IMPACT: An Inference-Driven Modeling Framework for Cost-Effective Incentive Allocation in Service Operations

Yiheng An
Wake Forest University
Winston-Salem, NC, USA
any@wfu.edu

Jia Li*
Wake Forest University
Winston-Salem, NC, USA
lijia@wfu.edu

Jeffrey D. Camm
Wake Forest University
Winston-Salem, NC, USA
cammjd@wfu.edu

Liang Hu
Alibaba Group
Hangzhou, China
luheng.hl@taobao.com

Qinqin Zhuge
Alibaba Group
Hangzhou, China
zhugeqinqin.zgqq@taobao.com

Bingxin Jia
Alibaba Group
Hangzhou, China
bingxin.jbx@taobao.com

Abstract

Performance-based incentives are extensively used to align employee effort with organizational objectives. Their effective use is particularly important in high-frequency service environments, where frontline agents directly shape the customer experience in real time. However, existing incentive systems for customer-facing employees rely largely on static, pre-defined rules and typically overlook the service context that predicts *when and where* incentives are most valuable. In this paper, we propose **I**nference-driven **M**odeling for **P**rescriptive **A**llocation of **C**onstrained **T**reatments (IMPACT), a novel framework that integrates causal machine learning, supervised learning, and constrained optimization to guide prescriptive allocation of targeted incentives in budget-constrained service operations. The application is developed in partnership with a customer service platform seeking to improve customer satisfaction by awarding cash bonuses to agents handling service cases, while controlling overall compensation costs. The framework flexibly adapts to high-dimensional contextual data and diverse intervention goals, making it broadly applicable to incentive allocation problems under resource constraints. Experimental results on production data demonstrate that IMPACT consistently outperforms context-free benchmarks, achieving higher customer satisfaction rates without increasing total incentive spending.

CCS Concepts

• **Information systems** → *Decision support systems*; • **Computing methodologies** → *Machine learning approaches.*

Keywords

personalized incentives, decision support systems, causal machine learning, prescriptive analytics, customer service

1 Introduction

Performance-based incentives, such as bonuses and commissions, remain an essential lever to guide employee effort toward organizational goals. In high-frequency customer service environments, the *when and where* of offering these incentives matter as much as the incentive itself: an incentive treatment may influence performance outcomes in one service session, but have minimal effect in another. However, conventional incentive plans are typically static and retrospective, uniformly rewarding aggregated past performance (e.g., weekly, monthly, or even annually) [18], while overlooking the contextual features that often better predicts the marginal value of an incentive in real time. This presents two interrelated problems: (1) inefficient allocation of limited incentive budgets, and (2) missed opportunities to proactively influence employee behavior and improve performance outcomes in real time.

In this research, we propose **I**nference-driven **M**odeling for **P**rescriptive **A**llocation of **C**onstrained **T**reatments (IMPACT), a model-based framework that proactively allocates cost-effective incentive interventions in time-sensitive service environments. The framework integrates causal machine learning (ML), supervised learning, and constrained optimization to guide personalized incentive decisions that adapt to specific service contexts. The acronym IMPACT highlights two design choices. The approach is *inference-driven* because every decision traces back to counterfactual uplift estimates rather than descriptive correlations or pre-defined business rules, and it is *prescriptive* because those estimates feed a downstream optimization model that recommends budget-feasible, objective-maximizing allocations.

The practical application is developed in collaboration with Alibaba Group, where we design a new incentive system for customer service agents. The objective is to improve customer satisfaction by selectively offering bonus opportunities to agents at the beginning of service sessions. This presents three key challenges: (1) Decision timing: bonus decisions must be made proactively *before* the conversation begins, requiring reliance solely on pre-interaction information; (2) Causal inference under partial information: the counterfactual performance outcome—what would have achieved under the alternative bonus condition—is unobservable, necessitating robust causal estimation; and (3) Budget constraints: total bonus spending is limited by a strict daily budget, requiring cost-effective allocation across sessions.

These challenges highlight the need for a solution that can effectively manage the complexity of inference, prediction, and allocation in a unified decision-support system. To address this, we design IMPACT grounded in an estimate–predict–optimize architecture: (i) an ensemble of causal learners estimates heterogeneous effects of bonuses using randomized experimental data; (ii) a supervised ML model predicts those effects for unseen cases using only pre-interaction features; and (iii) a robust optimization model allocates

*Corresponding author.

bonuses to maximize expected satisfaction improvements under a fixed daily budget.

We evaluate IMPACT using production data from more than 380K customer service sessions. Through a series of off-policy experiments, we show that our approach consistently outperforms both random and homogeneous incentive strategies. Across a broad budget range, the personalized incentive policies generated by IMPACT deliver the highest customer satisfaction rates.

