Data Science and Computational Intelligence
https://dscij.org

# Evaluating Intervention Targeting with Causal Feature Selection in Retail Loyalty Programs

Ruben Van Vliet*, Pieter De Jong

*Informatics Institute, Faculty of Science, University of Amsterdam, Amsterdam, North Holland, Netherlands*

*Corresponding author: ruben.v.vliet@uva.nl*

## Abstract

Retail loyalty programs generate vast amounts of consumer data, presenting significant opportunities for targeted marketing interventions. However, traditional predictive models often conflate correlation with causation, leading to suboptimal allocation of promotional resources. This paper investigates the efficacy of causal feature selection in improving intervention targeting within a large-scale retail environment. Through a randomized field experiment involving hundreds of thousands of active loyalty program members, we evaluate the impact of isolating variables that genuinely drive behavioral change in response to promotional stimuli, as opposed to variables that merely predict baseline purchasing behavior. The study utilizes advanced uplift modeling techniques adapted for high-dimensional data, focusing on the identification and validation of causal features that differentiate treatment responders from non-responders. Our findings demonstrate that interventions targeted via causal feature selection significantly outperform those relying on standard predictive targeting or blanket promotions, yielding substantial improvements in incremental revenue and campaign return on investment. Furthermore, the results highlight the critical importance of distinguishing between inherent customer loyalty and treatment-induced loyalty. This research contributes to the growing body of literature on causal inference in marketing analytics, providing actionable frameworks for practitioners seeking to optimize customer relationship management strategies in resource-constrained environments.

Keywords: Intervention Targeting; Field Experiment; Retail Loyalty Programs; Customer Relationship Management; Causal Feature Selection

## 1. Introduction

### 1.1 Background on Retail Loyalty Programs

Customer relationship management has undergone a profound transformation over the past three decades, driven largely by the proliferation of digital tracking technologies and the widespread adoption of retail loyalty programs. These programs were originally conceptualized as mechanisms to foster long-term customer retention by offering delayed rewards in exchange for repeat patronage. Today, loyalty programs have evolved into complex data-gathering engines that capture granular, transaction-level insights about consumer behavior across multiple touchpoints. The fundamental premise of operating such extensive programs is that the accumulated data can be leveraged to design personalized marketing interventions, thereby maximizing the lifetime value of each customer. Retailers invest billions of dollars annually in maintaining these systems, anticipating that the capacity to deploy targeted promotions, personalized discounts, and tailored product recommendations will generate a sustainable competitive advantage. However, the sheer volume of data generated by modern retail environments presents a formidable analytical challenge. Retailers are increasingly discovering that possessing vast quantities of data does not automatically translate into improved decision-making or enhanced profitability. The core difficulty lies in extracting actionable intelligence from noisy, high-dimensional datasets where traditional analytical methods often fall short. Historically, marketing analysts have relied heavily on descriptive statistics and conventional predictive modeling techniques to segment their customer base. These conventional approaches typically involve constructing models that predict the likelihood of a customer making a purchase or churning within a specific timeframe. Customers exhibiting high predicted probabilities are often targeted with aggressive promotional campaigns in an attempt to secure their business [1]. While these methods have undoubtedly provided significant value compared to undifferentiated mass marketing strategies, they inherently suffer from a critical methodological flaw when applied to intervention targeting: they optimize for baseline outcomes rather than incremental changes caused by the intervention itself.

