Applying Doubly Robust Policy Learning to the Bank Marketing Dataset

R26144097 李祐君

Dec 25, 2025

Content

1. Introduction

1.1 Background and Motivation

In the domain of bank direct marketing, the efficiency of resource allocation is paramount. Traditional marketing approaches often rely on "Average Treatment Effects" (ATE), assuming that a single strategy (e.g., calling everyone via cellular phone) works best for the average customer. However, this "one-size-fits-all" approach ignores **treatment effect heterogeneity**—the fact that different customer segments respond differently to various interventions based on their characteristics (e.g., age, economic indicators).

1.2 Objectives

The primary objective of this project is to move beyond outcome prediction and toward **causally optimal decision-making through policy learning**. Specifically, we aim to construct an interpretable, data-driven policy

$$\pi(X): X \rightarrow W,$$

that maps observable client characteristics X to an optimal marketing action W, with the goal of **maximizing overall welfare**, measured by the expected subscription rate to term deposits.

To achieve this goal, we adopt a two-stage causal inference framework. In the first stage, we employ Causal Forests to flexibly estimate heterogeneous treatment effects and the nuisance components required for causal identification. In the second stage, we use Policy Trees to transform these estimates into simple, interpretable decision rules that can be readily implemented in real-world marketing operations.

By combining the statistical efficiency of machine learning with the interpretability of decision trees, this project demonstrates how modern causal methods can be used not only to understand treatment heterogeneity, but also to derive actionable and transparent marketing policies.

2. Methodology

2.1 Introduction to Policy Learning and Welfare Maximization

In many fields, such as economics, marketing, and personalized medicine, researchers have increasingly recognized that estimating a single Average Treatment Effect (ATE) is often insufficient for guiding optimal decision-making. The central challenge today is therefore one of **policy learning**, which aims to determine an optimal, individualized decision rule—referred to as a policy π—that maps observable individual characteristics X to an action W to maximize the resulting outcome Y.

Formally, let $Y_i(w)$ denote the potential outcome for individual i if treatment w were assigned. The goal of policy learning is to identify the optimal policy π^* that maximizes social welfare, defined as the expected outcome that would be observed if the policy were applied to the entire population:

$$V(\pi) = \mathbb{E}[Y(\pi(X))]$$

In practice, the true population value $V(\pi)$ is unobservable because only a finite sample of size N is available. As a result, policy learning proceeds by Empirical Welfare Maximization (EWM), where the population expectation is replaced by its sample analogue. This means we seek the estimated policy $\hat{\pi}$ that maximizes the average welfare observed in our specific dataset.

2.2 The Doubly Robust (DR) Score

A primary challenge in policy learning with observational data—such as the Bank Marketing dataset—is the presence of **confounding bias**. Customers were not randomly assigned their marketing action (W); the bank's decision was likely influenced by the client's features (X). Consequently, a naive comparison of outcomes between clients contacted via different channels (e.g., cellular versus telephone) would generally yield biased estimates of causal effects.

To overcome this, policy learning relies on a robust method to evaluate the "reward" of each action. To address this issue, we employ the **Doubly Robust (DR) score**, which provides an unbiased estimate of the expected reward of assigning a specific action to a specific individual under standard unconfoundedness assumptions.

As formalized in the work of Athey & Wager (2017), the DR estimator achieves its "doubly robust" property by providing a two-pronged defense against model misspecification. It combines two separate prediction models, known as "nuisance components":

1. **The Outcome Regression Model ($\hat{\boldsymbol{\mu}}(\boldsymbol{X}, \boldsymbol{W})$):** Predicts the expected outcome Y given features X and action W.
2. **The Propensity Score Model ($\hat{\boldsymbol{e}}(\boldsymbol{X}, \boldsymbol{W})$):** Predicts the probability of receiving the observed action W given features X.

