Pontifícia Universidade Católica
do Rio de Janeiro

Leonardo Domingues

A non-parametric probabilistic counterfactual approach to assess a retailer's transactional potential

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia de Produção.

Advisor : Prof. Davi Michel Valladão
Co-advisor: Prof. Alexandre Street de Aguiar

Rio de Janeiro
July 2022

Pontifícia Universidade Católica
do Rio de Janeiro

Leonardo Domingues

A non-parametric probabilistic counterfactual approach to assess a retailer's transactional potential

Dissertation presented to the Programa de Pós–graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia de Produção. Approved by the Examination Committee:

Prof. Davi Michel Valladão
Advisor
Departamento de Engenharia Industrial – PUC-Rio

Prof. Alexandre Street de Aguiar
Co-advisor
Departamento de Engenharia Industrial – PUC-Rio

Prof. Paula Medina Maçaira Louro
Departamento de Engenharia Industrial – PUC-Rio

Prof. Cristiano Augusto Coelho Fernandes
Departamento de Engenharia Elétrica – PUC-Rio

Rio de Janeiro, July 21st, 2022

Leonardo Domingues

Leonardo Domingues currently is a master student candidate of Industrial Engineering Departament at PUC-Rio. His research focus on Revenue Management problems. He undergraduated in Chemical Engineering at Federal University of Alagoas in 2019.

Bibliographic data

Domingues, Leonardo

A non-parametric probabilistic counterfactual approach to assess a retailer's transactional potential / Leonardo Domingues; advisor: Davi Michel Valladão; co-advisor: Alexandre Street de Aguiar. – 2022.

45 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2022.

Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Contrafactual. 3. Churn não-contratual. 4. Gerenciamento de Receita. 5. Adquirente. I. Valladão, Davi. II. Aguiar, Alexandre Street de. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. IV. Título.

CDD: 658.5

Acknowledgments

Thank to my advisers Prof. Davi Michel Valladão and Prof. Alexandre Street de Aguiar for the stimulus and partnership to carry out this work.

To my parents Marta and Fernando, for their unconditional support during all this journey.

To my family and friends who supported me even with my absence in family life.

To my fiance Camilla, who accompanied me all this time with direct and indirect support.

To all colleagues, faculty and staff of the Department of PUC-Rio, for the fellowship, learning and support.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Domingues, Leonardo; Valladão, Davi (Advisor); Aguiar, Alexandre Street de (Co-Advisor). **A non-parametric probabilistic counterfactual approach to assess a retailer's transactional potential**. Rio de Janeiro, 2022. 45p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

In the payment industry context, a merchant acquirer is a firm that facilitates communication between a retailer (online or brick–and–mortar store) and the issuing banks. For an acquirer, it is crucial to determine the transactional potential of each retailer to guide proper pricing and risk management strategies. In this work, we propose a framework to properly assess the transactional potential of any retailer using the transactions of its peers. The proposed framework is based on the construction of a probabilistic counterfactual that uses non-parametric Nadaraya-Watson kernel regression to model differing seasonal patterns, trends and business cycles. We propose an integrated data processing methodology to separate and validate the data not affected by interventions to construct our non-parametric probabilistic counterfactual model. The proposed framework is a powerful decision support system for a merchant acquirer revenue management, with direct applications to pricing, churn detection and, more generally, revenue management. Empirical results corroborate the effectiveness of the method against relevant benchmarks.

Keywords

Counterfactual; Non-contractual Churn; Revenue Management; Acquirer.

Resumo

Domingues, Leonardo; Valladão, Davi; Aguiar, Alexandre Street de. **Uma abordagem contrafactual probabilística não paramétrica para avaliar o potencial transacional de um varejista**. Rio de Janeiro, 2022. 45p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

No contexto da indústria de adquirência, uma adquirente é uma empresa que facilita a comunicação entre um varejista (online ou loja física) e os bancos emissores. Para um adquirente, é crucial determinar o potencial transacional de cada varejista para orientar estratégias adequadas de precificação e gestão de risco. Neste trabalho, propomos uma estrutura para avaliar adequadamente o potencial transacional de qualquer varejista usando as transações de seus pares. A estrutura proposta é baseada na construção de um contrafactual probabilístico que usa a regressão não paramétrica do kernel Nadaraya-Watson para modelar diferentes padrões sazonais, tendências e ciclos de negócios. Propomos uma metodologia integrada de processamento de dados para separar e validar os dados não afetados por intervenções para construir nosso modelo contrafactual probabilístico não paramétrico. O framework proposto é um poderoso sistema de suporte à decisão para gestão de receitas de uma adquirente, com aplicações diretas para precificação, detecção de churn e, de forma mais geral, gerenciamento de receita. Os resultados empíricos corroboram a eficácia do método em relação aos benchmarks relevantes.

Palavras-chave

Contrafactual; Churn não-contratual; Gerenciamento de Receita; Adquirente.

