GLANCE: Global Actions in a Nutshell for Counterfactual Explainability

Loukas Kavouras[1*], Eleni Psaroudaki[1,2*], Konstantinos Tsopelas[1], Dimitrios Rontogiannis[3], Nikolaos Theologitis[1], Dimitris Sacharidis[4,5], Giorgos Giannopoulos[1], Dimitrios Tomaras[6], Kleopatra Markou[7], Dimitrios Gunopulos[7], Dimitris Fotakis[2,8], Ioannis Emiris[1,7]

[1]Information Management Systems Institute, Athena Research Center, Greece
[2]National Technical University of Athens, Greece
[3]Max Planck Institute for Software Systems, Kaiserslautern, Germany
[4]Université Libre de Bruxelles, Belgium
[5]FARI Institute, Belgium
[6]Athens University of Economics and Business, Greece
[7]National and Kapodistrian University of Athens, Greece
[8]Archimedes, Athena Research Center, Greece
kavouras@athenarc.gr, epsaroudaki@mail.ntua.gr, k.tsopelas@athenarc.gr, drontogi@mpi-sws.org, n.theologitis@athenarc.gr, dimitris.sacharidis@ulb.be, giann@athenarc.gr, tomaras@aueb.gr, klmark@di.uoa.gr, dg@di.uoa.gr, fotakis@cs.ntua.gr, emiris@athenarc.gr

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. High effectiveness, measured by the fraction of the population that is provided recourse, ensures 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 maximizing interpretability. The primary challenge, therefore, is to balance these trade-offs–maximizing effectiveness, minimizing cost, while maintaining a small number of actions. We introduce GLANCE, a versatile and adaptive algorithm that employs a novel agglomerative approach, jointly considering both the feature space and the space of counterfactual actions, thereby accounting for the distribution of points in a way that aligns with the model's structure. This design enables 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. Our extensive experimental evaluation demonstrates that GLANCE consistently shows greater robustness and performance compared to existing methods across various datasets and models.

Code — https://github.com/AutoFairAthenaRC/GLANCE
Extended version — https://arxiv.org/abs/2405.18921

1 Introduction

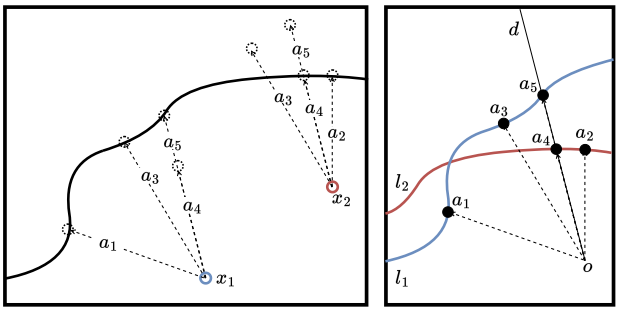
Machine learning models are increasingly deployed in critical domains such as loan approvals, hiring, and healthcare. This widespread adoption intensifies the need for transparency and interpretability in model decisions, requiring users to understand how their input features influence the outcomes and 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. 2021), their interpretability (Wachter, Mittelstadt, and Russell 2017), actionability (Ustun, Spangher, and Liu 2019), utility in fairness audits (Sharma, Henderson, and Ghosh 2020; Kavouras et al. 2023), etc. A counterfactual action, or simply an *action*, defines the specific feature changes that convert an unfavorable decision into a favorable one.

Traditionally, counterfactual explanations refer to *local explainability* tied to a specific negatively affected instance. However, many real-world scenarios require *global counterfactual explainability*, which provides shared, population-level explanations. While collecting all local counterfactuals could technically cover all affected individuals, this approach undermines interpretability, a core requirement of global explainability.

Building on prior research (Rawal and Lakkaraju 2020; Kanamori et al. 2022), we define Global Counterfactual Explanations (GCEs) as a small set of *global actions* designed to provide effective recourse for the affected population. Any global counterfactual solution must meet three objectives: (1) be composed of a small number of actions to ensure interpretability (small size), (2) minimize the cost of implementing those actions (low cost), and (3) offer recourse to as many affected individuals as possible (high effectiveness).

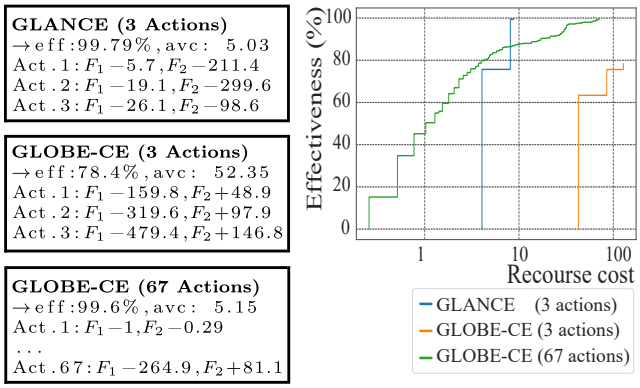
As noted by Branke (2008), the relationships between multiple optimization objectives are often complex, and aggregating them into a single objective, even though common in practice (Rawal and Lakkaraju 2020), can be problematic since they are typically non-commensurable. Framing GCEs as multi-objective optimization allows us to explore the in-

*These authors contributed equally.



(a) Feature space (b) Action space

Figure 1: A toy example depicting two negative instances x_1, x_2, and five actions. (a) The feature space; the line is the decision boundary. (b) The action space; l_1, l_2 depict the decision boundary from the perspective of x_1, x_2, respectively.