The estimated DR score for assigning action w to individual i is computed as:

$$\hat{\Gamma}_{i,w} = \hat{\mu}(X_i, w) + \frac{Y_i - \hat{\mu}(X_i, W_i)}{\hat{e}(X_i, W_i)} \mathbb{I}(W_i = w)$$

The term "Doubly Robust" implies that this estimator remains statistically consistent with the true welfare as long as at least one of these two nuisance models is correctly specified. This provides crucial security against the inevitable errors in real-world modeling.

2.3 Policy Tree Optimization

In our project, we first utilize the grf package (Generalized Random Forests) to accurately estimate the nuisance components ($\hat{\mu}$ and $\hat{e}$) and compute the DR scores. These DR scores serve as individual-level, unbiased estimates of potential rewards and are subsequently used as inputs to the “**policytree”** optimization algorithm.

The final step is to learn a simple and interpretable policy π, represented as a decision tree, that maximizes the sum of individual-level rewards across the sample. The algorithm solves the following weighted classification problem:

$$\hat{\pi} = arg \max_{\pi \in \Pi} \frac{1}{N} \sum_{i=1}^{N} \hat{\Gamma}_{i,\pi(Xi)}$$

By splitting the feature space based on these scores, the resulting tree identifies subgroups where a specific marketing action yields the highest causal uplift, directly implementing the goal of Empirical Welfare Maximization.

2.4 Methodological Workflow

Our empirical analysis implements a multi-action policy learning pipeline by integrating the **grf** and **policytree** packages, following the two-stage procedure proposed by Sverdrup et al. (2020).

First, we estimate the nuisance components using the multi_arm_causal_forest() function in the grf package, with observed features X, outcome Y, and treatment assignment W as inputs. This step produces flexible, non-parametric estimates of the outcome regression and propensity scores for each action.

Second, we compute individual-level doubly robust scores using the double_robust_scores() function. This yields a matrix $\Gamma \in \mathbb{R}^{N \times D}$, where each entry represents the estimated welfare for individual i under treatment assignment $W = d$.

Finally, the feature matrix X and the corresponding DR score matrix Γ are passed to the policy_tree() function. The algorithm performs an exact, recursive search to identify the optimal decision tree of a specified depth (1 and 2 depth in this project) that maximizes total empirical welfare across the sample.

3. Data Description and Preprocessing

3.1 Data Source and Introduction

This project analysis is based on the Bank Marketing dataset, originally published in the UCI Machine Learning Repository and redistributed via Kaggle. The dataset contains data from a direct marketing campaign (phone calls) of a Portuguese banking institution. The objective of the campaign was to encourage clients to subscribe to a term deposit.

The original purpose of this dataset is a predictive classification task, where the objective is to predict whether a client subscribes to a term deposit. In contrast, this project reframes the problem as a normative causal policy learning task, focusing on optimal decision-making rather than outcome prediction.

The dataset contains no missing values. The outcome variable Y indicates whether a client subscribed to a term deposit. Also, It divide input variables into four categories: bank client data(age, job, marital……), related with the last contact of the current campaign(contact, month, day_of_week), other attributes(campaign, pdays……) and social and economic context attributes (emp.var.rate: employment variation rate - quarterly indicator, cons.price.idx: consumer price inde)

3.2 Variable Definitions

To apply the Policy Tree, we categorize the variables into three distinct roles: Outcome (Y), Treatment (W), and Covariates (X).

- Outcome (Y):

  The outcome variable Y is a binary indicator of whether a client subscribed to the term deposit during the marketing campaign. We encoded Subscribed ("yes") as 1 and Did not subscribe ("no") as 0

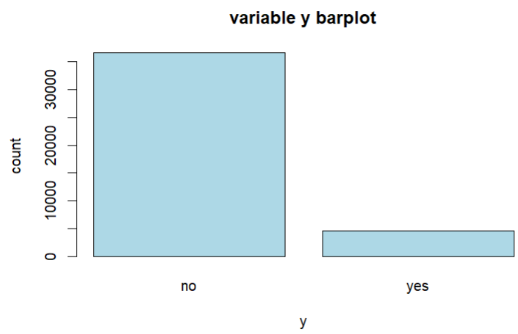

Figure 3.1 barplot of y

- Treatments (W):

  In this study, we designate a subset of variables related to the last contact of the current marketing campaign as treatment variables W. These variables represent managerial decisions that can be actively controlled by the bank at the time of contact, making them natural candidates for policy optimization. Specifically, we consider three distinct treatment dimensions and estimate separate policy learning models for each:

- Contact method: communication channel used to reach the client (cellular vs. telephone).