The key contributions of our work come from the following:

- We formulate the incentive targeting problem as a budget-constrained uplift-maximization task and characterize the key requirements that guide the development of deployable, personalized incentive policies.
- We develop a model-based framework (i.e., IMPACT) that integrates ML-based causal inference, real-time uplift prediction, and combinatorial optimization to deliver cost-effective incentives in service operations.
- We incorporate a protection function in the optimization module that guards against worst-case deviations in uplift predictions, which significantly enhances the robustness of policies under imperfect uplift signals.
- We demonstrate the practical effectiveness of the proposed IMPACT framework using production data and compare the performance with other model-free alternatives.

2 Related Work

2.1 ML for Personalized Interventions

Recent advances in ML have provided powerful tools for designing personalized interventions through modeling individual-level responsiveness. Uplift modeling, a key development in this area, seeks to quantify the incremental impact of actions (e.g., promotions, incentives) by estimating individual treatment effects (ITE) [20]. This ability provides opportunities for more effective intervention strategies, particularly in settings with rich covariates and substantial treatment heterogeneity, as actions can be directed toward individuals or units with the highest expected benefit. Techniques such as causal forests [1, 25, 26], meta-learners [16, 30], and doubly robust estimators [10, 22] have been explored to guide personalized interventions based on observational and/or experimental data. In parallel, software libraries such as CausalML [7], EconML [15], and DoubleML [3] have further lowered the barrier to deploying and operationalizing these methods in production environments.

Our study advances this stream of research in two ways. First, we address a key challenge in real-time decision support: some of the most informative signals for estimating treatment heterogeneity are only observed after the intervention, and thus cannot be used at decision time. To overcome this limitation, we introduce a supervised learning module that predicts post-intervention-informed treatment effects using only features observable before the decision is made. This preserves the quality of counterfactual estimates while ensuring full deployability in production environments. Second, we show how context-specific treatment effects can be translated into large-scale, operationalized bonus decisions in high-frequency service environments. Using detailed session-level information in real-time ITE predictions, our approach enables proactive incentive interventions.

2.2 Targeting under Constraints and Uncertainty

In many operational settings, targeting decisions must satisfy hard constraints, such as budgets, treatment quotas, or fairness requirements, while also managing uncertainty in the broader decision-making context. To address this complexity, recent research has combined predictive models with constrained optimization techniques [2, 5, 24] to derive real-time decisions. These methods have been applied in personalized promotions [11], targeted discounts [6], budget-constrained recommendation systems [24], and other domains. Our study connects to this stream by framing bonus planning as a constrained incentive targeting problem, where decisions are informed by learned uplift estimates.

While many prior applications focus on customer-facing interventions [9, 17, 27, 30], we focus on real-time incentive planning for frontline service agents, a novel setting characterized by employee-facing interventions and underexplored treatment heterogeneity. We propose IMPACT, a multi-step approach that leverages a *estimate-predict-optimize* pipeline to deliver targeted incentives. We demonstrate its effectiveness using real-world production data, showing how data-driven incentive targeting can be extended beyond conventional customer-facing applications. As part of IMPACT, we incorporate a robust optimization module that accounts for uncertainty in counterfactual predictions. Building on the Γ-robust framework [4], we employ a protection function to guard against worst-case deviations in the predicted uplift. This enhancement supports the delivery of effective incentive policies under noisy predictions.

3 Operational Setting and Constraints

Our research is conducted on a large-scale post-purchase customer service platform of Alibaba, where a *service session* refers to a live, text-based online session between a human agent and a customer, and a *case* refers to a customer-submitted issue being addressed. Each session is dedicated to resolving one case. In this setting, we focus on designing effective bonus strategies that reward agents for handling the most challenging cases, with the goal of improving customer satisfaction. To support this initiative, we leverage an in-house algorithm that classifies each incoming case as either "easy" or "difficult" at the onset of the corresponding session. Based on this classification, we decided to restrict bonus eligibility to cases identified as "difficult".

An service session begins when a customer is assigned to a human agent, initiating a service interaction in which the agent investigates and addresses the issue using a dedicated digital workbench interface. At this initiation stage, the case is classified as either "easy" or "difficult", as previously discussed. For cases identified as "difficult", the system must decide whether to offer the agent a cash bonus *opportunity*. If a bonus opportunity is offered, the agent sees a red envelope icon on the navigation bar along with a prompt message delivered through the live chat box of their workbench (Figure 1). If no bonus opportunity is offered, the interface remains unchanged. Each case can trigger at most one bonus opportunity, which is visible only to the agent and not to the customer.

Given the budget limitation, the monetary value of a bonus must be determined after the interaction, based on the customer's post-service satisfaction rating score, which is recorded on a 1-to-5 scale. A pre-defined business rule translates this score into a discrete payout amount. Thus, while the bonus opportunity is presented to the agent at the start of the session, the actual payout is finalized only after the session concludes. Importantly, to ensure the bonus opportunity has a sustained motivational impact throughout the service session, the intervention decision must be made before the service interaction begins. This requires that allocation decisions rely exclusively on pre-interaction information.

