168 factual explanations (Verma et al., 2020). An overview of methods on algorithmic recourse, which 169 provides explanations and recommendations to individuals impacted by automated decision-making 170 systems is presented by Karimi et al. (2020). These methods can be model-agnostic or model-171 specific (e.g., for trees Carreira-Perpiñán & Hada, 2021), focus on properties like diversity (Mothilal 172 et al., 2020) or feasibility (Ustun et al., 2019)), global or local. While local counterfactuals are well defined, global counterfactuals present challenges, as they must provide recourse for all individuals 173 within a specific group in the same manner while maintaining explainability and staying true to the 174 notion of local counterfactual, which focuses on minimal changes. Achieving this balance globally 175 presents significant difficulties. 176

177 Prior work does not make the distinction between explicit and implicit GCE. Regarding explicit GCE, Rawal & Lakkaraju (2020) introduced AReS, a framework for global counterfactual expla-178 nations, that jointly optimizes recourse correctness, coverage, and cost, providing an interpretable 179 summary of recourses, expressed in a two-level rule set. However, the AReS framework may fail to 180 cover the entire population. Ley et al. (2022) later improved the computational efficiency with Fast 181 AReS. In another direction, Kanamori et al. (2022) introduced CET, which partitions the space and 182 assigns an action to each part transparently and consistently. Although effective, its computational 183 complexity limits scalability. 184

Regarding implicit GCE, Warren et al. (2023) developed Group-CF, that generates counterfactuals 185 that seeks to maximize effectiveness, though it can result in higher costs. Ley et al. (2023) proposed GLOBE-CE, where global counterfactual explainability is defined differently, as a small set of action 187 directions along which individuals can "move" to achieve recourse. Therefore, GLOBE-CE attacks 188 a different optimization problem than the one we study here. For the toy example in Figure 1, the 189 single direction that minimizes the total recourse cost for  $x_1$  and  $x_2$  is the direction d depicted in 190 Figure 1b. This direction contains actions  $a_4$  and  $a_5$  that bring recourse to  $x_2$  and  $x_1$ , respectively, 191 with minimum cost along d. It is important to note that neither  $a_4$  nor  $a_5$  is optimal as a GCE (or 192 for local explainability). Even the set  $\{a_4, a_5\}$  is a suboptimal GCE—the set  $\{a_1, a_2\}$  dominates it 193 with equal size and effectiveness but lower total cost. In general, (1) directly translating the output 194 of GLOBE-CE into a GCE leads to numerous micro-actions (two in our example, but potentially as 195 many as the number of individuals), which reduces interpretability, (2) choosing a few actions along the optimal directions may result in suboptimal GCEs (in our example, any single action along d196 is dominated by a), and (3) since directions lack clear endpoints, they fail to specify the magnitude 197 of change required, leading to uncertainty for individuals seeking recourse and limiting real-world 198 applicability. 199

Other works include Carrizosa et al. (2024a;b), who use mixed-integer quadratic models for grouplevel explanations, and Koo et al. (2020), who employ Lagrangian methods. Research has also
extended to generating global counterfactuals for graphs (Huang et al., 2023) and auditing subgroup
fairness (Kavouras et al., 2023).

204 Global Explainability. Many works in global explainability are related to our approach. For in-205 stance, Chowdhury et al. (2022) summarize model logic by dividing the desired region of explana-206 tion features into subspaces based on similar logic, referred to as equi-explanation maps. Similarly, Lakkaraju et al. (2019) provides explanations for different subspaces characterized by specific fea-207 tures. Their method also allows users to customize the model explanations by selecting features 208 of interest, akin to the flexibility offered by T-GLANCE. While these works share a conceptual 209 focus on subspace-based explanations, they differ fundamentally from ours in their ultimate goals 210 and outcomes. Both primarily aim to summarize model behavior or provide faithful subspace-level 211 explanations, whereas GLANCE seeks to generate actionable global counterfactual explanations tai-212 lored to decision-making processes. Consequently, the direct comparison of these approaches with 213 T-GLANCE is not entirely applicable. 214

215 **Other Related Work.** Many clustering algorithms lead to clusters that are hard to explain. There is a plethora of work on explainable or interpretable clustering, mainly using decision trees that under

216 specific assumptions come with provable guarantees, including the works of Dasgupta et al. (2020); 217 Moshkovitz et al. (2020); Frost et al. (2020); Makarychev & Shan (2022); Bertsimas et al. (2021); 218 Laber & Murtinho (2021), etc. Another area of related work is multiobjective optimization and 219 Pareto optimality. We refer the reader to Branke (2008).

220 221 222

223

224

225

226

227

228

229

230

231 232 233

234

235

236

237 238

239

240 241 242

243 244

255

256

259

261

265 266 267

#### 2 PROBLEM FORMULATION

We consider a black box binary classifier  $h: \mathcal{X} \to \{-1, 1\}$ , where the positive outcome is favorable and the negative is unfavorable. We will focus on the set  $\mathcal{X}_{aff} \subseteq \mathcal{X}$  of adversely affected individuals, i.e., those who receive the unfavorable outcome. We denote as  $\mathbb{A}$  the set of all possible actions (which is potentially infinite), where an **action**  $a \in A$  is a set of changes to feature values, e.g.,  $a = \{$ country  $\rightarrow US,$  education-num  $\rightarrow +2\},$  which, when applied to an instance  $x \in \mathcal{X}_{aff}$ , results in a **counterfactual** instance x' = a(x). Every action a has a cost, denoted as cost(a, x), and is effective for an instance x if h(a(x)) = h(x') = 1. Let  $\mathbb{C} \subseteq \mathbb{A}$ . The recourse cost  $\mathsf{rc}(\mathbb{C}, x)$  of an instance x is the minimum cost incurred from an effective action in  $\mathbb{C}$ :

$$\operatorname{rc}(\mathbb{C}, x) = \min\{\operatorname{cost}(a, x) | a \in \mathbb{C} : h(a(x)) = 1\}$$

Let  $X_{\mathbb{C}} = \{x \in \mathcal{X}_{aff} | h(a(x)) = 1, a \in \mathbb{C}\}$  be the set of instances that flip their prediction using one of the actions in  $\mathbb{C}$ . Then the effectiveness, also known as coverage Ley et al. (2022), of  $\mathbb{C}$  for the affected instances  $\mathcal{X}_{aff}$  is defined as the percentage of  $\mathcal{X}_{aff}$  that managed to flip their prediction using one of the actions in  $\mathbb{C}$ :

$$\operatorname{eff}(\mathbb{C}, \mathcal{X}_{\operatorname{aff}}) = \frac{|X_{\mathbb{C}}|}{|\mathcal{X}_{\operatorname{aff}}|},$$

The cost of  $\mathbb{C}$  in  $\mathcal{X}_{aff}$  is defined as the **average recourse cost** of the instances in  $\mathcal{X}_{aff}$ :

$$\mathsf{avc}(\mathbb{C},\mathcal{X}_{\mathsf{aff}}) = \frac{\sum\limits_{x \in X_{\mathbb{C}}} \mathsf{rc}(\mathbb{C},x)}{|X_{\mathbb{C}}|}.$$

245 Finally, let size  $(\mathbb{C}) = |\mathbb{C}|$  denote the cardinality of a set  $\mathbb{C}$ .

246 As it is clear, multiple sets of actions can produce recourse for  $\mathcal{X}_{aff}$  and the quality of such a set is a 247 factor of the three notions we introduced: effectiveness, cost, and size. An ideal global counterfac-248 tual should maximize effectiveness while minimizing the cost and the size, based on the properties 249 that state-of-the-art works (Rawal & Lakkaraju, 2020; Kanamori et al., 2022; Ley et al., 2023; Huang 250 et al., 2023) have argued are essential. Formally, we define the global counterfactual explanation 251 problem as follows:

Problem 1 (Global Counterfactual Explanations (GCE)). Given a black box model h that classifies 253 the  $\mathcal{X}_{aff}$  instances to the negative class, our goal is to find the set  $\mathbb{C} \subseteq \mathbb{A}, \mathbb{C} \neq \emptyset$  that represents a 254 solution to the following multi-objective optimization:

 $\textit{minimize} \quad \left( \mathsf{size}(\mathbb{C}), \ -\mathsf{eff}(\mathbb{C}, \mathcal{X}_{\mathsf{aff}}), \ \mathsf{avc}(\mathbb{C}, \mathcal{X}_{\mathsf{aff}}) \right)$ 

257 The requirement for a small set of actions is to enhance the interpretability of the explanation. This 258 can also be expressed as a constraint to the set size we can afford. In this case, the third objective of Problem 1 is replaced by the constraint size  $(\mathbb{C}) \leq s$ , where s is a small positive integer. For the rest 260 of the paper, we will use the following problem formulation:

**Problem 2** (s-GCE). Given a black box model h that classifies the  $\mathcal{X}_{aff}$  instances to the negative 262 class, our goal is to find the set  $\mathbb{C} \subseteq \mathbb{A}$  that represents a solution to the following bi-objective 263 optimization problem: 264

$$\begin{array}{ll} \textit{minimize} & \left( -\operatorname{eff}(\mathbb{C}, \mathcal{X}_{\operatorname{aff}}), \operatorname{avc}(\mathbb{C}, \mathcal{X}_{\operatorname{aff}}) \right) \\ s.t. & \operatorname{size}(\mathbb{C}) \leq s \end{array}$$

If the set of actions  $\mathbb{A}$  is finite and explicitly given, s-CGE is NP-hard. For more details, we refer the 268 reader to Appendix A. For insights into how increasing the size impacts effectiveness and average 269 cost, and the importance of selecting actions to maintain a desirable balance, see Appendix B.

## Algorithm 1 C-GLANCE

-	
Inpu	t: $\mathcal{X}_{aff}$ , $m, k, s \{\mathcal{X}_{aff} := affected individuals, m := number of candidate actions to generate, k := initial number of clusters, s := number of clubal counterfactual actions}$
Outr	interest of global counterfactual actions
1:	$C \leftarrow \text{cluster}(\mathcal{X}_{\mathfrak{n}}, k)$ {Cluster $\mathcal{X}_{\mathfrak{n}}$ into k initial clusters}
2:	for $c \in C$ do
3:	$ca(c) \leftarrow actions(centroid(c), m)$ {For each cluster centroid generate m actions}
4:	end for
5:	while $ C  > s$ do
6:	$\{c_1, c_2\} \leftarrow \arg\min_{\{c_1, c_2\} \in C} (d_1(\operatorname{centroid}(c_1), \operatorname{centroid}(c_2)) + d_2(\operatorname{average}(\boldsymbol{ca}(c_1)), \operatorname{average}(\boldsymbol{ca}(c_2))))$
	$\{\operatorname{Find} c_1, c_2 \text{ that minimize } d_1 + d_2\}$
7:	$C \leftarrow \operatorname{merge}(\{c_1, c_2\} \in C) \{ \operatorname{Merge} c_1, c_2 \text{ into one cluster} \}$
8:	$ca(\{c_1, c_2\}) \leftarrow ca(c_1) \cup ca(c_2)$ {Merge their action sets}
9:	end while
10:	<b>return</b> $bca(ca(c))$ for each $c \in C$ {Return optimal action from each of the s clusters}

## 3 C-GLANCE

282

284

C-GLANCE (Algorithm 1) presents a clustering algorithm that follows an agglomerative approach. The first step is to establish multiple small clusters and determine representative diverse actions for each. The initial clusters will undergo merging, exploiting the notion of proximity in both the original feature space and the action space, until no more than *s* clusters exist. Finally, a single action will be extracted from each cluster with universal applicability across all instances. For the time complexity analysis we refer the reader to Appendix C.

Algorithm Description. (1) Initial clusters and actions generation. The feature space is partitioned into k clusters, by a clustering algorithm (e.g., k-means, line 1, alg. 1). We compute the *centroid* of each cluster and generate m diverse counterfactual actions for each centroid (line 3, alg. 1), employing any candidate counterfactual generation method (see Appendix G for the methods used in the experimental evaluation). We aim to efficiently and effectively explore the action space by generating actions from widely dispersed points within the feature space and guiding them in diverse directions that cross the decision boundary.

(2) Merging. Similar to agglomerative clustering, our approach merges clusters with the goal of 298 maximizing intra-cluster similarity and minimizing inter-cluster similarity with respect to both the 299 feature and action spaces. Specifically, similarity is computed by a metric D, which is the sum 300 of two distances:  $d_1: \mathcal{X}_{aff} \times \mathcal{X}_{aff} \to \mathbb{R}$ , the distance between the centroids of each cluster, and 301  $d_2: \mathbb{A} \times \mathbb{A} \to \mathbb{R}$ , the distance between the average counterfactual actions of each cluster; other 302 formulations for  $d_2$  are also possible, e.g., Wasserstein distance. Thus, clusters are combined based 303 on both their feature proximity and the proximity of their respective actions. Until the desired 304 number s of clusters is reached, we merge (line 7, alg. 1) the two clusters that minimize the total 305 distance  $d_1 + d_2$  (line 6, alg. 1). Merging two clusters means that their action sets are combined, 306 ensuring that the most effective actions for each cluster are retained (line 8, alg. 1).

307 (3) GCE extraction and optimality. For each of the s clusters, we extract a single candidate action, and return the set of these s actions (line 10, alg. 1). Assuming that the action generation and subsequent merging has succesfully grouped together individuals that can achieve recourse through similar cost-efficient actions, in the final step we target effectiveness, selecting at each cluster the optimal action, in terms of effectiveness, among those associated with the cluster.

Approach Strengths. C-GLANCE supports various clustering methods, cost metrics, and action generation techniques making it adaptable to different applications. In the App. Table 13, we present the results using various action generation methods, all of which lead to near-optimal solutions in terms of the effectiveness-cost tradeoff. Additionally, C-GLANCE demonstrates speed, robustness, and near-optimal solution quality across datasets, as shown in the experimental evaluation.

317 318

319

## 4 T-GLANCE

T-GLANCE (Algorithm 2) constructs a hierarchical partitionining of the feature space. Each partition level is derived by splitting one cell of the previous level partition along a feature, resulting in a decision tree like structure akin to classification and regression trees (CART) (Breiman et al., 1984; Bertsimas & Dunn, 2017). Each cell (whether an internal or leaf tree node) can be described as a conjunction of feature predicates, and is associated with a single counterfactual action, providing

## 324 Algorithm 2 T-GLANCE

Input:  $\mathcal{X}_{aff}$ , m, s, D'. { $\mathcal{X}_{aff}$  := affected individuals, m :=number of candidate actions to generate, s := maximum number of leaf nodes, 326 D', a subset of possible features } **Output:**  $\mathcal{P}$  {The tree and the actions assigned in each node.} 327  $\mathcal{P}_0 \leftarrow \{\mathcal{X}_{\mathsf{aff}}\}$  {The partition associated with the root of the tree.} 328 Assign to  $\mathcal{P}_0$  action  $a_0 = bca(ca(\mathcal{P}_0), \mathcal{P}_0)$  {Generate actions  $ca(\mathcal{P}_0)$ .  $a_0$  maximizes eq.(1)}  $\begin{array}{l} \text{level} \leftarrow 0, n_{nodes} \leftarrow 1, queue \leftarrow \mathcal{P}_0 \\ \text{while } n_{nodes} + 1 \leq s \text{ do} \\ \text{ if } queue = \emptyset \text{ then} \end{array}$ 3: 329 330 6:  $level \leftarrow level + 1, queue \leftarrow \mathcal{P}_{level}$ 331 7: end if 332 8: Pick A the first element in *queue*. 9:  $i \leftarrow \arg\max_{j \in D'} V(A, j) \ \{ \text{Choose the split that maximizes the criterion of eq.(2).} \}$ 333 10: if V(A, i) > 0 then 334 11: Split the cell A into cells  $A_k^i = \{x | x \in A \land x_i \in v_i(k)\}, k = 0, 1$ 335 12: Assign to  $A_k^i$  action  $a_k^i = bca(ca(A_k^i), A_k^i)$  { Generate actions  $ca(A_k^i)$ .  $a_k^i$  maximizes eq.(1). } 336 13:  $queue \leftarrow queue \setminus \{A\}, \mathcal{P}_{level+1} \leftarrow \mathcal{P}_{level+1} \cup \{A_0^i, A_1^i\}$ 14: else 337 15:  $queue \leftarrow queue \setminus \{A\}$  {Do not split this node and continue to the next.} 338 16: end if 17: end while 339 18:  $\mathcal{P}_{level+1} \leftarrow \mathcal{P}_{level+1} \cup queue$ 340 19: return  $\mathcal{P}_{level+1}$  { Return the tree and the assigned actions.} 341

342

thus *interpretable* counterfactual explanations at multiple granularity levels. T-GLANCE decides to
split a cell if more effective actions can be found for the child cells. Splits are performed along a
user-provided subset of features. Once a feature is used to split a cell, it is not considered again for
further splits. As a result, the height of the tree is limited by the cardinality of the user-provided
feature subset. We next present the necessary definitions to introduce Algorithm 2. For the time
complexity analysis, please refer to Appendix C.

**Definitions.** A partition  $\mathcal{P}$  of  $\mathbb{R}^d$  is a family of sets  $\{A_1, \ldots, A_l\}$  such that  $A_j \subseteq \mathbb{R}^d$ ,  $A_j \cap A_k = \emptyset$ 349 for all  $j, k \in [l]$  and  $\bigcup_{j=1}^{s} A_j = \mathbb{R}^d$ . Let  $\mathcal{P}$  a partition of  $\mathcal{X}_{aff}$ . Every element A of  $\mathcal{P}$  is called a 350 351 cell of  $\mathcal{P}$  or just a cell. Each cell can be further divided into two subcells  $A_0^i, A_1^i$  with respect to 352 any direction *i*, defined as  $A_0^i = \{x \in A | x_i \in v_i(0)\}$  and  $A_1^i = \{x \in A | x_i \in v_i(\hat{1})\}$ , where  $v_i(k)$ 353 represent the values that correspond to the specific split (e.g., 0 or 1 if it is a binary feature, specific 354 ranges for continuous features). Let ca(A) := actions(c, m) be the set of m generated actions 355 for the centroid c of the cell A. For simplicity, we will use bca(ca(A), A) to refer to the action in ca(A) with the highest effectiveness. Formally, 356

$$bca(\boldsymbol{ca}(A), A) = \underset{a \in \boldsymbol{ca}(A)}{\arg\max\{\mathsf{eff}(\{a\}, A)\}}$$
(1)

We define a local version of the *total effectiveness gain* achieved by a specific portion split i to a local cell A.

364 365

357

358 359

360

 $V(A,i) \triangleq \operatorname{eff}(\{\operatorname{bca}(\boldsymbol{ca}(A),A)\},A) - \overline{\operatorname{eff}}(\{\operatorname{bca}(\boldsymbol{ca}(A_k^i),A_k^i)\},A_k^i) \\ = \operatorname{eff}(\{\operatorname{bca}(\boldsymbol{ca}(A),A)\},A) - \frac{1}{|k|} \sum_{k=\{0,1\}} \operatorname{eff}(\{\operatorname{bca}(\boldsymbol{ca}(A_k^i),A_k^i)\},A_k^i)$ (2)

366 Algorithm Description. T-GLANCE selects a different direction for splitting at each iteration at 367 every cell in the current partition. Specifically, the algorithm begins with a single cluster containing 368 all affected instances and generates m diverse actions for the centroid, retaining the best action according to the criterion in eq. (1) (lines 1–2, alg. 2). The root then is established as the first cell 369 A to be examined, and the possible splits are examined based on the actions that were generated for 370 each split (lines 8–9, alg. 2). For simplicity, we present the algorithm with binary splits (branching 371 factor of 2), though it can be generalized to accommodate any branching factor. We proceed with 372 the split that maximizes eq. (2), provided it yields a nonnegative value, and add the new cells as 373 nodes of the tree to be examined in the next repetitions (lines 10–13, alg. 2); otherwise the node is 374 not split. Importantly, the set of directions allowed for splitting is a subset of the features, provided 375 by the user, allowing for controlled and interpretable partitioning of the data. 376

**Termination Criteria.** This process continues until one of the following termination criteria occurs: (1) a node has no remaining features to consider for splitting, (2) the maximum score of the actions

from all remaining features is negative, or (3) the specified number of leaf nodes has been reached (i.e., size constraints).

Approach Strengths. A key strength of T-GLANCE is its user-driven flexibility, allowing policy-381 makers or auditors to select the features that determine tree partitioning. By allowing users to guide 382 the feature selection, T-GLANCE can adapt to specific policy needs or audit requirements, making it highly customizable. T-GLANCE provides multi-level solutions, assigning a single optimal coun-384 terfactual action to each tree node, ensuring transparency and consistency across decision-making 385 processes. This enables both high-level and granular solutions for additional user flexibility, in con-386 tradiction with Kanamori et al. (2022), who optimize actions only at the leaf nodes. Additionally, 387 the method remains efficient, robust, and effective across various datasets and scenarios, as shown in 388 our evaluations, making it well-suited for real-world applications. Finally, unlike the stochastic local search used by Kanamori et al. (2022), which may not converge (in about 20% of our experiments), 389 T-GLANCE guarantees a solution at every node. 390

391 392

## 5 EXPERIMENTAL EVALUATION

393 394

395

## 5.1 EXPERIMENTAL SETTING

Baselines. We compare GLANCE framework methods against state-of-the-art methods in Global
 Counterfactual Explanations, specifically: AReS (using the Fast AReS implementation), CET,
 GroupCF, and GLOBE-CE, all of which are constrained by a predefined action set size. Details on
 the implementation used for each method can be found in Appendix F.

Datasets. We use four established benchmark datasets from previous research: COMPAS (Angwin et al., 2016), German Credit (Dua & Graff, 2019), Default Credit (Yeh & Lien, 2009), and HELOC (Brown et al., 2018). Additionally, we introduce the Adult dataset (Becker & Kohavi, 1996) for further evaluation. Details on the datasets and their preprocessing can be found in Appendix D.

404
 405
 406
 406
 407
 407
 408
 409
 409
 409
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400
 400

**Recourse Cost.** For computing recourse costs, we adhered to the guidelines established by Ley et al. (2023).

Reproducibility. All experiments were conducted on an in-house server with cloud infrastructure equipped with an Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz, 128 GB of RAM. No GPU acceleration was utilized during these experiments. We provide the code for the reproducibility of our experiments as supplementary material.

