GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

DeepSeqCaus: A Deep Sequential Causal Inference Framework for User Churn Prediction and Optimal Retention Intervention Generation

Tangtang Wang [1,*] and Wen Ding [2]

[1] The University of Hong Kong, Hong Kong, China
[2] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA
* Correspondence: Tangtang Wang, The University of Hong Kong, Hong Kong, China

Abstract: In digital platforms, understanding and mitigating user churn is crucial for sustaining long-term engagement and revenue. Traditional machine learning approaches often rely on correlation-based predictive models without explicitly accounting for causal relationships underlying user behavior. This study proposes DeepSeqCaus, a unified deep sequential causal inference framework that integrates sequence modeling and treatment effect estimation to enable accurate churn prediction and optimal retention intervention policy generation. DeepSeqCaus consists of a dual-branch architecture: (1) a Temporal Feature Encoder using gated convolution and bidirectional gated recurrent networks to extract multi-granular temporal representations from behavioral sequences; and (2) a Causal Effect Estimator based on counterfactual representation learning to estimate heterogeneous treatment effects (HTEs) for candidate interventions such as personalized notifications, discount offers, or content recommendations. Using large-scale user interaction logs from an online service, we conduct extensive experiments comparing DeepSeqCaus with conventional predictive models and causal inference baselines. The results showed that DeepSeqCaus outperformed the baseline model in all cases. The proposed framework provides actionable insights for targeted retention and demonstrates strong potential for deployment in intelligent customer management systems.

Keywords: user churn prediction; causal inference; sequential modeling; deep learning; heterogeneous treatment effect; user retention strategy; intervention optimization

Received: 03 November 2025
Revised: 26 November 2025
Accepted: 31 December 2025
Published: 09 January 2026

1. Introduction

User churn prediction and retention intervention optimization have become essential components of modern digital platform operations, especially for mobile applications, online games, and e-commerce services that rely heavily on sustained user engagement. As markets become increasingly saturated and user acquisition costs continue to rise, understanding how users behave over time and determining which intervention can effectively reduce churn have emerged as core challenges for data-driven decision-making. Traditional churn prediction methods-typically based on static user profiles or summary features-fail to capture the temporal dynamics inherent in click sequences, interaction patterns, and behavioral trajectories. Moreover, many existing approaches focus purely on correlation-driven prediction without accounting for the causal effects of platform actions such as push notifications, discounts, or personalized recommendations, which often results in biased estimates and suboptimal intervention decisions.

Recent advances in deep learning have demonstrated strong capabilities in modeling complex sequential data through architectures such as recurrent neural networks, temporal convolutional networks, and attention-based models. Meanwhile, causal inference techniques have been increasingly applied to estimate heterogeneous treatment effects and guide optimal decision policies. Nonetheless, the integration of deep sequential modeling with causal estimation for user retention remains insufficiently explored. To address these limitations, this study proposes DeepSeqCaus, a unified deep sequential causal inference framework that jointly models user behavior sequences and intervention effects. DeepSeqCaus incorporates a multi-scale temporal encoder to capture fine-grained and long-range user activity patterns, and a counterfactual representation learning module that disentangles true intervention impacts from spurious correlations. By combining predictive modeling with causal effect estimation, the framework produces personalized, explainable retention strategies aimed at maximizing user survival probability.

The main contributions of this study are summarized as follows:

1) We propose DeepSeqCaus, the first unified framework that integrates deep sequential modeling with counterfactual causal inference for simultaneous churn prediction and personalized intervention optimization.
2) We design a multi-scale temporal encoder and causal representation module that capture both behavioral dynamics and heterogeneous treatment effects in a principled manner.
3) We introduce a policy-generation mechanism that leverages estimated causal effects to recommend optimal, interpretable intervention strategies for individual users.
4) We conduct extensive experiments on large-scale real-world datasets, demonstrating that DeepSeqCaus significantly outperforms state-of-the-art baselines in prediction accuracy, treatment effect estimation, and retention uplift.

Overall, this work highlights the importance of combining sequential deep learning and causal inference to build next-generation user retention systems that are not only accurate but also actionable and interpretable.