(a) Actions (b) Effectiveness vs Recourse cost

Figure 2: Comparison of GCEs on the COMPAS dataset using an `XGBoost` model. (a) GCEs from `GLANCE` (s=3) and `GLOBE-CE` (s=3 & default s=67). (b) Effectiveness–recourse cost curves show the share of individuals (y-axis) achieving recourse below a cost (x-axis); area under the curve reflects the average recourse cost. Costs follow Sec. 5.

herent *trade-offs* between effectiveness and cost, especially when the solution size is constrained.

To understand this optimization problem, assume a 2-d numerical *feature space*, and consider the two affected instances x_1, x_2 depicted in Fig. 1a. Assume that the recourse cost equals the distance to the decision boundary, drawn as a line in the figure. Observe that a_1 (resp. a_2) is the local action providing recourse for x_1 (resp. x_2) at minimum cost.

Further, consider the *action space* depicted in Fig. 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 Fig. 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).

Finding GCEs gives rise to a different optimization problem than its local counterpart and requires a nuanced exploration of the action space. 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 set of GCEs. 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.

Key Contributions. First, we formally introduce Global Counterfactual Explanations (GCEs) as small, interpretable sets of actions that provide recourse to large population subgroups. A user study confirms that smaller action sets yield more intuitive and practical explanations.

Second, we formulate the problem of finding GCEs as a multi-objective optimization task, balancing effectiveness and cost under a size constraint. We compare our formulation to prior approaches, such as that of Ley, Mishra, and Magazzeni (2023), and highlight its interpretability advantages. We prove that a restricted version of the problem is NP-hard, motivating the need for efficient algorithms.

Third, we propose `GLANCE`, a novel algorithm that clusters individuals in both feature and action space to generate diverse and coherent counterfactuals. Our results show that proper use of clustering can lead to higher quality GCEs, contradicting the claims of Kanamori et al. (2022).

Finally, extensive experiments across models and datasets show that `GLANCE` *Pareto dominates* baselines, i.e., achieves higher effectiveness and lower cost, in 57% of cases (and is dominated in only 1%). A user study further supports its practical appeal even when it is not formally dominant.

2 Related Work

Counterfactual Explanations. There has been a plethora of work focusing on counterfactual explanations (see e.g., Guidotti 2024; Verma et al. 2024, and references therein). An overview of methods of algorithmic recourse, which provides explanations and recommendations to individuals impacted by automated decision-making systems, is presented by Karimi et al. (2021). These methods can be either model-agnostic (e.g., Wachter, Mittelstadt, and Russell 2017) or model-specific (e.g., for trees as in Carreira-Perpiñán and Hada 2021), may focus on specific properties such as diversity (Mothilal, Sharma, and Tan 2020), feasibility (Ustun, Spangher, and Liu 2019), or robustness (Stepka, Stefanowski, and Lango 2025), and may follow different methodological paradigms such as optimization-based approaches (Dandl et al. 2020; Karimi, Schölkopf, and Valera 2021) or instance-based approaches (Delaney,

Greene, and Keane 2021; Brughmans, Leyman, and Martens 2024; Smyth and Keane 2020), target local or global explanations. While local counterfactuals are well defined, global ones present challenges as they must provide recourse for all individuals within a specific group in the same manner while maintaining explainability and staying true to the notion of local counterfactuals, which focuses on minimal changes. Achieving this balance globally presents notable difficulties.

Global Counterfactuals. Rawal and Lakkaraju (2020) introduced AReS, a framework for global counterfactual explanations that jointly optimizes recourse correctness, coverage, and cost, providing an interpretable summary of recourses, expressed in a two-level rule set. However, the AReS framework may fail to cover the entire population. Ley, Mishra, and Magazzeni (2022) later improved the computational efficiency with Fast AReS. In another direction, Kanamori et al. (2022) introduced CET, which partitions the space and assigns an action to each part transparently and consistently. Although effective, its computational complexity limits scalability. Warren et al. (2024) developed Group-CF, which generates counterfactuals that seek to maximize effectiveness, though it can result in higher costs.

Actions vs Directions. Ley, Mishra, and Magazzeni (2023) proposed GLOBE-CE, where global counterfactual explainability is defined differently, as a small set of action *directions* along which individuals can "move" to achieve recourse. GLOBE-CE attacks a different optimization problem from ours. For the toy example in Fig. 1, the single direction that minimizes the total recourse cost for x_1 and x_2 is the direction d depicted in Fig. 1b. This direction contains actions a_4 and a_5 that bring recourse to x_2 and x_1, respectively, with minimum cost *along* d. However, neither a_4 nor a_5 is optimal as a GCE (or for local explainability). Even the set $\{a_4, a_5\}$ is a suboptimal GCE–the set $\{a_1, a_2\}$ dominates it with equal size and effectiveness but lower total cost. In general, (1) translating GLOBE-CE outputs into GCE leads to numerous micro-actions (two in our example, but potentially as many as the individuals' number), which reduces interpretability, (2) choosing a few actions along the optimal directions may result in suboptimal GCEs (any single action along d is dominated by a in Fig. 1), and (3) since directions lack clear endpoints, they fail to specify the required magnitude of change, creating uncertainty for individuals seeking recourse and limiting real-world applicability.

Therefore, GLOBE-CE relaxes the CGE formulation, ignoring size. Fig. 2 compares the output of GLANCE to that of GLOBE-CE. By default, GLOBE-CE (green line) produces 67 micro-actions. With such a wide range of actions available, most individuals can achieve recourse through low-cost actions, resulting in high effectiveness and low average cost–at the expense of interpretability. When restricted to three actions (orange line), GLOBE-CE experiences a sharp decline in effectiveness and a substantial increase in average cost. In contrast, GLANCE (blue line), thanks to its effective exploration of the action space, produces high-quality GCEs (with high effectiveness and low average cost) without sacrificing interpretability (just three actions).

