Optimizing Preferential Rate in Retail Lending with Causal Inference and Domain Adaptation

Jimyung Choi[1], Yujin Lee[1], Hyeryeong Oh[1], Sumin Shin[1], Jaehyun Kim[1], Wooyoung Kim[1], Kee-Eung Kim[2],

[1] Hyundai Capital Services, Korea
[2] Kim Jaechul Graduate School of AI, KAIST, Korea
{jimyung.choi, yujin8.lee, hyeryeong.oh, sumin.shin, jaehyun2.kim, wooyoung.kim}@hcs.com, kekim@kaist.ac.kr

Abstract

In retail lending, offering preferential interest rates is a core marketing instrument for balancing customer acquisition with portfolio profitability. Accurately predicting the effect of interest-rate discounts for each customer is pivotal for optimizing the discount strategy: offering overly generous discounts erodes margins, while insufficient discounts drive price-sensitive customers to defect. Off-the-shelf machine learning uplift models rarely respect the complex operational constraints of financial business, such as tiered rate grids, regulatory guard-rails, and marketing budget ceilings. We propose an integrated system that fuses causal inference and domain adaptation to produce constraint-aware, customer-specific discount recommendations. To further enhance practitioner adoption, a large language model layer translates model outputs into actionable narratives. Developed in Hyundai Capital Services, the system boosted transaction volume by 13%, demonstrating both technical soundness and material business impact.

Introduction

Finance companies, specializing in retail lending, operate interest-based business models centered on loan products. Historically, information asymmetry granted them pricing power and negotiating leverage over customers seeking loans. However, as product-comparison websites have made cross-institution pricing transparent, this advantage has eroded, elevating the importance of sophisticated marketing strategies. Although interest-rate discounts are central to marketing loan products, heterogeneity in customer responses across discount levels necessitates a systematic strategy that accounts for this variation to maximize effectiveness.

Conventional response-prediction models estimate who is how likely to respond, but they are vulnerable to spurious correlations arising from targeted assignment and confounding. For example, because deeper interest-rate discounts are preferentially offered to hard-to-convert, lower-score applicants, the raw data may show a negative association between discount depth and acceptance. Causal inference addresses this limitation by modeling the discount as a treatment and adjusting for confounders to estimate borrower-specific treatment effects, enabling personalized policies. However, its application to real-world finance faces the following practical constraints: first, there are restrictions arising from the structural pricing policy. Interest-rate discounts are not offered randomly but are determined by risk-based pricing, in which a base rate is set according to a customer's credit score, followed by marketing adjustments to the discount level. As a result, upper and lower bounds on allowable discounts exist for each credit tier, making it essential to optimize under these constraints. Second, we face a domain-mismatch problem driven by data specificity. As lenders target specific customer segments, their portfolios concentrate in narrow interest-rate ranges and product mixes. Consequently, models trained on general market data mainly operate outside the common support of the target segment and yield biased estimates of segment-specific response to discounts. Third, there is a need to optimally allocate limited marketing budgets. Marketing resources are limited, and customer responses to interest-rate discounts exhibit high heterogeneity. Therefore, accurately predicting each customer's response to the marketing offer is critical.

This paper proposes a domain-adaptation system for marketing loan products that incorporates these domain-specific constraints while leveraging causal inference to estimate individual customers' response sensitivities. The proposed system: (1) extracts key variables through causal feature selection, (2) builds a domain-specialized prediction model using an S-learner with knowledge distillation, and (3) uses the estimated sensitivities to derive an optimal offer strategy within a fixed budget. To further enhance practitioner adoption, an interpretation layer with a large language model (LLM) translates model outputs into intuitive narratives, supporting data-driven strategy formulation by the decision maker.

The main contributions of this study are as follows:

1. Proposes a novel marketing system that integrates causal inference and knowledge distillation while accounting for financial domain constraints.
2. Validates the effectiveness and soundness of each system component (causal feature selection, S-learner, knowledge distillation).
3. Demonstrates the quantitative performance of the proposed strategy through real-world A/B testing.

Beyond improving the predictive accuracy of discount-response sensitivities, we highlight that our system demonstrates real-world impact on marketing strategy: it delivered significant business gains in a live A/B test. This paper details the system design, presents the empirical evaluation and deployment outcomes, and outlines directions for future work.

Related Works

Predicting customer responses for marketing optimization has largely relied on statistical machine learning. Supervised learners, such as logistic regression, Random Forests (Breiman 2001), and XGBoost (Chen and Guestrin 2016), model associations between input attributes and outcomes to predict response probabilities (Adeleye et al. 2024). However, because these models are associative, they do not identify the individual-level causal effect of a specific treatment (i.e., uplift for an interest-rate discount), limiting their usefulness for budget-constrained resource allocation (Rubin 2005).