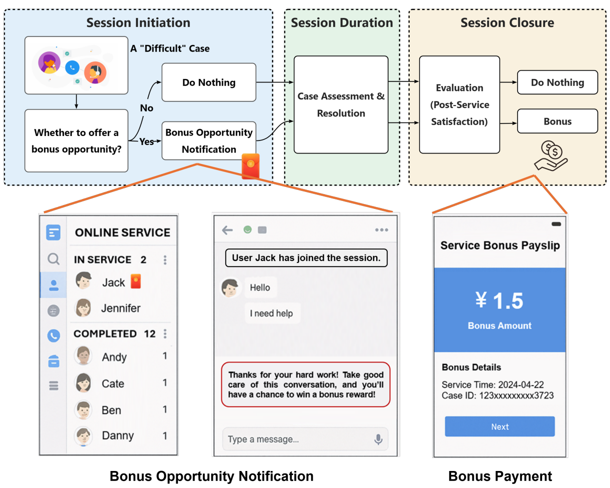


Figure 1: Bonus Allocation Workflow. When a bonus opportunity is triggered, the agent is notified by a red envelope icon in the navigation bar and a prompt message in the live chat window. The payment amount is determined after the interaction, based on the customer's satisfaction rating (1–5).

4 Proposed Framework

Motivated by the challenges identified in the bonus allocation workflow, we develop IMPACT, a modular framework that integrates causal ML, supervised learning, and robust optimization to support proactive incentive targeting (Figure 2). The following subsections describe the key components of the framework in detail.

4.1 Randomized Experiment

To provide unbiased training data, we first conducted a randomized controlled trial (RCT) on the customer service platform. From the full agent pool, we drew a stratified sample of 624 agents. Block randomization was applied based on agent ability scores to balance ability differences between groups. Within every quintile of agent ability, half of the agents were randomly assigned to the treatment group and half to the control group. Treatment-group agents (312 agents) were eligible to receive bonus prompts on "difficult" cases, whereas control-group agents (312 agents) never received bonus prompts, regardless of case difficulty. Customers were randomly assigned to agents via the platform's standard routing logic.

Let $i \in I$ denote a customer-agent service session (and associated case) created during the RCT. For every session, we recorded a binary treatment indicator T_i (1 = bonus prompt, 0 = no bonus prompt) and a binary outcome Y_i that equals 1 when the customer reports satisfaction (i.e., post-service rating $\geq$4) and 0 otherwise[1]. We also build a p-dimensional *context vector* $X_i = (X_i^{\text{pre}}, X_i^{\text{post}}) \in \mathbb{R}^p$, where X_i^{pre} collects pre-interaction features that are known *before* the intervention decision is made, including customer and agent demographics, case category, agent ability, agent workload, and the customer's projected value; X_i^{post} collects post-interaction features that realize *during* and/or *after* the session, such as response latency, number of conversation rounds, and session duration.

The customer-agent assignment approach guarantees that, within the "difficult" case segment, the intervention decision is statistically independent of both observed pre-interaction context features X_i^{pre} and the potential outcomes. Formally,

$$T_i \perp\!\!\!\perp \left(X_i^{\text{pre}},\, Y_i(0),\, Y_i(1)\right) \bigm| D_i = 1 \tag{1}$$

where $D_i = 1$ denotes a "difficult" case. This conditional independence [8] ensures that any difference in satisfaction outcomes between treatment and control groups for these "difficult" cases can be attributed solely to the bonus intervention, providing a valid basis for estimating heterogeneous treatment effects in the next stage of the pipeline.

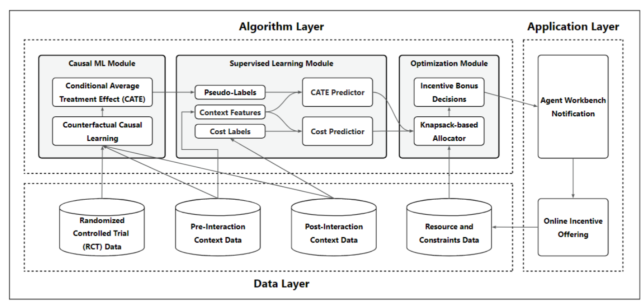


Figure 2: Overview of the IMPACT Framework. The data layer specifies the inputs required for modeling, including experimental, contextual, and operational data. The algorithm layer implements an estimate–predict–optimize pipeline to generate personalized incentive recommendations. The application layer translates these recommendations into agent-facing interventions.

