Uplift Modeling with Multi-Intervention Variables based on T-Learner

JiaLe Lei*

School of Software South China Normal University Foshan, China

*Corresponding author: leisir1211@gmail.com

Abstract. The paper addresses the limitations of traditional meta-learner models in the field of causal inference when dealing with multidimensional intervention variables and proposes an innovative modeling framework based on the traditional T-Learner framework. The paper proposes three different multi-intervention variable intervention strategies, namely the Link Strength Model, the Customer Value Model, and the Comprehensive Intervention Model, and selects XGBoost as the basic selector. The paper analyzes different evaluation indicators after model training, including the Qini curve, AUCC, Average Uplift/CATE, Uplift standard deviation, etc. The research results show that the comprehensive intervention model based on multidimensional comprehensive scoring has the most significant effect, which can effectively analyze the causal relationship between feature values and results under the condition of multiple intervention variables, and verifies the superiority of the proposed method in the estimation of intervention effects. This study provides an effective machine learning solution for causal inference in multi-intervention scenarios and has methodological significance for expanding the application boundaries of causal models.

Keywords: uplift modeling; multidimensional intervention variables; T-Learner; Machine Learning; XGBoost.

1. Introduction

In recent years, with the rise of artificial intelligence, machine learning has begun to play an important role in various industries. Traditional predictive models (such as response models) have focused more on the correlation and association between variables, answering the question of what.In various predictive analyses, traditional models such as logistic regression and random forests have long been the mainstream choices. These methods mainly establish correlations between features and specific results based on historical data, thus predicting whether a certain event will occur or not [1, 2].

However, in the actual process of solving problems, people are often not satisfied with just knowing the result in advance, but are more interested in knowing whether there is a certain causal relationship between specific variables and the result. The potential outcomes framework proposed by Rubin et al. provides a theoretical basis for identifying the causal effects of marketing interventions [3]. Under this framework, uplift modeling methods have emerged, with the core goal of estimating the Conditional Average Treatment Effect (CATE), which quantifies the net impact of marketing interventions on individual behavior [4]. Unlike traditional models, causal models attempt to infer the causal effect of one variable (intervention or treatment) on another variable (result). It answers the questions of Why or What if.

Meta-learners estimate CATE by combining traditional machine learning algorithms, which have become the most widely used method for improving modeling due to their simple implementation and strong interpretability. Among them, the S-Learner treats the treatment variable as a feature input to a single model, but may underestimate the heterogeneity treatment effect [5]; the T-Learner estimates CATE by separately constructing models for the treatment group and control group, but has the problem of error accumulation [6]; the X-Learner introduces a cross-training mechanism based on the T-Learner, performing better when there is an imbalance in the sample sizes of the treatment group and control group, as shown in the experiment by Künzel et al., where the X-Learner reduced the mean squared error (MSE) on multiple real datasets by about 12%–30% compared to the T-

Learner [5]; the R-Learner estimates CATE based on the orthogonalized machine learning method, showing excellent stability in the estimation of treatment effects, with its bias being 15% or more lower than that of traditional methods in highly confounded scenarios [7. 8].

Despite the great potential of Uplift Modeling, there is a significant limitation in current mainstream academic research and industrial practice, with the vast majority of methods based on the assumption of binary intervention variables. Künzel et al. proposed a better meta-algorithm, X-Learner, based on T-Learner, mainly discussing the case of binary treatment [9]. This means that the model can only answer whether the intervention variable has an impact on the outcome. However, many real-world academic and business scenarios are far from simple yes or no. Many intervention scenarios are inherently multidimensional, such as different channels and promotional strengths in marketing activities, which all affect users' willingness to purchase. These multidimensional intervention strategies together constitute a complex decision space.