To overcome this limitation, uplift modeling based on causal inference has recently gained significant attention. Uplift modeling aims to estimate the Conditional Average Treatment Effect (CATE), with representative approaches including the S/T/X/R-learner (Künzel et al. 2019; Nie and Wager 2021), Causal Forest (Wager and Athey 2018), as well as extensions to multiple treatments (Lechner 2018) and continuous treatments (Wan et al. 2022). In marketing scenarios where stronger treatment interventions are expected to yield higher response rates, research has also incorporated monotonicity constraints to improve the reliability of decisions (Zhou et al. 2024). However, applying these techniques to real-world finance still requires domain-specific adaptations that incorporate structural constraints, such as regulated pricing policies and the unique characteristics of financial data.

On the other hand, as neural networks have grown in depth and complexity, computational and memory demands have increased, driving research on transferring knowledge from large, high-performance models to smaller, deployment-efficient models. Hinton, Vinyals, and Dean (2015) introduced knowledge distillation, which is the process of transferring knowledge distilled from a teacher model to a student model. Subsequent work has identified the effects of knowledge distillation as label smoothing, transferring inter-class relationships, and gradient rescaling based on instance difficulty (Tang et al. 2020), while categorizing distilled knowledge into response-based, feature-based, and relation-based types (Gou et al. 2021). Beyond model compression, knowledge distillation serves as a generalization strategy through knowledge reconstruction and probability calibration, and can be beneficial even when teacher and student models share the same architecture (Furlanello et al. 2018). Romero et al. (2015) also demonstrated that transferring knowledge from a teacher model trained on a large general-domain dataset to a student model tailored to a specific domain improves generalization performance in data-constrained settings.

In summary, prior studies have highlighted the strengths of causal inference and knowledge distillation independently, yet addressing the CATE estimation problem in the financial domain requires an integrated approach that simultaneously accounts for complex domain constraints. Building on these insights, this study proposes a novel system that integrates causal inference and knowledge distillation within a domain-adaptation setting, offering a practical and innovative solution for CATE estimation in financial applications.

Proposed System

Our three-stage system estimates response sensitivity to discount offer and uses these estimates to optimize marketing strategies within real-world business constraints. The stages—causal feature selection, advanced model construction, and strategy optimization—are cohesively integrated, as illustrated in Figure 1.

Data Description

This study utilizes personal loan marketing campaign data from Hyundai Capital Services (HCS), the leading specialized credit financier of Korea and the captive finance company of Hyundai Motor Group, collected between November 2022 and October 2024. This two-year data ensures the model's ability to generalize across varying interest-rate levels driven by macroeconomic conditions. Except for the data from March to May 2024, which was reserved as an Out-of-Time (OOT) dataset, the remaining data was divided into training, validation, and test sets in a 6:2:2 ratio.

Initially, 564 candidate features were considered, encompassing customer credit, financial, and behavioral histories. This set was reduced to 303 features through a selection process based on statistical significance and business relevance. Finally, to ensure accurate estimation of causal effects, a causal inference method was applied to select approximately 183 final features for model development. Details of the feature selection procedure are presented in the following subsections, while Table 1 summarizes the key steps and results.

Stage 1. Causal Feature Selection

In this stage, we apply a two-step approach to select effective features for causal inference. First, potential causal relationships among variables are organized into a Directed Acyclic Graph (DAG) which visualizes causal dependencies and helps identify key confounders that may threaten the unconfoundedness assumption (Pearl 2009). This process reduces potential bias in the model. Second, we employ an Uplift Random Forest (Rzepakowski and Jaroszewicz 2012) to identify variables that directly contribute to heterogeneous treatment effects (HTE). The Uplift Random Forest modifies the splitting criterion of a standard Random Forest to maximize the divergence in outcome class distributions between treatment and control groups. Variables selected through this process are interpreted not as those that enhance predictive accuracy, but as those that most effectively differentiate treatment effects. Consequently, this two-step selection process ensures that the subsequent CATE estimation model is trained on meaningful and influential information.

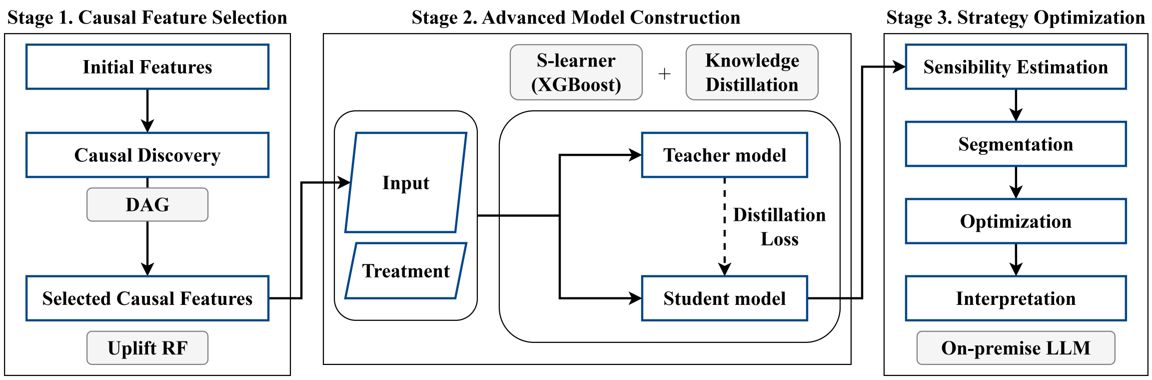


Figure 1: Overall architecture.