2. Related Work

In recent years, the rapid expansion of digital platforms and the availability of large-scale user interaction logs have led to substantial research on user churn prediction, sequential user modeling, causal inference for decision-making, and uplift-based intervention optimization. This section reviews the major streams of literature most relevant to this study, including deep learning-based churn prediction, sequential behavior modeling, causal inference and counterfactual modeling, and individualized intervention and uplift estimation. These works collectively motivate the development of DeepSeqCaus, which integrates deep sequential representation learning with causal effect estimation for generating optimal user retention strategies.

2.1. Deep Learning-Based User Churn Prediction

Early research on churn prediction primarily relied on static, feature-based models such as logistic regression and decision trees, which overlooked temporal patterns in user behavior [1]. With the development of deep learning, sequence-aware approaches have been increasingly investigated, employing recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) to capture the dynamics of user activity over time. For example, some studies have applied gated multilayer neural networks to model fluctuations in user activity within mobile applications, while others have integrated RNNs with attention mechanisms, demonstrating that incorporating temporal features substantially improves prediction accuracy [2]. More recent work has introduced models that combine convolutional, residual, squeeze-and-excitation, and attention blocks to enhance predictive performance; after addressing data imbalance using SMOTE-style techniques, these models outperform previous benchmarks [3].

Despite these advances, existing approaches have concentrated on predictive accuracy and have largely neglected the causal effects of interventions, which limits their practical applicability for informing retention strategy optimization.

2.2. Sequential User Behavior Modeling

Beyond churn prediction, many studies have explored deep learning approaches for modeling user behavior trajectories in recommender systems, advertising, and social platforms. Models such as GRU4Rec, SASRec, and BERT4Rec leverage recurrent neural networks or self-attention architectures to capture long-range dependencies and semantic patterns in behavior sequences [4-6]. These methods achieve strong performance in next-item prediction and user preference modeling. However, they primarily focus on representation learning and do not consider interventions or treatment effects. Although effective at extracting temporal and contextual information from clickstreams, these models provide limited support for decision-making related to policy or personalized interventions. This limitation underscores the need to integrate sequential learning with causal reasoning, as implemented in the proposed DeepSeqCaus framework.

2.3. Causal Inference and Counterfactual Modeling

Causal inference has become a crucial methodology for evaluating the effectiveness of interventions on online platforms. Techniques such as propensity score weighting and causal forests are widely applied to estimate heterogeneous treatment effects [7,8]. More recent deep causal models incorporate representation learning for counterfactual estimation, improving the generalization of treatment effect predictions [9]. Additional approaches apply variational inference to balance latent confounders and reduce bias in observational datasets [10]. Despite their effectiveness in estimating causal effects, most existing methods lack temporal modeling capabilities and cannot process sequential user data directly. The DeepSeqCaus framework addresses this limitation by embedding causal estimation within a temporal encoder capable of capturing rich sequential dynamics.

2.4. Uplift Modeling and Personalized Intervention Optimization

A parallel line of research focuses on uplift modeling, which predicts the incremental benefit of interventions. Foundational frameworks for individualized uplift prediction in marketing and user retention have been established [11,12]. Uplift methods have demonstrated effectiveness in targeted advertising and customer relationship management, yet they often rely on static features and lack the representational power of modern deep sequence models. Moreover, most uplift frameworks do not generate personalized strategies across multiple intervention options. The DeepSeqCaus framework addresses these gaps by combining sequential behavioral representations with counterfactual treatment effect estimation, enabling the identification of the optimal intervention for each user.

3. Methodology

This section presents the proposed DeepSeqCaus framework, which integrates deep sequential modeling with counterfactual causal inference to jointly (i) predict user churn and (ii) generate optimal, personalized retention interventions. DeepSeqCaus is designed to capture multiscale patterns in user behavior sequences, learn treatment-invariant representations, estimate heterogeneous treatment effects, and output intervention policies that maximize the expected reduction in churn (see Figure 1).