Therefore, this paper proposes an innovative modeling framework based on the meta-learner T-Learner, extending Uplift Modeling from binary intervention to a multi-intervention variable pattern, which can handle intervention variables from multiple business dimensions simultaneously. To this end, this paper designs three different combined intervention strategies, namely, the direct combination strategy, the weighted combination strategy, and the feature scoring strategy. The paper is based on a model training and analysis using a dataset from the UCI Portuguese Bank Telemarketing dataset. Therefore, the paper aims to explore a uplift modeling method applicable to multi-dimensional intervention variables, thereby enabling more precise estimation of the causal effects of complex strategy combinations, providing quantifiable decision support for multi-dimensional decision-making scenarios in the real world.

2. Methodology

To better conduct causal analysis, the paper selected appropriate intervention variables for strategy segmentation after processing the original dataset. These include: directly combining customer value and historical response status to create intervention strategies, combining contact frequency and call duration to create weighted intervention strategies, and creating intervention strategies through feature scoring based on multiple dimensions such as contact intensity, customer value, response potential, and time factors. The paper can determine which modeling strategy better demonstrates the causal relationship between intervention variables and outcomes through comparison and analysis of different strategies.

2.1. Dataset

This article uses a bank marketing dataset from UCI, which contains 45211 records of a Portuguese bank's telemarketing campaign. Each record has 16 features (including age, occupation, deposit balance, previous marketing result, etc.) and a target variable, whether to subscribe to a fixed-term deposit (y) [1].

2.2. Data Processing

This article first completes the basic processing of the dataset, including checking and deleting duplicate rows, checking for missing values, and encoding categorical variables using label encoding.

Next, the paper conducted preliminary data analysis on the processed data, including the statistical collection of main information from the dataset and the analysis of the correlation between features and target variables. The preliminary statistical and analytical results were displayed using visual charts, as shown in Fig. 1.