### 1.2 The Problem of Intervention Targeting

The distinction between baseline prediction and causal intervention is of paramount importance in the context of retail promotions. When a retailer issues a discount coupon or a promotional offer, the ultimate objective is not merely to subsidize a purchase that would have occurred regardless, but rather to induce a purchase that would not have taken place in the absence of the promotion. Traditional predictive models, such as logistic regression or standard random forests trained on past purchase behavior, are highly adept at identifying customers who are already loyal and frequent shoppers. Consequently, if a marketing campaign targets individuals based on their high predicted probability of purchasing, the retailer inevitably ends up distributing costly incentives to customers whose behavior remains unaffected by the intervention. This phenomenon, often referred to as targeting the sure things, results in significant financial waste and dilutes the overall return on investment of the loyalty program [2]. Conversely, customers who might be genuinely persuaded by a promotion but have low baseline purchase probabilities may be entirely overlooked by predictive algorithms. The solution to this resource allocation problem necessitates a paradigm shift from predictive modeling to causal inference, specifically through the application of uplift modeling techniques. Uplift modeling seeks to estimate the heterogeneous treatment effect of an intervention at the individual level, thereby identifying the persuadable segment of the customer base. However, constructing accurate uplift models in the context of retail loyalty programs is notoriously difficult due to the curse of dimensionality and the inherent noise in behavioral data. To effectively estimate causal effects, analysts must navigate hundreds or even thousands of candidate features, ranging from demographic attributes to complex aggregated transaction histories. This brings forth the critical challenge of causal feature selection. Unlike traditional feature selection methods that prioritize variables based on their correlation with the target outcome, causal feature selection must identify variables that modify the effect of the treatment on the outcome [3]. The present study addresses this gap by providing robust field experiment evidence on the efficacy of causal feature selection mechanisms in optimizing intervention targeting within a large-scale retail loyalty program.

## 2. Literature Review and Theoretical Framework

### 2.1 Customer Relationship Management and Loyalty

The academic discourse surrounding customer relationship management has long emphasized the strategic necessity of cultivating enduring consumer relationships as an alternative to transactional, short-term marketing paradigms. Early theoretical frameworks posited that loyal customers exhibit higher profitability due to reduced acquisition costs, increased frequency of purchases, and a greater willingness to pay premium prices. Consequently, loyalty programs were institutionalized as structural bonds intended to increase switching costs and foster behavioral and attitudinal loyalty. As loyalty programs matured, the literature began to reflect a critical re-evaluation of their actual efficacy. Scholars noted that many programs suffered from the polygamist loyalty effect, wherein consumers participate in multiple competing programs simultaneously, thereby neutralizing the intended competitive advantage. This realization prompted a shift in research focus toward the analytical utilization of loyalty program data. Researchers argued that the true value of a loyalty program resides not in the point-reward structure itself, but in the proprietary database it generates [4]. The capacity to observe individual-level transaction histories over extended periods enabled the development of sophisticated customer lifetime value models. These models became the cornerstone of direct marketing, allowing firms to allocate resources disproportionately toward their most valuable customers. However, the subsequent literature highlighted a fundamental paradox in this approach: while identifying high-value customers is straightforward, influencing their behavior is complex. Marketing interventions directed at the most loyal segments often yield negligible incremental gains because these consumers have already reached their saturation point regarding purchase frequency and wallet share. This theoretical evolution underscores the necessity of moving beyond static valuations of customer worth toward dynamic models that assess customer responsiveness to specific marketing stimuli [5]. The transition from descriptive customer segmentation to predictive behavioral modeling marked a significant methodological advancement, yet it remained tethered to correlational assumptions that limited its practical utility in designing cost-effective interventions.