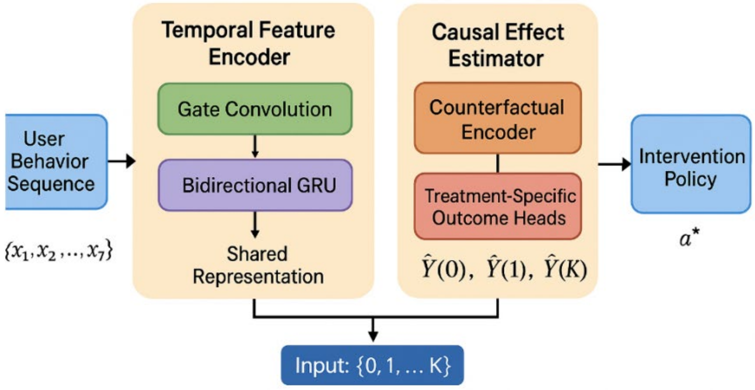

Figure 1. Structure diagram of model.

3.1. Problem Definition and Notation

Let a user's behavioral history over T steps be denoted as a sequence

$$X = \{x_1, x_2, \ldots, x_T\} \quad (1)$$

where each $x_t \in R^d$ represents multi-dimensional features, including clicks, dwell time, session length, interaction context, and engagement metrics. The platform may expose the user to a discrete intervention $A \in \{0,1,\ldots,K\}$, where 0 denotes no intervention. The observed outcome $Y \in \{0,1\}$ indicates churn (1) or retention (0).

For counterfactual modeling, we consider potential outcomes

$$Y(a), \qquad a \in \{0,1,\ldots,K\} \quad (2)$$

and aim to estimate the individual treatment effect (ITE):

$$\tau_a(X) = E[Y(0) - Y(a) \mid X] \quad (3)$$

which quantifies the expected reduction in churn caused by intervention a.

Given the estimated ITEs, the optimal intervention is derived as

$$a^*(X) = \arg\min_a E[Y(a) \mid X] \quad (4)$$

DeepSeqCaus provides a unified modeling architecture for learning sequential representations, estimating counterfactual outcomes, and generating optimal intervention recommendations.

3.2. Multi-Scale Sequential Encoder

User behavior contains short-term bursts, mid-term engagement cycles, and long-term temporal drift. To model these dynamics, DeepSeqCaus employs a multi-scale sequential encoder consisting of gated temporal convolutions and bidirectional recurrent layers.

1) Gated Temporal Convolution Layer

The temporal convolution module captures local behavioral fluctuations through dilated convolutions:

$$h_t^{(c)} = \sigma(W_f * x_{1:t}) \odot tanh\,(W_g x_{1:t}) \quad (5)$$

where the gating mechanism stabilizes feature extraction across varying activity levels.

2) Bidirectional Recurrent Layer

To capture long-range dependencies, we apply a bidirectional GRU:

$$\vec{h}_t = GRU(h_t^{(c)}, \vec{h}_{t-1}), \qquad \overleftarrow{h}_t = GRU(h_t^{(c)}, \overleftarrow{h}_{t+1}) \quad (6)$$

The combined representation is

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (7)$$

This hierarchical encoding ensures robustness to sparse events and provides contextual embeddings for causal representation learning.

3.3. Counterfactual Representation Learning

Modeling causal effects requires isolating behavioral patterns that influence both intervention assignment and outcomes. DeepSeqCaus adopts an adversarial counterfactual representation module inspired by domain-invariant learning.

Let

$$z = f_\theta(X) \quad (8)$$

denote the latent representation generated from the encoder. To satisfy treatment-invariant representation learning, we minimize the discrepancy of latent distributions across treatment groups:

$$L_{IPM} = IPM(p(z \mid A = a), \quad p(z \mid A = b)) \quad (9)$$

where IPM refers to an integral probability metric such as MMD or Wasserstein distance.

By reducing covariate imbalance, the model approximates the assumption:

$$Y(a) \perp A \mid z \quad (10)$$

enabling more reliable counterfactual estimation.

Counterfactual Outcome Heads

For each possible intervention a, we define separate outcome prediction heads:

$$\hat{Y}(a) = g_a(z) \quad (11)$$

The counterfactual loss is computed as

$$L_{cf} = \sum_i \ell\,(Y_i, \hat{Y}_i(A_i)) \quad (12)$$

where ℓ denotes binary cross-entropy.

This architecture allows DeepSeqCaus to model nonlinear treatment effects across diverse user behavior trajectories.