Step	# features	Selection Criteria
Total	564	- Internal Data (credit scores, loan history, macro index, price info, etc.)
1st	303	- Remove features with low Information Value (IV). - For highly correlated features, remove the one with the lower IV.
2nd	183	- Causal inference-based feature selection using Uplift Random Forest. - Identify features that best explain the difference in treatment effects, not just those with high predictive power.

Table 1: Feature selection steps.

Stage 2. Advanced Model Construction

S-learner. The core component of this study, the CATE estimation model, adopts the S-learner enhanced with knowledge distillation. In selecting the learner, we considered two key constraints of the business problem: (1) the ability to accommodate the *continuous* nature of interest-rate discounts, (2) compliance with the *monotonicity constraint*, which posits that customer response does not decline as the discount rate increases. A comparison of the structural characteristics of major uplift models, summarized in Table 2, confirmed that both the S-learner and the R-learner satisfy these requirements. Beyond these two constraints, however, we also considered *feasibility*, defined as the practical simplicity and interpretability of a learner in real financial applications. The S-learner, with its flexible and single-model structure, satisfies this criterion well. Based on these considerations, we chose the S-learner as our main approach. A performance comparison with the R-learner, which also meets the first two constraints, is presented later.

However, the standard S-learner has two primary limitations: (1) bias from confounders and (2) the risk of underestimating the treatment effect. We addressed these issues by constructing a DAG to identify and control confounders in advance, thereby reducing bias. In addition, the causal feature selection process prevents the model from overlooking treatment effects and improves the sensitivity of CATE estimation. Through these additional causal identification safeguards, our study overcomes the limitations of the S-learner and establishes a robust system that meets key business constraints.

Method	Continuous Treatment	Monotonicity	Feasibility
S-learner	Y	Y	Y
T-learner	N	N	Y
X-learner	N	N	N
R-learner	Y	Y	N
Causal forest (CF)	N	N	Y
Generalized CF	Y	N	N

Table 2: Key characteristics of meta-learners and causal forest-based methods.

Knowledge Distillation. Interest-rate bands vary across financial institutions due to differences in funding costs and risk management strategies. Banks with stable funding offer low-interest loans, whereas savings banks face higher funding costs and serve higher-risk customers, operating in higher-rate ranges. Our main competitors are peer institutions in sectors with similar interest-rate bands, which strongly affect our transaction volume. Accordingly, our primary target segment is customers choosing between HCS and competitor companies.

A model trained only on this segment, however, would ignore customer behavior in other sectors, while training on the entire market risks underfitting the target segment. To balance these trade-offs, we adopt a knowledge distillation (KD) approach.

Both the teacher and student models are S-learners with XGBoost as base estimator, but differ in training data and objectives. The teacher model is trained on the entire market using a multi-class setup, where the label indicates the sector of loan execution. Although customers may respond to multiple sectors, a single label is assigned according to a priority rule based on creditworthiness and accessibility. This allows the teacher model to capture general market re-

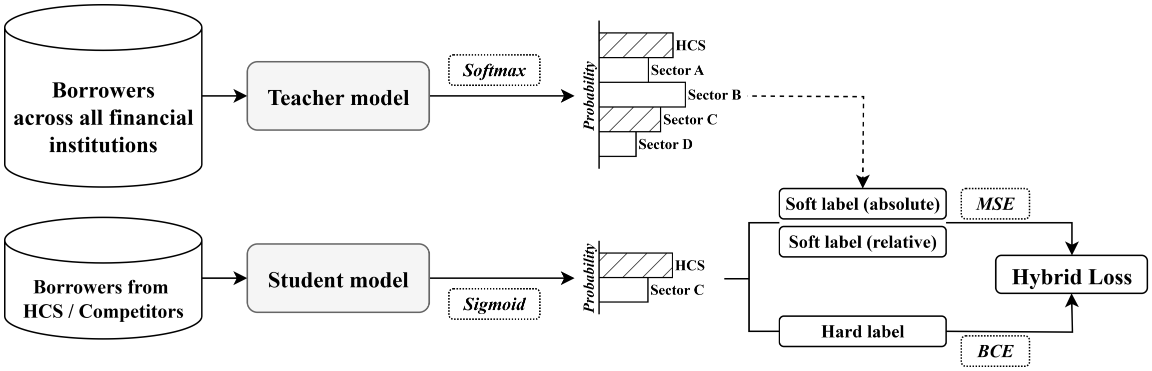


Figure 2: Overview of domain adaptation via knowledge distillation.