414 Running Time. In our experiments, GLOBE-CE and dGLOBE-CE were the fastest methods, con-415 sistently delivering solutions within 20 seconds across all datasets and models. C-GLANCE and T-GLANCE followed, typically completing in under 300 seconds, although T-GLANCE reached 416 up to 850 seconds in four cases. Other methods were slower, with Fast AReS ranging typically 417 between 150-400 seconds, and peaking at 1,400 seconds in some runs. The least computationally 418 efficient were Group-CF and CET, with maximum runtimes of 3,500 and 17,000 seconds, respec-419 tively, and CET failed to solve the underlying optimization problem after 20 hours of runtime for the 420 Adult dataset across all models. 421

421 422 423

## 5.2 EXPERIMENTAL EVALUATION

424 Table 1 presents the summarized results of all competing methods. We compare C-GLANCE and 425 T-GLANCE against the five other competitors across five datasets and three models, resulting in 426 75 head-to-head comparisons for each method, totaling 150. However, since CET failed to solve 427 the underlying optimization problem for the Adult dataset across all models, the final count is 72 428 comparisons per GLANCE method. Furthermore, we perform additional evaluations by categorizing 429 the competitors into implicit and explicit methods. Specifically, C-GLANCE is evaluated against implicit—Group-CF, GLOBE-CE, and dGLOBE-CE— resulting to 45 head-to-head comparisons. 430 Similarly, T-GLANCE is compared with explicit competitors — Fast AReS, CET— yielding 27 431 head-to-head comparisons. Recall that all reported results concern the s-GCE problem for s = 4;

Table 1: Evaluating the effectiveness and average cost of T-GLANCE against explicit (Fast AReS, CET) and C-GLANCE against implicit (Group-CF, GLOBE-CE, and dGLOBE-CE) GCE methods for s-GCE problem with s = 4. s-GCE solutions with effectiveness below 80% (applicability threshold) are highlighted in red. Non-robust GCEs, identified by either a standard deviation (std) in effectiveness greater than 5% across folds or a std in cost greater than half the average cost, are highlighted in blue.

MODELS	ALGORITHMS					DATA	SETS				
		ADI	JLT	COM	PAS	DEFAULT	CREDIT	GERMAN	CREDIT	HEI	LOC
		eff	avc	eff	avc	eff	avc	eff	avc	eff	avc
DNN	FAST ARES	$12.39 \pm 1.06$	$1.0 \pm 0.0$	55.0 ± 0.86	$1.21 \pm 0.09$	18.88 ± 2.16	$1.0 \pm 0.0$	52.39 ± 1.63	$1.0 \pm 0.0$	$12.19 \pm 0.58$	1.03 ±
	CET	$NAN \pm NAN$	$NAN \pm NAN$	$63.62 \pm 10.35$	$0.96 \pm 0.24$	$98.87 \pm 0.62$	$6.32 \pm 2.28$	97.3 ± 2.46	$1.58 \pm 0.54$	86.78 ± 10.62	8.67 ±
	T-GLANCE	$100.0 \pm 0.0$	$4.43 \pm 0.43$	$99.53 \pm 0.22$	$2.62 \pm 0.24$	$100.0 \pm 0.0$	$1.48 \pm 0.43$	$96.97 \pm 3.48$	$1.59 \pm 0.8$	99.8 ± 0.24	10.9 ±
	GROUP-CF	$100.0 \pm 0.0$	$10.08 \pm 0.03$	$100.0 \pm 0.0$	$4.48 \pm 2.53$	$79.6 \pm 20.79$	$1.53 \pm 0.62$	97.8 ± 4.4	$1.85 \pm 0.13$	80.4 ± 10.17	3.09 ±
	GLOBE-CE	$99.92 \pm 0.0$	$4.24 \pm 0.42$	$100.0 \pm 0.0$	$4.54 \pm 3.31$	76.94 ± 37.55	$5.14 \pm 0.35$	93.31 ± 3.48	$2.0 \pm 1.55$	42.72 ± 46.97	11.77 ±
	DGLOBE-CE	$99.92 \pm 0.0$	$10.89 \pm 1.37$	$100.0 \pm 0.0$	$7.96 \pm 3.91$	87.38 ± 18.69	$5.96 \pm 4.14$	$97.36 \pm 0.82$	$2.49 \pm 0.27$	99.96 ± 0.05	11.07 :
	C-GLANCE	$100.0 \pm 0.0$	$4.6 \pm 0.73$	$100.0 \pm 0.0$	$2.34 \pm 0.43$	$100.0 \pm 0.0$	$1.2 \pm 0.4$	95.31 ± 3.15	$1.25 \pm 0.33$	99.94 ± 0.05	11.24 ±
LR	FAST ARES	$11.74 \pm 2.4$	$1.0 \pm 0.0$	$62.5 \pm 1.82$	$1.24 \pm 0.14$	10.85 ± 5.45	$1.07 \pm 0.13$	75.27 ± 2.96	$1.0 \pm 0.0$	9.23 ± 1.24	1.12
	CET	NAN ± NAN	NAN ± NAN	$73.18 \pm 4.34$	$1.24 \pm 0.15$	$100.0 \pm 0.0$	$3.79 \pm 1.31$	$96.5 \pm 2.85$	$2.42 \pm 0.24$	$100.0 \pm 0.0$	3.57 ±
	T-GLANCE	$100.0 \pm 0.0$	$0.69 \pm 0.01$	$99.79 \pm 0.22$	$2.19 \pm 0.1$	$100.0 \pm 0.0$	$1.43 \pm 0.36$	$99.58 \pm 0.83$	$1.54 \pm 0.32$	$100.0 \pm 0.0$	1.58 ±
	GROUPCF	$100.0 \pm 0.0$	$1.71 \pm 0.39$	$100.0 \pm 0.0$	3.97 ± 2.38	95.4 ± 9.2	1.94 ± 1.2	97.6 ± 2.94	$9.34 \pm 3.85$	90.6 ± 3.93	2.4 ±
	GLOBE-CE	$99.92 \pm 0.0$	$2.68 \pm 0.17$	95.74 ± 8.52	$5.14 \pm 3.77$	$99.94 \pm 0.07$	3.42 ± 1.99	$57.09 \pm 20.03$	$0.75 \pm 1.04$	$99.9 \pm 0.0$	0.6 ±
	DGLOBE-CE	$99.92 \pm 0.0$	$5.91 \pm 0.93$	$100.0 \pm 0.0$	$6.71 \pm 0.23$	$99.94 \pm 0.07$	10.38 ± 7.76	$69.89 \pm 15.35$	$2.47 \pm 0.23$	$99.9 \pm 0.0$	1.63 ±
	C-GLANCE	$100.0 \pm 0.0$	$1.04 \pm 0.07$	$100.0 \pm 0.0$	$2.33 \pm 0.38$	$100.0 \pm 0.0$	$1.05 \pm 0.11$	$100.0 \pm 0.0$	$1.21 \pm 0.06$	$100.0 \pm 0.0$	1.55 ±
XGB	FAST ARES	$6.13 \pm 0.42$	$1.0 \pm 0.0$	59.83 ± 3.12	$1.1 \pm 0.05$	$31.86 \pm 5.12$	$1.05 \pm 0.04$	51.27 ± 1.57	$1.0 \pm 0.0$	8.49 ± 1.32	1.16 ±
	CET	NAN ± NAN	$NAN \pm NAN$	$58.4 \pm 9.3$	$1.06 \pm 0.24$	86.29 ± 9.94	$4.5 \pm 2.64$	$100.0 \pm 0.0$	$2.73 \pm 0.49$	86.78 ± 6.7	12.51 ±
	T-GLANCE	$99.86 \pm 0.14$	$1.8 \pm 0.51$	$99.02 \pm 0.93$	$2.6 \pm 0.46$	$94.6 \pm 3.13$	$2.75 \pm 1.42$	$99.46 \pm 1.08$	$1.33 \pm 0.42$	96.42 ± 2.45	24.85
	GROUPCF	96.8 ± 1.72	$1.41 \pm 0.54$	$100.0 \pm 0.0$	$4.06 \pm 2.1$	$95.2 \pm 1.6$	$1.41 \pm 0.64$	$100.0 \pm 0.0$	$5.78 \pm 4.11$	78.4 ± 5.82	5.63 ±
	GLOBE-CE	82.87 ± 12.14	$30.1 \pm 10.39$	$87.13 \pm 11.14$	9.75 ± 7.2	82.7 ± 7.26	$20.82 \pm 1.73$	$77.05 \pm 11.26$	$1.14 \pm 1.24$	27.66 ± 5.06	12.52 ±
	DGLOBE-CE	$93.76 \pm 1.98$	$64.76 \pm 1.29$	$99.84 \pm 0.31$	$12.46 \pm 3.42$	$97.47 \pm 0.82$	$42.58 \pm 3.57$	86.96 ± 9.79	$2.66 \pm 0.77$	77.64 ± 11.51	128.0
	C-GLANCE	$99.85 \pm 0.12$	$5.98 \pm 4.22$	$99.51 \pm 0.46$	$2.96 \pm 0.82$	$98.13 \pm 1.05$	$3.68 \pm 1.64$	$100.0 \pm 0.0$	$1.06 \pm 0.03$	$98.94 \pm 0.66$	19.99 :

Appendix H presents an extensive experimental evaluation, including: additional and more detailed results for s = 4; results for s = 8 for all methods and results under no size constraints for GLOBE-CE; results utilizing different counterfactual generation methods.

Pareto Dominance. We summarize method performance by determining whether one solution dom-inates another based on effectiveness and cost. Specifically, a solution  $\mathbb{C}$  of s-GCE Pareto dominates another solution  $\mathbb{C}'$  if it offers equal or better effectiveness and cost, and is strictly better in at least one of these objectives. As shown in Table 2, C-GLANCE dominates other methods in 39 out of 72 cases (54%), while is dominated only once by dGLOBE-CE (in HELOC-DNN-cf. Table 1). T-GLANCE dominates other methods in 30 out of 72 cases (41%), while it is dominated once by Group-CF (in Default Credit-XGB) and once by CET (in German Credit-DNN). In the three cases in which our method is dominated, the performance in terms of effectiveness is comparable. Overall, the GLANCE methods dominate other solutions in almost half of the cases (48%) and are dominated in only 2%. 

Solution Practicality. In the prior Pareto-dominance evaluation (Table 2) we compare solutions with optimal or near-optimal effectiveness to many solutions that exhibit unacceptable or unsatis-factory effectiveness. These lower-performing solutions are impractical for GCE, as the goal is to offer recourse to a large population segment. Solutions with low effectiveness fail to meet this goal, limiting their applicability in real-world scenarios. A solution that leaves a significant percentage of individuals without recourse undermines the very purpose of GCE, as noted by Ley et al. (2023). It is also important to note that achieving low recourse costs is easier for smaller subpopulations, especially those near the decision boundary, which explains the lower domination number when 

Table 2: Pareto domination evaluation of solutions, for s-GCE problem with s = 4. The table reports the rate (number of times over available comparisons) at which GLANCE methods dominate com-petitors (listed in the DOMINATES column) and the rate at which GLANCE methods are dominated by competitors (listed in the IS DOMINATED column). 

479					
480		C-GLANC	C-GLANCE (IMPLICIT)		CE (EXPLICIT)
481		DOMINATES	IS DOMINATED	DOMINATES	IS DOMINATED
482 483	FAST ARES (EXPLICIT) CET (EXPLICIT) GROUPCF (IMPLICIT) GLOBE-CE (IMPLICIT) DGLOBE-CE (IMPLICIT)	1/15 6/12 9/15 11/15 12/15	0/15 0/12 0/15 0/15 1/15	1/15 5/12 6/15 9/15 9/15	0/15 1/12 1/15 0/15 0/15
485	COMPETITORS (EXPLICIT) COMPETITORS (IMPLICIT) COMPETITORS (ALL)	32/45 39/72	1/45 1/72	6/27 30/72	1/27 2/72

486 Table 3: Pareto domination evaluation of solutions, for s-GCE problem with s = 4, after applying 487 the eff > 80% threshold. The table reports the rate (number of times over available comparisons) 488 at which GLANCE methods dominate competitors (listed in the DOMINATES column) and the rate at which GLANCE methods are dominated by competitors (listed in the IS DOMINATED column). 489

	C-GLAN	CE (IMPLICIT)	T-GLANC	CE (EXPLICIT)
	DOMINATES	IS DOMINATED	DOMINATES	IS DOMINATE
FAST ARES (EXPLICIT)	0/0	0/0	0/0	0/0
CET (EXPLICIT)	6/9	0/9	5/9	1/9
GROUPCF (IMPLICIT)	8/13	0/13	5/13	1/13
GLOBE-CE (IMPLICIT)	8/10	0/10	7/10	0/10
DGLOBE-CE (IMPLICIT)	10/13	1/13	7/13	0/13
COMPETITORS (EXPLICIT)			5/9	1/9
COMPETITORS (IMPLICIT)	26/36	1/36		
COMPETITORS (ALL)	32/45	1/45	24/45	2/45

497 498 499

501

comparing with explicit methods. We consider a solution to be practical if it achieves effectiveness 500 of at least 80%. In Table 1, we highlight the impractical solutions in red. GLANCE methods never return impractical solutions. In contrast, all the solutions returned by Fast AReS are deemed im-502 practical, while the other methods return 2–5 impractical solutions each. Table 3 summarizes the head-to-head comparisons only when considering practical solutions. GLANCE framework domi-504 nates other methods in a much larger percentage—over 62% of the cases. It remains dominated in the same three cases, maintaining the percentage of dominated cases at a low 3%. 505

506 **Explicit-Implicit Evaluation.** When evaluating our methods within their respective explicit or im-507 plicit categories, C-GLANCE dominates other implicit methods in 71% of cases (32 out of 45), 508 compared to a 54% domination rate across all methods. When focusing solely on practical solu-509 tions, this rate increases to 72% (26 out of 36). For T-GLANCE, the domination rate against explicit methods is 22% (6 out of 27), lower than the 48% achieved when considering all methods. This 510 drop is attributed to the high number of impractical solutions provided by competing explicit meth-511 ods. However, after removing impractical solutions, T-GLANCE dominates 56% of cases, slightly 512 higher than the 53% in head-to-head comparisons across all methods. Overall, the GLANCE frame-513 work demonstrates strong performance, dominating 52% of cases within categories (69% when 514 considering only practical solutions) while keeping the percentage of dominated cases at a low 4% 515 (3% when considering only practical solutions). 516

**Robustness.** We expect methods to be robust, *consistently* generating highly effective and low-cost 517 GCEs across different data splits, which is crucial for real-world deployment. Without robustness, 518 recourse actions can vary significantly, undermining trust and leading to unfair outcomes, especially 519 in critical areas like healthcare or finance. Evaluating the stability of effectiveness and cost metrics 520 across different folds is key to determining the practical applicability of a counterfactual explanation 521 method. Standard deviation measures this stability. An effectiveness deviation above 5% indicates 522 an inconsistency in providing recourse, while a cost deviation exceeding half the average suggests 523 unpredictable, impractical actions. These fluctuations make a solution unreliable and we highlight 524 them in red in Table 1. Figure 3 visualizes Table's 1 results, i.e., the effectiveness of each solution 525 vs. the cost (scaled to 0-1 values), along with their standard deviation. This figure provides a 526 clear comparison of the trade-offs between cost and effectiveness, underscoring the robustness and consistency of the methods across different scenarios. 527

528 529

530

### CONCLUSION 6

531 This paper introduces GLANCE, a robust and flexible framework for generating global counterfactual 532 explanations that effectively balances the trade-offs between effectiveness, cost, and interpretability. 533 An extensive experimental evaluation demonstrates that GLANCE constructs counterfactual expla-534 nations that are more effective and cheaper than those produced by the state of the art under size 535 constraints.

536

### REFERENCES 538

Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias. Ethics of Data and Analytics, pp. 254–264, 5 2016. doi: 10.1201/9781003278290-37.

540 541	Barry Becker and Ronny Kohavi. Adult. UCI Machine Learning Repository, 1996. DOI: https://doi.org/10.24432/C5XW20.
542 543 544	Dimitris Bertsimas and Jack Dunn. Optimal classification trees. <i>Machine Learning</i> , 106:1039–1082, 2017.
545 546 547	Dimitris Bertsimas, Agni Orfanoudaki, and Holly Wiberg. Interpretable clustering: an optimization approach. <i>Machine Learning</i> , 110(1):89–138, 2021.
548 549	Jürgen Branke. <i>Multiobjective optimization: Interactive and evolutionary approaches</i> , volume 5252. Springer Science & Business Media, 2008.
550 551 552	Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. Cart. Classification and regression trees, 1984.
553 554 555	Kyle Brown, Derek Doran, Ryan Kramer, and Brad Reynolds. HELOC applicant risk perfor- mance evaluation by topological hierarchical decomposition. <i>CoRR</i> , abs/1811.10658, 2018. URL http://arxiv.org/abs/1811.10658.
556 557 558 559	Miguel Á Carreira-Perpiñán and Suryabhan Singh Hada. Counterfactual explanations for oblique decision trees: Exact, efficient algorithms. In <i>Proceedings of the AAAI conference on artificial intelligence</i> , volume 35, pp. 6903–6911, 2021.
560 561 562	Emilio Carrizosa, Jasone Ramírez-Ayerbe, and Dolores Romero Morales. Generating collective counterfactual explanations in score-based classification via mathematical optimization. <i>Expert Systems with Applications</i> , 238:121954, 2024a.
563 564 565 566	Emilio Carrizosa, Jasone Ramírez-Ayerbe, and Dolores Romero Morales. Mathematical optimiza- tion modelling for group counterfactual explanations. <i>European Journal of Operational Research</i> , 2024b.
567 568 569	Tanya Chowdhury, Razieh Rahimi, and James Allan. Equi-explanation maps: concise and infor- mative global summary explanations. In <i>Proceedings of the 2022 ACM Conference on Fairness,</i> <i>Accountability, and Transparency</i> , pp. 464–472, 2022.
570 571 572	Sanjoy Dasgupta, Nave Frost, Michal Moshkovitz, and Cyrus Rashtchian. Explainable k-means and k-medians clustering. In <i>Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria</i> , pp. 12–18, 2020.
573 574 575	Dheeru Dua and C Graff. Uci machine learning repository. university of california, school of infor- mation and computer science, irvine, ca (2019), 2019.
576 577 578	Nave Frost, Michal Moshkovitz, and Cyrus Rashtchian. Exkmc: Expanding explainable k-means clustering. arXiv preprint arXiv:2006.02399, 2020.
579 580 581	Zexi Huang, Mert Kosan, Sourav Medya, Sayan Ranu, and Ambuj Singh. Global counterfactual ex- plainer for graph neural networks. In <i>Proceedings of the Sixteenth ACM International Conference</i> on Web Search and Data Mining, pp. 141–149, 2023.
582 583 584	Kentaro Kanamori, Takuya Takagi, Ken Kobayashi, and Yuichi Ike. Counterfactual explanation trees: Transparent and consistent actionable recourse with decision trees. In <i>International Conference on Artificial Intelligence and Statistics</i> , pp. 1846–1870. PMLR, 2022.
586 587 588	Amir-Hossein Karimi, Gilles Barthe, Bernhard Schölkopf, and Isabel Valera. A survey of algorithmic recourse: definitions, formulations, solutions, and prospects. <i>arXiv preprint arXiv:2010.04050</i> , 2020.
589 590 591	Loukas Kavouras, Konstantinos Tsopelas, Giorgos Giannopoulos, Dimitris Sacharidis, Eleni Psaroudaki, Nikolaos Theologitis, Dimitrios Rontogiannis, Dimitris Fotakis, and Ioannis Emiris. Fairness aware counterfactuals for subgroups. <i>arXiv preprint arXiv:2306.14978</i> , 2023.
592 593	Jaehoon Koo, Diego Klabjan, and Jean Utke. Inverse classification with limited budget and maxi- mum number of perturbed samples. arXiv preprint arXiv:2009.14111, 2020.

- Eduardo S Laber and Lucas Murtinho. On the price of explainability for some clustering problems. In *International Conference on Machine Learning*, pp. 5915–5925. PMLR, 2021.
- Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Jure Leskovec. Faithful and customizable
   explanations of black box models. In *Proceedings of the 2019 AAAI/ACM Conference on AI*,
   *Ethics, and Society*, pp. 131–138, 2019.
- Dan Ley, Saumitra Mishra, and Daniele Magazzeni. Global counterfactual explanations: Investigations, implementations and improvements. In *ICLR Workshop on Privacy, Accountability, Interpretability, Robustness, Reasoning on Structured Data*, 2022.
- Dan Ley, Saumitra Mishra, and Daniele Magazzeni. GLOBE-CE: A translation based approach for global counterfactual explanations. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 19315–19342. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr. press/v202/ley23a.html.
- Konstantin Makarychev and Liren Shan. Explainable k-means: don't be greedy, plant bigger trees!
   In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1629–1642, 2022.
- Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.
- Michal Moshkovitz, Sanjoy Dasgupta, Cyrus Rashtchian, and Nave Frost. Explainable k-means and
   k-medians clustering. In *International conference on machine learning*, pp. 7055–7065. PMLR, 2020.
- Ramaravind K Mothilal, Amit Sharma, and Chenhao Tan. Explaining machine learning classifiers
   through diverse counterfactual explanations. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 607–617, 2020.
- Kaivalya Rawal and Himabindu Lakkaraju. Beyond individualized recourse: Interpretable and interactive summaries of actionable recourses. *Advances in Neural Information Processing Systems*, 33:12187–12198, 2020.
- Shubham Sharma, Jette Henderson, and Joydeep Ghosh. CERTIFAI: counterfactual explanations
   for robustness, transparency, interpretability, and fairness of artificial intelligence models. *CoRR*,
   abs/1905.07857, 2019. URL http://arxiv.org/abs/1905.07857.
- Berk Ustun, Alexander Spangher, and Yang Liu. Actionable recourse in linear classification. In *Proceedings of the conference on fairness, accountability, and transparency*, pp. 10–19, 2019.
- Sahil Verma, Varich Boonsanong, Minh Hoang, Keegan E Hines, John P Dickerson, and Chirag
   Shah. Counterfactual explanations and algorithmic recourses for machine learning: A review.
   *arXiv preprint arXiv:2010.10596*, 2020.
- Sandra Wachter, Brent Mittelstadt, and Chris Russell. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31:841, 2017.
- Greta Warren, Mark T Keane, Christophe Gueret, and Eoin Delaney. Explaining groups of instances
   counterfactually for xai: a use case, algorithm and user study for group-counterfactuals. *arXiv preprint arXiv:2303.09297*, 2023.
- Ivy Yeh and Che-Hui Lien. The comparisons of data mining techniques for the predictive accuracy
   of probability of default of credit card clients. *Expert Systems with Applications*, 36:2473–2480,
   03 2009. doi: 10.1016/j.eswa.2007.12.020.
- 644

637

- 645
- 646 647

## 648 A HARDNESS

ments in  $\mathcal{U}$ .

