Designing Fair, Efficient, and Interpretable Policies for Prioritizing Homeless Youth for Housing Resources

Mohammad Javad Azizi, Phebe Vayanos,
Bryan Wilder, Eric Rice, Milind Tambe

University of Southern California, CAIS Center for Artificial Intelligence in Society

Abstract. We consider the problem of designing fair, efficient, and interpretable policies for prioritizing heterogeneous homeless youth on a waiting list for scarce housing resources of different types. We focus on point-based policies that use features of the housing resources (e.g., permanent supportive housing, rapid rehousing) and the youth (e.g., age, history of substance use) to maximize the probability that the youth will have a safe and stable exit from the housing program. The policies can be used to prioritize waitlisted youth each time a housing resource is procured. Rather than focusing on a specific policy structure and fairness criteria, our framework provides the policy-maker the flexibility to select both their desired structure for the policy and their desired fairness requirements. Our approach can thus explicitly trade-off interpretability and efficiency while ensuring that fairness constraints are met. We propose a flexible data-driven mixed-integer optimization formulation for designing the policy, along with an approximate formulation which can be solved efficiently for broad classes of interpretable policies using Bender's decomposition. We evaluate our framework using real-world data from the United States homeless youth housing system. We show that our framework results in policies that are more fair than the current policy in place and than classical interpretable machine learning approaches while achieving a similar (or higher) level of overall efficiency.

1 Introduction

This paper addresses the problem of designing policies for prioritizing heterogeneous allocatees on a waiting list for scarce resources of different types so as to maximize overall efficiency. The allocatees differ in their intrinsic characteristics which, combined with the characteristics of their assigned resource, impact the efficiency of the policy. We consider a policy-maker who is able to enforce adoption of the computed policy. However, since the allocated resources are viewed as *common property*, i.e., as belonging to all members of the community, the policy should satisfy certain fairness requirements while also being interpretable, making it easy to explain why a particular assignment was made.

We are particularly motivated by the problem of allocating housing to homeless youth. In the U.S., hundreds of thousands of homeless youth are forced to live in emergency shelters or on the streets, where they run a high risk of violence, substance abuse, and sexual exploitation [27]. To help support this vulnerable population, the U.S. government directs federal resources towards programs that assist homeless youth [21]. The Homeless Management Information System (HMIS) database collects information on these services. Analysis of the HMIS database has shown that providing housing to homeless individuals produces large gains in long-term health and stability [22, 5]. Unfortunately, the number of homeless youth in the U.S. far exceeds the housing resources available [9]. Moreover, once a house is procured, there are potentially hundreds of local homeless youth that are eligible for the resource.

Given the immense difference that housing programs can make for youth, policy-makers and communities must allocate these precious resources efficiently. Most communities employ a *Coordinated Entry System* (CES) in which organizations within the same community pool both their housing resources and youth. When a housing resource becomes available, the waitlisted youth are ranked based on a set of *priority rules* and the house is allocated to the highest ranking individual [9].[1] The current prioritization tool, the TAY Triage Tool,[2] ranks the youth based on a *vulnerability score* that relies on six key experiences that increase the risk of prolonged homelessness [23]. Thus, the current policy is not directly tied to desired outcomes (due mostly to lack of outcome data at the time of design). Instead, it is purely based on factors intrinsic to each youth that determine their ability to exit homelessness without supportive housing. The increasing availability of outcome data, combined with a strategic push to better coordinate housing resources [21], constitute a significant opportunity to improve the current policy to better match supply and demand for housing resources. We now summarize the main desiderata of such a policy:

(a) *Maximize Efficiency.* Given the scarcity of housing resources , it is critical to design an efficient policy explicitly tied to outcomes for allocating houses to the homeless youth. We thus seek to improve upon the efficiency of the current policy (which is not tied to outcomes) as measured in terms of the expected number of stably housed youth at the end of the intervention.

(b) *Ensure Fairness.* Housing resources constitute common property and can prove invaluable for the homeless youth. It is thus natural to seek an allocation policy that is in some sense fair. Since there is no universally accepted measure of fairness, the proposed framework should afford the policy-maker the flexibility to select the fairness criteria that they wishes to enforce. For example, one could require that the probability of a stable exit for a youth in the system should be equal across different races, or independent of the vul-

[1] To date there is no regulation in place that enforces the current policy. However, previous analysis has shown that communities follow this policy in practice [20].

[2] Transition Age Youth (TAY) is a Service Prioritization Decision Assistance Tool that can be accessed at `http://orgcode.nationbuilder.com/tools_you_can_use`. It is incorporated work from the TAY Triage Tool of Rice [23] which can be accessed at `http://www.csh.org/wp-content/uploads/2014/02/TAY_TriageTool_2014.pdf`.

nerability score of the youth so that, independently of their backgrounds and past experiences, youth are equally likely to transition into stable housing.

(c) *Customize Interpretability.* Currently, communities can decide to comply or not with a recommended allocation. Hence, policies should be interpretable: it should be easy to explain the structure of the policy and to justify a particular matching. For example, a policy which assigns priority based on a linear scoring rule with few features may be viewed as interpretable. From our discussions with communities across the country, it appears that much of the success of the current policy can be attributed to its interpretability. Since interpretability is subjective, it is desirable that the policy-maker be able to customize the structure of the policy.

Given the above desiderata, a natural question is: how to design classes of policies that conveniently trade-off between efficiency, fairness, and interpretability? We note that this question arises in many other contexts, e.g., in the design of policies for the U.S. Kidney Allocation System and the U.S. Public Housing Program. In this paper, we propose a framework for designing such policies that is applicable to all these contexts. We now summarize our main contributions:

(a) We introduce a data-driven framework for optimizing over interpretable policies, to which a policy-maker may add flexibly defined fairness constraints. We give a mixed-integer program for computing optimal policies.

(b) To enhance scalability for complex policy classes, we give an approximate solution approach which relaxes the problem to a form amenable to Bender's decomposition. This offers significant speedup and allows us to optimize over much more sophisticated policies (e.g., multi-level decision trees, compared to linear policies in previous work).

(c) We conduct an empirical evaluation using real-world data from homeless youth across the U.S. We compare to both the status-quo TAY prioritization as well an an array of approaches from the literature. Our exact approach offers significant improvements in fairness compared to previous approaches for optimizing the same class of models, while our approximate approach allows us to improve fairness in a complementary way, by using a more expressive class. In both cases, we obtain efficiency comparable to the best (unfair) alternatives, and better than the status-quo TAY.