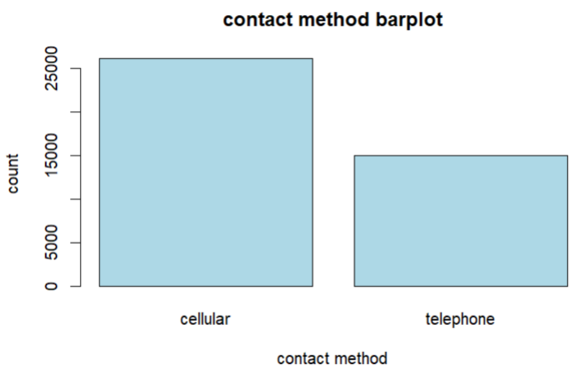

Figure 3.2 barplot of the contact method

- Contact timing (month/season): the month of the last contact, which we group into broader seasonal categories (Spring, Summer, and Winter) to reduce dimensionality.

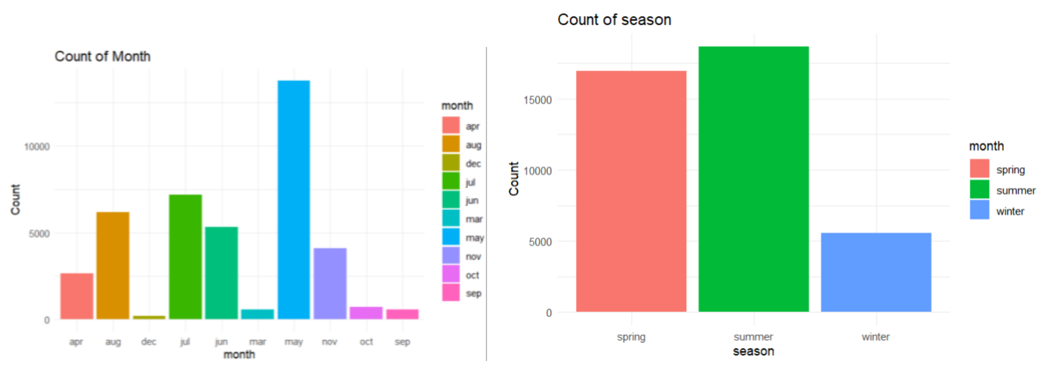

Figure3.3 barplot of month(original/after processing)

- Day of week: the weekday on which the contact occurred (Monday through Friday).

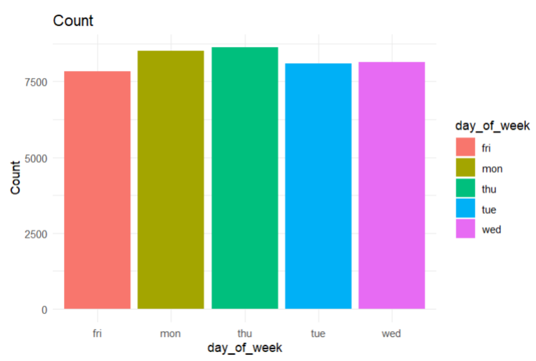

Figure 3.4 barplot of day_of_week

These variables are treated as policy-relevant actions because they can be deliberately chosen by the bank prior to initiating contact and are therefore suitable for causal policy evaluation.

- Covariates (X):

The covariate matrix X includes client demographic characteristics (e.g., age, job, marital status, education) as well as social and economic context attributes (e.g., euribor3m, consumer price index). These variables are assumed to be **pre-treatment characteristics** that influence both the marketing decision and the outcome, and are therefore used to control for confounding and to capture heterogeneity in treatment effects.