4.2 Causal ML Module

Building on the data collected from the RCT, the Causal ML module is designed to estimate the session-level treatment effects of bonus offerings, conditional on rich service contextual features. With the treatment indicator $T_i \in \{0, 1\}$ and binary outcome $Y_i \in \{0, 1\}$,

[1]Although we focus on customer satisfaction, Y_i can be flexibly re-defined as other binary or continuous key performance or operational metric, such as service quality, first-call resolution, future redial rate, allowing the framework to accommodate other intervention goals.

the session-level Conditional Average Treatment Effect (CATE) is defined as

$$\tau_i(x_i) = \mathbb{E}\big[Y_i \mid X_i = x_i,\, T_i = 1\big] - \mathbb{E}\big[Y_i \mid X_i = x_i,\, T_i = 0\big] \quad (2)$$

In our setting, the high-dimensional feature space and the lack of strong prior knowledge about the structure of treatment effect heterogeneity make it difficult to rely on any single learner to estimate Equation (2). Therefore, we follow the advocated practices [12, 21] and employ an *ensemble* of causal learners (i.e., Causal Forest, X-Learner, T-Learner, and Doubly-Robust Learner) using R-Stacking [12, 19]. As output, we obtain the uplift estimate $\widehat{\tau}_i$ and the corresponding variance estimates, which are used in subsequent policy evaluation.

To increase statistical efficiency, we use both X_i^{pre} and X_i^{post} in training causal learners. Although incorporating X_i^{post} significantly enhances estimation precision and statistical power during offline training, these features won't be available at the time of decision for upcoming cases. This operational constraint motivates the need for the subsequent Supervised Learning module.

4.3 Supervised Learning Module

The Supervised Learning module predicts the uplifts estimated from the Causal ML module, using only pre-interaction features available at decision time. Specifically, we fit two supervised models using the RCT data:

(i) CATE predictor. We treat the CATE estimates τ_i produced by the Causal ML module as *labels*, and learn a mapping $\tau_i = f(X_i^{\text{pre}})$ where X_i^{pre} denotes the vector of pre-interaction features introduced in Section 4.1. This design addresses a key challenge in real-time decision support: while post-interaction data are essential for efficiently estimating treatment heterogeneity, they are not available at the time of decision. By learning to predict these post-interaction-informed labels using only pre-interaction signals, the predictor enables deployability in real-time operations.

(ii) Cost predictor. Similarly, as illustrated in Figure 1, the monetary incentive cost c_i associated with case i under treatment $T = 1$ is revealed only after the service session ends and is thus also unavailable at decision time. To address this, we learn a second mapping $c_i = g(X_i^{\text{pre}})$ based on the same pre-interaction context to generate real-time cost predictions at the case level.

The framework places no restriction on the functional form of $f(\cdot)$ and $g(\cdot)$. At run time, the module consumes the pre-interaction context of each newly arriving "difficult" case, and outputs the pair $\big(\widehat{\tau}_i,\ \widehat{c}_i\big) = \Big(f(X_i^{\text{pre}}),\ g(X_i^{\text{pre}})\Big)$. These predictions serve as key input parameters for the downstream optimization module.

4.4 Optimization Module

The optimization module translates the uplift predictions from the CATE predictor into bonus allocation decisions. The objective is to maximize the expected satisfaction uplift while respecting the daily budget. For each newly arriving case i, the binary variable $w_i \in \{0, 1\}$ denote the bonus intervention decision, where $w_i = 1$ indicates a bonus prompt and $w_i = 0$ indicates no bonus prompt. Note that we observe the predicted satisfaction uplift score $\widehat{\tau}_i(x_i)$ and the predicted incentive cost $\widehat{c}_i$, as well as the remaining budget b_0 at the decision time.

However, since each uplift score $\widehat{\tau}_i(x_i)$ is a predicted value with clear variability, directly optimizing based on these estimates may lead to over-committing resources to cases with high nominal uplift but large prediction errors[2]. To mitigate the risks associated with uncertainty in the predicted satisfaction uplift, we propose a robust incentive allocation model following the Γ-approach [4], which enables flexible control over the degree of protection against prediction uncertainty by adjusting probabilistic bounds of constraint violations.

We assume that the expected satisfaction uplift score $\widehat{\tau}_i$ is a symmetric and bounded random variable. Its realized value is allowed to vary randomly within a deviation $\delta_i > 0$ around the average estimate $\bar{\tau}_i$, forming a box-type uncertainty set defined as $\widehat{\tau}_i \in [\bar{\tau}_i - \delta_i, \bar{\tau}_i + \delta_i]$. This box-type uncertainty set is widely used in robust optimization due to its tractability and ease of implementation when the true distribution of the uncertain parameter is unknown [14], which makes it well suited for our operational setting. Under this formulation, the incentive allocation model $z^R(\Gamma)$ is cast as a robust 0–1 knapsack problem that balances expected satisfaction gains against the risk induced by uncertain uplift predictions.

$$
\begin{aligned}
z^R(\Gamma) :=& \quad \max_{w} \sum_{i \in I} \bar{\tau}_i\, w_i - \Delta(w, \Gamma) && (3)\\
\text{s.t.}& \quad \sum_{i \in I} \widehat{c}_i\, w_i \le b_0 && (4)\\
& \quad w_i \in \{0, 1\} \quad \forall\, i \in I && (5)\\
\text{where}& \quad \Delta(w, \Gamma) := \max_{s} \sum_{i \in I} \delta_i\, w_i\, s_i && (6)\\
\text{s.t.}& \quad \sum_{i \in I} |s_i| \le \Gamma && (7)\\
& \quad |s_i| \le 1 \quad \forall\, i \in I && (8)
\end{aligned}
$$