Literature Review. Allocation problems have been studied extensively in computer science and operations research. Much of this work considers incentives issues, where agents may misreport their true preferences to obtain a better match [1, 26, 13, 8], and the focus is on satisfying axiomatic properties (e.g. Pareto optimality or strategy-proofness). We do not consider strategic reporting since information reported by the youth can generally be verified. Instead, we focus on balancing efficiency, fairness, and interpretability. None of these goals are considered in this previous work, and all are crucial features of our domain. Another line of research considers nonstrategic online resource allocation, e.g., in the "Adwords" setting [7, 4, 19]. The focus here is on algorithms which provably approximate the optimal efficiency. By contrast, our goal is to find exactly optimal policies out of a feasible set which is constrained by fairness and interpretability. Lastly, much previous work considers organ (e.g., kidney) allocation [25, 11, 24, 2]. Our paper is most closely related to that of Bertsimas et al. [2], who optimize the U.S. Kidney Allocation System over a class of linear policies. We improve upon their approach in several ways: *(i)* we propose an *exact* formulation of the allocation problem that enables us to guarantee fairness, while Bertsimas et al. use a heuristic method that cannot guarantee fairness; *(ii)* our model is *exact*, incorporating the order in which youth and housing resources arrive, to provide accurate prioritization; *(iii)* we consider larger classes of interpretable policies (e.g., based on decision trees). These contributions translate into substantial empirical improvement.

Our work is also related to recent applications of mixed-integer programming to machine learning [3, 17, 18]. Our approach uses a mixed-integer program (MIP) to optimize over classes of policies (linear models, decision trees) also used in the machine learning literature. Previous work has shown the promise of using MIPs in machine learning; however, we are not aware of any work using such techniques to construct policies for resource allocation.

Lastly, our work is related to interpretable machine learning. Many interpretable models have been proposed, including decision rules [29, 14], decision sets [12] and generalized additive models [15, 16]. In this work, motivated by the policies currently used in the homeless youth housing system and U.S. Kidney Allocation System, we build on decision trees, which have been used to create interpretable models in many contexts [28, 10, 6]. We make two contributions compared to this previous work. First, we introduce two new model classes which generalize decision trees to respectively allow more flexible branching structures and the use of a linear scoring policy at each node of the tree (Examples 3 and 4 of Section 2.3). Second, we use these models to parameterize the allocation policy itself rather than the learning system. Thus, the final policies produced by our system are interpretable, not just the predicted success probabilities.

Notation. We denote sets (resp. random variables) using uppercase blackboard bold (resp. uppercase script) font. We denote the indicator function with $\mathcal{I}(\cdot)$.

2 Model, Problem Statement, and Interpretable Policies

2.1 System Model

We model the homeless youth housing allocation system as an infinite stream of housing resources indexed by $h \in \{1, \ldots, \infty\}$ that must be allocated to an infinite stream of youth indexed by $y \in \{1, \ldots, \infty\}$. Associated with each housing resource h is a random feature vector $\mathscr{F}_h \in \mathbb{R}^{n_{\mathrm{h}}}$ which includes, without loss of generality, the (random) arrival time $\mathscr{A}_h \in \mathbb{R}$ of the house in the system and may also include e.g., the type of house (rapid rehousing, permanent supportive housing, etc.). Accordingly, associated with each youth y is a random feature vector $\mathscr{G}_y \in \mathbb{R}^{n_{\mathrm{y}}}$ which includes the arrival time $\mathscr{T}_y \in \mathbb{R}$ of the youth in the system and may also include e.g., the intrinsic characteristics of the youth (age, history of abuse, history of substance use, etc.). Not all youth are eligible for all housing resources. Whether a youth is compatible with a particular house can be determined based on the features of the house and the youth. We let $\mathbb{M}(\mathscr{F}_h) \in \mathbb{R}^{n_\mathrm{y}}$ denote the set of all youth feature vectors that are compatible with house h. For example, we may wish to enforce that $\mathbb{M}(\mathscr{F}_h) := \{\mathscr{G}_y \ : \ \mathscr{A}_h \geq \mathscr{T}_y\}$ so that a house must be allocated immediately upon arrival in the system. Thus, youth y is eligible for house h if and only if $\mathscr{G}_y \in \mathbb{M}(\mathscr{F}_h)$. The probability of a successful outcome (a safe and stable exit) when youth y is placed in house h is denoted by $p(\mathscr{G}_y, \mathscr{F}_h)$. Similarly the probability of a successful outcome if youth y is not offered a house is denoted by $\overline{p}(\mathscr{G}_y)$. We assume that both these quantities are perfectly known (can be accurately estimated from data). In our numerical experiments, see Section 5, we discuss how these quantities can be learned from the HMIS database.

Our aim is to design interpretable parametric point-based policies that prioritize the youth for housing resources so as to maximize overall welfare. In particular, we consider parametric policies with parameter vector $\beta \in \mathbb{R}^n$ that map the features of the youth and the house to a score, see Section 2.3 for examples of such policies. We denote the score obtained for a youth y and house h for a given parameter choice β by $\pi_\beta(\mathscr{G}_y, \mathscr{F}_h)$. Then, youth y will have priority over youth y' if $\pi_\beta(\mathscr{G}_y, \mathscr{F}_h) > \pi_\beta(\mathscr{G}_{y'}, \mathscr{F}_h)$. We assume that ties are broken using a suitable tie-braking rule (e.g., at random). We thus let $\mathscr{R}$ be a permutation of the set $\{1, \ldots, \infty\}$ where the quantity $\mathscr{R}(y)$ denotes the tie-breaking score of youth y: when $\pi_\beta(\mathscr{G}_y, \mathscr{F}_h) = \pi_\beta(\mathscr{G}_{y'}, \mathscr{F}_h)$, youth y will be given priority over youth y' if and only if $\mathscr{R}(y) > \mathscr{R}(y')$.

Given a parameter vector β we now formalize the allocation process. For $t \in [0, \infty]$, we let $\mathbb{Y}(t)$ denote the set of youth that are available in the system at time t. We omit the dependence of $\mathbb{Y}(t)$ on β to minimize notational overhead. Thus, $\mathbb{Y}(0)$ denotes the initial state of the system. Suppose that a youth y arrives in the system at time t. Then $\mathbb{Y}(t+) = \mathbb{Y}(t) \cup \{y\}$. Suppose instead that house h arrives in the system at time t. Then, the house will be allocated, among all the compatible youth, to the one with the highest score (accounting for the tie breaking rule). In particular, it will be assigned to the youth[3]

$$y^\star = \operatorname*{argmax}_y \left\{ \mathscr{R}(y) \ : \ y \in \operatorname*{argmax}_y \left\{ \pi_\beta(\mathscr{G}_y, \mathscr{F}_h) \ : \ \mathscr{G}_y \in \mathbb{M}(\mathscr{F}_h), \ y \in \mathbb{Y}(t) \right\} \right\}.$$

Subsequently, youth $y^\star$ leaves the system, i.e., $\mathbb{Y}(t+) = \mathbb{Y}(t) \backslash \{y^\star\}$. Thus, given β, the allocation system generates: *(i)* an infinite random sequence $\{(\mathscr{Y}_i(\beta), \mathscr{H}_i)\}_{i=1}^\infty$ of matches, where $\mathscr{H}_i \in \{1, \ldots, \infty\}$ denotes the ith allocated house and $\mathscr{Y}_i(\beta)$ the youth to which the ith house is allocated under the policy with parameters β, and *(ii)* a set $\lim_{t \to \infty} \mathbb{Y}(t)$ of youth that will never receive a house.