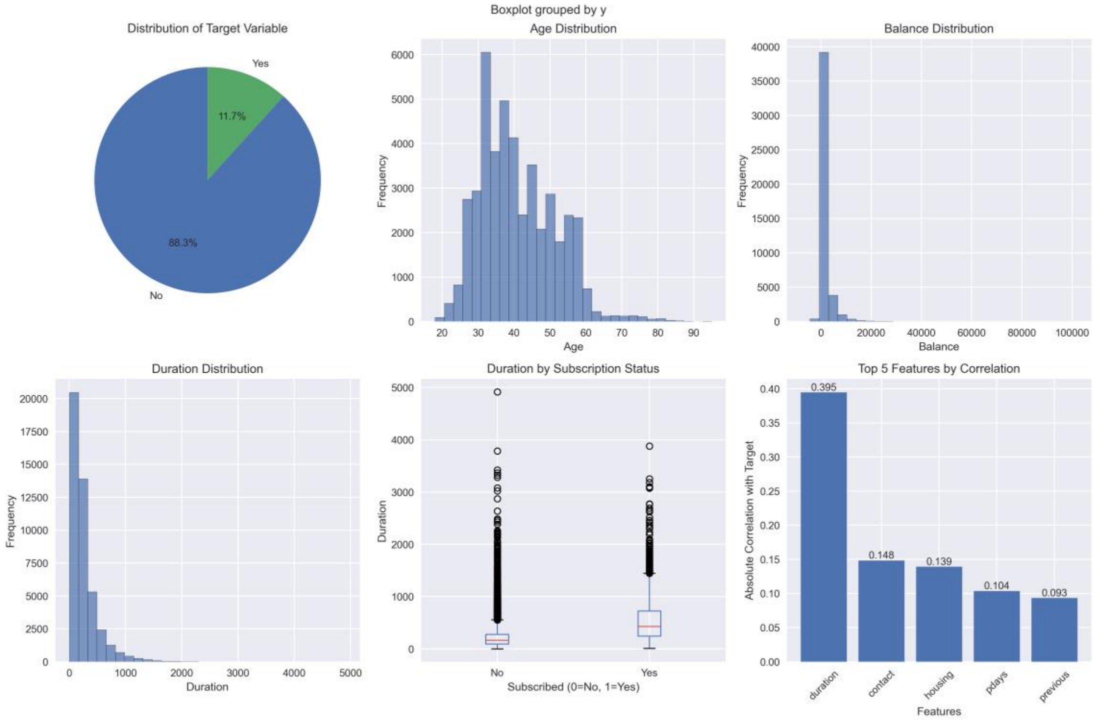

Fig. 1. Analysis result display chart

The six charts above display the following contents respectively, the pie chart of target variable distribution showing the proportion of "yes" and "no" categories, the histogram of age distribution - displays the distribution of customer ages, the histogram of balance distribution displaying the distribution of customer account balances, the histogram of contact duration distribution displaying the distribution of contact durations, the box plot of duration grouped by subscription status comparing the contact duration differences between subscribed and non-subscribed customers, the bar chart of top five most related features showing the five features with the highest correlation to the target variable.

2.3. Uplift Modeling

In terms of model selection, the paper adopts T-Learner and chooses XGBoost as the learner for training the models. The core of uplift modeling is to estimate the Conditional Average Treatment Effect (CATE), defined as follows :

$$\tau(x) = E[Y(1) - Y(0)|X = x] \tag{1}$$

Where: $Y(1)$ represents the potential outcome under the treatment condition. $Y(0)$ represents the potential outcome under the control condition (no treatment). X is a covariate or feature vector. $\tau(x)$ is the individual treatment effect for an individual with features x [5].

Modeling Process. T-Learner (also known as the two-tree estimator) is one of the fundamental methods in uplift modeling. It is conceptually simple yet powerful, making it a popular choice in practical applications.

The T-Learner method involves training two independent models: the treatment group model and the control group model.

$$\mu_1(x) = E[Y|X = x, W = 1] \tag{2}$$

$$\mu_0(x) = E[Y|X = x, W = 0] \tag{3}$$

In the above formulas, the paper defines the following variables: W is the treatment allocation variable (1 represents the treatment group, 0 represents the control group). $\mu_1(x)$ is the expected outcome of individuals with feature x in the treatment group. $\mu_0(x)$ is the expected outcome of individuals with feature x in the control group.

Subtract the result of the control group model from the result of the treatment group model, and the obtained difference ($\tau(x)$) is the value of CATE.

$$\tau(x) = \mu_1(x) - \mu_0(x) \tag{4}$$

Intervention Condition Design. The paper designed three different groups of combined intervention variables for model training. The paper trained a corresponding model for each group of multiple intervention variables, which were named the Contact Strength Model, Customer Value Model, and Comprehensive Intervention Model, respectively.

The first group of intervention variables (Contact Strength Model) adopted a combination of contact frequency and call duration. The paper set up three groups in this model. First is the control group (Treatment 0), which means a single contact with customers (campaign = 1). And the second one is the moderate intervention group (Treatment 1). This group includes customers who have contacted 2-3 times, and the total call duration does not exceed 6000 seconds. The last one is the strong intervention group (Treatment 2), which includes customers who contact more than 3 times or the total call duration exceeding 6000 seconds.

The second group of intervention variables (Customer Value Model) was based on a combination of customer value and historical response. The paper also set up 3 different groups in this model. The first one is the control group (Treatment 0), which contains new customers with low value(account balance ≤500 and no successful history). The second one is a standard contact group (Treatment 1), which involves medium-value, older customers (account balance: 500-1000 and a successful history). The third one is an active contact group (Treatment 2) with high-value customers (account balance >1000).

A comprehensive intervention variable design was used for multidimensional integrated scoring in the third group of intervention variables (Comprehensive Intervention Model). The core of this method is based on a comprehensive scoring of intervention variables across four dimensions.