As part of the objective, we introduce a protection function $\Delta(w, \Gamma)$, which is an inner optimization problem that captures the worst-case impact of uncertainty in the predicted satisfaction uplift. The function quantifies the maximum potential loss in uplift due to deviations, subject to a robustness level Γ. The auxiliary variables s_i, bounded by $0 \le s_i \le 1$, determine how much deviation is considered for each case in the protection function. The constraint (7) ensures that the total magnitude of deviations across all cases does not exceed Γ, allowing the model to balance performance and robustness. The parameter Γ governs the overall level of protection by limiting the aggregate deviation mass the model guards against. The intuition is that even if the predicted gains for a set of cases deviate adversely within a total deviation upper bound of Γ, the allocation still contributes a uplift of at least $\sum_i \bar{\tau}_i\, w_i - \Delta(w_i, \Gamma)$. As Γ increases, the model shifts from the nominal version that assumes no uncertainty ($\Gamma = 0$) to increasingly conservative stances that hedge against greater prediction risk.

The formulation of $z^R(\Gamma)$ is NP-Hard. The streaming nature of service platform traffic and latency requirements make exact optimization impractical for real-time operations. Therefore, we

[2] In contrast, incentive costs are drawn from a small set of discrete values. They are highly predictable from pre-interaction features and show very little variation across cases. For simplicity, we treat $\widehat{c}_i$ as known in the optimization model and apply robustness adjustments only to the more uncertain uplift predictions.

implement a batched incentive allocation algorithm (1) that follows a greedy heuristic commonly used in knapsack-based uplift optimization. For each incoming batch of M concurrent cases, the system (i) computes the total uncertainty level $\Delta_{tot} = \sum_{i \in M} \delta_i$, (ii) derives a scaling factor $\theta = \min\{1, \frac{\Gamma}{\Delta_{tot}}\}$ to normalize the robustness adjustment (iii) calculates the risk-adjusted uplift score $\tilde{\tau}_i = \bar{\tau} - \theta\delta_i$ for each active case, (iv) ranks each case by a priority score $\rho_i = \tilde{\tau}_i / \max\{\widehat{c}_i, \bar{c}\}$ in descending order, where $\bar{c}$ is a cost floor to prevent division by zero or instability from near-zero costs, and (v) iteratively assigns bonuses following this order until the residual budget is exhausted.

Algorithm 1: Batched Greedy Algorithm for Robust Incentive Targeting

Input: CATE predictor $f(x)$, cost predictor $g(x)$, batch of cases M, daily budget b, robustness level Γ
Output: Bonus assignment vector $w = \{w_1^*, w_2^*, \ldots, w_{|M|}^*\}$

$b_0 \leftarrow b$;
foreach *session* $i \in M$ **do**
  $\bar{\tau}_i, \delta_i \leftarrow f(x_i)$;
  $\widehat{c}_i \leftarrow g(x_i)$;
$\Delta_{\text{tot}} \leftarrow \sum_{i \in M} \delta_i$;
$\theta \leftarrow \min\left\{1, \frac{\Gamma}{\Delta_{\text{tot}}}\right\}$;
foreach *session* $i \in M$ **do**
  $\tilde{\tau}_i \leftarrow \bar{\tau}_i - \theta\delta_i$;
  $\rho_i \leftarrow \tilde{\tau}_i / \max\{\widehat{c}_i, \bar{c}\}$;
Sort sessions in M in descending order of ρ_i;
foreach *session* i *in sorted* M **do**
  if $b_0 - \widehat{c}_i \geq 0$ **then**
    $w_i^* \leftarrow 1$;
    $b_0 \leftarrow b_0 - \widehat{c}_i$;
  else
    $w_i^* \leftarrow 0$;

5 Evaluation

5.1 Experiment Setup

5.1.1 Production Data. We evaluate IMPACT and alternative policies on the six-week RCT data introduced in Section 4.1. The data contains 382,604 customer cases handled by 624 human agents on the service platform; 47,915 of these cases (12.5 %) are classified as "difficult" and hence eligible for bonus opportunities. We randomly split the entire dataset into three equal-sized subsets for training, validation, and testing. Stratified sampling is used to preserve the original treatment-to-control ratio in each split. The training set is used to fit model parameters, the validation set is used for hyperparameter tuning, and the test set is held out entirely for final policy evaluation.

5.1.2 Policies for Evaluation. We evaluate the effectiveness of the proposed framework by assessing four targeting policies. These include two context-free baselines and two IMPACT-based policy instantiations, which differ only in the CATE predictor used.