[3] By construction, there will be at most one youth in this set. Moreover, since there is a severe shortage of houses for homeless youth, we can assume without loss of generality that this set will never be empty.

2.2 Problem Statement

Given the model described in Section 2.1, the expected probability of a safe and stable exit across all youth is a complicated function of the parameters β and is expressible as

$$\mathcal{P}(\beta) := \mathbb{E}\left[\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(\mathscr{G}_{\mathscr{Y}_i(\beta)}, \mathscr{F}_{\mathscr{H}_i}) + \lim_{t\to\infty} \frac{1}{|\mathbb{Y}(t)|} \sum_{\mathscr{Y}\in\mathbb{Y}(t)} \overline{p}(\mathscr{G}_{\mathscr{Y}})\right],$$

where the expectation is taken with respect to the distribution $\mathbb{P}$ of the random features of the houses and the youth, which include their arrival times and determine permissible matchings. The first (second) part in the expression above corresponds to the probability that a randomly chosen youth that received (did not receive) a house will have a safe and stable exit under the matching.

From the desiderata of the policy described in the Introduction, we wish to be able to enforce flexible fairness requirements. These requirements take the form of set-based constraints on the random sequence $\{(\mathscr{Y}_i(\beta), \mathscr{H}_i)\}_{i=1}^{\infty}$ of matches. For example, we may require that, almost surely, the proportion of all houses that provide permanent support that go to individuals with high vulnerability scores is greater than 40%. We denote by $\mathbb{F}$ the set in which the sequence of matchings is required to lie. Then, choices of β are restricted to lie in the set

$$\mathbb{S} := \{\beta \in \mathbb{B} \;:\; \{(\mathscr{Y}_i(\beta), \mathscr{H}_i)\}_{i=1}^{\infty} \in \mathbb{F},\ \mathbb{P}\text{-a.s.}\},$$

where $\mathbb{B} \subseteq \mathbb{R}^n$ captures constraints that relate to interpretability of the policy (e.g., constraints on the maximum number of features employed, see Section 3.2).

The problem faced by the policy-maker can then be expressed compactly as

$$\text{maximize } \{\ \mathcal{P}(\beta) \;:\; \beta \in \mathbb{S}\ \}. \tag{1}$$

Unfortunately, Problem (1) is very challenging to solve since the relation between β and the random sequence $\{(\mathscr{Y}_i(\beta), \mathscr{H}_i)\}_{i=1}^{\infty}$ can be highly nonlinear, while the distribution of the features of the youth and the houses is unknown. In Section 3, we will propose a data-driven mixed-integer optimization approach for learning the parameters β of the policy in Problem (1).

2.3 Interpretable Policies

In what follows, we describe several policies that can be employed in our framework and that possess attractive interpretability properties.

Example 1 (Linear Scoring Policies). A natural choice of interpretable policy are linear (or affine) in the features of the houses and the youth. These are expressible as $\pi_\beta(\mathscr{G}, \mathscr{F}) := \beta^\top(\mathscr{G}, \mathscr{F})$, where one uses one-hot encoding to encode categorical features. We note that the feature vector can naturally be augmented by nonlinear functions of the features available in the dataset as one would do in standard linear regression, see Section 5 for examples that are pertinent to this work. To reinforce interpretability, one may wish to limit the number of permitted non-zero coefficients of β.

Example 2 (Decision-Tree-Based Scoring Policies). We refer to those policies that take the form of a tree-like structure (in the spirit of decision-trees in machine learning, see Introduction) as decision-tree-based scoring policies. In each internal node of the decision-tree, a "test" is performed on a *specific feature* (e.g., if the age of the youth is smaller than 18). Each branch represents the outcome of the test, and each leaf node represents a score (score assigned to all the youth that reached that leaf). Thus, each path from root to leaf represents a classification rule that assigns a unique score to each youth. All youth that reach the same leaf will have the same score. In these policies, the policy-maker selects the depth K of the tree. The vector β collects the set of features to branch on at each node and either the set of feature values that will be assigned to each branch (for categorical features) or the cut-off values (for quantitative features). Thus, these policies partition the space of features into 2^K disjoint subsets. Letting $\mathbb{S}_\ell$ denote the set of all feature values that belong to the ℓth subset and z_ℓ the score assigned to that subset, we have $\pi_\beta(\mathscr{G}, \mathscr{F}) := \sum_{\ell=1}^{2^K} z_\ell \mathcal{I}\left((\mathscr{G}, \mathscr{F}) \in \mathbb{S}_\ell\right)$. We note that, while the these policies are exponential in K, to maximize interpretability, K should be kept as small as possible. To improve interpretability further, one may require that each feature be branched on at most once.

Example 3 (Decision-Tree-Based Policies enhanced with Linear Branching). A natural variant of the policies from Example 2 is one where policies take again the form of a tree-like structure, but this time, each "test" involves a linear function of several features (e.g., whether a vulnerability measure of the youth is greater than 10). In this setting, the vector β collects the coefficients of the linear function at each node and the cut-off values of the branching.

Example 4 (Decision-Tree-Based Policies enhanced with Linear Leafing). Another variant of the policies from Example 2 is one where rather than having a common score for all youth that reach a leaf, instead, a different linear scoring rule is employed on each branch. Thus, $\pi_\beta(\mathscr{G}, \mathscr{F}) := \sum_{\ell=1}^{2^K} [\beta_{\mathrm{y},\ell}^\top \mathscr{G} + \beta_{\mathrm{h},\ell}^\top \mathscr{F}] \mathcal{I}\left((\mathscr{G}, \mathscr{F}) \in \mathbb{S}_\ell\right)$, and the parameters to be optimized are augmented with $\beta_{\mathrm{y},\ell}$ and $\beta_{\mathrm{h},\ell}$ for each ℓ.

In addition to the examples above, one may naturally also consider decision-tree-based policies enhanced with both linear branching and linear leafing.

3 Data-Driven Framework for Policy Calibration

In Section 2.1, we proposed a model for the homeless youth housing allocation system and a mathematical formulation (Problem (1)) of the problem of designing fair, efficient, and interpretable policies for allocating these scarce resources. This problem is challenging to solve as it requires knowledge of the distribution of the uncertain parameters. In this section, we propose a data-driven mixed-integer optimization formulation for learning the parameters β of the policy, thus approximating Problem (1).