3.3 Data Preprocessing

3.3.1 Handling "Unknown" Values

Although the dataset contains no missing values, several categorical variables include entries labeled as "unknown." To reduce ambiguity in customer profiles and to improve the interpretability of treatment effect heterogeneity, we remove observations with "unknown" values in key demographic variables, including marital status, education, housing loan status, and personal loan status. After this filtering step, the final sample consists of 30,488 observations.

3.3.2 Exclusion of Post-Treatment Variables

A fundamental requirement for valid causal inference is that covariates included in X must not be affected by the treatment assignment. Including post-treatment variables would introduce **post-treatment bias** and invalidate causal interpretation.

Accordingly, we exclude the following variables from the feature matrix:

- Duration: call duration is not known prior to initiating contact and is mechanically correlated with the outcome. Including this variable would result in severe data leakage and yield an infeasible policy.

- Campaign, pdays, previous, poutcome: these variables describe the execution or outcomes of the current or past marketing campaigns and are therefore influenced by treatment decisions.

By excluding these variables, we ensure that the covariate set consists only of information available **prior to the marketing decision**, which is essential for policy learning.

3.3.3 Feature Engineering

Since the policytree algorithm requires a numeric feature matrix, we apply one-hot encoding to all remaining categorical covariates.(e.g., job, education). This process converted categorical levels into binary dummy variables, resulting in a final feature matrix (X) with dimensions of 30488×26.

4. Results

4.1 Overview of Policy Evaluation

This section evaluates the performance of the proposed Policy Tree models under three treatment dimensions: **contact method**, **contact month (season)**, and **day of week**. For each dimension, we compare the expected reward of the learned personalized policies at tree depths 1 and 2 against a baseline **Best Fixed Strategy**, which assigns the same treatment to all individuals regardless of their characteristics.

The expected reward is defined as the predicted probability of subscription under the corresponding policy. Table 4.1 summarizes the performance comparison across all models.

Table 4.1 Performance Comparison (Expected Reward)

Model (Treatment)	Tree Depth	Baseline: Best Fixed Strategy	Policy Tree Reward	Lift (Absolute)	Lift (%)
Model 1: Contact	1	0.1384 (Always Cellular)	0.1529	+0.0145	10.47%
Model 1: Contact	2		0.1753	+0.0369	26.64%
Model 2: Month	1	0.7630 (All Summer)	(All Summer)	-	-
Model 2: Month	2		0.7630	+0.0001	0.01%
Model 3: Day	1	0.1586 (Always Tuesday)	0.1657	+0.0072	4.51%
Model 3: Day	2		0.1730	+0.0145	9.13%

(Note: The 'Best Fixed Strategy' refers to assigning the treatment with the highest average effect to all individuals. 'Lift' denotes the improvement of the personalized policy over this fixed baseline.)

Table 4.1 reports the expected rewards achieved by the Policy Tree models and the corresponding lifts relative to the best fixed strategy. The results show that the benefits of personalization vary substantially across treatment dimensions: sizable gains are observed for **contact method**, moderate gains for **day of week**, and little to no improvement for **contact month**.

4.2 Model-Specific Results and Interpretation

4.2.1. Model 1: Contact Method (Significant Heterogeneity)

This model examines the decision between contacting clients via **cellular phone** or **telephone**, a key operational choice in marketing campaigns. Under the Best Fixed Strategy, assigning all clients to cellular contact yields an expected subscription rate of 0.1384.

The Policy Tree models substantially outperform this benchmark by exploiting heterogeneity in client responsiveness. The depth-1 policy increases the expected reward to 0.1529, corresponding to a 10.47% lift, while the depth-2 policy further improves performance to 0.1753, representing a 26.64% increase over the fixed strategy. These results indicate pronounced treatment effect heterogeneity in communication channels.