Contents

List of figures

List of tables

1
Introduction

The way we pay for our purchases has tremendously evolved throughout time. The rise of FinTech (Financial Technology) companies has accelerated this trend, as conventional financial institutions attempt to stay relevant and viable in this new environment (JUENGERKES, 2016). Along with banks and credit card companies, merchant acquiring is one of the parties engaged in the payment process. Merchant acquirers are firms that facilitate communication between a retailer (online or brick–and–mortar store) and the banks that issue credit and debit cards. They conduct financial operations and distribute money to retailers from sales made through these channels.

For the Brazilian market in particular, there was a duopoly in the acquirers market until 2010. Each company had an exclusive deal for the use of the credit card brand (Visa or Master), forcing business owners to contract with both to meet their consumers' demands. After ending the duopoly, the market began to expand by attracting new competitors. As a result, the payment market grew significantly due to cheaper rates for customers.

Notwithstanding the fact that cash is still commonly used in Brazil, the number of credit card users is rapidly increasing, thanks to the rise in e-commerce in recent years and the COVID-19 epidemic. Faced with this situation, businesses have pushed to make payment transactions for their consumers quicker, more diverse, and secure, utilizing all of the payment industry's technologies.

In this competitive industry, companies provide a variety of technical solutions, in addition to competitive fees on all types of transactions. In this context, a proper revenue management system of an acquirer must support pricing decisions, churn detection and retention campaigns, and credit and fraud risks. As well we argue that all these issues depend on an accurate estimate of a counterfactual model for the transactional profile of each retailer.

In this work, we address the underlying issue of the above mentioned business problems: we systematically assess transactional potential of each retailer in a non-parametric data-driven framework. Thus, the main objectives and contributions of this work are:

1. To develop a complete framework to construct a non-parametric probabilistic counterfactual (NPPC) for a retailer's transactional potential based solely on the transactional data of validated peers. We propose an adaptive rolling horizon framework to dynamically identify locally

abnormal behaviors for a given retailer due to an observable (e.g., price changes) or non-observable intervention (e.g., competitors lowering prices). Our approach provides a probabilistic counterfactual range in which the retailer's daily transaction amount should be found if no intervention takes place.

2. To develop a thorough validation methodology to ensure that the selected peers compose a valid synthetic control group for the NPPC. The peers must be free of external interventions to constitute a valid synthetic control group and avoid jeopardizing the causal-effect assessment associated with non-observable interventions. In the acquiring context, we need to ensure that selected peers did not churn before using their transactional profiles as predictors. To do that, we propose dynamic verification steps that provide validated data that are close enough to the real-time data to provide a good fit and far enough to enable stable counterfactual estimates to allow treatment detection based on consistent violation of the counterfactual confidence intervals.

3. To propose the usage of the validated peers as predictors on a probabilistic kernel regression to build the NPPC. We interpret the current transaction of each validated peer as a potential outcome whose probability increases with the similarity to the target retailer. Our model inherits the mathematical properties of a Nadaraya-Watson regression, thereby the NPPC renders weakly universally consistent probabilistic estimator.

1.1 Illustrative examples

In this section, we provide illustrative applications of our framework to address revenue management issues of an acquirer. All illustrative examples in this section were generated applying the proposed framework to real data of selected retailers. The main goal of this section is to enlighten the reader with a high level perspective on how the counterfactual could support the decision making process of a merchant acquirer.

For instance, Figure 1.1 depicts the counterfactual and the observed transactions of a retailer that did not suffer any external or internal interventions. Indeed, the observed transactional behavior remains within the (probabilistic) prediction interval constructed solely based on the transactions of the retailer's peers. In this case, the acquirer would interpret that there is no causal effect (e.g., churn) since transactional fluctuations could be completely explained by the seasonal economic movements affecting jointly the retailer and its peers. It is important to note that the proposed framework captures the transaction seasonality in a non-parametric manner, i.e., without the need to explicitly model each and every possible pattern. This is crucial for a decision support system that automatically generates recommendations for a variety of retailers with differing transactional patterns. Next, we illustrate situations of churn and price renegotiation.