3.1 A Data-Driven Mixed Integer Optimization Problem

We assume that we have at our disposal a dataset that consists of: *(i)* a (finite) stream $\mathbb{H}$ of housing resources that became available in the past and their associated feature vectors $f_h \in \mathbb{R}^{n_{\rm h}}$, $h \in \mathbb{H}$; and *(ii)* a (finite) stream $\mathbb{Y}$ of youth waitlisted for a house and their associated feature vectors $g_y \in \mathbb{R}^{n_{\rm y}}$. We let α_h (resp. τ_y) denote the arrival time of house h (resp. youth y) in the system. For convenience, we define

$$\mathbb{C} := \{(y,h) \in \mathbb{Y} \times \mathbb{H} \; : \; g_y \in \mathbb{M}(f_h)\},$$

and also let $p_{yh} := p(g_y, f_h)$, $\overline{p}_y := \overline{p}(g_y)$, and $\rho_y := \mathscr{R}(y)$. Using this data, the problem of learning (estimating) the parameters β of the policy can be cast as a mixed-integer optimization problem. The main decision variables of the problem are the policy parameters β. Consider the MIP

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{y \in \mathbb{Y}} \left[\sum_{h \in \mathbb{H}} p_{yh} x_{yh} + \overline{p}_y \left(1 - \sum_{h \in \mathbb{H}} x_{yh} \right) \right] \\ \text{subject to} & \pi_{yh} = \pi_\beta(g_y, f_h), \quad \forall y \in \mathbb{Y}, \; h \in \mathbb{H} \\ & \forall y \in \mathbb{Y}, \; h \in \mathbb{H}, \\ & x_{yh} = 1 \Leftrightarrow \left\{ \begin{array}{l} (y,h) \in \mathbb{C}, \quad \displaystyle\sum_{h' \neq h : \alpha_{h'} \leq \alpha_h} x_{yh'} = 0, \quad \text{and} \\ \forall y' : (y', h) \in \mathbb{C} \;\; \text{and} \displaystyle\sum_{h' : \alpha_{h'} \leq \alpha_h} x_{y'h'} = 0, \\ (\pi_{yh} > \pi_{y'h}) \text{ or } (\pi_{yh} = \pi_{y'h} \text{ and } \rho_y > \rho_{y'}) \end{array} \right\} \\ & \beta \in \mathbb{B}, \; x \in \mathbb{F}, \; x_{yh} \in \{0,1\} \; \forall y \in \mathbb{Y}, \; h \in \mathbb{H}. \end{array} \tag{2}$$

In addition to β, the decision variables of the problem are the assignment variables x and the scoring variables π. Thus, x_{yh} indicates whether house h is allocated to youth y under the policy with parameters β and π_{yh} corresponds to the score of youth y for house h under the policy. The first (second) part of the objective function corresponds to the probability that youth y will be successful if they do (do not) receive a house under the policy with parameters β. The first constraint in the formulation defines the scoring variables in terms of the parameters β and the features of the youth and the house. The second constraint is used to define the assignment variables in terms of the scores: it stipulates that youth y will receive house h if and only if: *(i)* the two are compatible, *(ii)* youth y is still on the waitlist, and *(iii)* youth y has higher priority over all youth that have not yet been allocated a house in the sense that they score higher using the scoring policy dictated by β (combined with the tie-breaking rule).

Next, we show that if $\mathbb{F}$ is polyhedral, Problem (2) can be solved as a mixed-integer linear optimization problem provided one can define the scores π using linear inequalities. The main decision variables of this problem are the policy parameters β. Consider the MIP

$$\text{maximize} \quad \sum_{y \in \mathbb{Y}} \left[\sum_{h \in \mathbb{H}} p_{yh} x_{yh} + \overline{p}_y \left(1 - \sum_{h \in \mathbb{H}} x_{yh} \right) \right] \tag{3a}$$

$$\text{subject to} \quad \pi_{yh} = \pi_\beta(g_y, f_h) \quad \forall y \in \mathbb{Y}, h \in \mathbb{H} \tag{3b}$$

$$\sum_{h \in \mathbb{H}} x_{yh} \le 1 \quad \forall y \in \mathbb{Y}, \quad \sum_{y \in \mathbb{Y}} x_{yh} \le 1 \quad \forall h \in \mathbb{H} \tag{3c}$$

$$z_{yh} = \sum_{h' \in \mathbb{H} \setminus \{h\}} \mathcal{I}\left(\alpha_{h'} \le \alpha_h\right) x_{yh'} \quad \forall y \in \mathbb{Y}, h \in \mathbb{H} \tag{3d}$$

$$\pi_{yh} - \pi_{y'h} = v^+_{yy'h} - v^-_{yy'h} \quad \forall y, y' \in \mathbb{Y}, h \in \mathbb{H} \tag{3e}$$

$$v^+_{yy'h} \le M u_{yy'h} \quad \forall y, y' \in \mathbb{Y}, h \in \mathbb{H} \tag{3f}$$

$$v^-_{yy'h} \le M(1 - u_{yy'h}) \quad \forall y, y' \in \mathbb{Y}, h \in \mathbb{H} \tag{3g}$$

$$v^+_{yy'h} + v^-_{yy'h} \ge \epsilon(1 - u_{yy'h}) \quad \forall y, y' \in \mathbb{Y}, h \in \mathbb{H} \,:\, \rho_y > \rho_{y'} \tag{3h}$$

$$v^+_{yy'h} + v^-_{yy'h} \ge \epsilon u_{yy'h} \quad \forall y, y' \in \mathbb{Y}, h \in \mathbb{H} \,:\, \rho_{y'} > \rho_y \tag{3i}$$

$$x_{yh} \le u_{yy'h} + z_{y'h} \quad \forall y, y' \in \mathbb{Y}, h \in \mathbb{H} \tag{3j}$$

$$1 - z_{yh} \le \sum_{y' : (y', h) \in \mathbb{C}} x_{y'h} \quad \forall (y, h) \in \mathbb{C} \tag{3k}$$

$$x_{yh} = 0 \quad \forall y \in \mathbb{Y}, h \in \mathbb{H} \,:\, (y, h) \notin \mathbb{C} \tag{3l}$$

$$x \in \mathbb{F} \tag{3m}$$

$$v^+_{yy'h},\ v^-_{yy'h} \ge 0,\ u_{yy'h} \in \{0, 1\}, \quad \forall y, y' \in \mathbb{Y}, h \in \mathbb{H} \tag{3n}$$

$$x_{yh}, z_{yh} \in \{0, 1\} \quad \forall y \in \mathbb{Y}, h \in \mathbb{H}. \tag{3o}$$

In addition to the policy parameters β, the score variables π and assignment variables x, Problem (3) employs several auxiliary variables (z, v^+, v^-, and u) that are used to uniquely define the assignment variables x based on the scores π. The variables z indicate whether a youth is still waiting at the time a house arrives: $z_{yh} = 1$ if and only if youth y has been allocated a house on or before time α_h. The non-negative variables $v^+_{yy'h}$ and $v^-_{yy'h}$ denote the positive and negative parts of $\pi_{yh} - \pi_{y'h}$. Finally, the variables u are prioritization variables: $u_{yy'h} = 1$ if and only if either youth y has a higher score than youth y' for house h (i.e., $\pi_{yh} > \pi_{y'h}$) or they have the same score but youth y has priority due to tie-breaking (i.e., $\pi_{yh} = \pi_{y'h}$ and $\rho_y > \rho_{y'}$).