Other. Other works include the ones of Carrizosa, Ramírez-Ayerbe, and Morales (2024a,b), who use mixed-integer quadratic models for group-level explanations, and Koo, Klabjan, and Utke (2023), who employ Lagrangian methods. Research has also extended to generating global counterfactuals for graphs (Huang et al. 2023) and for auditing subgroup fairness (Kavouras et al. 2023; Fragkathoulas et al. 2025).

3 Problem Formulation and Hardness

We consider a black box binary classifier $h : \mathcal{X} \to \{-1, 1\}$, where the positive outcome is favorable and the negative outcome is unfavorable. We will focus on the set $\mathcal{X}_{\mathsf{aff}} \subseteq \mathcal{X}$ 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 $a \in \mathbb{A}$ is a set of changes to feature values, e.g., $a = \{$ country $\to$ US, education-num $\to +2\}$, which, when applied to an instance $x \in \mathcal{X}_{\mathsf{aff}}$, results in a counterfactual instance $x' = a(x)$. Every action a has a cost, denoted as $\mathsf{cost}(a, x)$, 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 $\mathbb{C}$:

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

Let $X_{\mathbb{C}} = \{x \in \mathcal{X}_{\mathsf{aff}} | h(a(x)) = 1, a \in \mathbb{C}\}$ 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, Mishra, and Magazzeni 2022), of $\mathbb{C}$ for the affected instances $\mathcal{X}_{\mathsf{aff}}$ is defined as the percentage of $\mathcal{X}_{\mathsf{aff}}$ that managed to flip their prediction using one of the actions in $\mathbb{C}$:

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

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

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

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

As it is clear, multiple sets of actions can produce recourse for $\mathcal{X}_{\mathsf{aff}}$, and the quality of such a set is a factor of the three notions we introduced: effectiveness, cost, and size. An ideal global counterfactual should maximize effectiveness while minimizing the cost and the size, based on the properties that state-of-the-art works (Rawal and Lakkaraju 2020; Kanamori et al. 2022; Ley, Mishra, and Magazzeni 2023; Huang et al. 2023) have argued are essential. The requirement for a small set of actions is to enhance the interpretability of the explanation. This can also be expressed as a constraint on the set size we can afford (Rawal and Lakkaraju 2020). Therefore, instead of minimizing the size, we constrain it to $\mathsf{size}(\mathbb{C}) \leq s$, where s is a small positive integer (set to four for our empirical evaluation). For the rest of the paper, we will use the following problem formulation:

Problem 1 (s-GCE)**.** *Given a black box model h that classifies the $\mathcal{X}_{\mathsf{aff}}$ instances to the negative class, our goal is to*

find the set $\mathbb{C} \subseteq \mathbb{A}$ that represents a solution to the following multiobjective optimization problem:

$$\textit{minimize} \quad \Big(-\,\mathsf{eff}(\mathbb{C}, \mathcal{X}_{\mathsf{aff}}),\ \mathsf{avc}(\mathbb{C}, \mathcal{X}_{\mathsf{aff}}) \Big)$$
$$\textit{s.t.} \quad \mathsf{size}(\mathbb{C}) \leq s$$

We next show that a very restricted case of s-GCE is computationally difficult in the worst case.

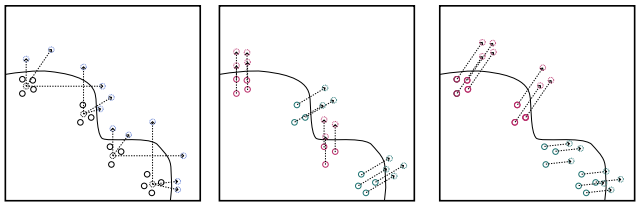
Theorem 2 (NP-hardness)**.** *The special case of s-GCE, where the model h and the set of allowable actions $\mathbb{A}$ are explicitly given and the cost is ignored, is NP-hard.*

We prove that the decision version of the following special case of s-GCE is NP-complete. The input consists of a model h, a finite set of affected instances $\mathcal{X}_{\mathsf{aff}} = \{x_1, \ldots, x_n\}$, a finite set of allowable actions $\mathbb{A} = \{a_1, \ldots, a_m\}$ and a positive fractional number $E \in \mathbb{Q}$. We seek to determine if there is a subset of actions $\mathbb{C} \subseteq \mathbb{A}$ with $\mathsf{size}(\mathbb{C}) \leq s$ and $\mathsf{eff}(\mathbb{C}, \mathcal{X}_{\mathsf{aff}}) \geq E$. We assume that the model h is defined only on $\mathcal{X}_{\mathsf{aff}} \cup \{a_i(x_j) \,|\, a_i \in \mathbb{A} \text{ and } x_j \in \mathcal{X}_{\mathsf{aff}}\}$ and its full description is given as part of the input. For the hardness part, we reduce Max s-Cover to the special case of s-GCE above. Max s-Cover is known to be NP-complete (Garey and Johnson 1979, SP5) and inapproximable in polynomial time (Feige 1998, Theorem 5.3). For the full proof, we refer the reader to the extended version of the paper.

The fact that the very restricted special case of s-GCE in Theorem 2, where the model h and the set of allowable actions $\mathbb{A}$ are explicitly given, is NP-hard (and NP-hard to approximate) indicates the computational challenges behind producing good enough solutions to s-GCE in practical settings, where we only have black-box access to h, the action cost is important and the set of allowable actions $\mathbb{A}$ is unknown and potentially infinite.