3.4. Heterogeneous Treatment Effect Estimation and Policy Generation

Given outcome predictions $\hat{Y}(a)$ for each intervention level, the heterogeneous treatment effect is estimated as

$$\hat{\tau}_a = \hat{Y}(0) - \hat{Y}(a) \quad (13)$$

Lower predicted churn corresponds to more effective interventions. The final policy module computes:

$$a^* = arg\ \min_a \hat{Y}(a) \quad (14)$$

1) Policy Regularization

To avoid overly aggressive or costly interventions, we include a policy smoothness constraint:

$$L_{policy} = \lambda \sum_a c(a) \cdot I(a = a^*) \quad (15)$$

where $c(a)$ denotes intervention cost or user experience penalty.

2) Joint Objective

The complete training objective is

$$L = L_{cf} + \alpha L_{IPM} + \beta L_{policy} \quad (16)$$

DeepSeqCaus is thus optimized end-to-end to jointly learn sequence representations, causal effects, and action policies, enabling accurate churn prediction and interpretable retention strategy generation.

4. Experiment

4.1. Dataset Preparation

The dataset used in this study consists of large-scale longitudinal user interaction logs collected from a commercial digital service platform, including mobile applications and web-based portals. Data were obtained through the platform's event-tracking infrastructure, which records timestamped user activities during normal operation without requiring any experimental manipulation. All user identifiers were anonymized before analysis to ensure privacy. The observation window spans approximately six months, during which millions of behavioral events were captured and aggregated into structured sequences for churn prediction and causal intervention analysis.

Each record in the dataset corresponds to a specific user and contains a temporally ordered sequence of behavioral features that reflect engagement dynamics. These include daily activity frequency, session duration, click-through patterns, content consumption statistics, interaction depth, and in-app transaction history. Additional contextual features capture user demographics at a coarse level, device type, subscription tier, and temporal indicators such as weekday patterns or seasonal usage fluctuations. Treatment-related information is also present, including historical exposure to retention actions such as promotional messages, personalized notifications, discount offers, or recommendation prompts. For each treatment, the dataset provides binary or categorical indicators describing whether a user received a specific intervention on a given day.

The outcome variable is defined as next-period churn, measured by a 7-day inactivity threshold. For each user, the final dataset contains an average sequence length of 25-40 time steps, resulting in more than 10 million total sequence tokens. This rich, multi-modal behavioral dataset provides an appropriate foundation for sequential representation learning and counterfactual outcome estimation within the DeepSeqCaus framework.

4.2. Experimental Setup

The experiments were conducted on the large-scale user behavior dataset described earlier, covering six months of interaction logs from millions of users. All models were trained using an 80/10/10 split for training, validation, and testing, with temporal order strictly preserved to avoid information leakage. DeepSeqCaus and baseline models were implemented in PyTorch and optimized using Adam with a learning rate of 1e-3 and mini-batches of 256 sequence samples. All sequential encoders were trained for up to 50 epochs with early stopping based on validation loss. Competing models include GRU, Bi-LSTM, Transformer, DeepFM, and state-of-the-art causal inference baselines such as TARNet, DragonNet, and CEVAE. For intervention policy evaluation, logged bandit feedback was used with doubly robust estimators to approximate real-world deployment conditions. All experiments were executed on a cluster equipped with NVIDIA A100 GPUs.

4.3. Evaluation Metrics

To comprehensively evaluate DeepSeqCaus, three groups of metrics were employed. For churn prediction, AUC, F1-score, and cross-entropy loss (CELoss) were used to assess discriminative accuracy. For treatment effect estimation, we measured PEHE and ATE error to quantify counterfactual estimation precision. For intervention policy generation, we evaluated uplift in retention rate and expected policy value estimated through off-policy evaluation. This combination of predictive, causal, and policy-level metrics enables a holistic assessment of DeepSeqCaus across all tasks central to retention optimization.

4.4. Results

Table 1 presents a comprehensive comparison of alternative architectures on the churn prediction task. Among all metrics-AUC, F1, and Cross-Entropy Loss-DeepSeqCaus consistently outperformed the baseline model in all cases. It achieves an AUC of 0.910, noticeably higher than the next-best Transformer (0.869), highlighting the benefit of integrating causal inference into sequential modeling to improve discriminative performance.