To analyze the complexity of the *s*-GCE problem, we suppose that the set of all actions is finite by discretizing continuous features.

Theorem 3 (NP-hardness). The s-GCE problem is NP-hard.

654 *Proof.* We establish the NP-hardness of the *s*-GCE problem by reducing the well-known *Max k-Coverage* problem to it.
 656

The Max k-Coverage problem is defined as follows: Given a universe  $\mathcal{U}$  of elements and a collection of subsets  $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$  where each  $S_i \subseteq \mathcal{U}$ , along with an integer k, the goal is to select at most k subsets from  $\mathcal{S}$  such that the total number of covered elements is maximized.

To perform the reduction, we proceed by mapping each subset  $S_i$  to an action  $a_i$  in the s-GCE problem:

- 662
- 663
- 64
- 665 666 667

668

669 670 • Coverage Mapping: The effectiveness of an action  $a_i$  in the *s*-GCE problem is defined as the elements in  $S_i$ .

• Actions and Subsets Correspondence: Each action  $a_i$  corresponds to a subset  $S_i$  of ele-

• **Ignoring Costs:** We ignore the cost associated with actions in the *s*-GCE problem, focusing solely on coverage.

Under this mapping, the objective of selecting at most k actions in the *s*-GCE problem to maximize the coverage of elements aligns exactly with the objective of the Max k-Coverage problem.

675

676

677

678

Given that even the relaxed version of the *s*-GCE problem that ignores costs is NP-hard, identifying the optimal solution for the *s*-GCE is computationally infeasible. Moreover, the action space, although finite, may be too large to process efficiently. Therefore, our approach involves generating a manageable subset  $\mathbb{A}' \subseteq \mathbb{A}$  of actions before extracting the final *s* actions. In the next section, we will introduce C-GLANCE, an efficient heuristic designed to solve the *s*-GCE problem.

679 680 681

682

683 684

685

606

## B SIZE-COST AND SIZE-EFFECTIVENESS TRADE-OFFS

Let  $\mathbb{C}$  be a set of actions. Let a new action a and the  $\mathbb{C}' = \mathbb{C} \cup \{a\}$ . We analyze how increasing the size (size) of the solution, impacts the effectiveness (eff) and the average cost (avc).

000		
687	Case 1	The new action does not flip any new instances.
688		In this scenario, the effectiveness remains the same.
689		• If the new action is cheaper than the current actions:
690		Some instances already receiving recourse from the current actions may now be ad-
691		dressed by the cheaper new action, reducing the average cost.
692		• If the new action is costlier than the current actions:
693		The current actions will continue to provide recourse for the instances they already
694		cover, leaving the average cost unchanged.
695	Case 2	The new action flips new instances.
696		In this scenario, the effectiveness increases, as more instances now receive recourse.
697		• If the cost of the new action is lower than the current average cost:
698		The new action provides recourse for additional instances at a lower cost, thereby
699		reducing the average cost.
700		• If the cost of the new action is higher than the current average cost:
701		The inclusion of the new action raises the average cost, as the costlier new instances contribute positively to the total, increasing the average cost.

702 This analysis highlights the trade-offs between size, effectiveness, and cost, emphasizing the im-703 portance of carefully selecting new actions to achieve a desirable balance. Notably, achieving near-704 optimal effectiveness (close to 100%) often requires adding actions that address outliers, which may 705 have higher costs. However, one of the strengths of our method is its ability to achieve near-optimal 706 effectiveness with relatively low average costs across most datasets.

#### 708 С TIME COMPLEXITY ANALYSIS 709

710 Let n be the number of instances, k the initial number of clusters, s the number of global counter-711 factual actions, m the number of candidate actions to generate, and d the number of features. Let 712  $\mathcal{T}_{CF}(d, \text{model})$  be the time of generating a single candidate counterfactual action and  $\mathcal{T}_{\text{model}}$  be the 713 prediction time of the black-box model. 714

715 C.1 C-GLANCE 716

C.2 T-GLANCE

717 Let's assume that k-means is used for the clustering and Iterations are the number of iterations required for convergence. Then C-GLANCE time complexity is: 718

• Merging Clusters:  $\mathcal{O}((k-s) \cdot k^2 \cdot d + (k-s) \cdot (n \cdot d + m \cdot (k-s) \cdot d))$ 

• Clustering (K-means):  $\mathcal{O}(n \cdot k \cdot d \cdot \text{Iterations})$ 

• Action Generation:  $\mathcal{O}(k \cdot m \cdot \mathcal{T}_{CF}(d, \text{model}))$ 

- Finding pair of closest clusters:

\* For one iteration:  $\mathcal{O}(k^2 \cdot d)$ 

- Merging pair of closest clusters:

After simplifying common terms, the total complexity is:

Let  $d' \leq d$  be the number of the selected features.

\* For all iterations:  $\mathcal{O}((k-s) \cdot k^2 \cdot d)$ 

\* For one iteration:  $\mathcal{O}(n \cdot d + m \cdot (k - s) \cdot d)$ 

\* For all iterations:  $\mathcal{O}((k-s) \cdot n \cdot d + m \cdot (k-s) \cdot d)$ 

• Evaluation and selection of final actions:  $\mathcal{O}(m \cdot k \cdot n \cdot d + n \cdot \mathcal{T}_{\text{model}})$ 

 $\mathcal{O}(n \cdot k \cdot d \cdot \text{Iterations} + k \cdot m \cdot \mathcal{T}_{CF}(d, \text{model}) + d \cdot (k - s) \cdot k^2 + n \cdot k \cdot m \cdot (d + \mathcal{T}_{\text{model}}))$ 

• Initial action assignment (root):  $\mathcal{O}\Big(m\big(\mathcal{T}_{CF}(d, \text{model}) + n \cdot d + n \cdot \mathcal{T}_{\text{model}}\big)\Big)$ 

- Search-and-split a single node:  $\mathcal{O}\left(d' \cdot m \cdot \left(n \cdot (d + \mathcal{T}_{\text{model}}) + \mathcal{T}_{CF}(d, model)\right)\right)$ 

\* Brief Explanation: For every possible split, generate m actions on both children

and evaluate them. Given these, the rest (finding maximum effectiveness action,

• Tree construction:  $\mathcal{O}\left(2^{d'} \cdot d' \cdot m \cdot \left(n \cdot (d + \mathcal{T}_{\text{model}}) + \mathcal{T}_{CF}(d, model)\right)\right)$ 

724

727

728 729

730 731

732 733 734

735

754

755

The total complexity is dominated of course by the tree construction term and is equal to:  $\mathcal{O}\left(2^{d'} \cdot d' \cdot m \cdot \left(n \cdot (d + \mathcal{T}_{\text{model}}) + \mathcal{T}_{CF}(d, model)\right)\right)$ 

computing V(A, i) are much faster.

– Maximum number of nodes:  $\mathcal{O}(2^{d'})$ 

## 756 D DATASETS & PREPROCESSING

We use five publicly available datasets as benchmarks. Our choice is based on their established use in previous works. Table 4 summarizes the datasets' information, including the number of instances, the number of categorical and continuous features, the input dimensions (i.e., the number of continuous features plus the number of categorical after a preprocessing step using one-hot-encoding), and the number of instances used for training and testing, evaluated using a 5-fold cross-validation

The dataset preprocessing follows the approach in Ley et al. (2023), using the publicly available repository at https://github.com/danwley/GLOBE-CE/. For completeness, we also provide a short description of each dataset below.

strategy. This approach illustrates the resilience of our method across various splits.

767

763

Adult The Adult dataset (also known as "Census Income" dataset) is designed for the task of pre dicting whether an individual's income exceeds \$ 50K/yr, based on census data. The data as well as
 more detailed information can be obtained at https://archive.ics.uci.edu/dataset/
 2/adult.

For the preprocessing of this dataset, we first drop the "education-num" feature due to redundancy with the "education" feature (which takes on string values). Next, we remove missing values and map the class labels '<=50K' and '>50K' to 0 and 1, respectively. Additional minor transformations, mostly involving data types, are applied and can be reviewed in our source code.

776 777

COMPAS The COMPAS dataset (Correctional Offender Management Profiling for Alternative Sanctions) Angwin et al. (2016) is available at https://github.com/propublica/compas-analysis/blob/master/compas-scores-two-years.csv. Detailed description and information on the dataset can be found at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. It categorizes recidivism risk based on several factors, including race.

For the preprocessing of this dataset, we drop the "days\_b\_screening\_arrest" feature, as it contains missing values. We also turn jail-in and jail-out dates to durations and turn negative durations to 0. Some additional filters are taken from the COMPAS analysis by ProPublica. Finally, the target variable's values are transformed into the canonical 0 for the negative class and 1 for the positive class.

789 German Credit The German Credit dataset Dua & Graff (2019) classifies people described
 790 by a set of attributes as good or bad credit risks. A detailed description and the dataset can
 791 be found in https://archive.ics.uci.edu/ml/datasets/statlog+(german+
 792 credit+data).

The only preprocessing step we performed for this dataset was the transformation of the target variable's values into 0 - 1.

Default Credit The Default Credit dataset Yeh & Lien (2009) is designed to classify the risk of default on customer payments, aiming to support the development and assessment of models for predicting creditworthiness and the likelihood of loan default. It can be obtained at https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients.

To properly work with this dataset, we needed to drop the "ID" feature, since it holds no useful information, and transform the target labels into the canonical 0 - 1 values.

HELOC The HELOC (Home Equity Line of Credit) dataset Brown et al. (2018) contains anonymized information about home equity line of credit applications made by real home-owners, classifying credit risk. It is available at https://community.fico.com/s/explainable-machine-learning-challenge. All the features on this dataset are numeric.

A substantial percentage of these features' values are missing, so the main preprocessing step we performed here was to remove rows where all values are missing, and then replace all remaining

missing values with the median of the respective feature. Other than that, we only needed to transform target labels to 0-1.

Table 4: Summary of the datasets used in our experiments. Specifically, we list the number of instances, input dimensions (i.e., the number of continuous features plus the number of categorical after the one-hot-encoding preprocessing step performed internally by the models), the number of categorical and continuous features, and the number of instances used for training and testing, evaluated using a 5-fold cross-validation strategy.

DATASET	NO. INSTANCES	INPUT DIM.	CATEGORICAL	CONTINUOUS	TRAIN	TEST
ADULT	30161	102	8	5	24128	6033
COMPAS	6172	15	4	2	4937	1235
DEFAULT CREDIT	30000	91	9	14	24000	6000
GERMAN CREDIT	1000	71	17	3	800	200
HELOC	9871	23	0	23	7896	1975

## E MODELS AND HYPERPARAMETERS

In our experimental evaluation, we utilize three distinct models: XGBoost (XGB), Logistic Regression (LR), and Deep Neural Networks (DNNs). Following Ley et al. (2023), we maintain an 80:20 train-test split, but, instead of splitting once, we perform 5-fold cross-validation, using each fold as a test set while training on the remaining four. The distinctive hyperparameters for each model are described in detail in this section. Additionally, we showcase the performance metric (accuracy), with all reported accuracies presented as the mean and standard deviation across the folds, providing a standardized foundation for comparative analysis of our methodologies.

**XGBoost (XGB)** Implementation from the common  $xgboost^1$  library. Hyperparameter values for each dataset and the model's accuracy on the test set are shown in table 5.

 Table 5: XGBoost Hyperparameter Configurations.

DATASET	Depth	ESTIMATORS	$\gamma, lpha, \lambda$	TEST ACCURACY
ADULT COMPAS DEFAULT CREDIT GERMAN CREDIT HELOC	6 4 10 6 6	100 100 200 500 100	$0,0,1 \\ 1,0,1 \\ 2,4,1 \\ 0,0,1 \\ 4,4,1$	$\begin{array}{c} 86.62\% \pm 0.18\% \\ 67.95\% \pm 1.73\% \\ 81.35\% \pm 0.18\% \\ 76.9\% \pm 0.86\% \\ 72.97\% \pm 1.14\% \end{array}$

**Logistic Regression (LR)** Implementation from the common sklearn<sup>2</sup> library. Hyperparameter values for each dataset and the model's accuracy on the test set are shown in table 6.

Table 6: Logistic Regression Hyperparameter Configurations.

DATASET	MAX ITER.	CLASS WEIGHTS(0:1)	TEST ACCURACY
ADULT COMPAS DEFAULT CREDIT GERMAN CREDIT HELOC	100 1000 1000 1000 3000	1:1 1:1 1:1 1:1	$\begin{array}{c} 79.09\% \pm 0.12\% \\ 66.69\% \pm 0.97\% \\ 82.08\% \pm 0.19\% \\ 74.80\% \pm 1.91\% \\ 73.06\% \pm 0.96\% \end{array}$

**Deep Neural Network (DNN)** Implementation using pytorch<sup>3</sup> library. Hyperparameter values for each dataset and the model's accuracy on the test set are shown in table 7.

<sup>1</sup>https://xgboost.readthedocs.io/en/stable/

<sup>2</sup>https://scikit-learn.org/stable/modules/generated/sklearn.linear\_

<sup>863</sup> model.LogisticRegression.html

<sup>&</sup>lt;sup>3</sup>https://pytorch.org/docs/stable/index.html

DATASET	HIDDEN LAYERS WIDTH	TEST ACCURACY
ADULT COMPAS DEFAULT CREDIT GERMAN CREDIT HELOC	$ \begin{array}{c} [64, 32, 16]\\[64, 32, 16]\\[64, 32, 16, 8]\\[8, 4]\\[64, 32, 16, 8]\end{array} $	$\begin{array}{c} 84.13\% \pm 0.30\% \\ 66.98\% \pm 1.27\% \\ 81.37\% \pm 0.13\% \\ 71.3\% \pm 2.27\% \\ 70.97\% \pm 0.54\% \end{array}$

Table 7: Deep Neural Network Hyperparameter Configurations.

## F COMPETING METHODS

**GLOBE-CE.** The GLOBE-CE method, introduced by Ley et al. (2023) was executed based on the implementation available on their GitHub (https://github.com/danwley/GLOBE-CE).

Table 8: GLOBE-CE results without constraining the action set size

MODELS	DATASETS														
		(		DEFAULT CREDIT			Germ	GERMAN CREDIT			HELOC				
	eff	avc	size	eff	avc	size	eff	avc	size	eff	avc	size	eff	avc	size
DNN	$100.0 \pm 0.0$	$1.24 \pm 0.06$	322.4	$100.0 \pm 0.0$	$1.11 \pm 0.6$	186.2	98.9 ± 0.73	$1.1 \pm 0.04$	18.4	95.69 ± 2.42	$1.12 \pm 0.09$	4.2	$84.32 \pm 19.32$	$2.63 \pm 1.03$	224.8
LR	$100.0 \pm 0.0$	$1.2 \pm 0.03$	545.2	$98.92 \pm 8.52$	$1.41 \pm 0.19$	97.8	$99.94 \pm 0.07$	$1.08 \pm 0.01$	4.6	71.29 ± 14.27	$1.05 \pm 0.09$	1.5	$100.00 \pm 0.0$	$0.44 \pm 0.06$	512.8
XGB	$93.76 \pm 0.91$	$4.28 \pm 4.46$	46.8	$92.34 \pm 8.87$	$1.22 \pm 0.27$	42.6	$94.62 \pm 2.85$	$1.05 \pm 0.28$	40.6	84.97 ± 10.46	$1.17 \pm 0.15$	2.2	$30.32 \pm 4.07$	$2.11 \pm 0.47$	17

Additional results of GLOBE-CE can be found in Table 8. Here we do not constrain the set of actions and we observe better performance, especially in terms of cost. However, these solutions do not represent solutions for *s*-GCE problem, since the solution size can be up to an average of 545.2 actions using 5-fold cross-validation (in Adult-LR case).

dGLOBE-CE. The dGLOBE-CE method, introduced by Ley et al. (2023), was executed based on
 the implementation available on their GitHub repository (https://github.com/danwley/
 GLOBE-CE).

AReS. The AReS method, first introduced by Rawal & Lakkaraju (2020), was executed using the enhanced Fast AReS version implemented by Ley et al. (2023). The code for this version was also obtained from the same GitHub repository (https://github.com/danwley/GLOBE-CE).
We used the default hyperparameters provided in this implementation.

GroupCF. The GroupCF method, introduced by Warren et al. (2023), was implemented using custom code due to some ambiguity in the instructions provided in the original GitHub repository (https://github.com/e-delaney/group\_cfe), which made the original implementation challenging to follow.

**CET.** The CET method, introduced by Kanamori et al. (2022), was executed using the original implementation available on GitHub (https://github.com/kelicht/cet). We had to replace the solver they used for the MIP solved at each node of their tree. IBM CPLEX solver is proprietary and we replaced it with the GUROBI solver, for which we had an available license. We used the default hyperparameters provided in this implementation.

## G CANDIDATE COUNTERFACTUAL ACTION GENERATORS

In this section, we outline the methods used to generate candidate counterfactual explanations and
present comparative results employing different generation methods. To demonstrate the modularity
of our framework, we developed and utilized various methods, showing that our approach is not tied
to a specific counterfactual generation technique. The comparative results of these experiments are
shown in Tables 13 and 14. For the results in Table 1 of the main paper and Table 10, DiCE (Mothilal et al., 2020) was used to generate candidate counterfactuals.

#### 918 G.1 DICE 919

920 Detailed information on its methodology can be found in the corresponding paper by Mothilal et al. 921 (2020), and we encourage interested readers to refer to it for a deeper understanding.

922 For the scope of this paper, it suffices to say that DiCE offers a number of alternative algorithms 923 for finding local counterfactuals. In our experiments, we use the 'random' algorithm, which selects 924 features to change at random and tests whether the change results in a successful counterfactual. 925

The other algorithms were either unsuitable for a black-box setting (e.g., requiring gradients), com-926 putationally inefficient (e.g., genetic algorithm), or failed to generate cost-efficient counterfactual 927 (e.g., KDTree) during our initial experimentation. 928

929 G.2 RANDOM SAMPLING 930

931 We developed a method called 'Random Sampling'. To find counterfactuals for an affected instance, 932 this method iteratively modifies its features one at a time. The process begins by randomly altering 933 one feature at a time, generating multiple new candidate instances. Each modified instance is then 934 evaluated by querying the black-box model to determine whether it qualifies as a valid counterfac-935 tual. The method proceeds to modify additional features, building a set of potential counterfactuals.

- 936 Key differences between our method and DiCE's random approach include: 937
  - 1. We focus on modifying only the top  $k_f$  most important features (using permutation feature importance), set to 3 in all experiments.
  - 2. For categorical features, only the top  $k_c$  most frequent categories among unaffected individuals are considered replacement candidates (set to 10 for all experiments).
  - 3. We also introduce vectorization in certain operations, improving computational efficiency over DiCE's implementation.

## G.3 NEAREST NEIGHBORS

This method is implemented by storing all unaffected individuals in memory. When queried to provide k counterfactuals for an affected individual, it retrieves the k nearest neighbors from the set of unaffected instances based on their proximity to the affected individual. This approach ensures that the generated counterfactuals are valid and closely aligned with the original instance, improving the relevance of the recommendations.

G.4 NEAREST NEIGHBORS SCALED

This algorithm closely resembles the 'Nearest Neighbors' one but introduces a key enhancement. Rather than returning the k nearest neighbors, it performs a localized search along the multidimen-956 sional line connecting the affected individual and any neighbor. We sample points along this line segment, and if any are classified as positive by the model, they are returned instead, as they are closer to the affected instance, potentially offering a more cost-efficient recourse.

959 960 961

962

964

938

939

940

941

942

943

944 945

946 947

948

949

950

951

952 953

954

955

957

958

### Η **EXPERIMENTAL RESULTS**

963 H.1 EXPERIMENTAL PROCEDURE

We evaluate our algorithms by comparing them with the following methods: Fast AReS by Ley 965 et al. (2022), CET by Kanamori et al. (2022), GroupCF by Warren et al. (2023), GLOBE-CE and 966 dGLOBE-CE by Ley et al. (2023). 967

968 All of the methods are constrained to solve s-GCE problem, with s = 4 and s = 8. These constraints are applied differently depending on the method. In C-GLANCE and GroupCF, we configure the 969 resulting number clusters to match the target sizes. GLOBE-CE is constrained by a maximum of 4 970 or 8 scalars, while dGLOBE-CE is limited to 4 or 8 directions, with up to two scalars per direction. 971 Lastly, T-GLANCE and CET are constrained with a maximum of 4 or 8 leaves.

In the main paper, we present summarized results for s = 4 in Table 1, with complementary results for s = 8 in Table 10 of the following section. The summarized results include the effectiveness and cost of each solution for all the dataset and model combinations. In Tables 15 to 19, we provide detailed results for each model, adding runtime and the number of actions per dataset and method alongside effectiveness and cost.

978 H.2 EXPERIMENTS FOR EFFECTIVENESS-COST TRADE-OFFS

As shown in Table 1, our methods significantly outperform CET and Fast AReS in terms of effectiveness. Fast AReS fails to exceed 80% effectiveness in any case, and CET achieves near-optimal effectiveness in less than half the cases, sometimes failing to find a solution altogether. These low effectiveness levels make fair comparisons challenging, as solving for a small subset of the population is inherently easier.

To address this challenge, we demonstrate in Table 9 that by intentionally lowering our solutions' effectiveness and selecting cost-efficient actions, our approach can still dominate other methods, achieving both greater effectiveness and lower cost. This adaptability underscores the strength of our methodology in maintaining superior performance even when optimizing for different trade-offs.

Our algorithm's default strategy selects optimal solutions based on maximum effectiveness. How ever, it also supports alternative strategies, such as selecting the lowest-cost solution, the lowest-cost solution above a specified effectiveness threshold, or the most effective solution below a specified cost. For the experiments presented Table 9, we used the "lowest cost above a certain effectiveness threshold" strategy in C-GLANCE, setting thresholds based on the effectiveness levels achieved by other methods. This flexibility highlights the algorithm's ability to align with various evaluation scenarios and user-defined trade-offs.