A key step of any approach is to efficiently generate a representative subset $\mathbb{A}' \subseteq \mathbb{A}$ of candidate actions, significantly larger than s, from which the final set of s actions can be carefully chosen. Since actions that perform well locally may not generalize well as GCEs (as discussed, e.g., in Fig. 1), a myopic action selection (either as representative actions in $\mathbb{A}'$ or as actions in the final solution $\mathbb{C}$) may fail to produce results anywhere close to optimal. To further clarify this point, we note that effectiveness is a nondecreasing submodular function of the chosen actions set $\mathbb{C}$ (and so well-fit for the standard greedy approach), but the average cost function is non-monotone in $\mathbb{C}$, because its definition averages over the set of instances that receive recourse under $\mathbb{C}$. At a conceptual level, increasing the size of $\mathbb{C}$ may either leave effectiveness mostly unchanged, while potentially reducing the average cost, or improve effectiveness at an increased cost for the instances just added to $X_{\mathbb{C}}$ (which might increase the average cost). Therefore, the addition of new actions must carefully balance gains in effectiveness without disproportionately increasing the average cost.

The discussion above aims to underscore the challenging trade-off between size, effectiveness, and average cost. Our method addresses this challenge by employing clustering in both feature and action space to ensure diversity and global representativeness of the preselected action set $\mathbb{A}'$, thus maintaining high effectiveness and low average cost.



(a) (b) (c)

Figure 3: Intuition behind clustering approaches. (a) First, GLANCE generates diverse candidate actions from the centroids of feature-based clusters. (b) Then, GLANCE merges clusters based on similarity in both feature and action space, grouping instances that may be further apart but can be explained by similar actions. (c) Clustering solely in feature space can yield suboptimal global actions, trading high effectiveness for high cost or low cost for low effectiveness.

4 GLANCE

We present GLANCE, a novel algorithm for solving Problem 1. GLANCE is described in Algorithm 1 in pseudocode.

Algorithm Description. The algorithm operates in two phases. *(1) Diverse Counterfactual Actions Generation.* In the first phase, the algorithm produces a set of counterfactual actions while balancing the action diversity and sparsity (i.e., the number of actions). The goal is 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. This is achieved by employing a clustering algorithm (e.g., k-means, line 1, alg. 1) to partition the feature space into k clusters, based on the assumption that nearby points are likely to share similar views of the boundary. 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. Fig. 3a shows an example with four fine-grained clusters, where three diverse actions are generated per centroid. Each action represents an alternative way to bring recourse to cluster members, giving GLANCE the flexibility to later select among them.

(2) Extracting an optimal set of CGEs. In the second phase, we identify a final set of s actions from those produced in Phase 1. The approach combines the most similar clusters with the similarity function taking into account both their feature proximity and the proximity of their respective actions. Specifically, similarity is computed by a metric D, which is the sum of two components: $d_1 : \mathcal{X}_{\mathsf{aff}} \times \mathcal{X}_{\mathsf{aff}} \to \mathbb{R}$, the distance between the centroids of each cluster, and $d_2 : 2^{\mathbb{A}} \times 2^{\mathbb{A}} \to \mathbb{R}$, the distance between the action set of the clusters. The distance d_2 can be defined as the Wasserstein distance of the sets, though simpler formulations are also possible, e.g., computing the distance between the average counterfactual actions of each cluster; in the remainder of the paper, we adopt the latter formulation of d_2 for computational efficiency. When similar actions bring recourse to individuals, it is easier to identify a common low-cost, highly ef-

Algorithm 1 GLANCE

Input: $\mathcal{X}_{\text{aff}}$, m, k, s {$\mathcal{X}_{\text{aff}}$:= affected individuals , m := number of candidate actions to generate, k := initial number of clusters, s := number of global counterfactual actions}
Output: s global counterfactual actions

1: $C \leftarrow \texttt{cluster}(\mathcal{X}_{\text{aff}}, k)$ {Cluster $\mathcal{X}_{\text{aff}}$ into k initial clusters}
2: **for** $c \in C$ **do**
3: $\boldsymbol{ca}(c) \leftarrow \texttt{actions}(\texttt{centroid}(c), m)$ {For each cluster centroid generate m actions}
4: **end for**
5: **while** $|C| > s$ **do**
6: $\{c_1, c_2\} \leftarrow (\underset{\{c_1,c_2\} \subseteq C}{\text{argmin}} (d_1(\texttt{centroid}(c_1), \texttt{centroid}(c_2)) + d_2(\boldsymbol{ca}(c_1), \boldsymbol{ca}(c_2))))$ {Find c_1,c_2 minimizing d_1+d_2}
7: $C \leftarrow \texttt{merge}(\{c_1, c_2\} \in C)$ {Replace c_1,c_2 in C with the merged cluster}
8: $\boldsymbol{ca}(\{c_1, c_2\}) \leftarrow \boldsymbol{ca}(c_1) \cup \boldsymbol{ca}(c_2)$ {Merge their action sets}
9: **end while**
10: **return** best action from each of the s clusters

fective action, even though they are dissimilar feature-wise. Fig. 3b shows an example where red-colored individuals are grouped through affinity either in the feature or in the action space. Until the desired number s of clusters is reached, the algorithm iteratively merges (line 7, alg. 1) the two clusters that minimize the total distance d_1+d_2 (line 6, alg. 1). Merging two clusters also merges their corresponding action sets, ensuring that the most effective actions for each cluster are retained (line 8, alg. 1).