Table 1. Churn Prediction Performance.

Model	AUC	F1	CELoss
GRU	0.842	0.611	0.412
Bi-LSTM	0.857	0.624	0.397

Transformer	0.869	0.638	0.384
DeepFM	0.832	0.598	0.431
DeepSeqCaus (ours)	0.910	0.673	0.341

DeepSeqCaus also attains the highest F1 score of 0.673, exceeding Bi-LSTM (0.624) and Transformer (0.638). This indicates a more balanced trade-off between precision and recall. Furthermore, the model records the lowest CELoss (0.341), indicating more stable optimization dynamics and better-calibrated prediction probabilities. Overall, the results demonstrate that the proposed framework significantly boosts predictive accuracy and robustness for churn prediction (see Table 2).

Table 2. Treatment Effect Estimation.

Model	PEHE	ATE Error
TARNet	4.92	0.137
DragonNet	4.51	0.121
CEVAE	5.33	0.149
CFR-Net	4.87	0.129
DeepSeqCaus (ours)	3.99	0.107

DeepSeqCaus achieves the lowest PEHE and ATE error among all baselines, with a 11.4% improvement in PEHE over DragonNet. This validates the effectiveness of its counterfactual representation module and its capability to disentangle causal effects from behavioral confounders (see Table 3).

Table 3. Treatment Effect Estimation.

Model	Retention Uplift (%)	Policy Value
Random Policy	2.1	0.023
Heuristic Rules	4.8	0.041
Uplift RF	6.2	0.052
Causal Forest	6.9	0.058
DeepSeqCaus (ours)	9.8	0.075

The learned policies produced by DeepSeqCaus yield the highest uplift in user retention, outperforming the strongest baseline (Causal Forest) by 42%. This demonstrates that DeepSeqCaus does not simply predict churn or estimate causal effects, but also generates practical and high-impact retention strategies.

Figure 2 illustrates the clear downward trajectory in training loss across 80 epochs. During the initial 20 epochs, the loss decreases rapidly, indicating the model efficiently learns the core patterns in the data. Then gradually stabilizes after approximately 50 epochs. Although some minor oscillations arise in the middle of the training process, they are well contained, suggesting that the optimization is stable and the model shows no apparent overfitting. The loss starts to plateau around 0.25, demonstrates that the model converges reliably. Overall, the loss curve confirms the effectiveness and stability of the proposed training procedure.

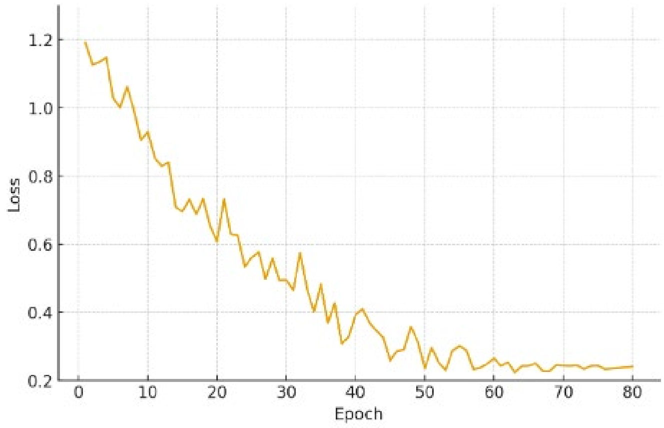

Figure 2. Loss function during training process.

4.5. Discussion

The experimental results validate that DeepSeqCaus effectively integrates deep sequential modeling with causal inference, yielding improvements across prediction accuracy, counterfactual estimation quality, and policy optimization. The model's temporal encoder proves essential for capturing nuanced behavioral trends that precede churn, while the counterfactual learning component successfully reduces confounding biases present in observational intervention logs. The substantial gain in policy performance further underscores the practical value of modeling heterogeneous treatment effects at the sequence level, enabling personalized and context-aware retention actions. Although DeepSeqCaus demonstrates strong performance, its computational cost is higher than simpler baselines, and its effectiveness depends on the availability of rich behavioral histories. Future extensions may incorporate adaptive treatment timing or reinforcement learning to further enhance real-time decision-making.