Problems (2) and (3) share the same objective function. An interpretation of the constraints in Problem (3) is as follows. Constraint (3b) is used to define the variables π_{yh}. Constraints (3c) are classical matching constraints. Constraint (3d) is used to define the variables z. Constraints (3e)-(3i) are used to define the prioritization variables u in term of the scores π: constraint (3e) defines v^+_{yh} and v^-_{yh} as the positive and negative parts of $\pi_{yh} - \pi_{y'h}$, respectively. Constraints (3f) and (3g) stipulate that $u_{yy'h}$ must be 1 if $\pi_{yh} > \pi_{y'h}$ and must be 0 if $\pi_{yh} < \pi_{y'h}$. Constraints (3h) and (3i) ensure that if π_{yh} and $\pi_{y'h}$ are equal then $u_{yy'h} = 1$ if and only if $\rho_y > \rho_{y'}$. Constraint (3j) stipulates that youth y cannot receive house h if there is another youth y' that is still waiting for a house and that has priority for house h over y. Constraint (3k) ensures that if a youth that is compatible with a house has not been served at the time a house arrives, then the house must be assigned to a compatible youth. Finally, constraint (3l) ensures that youth are only assigned houses they are eligible for.

If the scoring variables π can be defined in terms of the policy parameters β (constraint (3b)) using integer linear constraints, then Problem (3) is an MILP.

3.2 Expressing the Policy Values using Integer Linear Constraints

We now show that for all the interpretable policies from Section 2.3, the scoring variables π can be defined using finitely many integer linear constraints, implying that Problem (3) reduces to a mixed-integer linear program if $\mathbb{F}$ is polyhedral.

Example 5 (Linear Scoring Policies). In the case of the linear policies (Example 1), constraint (3b) is equivalent to

$$\pi_{yh} = \beta^\top (g_y, f_h) \quad \forall y \in \mathbb{Y},\ h \in \mathbb{H}. \tag{4}$$

To increase interpretability, one may impose a limit K on the number of features employed in the policy by letting

$$\mathbb{B} = \left\{ \beta \in \mathbb{R}^n \ :\ \exists \kappa \in \{0,1\}^n \text{ with } \sum_{i=1}^{n} \kappa_i \leq K,\ |\beta_i| \leq \kappa_i,\ i = 1, \ldots, n \right\},$$

where $\kappa_i = 1$ if and only if the ith feature is employed.

Example 6 (Decision-Tree-Based Scoring Policies). For decision-tree-based scoring policies (Example 2), constraint (3b) is equivalent to

$$\pi_{yh} = \sum_{\ell \in \mathbb{L}} z_\ell x_{yh\ell} \quad \forall y \in \mathbb{Y},\ h \in \mathbb{H}, \tag{5}$$

where $\mathbb{L}$ denotes the set of all leafs in the tree, the variables x are leaf assignment variables such that $x_{yh\ell} = 1$ if and only if the feature vectors of youth y and house h belong to leaf ℓ, and z are score variables such that z_ℓ corresponds to the score assigned to leaf ℓ. The above constraint is bilinear but can be linearized using standard techniques. Next, we illustrate that the leaf assignment variables can be defined using a system of integer linear inequalities.

Let $\mathbb{I}_\mathrm{c}$ and $\mathbb{I}_\mathrm{q}$ denote the sets of all categorical and quantitative features (of both the youth and the houses), respectively. Also, let $\mathbb{I} := \mathbb{I}_\mathrm{c} \cup \mathbb{I}_\mathrm{q}$. Denote with d_{yhi} the value attained by the ith feature of the pair (y, h) and for $i \in \mathbb{I}_\mathrm{c}$ let $\mathbb{S}_i$ collect the possible levels attainable by feature i. Finally, let $\mathbb{V}$ denote the set of all branching nodes in the tree and for $\nu \in \mathbb{V}$, let $\mathbb{L}^\mathrm{r}(\nu)$ (resp. $\mathbb{L}^\mathrm{l}(\nu)$) denote all the leaf nodes that lie to the right (resp. left) of node ν.

Consider the system

$$\sum_{i \in \mathbb{I}} p_{\nu i} = 1 \quad \forall \nu \in \mathbb{V} \tag{6a}$$

$$q_\nu - \sum_{i\in\mathbb{I}_\mathrm{q}} p_{\nu i} d_{yhi} = g^+_{yh\nu} - g^-_{yh\nu} \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H} \tag{6b}$$

$$g^+_{yh\nu} \le M w^\mathrm{q}_{yh\nu} \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H} \tag{6c}$$

$$g^-_{yh\nu} \le M(1 - w^\mathrm{q}_{yh\nu}) \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H} \tag{6d}$$

$$g^+_{yh\nu} + g^-_{yh\nu} \ge \epsilon(1 - w^\mathrm{q}_{yh\nu}) \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H} \tag{6e}$$

$$x_{yh\ell} \le 1 - w^\mathrm{q}_{yh\nu} + \sum_{i\in\mathbb{I}_\mathrm{c}} p_{\nu i} \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H},\ \ell\in\mathbb{L}^\mathrm{r}(\nu) \tag{6f}$$

$$x_{yh\ell} \le w^\mathrm{q}_{yh\nu} + \sum_{i\in\mathbb{I}_\mathrm{c}} p_{\nu i} \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H},\ \ell\in\mathbb{L}^\mathrm{l}(\nu) \tag{6g}$$

$$s_{\nu ik} \le p_{\nu i} \quad \forall \nu\in\mathbb{V},\ i\in\mathbb{I}_\mathrm{c},\ k\in\mathbb{S}_i \tag{6h}$$

$$w^\mathrm{c}_{yh\nu} = \sum_{i\in\mathbb{I}_\mathrm{c}} \sum_{k\in\mathbb{S}_i} s_{\nu ik} \mathcal{I}\left(d_{yhi} = k\right) \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H} \tag{6i}$$