Finally, for each of the s clusters, we select a single action and return the compact set of s global counterfactual actions (line 10, alg. 1). Assuming that action generation and subsequent merging have successfully grouped individuals that can achieve recourse through similar cost-efficient actions, the final step prioritizes effectiveness: we select, for each cluster, the optimal action in terms of effectiveness, among those associated with the cluster.

Approach Strengths. As noted by Kanamori et al. (2022), coarse-grained clustering approaches based solely on feature similarity lead to inadequate global actions that suffer either in cost or effectiveness (see Fig. 3c). GLANCE mitigates these limitations by jointly considering both feature and action-space proximity, enabling the identification of cost-effective recourse actions.

Although GLANCE explicitly prioritizes effectiveness, it inherently balances the effectiveness-cost trade-off. During the initial stage, refined clustering leads to diverse and typically low-cost candidate actions. In the second phase, the algorithm favors actions with maximal effectiveness, occasionally choosing higher-cost actions to achieve full coverage. For example, it may prefer an action with 100% effectiveness and an average recourse cost of 2.3 over one with 98% effectiveness and an average recourse cost of 1. However, affected individuals are ultimately free to select their preferred option among the final and applicable actions, often reducing the average recourse cost in practice.

While alternative selection strategies could explicitly incorporate cost, we deliberately avoided such heuristics. Recourse cost is often domain-specific, and GLANCE is designed to remain flexible for practitioners to adapt based on context and constraints.

Finally, GLANCE is modular and broadly applicable: it supports various clustering methods, cost metrics, and action generation techniques. Different clustering methods and action generation techniques consistently lead to a near-optimal effectiveness-cost trade-off. Across all datasets and models, GLANCE demonstrates fast runtimes, robust performance, and high-quality solutions.

Time Complexity. Let n denote the number of instances, k the number of initial clusters, s global counterfactual actions, m the number of candidate actions, and d the number of features, respectively. Define $\mathcal{T}_{\text{CF}}(d, \text{model})$ as the generation time for a single candidate counterfactual action and $\mathcal{T}_{\text{model}}$ as the black-box model's prediction time. With k-means clustering requiring I iterations to converge, the total time complexity of GLANCE is:

$$\mathcal{O}\big(kndI + km\mathcal{T}_{\text{CF}}(d, \text{model}) + k^2(k-s)d + knm\mathcal{T}_{\text{model}}\big)$$

5 Evaluation

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.

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

Models. We trained three different model types: XGBoost (XGB), Logistic Regression (LR), and Deep Neural Network (DNN). We used 5-fold cross-validation to also evaluate the robustness of the results.

Recourse Cost. Given the complexity of defining and computing the recourse cost for each action, we do not focus on the cost estimation and adhere to the guidelines established by Ley, Mishra, and Magazzeni (2023). Specifically, the cost in eq. (1) is defined as the L_1 distance between the individual and the counterfactual point, i.e., $cost(a, x) = ||x - a(x)||_1$. For categorical features, this translates to the Hamming distance. For numerical features, this translates to the sum of the absolute differences in their values. However, due to some preprocessing (following the practices in Ley, Mishra, and Magazzeni (2023)), the numerical features are split into 10 equal bins and their values are normalized to the bin size, resulting in a cost of one unit per decile.

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.