- **50/50 Random Assignment (50/50)**: The benchmark that offers a bonus opportunity to exactly half of the "difficult" cases chosen at random until the daily budget is exhausted. This policy reflects a simple heuristic that does not incorporate any contextual information.
- **Homogeneous Treatment (HT)**: The benchmark that allocates a bonus opportunity to every "difficult" case uniformly until the daily budget is exhausted, disregarding contextual heterogeneity. HT reflects a rule-based approach historically adopted by the platform.
- **IMPACT-RF**: A deployable IMPACT-based policy generated by the proposed framework. It solely leverages pre-interaction features to generate uplift predictions using a Random Forest regressor, and feeds the resulting prediction scores into the downstream knapsack optimizer.
- **IMPACT-LASSO**: Identical to the IMPACT-RF policy, except that the CATE predictor is replaced by an L_1-regularized linear model. The resulting sparse coefficient vector yields a transparent scoring rule and enables faster inference, while maintaining generalizability across cases.

5.1.3 Evaluation Method. Due to fairness concerns, the live testing of alternative employee-facing incentive policies is often risky. Following established practices [23], we instead adopt an *off-policy* evaluation strategy that leverages the randomization in the RCT data to estimate policy performance without additional experimentation. We define a policy $\pi : \mathcal{X} \to \{0, 1\}$ as a mapping from session-level context features x_i to a binary treatment recommendation, where $\pi(x_i) = 1$ indicates that a bonus opportunity is offered for session i, while $\pi(x_i) = 0$ indicates no bonus opportunity is offered. The performance of any policy $\pi(\cdot)$ is estimated on the hold-out sample with the inverse propensity score-weighted (IPS) estimator as described in [13]:

$$\hat{R}_{\text{IPS}}(\pi) = \frac{1}{N} \sum_{i \in I} \left[\frac{1 - w_i}{1 - \hat{e}_i} \left(1 - \pi(x_i)\right) y_{i,T=0} + \frac{w_i}{\hat{e}_i} \pi(x_i)\, y_{i,T=1} \right] \quad (9)$$

where w_i is the realized treatment assignment in the experiment, $\hat{e}_i = \Pr(T_i = 1 | x_i)$ is the experimental propensity score. $y_{i,T=0}$ and $y_{i,T=1}$ are the observed outcomes under control and treatment, respectively, and N is the number of hold-out sessions. The weighting scheme corrects for the mismatch between the experimental randomization and the policy recommendation, rendering $\hat{R}_{\text{IPS}}(\pi)$ an *unbiased* estimate of the mean satisfaction that would have been obtained had π been implemented.

5.1.4 Evaluation Metrics. We report two metrics to evaluate policy performance on the hold-out data.

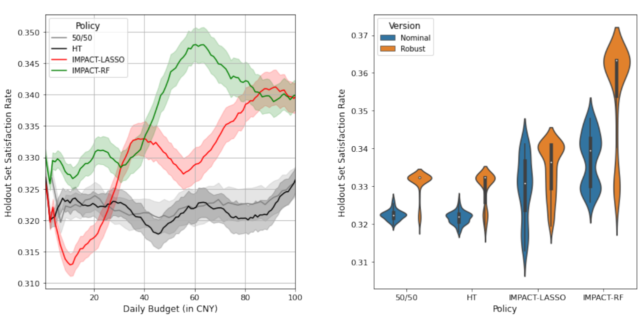
(i) Expected Satisfaction Rate (ESR). This is our primary domain-specific outcome, estimated using the IPS estimator described in Equation (9). It belongs to the broader class of Expected Outcome Metrics (EOMs) commonly used in uplift modeling and policy evaluation [28, 29]. In our setting, it provides a direct measure of how well a given policy improves service outcomes among "difficult" service cases.

(ii) Area Under the Cost Curve (AUCC). This metric aggregates the incremental satisfaction gains achieved relative to incremental incentive spending. It captures each policy's ability to prioritize high-impact cases under varying budget constraints in the binary treatment setting [9, 29].

5.1.5 Implementation Details. All policies described in Section 5.1.2 are evaluated over a grid of daily budget levels ranging from 0 to ¥100, in increments of ¥5. This normalized range is set to ensure that our evaluation fully covers the realistic operating range. All causal ML algorithms described in Section 4.2 are implemented using the EconML package [15]. Hyper-parameters for both the causal learners and supervised prediction models within the IMPACT pipeline are selected via grid search, using performance on the validation set measured by mean-squared error (MSE). We use a logistic regression model with an L-1 penalty to predict incentive costs from a discrete set of bonus levels. Evaluation metrics are computed using 100 bootstrap samples, and 95% confidence intervals (CIs) are reported to quantify uncertainty.

5.2 Experimental Results

To protect confidentiality, all reported values in this section have been normalized to mask actual amounts while preserving relative magnitudes. Figure 3a illustrates the expected satisfaction rate of each policy across the range of daily budgets. The IMPACT-based policies outperform the two context-free baselines across most of the budget grid. Notably, IMPACT-RF maintains the highest satisfaction rates over a wide budget range, indicating strong out-of-sample performance. Figure 3b examines the impact of robustness adjustments in the optimization module by comparing the nominal model ($\Gamma = 0$) with a robust variant ($\Gamma = 2$) over the same budget range (0-¥100). Policies derived from the robust model generally achieve higher satisfaction rates, showing the practical value of accounting for uplift uncertainty in the incentive allocation logic.