Attribute	Teacher model dataset	Student model dataset
Population	All loan records from the entire market	Loan records from HCS and competitor companies
Treatment	Offered interest-rate discount	Offered interest-rate discount
Prediction target	Loan execution by financial sector (multiclass)	Loan execution for our company (binary)
Target class	HCS, Sector A, Sector B, Sector C, Sector D	HCS, Sector C
Dataset size	1.9 M	1.3 M

Table 3: Dataset configuration for training the teacher and student models.

sponse patterns and sector selection tendencies.

The student model, in contrast, is trained as a binary classifier using only HCS and Sector C (a competitor sector with similar interest-rate bands, anonymized due to internal confidentiality). This design focuses the model on learning the precise decision boundary within the competitive space where interest-rate bands overlap. Because real-world operations involve customers with diverse characteristics, the student model incorporates the teacher model's soft labels during training.

This approach calibrates the student model's response estimates, enabling more sophisticated and generalized predictions across a broader range of customer profiles. Figure 2 illustrates the specific KD architecture, and Table 3 summarizes the dataset configurations. Again, finance sector names are anonymized for confidentiality.

The student model receives two soft labels from the teacher model's output: (1) the absolute probability that a customer will select our product in the market, and (2) the relative preference between our product and those from Sector C. The calculation of each soft label is as follows:

$$\tilde{y}_{\text{absolute}} = p_T^{\text{hcs}}$$

$$\tilde{y}_{\text{relative}} = \frac{\exp(z_{\text{hcs}})}{\exp(z_{\text{hcs}}) + \exp(z_{\text{c}})}$$

$$z_k = \log\left(\frac{p_T^k}{1 - p_T^k}\right)$$

where p_T^k denotes the softmax-based output probability of the teacher model, with $k \in \{\text{hcs, a, b, c, d}\}$ corresponding to HCS and Sectors A–D. The student model is trained using the two soft labels along with the actual binary response label. The hybrid loss function is defined as the weighted sum of three components:

$$\begin{aligned} L_{\text{hybrid}} &= (1 - \alpha_1 - \alpha_2) \cdot \text{BCE}(y, p_S) \\ &+ \alpha_1 \cdot \text{MSE}(\tilde{y}_{\text{absolute}}, p_S) \\ &+ \alpha_2 \cdot \text{MSE}(\tilde{y}_{\text{relative}}, p_S) \end{aligned}$$

where $y \in \{0, 1\}$ is the actual binary response label for our product, p_S is the sigmoid-based output probability of the student model, and $\alpha_1, \alpha_2 \in [0, 1]$ are hyperparameters controlling the relative importance of each soft label.

Stage 3. Strategy Optimization

The final stage of the system derives an optimal interest-rate discount strategy from the estimated customer's response sensitivity, aiming to maximize business impact while satisfying the operational constraints of the financial domain. This stage is carried out through four steps: (1) estimating response sensitivity, (2) segmenting customers, (3) performing constrained optimization, and (4) generating interpretations and strategic insights.

Estimation of Customer Response Sensitivity. Using the trained student model, we quantify the response sensitivity of each customer. Response sensitivity measures how a customer's response probability changes with variations in the discount level. While the discount is treated as a continuous variable for model learning, it is discretized into bands to enhance statistical stability and practitioner interpretability. Let $p_S(i, t)$ denote the predicted response probability of customer i at discount band t. The sensitivity score is defined as the average of marginal increase in the response rates:

$$RS(i) = \frac{1}{B-1} \sum_{t=0}^{B-1} \frac{p_S(i, t+1) - p_S(i, t)}{bp(t+1) - bp(t)}$$

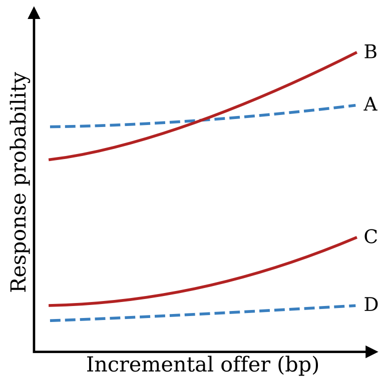


Figure 3: Example curves of customer-level response sensitivity to interest-rate discounts, illustrating four customer types: high-response & insensitive (A), high-response & sensitive (B), low-response & sensitive (C), and low-response & insensitive (D).

where B is the total number of bands and $bp(t)$ is the basis point discount of band t. A higher value implies that the customer is more sensitive to changes in the treatment.