$$x_{yh\ell} \le w^\mathrm{c}_{yh\nu} + \sum_{i\in\mathbb{I}_\mathrm{q}} p_{\nu i} \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H},\ \ell\in\mathbb{L}^\mathrm{r}(\nu) \tag{6j}$$

$$x_{yh\ell} \le 1 - w^\mathrm{c}_{yh\nu} + \sum_{i\in\mathbb{I}_\mathrm{q}} p_{\nu i} \quad \forall \nu\in\mathbb{V},\ y\in\mathbb{Y},\ h\in\mathbb{H},\ \ell\in\mathbb{L}^\mathrm{l}(\nu) \tag{6k}$$

$$\sum_{\ell\in\mathbb{L}} x_{yh\ell} = 1 \quad \forall y\in\mathbb{Y},\ h\in\mathbb{H} \tag{6l}$$

in variables $q_\nu \in \mathbb{R}$, $g^+_{yh\nu},\ g^-_{yh\nu} \in \mathbb{R}_+$, and $x_{yh\ell},\ p_{\nu i},\ w^\mathrm{q}_{yh\nu},\ w^\mathrm{c}_{yh\nu},\ s_{\nu ik} \in \{0,1\}$ for all $y\in\mathbb{Y}$, $h\in\mathbb{H}$, $\ell\in\mathbb{L}$, $\nu\in\mathbb{V}$, $i\in\mathbb{I}$, $k\in\mathbb{S}_i$.

An interpretation of the variables is as follows. The variables p indicate the feature that we branch on at each node. Thus, $p_{\nu i} = 1$ if and only if we branch on feature i at node ν. The variables q_ν, $g^+_{yh\nu}$, $g^-_{yh\nu}$, and $w^\mathrm{q}_{yh\nu}$ are used to bound $x_{yh\ell}$ based on the branching decisions at each node ν, whenever branching is performed on a *quantitative* feature at that node. The variable q_ν corresponds to the cut-off value at node ν. The variables $g^+_{yh\nu}$ and $g^-_{yh\nu}$ represent the positive and negative parts of $q_\nu - \sum_{i\in\mathbb{I}_\mathrm{q}} p_{\nu i} d_{yhi}$, respectively. Whenever branching occurs on a quantitative feature, the variable $w^\mathrm{q}_{yh\nu}$ will equal 1 if and only if $q_\nu \ge \sum_{i\in\mathbb{I}_\mathrm{q}} p_{\nu i} d_{yhi}$, in which case the data point (y,h) must go left in the branch. The variables $w^\mathrm{c}_{yh\nu}$ and $s_{\nu ik}$ are used to bound $x_{yh\ell}$ based on the branching decisions at each node ν, whenever branching is performed on a *categorical* feature at that node. Whenever we branch on categorical feature $i\in\mathbb{I}_\mathrm{c}$ at node ν, the variable $s_{\nu ik}$ equals 1 if and only if the points such that $d_{yhi} = k$ must go left in the branch. If we do not branch on feature i, then the variable $s_{\nu ik}$ will equal zero. The variable $w^\mathrm{c}_{yh\nu}$ will equal 1 if and only if we branch on a categorical feature at node ν and data point (y,h) must go left at the node.

An interpretation of the constraints is as follows. Constraint (6a) ensures that only one variable is branched on at each node. Constraints (6b)-(6g) are used to bound $x_{yh\ell}$ based on the branching decisions at each node ν, whenever branching is performed on a *quantitative* feature at that node. Constraints (6b)-(6e) are used to define $w^{\mathrm{q}}_{yh\nu}$ to equal 1 if and only if $q_\nu \geq \sum_{i\in\mathbb{I}_{\mathrm{q}}} p_{\nu i} d_{yhi}$. Constraint (6f) stipulates that if we branch on a quantitative feature at node ν and data point (y,h) goes left at the node (i.e., $w^{\mathrm{q}}_{yh\nu}=1$), then the data point cannot reach any leaf node that lies to the right of the node. Constraint (6g) is symmetric to (6f) for the case when the data point goes right at the node. Constraints (6h)-(6k) are used to bound $x_{yh\ell}$ based on the branching decisions at each node ν, whenever branching is performed on a *categorical* feature at that node. Constraint (6h) stipulates that if we do not branch on feature i at node ν, then $s_{\nu ik}=0$. Constraint (6i) is used to define $w^{\mathrm{c}}_{yh\nu}$ such that it is equal to 1 if and only if we branch on a particular feature i, the value attained for that feature by data point (y,h) is k and data points with feature value k are assigned to the left branch of the node. Constraints (6j) and (6k) mirror constraints (6f) and (6g), respectively, for the case of categorical features.

Example 7 (Decision-Tree-Based Policies enhanced with Linear Branching). For decision-tree-based policies enhanced with linear branching (Example 3), constraint (3b) can be expressed in terms of linear inequalities using a variant of the formulation from Example 6. Specifically, one can convert the dataset to have only quantitative features using one hot encoding and subsequently enforce constraints (5) and (6b)-(6g) to achieve the desired model.

Example 8 (Decision-Tree-Based Policies enhanced with Linear Leafing). For decision-tree-based policies enhanced with linear leafing (Example 4), constraint (3b) can be expressed in terms of linear inequalities using a variant of the formulation from Example 6 by replacing constraint (5) with

$$\pi_{yh} = \sum_{\ell\in\mathbb{L}} [\beta_{\mathrm{y},\ell}^\top g_y + \beta_{\mathrm{h},\ell}^\top f_h] x_{yh\ell} \quad \forall y\in\mathbb{Y},\ h\in\mathbb{H}. \tag{7}$$

4 Approximate Solution Approach

Albeit exact, the data-driven MIP (3) scales with the number of youth and houses in the system. In this section, we propose an approximate solution approach that relies upon and generalizes the one from [2] to decision-tree-based policies, see Examples 1-4. Consider the following relaxation of Problem (3).

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{y\in\mathbb{Y}} \left[\sum_{h\in\mathbb{H}} p_{yh} x_{yh} + \overline{p}_y \left(1 - \sum_{h\in\mathbb{H}} x_{yh} \right) \right] \\ \text{subject to} & \displaystyle\sum_{h\in\mathbb{H}} x_{yh} \leq 1 \quad \forall y\in\mathbb{Y}, \quad \sum_{y\in\mathbb{Y}} x_{yh} \leq 1 \quad \forall h\in\mathbb{H} \\ & x_{yh} = 0 \quad \forall y\in\mathbb{Y},\ h\in\mathbb{H}\ :\ (y,h)\notin\mathbb{C} \\ & x\in\mathbb{F},\ x_{yh}\geq 0 \quad \forall y\in\mathbb{Y},\ h\in\mathbb{H} \end{array} \tag{8}$$

Contrary to Problem (3) in which the matching is guided by the policy with parameters β, this formulation allows for arbitrary matches. Moreover, integrality constraints on x are relaxed so that x_{yh} can now be interpreted as the probability that house h is offered to youth y. For convenience, we henceforth assume that the set of fair matchings is expressible as $\mathbb{F} := \{x : Ax \leq b\}$ for some matrix A and vector b. Moreover, we let λ denote the vector of optimal dual multipliers of the fairness constraints in (8).