$$S(x) = w_1 I_1(x) + w_2 I_2(x) + w_3 I_3(x) + w_4 I_4(x) \tag{5}$$

Among them, $\mathrm{S(x)}$ is the comprehensive score, $\mathrm{w_i}$is the weight of the i-th dimension, and $\mathrm{I_i(x)}$ is the indicator value of the i-th dimension. The four evaluation indicators respectively represent Contact Strength Indicator ($\mathrm{I_1(x)}$),Customer Value Indicator($\mathrm{I_2(x)}$),Response Potential Indicator($\mathrm{I_3(x)}$) and Time Factor($\mathrm{I_4(x)}$).Additionally, variables like $\mathrm{campaign(x)}$, $\mathrm{duration(x)}$, $\mathrm{balance(x)}$ and $\mathrm{previous(x)}$ are eigenvalues in dataset.

$$I_1(x) = \frac{campaign(x) \times duration(x)}{60} \tag{6}$$

$$I_2(x) = balance(x) + previous(x) \times 50 \tag{7}$$

$$I_3(x) = \begin{cases} 1 & if\ pdays(x) = -1 \\ 0 & otherwise \end{cases} \tag{8}$$

$$I_4(x) = \begin{cases} 1 & if\ month(x) \in \{4,5,8,11\} \\ 0 & otherwise \end{cases} \tag{9}$$

What's more, the paper set the following comprehensive scoring rules. Firstly, in Contact Strength ($\mathrm{I_1(x)}$), Low (≤50) = 1 point, Medium (≤150) = 2 points, High (>150) = 3 points. In Customer Value($\mathrm{I_2(x)}$), Low (≤500) = 1 point, Medium (≤2000) = 2 points, High (>2000) = 3 points. In Response Potential ($\mathrm{I_3(x)}$), Yes = 1 point and No = 0 points. Finally in Time Factor ($\mathrm{I_4(x)}$), Yes = 1 point and No = 0 points.

After a comprehensive evaluation, the paper formed the final intervention groups. The first one is the Control Group (Treatment 0). In this group, if the total score ≤3, the paper will define it as a low-intensity intervention. The second group is the Medium-intensity Intervention Group (Treatment 1). When the total score = 4-6, it will be considered a Medium-intensity intervention. Then the last one is the High-intensity Intervention Group (Treatment 2). If the total score >6, this group will be defined as a High-intensity intervention.

3. Results

After the model training, various evaluation indicators were obtained, and the paper integrated them into a table, as shown in Table I.

Table 1. Comparison of T-Learner Performance

Group	Intervention Group	Average Uplift/CATE	Uplift standard deviation	AUUC	Framework	Group
1	treatment_1	-0.0215	0.1009	0.1307	T-Learner	1
1	treatment_2	-0.0303	0.1262	0.0831	T-Learner	1
2	treatment_1	-0.1872	0.2238	0.1447	T-Learner	2
2	treatment_2	-0.1610	0.2136	0.1352	T-Learner	2
3	treatment_1	-0.0355	0.1258	0.1947	T-Learner	3
3	treatment_2	0.1091	0.3126	0.0352	T-Learner	3

Overall, the model trained with the integrated intervention variables obtained the highest positive Uplift value (0.1091) in the high-intensity intervention group (treatment_2), indicating that this model performs best in identifying customers who respond well to positive interventions. For other intervention variable groups, the paper found that the models trained with them showed relatively stable performance overall, but all the Uplift values were negative for all intervention groups. The model trained with customer value intervention variables showed the most negative effects, indicating that intervention strategies solely based on customer value may not be applicable. The model trained with integrated intervention variables showed the most diverse performance, with both negative and positive Uplift values, suggesting that considering multiple factors more comprehensively can more accurately identify customers with different response types.

Moreover, from the Qini curve (Figs. 2-4), it can be observed that the integrated intervention model has the best discriminative power in the high-intensity intervention group, while the customer value model's curve is relatively flat and has weaker discriminative power. Additionally, the contact strength model has some discriminative power in the moderate-intensity intervention group.