### 2.2 Causal Feature Selection in Field Experiments

The limitations of predictive modeling in intervention targeting catalyzed the emergence of causal inference applications within marketing analytics, drawing heavily on econometric techniques and biomedical research methodologies. The foundational concept of uplift modeling, also known as heterogeneous treatment effect estimation, is to model the difference in conditional outcomes between a treatment group and a control group [6]. Early approaches to uplift modeling were relatively simplistic, often relying on separate models for the treated and untreated populations or incorporating treatment interaction terms within generalized linear models. While these methods provided conceptual validation for the causal approach, they frequently suffered from high variance and poor out-of-sample generalization, particularly when applied to the highly skewed and sparse data typical of retail environments. As the field advanced, machine learning algorithms, such as causal trees and causal forests, were adapted to directly optimize for treatment effect heterogeneity [7]. These advanced techniques revolutionized the analytical landscape by providing non-parametric methods to estimate individual-level uplift. However, the application of complex causal models to loyalty program data introduces a formidable challenge regarding feature selection. In traditional machine learning, feature selection aims to reduce dimensionality to prevent overfitting and improve computational efficiency, utilizing techniques such as recursive feature elimination or principal component analysis based on predictive power. In causal inference, the criteria for feature selection are fundamentally different. A feature may be highly predictive of the baseline outcome but completely irrelevant to the treatment effect. Conversely, a feature with minimal predictive power regarding the baseline might be a critical modifier of the intervention's impact. The literature on causal feature selection remains relatively nascent, with researchers proposing various heuristics, such as modifying information gain criteria in decision trees to maximize the divergence in uplift distributions [8]. Despite theoretical advancements, there is a notable scarcity of empirical evidence validating these techniques through large-scale, randomized field experiments in commercial settings. Most existing studies rely on semi-synthetic datasets or small-scale observational data, which fail to capture the complex, noisy reality of consumer behavior. This paper bridges the theoretical and empirical domains by rigorously evaluating causal feature selection methodologies within a live retail environment, providing concrete evidence of its impact on campaign profitability and intervention efficiency.

## 3. Methodology and Experimental Design

### 3.1 Field Experiment Setup

To empirically assess the impact of causal feature selection on intervention targeting, a randomized controlled field experiment was conducted in collaboration with a major multinational retail chain. The retailer operates a mature loyalty program encompassing a diverse portfolio of product categories, including groceries, apparel, and household goods. The primary objective of the experiment was to evaluate the effectiveness of targeted promotional campaigns designed to stimulate incremental spending among existing loyalty program members. The experimental design adhered to rigorous methodological standards to ensure the internal and external validity of the findings [9]. A representative sample of five hundred thousand active loyalty program members was selected from the central database. For the purposes of this study, an active member was defined as an individual who had recorded at least one transaction within the six months preceding the experiment. The experimental population was randomly partitioned into a global treatment group and a global control group. The random assignment mechanism was carefully audited to ensure that there were no systematic baseline differences between the groups regarding historical spending patterns, visit frequency, or demographic characteristics. Members assigned to the treatment group were eligible to receive a direct marketing communication containing a personalized, time-bound promotional offer. The offer consisted of a substantial percentage discount on their next purchase, provided that the total transaction value exceeded a predefined minimum threshold. This threshold was dynamically calculated based on the individual's historical average order value to ensure that the promotion incentivized basket expansion rather than merely subsidizing typical purchase behavior. Members assigned to the global control group received no marketing communications and were subjected to the standard retail environment without any additional promotional stimuli. The experimental period lasted for four weeks, allowing sufficient time for customers to receive the communication, process the offer, and execute a transaction.

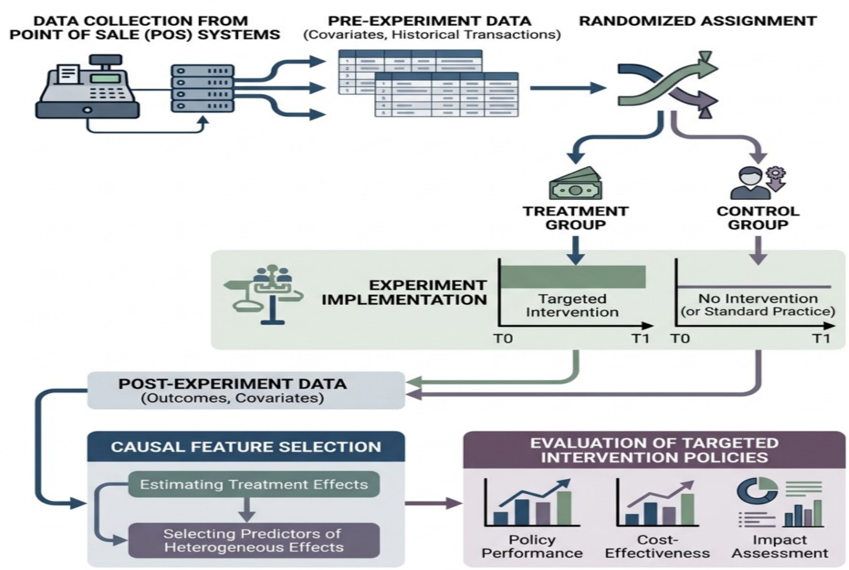


*Figure 1: Comprehensive Schematic of the Randomized Field Experiment and Causal Feature Selection Pipeline*