In the spirit of [2], we define the quantity $C_{yh} := p_{yh} - \overline{p}_y - (\lambda^\top A)_{(y,h)}$ and propose to learn β to approximate C_{yh} using π_{yh}, the score for this match. This problem is expressible as

$$\text{minimize} \left\{ \sum_{y \in \mathbb{Y}} \sum_{h \in \mathbb{H}} |C_{yh} - \pi_{yh}| \ : \ \text{Constraint (3b)} \right\}. \tag{9}$$

As discussed in Section 3.2, for all interpretable policies proposed in Section 2.3, constraint (3b) is equivalent to a finite set of linear inequality constraints involving binary variables. Thus, Problem (9) is equivalent to an MILP (and LP for the case of linear policies). Problem (9) is significantly more tractable than (3). While it can still be challenging to solve for large datasets, in the case of tree-based policies (Examples 2, 3, and 4), it presents an attractive decomposable structure that enables us to solve it efficiently using Bender's decomposition. Thus, x, z, and π are variables of the subproblem and all other variables are decided in the master. Note that it can be shown that integrality constraints in the subproblem may be relaxed to yield an equivalent problem.

5 Numerical Study

We showcase the performance of our approach to design allocation policies for the U.S. homeless youth. We build interpretable policies that aim to maximize efficiency while ensuring fairness across vulnerability (NST) scores and across races, in turn. We use real-world data (10,922 homeless youth and 3474 housing resources) from the HMIS database obtained from Ian De Jong as part of a working group called "Youth Homelessness Data, Policy, Research." The dataset includes 54 features for the youth and each house is of one of two types (rapid rehousing (RRH) or permanent supportive housing (PSH)), see [5]. A youth is considered to have a successful outcome if they are housed one year later. We use 80% of data for training and show results on a 20% testing set. We use the training set to learn (using CART) the success probabilities that are fed in our models and to identify the five most significant features. We compare our proposed approach to several baselines: *(i)* the status-quo policy TAY; *(ii)* random allocation (Random); *(iii)* the (interpretable) machine learning approaches without fairness from [5] (Linear and Logistic Regression and CART); *(iv)* the linear scoring policies with relaxed fairness constraints originally proposed in [2] (Linear RF). To these baselines, we add: *(i)* Decision-tree-based policies with relaxed fairness constraints (Decision-Tree RF); *(ii)* Decision-tree-based policies with linear leafing (depth 1 and 2) with relaxed fairness constraints (Decision-Tree LL RF); *(iii)* Linear scoring policies with (exact) fairness constraints computed using MILP (3) (Linear EF).

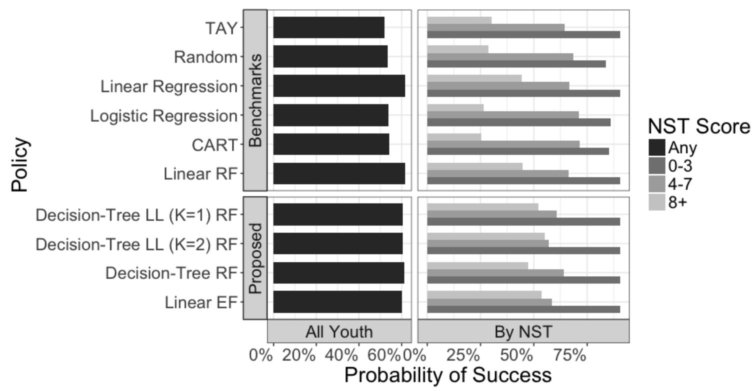

Fig. 1: Success probability across all youth (left) and by vulnerability level (right) when fairness across vulnerability levels is desired.

Fairness across NST Scores. Motivated by TAY which provides the most supportive resources to the most vulnerable youth, we enforce fairness with respect to vulnerability score: independently of their score, youth should be equally likely to transition to a fair and stable exit. We enforce fairness across two groups which were found to have very different chances to remain homeless in the long run: youth with scores 4-7 and 8+, respectively. Youth with scores below 4 are excluded since they have a *higher* estimated success probability when not offered housing. Figure 1 shows the success probability of youth under each policy. The baselines TAY, Random, Logistic Regression, and CART are all very unfair: the probability of success for youth with scores 8+ is uniformly below 30%, while lower risk youth with scores 4-7 have success probability higher than 60%. Linear Regression performs considerably better and introducing relaxed fairness constraints (Linear RF) does not yield any improvement. Our proposed policies outperform all benchmarks in terms of fairness and do so at marginal cost to overall efficiency. Figure 2 shows the percentage of each type of house allocated to each group under each policy. The current policy allocates the most resource-intensive resources (PSH) to the highly vulnerable individuals and the RRH resources to individuals scoring 4-7. Our analysis however shows that some high risk individuals can improve their chances of a successful outcome by receiving an RRH resource. Thus, our policies allocate some RRH (resp. PSH) houses to high (resp. low) risk individuals, resulting in policies that *benefit the most vulnerable* youth, see Figures 1 and 2. Lastly, Table 1 shows the runtime required to solve each problem.[4] Exact formulations require more runtime than approximations, and more sophisticated policies require greater runtime. Moreover, there are significant benefits in employing our proposed decomposition approach.

Fairness across Races. Motivated by the desire to avoid racial discrimination, we seek policies that are fair across races. The results are summarized in Figure 3

[4] These experiments were run on a 2.0GHz Intel Core i7 processor machine with 4GB RAM and all optimization problems were solved with Gurobi v7.0.

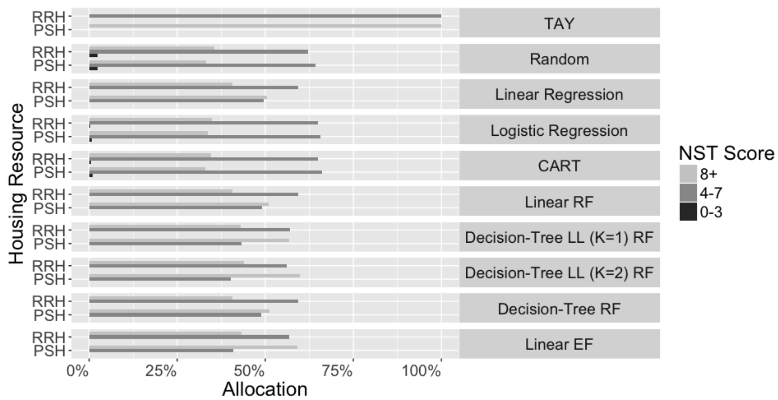

Fig. 2: Housing resources allocated by vulnerability level when fairness across vulnerability levels is desired.