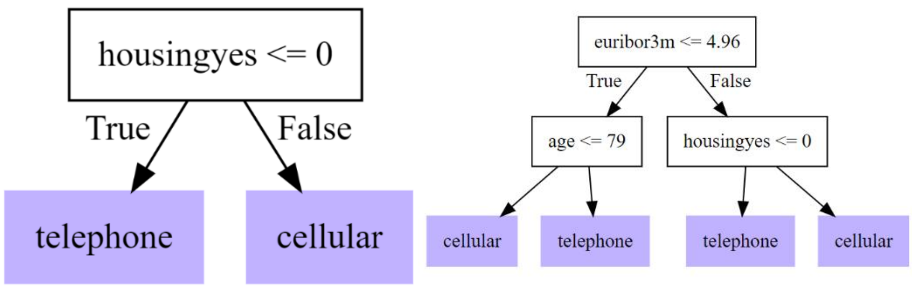

Figure4.1 Policy tree with contact method(depth1/2)

Figure 4.1 presents the learned Policy Trees at depths 1 and 2. In the depth-1 tree, the optimal contact method is determined by **housingyes**, which denotes whether a client has an existing housing loan. Clients without a housing loan (housingyes ≤ 0) are optimally contacted via telephone, whereas clients with a housing loan (housingyes > 0) respond more favorably to cellular communication.

Allowing for greater policy complexity, the depth-2 Policy Tree introduces additional splits based on macroeconomic and demographic variables. The first split is based on **euribor3m**, which denotes the three-month Euribor rate and captures prevailing monetary conditions. When interest rates are relatively low (euribor3m ≤ 4.96), the optimal contact method depends on **age**, which denotes the client's age in years. Younger and middle-aged clients (age ≤ 79) are optimally contacted via cellular phone, whereas older clients (age > 79) are better reached through telephone contact.

When interest rates are higher (euribor3m > 4.96), housing loan status again becomes the key determinant: clients without a housing loan (housingyes ≤ 0) are optimally contacted via telephone, whereas clients with a housing loan (housingyes > 0) respond more favorably to cellular communication.

4.2.2 Model 2: Season (Homogeneous Effects)

This model evaluates whether the effectiveness of marketing campaigns varies across seasons and whether such variation can be exploited through personalized policies. Under the Best Fixed Strategy, contacting clients during the **Summer** achieves the highest average subscription rate, with an expected reward of 0.7630.

The Policy Tree algorithm, even at depth 2, delivers only a negligible improvement over this baseline (+0.01%). This result indicates that the **seasonality effect is largely homogeneous across the population**. In other words, Summer appears to be the optimal contact period for nearly all clients, regardless of their individual characteristics.

As a result, there is little scope for personalization along this dimension. Implementing a complex targeting policy for contact timing by season would therefore generate minimal additional value relative to a simple, uniform strategy.

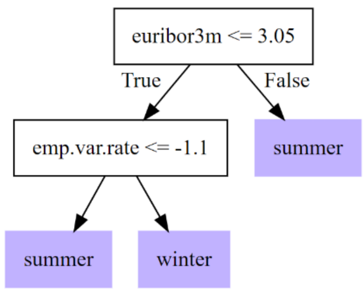

Figure4.2 Policy tree with contact month(only depth2)

Figure 4.2 presents the learned Policy Tree for contact month at depth 2, illustrating the absence of meaningful splits and reinforcing the conclusion of limited treatment effect heterogeneity.

4.2.3 Model 3: Day of Week (Moderate Optimization)

This model focuses on optimizing the **day of the week** for contacting clients. Under the Best Fixed Strategy, assigning all contacts to **Tuesday** yields an expected reward of 0.1586.

The Policy Tree identifies moderate but consistent improvements through personalization. At depth 2, the learned policy achieves an expected reward of 0.1730, corresponding to a 9.13% lift over the fixed strategy. The results suggest the presence of moderate treatment effect heterogeneity. From a managerial perspective, optimizing the day of contact can still generate meaningful efficiency improvements, particularly in the allocation of operational resources such as call center staffing and scheduling.