Dataset	Method	DNN		LR		XGB	
		eff	avc	eff	avc	eff	avc
Adult	Fast AReS	12.39 ± 1.06	1.0 ± 0.0	11.74 ± 2.4	1.0 ± 0.0	6.13 ± 0.42	1.0 ± 0.0
	CET	timeout	timeout	timeout	timeout	timeout	timeout
	Group-CF	100.0 ± 0.0	10.08 ± 0.03	100.0 ± 0.0	1.71 ± 0.39	96.8 ± 1.72	1.41 ± 0.54
	GLOBE-CE	99.92 ± 0.0	3.34 ± 0.29	99.92 ± 0.0	2.34 ± 0.31	82.88 ± 12.13	22.8 ± 7.87
	dGLOBE-CE	99.92 ± 0.0	10.89 ± 1.37	99.92 ± 0.0	5.91 ± 0.93	93.76 ± 1.98	64.76 ± 1.29
	GLANCE	100.0 ± 0.0	4.6 ± 0.73	100.0 ± 0.0	1.04 ± 0.07	99.85 ± 0.12	4.9 ± 3.41
COMPAS	Fast AReS	55.0 ± 0.86	1.21 ± 0.09	62.5 ± 1.82	1.24 ± 0.14	59.83 ± 3.12	1.1 ± 0.05
	CET	63.62 ± 10.35	0.96 ± 0.24	73.18 ± 4.34	1.24 ± 0.15	58.4 ± 9.3	1.06 ± 0.24
	Group-CF	100.0 ± 0.0	4.48 ± 2.53	100.0 ± 0.0	3.97 ± 2.38	100.0 ± 0.0	4.06 ± 2.10
	GLOBE-CE	100.0 ± 0.0	2.82 ± 1.06	95.74 ± 8.52	2.91 ± 0.57	87.17 ± 11.09	4.73 ± 0.92
	dGLOBE-CE	100.0 ± 0.0	7.96 ± 3.91	100.0 ± 0.0	6.71 ± 0.23	99.84 ± 0.31	12.46 ± 3.42
	GLANCE	100.0 ± 0.0	2.34 ± 0.43	100.0 ± 0.0	2.33 ± 0.38	99.51 ± 0.46	2.96 ± 0.82
Default Credit	Fast AReS	18.88 ± 2.16	1.0 ± 0.0	10.85 ± 5.45	1.07 ± 0.13	31.86 ± 5.12	1.05 ± 0.04
	CET	98.87 ± 0.62	6.32 ± 2.28	100.0 ± 0.0	3.79 ± 1.31	86.29 ± 9.94	4.5 ± 2.64
	Group-CF	79.6 ± 20.79	1.53 ± 0.62	95.4 ± 9.2	1.94 ± 1.2	95.2 ± 1.6	1.41 ± 0.64
	GLOBE-CE	81.19 ± 35.33	3.76 ± 1.35	99.94 ± 0.07	2.91 ± 1.55	83.69 ± 6.72	17.211 ± 2.22
	dGLOBE-CE	87.38 ± 18.69	5.96 ± 4.14	99.94 ± 0.07	10.38 ± 7.76	97.47 ± 0.82	42.58 ± 3.57
	GLANCE	100.0 ± 0.0	1.20 ± 0.40	100.0 ± 0.0	1.05 ± 0.11	98.13 ± 1.05	3.68 ± 1.64
German Credit	Fast AReS	52.39 ± 1.63	1.0 ± 0.0	75.27 ± 2.96	1.0 ± 0.0	51.27 ± 1.57	1.0 ± 0.0
	CET	97.3 ± 2.46	1.58 ± 0.54	96.5 ± 2.85	2.42 ± 0.24	100.0 ± 0.0	2.73 ± 0.49
	Group-CF	97.8 ± 4.4	1.85 ± 0.13	97.6 ± 2.94	9.34 ± 3.85	100.0 ± 0.0	5.78 ± 4.11
	GLOBE-CE	95.12 ± 2.04	2.11 ± 0.18	57.09 ± 20.03	2.27 ± 0.33	77.05 ± 11.26	2.52 ± 0.33
	dGLOBE-CE	97.36 ± 0.82	2.49 ± 0.27	69.89 ± 15.35	2.47 ± 0.23	86.96 ± 9.79	2.66 ± 0.77
	GLANCE	95.31 ± 3.15	1.25 ± 0.33	100.0 ± 0.0	1.21 ± 0.06	100.0 ± 0.0	1.06 ± 0.03
HELOC	Fast AReS	12.19 ± 0.58	1.03 ± 0.05	9.23 ± 1.24	1.12 ± 0.10	8.49 ± 1.32	1.16 ± 0.13
	CET	86.78 ± 10.62	8.67 ± 3.25	100.0 ± 0.0	3.57 ± 1.48	86.78 ± 6.70	12.51 ± 2.75
	Group-CF	80.4 ± 10.17	3.09 ± 0.91	90.6 ± 3.93	2.40 ± 1.38	78.4 ± 5.82	5.63 ± 1.93
	GLOBE-CE	47.18 ± 45.02	20.44 ± 24.18	99.9 ± 0.0	0.66 ± 0.10	28.33 ± 5.14	32.73 ± 0.48
	dGLOBE-CE	99.96 ± 0.05	11.07 ± 8.6	99.9 ± 0.0	1.63 ± 0.35	77.64 ± 11.51	128.0 ±0.0
	GLANCE	99.94 ± 0.05	11.24 ± 1.37	100.0 ± 0.0	1.55 ± 0.54	98.94 ± 0.66	19.99 ± 1.91

Table 1: Evaluating the effectiveness and average recourse cost of GLANCE against Fast AReS, CET, Group-CF, GLOBE-CE, and dGLOBE-CE) GCE methods for s-GCE problem with $s = 4$. s-GCE solutions with effectiveness below 80% (practicality 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 recourse cost, are highlighted in blue.

Running Time. In our experiments, GLOBE-CE and dGLOBE-CE were the fastest methods, typically delivering solutions in seconds, but required up to 30 seconds on datasets with many categorical features due to mandatory one-hot encoding, impacting scalability. GLANCE followed, typically completing in under 300 seconds. Other methods were slower, with Fast AReS ranging typically between 150–400 seconds, and peaking at 1,400 seconds in some runs. The least computationally efficient were Group-CF and CET, with maximum runtimes of 3,500 and 17,000 seconds, respectively. CET failed to solve the underlying optimization problem after 20 hours of runtime for the Adult dataset across all models.

Experimental Evaluation

Table 1 presents the summarized results of all competing methods. We compare GLANCE against the five other competitors across five datasets and three models, resulting in 75 head-to-head comparisons. However, since CET failed to solve the underlying optimization problem for the Adult dataset across all models, the final count is 72 comparisons per GLANCE method. Recall that all reported results concern the s-GCE problem for $s = 4$; The extended version of the paper presents full experimental details, along with additional results.

Pareto Dominance Evaluation. We summarize method performance by determining whether one solution dominates 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, GLANCE dominates other methods in 41 out of 72 cases (57%). It is dominated only once by dGLOBE-CE (in HELOC-DNN—cf. Table 1), where the performance of GLANCE is competitive in both terms of effectiveness and cost.

Solution Practicality. In the prior Pareto-dominance evaluation (Table 2), we compare solutions with optimal or near-optimal effectiveness to many that exhibit insufficient effec-

GLANCE	Fast AReS	CET	GroupCF	GLOBE-CE	dGLOBE-CE	$\sum_{\text{overall}}$
dominates	1/15	6/12	9/15	13/15	12/15	41/72
is dominated	0/15	0/12	0/15	0/15	1/15	1/72

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 method dominates and is dominated by competitor methods.

tiveness. 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, Mishra, and Magazzeni (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 low domination scores for Fast AReS and CET. To address this, we conduct additional experiments in which we cap GLANCE's effectiveness. Under these conditions, GLANCE achieves lower costs and dominates all baseline methods.