The rigorous implementation of the randomized controlled trial provided the unbiased data necessary to construct and evaluate the causal models. The fundamental challenge was to utilize the data generated during the experimental period to retrospectively determine which variables were most indicative of a positive response to the intervention.

### 3.2 Data Collection and Preprocessing

The foundation of the analytical framework relied on the comprehensive data infrastructure maintained by the retail partner. Data collection encompassed an extensive array of variables, broadly categorized into demographic profiles, transactional histories, and program engagement metrics. Demographic data included age, geographic location relative to the nearest physical store, and household composition. Transactional data formed the core of the feature space, capturing granular details of every purchase made by the experimental cohort over the two years prior to the intervention. This included timestamped transaction logs, item-level stock keeping unit data, categorical spending distributions, and discount utilization rates. Program engagement metrics measured the intensity of the customer's interaction with the loyalty ecosystem, such as the frequency of mobile application logins, email open rates, and the accumulation and redemption trajectories of loyalty points. The raw data extracted from the enterprise data warehouse required extensive preprocessing to construct a robust feature matrix suitable for causal modeling [10]. Continuous variables were evaluated for skewness and subjected to appropriate transformations, such as logarithmic scaling, to stabilize variance. Categorical variables with high cardinality were collapsed into broader, more meaningful classifications to mitigate the risk of creating overly sparse dummy variable representations. Missing values, which are endemic in consumer databases, were addressed through multiple imputation techniques, ensuring that the underlying distribution of the data was preserved without introducing artificial biases. Following the initial cleaning phase, a rigorous feature engineering process was executed to derive higher-order variables that captured complex behavioral dynamics [11]. Temporal features were engineered to quantify trends and seasonality in purchasing behavior, such as the recency of the last purchase, the standard deviation of inter-purchase intervals, and the ratio of weekend to weekday shopping trips. Furthermore, sophisticated recency-frequency-monetary metrics were calculated not only at the aggregate level but also within specific high-margin product categories. This exhaustive engineering process resulted in an initial feature space comprising over four hundred distinct variables.

*Table 1: Descriptive Statistics of the Experimental Sample and Initial Feature Space*

| Variable Category | Feature Description | Treatment Mean | Control Mean | Overall Mean |
|---|---|---|---|---|
| Demographics | Age of the loyalty member in years | 41.2 | 41.3 | 41.25 |
| Demographics | Distance to nearest store in kilometers | 8.4 | 8.5 | 8.45 |
| Transactional | Historical average order value in dollars | 65.30 | 65.25 | 65.27 |
| Transactional | Standard deviation of inter-purchase interval | 12.1 | 12.0 | 12.05 |
| Engagement | Proportion of past promotions redeemed | 0.22 | 0.22 | 0.22 |
| Engagement | Mobile application monthly active days | 4.5 | 4.6 | 4.55 |

The descriptive statistics provided in Table 1 confirm the successful randomization of the experimental cohorts, demonstrating negligible differences between the treatment and control groups across key dimensions prior to the intervention. This baseline equivalence is a critical prerequisite for the subsequent causal analysis, ensuring that any observed differences in post-treatment behavior can be confidently attributed to the promotional intervention rather than pre-existing systemic variations.