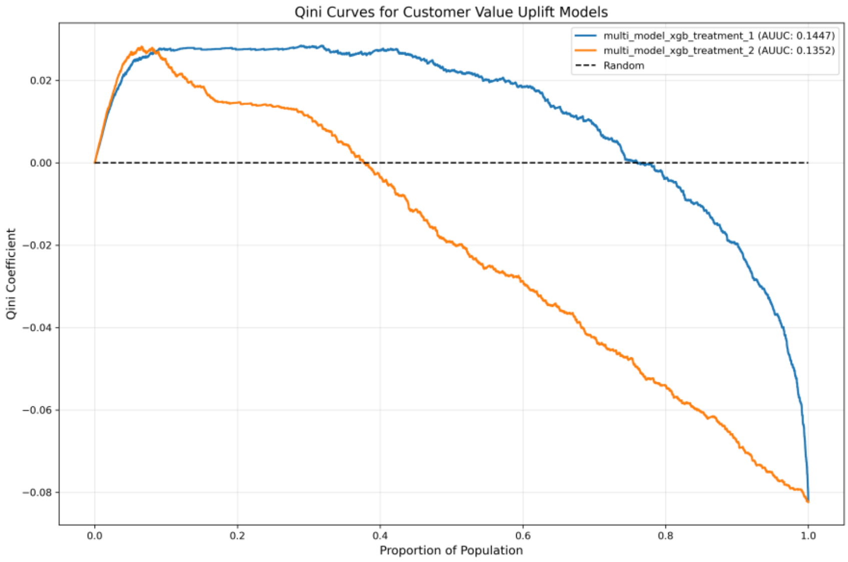

Fig. 2. Qini curve plots for customer value uplift models (Photo/Picture credit: Original)

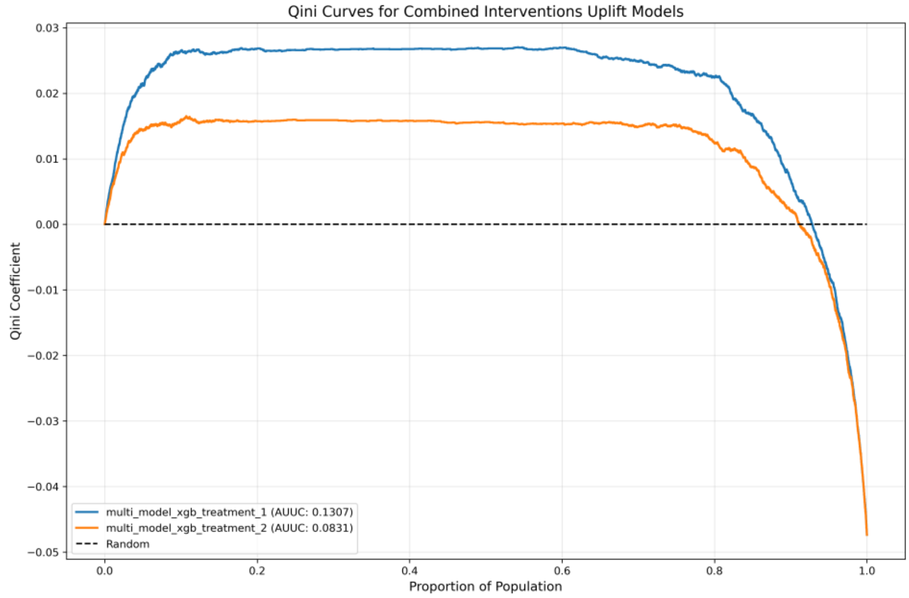

Fig. 3. Qini curve plots for combined interventions uplift models (Photo/Picture credit: Original)