We consider a solution to be practical if it achieves effectiveness of at least 80%. This number could be customizable depending on the dataset and the criticality of the application. Previous papers (Ley, Mishra, and Magazzeni 2023) used smaller thresholds that were too low to be meaningful. In Table 1, impractical solutions are shown in red. GLANCE never produces impractical solutions, whereas all Fast AReS outputs are impractical, and the remaining methods return 2–5 impractical solutions each.

Robustness. We expect methods to be robust, *consistently* generating highly effective and low-cost GCEs across different data splits, which is crucial for real-world deployment. Without robustness, recourse actions can vary significantly, undermining trust and leading to unfair outcomes, especially in critical areas like healthcare or finance. Evaluating the stability of effectiveness and cost metrics across different folds is key to determining the practical applicability of a counterfactual explanation method. Standard deviation measures this stability. An effectiveness deviation above 5% indicates an inconsistency in providing recourse, while a cost deviation exceeding half the average suggests unpredictable actions; since cost scales vary by dataset, uniform robustness thresholds are not always appropriate. These fluctuations make a solution unreliable, and we highlight them in blue in Table 1. GLANCE is fully robust in effectiveness (15/15) and nearly so in cost (14/15). Even when excluding low-effectiveness (<80%) solutions, it remains the top performer; the next best methods (dGLOBE-CE, Group-CF) achieve only 11/13 and 10/13.

6 User Study

To examine how humans evaluate Global Counterfactual Explanations (GCEs), we conducted an online user study following best practices in (Ley, Mishra, and Magazzeni 2023; Chowdhury, Rahimi, and Allan 2022; Warren et al. 2024). The study had two parts and three primary objectives: (1) to assess how participants weigh trade-offs among effectiveness, average recourse cost, and size; (2) to validate metrics of practicality and robustness, by analyzing the sensitivity of user preferences in cost and effectiveness variance; and (3) to compare the perceived quality of GLANCEagainst baselines in dominated and non-dominated scenarios. We recruited 55 participants–mainly PhD students and ML researchers–from six countries. Full details appear in the extended version of the paper; below, we summarize the key findings.

In the first part of the study, participants ranked GCEs produced by three methods: GLANCE with $s = 3$, by GLOBE-CE with $s = 3$, and by GLOBE-CE in its default configuration (sizes ranging from 58 to 526). Aggregating rankings with the Borda Count (Fishburn and Gehrlein 1976) yielded the order: GLANCE $\mid s = 3 \succ$ GLOBE-CE $\succ$ GLOBE-CE $\mid s$=3 with total Borda scores of 473, 277, and 241, respectively. GLANCE was significantly preferred over both variations of GLOBE-CE. Despite higher effectiveness and lower cost, the default GLOBE-CE was less preferred, as participants favored smaller, more interpretable action sets. This highlights the central role of explanation size in perceived usefulness, even when effectiveness and cost trade-offs are present.

In the second part, participants made pairwise comparisons between anonymized methods under various trade-off scenarios. GLANCE was unanimously preferred over impractical baselines and chosen by 74.5% of participants when formally dominated by a competitor, with participants citing its lower variance as the rationale. In non-dominated scenarios, it was preferred 71.5% of the time. These preferences were statistically significant ($p < 0.01$), indicating that robustness, and especially, lower variance, can outweigh strict numerical dominance in human evaluations.

Participants' justifications revealed nuanced reasoning: over half prioritized effectiveness overall, but nearly all carefully considered trade-offs, including robustness in close cases. Some found cost difficult to interpret without domain context, underscoring the need for flexible cost metrics. These findings validate our evaluation criteria and highlight the practical value of concise, stable, and interpretable action sets, such as those produced by GLANCE.

7 Conclusion

This paper presents GLANCE, a flexible framework for generating global counterfactual explanations that optimize the trade-off between effectiveness, cost, and interpretability. Extensive experiments show that GLANCE outperforms state-of-the-art methods by producing more effective and cost-efficient counterfactuals, under size constraints. A user study further confirms that the generated explanations are more interpretable, highlighting the practical advantages of our approach.

Acknowledgments

This work was conducted while Dimitrios Rontogiannis was affiliated with the Athena Research Center and the National and Kapodistrian University of Athens.

This work has been partially supported by the European Union's Horizon Europe research and innovation programme under Grant Agreements No. 101070568 (AutoFair), No. 101181895 (WiseFood), and No. 101135826 (AIDAPT) and by the CoDiet project, which is funded by the European Union under Horizon Europe (grant number 101084642) and supported by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant number 101084642).

References

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2016. Machine Bias. *Ethics of Data and Analytics*, 254–264.

Becker, B.; and Kohavi, R. 1996. Adult. UCI Machine Learning Repository, https://doi.org/10.24432/C5XW20. Accessed: 2025-12-06.

Branke, J. 2008. *Multiobjective optimization: Interactive and evolutionary approaches*, volume 5252. Springer Science & Business Media.

Brown, K.; Doran, D.; Kramer, R.; and Reynolds, B. 2018. HELOC Applicant Risk Performance Evaluation by Topological Hierarchical Decomposition. arXiv:1811.10658.

Brughmans, D.; Leyman, P.; and Martens, D. 2024. Nice: an algorithm for nearest instance counterfactual explanations. *Data mining and knowledge discovery*, 38(5): 2665–2703.