Customer Segmentation. Based on the response sensitivity score, we segment the entire campaign population into strategic groups to prioritize marketing efforts. The conceptual customer types shown in Figure 3 are interpreted as follows: *Sure Things* — customers with $RS(i) \approx 0$ and a consistently high $p_S(i,t)$ across all bands (A); *Persuadables* — customers with a high $RS(i)$ (B and C); *Lost Causes* — customers with $RS(i) \approx 0$ and a consistently low $p_S(i,t)$ (D).

Constrained Optimization. The final campaign is determined by balancing transaction volume and portfolio profit, both of which vary with the sensitivity. This balance must be achieved while adhering to the upper and lower interest-rate limits for each credit tier and remaining within the allocated marketing budget. The campaign's objective can be tailored by the marketing department, for example, to maximize transaction volume or to maximize portfolio profit. A transaction volume–oriented strategy prioritizes customers with the highest sensitivity, whereas a profit-oriented strategy allocates discounts to those with the greatest incremental profit per unit of discount.

The graph in Figure 4 depicts the optimal curve of transaction volume versus portfolio profit and illustrates how the expected outcome point shifts under different strategic scenarios. This process is structured to persuade those responsible for marketers, including individuals without AI expertise, which is advantageous as it enables practical, operations-oriented campaign design. Furthermore, depending on their objectives, key corporate decision-makers can use these sensitivity scores to flexibly develop a range of strategic scenarios.

Interpretation and Strategic Insight Derivation. Finally, we include an interpretation step using an LLM to derive the business meaning of each segment beyond statistical characteristics. To ensure data security and maintain an independent analysis environment, a latest open-weight model was deployed on-premise. Structured inputs included quantitative information (demographics, key feature statistics) and qualitative information (anonymized cases) for each segment. Based on this input, the LLM generates (a) data-driven narrative summaries, (b) representative personas, and (c) strategic marketing directions. This process plays a key role in transforming model predictions into actionable insights that can be intuitively understood and applied by marketers.

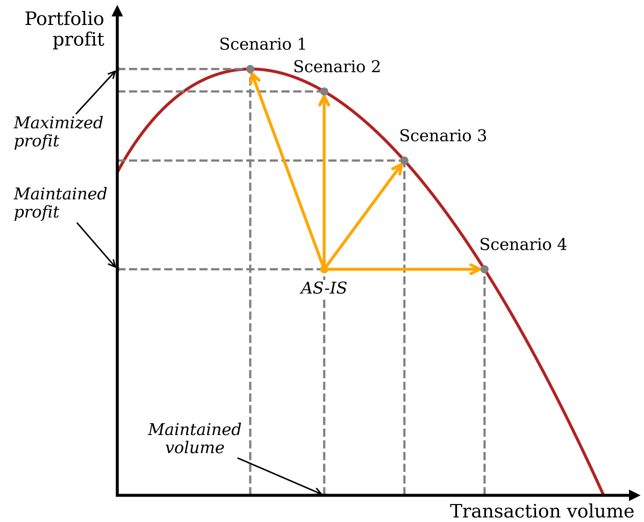
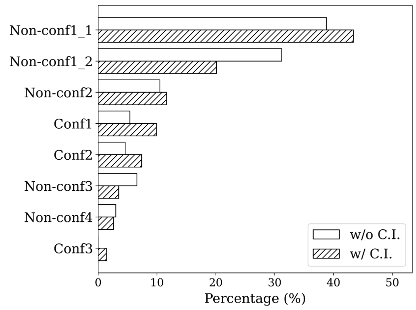


Figure 4: Expected transaction volume and portfolio profit under different scenarios. *AS-IS* marks the existing strategy; scenarios 1, 2, and 4 aim to maximize profit, maintain current volume while increasing profit, and maintain current profit while increasing volume, respectively; scenario 3 aims to increase both metrics.

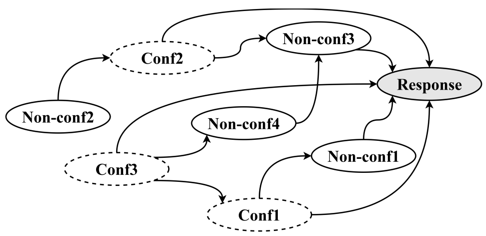
By integrating causal feature selection, domain-specialized prediction, and sensitivity-based strategy optimization, this system introduces a practical approach that considers strategic feasibility, moving beyond the prediction-centric limitations of traditional uplift modeling. The proposed system unifies the process from model development to strategy formulation and establishes a comprehensive decision-making system that connects data-driven insights to actual business execution. In the next section, we present experimental validation demonstrating how the system improves performance and profitability in a real-world marketing environment.

Experimental Analysis

In this section, we present quantitative and qualitative evaluations to validate the effectiveness of the key components of the proposed system. Specifically, we assess the efficacy of the feature selection process using the uplift random forest, evaluate the S-learner combined with the knowledge distillation approach, and conduct an in-depth analysis and interpretation of the results using an LLM.