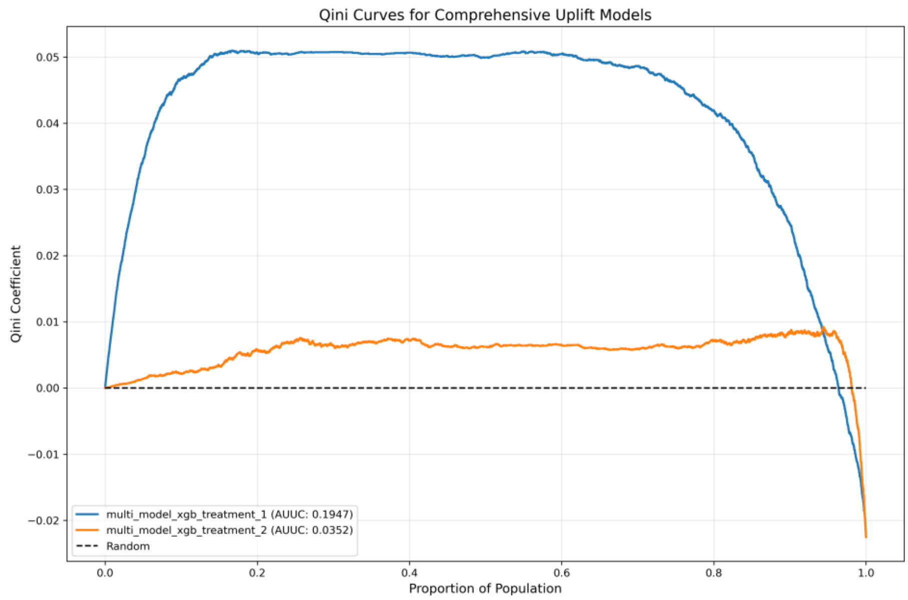

Fig. 4. Qini curve plots for comprehensive uplift models (Photo/Picture credit: Original)

It can be seen from the curve in Fig. 2 that the curve is relatively flat, indicating that the model's predictive ability is not very good.

Compared to Fig. 2, the curve in Fig. 3 is steeper, indicating that the model's predictive ability is better than that in Fig. 2.

Fig. 4 shows that the group of treatment 2 exhibits excellent results. We can observe that the curve of this group is significantly better than that of all other models, which proves that the multidimensional comprehensive score in the paper can more accurately identify potential high-value customers.

From the model comparison chart, it can be seen that the performance differences among the models are significant under different intervention groups. However, the high-intensity intervention group of the comprehensive intervention model performs the best, consistent with the numerical results.

In summary, the experimental results show that the comprehensive intervention strategy integrating multidimensional business information can identify the truly sensitive customer groups for marketing activities more comprehensively and accurately than other intervention strategies. This means that the comprehensive intervention model proposed in the paper can effectively identify causal relationships under the scenario of multiple intervention variables, proving that the comprehensive intervention strategy proposed in the paper has a relatively high feasibility.

4. Discussion

After a series of design and model training, the paper finds that the multi-intervention strategy based on the comprehensive intervention model has a significant effect on causal prediction. This fully demonstrates the advantages of the T-Learner framework and the comprehensive intervention model, not only providing a new perspective for solving the challenges of bank marketing [8] but also highlighting the tremendous benefits brought by the multi-intervention variable modeling framework. However, the model proposed in this paper still has shortcomings. Firstly, the dataset used in the paper has a limited amount of data, which limits the further optimization of the model. Secondly, there are still doubts about the scientificity of the comprehensive scoring strategy set in the paper. Whether the generalization ability of the model can be improved by using more intervention variables in combination remains to be observed.

But fortunately, many previous studies have provided ideas for the development of subsequent research on the paper. Devriendt et al. point out that different methods of Uplift Modeling have their own advantages and disadvantages, and their effectiveness is highly dependent on the application background [10]. Bojinov et al. emphasized the challenges and importance of causal inference in time series data, providing a theoretical basis for studying the long-term and delayed effects of marketing activities [11]. Zhao et al. expanded the framework of traditional binary interventions, providing a theoretical basis for simultaneously evaluating the effects of various marketing channels such as phone calls, emails, and SMS [12].

In the future, the paper will conduct more in-depth research based on the comprehensive intervention model, seek more diverse datasets, integrate more scientific indicators, and construct a new algorithm framework for more comprehensive causal analysis. In addition, the paper will also focus on other advanced algorithms in the academic circle, integrate them into the framework, and continuously optimize and improve them.