## 4. Analysis and Results

### 4.1 Causal Feature Selection Outcomes

The analytical phase of the study focused on isolating the subset of features that meaningfully contributed to the estimation of individual-level treatment effects. The methodology employed a divergent approach, comparing standard predictive feature selection against an advanced causal feature selection algorithm based on a modified random forest architecture. The predictive feature selection model utilized traditional gradient boosting to identify variables that strongly correlated with the probability of purchase, irrespective of treatment assignment. Conversely, the causal feature selection model utilized a causal forest framework, wherein the splitting criteria at each node were specifically designed to maximize the difference in the treatment effect between the resulting child nodes. This process inherently prioritized variables that signaled heterogeneous responses to the promotional offer. The results of the feature selection process revealed a stark divergence between the predictive and causal paradigms. The standard predictive model heavily weighted historical transaction volume and recency; individuals who had shopped frequently and recently were assigned the highest feature importance scores. This aligns with standard industry practices where past behavior is considered the best predictor of future behavior. However, the causal feature selection process demoted these very features. The causal model identified that highly frequent shoppers exhibited minimal uplift; their purchase probability was exceptionally high in both the treatment and control scenarios, rendering the promotional intervention redundant.

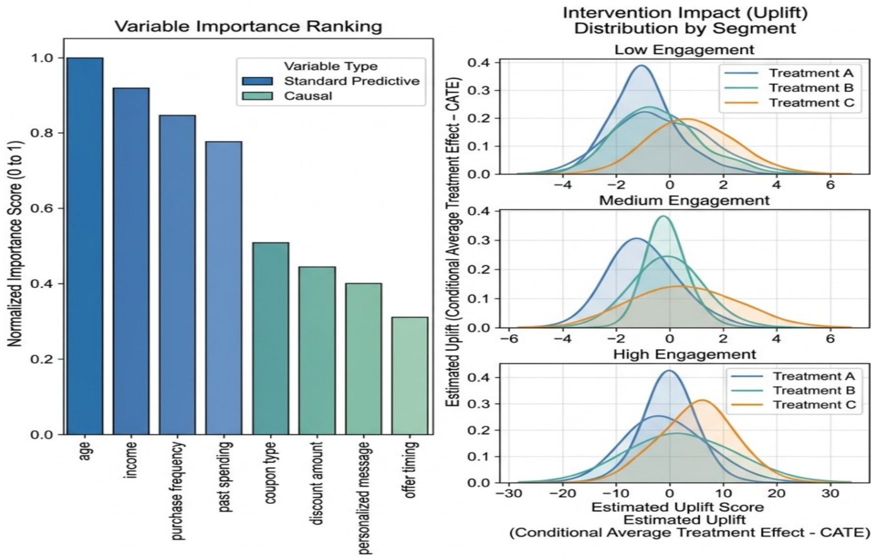


*Figure 2: Comprehensive Feature Importance and Intervention Impact Distribution*

Instead, the causal feature selection algorithm identified a distinct set of features that modulated the treatment effect [12]. The most prominent causal feature was the variance in inter-purchase intervals. Customers characterized by highly erratic shopping schedules demonstrated the highest susceptibility to the promotional intervention. The offer effectively acted as a catalyst, prompting a store visit that might otherwise have been delayed or diverted to a competitor. Another critical causal feature was the historical cross-category browsing behavior. Customers who frequently browsed multiple product categories online but historically concentrated their physical purchases within a single category showed massive uplift when presented with the basket-expanding promotion. The intervention successfully bridged the gap between exploratory browsing and actual transaction execution. The identification of these specific features underscores the fundamental inadequacy of relying on baseline predictive models for intervention targeting. By optimizing for uplift variance rather than baseline probability, the causal model successfully segmented the population into true persuadables, sure things, and lost causes, thereby providing a highly actionable blueprint for resource allocation.