(a) Comparison of feature importance with/without **C**ausal **I**nference (C.I.)



(b) Directed Acyclic Graph

Figure 5: Causal feature selection results. In accordance with the data security policy, confounders identified by the DAG are denoted as Conf1, Conf2, etc., while other variables are denoted as Non-conf1, Non-conf2, etc.

Effectiveness of Feature Selection Based on Causal Inference

To evaluate the effectiveness of the proposed causal feature selection, we compared the feature importance identified by a conventional predictive model and by the causal inference–based uplift model. Figure 5(a) highlights the difference between the two approaches. While the conventional model (w/o C.I.) primarily selected Non-conf variables that are directly correlated with the outcome, the uplift model (w/ C.I.) revealed that Conf variables, identified in advance through the DAG in Figure 5(b), serve as the key drivers of the sensitivity (treatment effect). These core confounders were mainly associated with macroeconomic indicators and market interest rates. This finding is significant because it goes beyond simple correlations and identifies the variables that answer the causal question, "Which factors maximize the effectiveness of a marketing strategy?"—a question that traditional predictive models cannot adequately address.

Monotonic Base-learner Comparison for S-learner

While simple, the S-learner readily exhibited robustness in predicting the customer response. In order to verify this, we compared it with a more sophisticated R-learner, also trained with monotonicity and continuous treatment conditions. For the sake of simplicity, KD was not applied; both models were trained and evaluated on the same dataset used for the student model. All three components of the R-learner (outcome, propensity, and final CATE models) used XGBoost as the base learner, consistent with the S-learner. Performance was measured by how closely the estimated average CATE at the credit tier × offer (bp) segment level matched the actual uplift in response rates, as evaluated by the Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Method	MSE	MAE
S-learner	**0.0194**	0.1165
R-learner	0.0198	**0.1109**

Table 4: Performance comparison between S-learner and R-learner.

According to Table 4, the S-learner showed negligible differences from the R-learner in MSE and MAE. While their performance was nearly equivalent, the S-learner provides practical advantages in real business settings, including simpler implementation and operation as well as easier interpretation of results.

Effectiveness of Knowledge Distillation

As described earlier, to mitigate domain mismatch and learn a sharper decision boundary, we employ KD: a generalist teacher model trained on market-wide data provides soft targets to a specialist student model trained on our proprietary in-domain data (our portfolio and customers of competing institutions). To evaluate its effectiveness, we compared five model configurations, varying the application of KD and the classification structure of the teacher model (multiclass vs. binary).

We used two performance metrics:

- AUC (Area Under the ROC curve): A standard metric for evaluating binary classification models. We report AUC separately for the entire campaign population (AUC1) and for the subset consisting of our company and customers of competing institutions (AUC2).
- STCE (Segment-level Treatment Calibration Error): A metric defined in this study to measure predictive consistency at the offer-unit (bp) level of practical importance. STCE is calculated as the weighted average of the absolute deviation between the mean predicted probability and the actual response rate at the credit tier × offer (bp) segment level, weighted by the number of customers in each segment:

$$\mathrm{STCE} = \frac{\sum_g n_g \cdot |\bar{p}_g - \bar{y}_g|}{\sum_g n_g}$$

where $\bar{p}_g$ and $\bar{y}_g$ are the mean predicted probability and actual response rate in segment g, respectively, and n_g is the number of customers in that segment.

The performance results of the comparison are summarized in Table 5. The student-only model achieved the highest performance on its own training population (AUC2 of 93.49) but also recorded the highest prediction error (STCE), raising concerns about stability and overfitting in real business applications. In contrast, models using

Model configuration	AUC1	AUC2	STCE
Without KD			
Direct — Student only	95.24	**93.49**	0.00162
Direct — Binary teacher	95.26	92.63	0.00140
Direct — Multiclass teacher	95.26	92.76	0.00134
With KD			
KD — Binary teacher	**95.30**	93.08	0.00078
KD — Multiclass teacher	95.29	93.17	**0.00072**

Table 5: Performance comparison of five model configurations with and without knowledge distillation (KD). "Direct" denotes models trained without KD, and "KD" denotes student models trained using KD from the specified teacher.

KD demonstrated improved stability, as indicated by reduced STCE. Notably, the student model with a Multiclass teacher achieved the lowest STCE while maintaining strong generalization performance on the entire campaign population (AUC1). This advantage arises because, unlike the Binary teacher, which provides fragmented information, the Multiclass teacher learns and transfers richer, more structured information about customers' financial behaviors, thereby enhancing the student model's generalization and stability. These results demonstrate that combining the teacher model's macro-level market insights with the student model's target-segment precision produces an improved model, balancing market-wide generalization, competitive-segment precision, and consistency in practical strategic units.

Interpretation Using LLM

Finally, to verify the interpretability and practicality of the system, we developed a process to comparatively analyze the profiles of the optimized customer segments using a large language model (LLM).