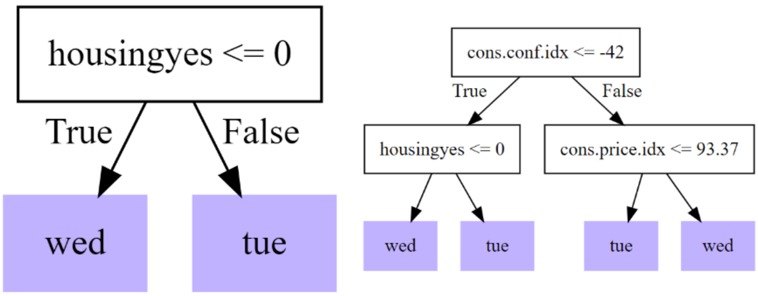

Figure4.3 Policy tree with day_of_week(depth1/2)

Figure 4.3 displays the learned Policy Trees for this treatment dimension. In the depth-1 tree, the optimal contact day is determined by **housingyes**, which denotes whether a client has an existing housing loan. Clients without a housing loan (housingyes $\leq$ 0) are optimally contacted on Wednesday, whereas clients with a housing loan (housingyes > 0) respond more favorably to Tuesday contact.

The depth-2 Policy Tree introduces additional heterogeneity driven by macroeconomic indicators. The first split is based on **cons.conf.idx**, which denotes the consumer confidence index, a monthly indicator of households' economic sentiment. When consumer confidence is relatively low (cons.conf.idx ≤ -42), housing loan status remains the key determinant: clients without a housing loan (housingyes ≤ 0) are optimally contacted on Wednesday, whereas clients with a housing loan (housingyes > 0) respond more favorably to Tuesday contact.

When consumer confidence is higher (cons.conf.idx > -42), the optimal contact day depends on **cons.price.idx**, which denotes the consumer price index, a monthly indicator of the overall price level. Clients are optimally contacted on Tuesday when the consumer price index is relatively low (cons.price.idx ≤ 93.37), and on Wednesday otherwise.

4.3 Summary of Managerial Implications

The results across the three models reveal a clear hierarchy in the value of personalization. Contact method selection exhibits the strongest and most economically significant heterogeneity, yielding substantial improvements in expected subscription rates. Day-of-week optimization shows moderate but consistent gains, suggesting that temporal targeting within the week can enhance operational efficiency when combined with client-level information. In contrast, contact timing by season displays little heterogeneity, indicating that personalization along this dimension offers minimal additional value.

From a managerial perspective, these findings imply that policy learning should be selectively applied. Resources and algorithmic complexity should be prioritized for decisions where heterogeneity is pronounced and actionable—most notably, the choice of communication channel. Secondary gains can be achieved by refining contact schedules during the week, while seasonal targeting can be handled effectively with simple, uniform rules.

More broadly, this analysis illustrates a central insight of causal policy learning: personalization is not universally beneficial, but when heterogeneity aligns with interpretable economic factors—such as financial commitments and macroeconomic conditions—it can deliver meaningful and implementable improvements in decision-making.

5. Conclusion and Discussion

In this project, we apply modern causal policy learning methods to optimize decision-making in a bank direct marketing context. Moving beyond traditional outcome prediction and average treatment effect analysis, we adopt a welfare-maximization framework that explicitly accounts for treatment effect heterogeneity. By combining doubly robust estimation with interpretable Policy Trees, the analysis translates causal insights into transparent and operational decision rules.

5.1 Summary of Findings

The empirical results yield three main conclusions. First, personalization is highly effective for selecting the contact method. The Policy Tree model for communication channel choice achieves substantial welfare gains relative to the best fixed strategy, with the depth-2 policy delivering a 26.64% improvement in expected subscription probability. These gains are driven by clear and interpretable heterogeneity associated with housing loan status, age, and macroeconomic conditions, indicating that communication preferences vary meaningfully across client segments.