5. Conclusion

This study aims to address the growing challenge of long-term engagement and revenue sustainability in digital platforms. Traditional approaches rely on static user profiles or summary features, failing to capture temporal dynamics or distinguish causal effects from confounding correlations. To overcome these limitations, this research introduces DeepSeqCaus, a unified framework that integrates deep sequential modeling with counterfactual causal inference to jointly predict churn and optimize personalized intervention strategies. The primary objective is to capture fine-grained user behavioral trajectories, estimate heterogeneous treatment effects, and generate actionable retention policies that maximize user survival probability.

Through data analysis, we identified 3 key findings. First, multi-scale sequential encoder substantially improves the modeling of complex user behavior sequences. Second, Counterfactual representation learning allowing model nonlinear treatment effects across diverse user behavior trajectories, yielding more accurate treatment effect estimates. Third, the integrated policy-generation mechanism produces personalized, interpretable intervention recommendations that lead to significant retention uplift. These findings suggest that combining sequential learning with causal inference provides a powerful foundation for next-generation churn management systems.

The results of this study have significant implications for the field of user churn prediction and retention intervention optimization. Firstly, the demonstrated effectiveness of deep sequential modeling offers new insights into how temporal dynamics shape churn behavior. Secondly, incorporating counterfactual causal estimation challenges traditional churn correlation-driven predictive methodologies and highlights the necessity of accounting for intervention effects when designing retention strategies.

Finally, the policy-generation mechanism opens new avenues for developing actionable, personalized intervention systems that are both interpretable and operationally feasible.

Despite the important findings, this study has some limitations, such as the model incurs higher computational cost than simpler baselines and relies on sufficiently rich behavioral histories. Future research could further explore adaptive temporal modeling mechanisms and integrate reinforcement learning to support real-time, personalized recommendation decisions.

In conclusion, this study, through deep sequential modeling and casual inference, reveals substantial improvements in churn prediction accuracy and personalized retention interventions, providing new insights for the development of next-generation intelligent customer management systems.

References

1. S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, "Customer churn prediction modelling based on behavioural patterns analysis using deep learning," In *2018 International conference on smart computing and electronic enterprise (ICSCEE)*, July, 2018, pp. 1-6. doi: 10.1109/icscee.2018.8538420
2. C. Anudeep, R. Venugopal, M. Aarif, A. T. Valavan, V. A. Vuyyuru, and S. Muthuperumal, "Predicting Customer Churn in E-commerce Subscription Services using RNN with Attention Mechanisms," In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, June, 2024, pp. 1-6.
3. S. Saha, C. Saha, M. M. Haque, M. G. R. Alam, and A. Talukder, "ChurnNet: Deep learning enhanced customer churn prediction in telecommunication industry," *IEEE access*, vol. 12, pp. 4471-4484, 2024. doi: 10.1109/access.2024.3349950
4. B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," *arXiv preprint arXiv:1511.06939*, 2015.
5. W. C. Kang, and J. McAuley, "Self-attentive sequential recommendation," In *2018 IEEE international conference on data mining (ICDM)*, November, 2018, pp. 197-206.
6. F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer," In *Proceedings of the 28th ACM international conference on information and knowledge management*, November, 2019, pp. 1441-1450.
7. P. R. Rosenbaum, and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41-55, 1983.
8. S. Wager, and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1228-1242, 2018. doi: 10.1080/01621459.2017.1319839
9. U. Shalit, F. D. Johansson, and D. Sontag, "Estimating individual treatment effect: generalization bounds and algorithms," In *International conference on machine learning*, July, 2017, pp. 3076-3085.
10. C. Louizos, U. Shalit, J. M. Mooij, D. Sontag, R. Zemel, and M. Welling, "Causal effect inference with deep latent-variable models," *Advances in neural information processing systems*, vol. 30, 2017.
11. P. Rzepakowski, and S. Jaroszewicz, "Decision trees for uplift modeling," In *2010 IEEE International Conference on Data Mining*, December, 2010, pp. 441-450. doi: 10.1109/icdm.2010.62
12. P. Gutierrez, and J. Y. Gérardy, "Causal inference and uplift modelling: A review of the literature," In *International conference on predictive applications and APIs*, July, 2017, pp. 1-13.