The LLM input consisted of summarized statistical features for 990,000 customers across four key segments classified by their sensitivity levels (e.g., *high-response & high-sensitivity*, *high-response & low-sensitivity*, etc.). To minimize hallucination and ensure consistent analysis, we provided structured information for each segment: (1) demographic distributions, (2) summary statistics (mean, median, and standard deviation, etc.) of the 30 features with the highest average SHAP values. In addition, anonymized individual cases were included as few-shot examples to help the LLM interpret the specific context. An example of the analysis results is presented in Table 6.

The significance of this process lies in its ability to precisely identify and analyze the *high-sensitivity* group, the primary target for concentrating limited resources. In particular, we identified potential persuadable customers (*low-response & high-sensitivity*) who exhibit low spontaneous response rates but react strongly to targeted marketing interventions, revealing opportunities for new market creation. While the combination of *low-response* and *high-sensitivity* may appear contradictory, a case identified by the model illustrates this pattern: customers in this segment were actively using loans from competitors with a large *estimated rate gap*. Although their likelihood of switching spontaneously is low—due to the absence of major dissatisfaction with their current product—they have the latent sensitivity to switch to our loan if offered a significantly lower interest rate. Our proposal can serve as the decisive trigger for activating this latent sensitivity.

Segment	Low-response & Low-sensitivity
Narrative summary	Customers with very high A and extensive experience with B-sector products. They tend to have low C and maintain stable D patterns.
Persona	*Indifferent & Stable Customer* — Shows low interest in our marketing messages or discount offers, with a low likelihood of behavioral change.
Strategic direction	Strengthen E products and formulate strategy: Develop targeted marketing approaches, such as promoting E-related services.

Table 6: Anonymized example of one segment interpretation generated by the LLM-based analysis.

Ultimately, this analysis enabled us to understand the customer characteristics and credit transaction patterns of each segment, confirm significant differences between them, and have a domain expert (marketer) validate the plausibility of the results. This approach is significant because it not only maximizes analytical efficiency by automating the previously manual task of model interpretation but also empowers marketers without AI expertise to design and execute sophisticated, data-driven strategies.

Application Use

Live Experiment Setting

To evaluate the business impact of the proposed sensibility-based optimization system, we conducted an A/B test over a one-month period in December 2024. The hypothesis was "By identifying response–sensitive customers and concentrating discount resources on them, the proposed model will achieve a greater increase in loan origination volume than the existing strategy, under the same budget constraints." Customers were randomly assigned to two groups, with the total discount budget controlled to be identical for both groups:

- Control Group: Applied the existing human-driven interest-rate discount strategy.
- Test Group: Applied the interest-rate discounts generated by the proposed system.

Results and Analysis

Table 7 summarizes the results of the A/B test. The experiment first classified customers into response–sensitive and response–insensitive groups using the model output. Within each group, only a subset of customers received an additional discount (Test), while the remainder did not (Control).

- **Response–sensitive group**: Providing additional discounts increased total loan origination volume by

	Sensitive group			Insensitive group		
	Control	Test	Gap	Control	Test	Gap
Base rate	12.80%	12.80%	0.0%p	12.60%	12.60%	0.0%p
Rate discount	0.60%	1.30%	0.7%p	0.50%	1.10%	0.6%p
Response rate	1.23%	1.36%	**0.14%p**	0.42%	0.41%	**-0.01%p**
Volume uplift	–	–	**16.80%**	–	–	**0.00%**

Table 7: A/B test results for response–sensitive and –insensitive groups, with metrics for Control and Test subgroups and their differences (Gap).

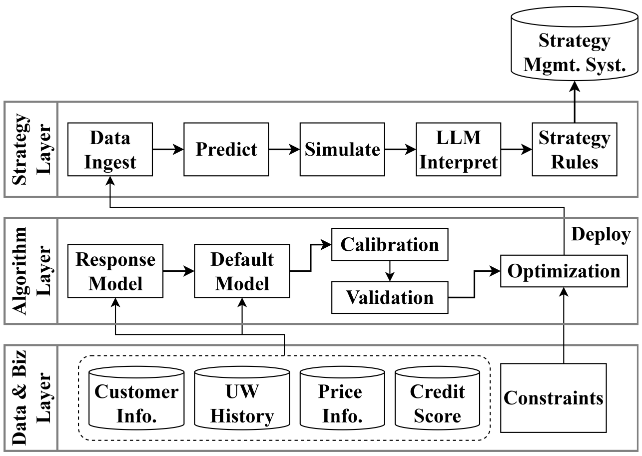


Figure 6: Operational Architecture

+16.8% compared to the control subgroup under the existing discount strategy, and the difference was statistically significant.

- **Response–insensitive group**: No meaningful change in loan origination volume was observed, even with additional discounts.