Fairness Constraints	Type of Policy	Decomposition Used	Solver Time (Seconds)
Relaxed	Linear	N/A	932.57
Relaxed	Decision-Tree	Yes (No)	3570.12 (7105.11)
Relaxed	Decision-Tree LL	Yes (No)	9031.32 (14045.45)
Exact	Linear	N/A	36400.98

Table 1: Solver times for the proposed approaches for solving to optimality when fairness across vulnerability levels is desired.

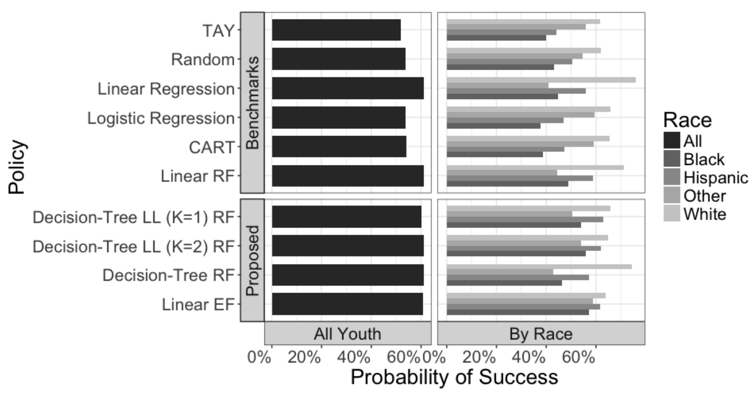

Fig. 3: Success probability across all youth (left) and by race (right) when fairness across races is desired

which shows that the current policy and classical machine learning approaches are unfair, with "Whites" having higher success probability than "Blacks" and "Hispanics." In contrast, our proposed policies, in particular Linear EF outperform significantly the state of the art at marginal cost to overall efficiency.

Bibliography

[1] M. Adamczyk, P. Sankowski, and Q. Zhang. Efficiency of truthful and symmetric mechanisms in one-sided matching. In *International Symposium on Algorithmic Game Theory*, pages 13–24. Springer, 2014.

[2] D. Bertsimas, V. F. Farias, and N. Trichakis. Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research*, 61(1):73–87, 2013.

[3] D. Bertsimas, A. King, and R. Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44(2):813–852, 2016.

[4] N. Buchbinder, K. Jain, and J. Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. *Algorithms–ESA 2007*, pages 253–264, 2007.

[5] H. Chan, E. Rice, P. Vayanos, M. Tambe, and M. Morton. Evidence from the past: Ai decision aids to improve housing systems for homeless youth. In *AAAI 2017 Fall Symposium Series*, 2017.

[6] Z. Che, S. Purushotham, R. Khemani, and Y. Liu. Interpretable deep models for icu outcome prediction. In *AMIA Annual Symposium Proceedings*, volume 2016, page 371. American Medical Informatics Association, 2016.

[7] N. R. Devanur, K. Jain, B. Sivan, and C. A. Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 29–38. ACM, 2011.

[8] S. Dughmi and A. Ghosh. Truthful assignment without money. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 325–334. ACM, 2010.

[9] Housing and Urban Development (HUD). Coordinated Entry Policy Brief. https://www.hudexchange.info/resources/documents/ Coordinated-Entry-Policy-Brief.pdf, 2015.

[10] L.-T. Huang, M. M. Gromiha, and S.-Y. Ho. iptree-stab: interpretable decision tree based method for predicting protein stability changes upon mutations. *Bioinformatics*, 23(10):1292–1293, 2007.

[11] N. Kong, A. J. Schaefer, B. Hunsaker, and M. S. Roberts. Maximizing the efficiency of the U.S. liver allocation system through region design. *Management Science*, 56(12):2111–2122, 2010.

[12] H. Lakkaraju, S. H. Bach, and L. Jure. Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1675–1684. ACM, 2016.

[13] J. D. Leshno. Dynamic matching in overloaded waiting lists. 2017.

[14] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics*, 9(3):1350–1371, 2015.

[15] Y. Lou, R. Caruana, and J. Gehrke. Intelligible models for classification and regression. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 150–158. ACM, 2012.

[16] Y. Lou, R. Caruana, J. Gehrke, and G. Hooker. Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 623–631. ACM, 2013.

[17] R. Mazumder and P. Radchenko. The discrete dantzig selector: Estimating sparse linear models via mixed integer linear optimization. *IEEE Transactions on Information Theory*, 63(5):3053–3075, 2017.

[18] R. Mazumder, P. Radchenko, and A. Dedieu. Subset selection with shrinkage: Sparse linear modeling when the snr is low. *arXiv preprint arXiv:1708.03288*, 2017.

[19] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22, 2007.

[20] The U.S. Department of Housing, OFFICE OF COMMUNITY PLANNING Urban Development, and DEVELOPMENT. The 2016 annual homeless assessment report (ahar) to congress, 2016.

[21] United States Interagency Council on Homelessness. Opening doors: federal strategic plan to prevent and end homelessness, 2010.

[22] C. L. Pearson. The applicability of housing first models to homeless persons with serious mental illness: Final report. Technical report, U.S. Dept. of Housing and Urban Development, Office of Policy Development and Research, 2007.

[23] E. Rice. Assessment Tools for Prioritizing Housing Resources for Homeless Youth. https://static1.squarespace.com/static/56fb3022d210b8911 56b3948/t/5887e0bc8419c20e9a7dfa81/1485299903906/ Rice-Assessment-Tools-for-Youth-2017.pdf, 2017.

[24] X. Su and S. A. Zenios. Patient choice in kidney allocation: A sequential stochastic assignment model. *Operations research*, 53(3):443–455, 2005.

[25] X. Su and S. A. Zenios. Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. *Management Science*, 52(11):1647–1660, 2006.

[26] N. Thakral. Matching with stochastic arrival. In *EC*, page 343, 2015.

[27] P. A. Toro, T. M. Lesperance, and J. M. Braciszewski. *The heterogeneity of homeless youth in America: Examining typologies*. National Alliance to End Homelessness: Washington, DC, 2011.

[28] G. Valdes, J. Marcio Luna, E. Eaton, and C. B. Simone. Mediboost: a patient stratification tool for interpretable decision making in the era of precision medicine. *Scientific reports*, 6, 2016.

[29] T. Wang, C. Rudin, F. Doshi-Velez, Y. Liu, E. Klampfl, and P. MacNeille. A bayesian framework for learning rule sets for interpretable classification. *Journal of Machine Learning Research*, 18(70):1–37, 2017.