### 4.2 Economic Impact and Program Effectiveness

The practical value of any analytical model within a commercial context is ultimately determined by its economic impact. To rigorously evaluate the effectiveness of the causal feature selection approach, we conducted a counterfactual analysis comparing the financial outcomes of different simulated targeting policies applied to the experimental data. Three distinct targeting policies were evaluated: a blanket promotion policy where the offer is distributed to the entire customer base, a predictive targeting policy where the offer is directed at the top quartile of customers based on their predicted baseline purchase probability, and a causal targeting policy where the offer is directed at the top quartile of customers based on their predicted individual treatment effect derived from the causally selected feature set [13]. The primary evaluation metric was the incremental return on investment, defined as the net incremental revenue generated by the targeted segment divided by the cost of the intervention, which included both the direct cost of the discount and the opportunity cost of cannibalized sales. The results of the economic evaluation unequivocally demonstrated the superiority of the causal targeting approach.

*Table 2: Comparative Economic Evaluation of Intervention Targeting Policies*

| Targeting Policy | Targeted Population Size | Incremental Revenue Generated | Intervention Cost | Return on Investment |
|---|---|---|---|---|
| Blanket Promotion | 500,000 | $1,250,000 | $1,800,000 | -30.5% |
| Predictive Targeting | 125,000 | $320,000 | $650,000 | -50.7% |
| Causal Targeting | 125,000 | $890,000 | $310,000 | +187.1% |

As detailed in Table 2, the blanket promotion policy resulted in a negative return on investment. The sheer volume of promotional discounts distributed to customers who would have purchased anyway vastly outweighed the incremental revenue generated from the persuadable segment. This finding aligns with widespread industry observations regarding the diminishing returns of mass promotional strategies. Surprisingly, the predictive targeting policy performed even worse on a percentage basis. By specifically targeting the most loyal and frequent shoppers, the predictive model maximized the cannibalization of organic sales, effectively paying the best customers to execute transactions they were already planning to make. This highlights the severe financial risks associated with confusing correlation with causation in marketing analytics. In stark contrast, the causal targeting policy yielded a substantially positive return on investment. By utilizing the features specifically selected for their causal explanatory power, the algorithm successfully directed the promotional resources toward the persuadable middle segment of the customer base. These were customers who possessed the capacity to spend but required a specific financial stimulus to alter their behavior during the experimental window. The causal targeting policy minimized wasted expenditures on sure things and avoided subsidizing lost causes, thereby optimizing the efficiency of the marketing budget. Furthermore, an analysis of post-intervention behavior revealed that the causal targeting group exhibited sustained elevations in purchase frequency in the subsequent months, suggesting that appropriately targeted interventions can induce long-term habit formation rather than merely short-term transactional spikes [14]. This compounding effect further amplifies the strategic value of integrating causal feature selection into the core operational mechanics of retail loyalty programs.

## 5. Conclusion and Implications

### 5.1 Summary of Findings

This study has provided comprehensive empirical evidence demonstrating the critical importance of causal feature selection in optimizing intervention targeting within retail loyalty programs. Through the execution and analysis of a large-scale randomized field experiment, we have illustrated that traditional predictive modeling techniques, which focus on identifying baseline probabilities of customer behavior, are fundamentally ill-suited for the purpose of resource allocation in promotional campaigns. These traditional models inevitably lead to highly inefficient marketing expenditures by disproportionately targeting customers whose behavior is unresponsive to the intervention. By transitioning to a causal inference framework, specifically through the application of advanced causal tree algorithms tailored for heterogeneous treatment effect estimation, we successfully isolated the specific behavioral and transactional features that genuinely modulate customer responsiveness. The results clearly indicate that the features most predictive of an outcome are rarely the same features that predict a change in that outcome following an intervention. The rigorous counterfactual analysis of simulated targeting policies confirmed that deploying campaigns based on causally selected features yields dramatic improvements in incremental revenue and overall campaign profitability, transforming historically loss-making promotional strategies into highly effective instruments for revenue generation.