Figure 3: Comparison of effectiveness and average cost, normalized with the maximum cost achieved in each dataset/model combination) for the solution of *s*-GCE with s = 4. Standard deviations are represented by error bars. The red horizontal lines represent the eff > 80% threshold for evaluating the practicality of the solutions.

Table 9: Experiments when C-GLANCE effectiveness is intentionally reduced (provided as threshold) to find dominating solutions toward the lower-effectiveness solutions of other methods.

MODELS	CSOMPETITORS						DATA	SETS					
			AD	ULT			DEFAULT	CREDIT		HELOC			
		COMPETITOR eff	COMPETITOR avc	C-GLANCE eff	C-GLANCE ave	COMPETITOR eff	COMPETITOR ave	C-GLANCE eff	C-GLANCE avc	COMPETITOR eff	COMPETITOR avc	C-GLANCE eff	C-GLANCE av
DNN	FAST ARES CET GROUPCF GLOBE-CE DGLOBE-CE	$\begin{array}{c} 12.39 \pm 1.06 \\ NAN \\ 100.0 \pm 0.0 \\ 99.92 \pm 0.0 \\ 99.92 \pm 0.0 \end{array}$	$\begin{array}{c} 1.0 \pm 0.0 \\ \mathrm{NAN} \\ 10.08 \pm 0.03 \\ 4.24 \pm 0.42 \\ 10.89 \pm 1.37 \end{array}$	$\begin{array}{c} 26.59 \pm 10.5 \\ N \wedge N \\ 100.0 \pm 0.0 \\ 99.96 \pm 0.01 \\ 99.96 \pm 0.01 \end{array}$	$\begin{array}{c} 0.66 \pm 0.23 \\ NAN \\ 1.73 \pm 0.06 \\ 1.71 \pm 0.05 \\ 1.71 \pm 0.05 \end{array}$	$\begin{array}{c} 18.88 \pm 2.16 \\ 98.87 \pm 0.62 \\ 79.6 \pm 20.79 \\ 76.94 \pm 37.55 \\ 87.38 \pm 18.69 \end{array}$	$\begin{array}{c} 1.0\pm0.0\\ 6.32\pm2.28\\ 1.53\pm0.62\\ 5.14\pm0.35\\ 5.96\pm4.14 \end{array}$	$\begin{array}{c} 21.54 \pm 4.11 \\ 99.83 \pm 0.1 \\ 81.44 \pm 2.18 \\ 81.44 \pm 2.18 \\ 91.22 \pm 3.87 \end{array}$	$\begin{array}{c} 0.3 \pm 0.1 \\ 0.82 \pm 0.04 \\ 0.8 \pm 0.08 \\ 0.8 \pm 0.08 \\ 0.81 \pm 0.07 \end{array}$	$\begin{array}{c} 12.19\pm0.58\\ 86.78\pm10.62\\ 80.4\pm10.17\\ 42.72\pm46.97\\ 99.96\pm0.05 \end{array}$	$\begin{array}{c} 1.03 \pm 0.05 \\ 8.67 \pm 3.25 \\ 3.09 \pm 0.91 \\ 11.77 \pm 15.87 \\ 11.07 \pm 8.6 \end{array}$	$\begin{array}{c} 21.32\pm 5.13\\ 92.67\pm 4.69\\ 82.31\pm 1.81\\ 45.49\pm 2.1\\ 100.0\pm 0.0 \end{array}$	$\begin{array}{c} 0.11 \pm 0.11 \\ 3.52 \pm 1.09 \\ 3.09 \pm 0.9 \\ 1.19 \pm 0.71 \\ 4.22 \pm 1.10 \end{array}$
LR	FAST ARES CET GROUPCF GLOBE-CE DGLOBE-CE	$\begin{array}{c} 11.74 \pm 2.4 \\ NAN \\ 100.0 \pm 0.0 \\ 99.92 \pm 0.0 \\ 99.92 \pm 0.0 \end{array}$	$\begin{array}{c} 1.0 \pm 0.0 \\ \text{NAN} \\ 1.71 \pm 0.39 \\ 2.68 \pm 0.17 \\ 5.91 \pm 0.93 \end{array}$	$57.91 \pm 24.77 \\ NAN \\ 100.0 \pm 0.0 \\ 99.96 \pm 0.02 \\ 99.96 \pm 0.02 \\ $	$\begin{array}{c} 0.63 \pm 0.26 \\ NAN \\ 0.78 \pm 0.06 \\ 0.78 \pm 0.06 \\ 0.78 \pm 0.06 \end{array}$	$\begin{array}{c} 10.85 \pm 5.45 \\ 100.0 \pm 0.0 \\ 95.4 \pm 9.2 \\ 99.94 \pm 0.07 \\ 99.94 \pm 0.07 \end{array}$	$\begin{array}{c} 1.07 \pm 0.13 \\ 3.79 \pm 1.31 \\ 1.94 \pm 1.2 \\ 3.42 \pm 1.99 \\ 10.38 \pm 7.76 \end{array}$	$\begin{array}{c} 65.14 \pm 15.62 \\ 100.0 \pm 0.0 \\ 99.05 \pm 0.38 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \end{array}$	$\begin{array}{c} 0.55 \pm 0.11 \\ 0.73 \pm 0.07 \\ 0.71 \pm 0.07 \\ 0.73 \pm 0.07 \\ 0.73 \pm 0.07 \end{array}$	$\begin{array}{c} 9.23 \pm 1.24 \\ 100.0 \pm 0.0 \\ 90.6 \pm 3.93 \\ 99.9 \pm 0.0 \\ 99.9 \pm 0.0 \end{array}$	$\begin{array}{c} 1.12 \pm 0.1 \\ 3.57 \pm 1.48 \\ 2.4 \pm 1.38 \\ 0.6 \pm 0.54 \\ 1.63 \pm 0.35 \end{array}$	$\begin{array}{c} 71.55 \pm 3.72 \\ 100.0 \pm 0.0 \\ 95.18 \pm 1.75 \\ 99.92 \pm 0.04 \\ 99.92 \pm 0.04 \end{array}$	$\begin{array}{c} 0.43 \pm 0.04 \\ 0.72 \pm 0.11 \\ 0.6 \pm 0.08 \\ 0.7 \pm 0.11 \\ 0.7 \pm 0.11 \end{array}$
XGB	FAST ARES CET GROUPCF GLOBE-CE DGLOBE-CE	$\begin{array}{c} 6.13 \pm 0.42 \\ \text{NAN} \\ 96.8 \pm 1.72 \\ 82.87 \pm 12.14 \\ 93.76 \pm 1.98 \end{array}$	$1.0 \pm 0.0$ NAN $1.41 \pm 0.54$ $30.1 \pm 10.39$ $64.76 \pm 1.29$	$\begin{array}{c} 31.16\pm8.02\\ NAN\\ 97.16\pm0.41\\ 94.04\pm0.46\\ 94.04\pm0.46\end{array}$	$\begin{array}{c} 0.66 \pm 0.13 \\ NAN \\ 0.81 \pm 0.02 \\ 0.77 \pm 0.01 \\ 0.77 \pm 0.01 \end{array}$	$\begin{array}{c} 31.86 \pm 5.12 \\ 86.29 \pm 9.94 \\ 95.2 \pm 1.6 \\ 82.7 \pm 7.26 \\ 97.47 \pm 0.82 \end{array}$	$\begin{array}{c} 1.05 \pm 0.04 \\ 4.5 \pm 2.64 \\ 1.41 \pm 0.64 \\ 20.82 \pm 1.73 \\ 42.58 \pm 3.57 \end{array}$	$\begin{array}{c} 40.03 \pm 7.82 \\ 89.31 \pm 2.77 \\ 95.73 \pm 0.18 \\ 83.77 \pm 0.84 \\ 97.64 \pm 0.12 \end{array}$	$0.13 \pm 0.04$ $0.73 \pm 0.06$ $1.01 \pm 0.05$ $0.63 \pm 0.08$ $1.37 \pm 0.25$	$\begin{array}{c} 8.49 \pm 1.32 \\ 86.78 \pm 6.7 \\ 78.4 \pm 5.82 \\ 27.66 \pm 5.06 \\ 77.64 \pm 11.51 \end{array}$	$\begin{array}{c} 1.16 \pm 0.13 \\ 12.51 \pm 2.75 \\ 5.63 \pm 1.93 \\ 12.52 \pm 32.48 \\ 128.0 \pm 0.0 \end{array}$	$\begin{array}{c} 9.66 \pm 1.12 \\ 90.72 \pm 3.35 \\ 82.46 \pm 6.55 \\ 28.64 \pm 0.99 \\ 82.46 \pm 6.55 \end{array}$	$\begin{array}{c} 0.4 \pm 0.05 \\ 11.53 \pm 2.94 \\ 8.45 \pm 3.1 \\ 0.85 \pm 0.19 \\ 8.45 \pm 3.1 \end{array}$

### H.3 RESULTS FOR S = 8

Table 10 presents the comparison of our methods, C-GLANCE and T-GLANCE, with other state-ofthe-art methods for a maximum of 8 actions. We distinguish the results of the explicit (T-GLANCE, Fast AReS, CET) and implicit (C-GLANCE, Group-CF, GLOBE-CE, and dGLOBE-CE), how-ever each of our methods can be compared to all competitor GCE methods, regardless of being implicit or explicit. The results indicate that C-GLANCE and T-GLANCE demonstrate superior per-formance in most cases, achieving nearly 100% effectiveness with minimal and stable costs across all datasets and models. Exceptions are noted in the HELOC DNN and XGB experiments, where all methods report higher costs due to the dataset's numeric-only features. In contrast, Fast AReS struggles significantly, particularly on the Adult dataset, showing effectiveness as low as 13.8% for DNN, 12.51% for LR, and 7.19% for XGB, explaining its lower costs. Other methods, such as GroupCF, CET, GLOBE-CE, and dGLOBE-CE, generally perform well in terms of effectiveness but in almost all cases incur higher costs compared to C-GLANCE and T-GLANCE.

Table 10: Evaluating the effectiveness and average cost of T-GLANCE against explicit (Fast AReS, CET) and C-GLANCE against implicit (Group-CF, GLOBE-CE, and dGLOBE-CE) GCE methods for s-GCE problem with s = 8. GCE solutions with effectiveness below 80% are high-lighted in red. Non-robust GCEs, identified by either a standard deviation (std) in effectiveness greater than 5% across folds or a std in cost greater than half the average cost, are highlighted in blue.

MODELS	ALGORITHMS					DAT	FASETS				
		AD	ULT	сом	PAS	DEFAULT	CREDIT	GERMAN	CREDIT	HE	LOC
		eff	avc	eff	avc	eff	avc	eff	avc	eff	avc
DNN	FAST ARES CET T-GLANCE	$13.8 \pm 0.92$ NAN ± NAN $100.0 \pm 0.0$	$1.0 \pm 0.0$ NAN ± NAN $3.77 \pm 0.27$	$63.02 \pm 1.58$ 74.66 ± 5.33 99.65 ± 0.15	$\begin{array}{c} 1.11 \pm 0.04 \\ 1.02 \pm 0.07 \\ 2.18 \pm 0.15 \end{array}$	$\begin{array}{c} 23.75 \pm 1.81 \\ 98.53 \pm 2.11 \\ 100.0 \pm 0.0 \end{array}$	$\frac{1.02 \pm 0.03}{3.68 \pm 1.46}\\1.25 \pm 0.38$	$\begin{array}{c} 65.47 \pm 2.2 \\ 100.0 \pm 0.0 \\ 96.65 \pm 4.08 \end{array}$	$1.0 \pm 0.0$ $1.27 \pm 0.37$ $0.94 \pm 0.3$	$\frac{15.44 \pm 1.91}{99.36 \pm 0.92}$ 99.78 ± 0.21	$1.05 \pm 0.06$ $8.86 \pm 2.99$ $9.92 \pm 1.83$
	GROUPCF	$100.0 \pm 0.0$ $100.0 \pm 0.0$	11.54 ± 2.98	$100.0 \pm 0.0$ $100.0 \pm 0.0$	$5.65 \pm 3.14$	84.8 ± 13.23	$2.15 \pm 0.98$ 2.73 ± 0.85	99.6 ± 0.8	$2.49 \pm 0.82$ 1.89 ± 0.38	$84.8 \pm 12.32$	1.98 ± 0.5
	DGLOBE-CE C-GLANCE	$100.0 \pm 0.0$ $100.0 \pm 0.0$ $100.0 \pm 0.0$	$11.85 \pm 1.53$ $3.96 \pm 0.56$	$100.0 \pm 0.0$ $100.0 \pm 0.0$ $100.0 \pm 0.0$	$7.28 \pm 1.25$ $1.77 \pm 0.16$	99.95 ± 0.04 100.0 ± 0.0	$6.77 \pm 3.67$ $1.2 \pm 0.4$	98.09 ± 1.72 96.43 ± 3.6	$2.34 \pm 0.17$ $0.86 \pm 0.18$	$100.0 \pm 0.0$ 99.98 ± 0.04	$6.65 \pm 0.15$ $9.42 \pm 2.22$
LR	FAST ARES	12.51 ± 2.44	$1.0 \pm 0.0$	$68.19 \pm 0.6$	$1.03 \pm 0.01$	$11.67 \pm 6.18$	$1.04 \pm 0.08$	$71.9 \pm 2.65$	$1.0 \pm 0.0$	$10.95 \pm 1.73$	$1.11 \pm 0.07$
	T-GLANCE	$100.0 \pm 0.0$	0.79 ± 0.16	99.71 ± 0.16	$1.94 \pm 0.11$	100.0 ± 0.0	$1.13 \pm 0.18$	99.55 ± 0.91	$1.34 \pm 0.22$	$100.0 \pm 0.0$ $100.0 \pm 0.0$	$1.52 \pm 0.32$
	GROUPCF	$100.0 \pm 0.0$ $100.0 \pm 0.0$	$2.33 \pm 1.91$ 1.67 ± 0.11	$100.0 \pm 0.0$ 97 74 + 8 52	$4.67 \pm 2.29$ 5 22 + 3 3	$100.0 \pm 0.0$ 99.94 ± 0.07	$3.98 \pm 2.86$ 2.24 ± 0.98	98.9 ± 1.65	$10.56 \pm 2.43$ 2 25 ± 0 33	96.6±1.96	$1.98 \pm 0.5$ 0.58 ± 0.5
	DGLOBE-CE	$100.0 \pm 0.0$	$6.12 \pm 0.64$	$100.0 \pm 0.0$	$6.71 \pm 0.23$	99.94 ± 0.07	$10.38 \pm 7.76$	69.89 ± 15.35	$2.47 \pm 0.23$	99.9 ± 0.0	$1.63 \pm 0.35$
	C-GLANCE	$100.0 \pm 0.0$	$1.03 \pm 0.07$	$100.0 \pm 0.0$	$1.69 \pm 0.06$	100.0 ± 0.0	$1.05 \pm 0.11$	$100.0 \pm 0.0$	$1.18 \pm 0.05$	$100.0 \pm 0.0$	$1.2 \pm 0.12$
XGB	FAST ARES CET	7.19 ± 0.3 NAN ± NAN	$\frac{1.0 \pm 0.0}{\text{NAN} \pm \text{NAN}}$	$66.07 \pm 2.5$ $68.62 \pm 15.67$	$1.14 \pm 0.1$ $1.3 \pm 0.52$	$36.88 \pm 5.46$ 90.01 ± 10.44	$\frac{1.07 \pm 0.06}{2.61 \pm 1.06}$	$\frac{68.88 \pm 1.27}{100.0 \pm 0.0}$	$\frac{1.0 \pm 0.0}{2.18 \pm 0.37}$	$\frac{11.22 \pm 2.05}{90.25 \pm 2.51}$	$\frac{1.09 \pm 0.09}{12.99 \pm 3.49}$
	T-GLANCE	$99.94 \pm 0.04$	$1.65 \pm 0.39$	99.39 ± 0.26	$2.41 \pm 0.4$	94.1±3.57	$1.85 \pm 0.16$	99.46 ± 1.08	$1.07 \pm 0.16$	98.25 ± 1.85	$18.75 \pm 4.92$
	GROUPCF GLOBE-CE	99.4 ± 1.2 85.76 ± 9.11	$1.88 \pm 1.36$ 13 48 + 4 71	$100.0 \pm 0.0$ 89 17 + 11 09	$4.08 \pm 2.15$ 9.78 + 6.35	97.8±0.75 84.94±6.59	$1.92 \pm 1.33$ 11.21 + 1.29	$100.0 \pm 0.0$ 80.25 ± 9.62	$4.05 \pm 2.55$ 2.51 ± 0.33	81.4 ± 4.22 28 88 ± 4 71	$5.14 \pm 1.46$ 12 51 + 30 88
	DGLOBE-CE C-GLANCE	$94.38 \pm 2.01$ 99.87 ± 0.09	$62.37 \pm 3.07$ $3.85 \pm 2.33$	$99.96 \pm 0.08$ $99.83 \pm 0.24$	$9.17 \pm 2.01$ 2.06 ± 0.46	$99.08 \pm 0.66$ $98.87 \pm 0.59$	$28.73 \pm 9.8$ 2.14 ± 0.18	88.17 ± 8.2 100.0 ± 0.0	$2.64 \pm 0.78$ 1.02 ± 0.04	$81.95 \pm 10.38$ 99.0 ± 0.55	$128.0 \pm 0.0$ $16.65 \pm 2.44$

We repeat the experimental evaluation utilizing Pareto-dominance, now for s = 8. As shown in Ta-ble 11 C-GLANCE dominates other methods in 39 out of 72 cases (54%) while is dominated once by GLOBE-CE (in Adult-DNN-cf. Table 10) and once by dGLOBE-CE (in Heloc-DNN). T-GLANCE dominates other methods in 34 out of 72 comparisons, and is dominated one by Group-CF (in Default Credit-XGB) and once by CET (in German Credit-DNN). Overall, the GLANCE methods dominate other solutions in almost half of the cases (48%) and are dominated in only 3%. 

Table 12 summarizes the head-to-head comparisons only when considering practical solutions. GLANCE framework dominates other methods in a much larger percentage—over 62% of the cases.



Figure 4: Comparison of effectiveness and average cost, normalized with the maximum cost achieved in each dataset/model combination) for the solution of s-GCE with s = 8. Standard deviations are represented by error bars. The red horizontal lines represent the eff > 80% threshold for evaluating the practicality of the solution. 

Table 11: Pareto domination evaluation of solutions, for s-GCE problem with s = 8. This table reports the rate (number of times over available comparisons) at which GLANCE methods domi-nate competitors (listed in the DOMINATES column) and the rate at which GLANCE methods are dominated by competitors (listed in the IS DOMINATED column). 

	C-CLANC	CE (IMPLICIT)	T-CLANC	CE (EXPLICIT)
-	Dominates	IS DOMINATED	DOMINATES	IS DOMINATED
FAST ARES (EXPLICIT)	1/15	0/15	2/15	0/15
CET (EXPLICIT)	6/12	0/12	5/12	0/12
GROUPCF (IMPLICIT)	9/15	0/15	7/15	0/15
GLOBE-CE (IMPLICIT)	12/15	1/15	11/15	1/15
DGLOBE-CE (IMPLICIT)	11/15	1/15	9/15	1/15
COMPETITORS (EXPLICIT)			7/27	0/27
COMPETITORS (IMPLICIT)	32/45	2/45		
COMPETITORS (ALL)	39/72	2/72	34/72	2/72

Table 12: Pareto domination evaluation of solutions, for s-GCE problem with s = 8, after applying the eff > 80% threshold. This table reports the rate (number of times over available comparisons) at which GLANCE methods dominate competitors (listed in the DOMINATES column) and the rate at which GLANCE methods are dominated by competitors (listed in the IS DOMINATED column). 

1123		C-CLANO	CE (implicit)	T-CLANC	CE (EXPLICIT)
1124		DOMINATES	IS DOMINATED	DOMINATES	IS DOMINATED
1125	FAST ARES (EXPLICIT)	0/0	0/0	0/0	0/0
1126	CET (EXPLICIT)	6/10	0/10	5/10	0/10
1127	GROUPCF (IMPLICIT) GLOBE-CE (IMPLICIT)	9/15 10/12	0/15 1/12	7/15 9/12	0/15 1/12
1128	DGLOBE-CE (IMPLICIT)	10/14	1/14	8/14	1/14
1129	COMPETITORS (EXPLICIT)			5/10	0/10
1130	COMPETITORS (IMPLICIT)	29/41	2/41	20/51	2/51
1131	COMPETITORS (ALL)	55751	2/31	29731	2131

## It remains dominated in the same four cases, maintaining the percentage of dominated cases at a low 4%.

Finally, when considering only implicit methods, C-GLANCE dominates competitor methods in 1137 71% of cases (32 out of 45), compared to a 54% domination rate across all methods. When focusing 1138 solely on practical solutions, this rate stays the same at 71 % (29 out of 41). For T-GLANCE, the 1139 domination rate against explicit methods is 26% (7 out of 27), lower than the 47% achieved when 1140 considering all methods. This drop is attributed to the high number of impractical solutions provided 1141 by competing explicit methods. However, after removing impractical solutions, T-GLANCE domi-1142 nates 50% of cases, lower than the 57% in head-to-head comparisons across all methods. Overall, 1143 the GLANCE framework demonstrates strong performance, dominating 54% of cases within categories (67% of cases when considering only practical solutions) while keeping the percentage of 1144 dominated cases at a low 2% (4% when considering only practical solutions). 1145

## H.4 COMPERATIVE EVALUATION WITH DIFFERENT COUNTERFACTUAL GENERATION METHODS

Table 13: Evaluation of the effectiveness and cost of C-GLANCE and T-GLANCE, utilizing different candidate action generation methods for *s*-GCE when s = 4. The best metrics are shown in **bold**.