Second, personalization provides moderate but consistent improvements for optimizing the day of the week of contact. While the magnitude of the gains is smaller than for contact method selection, the results nonetheless demonstrate non-negligible heterogeneity linked to clients' financial commitments and prevailing economic conditions, as captured by the consumer confidence and consumer price indices.

Third, in contrast to these dimensions, contact timing by season exhibits little scope for personalization. The Policy Tree models fail to outperform the best fixed strategy, suggesting that seasonal effects are largely homogeneous across the population. This result highlights that not all managerial decisions benefit from individualized targeting.

5.2 Managerial Implications

From a managerial perspective, the findings imply a clear prioritization of personalization efforts. The choice of contact method should be the primary focus of policy optimization, as it offers the largest and most robust gains in expected subscription rates. Implementing a simple, data-driven rule to assign communication channels based on observable client characteristics represents a low-cost operational change with potentially large returns.

Personalization of the day of the week can serve as a complementary strategy. Although the gains are more modest, aligning contact schedules with client profiles and economic conditions can improve operational efficiency, particularly in call center staffing and workload planning. In contrast, seasonal targeting does not require complex algorithms; concentrating marketing efforts during high-performing months is sufficient.

More broadly, these results illustrate how interpretable policy learning tools can guide managerial decision-making by identifying where personalization is most valuable and where simplicity is preferable.

5.3 Limitations

Several limitations of this analysis should be acknowledged. First, the study relies on observational data and the assumption of unconfoundedness. Although doubly robust estimation mitigates bias arising from model misspecification, the presence of unobserved confounders cannot be ruled out. Factors such as call content, agent behavior, or unmeasured aspects of client financial history may influence both treatment assignment and outcomes.

Second, the external validity of the learned policies may be limited. The data originate from a single Portuguese bank during a specific economic period, and the estimated thresholds in the policy trees reflect that context. Applying the same decision rules in different institutional or macroeconomic environments would require re-estimation.

Third, the analysis considers each treatment dimension in isolation. In practice, marketing decisions may involve joint or sequential choices—such as simultaneously selecting the contact method and timing—which are not captured in the current framework.

5.4 Future Directions

Future research could extend this work in several directions. First, field experiments or A/B testing could be conducted to validate the estimated welfare gains in a live operational setting. Second, dynamic policy learning methods could be employed to model sequential decision-making across multiple contacts or campaigns. Third, extending the framework to accommodate continuous or high-dimensional treatment spaces—such as call duration or intensity—would further enhance its practical relevance.

Overall, this project demonstrates that causal policy learning, when combined with interpretable models and robust estimation techniques, provides a powerful and actionable framework for improving decision-making in applied settings. By focusing on welfare maximization and transparency, Policy Trees offer a practical bridge between causal inference and real-world management.

References

Sverdrup, E., Kanodia, A., Zhou, Z., Athey, S., & Wager, S. (2020). *policytree: Policy learning via doubly robust empirical welfare maximization over trees*. *Journal of Open Source Software, 5*(50), 2232.

Athey, S., & Wager, S. (2017). Efficient policy learning. *arXiv preprint arXiv:1702.02896.*

Sverdrup, E., Uminsky, D., & Wager, S. (2020). Policy learning with observational data. *Econometrica, 88*(1), 273–308.

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems, 62*, 22–31.

Appendix. Reproducibility and Code Availability

To ensure transparency and reproducibility, all data preprocessing scripts, model estimation code, and policy learning implementations used in this project are publicly available in a GitHub repository:

GitHub Repository:

https://github.com/chiiillllll/policy-tree/tree/main

The repository contains the raw datasets, data cleaning and feature engineering scripts, causal forest estimation code, doubly robust score computation, policy tree estimation, and visualization of policy results.

The following code snippets illustrate the key implementation steps. The complete pipeline is available in the GitHub repository.

```
multi.forest <- grf::multi_arm_causal_forest(X, Y, W)

DR.scores <- double_robust_scores(multi.forest)

action_names <- levels(W)

tr <- policy_tree(X, DR.scores, depth = 1)

plot(tr, leaf.labels = action_names)
```