Carreira-Perpiñán, M. Á.; and Hada, S. S. 2021. Counterfactual explanations for oblique decision trees: Exact, efficient algorithms. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 6903–6911.

Carrizosa, E.; Ramírez-Ayerbe, J.; and Morales, D. R. 2024a. Generating collective counterfactual explanations in score-based classification via mathematical optimization. *Expert Systems with Applications*, 238: 121954.

Carrizosa, E.; Ramírez-Ayerbe, J.; and Morales, D. R. 2024b. Mathematical optimization modelling for group counterfactual explanations. *European Journal of Operational Research*.

Chowdhury, T.; Rahimi, R.; and Allan, J. 2022. Equi-explanation maps: concise and informative global summary explanations. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 464–472.

Dandl, S.; Molnar, C.; Binder, M.; and Bischl, B. 2020. Multi-objective counterfactual explanations. In *International conference on parallel problem solving from nature*, 448–469. Springer.

Delaney, E.; Greene, D.; and Keane, M. T. 2021. Instance-based counterfactual explanations for time series classification. In *International conference on case-based reasoning*, 32–47. Springer.

Dua, D.; and Graff, C. 2019. UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2019).

Feige, U. 1998. A Threshold of $\ln n$ for Approximating Set Cover. *Journal of the ACM*, 45(4): 634–652.

Fishburn, P. C.; and Gehrlein, W. V. 1976. Borda's rule, positional voting, and Condorcet's simple majority principle. *Public Choice*, 79–88.

Fragkathoulas, C.; Papanikou, V.; Pitoura, E.; and Terzi, E. 2025. FACEGroup: Feasible and Actionable Counterfactual Explanations for Group Fairness. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 41–59. Springer.

Garey, M.; and Johnson, D. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman. ISBN 0-7167-1044-7.

Guidotti, R. 2024. Counterfactual explanations and how to find them: literature review and benchmarking. *Data Mining and Knowledge Discovery*, 38(5): 2770–2824.

Huang, Z.; Kosan, M.; Medya, S.; Ranu, S.; and Singh, A. 2023. Global counterfactual explainer for graph neural networks. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 141–149.

Kanamori, K.; Takagi, T.; Kobayashi, K.; and Ike, Y. 2022. Counterfactual explanation trees: Transparent and consistent actionable recourse with decision trees. In *International Conference on Artificial Intelligence and Statistics*, 1846–1870. PMLR.

Karimi, A.-H.; Barthe, G.; Schölkopf, B.; and Valera, I. 2021. A survey of algorithmic recourse: definitions, formulations, solutions, and prospects. arXiv:2010.04050.

Karimi, A.-H.; Schölkopf, B.; and Valera, I. 2021. Algorithmic recourse: from counterfactual explanations to interventions. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 353–362.

Kavouras, L.; Tsopelas, K.; Giannopoulos, G.; Sacharidis, D.; Psaroudaki, E.; Theologitis, N.; Rontogiannis, D.; Fotakis, D.; and Emiris, I. 2023. Fairness aware counterfactuals for subgroups. *Advances in Neural Information Processing Systems*, 36: 58246–58276.

Koo, J.; Klabjan, D.; and Utke, J. 2023. An inverse classification framework with limited budget and maximum number of perturbed samples. *Expert Systems with Applications*, 212: 118761.

Ley, D.; Mishra, S.; and Magazzeni, D. 2022. Global Counterfactual Explanations: Investigations, Implementations and Improvements. In *ICLR Workshop on Privacy, Accountability, Interpretability, Robustness, Reasoning on Structured Data*.

Ley, D.; Mishra, S.; and Magazzeni, D. 2023. GLOBE-CE: A Translation Based Approach for Global Counterfactual Explanations. In Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; and Scarlett, J., eds., *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 19315–19342. PMLR.

Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267: 1–38.

Mothilal, R. K.; Sharma, A.; and Tan, C. 2020. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 607–617.

Rawal, K.; and Lakkaraju, H. 2020. Beyond individualized recourse: Interpretable and interactive summaries of actionable recourses. *Advances in Neural Information Processing Systems*, 33: 12187–12198.

Sharma, S.; Henderson, J.; and Ghosh, J. 2020. Certifai: A common framework to provide explanations and analyse the fairness and robustness of black-box models. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 166–172.

Smyth, B.; and Keane, M. T. 2020. Good Counterfactuals and Where to Find Them: A Case-Based Technique for Generating Counterfactuals for Explainable AI (XAI). *ICCBR 2020: Case-Based Reasoning Research and Development*.

Stepka, I.; Stefanowski, J.; and Lango, M. 2025. Counterfactual Explanations with Probabilistic Guarantees on their Robustness to Model Change. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1277–1288. ACM.

Ustun, B.; Spangher, A.; and Liu, Y. 2019. Actionable recourse in linear classification. In *Proceedings of the conference on fairness, accountability, and transparency*, 10–19.

Verma, S.; Boonsanong, V.; Hoang, M.; Hines, K.; Dickerson, J.; and Shah, C. 2024. Counterfactual explanations and algorithmic recourses for machine learning: A review. *ACM Computing Surveys*, 56(12): 1–42.

Wachter, S.; Mittelstadt, B.; and Russell, C. 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.*, 31: 841.

Warren, G.; Delaney, E.; Guéret, C.; and Keane, M. T. 2024. Explaining multiple instances counterfactually: User tests of group-counterfactuals for xai. In *International Conference on Case-Based Reasoning*, 206–222. Springer.

Yeh, I.; and Lien, C.-H. 2009. 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.