MODELS	ALGORITHMS					DAT	ASETS				
		ADU	JLT	COM	PAS	DEFAULT	CREDIT	GERMAN	CREDIT	HE	LOC
DNN	C-GLANCE - DICE C-GLANCE - NEARESTNEIGHBORS C-GLANCE - NEARESTNEIGHBORSSCALED C-GLANCE - RANDOMSAMPLING	100.0 ± 0.0 100.0 ± 0.0 99.99 ± 0.03 98.01 ± 0.66	$4.6 \pm 0.73 \\ 6.04 \pm 0.35 \\ 5.04 \pm 0.51 \\ 2.03 \pm 0.01$	$100.0 \pm 0.0 \\ 100.0 \\ 100.0 \\ 100.0 \\ 100.0 \\ 100.0 \\ 100.0 \\ 100.0 \\ $	$2.34 \pm 0.43 2.27 \pm 0.31 1.74 \pm 0.38 2.08 \pm 0.21$	$100.0 \pm 0.0 \\ 100.$	1.2 ± 0.4 4.07 ± 0.08 3.02 ± 0.58 1.0 ± 0.0	99.46 ± 1.08 100.0 ± 0.0 100.0 ± 0.0 99.43 ± 1.14	$1.22 \pm 0.41 \\ 4.9 \pm 0.69 \\ 2.1 \pm 0.4 \\ 1.0 \pm 0.2$	99.94 ± 0.05 99.82 ± 0.15 99.79 ± 0.3 94.02 ± 4.28	11.24 ± 1.37 20.18 ± 3.28 18.88 ± 2.11 2.53 ± 1.12
	T-GLANCE - DICE T-GLANCE - NEARESTNEIGHBORS T-GLANCE - NEARESTNEIGHBORSSCALED T-GLANCE - RANDOMSAMPLING	<b>100.0 ± 0.0</b> 99.96 ± 0.08 99.97 ± 0.04 97.57 ± 0.36	$\begin{array}{c} 4.43 \pm 0.43 \\ 9.4 \pm 1.65 \\ 5.92 \pm 1.06 \\ \textbf{2.91} \pm \textbf{0.05} \end{array}$	$\begin{array}{c} 99.53 \pm 0.22 \\ 99.52 \pm 0.29 \\ \textbf{99.88} \pm \textbf{0.16} \\ 99.65 \pm 0.08 \end{array}$	$\begin{array}{c} 2.62 \pm 0.24 \\ 2.71 \pm 0.19 \\ 2.47 \pm 0.35 \\ \textbf{2.28 \pm 0.4} \end{array}$	<b>100.0 ± 0.0</b> 99.95 ± 0.06 99.73 ± 0.37 99.97 ± 0.05	$\begin{array}{c} 1.48 \pm 0.43 \\ 4.13 \pm 0.38 \\ 3.22 \pm 0.32 \\ \textbf{1.0} \pm \textbf{0.0} \end{array}$	$\begin{array}{c} 96.97 \pm 3.48 \\ \textbf{100.0} \pm \textbf{0.0} \\ \textbf{100.0} \pm \textbf{0.0} \\ \textbf{99.53} \pm 0.93 \end{array}$	$\begin{array}{c} 1.59 \pm 0.8 \\ 8.54 \pm 1.9 \\ 2.35 \pm 0.44 \\ \textbf{1.06} \pm \textbf{0.46} \end{array}$	<b>99.8 ± 0.24</b> 93.49 ± 2.03 95.32 ± 1.19 95.9 ± 1.85	$\begin{array}{c} 10.9 \pm 1.22 \\ 17.71 \pm 2.06 \\ 18.97 \pm 2.08 \\ \textbf{6.45 \pm 1.31} \end{array}$
LR	C-GLANCE - DICE C-GLANCE - NEARESTNEIGHBORS C-GLANCE - NEARESTNEIGHBORSSCALED C-GLANCE - RANDOMSAMPLING	$\begin{array}{c} 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \end{array}$	$\begin{array}{c} 1.04 \pm 0.07 \\ 3.76 \pm 0.96 \\ 3.19 \pm 0.51 \\ \textbf{0.72 \pm 0.12} \end{array}$	$\begin{array}{c} 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0 \end{array}$	$\begin{array}{c} 2.33 \pm 0.38 \\ 2.53 \pm 0.32 \\ \textbf{2.14} \pm \textbf{0.15} \\ 2.21 \pm 0.2 \end{array}$	$\begin{array}{c} 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 99.94 \pm 0.12 \end{array}$	$\begin{array}{c} 1.05 \pm 0.11 \\ 1.78 \pm 0.39 \\ 1.11 \pm 0.14 \\ \textbf{1.0} \pm \textbf{0.0} \end{array}$	$\begin{array}{c} 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0 \end{array}$	1.21 ± 0.06 4.21 ± 0.56 2.02 ± 0.27 1.31 ± 0.15	$\begin{array}{c} \textbf{100.0 \pm 0.0} \\ 99.94 \pm 0.05 \\ 99.94 \pm 0.12 \\ \textbf{100.0 \pm 0.0} \end{array}$	$\begin{array}{c} 1.55 \pm 0.54 \\ 21.52 \pm 1.33 \\ 20.95 \pm 3.47 \\ \textbf{1.4 \pm 0.27} \end{array}$
	T-GLANCE - DICE T-GLANCE - NEARESTNEIGHBORS T-GLANCE - NEARESTNEIGHBORSSCALED T-GLANCE - RANDOMSAMPLING	<b>100.0 ± 0.0</b> 99.92 ± 0.04 98.64 ± 1.3 <b>100.0 ± 0.0</b>	0.69 ± 0.01 3.52 ± 1.37 2.98 ± 0.88 0.71 ± 0.12	$\begin{array}{c} 99.79 \pm 0.22 \\ 99.79 \pm 0.22 \\ \textbf{99.83 \pm 0.15} \\ 99.71 \pm 0.16 \end{array}$	2.19 ± 0.1 3.51 ± 0.19 3.45 ± 0.25 2.3 ± 0.28	$\begin{array}{c} 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 99.97 \pm 0.06 \end{array}$	$\begin{array}{c} 1.43 \pm 0.36 \\ 2.98 \pm 0.28 \\ 2.1 \pm 0.18 \\ \textbf{1.04 \pm 0.07} \end{array}$	99.58 ± 0.83 100.0 ± 0.0 100.0 ± 0.0 98.71 ± 1.05	$\begin{array}{c} 1.54 \pm 0.32 \\ 9.66 \pm 1.55 \\ 2.14 \pm 0.26 \\ \textbf{1.46 \pm 0.25} \end{array}$	<b>100.0 ± 0.0</b> 99.31 ± 0.92 99.54 ± 0.68 <b>100.0 ± 0.0</b>	$\begin{array}{c} 1.58 \pm 0.37 \\ 19.48 \pm 2.53 \\ 18.15 \pm 2.02 \\ \textbf{1.57 \pm 0.66} \end{array}$
XGB	C-GLANCE - DICE C-GLANCE - NEARESTNEIGHBORS C-GLANCE - NEARESTNEIGHBORSSCALED C-GLANCE - RANDOMSAMPLING	$\begin{array}{c} 99.85 \pm 0.12 \\ \textbf{100.0} \pm \textbf{0.01} \\ 99.98 \pm 0.03 \\ 99.85 \pm 0.19 \end{array}$	$5.98 \pm 4.22 4.53 \pm 0.56 4.12 \pm 0.31 2.1 \pm 0.19$	99.51 ± 0.46 99.6 ± 0.34 99.44 ± 0.54 99.88 ± 0.16	$\begin{array}{c} 2.96 \pm 0.82 \\ 2.95 \pm 0.58 \\ 3.0 \pm 0.33 \\ \textbf{2.83 \pm 0.76} \end{array}$	98.13 ± 1.05 100.0 ± 0.0 100.0 ± 0.0 91.52 ± 4.44	$3.68 \pm 1.64$ $3.8 \pm 0.44$ $3.09 \pm 0.42$ $2.25 \pm 1.06$	$\begin{array}{c} 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0\\ 100.0\pm 0.0\\ \end{array}$	1.05 ± 0.02 3.67 ± 0.58 1.75 ± 0.2 1.14 ± 0.07	$\begin{array}{c} 98.94 \pm 0.66 \\ 99.51 \pm 0.42 \\ \textbf{99.68} \pm \textbf{0.13} \\ 82.75 \pm 5.2 \end{array}$	$\begin{array}{c} 19.99 \pm 1.91 \\ 19.3 \pm 2.66 \\ 20.92 \pm 3.96 \\ \textbf{9.6 \pm 1.84} \end{array}$
	T-GLANCE - DICE T-GLANCE - NEARESTNEIGHBORS T-GLANCE - NEARESTNEIGHBORSSCALED T-GLANCE - RANDOMSAMPLING	$99.86 \pm 0.14$ $99.87 \pm 0.22$ $99.68 \pm 0.25$ $99.61 \pm 0.66$	1.8 ± 0.51 9.44 ± 1.71 9.04 ± 1.75 2.03 ± 0.3	99.02 ± 0.93 98.99 ± 0.54 99.24 ± 0.67 98.67 ± 1.0	$2.6 \pm 0.46$ $3.25 \pm 0.33$ $3.25 \pm 0.41$ $2.53 \pm 0.34$	94.6 ± 3.13 100.0 ± 0.0 100.0 ± 0.0 87.04 ± 7.36	$2.75 \pm 1.42$ $4.35 \pm 0.47$ $3.93 \pm 0.53$ $2.18 \pm 1.65$	99.46 ± 1.08 100.0 ± 0.0 100.0 ± 0.0 99.33 ± 1.33	$1.33 \pm 0.42$ $7.8 \pm 1.26$ $2.35 \pm 0.36$ $1.23 \pm 0.21$	96.42 ± 2.45 94.94 ± 1.91 96.07 ± 1.85 79.63 ± 7.83	$24.85 \pm 8.4$ $18.4 \pm 1.41$ $20.39 \pm 1.69$ $8.02 \pm 1.43$

Table 14: Evaluation of the effectiveness and cost of C-GLANCE and T-GLANCE, utilizing different candidate action generation methods for s-GCE when s = 8. The best metrics are shown in bold.

MODELS	ALGORITHMS					DAT/	SETS				
		ADU	JLT avc	COM eff	IPAS avc	DEFAULT eff	CREDIT	GERMAN eff	CREDIT	HE eff	LOC
DNN	C-GLANCE - DICE C-GLANCE - NEARESTNEIGHBORS C-GLANCE - NEARESTNEIGHBORSSCALED C-GLANCE - RANDOMSAMPLING	$\begin{array}{c} 100.0 \pm 0.0 \\ 100.0 \pm 0.01 \\ 100.0 \pm 0.0 \\ 95.36 \pm 3.8 \end{array}$	$\begin{array}{c} 3.96 \pm 0.56 \\ 4.99 \pm 0.62 \\ 4.44 \pm 0.15 \\ \textbf{2.84 \pm 0.8} \end{array}$	$ \begin{array}{c} 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \end{array} $	$\begin{array}{c} 1.77 \pm 0.16 \\ 1.71 \pm 0.18 \\ \textbf{1.08} \pm \textbf{0.08} \\ 1.74 \pm 0.22 \end{array}$	$100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 99.82 \pm 0.29$	$\begin{array}{c} 1.2 \pm 0.4 \\ 3.84 \pm 0.26 \\ 2.5 \pm 0.24 \\ \textbf{1.07 \pm 0.14} \end{array}$	$\begin{array}{c} 99.46 \pm 1.08 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \end{array}$	$\begin{array}{c} 0.78 \pm 0.27 \\ 3.94 \pm 0.39 \\ 1.65 \pm 0.14 \\ \textbf{0.7 \pm 0.18} \end{array}$	<b>99.98 ± 0.04</b> 99.75 ± 0.26 99.71 ± 0.15 98.55 ± 0.45	9.42 ± 2.2 16.07 ± 1. 16.17 ± 1. 10.01 ± 0.
	T-GLANCE - DICE T-GLANCE - NEARESTNEIGHBORS T-GLANCE - NEARESTNEIGHBORSSCALED T-GLANCE - RANDOMSAMPLING	$\begin{array}{c} \textbf{100.0 \pm 0.0} \\ 99.95 \pm 0.08 \\ 99.81 \pm 0.16 \\ 96.76 \pm 0.39 \end{array}$	$\begin{array}{c} 3.77 \pm 0.27 \\ 7.71 \pm 1.11 \\ 7.56 \pm 1.74 \\ \textbf{2.87 \pm 0.06} \end{array}$	$\begin{array}{c} 99.65 \pm 0.15 \\ 99.79 \pm 0.14 \\ \textbf{99.88} \pm \textbf{0.1} \\ 99.58 \pm 0.22 \end{array}$	$\begin{array}{c} 2.19 \pm 0.15 \\ 2.53 \pm 0.17 \\ 2.24 \pm 0.39 \\ \textbf{2.18 \pm 0.45} \end{array}$	$\begin{array}{c} \mathbf{100.0 \pm 0.0} \\ \mathbf{100.0 \pm 0.0} \\ 99.97 \pm 0.05 \\ \mathbf{100.0 \pm 0.0} \end{array}$	$\begin{array}{c} 1.25 \pm 0.38 \\ 4.18 \pm 0.29 \\ 3.47 \pm 0.72 \\ \textbf{1.16 \pm 0.31} \end{array}$	$\begin{array}{c} 96.65 \pm 4.08 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \end{array}$	$\begin{array}{c} 0.94 \pm 0.3 \\ 7.54 \pm 0.97 \\ 2.11 \pm 0.57 \\ \textbf{0.86 \pm 0.38} \end{array}$	<b>99.78 ± 0.21</b> 96.48 ± 0.9 97.48 ± 0.89 96.76 ± 4.58	9.42 ± 1. 19.11 ± 1 20.98 ± 1 7.7 ± 2.5
LR	C-GLANCE - DICE C-GLANCE - NEARESTNEIGHBORS C-GLANCE - NEARESTNEIGHBORSSCALED C-GLANCE - RANDOMSAMPLING	$\begin{array}{c} 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \end{array}$	$\begin{array}{c} 1.03 \pm 0.07 \\ 3.22 \pm 0.49 \\ 2.96 \pm 0.62 \\ \textbf{0.67} \pm \textbf{0.07} \end{array}$	$\begin{array}{c} 100.0\pm0.0\\ 100.0\pm0.0\\ 100.0\pm0.0\\ 100.0\pm0.0\\ \end{array}$	1.69 ± 0.06 1.9 ± 0.12 1.81 ± 0.18 1.78 ± 0.25	$\begin{array}{c} 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 99.94 \pm 0.12 \end{array}$	$\begin{array}{c} 1.05 \pm 0.11 \\ 1.39 \pm 0.11 \\ 1.01 \pm 0.0 \\ \textbf{1.0 \pm 0.0} \end{array}$	$\begin{array}{c} 100.0\pm0.0\\ 100.0\pm0.0\\ 100.0\pm0.0\\ 100.0\pm0.0\\ 100.0\pm0.0 \end{array}$	1.15 ± 0.03 3.98 ± 0.55 1.7 ± 0.2 1.19 ± 0.05	100.0 ± 0.0 99.88 ± 0.14 99.88 ± 0.15 100.0 ± 0.0	1.2 ± 0. 17.3 ± 1 17.39 ± 0 <b>1.19 ± 0</b>
	T-GLANCE - DICE T-GLANCE - NEARESTNEIGHBORS T-GLANCE - NEARESTNEIGHBORSSCALED T-GLANCE - RANDOMSAMPLING	<b>100.0 ± 0.0</b> 99.92 ± 0.04 98.64 ± 1.3 <b>100.0 ± 0.0</b>	$\begin{array}{c} 0.69 \pm 0.01 \\ 3.52 \pm 1.37 \\ 2.98 \pm 0.88 \\ \textbf{0.63 \pm 0.09} \end{array}$	99.71 ± 0.16 99.79 ± 0.22 99.83 ± 0.15 99.71 ± 0.16	1.94 ± 0.11 3.51 ± 0.19 3.45 ± 0.25 2.46 ± 0.32	100.0 ± 0.0 100.0 ± 0.0 100.0 ± 0.0 99.97 ± 0.06	$\begin{array}{c} 1.13 \pm 0.18 \\ 2.98 \pm 0.28 \\ 2.1 \pm 0.18 \\ \textbf{1.02 \pm 0.02} \end{array}$	$\begin{array}{c} 99.55 \pm 0.91 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \\ 100.0 \pm 0.0 \end{array}$	$\begin{array}{c} 1.34 \pm 0.22 \\ 7.47 \pm 2.24 \\ 1.82 \pm 0.29 \\ \textbf{1.27} \pm \textbf{0.17} \end{array}$	100.0 ± 0.0 99.31 ± 0.92 99.54 ± 0.68 100.0 ± 0.0	1.52 ± 0 19.48 ± 2 18.15 ± 2 1.62 ± 0
XGB	C-GLANCE - DICE C-GLANCE - NEARESTNEIGHBORS C-GLANCE - NEARESTNEIGHBORSSCALED C-GLANCE - RANDOMSAMPLING	$99.87 \pm 0.09$ $100.0 \pm 0.0$ $99.98 \pm 0.03$ $99.76 \pm 0.37$	$3.85 \pm 2.33$ $4.43 \pm 0.55$ $3.9 \pm 0.41$ $2.13 \pm 0.17$	$\begin{array}{c} 99.83 \pm 0.24 \\ 99.6 \pm 0.34 \\ 99.48 \pm 0.57 \\ \textbf{99.84 \pm 0.24} \end{array}$	$\begin{array}{c} 2.06 \pm 0.46 \\ \textbf{1.8} \pm \textbf{0.16} \\ 1.88 \pm 0.09 \\ 2.06 \pm 0.21 \end{array}$	98.87 ± 0.59 100.0 ± 0.0 100.0 ± 0.0 93.86 ± 2.69	$\begin{array}{c} 2.14 \pm 0.18 \\ 3.2 \pm 0.45 \\ 2.4 \pm 0.31 \\ \textbf{1.51 \pm 0.52} \end{array}$	$\begin{array}{c} 100.0\pm0.0\\ 100.0\pm0.0\\ 100.0\pm0.0\\ 100.0\pm0.0\\ \end{array}$	0.97 ± 0.08 3.63 ± 0.58 1.62 ± 0.13 1.06 ± 0.1	99.0 ± 0.55 99.49 ± 0.47 99.72 ± 0.17 84.76 ± 4.77	16.65 ± 2 15.76 ± 1 18.58 ± 1 <b>7.11 ± 1</b>
	T-GLANCE - DICE T-GLANCE - NEARESTNEIGHBORS T-GLANCE - NEARESTNEIGHBORSSCALED T-GLANCE - NEARESTNEIGHBORSSCALED	99.94 ± 0.04 99.87 ± 0.22 99.68 ± 0.25	$1.65 \pm 0.39$ 9.44 ± 1.71 9.04 ± 1.75	99.39 ± 0.26 98.99 ± 0.54 99.24 ± 0.67	$2.41 \pm 0.4$ 3.25 ± 0.33 3.25 ± 0.41	94.1 ± 3.57 100.0 ± 0.0 100.0 ± 0.0	$1.85 \pm 0.16$ $4.35 \pm 0.47$ $3.93 \pm 0.53$ $2.55 \pm 1.20$	99.46 ± 1.08 100.0 ± 0.0 100.0 ± 0.0	1.07 ± 0.16 5.39 ± 1.33 2.09 ± 0.28	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	18.75 ± 4 18.4 ± 1 20.39 ± 1

1184

1146

1149

1152

1166 1167

1168

1169 1170

1185Tables 13 and 14 present a comprehensive comparison of C-GLANCE and T-GLANCE utilizing the<br/>counterfactual generation methods described in Appendix G, i.e., DiCE, NearestNeighbors, Near-<br/>estNeighborsScaled, and RandomSampling, for the s-GCE problem with s = 4 and s = 8, respec-<br/>tively. Across all datasets and models, both C-GLANCE and T-GLANCE consistently exhibit superb

1188 performance in terms of effectiveness, frequently achieving near 100%. Since the performance in 1189 terms of effectiveness is comparable, we will focus on the cost our algorithms achieved under the 1190 different generation methods. The best performance is observed with the RandomSampling method, 1191 achieving the best results in 41 out of 60 cases. DiCE, which is regarded as state-of-the-art, fol-1192 lows, achieves the lowest cost in 15 out of 60 cases. NearestNeighbors and NearestNeighborsScaled achieve the best cost in only 1 out of 60 and 3 out of 60 cases respectively. In general, C-GLANCE 1193 and T-GLANCE, can demonstrate great performance under all action generation methods, maintain-1194 ing high effectiveness and low costs in almost all cases. 1195

- 1196
- H.5 DETAILED RESULTS 1197
- 1198

Table 15: Detailed results for the solution of s-GCE (s = 4 and s = 8) for COMPAS dataset. 1199 The table reports effectiveness, cost, size, and runtime, including their standard deviations for each 1200 method. 1201

METHOD / MODEL		DN	Ň			LR				XGE		
	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME
FAST ARES-4	$55.0 \pm 0.86$	$1.21 \pm 0.09$	4	59.26 ± 26.71	$62.5 \pm 1.82$	$1.24 \pm 0.14$	4	$11.52 \pm 5.39$	59.83 ± 3.12	$1.1 \pm 0.05$	4	$12.88 \pm 5.5$
CET - 4	$63.62 \pm 10.35$	$0.96 \pm 0.24$	7.0	1339.42 ± 247.04	$73.18 \pm 4.34$	$1.24 \pm 0.15$	4.0	$663.22 \pm 45.34$	$58.4 \pm 9.3$	$1.06 \pm 0.24$	3.0	3280.17 ± 834.28
GROUPCF - 4	$100.0 \pm 0.0$	$4.48 \pm 2.53$	4	$173.24 \pm 2.53$	$100.0 \pm 0.0$	$3.97 \pm 2.38$	4	$45.04 \pm 2.38$	$100.0 \pm 0.0$	$4.06 \pm 2.1$	4	$36.07 \pm 2.1$
GLOBE-CE - 4	$100.0 \pm 0.0$	$4.54 \pm 3.31$	3.0	$1.09 \pm 0.06$	$95.74 \pm 8.52$	$5.14 \pm 3.77$	2.6	$0.42 \pm 0.01$	$87.13 \pm 11.14$	$9.75 \pm 7.2$	1.6	$1.06 \pm 0.01$
DGLOBE-CE - 4	$100.0 \pm 0.0$	$7.96 \pm 3.91$	3.4	$1.55 \pm 0.07$	$100.0 \pm 0.0$	$6.71 \pm 0.23$	4.0	$0.61 \pm 0.01$	$99.84 \pm 0.31$	$12.46 \pm 3.42$	3.4	$1.53 \pm 0.03$
C-GLANCE-4	$100.0 \pm 0.0$	$2.34 \pm 0.43$	4.0	$177.41 \pm 0.0$	$100.0 \pm 0.0$	$2.33 \pm 0.38$	4.0	$82.09 \pm 0.0$	$99.51 \pm 0.46$	$2.96 \pm 0.82$	4.0	$286.26 \pm 0.0$
T-GLANCE-4	$99.53 \pm 0.22$	$2.62\pm0.24$	3.0	$72.01 \pm 0.0$	$99.79 \pm 0.22$	$2.19 \pm 0.1$	3.0	$50.14 \pm 0.0$	$99.02 \pm 0.93$	$2.6 \pm 0.46$	3.0	$156.0 \pm 0.0$
FAST ARES-8	$63.02 \pm 1.58$	$1.11 \pm 0.04$	8	65.84 ± 29.93	$68.19 \pm 0.6$	$1.03 \pm 0.01$	8	$15.2 \pm 6.9$	66.07 ± 2.5	$1.14 \pm 0.1$	8	$17.69 \pm 8.05$
CET - 8	74.66 ± 5.33	$1.02 \pm 0.07$	8.0	656.64 ± 119.51	$82.51 \pm 2.97$	$1.51 \pm 0.23$	6.0	$442.76 \pm 64.47$	$68.62 \pm 15.67$	$1.3 \pm 0.52$	6.0	1787.18 ± 360.63
GROUPCF - 8	$100.0 \pm 0.0$	$5.65 \pm 3.14$	8	$210.34 \pm 3.68$	$100.0 \pm 0.0$	$4.67 \pm 2.29$	8	$48.53 \pm 2.29$	$100.0 \pm 0.0$	$4.08 \pm 2.15$	8	$37.53 \pm 2.15$
GLOBE-CE - 8	$100.0 \pm 0.0$	$4.31 \pm 2.84$	6.8	$1.06 \pm 0.04$	$97.74 \pm 8.52$	$5.22 \pm 3.3$	5.8	$0.45 \pm 0.02$	89.17 ± 11.09	$9.78 \pm 6.35$	3.8	$1.08 \pm 0.03$
DGLOBE-CE - 8	$100.0 \pm 0.0$	$7.28 \pm 1.25$	6.8	$2.3 \pm 0.09$	$100.0 \pm 0.0$	$6.71 \pm 0.23$	8.0	$0.88 \pm 0.01$	$99.96 \pm 0.08$	$9.17 \pm 2.01$	5.6	$2.23 \pm 0.03$
C-GLANCE-8	$100.0 \pm 0.0$	$1.77 \pm 0.16$	8.0	$177.35 \pm 0.0$	$100.0 \pm 0.0$	$1.69 \pm 0.06$	8.0	$84.11 \pm 0.0$	$99.83 \pm 0.24$	$2.06 \pm 0.46$	8.0	$173.1 \pm 0.0$
T-GLANCE-8	$99.65 \pm 0.15$	$2.18 \pm 0.15$	4.0	$21.48 \pm 0.64$	$99.71 \pm 0.16$	$1.94 \pm 0.11$	4.0	$13.06 \pm 0.46$	$99.39 \pm 0.26$	$2.41 \pm 0.4$	4.0	$40.05 \pm 4.5$

1211

1215

1212 Table 15 presents a comparative analysis of all algorithms for the COMPAS dataset, revealing dis-1213 tinct performance patterns. CET and Fast AReS, both with 4 and 8 actions, show very low costs 1214 across all models; however, this is largely attributable to their relatively low effectiveness scores.