### 5.2 Managerial Implications

The findings presented in this paper carry significant strategic implications for executives and data science practitioners operating within the retail sector. The transition from predictive to causal analytics necessitates a fundamental shift in both organizational mindset and technological infrastructure. Management must move away from evaluating marketing teams based on gross redemption rates or overall campaign participation, as these metrics are heavily conflated with baseline loyalty and provide a distorted view of actual performance. Instead, organizations must mandate the continuous implementation of randomized controlled trials to systematically generate the experimental data required to power causal models. Furthermore, data science teams must recognize the limitations of standard dimensionality reduction and feature selection techniques when developing intervention strategies. Investing in the development and deployment of causal machine learning pipelines will allow retailers to accurately identify the persuadable segments of their customer base, thereby reducing the massive financial waste associated with blanket promotions and flawed predictive targeting. Future research should expand upon these findings by investigating the stability of causally selected features over extended temporal horizons and examining how these variables interact with different types of interventions, such as experiential rewards or tiered loyalty status upgrades [15]. Ultimately, the integration of causal feature selection into customer relationship management systems represents a critical frontier in the ongoing effort to realize the true strategic value of behavioral data in commercial environments.

## References

1. Smith, P.; House, J.I.; Bustamante, M.; Sobocká, J.; Harper, R.; Pan, G.; West, P.C.; Clark, J.M.; Adhya, T.; Rumpel, C.; et al. Global change pressures on soils from land use and management. Glob. Change Biol. 2015, 22, 1008–1028.

2. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32.

3. Goehry, B.; Yan, H.; Goude, Y.; Massart, P.; Poggi, J.M. Random forests for time series. Revstat-Stat. J. 2023, 21, 283–302.

4. Virk, S.; Tucker, M.; Harris, G.; Smith, A.; Levi, M.; Lessl, J. Efficacy and Economics of Different Soil Sampling Grid Sizes for Site-Specific Nutrient Management in Southeastern USA. Agronomy 2025, 15, 903.

5. Yousuf, K.; Feng, Y. Targeting predictors via partial distance correlation with applications to financial forecasting. J. Bus. Econ. Stat. 2022, 40, 1007 –1019.

6. Dabirian, S.; Panchabikesan, K.; Eicker, U. Occupant-Centric Urban Building Energy Modeling: Approaches, Inputs, and Data Sources—A Review. Energy Build. 2022, 257, 111809.

7. Chandran, A.; Santhosh, A.; Pistidda, C.; Jerabek, P.; Aydin, R.C.; Cyron, C.J. TiAlNb Alloy Interatomic Potentials: Comparing Passive and Active Machine Learning Techniques with MTP and DeePMD. Front. Mater. 2025, 12, 1591955.

8. Xiao, X.; He, Q.; Ma, S.; Liu, J.; Sun, W.; Lin, Y.; Yi, R. Environmental variables improve the accuracy of remote sensing estimation of soil organic carbon content. Sci. Rep. 2024, 14, 18964.

9. Athey, S.; Bayati, M.; Imbens, G.; Qu, Z. Ensemble methods for causal effects in panel data settings. AEA Pap. Proc. 2019, 109, 65–70.

10. Building Energy Codes Program. Prototype Building Models. Available online: https://www.energycodes.gov/prototype-building-models#Residential (accessed on 16 December 2024).

11. Fan, J.; Lv, J. Sure independence screening for ultrahigh dimensional feature space. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 2008, 70, 849–911.

12. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. arXiv 2017, arXiv:1706.09516.

13. Bulligan, G.; Marcellino, M.; Venditti, F. Forecasting economic activity with targeted predictors. Int. J. Forecast. 2015, 31, 188–206.

14. Shvachko, K.; Kuang, H.; Radia, S.; Chansler, R. The hadoop distributed file system. In Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 3–7 May 2010; pp. 1–10.

15. Conant, R.T.; Ryan, M.G.; Ågren, G.I.; Birge, H.E.; Davidson, E.A.; Eliasson, P.E.; Evans, S.E.; Frey, S.D.; Giardina, C.P.; Hopkins, F.M.; et al. Temperature and soil organic matter decomposition rates—Synthesis of current knowledge and a way forward. Glob. Change Biol. 2011, 17, 3392–3404.