5. Conclusion

To address the issue that traditional uplift modeling cannot effectively handle multiple intervention variables, the paper proposes an integrated intervention strategy based on a scoring mechanism and fully demonstrates its reliability, solving the problem that native meta-learners cannot directly handle multiple intervention variables. The paper uses a bank dataset for training, and through the above results analysis, proves that the integrated intervention model can effectively identify customer groups sensitive to marketing interventions in the case of multiple intervention variables.

The combination analysis of multiple intervention variables reveals heterogeneous treatment effects of different marketing strategies. Although no model performs the best on all evaluation indicators, the different results that can be brought about by different intervention variable designs fully demonstrate that users with different characteristics have different responses to the same bank telephone marketing strategy. Therefore, it is crucial to customize marketing strategies suitable for users with different characteristics. For example, in almost all groups, the value of Average Uplift is negative, indicating that excessive contact may lead to a decrease in users' "liking" for the bank's marketing strategy, which in turn affects subscription results.

The contribution of this study is mainly reflected in two aspects: theory and practice. In terms of theoretical methods, this study expands the application scope of Uplift Modeling from simple binary interventions to the field of multi-intervention variables, providing feasible ideas and frameworks for subsequent research on how to handle more complex and realistic scenarios. In terms of business practice, the model proposed in the paper provides a feasible data analysis method for different industry fields. Data analysts can comprehensively consider more intervention variables to scientifically measure the causal relationship between features and results, thereby effectively identifying customer groups that respond well to positive interventions.

References

[1] P. Gutierrez and J. Y. Gérardy, “Causal inference and uplift modelling: A review of the literature,” in Int. Conf. Predictive Appl. APIs, Jul. 2017, pp. 1–13. PMLR.

[2] D. Olaya, K. Coussement, and W. Verbeke, “A survey and benchmarking study of multitreatment uplift modeling,” Data Min. Knowl. Discov., vol. 34, no. 2, pp. 273–308, 2020.

[3] D. B. Rubin, “Estimating causal effects of treatments in randomized and nonrandomized studies,” J. Educ. Psychol., vol. 66, no. 5, p. 688, 1974.

[4] N. J. Radcliffe and P. D. Surry, “Real-world uplift modelling with significance-based uplift trees,” White Paper TR-2011-1, Stochastic Solutions, pp. 1–33, 2011.

[5] S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu, “Metalearners for estimating heterogeneous treatment effects using machine learning,” Proc. Natl. Acad. Sci. U.S.A., vol. 116, no. 10, pp. 4156–4165, 2019.

[6] Y. Zhao, X. Fang, and D. Simchi-Levi, “Uplift modeling with multiple treatments and general response types,” in Proc. 2017 SIAM Int. Conf. Data Min., Jun. 2017, pp. 588–596. Society for Industrial and Applied Mathematics.

[7] X. Nie and S. Wager, “Quasi-oracle estimation of heterogeneous treatment effects,” Biometrika, vol. 108, no. 2, pp. 299–319, 2021.

[8] B. Barile, M. Forti, A. Marrocco, and A. Castaldo, “Causal impact evaluation of occupational safety policies on firms’ default using machine learning uplift modelling,” Sci. Rep., vol. 14, no. 1, p. 10380, 2024.

[9] S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu, “Metalearners for estimating heterogeneous treatment effects using machine learning,” Proc. Natl. Acad. Sci. U.S.A., vol. 116, no. 10, pp. 4156–4165, 2019.

[10] F. Devriendt, D. Moldovan, and W. Verbeke, “A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics,” Big Data, vol. 6, no. 1, pp. 13–41, 2018.

[11] I. Bojinov and N. Shephard, “Time series experiments and causal estimands: exact randomization tests and trading,” J. Am. Stat. Assoc., vol. 114, no. 528, pp. 1665–1682, 2019.

[12] Y. Zhao, X. Fang, and D. Simchi-Levi, “Uplift modeling with multiple treatments and general response types,” in Proc. 2017 SIAM Int. Conf. Data Min., Jun. 2017, pp. 588–596. Society for Industrial and Applied Mathematics.