1216 Methods like GLOBE-CE, dGLOBE-CE display significantly higher effectiveness, mostly above 1217 90% across all models, GroupCF yields perfect effectiveness of 100%. However, these methods come with higher costs, indicating that increased effectiveness is often tied to a rise in cost. 1218

1219 On the other hand, C-GLANCE and T-GLANCE methods consistently achieve near-perfect or perfect 1220 effectiveness ( $\sim 100\%$ ) across all models while maintaining much lower costs. This balance of 1221 high effectiveness and low cost makes C-GLANCE and T-GLANCE the optimal methods for the 1222 COMPAS dataset.

1223 Table 16: Detailed results for the solution of s-GCE (s = 4 and s = 8) for German Credit dataset. 1224 The table reports effectiveness, cost, size, and runtime, including their standard deviations for each 1225 method. 1226

METHOD / MODEL		DN	Ň			LR				XGB		
	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME
FAST ARES-4	$52.39 \pm 1.63$	$1.0 \pm 0.0$	4	$119.34 \pm 54.6$	$75.27 \pm 2.96$	$1.0 \pm 0.0$	4	$47.88 \pm 23.06$	$51.27 \pm 1.57$	$1.0 \pm 0.0$	4	$55.01 \pm 27.2$
CET - 4	$97.3 \pm 2.46$	$1.58 \pm 0.54$	3.0	230.36 ± 47.15	$96.5 \pm 2.85$	$2.42 \pm 0.24$	3.0	237.17 ± 11.2	$100.0 \pm 0.0$	$2.73 \pm 0.49$	3.0	$347.41 \pm 30.2$
GROUPCF - 4	$97.8 \pm 4.4$	$1.85 \pm 0.13$	4	$25.54 \pm 0.13$	$97.6 \pm 2.94$	$9.34 \pm 3.85$	4	$7.91 \pm 3.85$	$100.0 \pm 0.0$	$5.78 \pm 4.11$	4	$5.87 \pm 4.11$
GLOBE-CE - 4	$93.31 \pm 3.48$	$2.0 \pm 1.55$	2.0	$1.02 \pm 0.01$	$57.09 \pm 20.03$	$0.75 \pm 1.04$	1.0	$0.44 \pm 0.0$	$77.05 \pm 11.26$	$1.14 \pm 1.24$	1.0	$1.34 \pm 0.03$
DGLOBE-CE - 4	$97.36 \pm 0.82$	$2.49 \pm 0.27$	4.0	$1.57 \pm 0.01$	$69.89 \pm 15.35$	$2.47 \pm 0.23$	3.2	$0.7 \pm 0.01$	$86.96 \pm 9.79$	$2.66 \pm 0.77$	3.6	$2.09 \pm 0.04$
C-GLANCE-4	95.31 ± 3.15	$1.25 \pm 0.33$	4.0	$78.45 \pm 0.0$	$100.0 \pm 0.0$	$1.21 \pm 0.06$	4.0	$61.58 \pm 0.0$	$100.0 \pm 0.0$	$1.06 \pm 0.03$	4.0	$67.36 \pm 0.0$
T-GLANCE-4	$96.97 \pm 3.48$	$1.59 \pm 0.8$	3.0	$85.12 \pm 0.0$	$99.58 \pm 0.83$	$1.54 \pm 0.32$	3.0	$75.64 \pm 0.0$	$99.46 \pm 1.08$	$1.33 \pm 0.42$	3.0	$185.11 \pm 0.0$
FAST ARES-8	$65.47 \pm 2.2$	$1.0 \pm 0.0$	8	$415.74 \pm 175.82$	$71.9 \pm 2.65$	$1.0 \pm 0.0$	8	59.8 ± 32.65	$68.88 \pm 1.27$	$1.0 \pm 0.0$	8	64.77 ± 29.6
CET - 8	$100.0 \pm 0.0$	$1.27 \pm 0.37$	1.0	$142.29 \pm 27.72$	$98.75 \pm 1.67$	$2.02 \pm 0.33$	6.0	$156.14 \pm 15.7$	$100.0 \pm 0.0$	$2.18 \pm 0.37$	5.0	$239.2 \pm 25.8$
GROUPCF - 8	$99.6 \pm 0.8$	$2.49 \pm 0.82$	8	$28.93 \pm 0.82$	$98.9 \pm 1.65$	10.56± 2.43	8	$15.6 \pm 2.33$	$100.0 \pm 0.0$	$4.05 \pm 2.55$	8	$7.84 \pm 2.55$
GLOBE-CE - 8	$95.37 \pm 2.9$	$1.89 \pm 0.38$	2.0	$1.03 \pm 0.01$	$58.42 \pm 18.45$	$2.25 \pm 0.33$	1.0	$0.45 \pm 0.0$	$80.25 \pm 9.62$	$2.51 \pm 0.33$	1.0	$1.36 \pm 0.05$
DGLOBE-CE - 8	$98.09 \pm 1.72$	$2.34 \pm 0.17$	7.8	$2.34 \pm 0.03$	$69.89 \pm 15.35$	$2.47 \pm 0.23$	5.6	$1.02 \pm 0.04$	$88.17 \pm 8.2$	$2.64 \pm 0.78$	6.2	$3.06 \pm 0.04$
C-GLANCE-8	$96.43 \pm 3.6$	$0.86 \pm 0.18$	8.0	$78.38 \pm 0.0$	$100.0 \pm 0.0$	$1.18 \pm 0.05$	8.0	$63.1 \pm 0.0$	$100.0 \pm 0.0$	$1.02 \pm 0.04$	8.0	$68.96 \pm 0.0$
I-GLANCE-8	$96.65 \pm 4.08$	$0.94 \pm 0.3$	4.0	$32.19 \pm 1.34$	$99.55 \pm 0.91$	$1.34 \pm 0.22$	3.8	$29.11 \pm 1.18$	$99.46 \pm 1.08$	$1.07 \pm 0.16$	3.8	$68.67 \pm 1.5$

1236

1237 Table 16 presents a comparative analysis of all algorithms for the German Credit dataset. Fast 1238 AReS remains the weakest method overall, continuing to exhibit poor effectiveness. Notably, while CET incurs a slightly higher cost than on the COMPAS dataset, it achieves significantly greater ef-1239 fectiveness, and the increase in cost is quite reasonable. The GLOBE-CE and dGLOBE-CE methods 1240 demonstrate mixed results. They perform relatively well under the DNN model, achieving effective-1241 ness scores above 90%, but their effectiveness declines when applied to the LR and XGB models. On 1242 the other hand, T-GLANCE and C-GLANCE continue to excel due to their robustness and consistent 1243 high performance. These methods consistently reach near-perfect effectiveness ( $\sim 100\%$ , and only 1244 for DNN 95%-97%) across all models while maintaining very low costs. Notably, C-GLANCE-8 1245 achieves the lowest cost of 0.86 under the DNN model, while still maintaining high effectiveness.

Table 17: Detailed results for the solution of s-GCE (s = 4 and s = 8) for Default Credit dataset. 1247 The table reports effectiveness, cost, size, and runtime, including their standard deviations for each 1248 method. 1249

METHOD / MODEL		DNI	Ň			LR				XGE	3	
	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME
FAST ARES-4	$18.88 \pm 2.16$	$1.0 \pm 0.0$	4	789.48 ± 306.69	$10.85 \pm 5.45$	$1.07 \pm 0.13$	4	$244.78 \pm 128.93$	$31.86 \pm 5.12$	$1.05 \pm 0.04$	4	226.08 ± 107.9
CET - 4	$98.87 \pm 0.62$	$6.32 \pm 2.28$	3.0	4404.58 ± 453.7	$100.0 \pm 0.0$	$3.79 \pm 1.31$	1.0	3711.21 ± 164.94	$86.29 \pm 9.94$	$4.5 \pm 2.64$	3.0	7982.05 ± 668.7
GROUPCF - 4	79.6 ± 20.79	$1.53 \pm 0.62$	4	$2415.7 \pm 0.62$	$95.4 \pm 9.2$	$1.94 \pm 1.2$	4	$306.76 \pm 1.2$	$95.2 \pm 1.6$	$1.41 \pm 0.64$	4	$991.53 \pm 0.64$
GLOBE-CE-4	76.94 ± 37.55	$5.14 \pm 0.35$	1.4	$3.33 \pm 0.2$	$99.94 \pm 0.07$	$3.42 \pm 1.99$	1.6	$1.76 \pm 0.04$	$82.7 \pm 7.26$	$20.82 \pm 1.73$	2.6	$2.44 \pm 0.05$
DGLOBE-CE - 4	$87.38 \pm 18.69$	$5.96 \pm 4.14$	3.4	$4.98 \pm 0.21$	$99.94 \pm 0.07$	$10.38 \pm 7.76$	1.2	$2.54 \pm 0.07$	$97.47 \pm 0.82$	$42.58 \pm 3.57$	4.0	$3.48 \pm 0.09$
C-GLANCE-4	$100.0 \pm 0.0$	$1.2 \pm 0.4$	4.0	$290.89 \pm 0.0$	$100.0 \pm 0.0$	$1.05 \pm 0.11$	4.0	$243.45 \pm 0.0$	$98.13 \pm 1.05$	$3.68 \pm 1.64$	4.0	$319.45 \pm 0.0$
T-GLANCE-4	$100.0 \pm 0.0$	$1.48\pm0.43$	3.0	$775.09 \pm 0.0$	$100.0 \pm 0.0$	$1.43 \pm 0.36$	3.0	$876.08 \pm 0.0$	$94.6 \pm 3.13$	$2.75 \pm 1.42$	3.0	$688.19 \pm 0.0$
FAST ARES-8	$23.75 \pm 1.81$	$1.02 \pm 0.03$	8	428.04 ± 199.29	$11.67 \pm 6.18$	$1.04 \pm 0.08$	8	272.64 ± 120.89	$36.88 \pm 5.46$	$1.07 \pm 0.06$	8	280.88 ± 135.3
CET - 8	$98.53 \pm 2.11$	$3.68 \pm 1.46$	4.0	$2500.02 \pm 328.94$	$100.0 \pm 0.0$	$2.0 \pm 1.36$	2.0	1795.08 ± 182.55	$90.01 \pm 10.44$	$2.61 \pm 1.06$	5.0	5105.11 ± 923.2
GROUPCF - 8	$84.8 \pm 13.23$	$2.15 \pm 0.98$	8	$2476.21 \pm 0.82$	$100.0 \pm 0.0$	$3.98 \pm 2.86$	8	$279.86 \pm 2.86$	$97.8 \pm 0.75$	$1.92 \pm 1.33$	8	631.31 ± 1.33
GLOBE-CE - 8	93.25 ± 11.26	$2.73 \pm 0.85$	3.6	$3.49 \pm 0.27$	$99.94 \pm 0.07$	$2.24 \pm 0.98$	1.8	$1.75 \pm 0.05$	$84.94 \pm 6.59$	$11.21 \pm 1.29$	5.2	$2.47 \pm 0.06$
DGLOBE-CE - 8	$99.95 \pm 0.04$	6.77 ± 3.67	4.8	$6.79 \pm 0.16$	$99.94 \pm 0.07$	$10.38 \pm 7.76$	1.2	$3.51 \pm 0.07$	$99.08 \pm 0.66$	$28.73 \pm 9.8$	8.0	$5.05 \pm 0.12$
C-GLANCE-8	$100.0 \pm 0.0$	$1.2 \pm 0.4$	8.0	$289.46 \pm 0.0$	$100.0 \pm 0.0$	$1.05 \pm 0.11$	8.0	$226.96 \pm 0.0$	$98.87 \pm 0.59$	$2.14 \pm 0.18$	8.0	$257.39 \pm 0.0$
T-GLANCE-8	$100.0 \pm 0.0$	$1.25 \pm 0.38$	3.8	$177.22 \pm 0.46$	$100.0 \pm 0.0$	$1.13 \pm 0.18$	3.8	$167.72 \pm 7.88$	$94.1 \pm 3.57$	$1.85 \pm 0.16$	4.0	191.86 ± 4.82

Table 17 provides a comparative analysis of all algorithms for the Default Credit dataset. 1261 C-GLANCE and T-GLANCE methods exhibit perfect effectiveness scores across both the DNN and 1262 LR models while maintaining the lowest costs, second only to Fast AReS, highlighting their su-1263 perior cost-effectiveness.

1264 In the XGB model, GroupCF-4 dominates T-GLANCE-4 and achieves smaller cost compared to 1265 C-GLANCE-4, which, however, has better performance in terms of effectiveness. Concerning the 1266 case s = 8, GroupCF-8 has slightly better cost but worse effectiveness than C-GLANCE-8, and 1267 slightly better effectiveness but worse cost than T-GLANCE-8. Overall, our methods ensure com-1268 petitive, robust and highly efficient performance.

1270 Table 18: Detailed results for the solution of s-GCE (s = 4 and s = 8) for HELOC dataset. The table reports effectiveness, cost, size, and runtime, including their standard deviations for each method. 1271

METHOD / MODEL		DN	Ň			LI	t			XGB	3	
	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME	eff	avc	size	RUNTIME
FAST ARES-4	$12.19 \pm 0.58$	$1.03 \pm 0.05$	4	1562.21 ± 656.59	$9.23 \pm 1.24$	$1.12 \pm 0.1$	4	961.23 ± 541.41	$8.49 \pm 1.32$	$1.16 \pm 0.13$	4	963.0 ± 492.98
CET - 4	$86.78 \pm 10.62$	$8.67 \pm 3.25$	1.0	11090.16 ± 4597.13	$100.0 \pm 0.0$	$3.57 \pm 1.48$	2.0	2662.77 ± 739.88	$86.78 \pm 6.7$	$12.51 \pm 2.75$	2.0	17339.7 ± 1363.4
GROUPCF - 4	$80.4 \pm 10.17$	$3.09 \pm 0.91$	4	$468.52 \pm 0.91$	$90.6 \pm 3.93$	$2.4 \pm 1.38$	4	$209.26 \pm 1.38$	$78.4 \pm 5.82$	$5.63 \pm 1.93$	4	$201.24 \pm 1.93$
GLOBE-CE - 4	$42.72 \pm 46.97$	$11.77 \pm 15.87$	1.4	$1.43 \pm 0.13$	$99.9 \pm 0.0$	$0.6 \pm 0.54$	3.0	$0.65 \pm 0.06$	$27.66 \pm 5.06$	$12.52 \pm 32.48$	2.4	$1.18 \pm 0.02$
DGLOBE-CE - 4	$99.96 \pm 0.05$	$11.07 \pm 8.6$	2.0	$2.2 \pm 0.14$	$99.9 \pm 0.0$	$1.63 \pm 0.35$	4.0	$0.94 \pm 0.05$	$77.64 \pm 11.51$	$128.0 \pm 0.0$	4.0	$1.77 \pm 0.03$
C-GLANCE-4	$99.94 \pm 0.05$	$11.24 \pm 1.37$	4.0	307.06 ± 0.0	$100.0 \pm 0.0$	$1.55 \pm 0.54$	4.0	$214.6 \pm 0.0$	$98.94 \pm 0.66$	$19.99 \pm 1.91$	4.0	$183.45 \pm 0.0$
T-GLANCE-4	$99.8 \pm 0.24$	$10.9 \pm 1.22$	3.0	$135.37 \pm 0.0$	$100.0 \pm 0.0$	$1.58 \pm 0.37$	3.0	$108.79 \pm 0.0$	$96.42 \pm 2.45$	$24.85 \pm 8.4$	3.0	$72.95 \pm 0.0$
FAST ARES-8	$15.44 \pm 1.91$	$1.05 \pm 0.06$	8	1070.19 ± 558.28	$10.95 \pm 1.73$	$1.11 \pm 0.07$	8	901.3 ± 493.42	$11.22 \pm 2.05$	$1.09 \pm 0.09$	8	829.62 ± 420.35
CET - 8	99.36 ± 0.92	$8.86 \pm 2.99$	4.0	6758.87 ± 3384.51	$100.0 \pm 0.0$	$2.99 \pm 1.3$	2.0	$1628.08 \pm 344.57$	$90.25 \pm 2.51$	$12.99 \pm 3.49$	6.0	9108.94 ± 1624.21
GROUPCF - 8	$84.8 \pm 12.32$	$1.98 \pm 0.5$	8	501.45 ± 2.5	$96.6 \pm 1.96$	$1.98 \pm 0.5$	8	$249.14 \pm 0.5$	$81.4 \pm 4.22$	$5.14 \pm 1.46$	8	$197.68 \pm 1.46$
GLOBE-CE - 8	$45.49 \pm 44.72$	$14.19 \pm 14.14$	3.4	$1.44 \pm 0.02$	$99.9 \pm 0.0$	$0.58 \pm 0.5$	7.0	$0.64 \pm 0.02$	$28.88 \pm 4.71$	$12.51 \pm 30.88$	3.0	$1.18 \pm 0.02$
DGLOBE-CE - 8	$100.0 \pm 0.0$	$6.65 \pm 0.15$	4.4	$3.2 \pm 0.22$	$99.9 \pm 0.0$	$1.63 \pm 0.35$	8.0	$1.35 \pm 0.04$	$81.95 \pm 10.38$	$128.0 \pm 0.0$	8.0	$2.55 \pm 0.04$
C-GLANCE-8	$99.98 \pm 0.04$	$9.42 \pm 2.22$	8.0	$301.47 \pm 0.0$	$100.0 \pm 0.0$	$1.2 \pm 0.12$	8.0	$217.9 \pm 0.0$	$99.0 \pm 0.55$	$16.65 \pm 2.44$	8.0	$182.83 \pm 0.0$
T-GLANCE-8	$99.78 \pm 0.21$	$9.92 \pm 1.83$	4.0	$40.64 \pm 0.43$	$100.0 \pm 0.0$	$1.52 \pm 0.32$	3.0	$39.0 \pm 0.89$	$98.25 \pm 1.85$	$18.75 \pm 4.92$	4.0	$75.9 \pm 0.8$

1281 1282

1246

1250

1259

1269

1272

Table 18 presents a comparative analysis of all algorithms for the Heloc dataset, which, due to its 1283 exclusively numeric features, poses challenges in achieving low costs. Despite this complexity, the 1284 methods T-GLANCE, C-GLANCE, and dGLOBE-CE stand out by delivering near-perfect or perfect 1285 effectiveness in the DNN models, while T-GLANCE, C-GLANCE, dGLOBE-CE, and GLOBE-CE 1286 excel in the LR models. dGLOBE-CE-8 provides the best combination of effectiveness and cost 1287 for DNN, and GLOBE-CE-8 achieves the same for LR. For the XGB models, T-GLANCE and 1288 C-GLANCE lead in effectiveness across all methods. Overall, our methods demonstrate strong and 1289 consistent performance, unlike other methods, where results fluctuate considerably.

1290 Table 19 presents a comparative analysis of all algorithms for the Adult dataset. The Fast ARes 1291 method consistently performs poorly, achieving notably low effectiveness scores across all models. 1292

It is also noteworthy that in the Adult dataset which is the largest of all in terms of instances and 1293 dimensions, the CET method did not run properly due to exceedingly high runtimes or infeasibility 1294 determined by the Gurobi optimizer. This is indicated by the ">20h or Infeasible (on Gurobi)" note 1295 present in the runtime column for both CET-4 and CET-8 across all models.

Table 19: Detailed results for the solution of *s*-GCE (s = 4 and s = 8) for Adult dataset. The table reports effectiveness, cost, size, and runtime, including their standard deviations for each method.