(a) ESR by Daily Budget **(b) ESR by Model Robustness**

Figure 3: Off-Policy Evaluation Results Across Budget Levels. (a) We depict the ESR measured in the hold out data across daily budgets. Each curve represents a different targeting policy. Satisfaction rates are based on the IPS estimator (Equation 9). Shaded regions denote 95% bootstrap CIs. (b) We compare hold-out ESR under nominal and robust model versions for each targeting policy. Violin plots summarize 100 bootstrap samples over the same budget grid.

To further assess overall effectiveness across the entire budget range, Table 1 reports the AUCC for each policy. This aggregate metric reflects how effectively each policy transforms incremental spending into satisfaction gains. IMPACT-RF achieves the highest AUCC, closely followed by IMPACT-LASSO, highlighting the value of personalization based on machine-learned heterogeneity. In contrast, the context-free baselines perform substantially worse, indicating that indiscriminate allocation leaves most of the budget's potential untapped.

Table 1: Experimental Results (AUCC). We report the mean and margin of error for each policy, based on 100 bootstrap samples. Improvement is calculated as the percentage change relative to the platform's historical approach (HT).

Policy	AUCC	Improvement
HT	0.2114 ± 0.0009	/
50/50	0.1770 ± 0.0254	-16.27%
IMPACT-LASSO	0.5228 ± 0.0276	147.30%
IMPACT-RF	0.5455 ± 0.0287	158.04%

We also conduct point-wise comparisons across policies under a consistent daily budget, fixed at a representative intermediate value of ¥60. Table 2 reports IPS-based satisfaction estimates for each policy in this setting. The two context-free baselines (HT and 50/50) yield similar satisfaction rates of approximately 32.19%. IMPACT-LASSO improves performance to 32.86%. IMPACT-RF achieves the highest satisfaction rate of 34.79%, corresponding to an 8.07% improvement over the platform's historical approach (HT).

Table 2: Experimental Results (ESR). We report the mean and margin of error for each policy, based on 100 bootstrap samples and a fixed budget of ¥60. Improvement is calculated as the percentage change relative to the platform's historical approach (HT).

Policy	ESR	Budget	Improvement
HT	0.3219 ± 0.0024	¥60	/
50/50	0.3219 ± 0.0029	¥60	0.00%
IMPACT-LASSO	0.3286 ± 0.0027	¥60	2.09%
IMPACT-RF	0.3479 ± 0.0026	¥60	8.07%

6 Conclusions

This study introduces IMPACT, an inference-driven approach for budget-constrained incentive allocation in customer service operations. We demonstrate how machine-learned treatment heterogeneity can be operationalized in high-frequency service settings, and how robustness adjustments can support real-time resource allocation while preserving necessary control when uplift predictions are noisy. The established pipeline offers a validated, deployable tool for customer service practices, enabling proactive, cost-effective incentive targeting that improves employee performance, customer benefits, and operational efficiency.

Acknowledgments

This work is in part supported by the Center for Analytics Impact (School of Business, Wake Forest University). The authors thank Prof. Fan Yang (Department of Computer Science, Wake Forest University), and colleagues at Wake Forest University and Alibaba Group for their valuable feedback and advice during the preparation of the manuscript.

References

[1] Meng Ai, Biao Li, Heyang Gong, Qingwei Yu, Shengjie Xue, Yuan Zhang, Yunzhou Zhang, and Peng Jiang. 2022. Lbcf: A large-scale budget-constrained causal forest algorithm. In *Proceedings of the ACM Web Conference 2022*. 2310–2319.

[2] Javier Albert and Dmitri Goldenberg. 2022. E-commerce promotions personalization via online multiple-choice knapsack with uplift modeling. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2863–2872.

[3] Philipp Bach, Victor Chernozhukov, Malte S. Kurz, and Martin Spindler. 2022. DoubleML – An Object-Oriented Implementation of Double Machine Learning in Python. *Journal of Machine Learning Research* 23, 53 (2022), 1–6. http://jmlr.org/papers/v23/21-0862.html

[4] Dimitris Bertsimas and Melvyn Sim. 2004. The price of robustness. *Operations research* 52, 1 (2004), 35–53.

[5] Semih Cayci, Atilla Eryilmaz, and Rayadurgam Srikant. 2020. Budget-constrained bandits over general cost and reward distributions. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 4388–4398.

[6] Fanglin Chen, Xiao Liu, Bo Tang, Feiyu Xiong, Serim Hwang, and Guomian Zhuang. 2022. BCRLSP: An Offline Reinforcement Learning Framework for Sequential Targeted Promotion. *arXiv preprint arXiv:2207.07790* (2022).