These results demonstrate that the proposed model successfully identified customer segments that are responsive to marketing interventions and adjust their behavior accordingly. Based on these findings, we deployed a full-scale optimization strategy focusing discounts exclusively on the sensitive group. As a result, total loan origination volume increased by approximately 13% (multi-billion KRW scale) compared to the previous human expert strategy. This confirms that the proposed system goes beyond simple response prediction and can lead to measurable revenue improvements through strategic optimization in real-world business environments.

Deployment and Maintenance

The proposed system has been deployed in a real business environment, as illustrated in Figure 6, and operates on a monthly automated batch basis via the company's inhouse MLOps platform. Using monthly updated customer and market data, the pre-trained model calculates the optimal interest rate for each customer, runs simulations, and finalizes the strategy. The resulting strategy is then automatically delivered to the marketing system (Strategy Management System) for execution in that month's campaign.

In addition, to maintain adaptability and predictive performance in an ever-changing market environment, we operate a systematic maintenance strategy. At the algorithm layer, the model is regularly re-trained, incorporating newly accumulated data. During this process, the influence of key variables sensitive to external market changes—such as interest rates—is re-examined and validated to ensure the model's up-to-dateness and robustness.

Conclusion

This paper presented a novel causal AI system for optimizing marketing resource allocation under the practical constraints of the credit finance domain. By combining causal feature selection with an S-learner–based knowledge distillation approach, the model precisely estimated customers' sensitivity to interest-rate discounts and achieved a significant business outcome—an increase of 13% in total loan origination volume. This achievement reflects the integrated contributions of each stage of the system. The causal feature selection process ensured that the model was built on genuine causal relationships, while knowledge distillation injected broad market intelligence to overcome the limitations of domain-specific data. Furthermore, the LLM–based interpretation layer enhanced the interpretability and practical applicability of the model. Finally, the value and scalability of the system have been demonstrated in the real-world business operation. The validated system has been successfully extended to various personal loan products at Hyundai Capital Services.

Future work will extend this system in multiple directions: we aim to evolve it into a multi-objective optimization structure that simultaneously considers multiple treatment variables, e.g. loan limits and distribution channels, in addition to interest rates, ultimately developing it into an integrated decision support system encompassing the company's overall marketing strategy operations. We also plan to relax the limitations of a batch-based, static optimization pipeline by developing an online learning layer. Specifically, for products requiring rapid responses to market dynamics or competitor strategies, we will investigate reinforcement learning methods that optimize discount policies in real time while jointly controlling origination growth and credit risk.

References

Adeleye, R. A.; Awonuga, K. F.; Asuzu, O. F.; Ndubuisi, N. L.; and Tubokirifuruar, T. S. 2024. Digital marketing analytics: A review of strategies in the age of big data and AI. *World Journal of Advanced Research and Reviews*, 21(2): 073–084.

Breiman, L. 2001. Random forests. *Machine learning*, 45(1): 5–32.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Furlanello, T.; Lipton, Z.; Tschannen, M.; Itti, L.; and Anandkumar, A. 2018. Born again neural networks. In *International conference on machine learning*, 1607–1616. PMLR.

Gou, J.; Yu, B.; Maybank, S. J.; and Tao, D. 2021. Knowledge distillation: A survey. *International journal of computer vision*, 129(6): 1789–1819.

Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Künzel, S. R.; Sekhon, J. S.; Bickel, P. J.; and Yu, B. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10): 4156–4165.

Lechner, M. 2018. Modified causal forests for estimating heterogeneous causal effects. *arXiv preprint arXiv:1812.09487*.

Nie, X.; and Wager, S. 2021. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2): 299–319.

Pearl, J. 2009. *Causality*. Cambridge university press.

Romero, A.; Ballas, N.; Kahou, S. E.; Chassang, A.; Gatta, C.; and Bengio, Y. 2015. FitNets: Hints for Thin Deep Nets. arXiv:1412.6550.

Rubin, D. B. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American statistical Association*, 100(469): 322–331.

Rzepakowski, P.; and Jaroszewicz, S. 2012. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2): 303–327.

Tang, J.; Shivanna, R.; Zhao, Z.; Lin, D.; Singh, A.; Chi, E. H.; and Jain, S. 2020. Understanding and improving knowledge distillation. *arXiv preprint arXiv:2002.03532*.

Wager, S.; and Athey, S. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

Wan, S.; Zheng, C.; Sun, Z.; Xu, M.; Yang, X.; Zhu, H.; and Guo, J. 2022. GCF: Generalized Causal Forest for Heterogeneous Treatment Effect Estimation in Online Marketplace. arXiv:2203.10975.

Zhou, H.; Huang, R.; Li, S.; Jiang, G.; Zheng, J.; Cheng, B.; and Lin, W. 2024. Decision focused causal learning for direct counterfactual marketing optimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 6368–6379.