METHOD / MODEL		DNN				LR				XGB	
FAST ARES - 4 1	12.39 ± 1.06 1.0	avc size 0±0.0 4.0	266.51 ± 127.13	ett 11.74 ± 2.4	avc 1.0 ± 0.0	4.0	124.2 ± 57.64	6.13 ± 0.42	1.0 ± 0.0	4.0	108.66 ± 57.21
GROUPCF - 4 GLOBE-CE - 4	$100.0 \pm 0.0$ 10.08 99.92 ± 0.0 4.24 00.02 ± 0.0 10.80	08±0.03 4.0 4±0.42 3.0	3224.75 ± 0.03 6.6 ± 0.22 0.74 ± 0.21	$100.0 \pm 0.0$ $99.92 \pm 0.0$ $00.02 \pm 0.0$	$1.71 \pm 0.39$ $2.68 \pm 0.17$ $5.01 \pm 0.02$	4.0	925.25 ± 0.39 6.48 ± 0.17 0.36 ± 0.06	96.8 ± 1.72 82.87 ± 12.14	$1.41 \pm 0.54$ 30.1 ± 10.39	4.0 2.4	596.57±0.54 5.44±0.18 8.26±0.14
-GLANCE - 4 -GLANCE-4	100.0 ± 0.0 4.43	5±0.73 4.0 3±0.43 3.0	246.82 ± 0.0 410.47 ± 0.0	100.0 ± 0.0 100.0 ± 0.0	$1.04 \pm 0.07$ $0.69 \pm 0.01$	4.0 3.0	194.83 ± 0.0 422.45 ± 0.0	99.85 ± 0.12 99.86 ± 0.14	5.98 ± 4.22 1.8 ± 0.51	4.0 3.0	$3.20 \pm 0.14$ $134.27 \pm 0.0$ $334.14 \pm 0.0$
AST ARES - 8 ET - 8 ROUPCF - 8	13.8±0.92 1.0 100.0±0.0 11.54	$0 \pm 0.0$ 8.0 - - >: $54 \pm 2.98$ 8.0	271.04 ± 126.28 OH OR INFEASIBLE (ON GUROBI) 3539.35 ± 2.98	12.51 ± 2.44 100.0 ± 0.0	1.0 ± 0.0 - 2.33 ± 1.91	8.0 - > 8.0	157.94 ± 73.44 >20H OR INFEASIBLE (ON GUROBI) 872.1 ± 1.91	7.19 ± 0.3	1.0±0.0 - 1.88±1.36	8.0 - 8.0	142.82 ± 67.84 >20H OR INFEASIBLE (ON GUROBI) 1705.8 ± 1.36
BE-CE - 8 OBE-CE - 8 ANCE - 8	$100.0 \pm 0.0$ 2.54 $100.0 \pm 0.0$ 11.85 $100.0 \pm 0.0$ 3.96	4±0.14 6.6 85±1.53 6.4 6±0.56 8.0	$6.56 \pm 0.18$ 14.09 ± 0.4 240.16 ± 0.0	$100.0 \pm 0.0$ $100.0 \pm 0.0$ $100.0 \pm 0.0$	$1.67 \pm 0.11$ $6.12 \pm 0.64$ $1.03 \pm 0.07$	7.0 8.0 8.0	$6.41 \pm 0.08$ $13.57 \pm 0.22$ $179.43 \pm 0.0$	85.76 ± 9.11 94.38 ± 2.01 99.87 ± 0.09	13.48 ± 4.71 62.37 ± 3.07 3.85 ± 2.33	3.8 7.4 8.0	$5.58 \pm 0.11$ 12.01 $\pm 0.33$ 126.65 $\pm 0.0$
E-8	100.0 ± 0.0 3.77	7 ± 0.27 4.0	119.66 ± 0.69	$100.0\pm0.0$	0.79 ± 0.16	4.0	122.45 ± 0.63	$99.94 \pm 0.04$	$1.65 \pm 0.39$	4.0	128.22 ± 1.21
n the of	ther han	nd, Gro	upCF, C-GI	LANCE	E, T−0	GLAI	NCE, and GL	OBE-C	CE me	thoc	is all perform
tably w	/ell in te	erms of	effectiveness	. Und	ler the	: DN	NN model, $T$ -	-GLAN 2 77 m	CE-4 8	ina volu	T-GLANCE-8
and C-0	CLANCE	= -8  mirror	or this perfect	100%, t effeci	, with tivene	costa ss a	lso achieving	100%	with s	imil	arly good cost
rformai	nce. Not	tably. GI	OBE-CE-8	chieve	es the l	best	cost performa	nce un	der the	DN	IN model with
effectiv	veness s	core of 1	100% and a c	ost of	2.54.					21	
	1.1				11100	0	C	· · · · 11	. 11		1.1.1.1.000
footivon	C model,	l, I-GLA	ANCE-4 and of 0.60 and	1-GL	ANCE	-ð p	alv C CI A	MCE al	y well	, ac. for	meving 100%
ith 1000	% effect	tiveness	and costs 1	0.79, 04 and	1 1 03	Si	milarly in th	ne XGF	so per	el "	T-GLANCE-4
-GLANC	CE-8. C	-GLANC	E=4. and $C$ -	-GLAN	JCE-8	mai	intain high e	ffective	ness s	core	es of 99.86%.
9.94%, 9	99.85%.	and 99.8	37%, with real	ally go	od res	spect	tive costs.				,
	Justina	the ment		and the	- <del>6</del>	data	col	IDA C	Commo		Pradit Dafault
pon eva redit H	FLOC 4	and Adu	It Income	it is e	e live vident	uala that	t our methods	T = CI		lli C Lana	
nsistent	tlv demo	onstrate	concrete perf	ormar	nce T	hev	achieve near-	nerfect	or net	rfec	t effectiveness
ores wh	nile mai	intaining	low costs a	cross	differe	ent r	nodels, such	as DN	N, LR	, ar	nd XGB. This
bustnes	s and co	onsistenc	y underline	their e	efficie	ncy a	and practicali	ty in h	andlin	g di	iverse datasets
nd comp	lexities.		•			•				0	
ΔΙΙ	SED ST	ר אמווד		ном	V GI	OR/		EDEVC		r	
	I A NAT		DE DEDCI			ODr		LINIA	IUA		
LAP	LANAI	IIONS A	AKE FEKU		)						
1 DES	EIGN										
I DES											
le have d	designed	d a user s	study, with th	e follo	owing	goal	s.				
	Ŭ		₩ 1		0	-					
• 1	Assess (a	(a) how o	lifferent cond	cepts f	for glo	bal	counterfactua	l expla	inabili	ty a	are understood
8	and (b) v	which co	ncepts are pr	eferre	d by <i>ii</i>	ıdivi	iduals who se	ek reco	ourse.		
• 4	Assess (	(a) how a	lifferent con	cepts f	for glo	bal	counterfactua	l expla	inabili	ty a	are understood
а	and $(b)^{2}$	which c	oncepts are	preferi	red by	the	system/mode	el owne	er who	see	eks to provide
1	recourse	e to the a	ffected popul	lation.							
• /	Assess h	now peor	le evaluate tl	he trac	le-off	betw	een effectiver	iess an	d aver	age	recourse cost.
ŕ		r- r			55					0-	
he user s	study ha	as three p	arts, with ea	ch hav	ing a	diffe	erent goal.				
1.1 T	HE VIEV	WPOINT	OF INDIVID	UALS	SEEK	ING	RECOURSE				
							<b>.</b>				
he study	<sup>'</sup> puts pa	articipant	s in the follo	wing l	nypoth	netic	al situation. "	Consid	ler a co	omp	any that at the
IU OI the	e year g	gives a b	onus to some	e or it	s emp	toye	can do differ	e some	emplo	ber ber	es that did not
cccive til	t vear i	e they	want to achie	NIIUW	wiidt   ourse	ncy The	v have the op	tion to	be inve	ncy nve	d in additional
projects to	o increa	se their	chances of re	eceivin	ig the	bon	us. Working	in an a	dditio	nal 1	project, means
hat their	work ho	ours will	increase."		0 110	5511	as. monthing	u		ł	project, means

The study then asks three questions.

**Q1** "Suppose you're an employee and have the option to be involved in one of these projects.

A. All employees working on Project A received a bonus. Among them, the average number of additional working time per week was 5 hours.

B. All employees working on Project B received a bonus. No employee worked more than 8 addi-tional hours per week.

1356 Which project would you choose based on the information provided?"

The goal of Q1 is to understand which type of information people prefer. A is similar to the information counterfactual directions (as in Ley et al. (2023)) provides, while B is similar to the information counterfactual actions provide.

13611362 Q2 "Suppose you're an employee and have the option to be involved in one of these projects.

1363 C. 75% of the employees working on Project C received a bonus. Among them, the average number1364 of additional working time per week was 4 hours.

D. 75% of the employees working on Project D received a bonus. No employee worked more than
 6 additional hours per week.

1368 Which project would you choose based on the information provided?"

1369 Q2 has the same goal with Q1, but it emphasizes that information comes with no guarantees.

1371 Q3 "Suppose you're an employee and you consider working on Project E to increase your chances of getting the bonus. There are three pieces of information about Project E.

1373
 1374
 1375
 Info-1. For employees working on Project E who received the bonus, the average number of additional working time per week was 5 hours.

1376 Info-2. All employees working on Project E who logged at least 10 additional working hours received the bonus.

Info-3. 60% of employees working on Project E who logged at least 5 additional working hoursreceived the bonus.

Rank the pieces of information from the most helpful to the least helpful."

Q3 aims to understand which type of information people find most helpful. Info-1 is similar to the information counterfactual directions provide, while Info-2 and Info-3 are similar to the information counterfactual actions provide.

1385 1386

1370

I.1.2 THE VIEWPOINT OF SYSTEM OWNERS SEEKING TO PROVIDE RECOURSE

This part puts the participants in the role of the system owner. "Assume you are the company CEO, summarizing the benefits of project opportunities.

This year, two types of employees, x1 and x2, did not earn a bonus. These employee types differ in how they allocate their work hours:

1392 1393

1394

• x1 employees spend more time in the office and less in the field (i.e., outside the office).

• x2 employees spend more time in the field and less in the office.

The number of employees who did not earn a bonus is evenly split between these two groups."

Then, with reference to Figure 5, the study provides this additional information. "Employee types are represented as points on the office hours-field hours plane, shown in the left image. Projects are depicted as directions on this plane, indicating the need for additional office and field hours. In some cases, it is more intuitive to view projects from the perspective of the employee types, represented as the origin o, as shown in the right image. "

1402 Then the user study asks two questions.

1403

Q1 "There are three possible projects that employees can work on:



Each algorithm is evaluated on two aspects:

Effectiveness: The percentage of the population for whom the algorithm provides recourse. The higher (up to 100%) the effectiveness is, the better the algorithm.
Average Cost: The average cost incurred by individuals who achieve recourse. The lower

(down to 0) the cost is, the better the algorithm.

The images display the effectiveness and average cost for an algorithm as a point with error bars;
the error bars denote the standard deviation across multiple evaluations. The closer a point is to the top left corner, the better the algorithm."

We then ask two types of questions for each figure, depicting either the four methods for implicit GCE or the three methods for explicit GCE.

1469
 1470
 1471
 1471
 Q1 "Based on the evaluation results presented in the figure, rank the three/four algorithms from best (1st position) to worst (3rd/4th position)."

1472

1458

1459

1460

1461

1462

1473

**Q2** "How many times (out of 15) is Algo 0/1/2/3 the better algorithm?"

These questions aim to understand how people combine the two aspects of effectiveness and average cost to assess the quality of global counterfactual explanations.

1476

1477 I.2 RESULTS 1478

Following the best practices outlined in Ley et al. (2023); Chowdhury et al. (2022), we recruited 40
participants consisting of (a) master's students in a data science program who attended lectures on
machine learning explainability and (b) PhD students from various machine learning domains with
extensive knowledge of explainability. After excluding six participants due to inconsistent responses
to repeated questions and/or incomplete rankings, we analyzed the responses of the remaining 34
participants.

1485 In the first part of the study, where participants assumed the role of individuals seeking recourse, we 1486 find that *participants prefer information conveyed through concrete counterfactual actions* (similar to GCEs produced by methods like ARES Rawal & Lakkaraju (2020) and this work) over counter-1487 factual directions represented by average-cost actions (akin to GCEs generated by GLOBE-CE Ley 1488 et al. (2023)). Specifically, for Q1, 56% of participants preferred information tied to concrete actions 1489 (project B) over average-cost actions (project A). This preference increased to 71% in Q2 when un-1490 certainty rose (effectiveness decreased from 100% to 75%). When asked to rank their preference 1491 among three types of information in Q3, 62% of participants rated Info-1, which represents an opti-1492 mal solution to the 1-GCE problem (similar to action c in Figure 5), as the most helpful. Conversely, 1493 41% ranked Info-2, which represents the average-cost action along the optimal direction as defined 1494 in Ley et al. (2023) (similar to action b in Figure 5), as the least helpful. Additionally, Info-3, an-1495 other optimal solution to the 1-GCE problem (similar to action a in Figure 5), was considered more 1496 helpful than Info-2.

1497 In the second part of the study, where participants assumed the role of system owners aiming to 1498 provide recourse to adversely affected individuals, we again found that participants preferred in-1499 formation conveyed through concrete counterfactual actions rather than counterfactual directions. 1500 Specifically, when asked (Q1) to choose among three projects corresponding to directions A, B, 1501 and C in Figure 5, 65% favored projects/directions A or C (both optimal 1-GCE solutions) over 1502 project/direction B (aligned with the optimal direction as defined in Ley et al. (2023)), with C1503 emerging as the clear favorite, selected 50% of the time. A similar pattern was observed in project placement decisions (Q2), involving actions a, b, and c in Figure 5, where 60% of participants 1504 preferred the optimal actions a and c over the average-cost action b. 1505

In the third part of the study, participants were asked to evaluate which method for global counterfactual explanations is superior based on performance in terms of average cost and effectiveness.
We found that *participants unanimously agree that the GLANCE methods outperform others*. Specifically, in the first two questions, participants assessed explicit GCE methods (T-GLANCE, ARES,

- 1510 CET). All participants (100%) identified T-GLANCE as the superior method, finding it outperformed
- the others approximately 13.41 out of 15 times (as illustrated in Figure 4). In the next two questions, participants evaluated implicit GCE methods (C-GLANCE, GLOBE-CE, dGLOBE-CE, GroupCF).

MODELS	ALGORITHMS		DATA	SETS	
		COM	PAS	GERMAN	CR
		EFF	AVC	EFF	
DNN	T-GLANCE	$99.53 \pm 0.22$	$2.62 \pm 0.24$	96.97 ± 3.48	1
	C-GLANCE	$100.0 \pm 0.0$	$2.34 \pm 0.43$	$95.31 \pm 3.15$	1.
	LOCAL	$100.0\pm0.0$	$2.75 \pm 0.53$	$100.0 \pm 0.0$	4.
LR	T-GLANCE	$99.79 \pm 0.22$	$2.19 \pm 0.1$	$99.58 \pm 0.83$	1.
	C-GLANCE	$100.0 \pm 0.0$	$2.33 \pm 0.38$	$100.0 \pm 0.0$	1.
	LOCAL	$100.0\pm0.0$	$2.73 \pm 0.23$	$100.0 \pm 0.0$	4
XGB	T-GLANCE	$99.02 \pm 0.93$	$2.6 \pm 0.46$	99.46 ± 1.08	1.
	C-GLANCE	$99.51 \pm 0.46$	$2.96 \pm 0.82$	$100.0 \pm 0.0$	1
	LOCAL	$100.0 \pm 0.0$	$2.87 \pm 0.27$	$100.0 \pm 0.0$	5.
				1	

### Table 20: Local vs global counterfactuals

Again, all participants (100%) selected C-GLANCE as the superior method, noting it outperformed the others approximately 10.81 out of 15 times (as depicted in Figure 3).

#### 1533 ADDRESSING POTENTIAL EDGE CASES IN RECOURSE ALGORITHMS J 1534

1535 While the proposed algorithms are designed to provide robust and actionable recourse, certain edge 1536 cases may pose challenges: 1537

1538 Highly nonlinear decision boundaries might cause the model's decision boundary to vary significantly across small regions, leading to oversimplified recourse suggestions. We address this issue 1539 by using proximity-aware clustering and local counterfactual action generation to better approximate 1540 the decision boundary, ensuring more accurate and actionable recommendations. 1541

1542 Sparse data regions or outlier points might result in unrealistic, infeasible, or overly tailored recourse 1543 actions. To mitigate these challenges, we generate many clusters and diverse counterfactuals, ensuring that feasible solutions are identified without being disproportionately influenced by outliers. 1544 This approach enhances robustness and broad applicability. 1545

1546 Ambiguity in selecting the "optimal" action can arise when multiple actions have similar costs and 1547 effectiveness, potentially leading to overly complex recourse suggestions. We resolve this by em-1548 ploying predefined criteria that prioritize actionable and interpretable solutions while maintaining a 1549 balance between cost and effectiveness.

1550 1551

1512

1527 1528 1529

1530

1531 1532

#### 1552 Κ LOCAL VS. GLOBAL COUNTERFACTUAL COSTS

1553 1554

To illustrate the challenges in defining optimal solutions for global counterfactual explanations, we 1555 conducted an experiment comparing the cost of local counterfactuals to the global counterfactual 1556 actions generated by our method. Using the Compas and German Credit datasets, we generated all 1557 possible local counterfactuals for each instance and calculated the average cost across the datasets. 1558 The results are presented in Table 20 under the row "LOCAL". 1559

Surprisingly, despite generating only 4 global counterfactual actions, our method achieved signif-1560 icantly lower average costs than the local counterfactuals, even though both were generated using 1561 the same local counterfactual method. This counterintuitive result highlights the inherent variability 1562 and complexity in local counterfactual generation, emphasizing the difficulty in defining an optimal 1563 solution at both local and global levels. 1564

This experiment demonstrates the effectiveness of our approach in achieving cost-efficient global 1565 counterfactuals while reinforcing the challenges of theoretical analysis in this domain.

MODEL	INITIAL CLUSTERS DIVERSE ACTIONS	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	10
DNN	5	99.91	100.00	99.86	99.97	100.00	99.94	100.00	99.97	99.97	10
	10	99.91	100.00	99.88	99.97	99.97	100.00	99.97	99.97	99.97	- 99
	15	99.89	100.00	99.94	100.00	99.97	99.94	99.97	99.97	100.00	- 10
	20	99.91	100.00	99.97	100.00	99.97	99.97	100.00	99.97	99.94	10
	25	99.89	100.00	100.00	99.97	99.97	99.97	99.94	100.00	99.97	9
	30	99.94	100.00	99.97	100.00	99.97	100.00	100.00	100.00	100.00	10
	35	99.94	99.97	99.97	99.97	99.97	100.00	99.97	99.97	99.97	10
	40	99.94	100.00	100.00	99.97	99.97	99.97	99.97	99.97	100.00	10
	45	99.94	100.00	99.94	100.00	100.00	99.97	99.97	99.97	100.00	10
	50	99.91	100.00	99.97	99.97	99.97	99.97	100.00	100.00	100.00	10
LR	5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	10
	10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	10
	15	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1
	20	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	10
	25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1
	30	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1
	35	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1
	40	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1
	45	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1
	50	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1
XGB	5	94.42	94.39	95.48	96.50	95.53	94.95	96.29	96.14	96.42	9
	10	94.03	96.08	97.38	95.99	95.85	98.24	96.79	96.94	95.54	9
	15	95.95	95.97	97.73	96.30	97.39	98.01	97.97	97.94	97.92	9
	20	96.02	96.62	97.79	96.19	98.54	98.54	98.62	98.52	98.11	9
	25	96.66	95.72	98.57	96.84	98.62	98.65	98.57	97.76	98.21	9
	30	90.68	90.69	98.62	90.31	98.49	98.67	98.46	98.49	98.43	9
	33	90.73	97.17	99.00	97.30	98.07	98.85	98.40	98.84	98.38	9
	40	90.94	97.09	90.02	97.44	98.90	98.03	98.92	99.07	98.80	- 9
	43	91.10	20.31	20.02	91.09	20.82	99.00	20.44	20.09	90.82	2

### Table 21: Default Credit Effectiveness Mean

## L IMPACT OF INITIAL CLUSTERS AND DIVERSE ACTIONS

To evaluate the influence of key parameters in C-GLANCE on solution quality, interpretability, and runtime, we analyzed the role of initial clusters and diverse candidate actions. The product of these parameters defines the total number of generated actions, which directly impacts the algorithm's performance.

Increasing the number of generated actions often improves effectiveness but typically comes at the expense of increased cost. However, selecting a subset of actions with optimal effectiveness tends to reduce overall cost. This pattern is clearly observed in Tables 21 to 24.

Regarding individual parameter effects, increasing the number of initial clusters proves particularly beneficial for larger datasets with widely distributed points. This approach allows for better grouping in the feature space, thereby enhancing the relevance and feasibility of the generated counterfactual actions. Conversely, increasing the diversity of candidate actions is more advantageous for complex models with intricate decision boundaries. A higher diversity ensures that the generated counterfactual actions are well-aligned with the model's structure, leading to better adaptation to the complexities of the decision boundary.