[7] Huigang Chen, Totte Harinen, Jeong-Yoon Lee, Mike Yung, and Zhenyu Zhao. 2020. Causalml: Python package for causal machine learning. *arXiv preprint arXiv:2002.11631* (2020).

[8] A Philip Dawid. 1979. Conditional independence in statistical theory. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 41, 1 (1979), 1–15.

[9] Shuyang Du, James Lee, and Farzin Ghaffarizadeh. 2019. Improve user retention with causal learning. In *The 2019 ACM SIGKDD Workshop on Causal Discovery*. PMLR, 34–49.

[10] Paul B Ellickson, Wreetabrata Kar, and James C Reeder III. 2023. Estimating marketing component effects: Double machine learning from targeted digital promotions. *Marketing Science* 42, 4 (2023), 704–728.

[11] Dmitri Goldenberg, Javier Albert, Lucas Bernardi, and Pablo Estevez. 2020. Free lunch! retrospective uplift modeling for dynamic promotions recommendation within roi constraints. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 486–491.

[12] Kevin Wu Han and Han Wu. 2022. Ensemble method for estimating individualized treatment effects. *arXiv preprint arXiv:2202.12445* (2022).

[13] Günter J Hitsch, Sanjog Misra, and Walter W Zhang. 2024. Heterogeneous treatment effects and optimal targeting policy evaluation. *Quantitative Marketing and Economics* 22, 2 (2024), 115–168.

[14] Hoon Jang. 2019. A decision support framework for robust R&D budget allocation using machine learning and optimization. *Decision Support Systems* 121 (2019), 1–12.

[15] Maggie Hei Greg Lewis Paul Oka Miruna Oprescu Vasilis Syrgkanis Keith Battocchi, Eleanor Dillon. 2019. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/py-why/EconML. Version 0.x.

[16] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences* 116, 10 (2019), 4156–4165.

[17] Aurélie Lemmens and Sunil Gupta. 2020. Managing churn to maximize profits. *Marketing Science* 39, 5 (2020), 956–973.

[18] Stefan A Lipman. 2024. One size fits all? Designing financial incentives tailored to individual economic preferences. *Behavioural Public Policy* 8, 2 (2024), 264–278.

[19] Xinkun Nie and Stefan Wager. 2021. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* 108, 2 (2021), 299–319.

[20] Diego Olaya, Kristof Coussement, and Wouter Verbeke. 2020. A survey and benchmarking study of multitreatment uplift modeling. *Data Mining and Knowledge Discovery* 34 (2020), 273–308.

[21] Alejandro Schuler, Michael Baiocchi, Robert Tibshirani, and Nigam Shah. 2018. A comparison of methods for model selection when estimating individual treatment effects. *arXiv preprint arXiv:1804.05146* (2018).

[22] Wei Shi, Chen Fu, Qi Xu, Sanjian Chen, Jizhe Zhang, Qinqin Zhu, Zhigang Hua, and Shuang Yang. 2024. Ads Supply Personalization via Doubly Robust Learning. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 4874–4881.

[23] Duncan Simester, Artem Timoshenko, and Spyros I Zoumpoulis. 2020. Efficiently evaluating targeting policies: Improving on champion vs. challenger experiments. *Management Science* 66, 8 (2020), 3412–3424.

[24] Zexu Sun, Hao Yang, Dugang Liu, Yunpeng Weng, Xing Tang, and Xiuqiang He. 2024. End-to-end cost-effective incentive recommendation under budget constraint with uplift modeling. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 560–569.

[25] Moritz von Zahn, Kevin Bauer, Cristina Mihale-Wilson, Johanna Jagow, Maximilian Speicher, and Oliver Hinz. 2024. Smart green nudging: Reducing product returns through digital footprints and causal machine learning. *Marketing Science* (2024).

[26] Shu Wan, Chen Zheng, Zhonggen Sun, Mengfan Xu, Xiaoqing Yang, Hongtu Zhu, and Jiecheng Guo. 2022. Gcf: Generalized causal forest for heterogeneous treatment effect estimation in online marketplace. *arXiv preprint arXiv:2203.10975* (2022).

[27] Jason Shuo Zhang, Benjamin Howson, Panayiota Savva, and Eleanor Loh. 2024. DISCO: An End-to-End Bandit Framework for Personalised Discount Allocation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 33–49.

[28] Yan Zhao, Xiao Fang, and David Simchi-Levi. 2017. Uplift modeling with multiple treatments and general response types. In *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 588–596.

[29] Hao Zhou, Shaoming Li, Guibin Jiang, Jiaqi Zheng, and Dong Wang. 2023. Direct heterogeneous causal learning for resource allocation problems in marketing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 5446–5454.

[30] Will Y Zou, Shuyang Du, James Lee, and Jan Pedersen. 2020. Heterogeneous causal learning for effectiveness optimization in user marketing. *arXiv preprint arXiv:2004.09702* (2020).