1620

MODEL	INITIAL CLUSTE DIVERSE ACTIO	ERS DNS	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0
DNN		5	4.85	3.15	1.45	1.58	1.40	2.27	1.13	2.65	1.41
		10	2.91	2.53	1.98	1.41	1.59	1.39	1.33	1.51	1.03
		20	2.76	1.90	1.19	1.23	1.21	1.00	1.03	1.08	1.02
		25	3.91	1.96	1.27	1.40	1.11	1.01	1.29	1.25	1.02
		30	2.16	1.90	1.27	1.40	1.01	1.01	1.21	1.28	1.01
		35	2.16	1.62	1.27	1.39	1.00	1.01	1.29	1.07	1.00
		40	1.82	1.47	1.21	1.23	1.00	1.01	1.02	1.19	1.01
		45 50	1.78	1.54 1.34	1.21	1.07	1.01 1.01	$1.00 \\ 1.00$	$1.02 \\ 1.22$	1.07	1.00
LR		5	1.93	1.43	1.33	1.47	1.45	1.48	1.30	1.27	1.32
		15	1.91	1.13	1.10	1.34	1.43	1.29	1.20	1.09	1.02
		20	1.22	1.01	1.27	1.35	1.23	1.24	1.20	1.11	0.99
		25	1.20	1.13	1.19	1.10	1.11	1.32	1.19	1.05	1.17
		30	1.20	1.00	1.27	1.13	1.22	1.20	1.09	1.19	1.21
		35	1.11	1.00	1.16	1.12	1.15	1.17	1.05	0.99	1.07
		40	1.11	1.00	1.20	1.16	1.10	1.11	1.11	1.00	0.99
		45 50	$1.11 \\ 1.11$	$1.00 \\ 1.00$	1.11	1.07	1.08	1.06	1.18 $1.06$	0.97	1.21
XGB		5	3.44	3.43	2.83	2.76	3.30	2.99	2.67	2.67	2.65
		10	2.97	2.81 2.47	2.91	2.80	3.08	4.05	2.97	2.89	2.95
		20	2.98	3.35	3.81	3.10	4.08	4.38	3.69	3.37	3.17
		25	2.95	3.15	3.81	3.03	2.63	3.29	4.26	3.75	3.75
		30	3.42	3.82	3.97	3.16	4.59	2.97	4.48	3.23	3.57
		35	2.65	3.01	4.18	3.06	5.12	4.24	3.27	3.43	4.10
		40	2.43	3.09	3.39	3.98	3.54	4.57	4.12	3.52	3.74
		45 50	2.57	3.83	2.39	4.03	2.84	4.84	5.00 4.00	4.11	3.47 1 14
			Table	23: HI	ELOC	Effect	iveness	Mean			
Model	INITIAL CLUSTERS DIVERSE ACTIONS	10.0	Table           20.0	23: HI	ELOC	Effect	iveness	60.0	70.0	80.0	90.0
Model	INITIAL CLUSTERS DIVERSE ACTIONS	10.0 99.21	20.0 99.6	23: HI	ELOC	Effect 40.0 99.34	50.0	Mean 60.0 99.46 99.92	70.0 99.88 99.92	80.0 99.69 90.86	90.0
Model	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15	10.0 99.21 99.41 99.36	Table           20.0           99.6           99.6           99.0	23: HI 0 30 57 99 56 99 98 99	ELOC 0.0	Effect 40.0 99.34 99.28 99.27	50.0 99.53 99.43 99.39	Mean 60.0 99.46 99.92 99.94	70.0 99.88 99.92 99.92	80.0 99.69 99.86 99.73	90.( 99.8 99.9 99.9
Model	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25	10.0 99.21 99.41 99.36 99.72	20.0 20.0 99.6 99.0 99.0	23: HI	ELOC 0.0 -	Effect 40.0 99.34 99.28 99.27 99.67 99.67	50.0 99.53 99.43 99.39 99.92	Mean 60.0 99.46 99.92 99.94 99.89 90.02	70.0 99.88 99.92 99.92 99.98 99.98 90.98	80.0 99.69 99.86 99.73 99.92	90.0 99.8 99.9 99.9 99.9
Model	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30	10.0 99.21 99.41 99.36 99.72 99.74 99.78	20.0 99.6 99.0 99.0 99.0 99.0 99.0	23: HI 0 30 67 99 16 99 18 99 18 99 18 99 18 99 18 99 18 99 18 99 18 99	ELOC	Effect 40.0 19.34 19.28 19.27 19.67 19.39 19.98	50.0 99.53 99.43 99.39 99.92 99.92 99.98 99.98	60.0 99.46 99.92 99.94 99.89 99.92 100.00	70.0 99.88 99.92 99.92 99.98 99.98 99.95	80.0 99.69 99.86 99.73 99.92 99.96 99.98	90.0 99.8 99.9 99.9 99.9 99.9 99.9
Model	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.78	20.0 99.6 99.0 99.0 99.0 99.0 99.0 99.0	<b>23: HF</b> 0 300 17 99 16 99 18 99 18 99 18 99 18 99 18 99 18 99 18 99 10 99	ELOC	Effect 40.0 99.34 99.27 99.27 99.67 99.39 99.98 99.98	50.0 99.53 99.43 99.92 99.92 99.98 99.98 100.00	Mean 60.0 99.46 99.92 99.94 99.92 100.00 100.00 00.00	70.0 99.88 99.92 99.98 99.98 99.98 99.98 99.95 99.98	80.0 99.69 99.86 99.73 99.92 99.96 99.98 99.98 99.98	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model dnn	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.78 99.78 99.76	20.4           99.6           99.0           99.0           99.0           99.0           99.1           98.8           99.1	<b>23:</b> HF 0 30 17 99 16 99 18 99 18 99 18 99 18 99 18 99 18 99 18 99 19 99 19 99 2 99	ELOC	Effect 40.0 99.34 99.28 99.27 99.67 99.98 99.98 90.98 00.00	50.0 99.53 99.43 99.92 99.98 99.98 100.00 100.00 99.98	Mean 99.46 99.92 99.94 99.92 100.00 100.00 99.98 99.98	70.0 99.88 99.92 99.92 99.98 99.98 99.98 99.98 99.98 99.98	80.0 99.69 99.86 99.73 99.92 99.96 99.98 99.98 99.98 99.98	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model dnn	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50	10.0 99.21 99.36 99.72 99.74 99.78 99.78 99.76 99.76 99.76	20.1 20.1 99.6 99.0 99.0 99.0 99.0 99.1 98.8 99.1 99.1	<b>23:</b> HI 0 300 77 99 16 99 18 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99	ELOC 0.0 61 9 22 9 21 9 14 9 23 9 27 9 60 10 65 10 65 10	Effect 40.0 19.34 19.28 19.27 19.67 19.39 19.98 00.00 00.00 00.00 00.00	50.0 99.53 99.43 99.92 99.98 99.98 100.00 100.00 99.98 99.98	Mean 99.46 99.92 99.94 99.92 100.00 100.00 99.98 100.00	70.0 99.88 99.92 99.98 99.98 99.98 99.98 99.98 99.98 100.00 99.94	80.0 99.69 99.86 99.92 99.96 99.98 99.98 99.98 99.98 99.98	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model dnn	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.78 99.76 99.76 99.76 100.00	20.0           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.1           98.8           99.1           99.1           99.1           99.1           99.1           99.1           99.1	23: HI 0 30 77 99 66 99 18 99 18 99 18 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99 2 99 2 99 2 99 0 100	ELOC 0.0 61 9 22 9 21 9 14 9 23 9 27 9 60 1 65 1 0.00 1	Effect 40.0 99.34 99.27 99.67 99.98 99.98 99.98 00.00 00.00 00.00 00.00	50.0 99.53 99.43 99.92 99.98 99.98 100.00 100.00 99.98 99.98 100.00	Mean 99.46 99.92 99.94 99.92 100.00 100.00 99.98 100.00 100.00 100.00	70.0 99.88 99.92 99.92 99.98 99.98 99.98 99.94 100.00 99.94 100.00	80.0 99.69 99.86 99.73 99.92 99.96 99.98 99.98 99.98 99.98 99.98 99.98	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model dnn	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.76 99.76 99.76 100.00 100.00	20.4           20.4           99.6           99.0           99.0           99.0           99.0           99.1           100.4           100.4	23: HI 0 30 77 99 16 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99 2 99 100 100 100 100 100 100 100 1	ELOC 0.0 61 9 22 9 21 9 14 9 23 9 27 9 60 10 65 10 65 10 65 10 0.00 10 0.00 10	Effect 40.0 99.34 99.27 99.67 99.98 99.98 99.98 00.00 00.00 00.00 00.00 00.00	50.0 99.53 99.43 99.99 99.92 99.98 99.98 100.00 99.98 99.98 100.00 100.00 100.00	Mean 99.46 99.92 99.94 99.92 100.00	70.0 99.88 99.92 99.98 99.98 99.98 99.98 99.94 100.00 99.94 100.00 100.00	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.00 100.00	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model dnn	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 30 35 40 45 50 5 10 15 20	10.0 99.21 99.34 99.72 99.74 99.78 99.76 99.76 99.76 100.00 100.00 100.00	20.1           20.2           99.6           99.0           99.0           99.0           99.0           99.0           99.1           98.8           99.1           100.4           100.4           100.4	23: HI 0 30 77 99 16 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99 2 99 100 100 00 100 00 100	ELOC 0.0 61 9 22 9 14 9 14 9 14 9 23 9 27 9 60 1 65 1 65 1 0.00 1 0.00 1 0.00 1	Effect 40.0 99.34 99.27 99.27 99.39 99.98 99.98 90.00 00.00	50.0 99.53 99.43 99.39 99.92 99.98 99.98 100.00 99.98 99.98 100.00 100.00 100.00	Mean 99.46 99.92 99.94 99.89 99.92 100.00 100.00 100.00 100.00 100.00 100.00	70.0 99.88 99.92 99.98 99.98 99.98 99.98 99.94 100.00 99.94 100.00 100.00	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.00 100.00 100.00	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model dnn	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 30 35 40 45 50 5 10 15 20 25	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.76 99.76 99.76 100.00 100.00 100.00 100.00	20.1           20.2           99.6           99.0           99.0           99.0           99.0           99.1           98.8           99.1           100.4           100.4           100.5	23: HI 0 30 77 99 76 99 78 99 78 99 78 99 79 99 79 99 79 99 79 99 79 99 79 99 70 100 70 100 700	ELOC 	Effect 40.0 99.34 99.27 99.27 99.98 99.98 90.00 00.00	50.0 99.53 99.43 99.39 99.92 99.98 99.98 100.00 99.98 99.98 100.00 100.00 100.00 100.00	Mean 99.46 99.92 99.94 99.92 100.00	70.0 99.88 99.92 99.98 99.98 99.98 99.98 99.94 100.00 99.94 100.00 100.00 100.00 100.00	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.00 100.00 100.00 100.00	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35	10.0 99.21 99.36 99.72 99.74 99.78 99.78 99.78 99.78 99.78 99.76 99.76 99.76 100.00 100.00 100.00 100.00	20.1           20.4           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.1           98.8           99.1           98.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           90.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           100.4           100.4           100.4           100.4           100.4           100.4	23: HI 0 30 77 99 16 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99 100 100 00 100 00 1	ELOC 0.0 61 9 22 9 21 9 21 9 21 9 21 9 23 9 20 9 20 9 20 9 21 9 20 9 2	Effect 40.0 99.34 99.27 99.27 99.67 19.39 99.98 00.00	50.0 99.53 99.43 99.92 99.98 99.98 100.00 99.98 100.00 100.00 100.00 100.00 100.00	Mean           60.0           99.46           99.92           99.94           99.89           99.92           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00	70.0 99.88 99.92 99.98 99.95 99.98 99.94 100.00 99.94 100.00 100.00 100.00 100.00	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 100.00 100.00 100.00 100.00 100.00	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 25 30 35 40	10.0 99.21 99.41 99.72 99.74 99.78 99.78 99.76 99.76 99.76 100.00 100.00 100.00 100.00 100.00	20.1           20.4           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.1           98.8           99.1           98.8           99.1           99.1           99.1           99.1           100.4           100.4           100.4           100.4           100.4           100.4           100.4	23: HI 0 30 77 99 16 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99 100 100 00 100 00 1	ELOC 0.0 61 9 22 9 21 9 21 9 21 9 23 9 23 9 27 9 60 10 65 10 0.00 10 0.	Effect 40.0 99.34 99.27 99.27 99.27 99.98 00.00	50.0 99.53 99.43 99.39 99.92 99.98 100.00 99.98 100.00 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00	Mean           60.0           99.46           99.92           99.94           99.89           99.92           100.00           99.98           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 99.94 100.00 100.00 100.00 100.00 100.00 100.00	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 100.00 100.00 100.00 100.00 100.00 100.00	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
MODEL DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 5 10 5 10 5 50 5 50 50	10.0 99.21 99.41 99.72 99.74 99.78 99.76 99.76 99.76 100.00 100.00 100.00 100.00 100.00 100.00 100.00	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.1           98.8           99.1           98.1           99.1           100.4           100.4           100.4           100.4           100.4           100.4           100.4           100.4           100.4           100.4	23: HI 0 30 77 99 16 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99 100 100 00 100 00 1	ELOC 0.0 61 9 22 9 21 9 21 9 21 9 23 9 23 9 27 9 60 10 65 10 0.00 10 0.	Effect 40.0 99.34 99.27 99.67 99.99 99.98 00.00	50.0 99.53 99.43 99.39 99.92 99.98 100.00 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00	Mean           60.0           99.46           99.92           99.94           99.98           99.99           100.00	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 99.94 100.00 100.00 100.00 100.00 100.00 100.00 100.00	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
MODEL DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 25 30 35 40 45 50 50	10.0 99.21 99.41 99.72 99.74 99.78 99.76 99.76 99.76 99.76 100.00 100000000	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           100.4           100.5           100.4           100.4           100.5           100.4	23: HI 0 30 77 99 16 99 18 99 18 99 18 99 18 99 18 99 19 99 2 99 2 99 2 99 100 100 00 100 00 1	ELOC 0.0 61 9 22 9 21 9 21 9 21 9 23 9 23 9 20 14 23 9 20 14 23 9 20 14 9 20 14 10 10 10 10 10 10 10 10 10 10	Effect 40.0 99.34 99.27 99.67 99.98 00.00	50.0 99.53 99.39 99.92 99.98 99.98 99.98 100.00	Mean           60.0           99.46           99.92           99.94           99.93           100.00	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
MODEL DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 25 30 35 40 45 50 5 10	10.0 99.21 99.41 99.72 99.74 99.78 99.76 99.76 99.76 99.76 100.00 100000000	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           99.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           90.1           100.4           100.4           100.4           91.5           95.7	23: HI 0 30 77 99 76 99 78 99 78 99 78 99 79 99 79 99 72 99 70 100 70 1000 70 100 70 100 70 1000 70 1000 7	ELOC 0.0 61 9 22 9 14 9 14 9 23 9 27 9 60 19 65 19 0.00	Effect 40.0 99.34 99.27 99.67 99.98 00.00	50.0 99.53 99.43 99.92 99.98 99.98 99.98 100.00 100000000	Mean           60.0           99.46           99.92           99.94           99.93           100.00           99.53	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 99.94 100.00 100000000	80.0 99.69 99.86 99.73 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.00 100000000	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
MODEL DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 40 45 50 25 30 35 40 40 45 50 25 50 25 50 25 50 50 50 50 50 50 50 50 50 50 50 50 50	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.76 99.76 99.76 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 87.66 93.13	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.1           100.4           100.4           100.4           100.4           100.5           91.5           95.7           96.5	23: HI 0 30 7 99 6 99 8 99 8 99 8 99 2 99 2 99 2 99 2 99 2 99 2 99 2 99 0 100 00 100 0	ELOC 0.0 61 9 22 9 21 9 21 9 21 9 23 9 27 9 60 14 65 14 65 14 0.00 14 0	Effect 40.0 99.34 99.27 99.67 99.98 00.00	50.0 99.53 99.43 99.92 99.98 99.98 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 95.33 97.86 98.05	Mean           60.0           99.46           99.92           99.94           90.93           99.94           100.00           00.00           00.00	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 97.12 98.39 98.89 99.94	80.0 99.69 99.73 99.96 99.98 99.98 99.98 99.98 99.98 99.98 99.98 100.00 100000000	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
MODEL DNN LR XGB	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 30 35 30 30 35 30 30 35 30 30 35 30 30 35 30 30 35 30 30 30 35 30 30 30 35 30 30 35 30 30 35 30 30 35 30 30 30 35 30 30 30 35 30 30 30 35 30 30 30 35 30 30 30 30 30 30 30 30 30 30 30 30 30	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.76 99.76 99.76 99.76 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 87.66 90.45 93.13 94.71	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.1           100.4           100.4           100.4           100.4           100.5           91.5           95.7           97.8           97.8           97.8           97.8           97.8           97.8	23: HI 0 30 7 99 6 99 8 99 8 99 8 99 2 99 2 99 2 99 2 99 2 99 2 99 2 99 0 100 00 100 0	ELOC 0.0 	Effect 40.0 99.34 99.27 99.67 99.98 90.00 00.00	50.0 99.53 99.43 99.39 99.92 99.98 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 95.33 97.86 98.95 98.98	Mean 99.46 99.92 99.94 99.99 99.92 100.00 1	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 99.94	80.0 99.69 99.73 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.00 97.34 99.88 99.98	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
MODEL DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 30 35 40 45 50 50 50 30 35 40 45 50 50 50 50 50 50 50 50 50 50 50 50 50	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.76 99.76 99.76 99.76 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 87.66 90.45 93.13 94.71	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.1           100.4           100.4           100.4           100.4           100.5           95.7           97.8           97.7           96.5           97.8           97.7	23: HI 0 30 7 99 6 99 8 99 8 99 8 99 2 99 3 00 100 000 100 000 1000 1	ELOC 0.0 61 9 22 9 21 9 21 9 21 9 21 9 21 9 22 9 21 9 22 9 23 9 29 5 29 9 29 9 29 9 29 5 29 9 29 9 29 9 29 5 29 9 29 9 29 9 29 9 29 5 29 9 29 5 29 9 29 5 29 5 29 9 20 9 2	Effect 40.0 99.34 99.27 99.67 99.98 99.98 00.00	50.0 99.53 99.43 99.39 99.92 99.98 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 95.33 97.86 98.98 98.98 98.82	Mean           60.0           99.46           99.92           99.94           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           100.00           98.57           98.40           98.94           99.00	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 100	80.0 99.69 99.73 99.92 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.000	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
MODEL DNN LR	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 40 45 50 50 50 50 50 50 50 50 50 50 50 50 50	10.0 99.21 99.41 99.36 99.72 99.74 99.78 99.76 99.76 99.76 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 87.66 93.13 94.71 100.45 95.23 96.14 95.26 85.23	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.1           100.4           100.5           91.5           95.7           96.5           97.8           97.7           98.1           99.2	23: HI 0 30 7 99 6 99 8 99 8 99 8 99 8 99 2 99 3 99 4 99 5 99 6 0 100 00 1000	ELOC 0.0 	Effect 40.0 99.34 99.27 99.67 99.98 99.98 90.00 00.00	50.0 99.53 99.43 99.39 99.92 99.98 99.98 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 95.33 97.86 98.98 98.98 98.98 98.62 99.02	Mean 99.46 99.92 99.94 99.94 99.92 100.00 1	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 99.94	80.0 99.69 99.73 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.00 97.34 98.82 99.98 88.83 99.98	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9
Model DNN LR XGB	INITIAL CLUSTERS DIVERSE ACTIONS 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 40 45 50 50 50 50 50 50 50 50 50 50 50 50 50	10.0 99.21 99.41 99.72 99.72 99.74 99.78 99.76 99.76 99.76 99.76 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 87.66 93.13 94.71 95.23 94.71 95.23 95.23	Table           20.1           99.6           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.0           99.1           99.1           98.8           99.1           100.4           100.4           100.4           100.4           100.5           91.5           95.7.7           96.5           97.8           97.7           98.1           97.8           97.7           98.1           97.8           97.7           98.1           98.6           98.7	23: HI 0 30 77 99 76 99 78 99 78 99 79 99 2 99 3 00 100 000 100 000 70 000 70 0000	ELOC 0.0 	Effect 40.0 99.34 99.27 99.67 99.98 99.98 90.00 00.00	50.0 99.53 99.43 99.92 99.98 99.98 99.98 100.00 99.98 99.88 99.88 99.88 99.88 99.98	Mean 99.46 99.92 99.94 99.99 99.92 100.00 1	70.0 99.88 99.92 99.98 99.98 99.98 99.94 100.00 100	80.0 99.69 99.86 99.73 99.92 99.96 99.98 99.98 99.98 99.98 99.98 99.98 100.000 100.000 100.00000000	90.0 99.8 99.9 99.9 99.9 99.9 99.9 99.9

## Table 22: Default Credit Average Cost Mean

MODEL	INITIAL CLUSTERS DIVERSE ACTIONS	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	100.
DNN	5	9.03	7.44	10.79	9.14	11.58	10.49	10.20	11.21	9.67	10.5
	10	10.97	10.12	10.58	9.85	9.97	10.77	10.95	11.35	10.37	11.2
	15	10.54	11.09	10.96	10.35	10.36	10.23	10.57	11.17	9.83	9.4
	20 25	10.85	11.27	11.35	10.38	11.87	12.11	10.98	11.87	10.19	10.5
	25	10.80	11.30	11.80	10.13	10.70	11.54	11.51	11.23	10.09	0.5
	35	11.70	11.06	10.86	11.70	11.97	12.22	10.50	11.08	10.11	10.8
	40	12.03	11.06	11.01	11.29	11.56	12.22	10.87	11.48	10.12	10.8
	45	12.03	11.25	11.63	12.47	11.59	12.65	11.90	9.85	10.21	8.8
	50	12.03	11.20	11.62	12.17	11.59	11.64	11.67	10.70	11.39	11.5
LR	5	2.67	2.40	1.92	1.77	1.78	1.92	2.00	1.72	1.66	1.68
	10	2.09	2.04	1.88	1.03	1.79	1.00	1.00	1.30	1.37	1.3
	20	1.80	1.56	1.59	1.60	1.54	1.52	1.59	1.53	1.47	1.2
	25	1.80	1.65	1.62	1.51	1.51	1.52	1.50	1.48	1.43	1.4
	30	1.69	1.62	1.54	1.51	1.57	1.38	1.52	1.52	1.43	1.3
	35	1.69	1.62	1.57	1.41	1.49	1.24	1.43	1.40	1.49	1.3
	40	1.66	1.53	1.41	1.38	1.52	1.28	1.34	1.39	1.40	1.3
	43 50	1.64	1.33	1.30	1.40	1.46	1.30	1.43	1.39	1.23	1.3
XGB	5	12.22	14.05	15.44	14.77	15.91	16.25	17.14	17.13	17.87	18.2
	10	13.83	16.14	16.51	18.75	17.54	17.57	20.25	20.62	21.32	25.7
	15	15.33	18.93	17.05	17.40	18.07	21.27	19.51	21.48	27.93	18.6
	20	14.92	20.48	10.79	20.42	21.98	25.15	22.44	20.12	22.97	23.0
	30	15.03	20.40	18.35	21 33	23.88	24.85	23.86	23.30	24.39	24.0
	35	17.16	22.75	20.80	21.62	23.70	27.08	24.47	20.27	25.74	24.1
	40	16.58	25.14	21.67	23.77	24.82	29.50	26.66	23.40	26.58	29.4
	45	18.55	25.83	22.86	25.32	25.87	26.71	25.22	25.62	22.86	27.0
	50	18.75	23.21	26.56	24.85	28.62	27.21	24.82	28.62	24.27	26.1