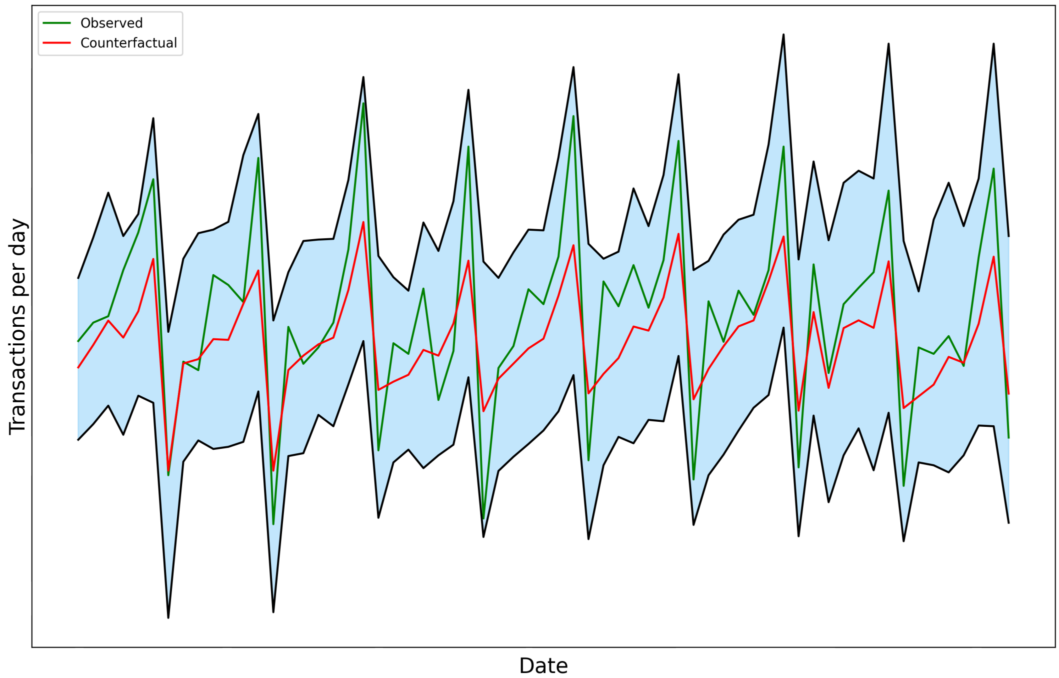


Figure 1.1: Measurement the prediction interval by NPPC.

In the context of an acquirer, churn cannot be detected in a straightforward manner. Very few retailers actually formalize the contract termination, but many change acquirers due to more competitive products and fees. Then, it is important to assess if a drop in the transactional profile is “natural” (due to seasonal economic movements) or due to an attack (e.g., a competitor offering lower fees). In Figure 1.2, we show that end-to-end monitoring is made possible by our framework, providing rapid churn detection. Note that the causal-effect is properly assess even though the associate intervention (e.g., competitors offering lower fees) is not directly observed. This is critical information for retention campaigns as well as for proper pricing strategies.

At last, we illustrate the opposite case where the retailer uses two different acquirers at the beginning of the analysis. After a proper price (fees) renegotiation, we may observe the full transactional potential of the retailer, see Figure 1.3. The intervention (e.g., renegotiation event) is indicated by the vertical black line. With this in hand, it is possible to estimate a price response function, which is critical for any revenue optimization system. We argue that

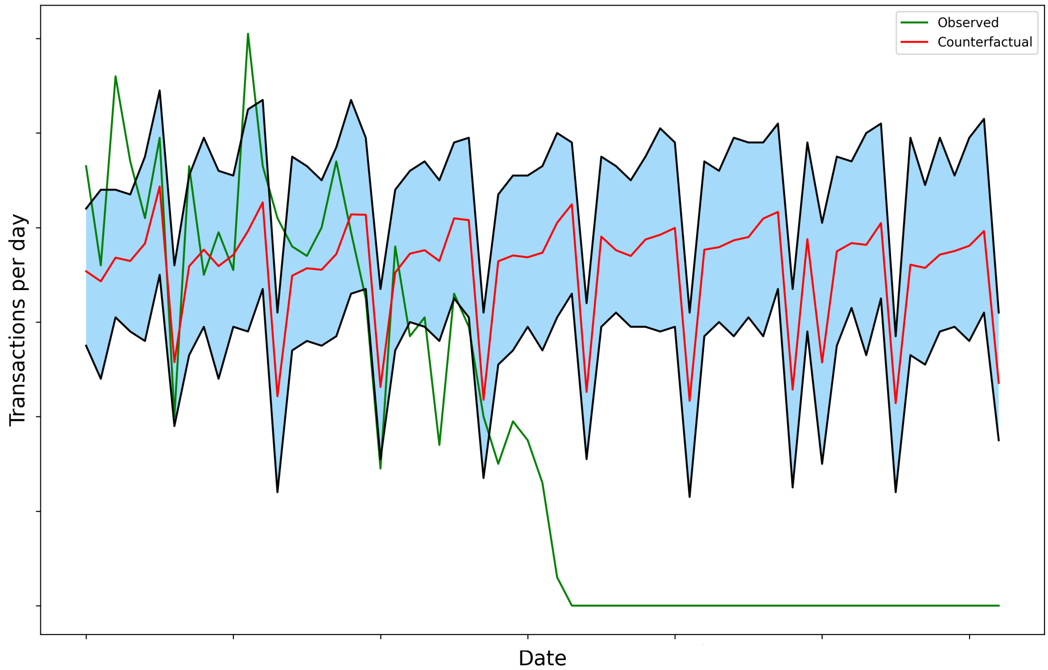


Figure 1.2: Illustration of how the counterfactual can be used to identify churn.

counterfactual estimation is a powerful tool to support revenue management decisions. Besides churn and pricing, it can also be used to help determining the Credit or Behavioral Score model, as well as fraud detection systems.

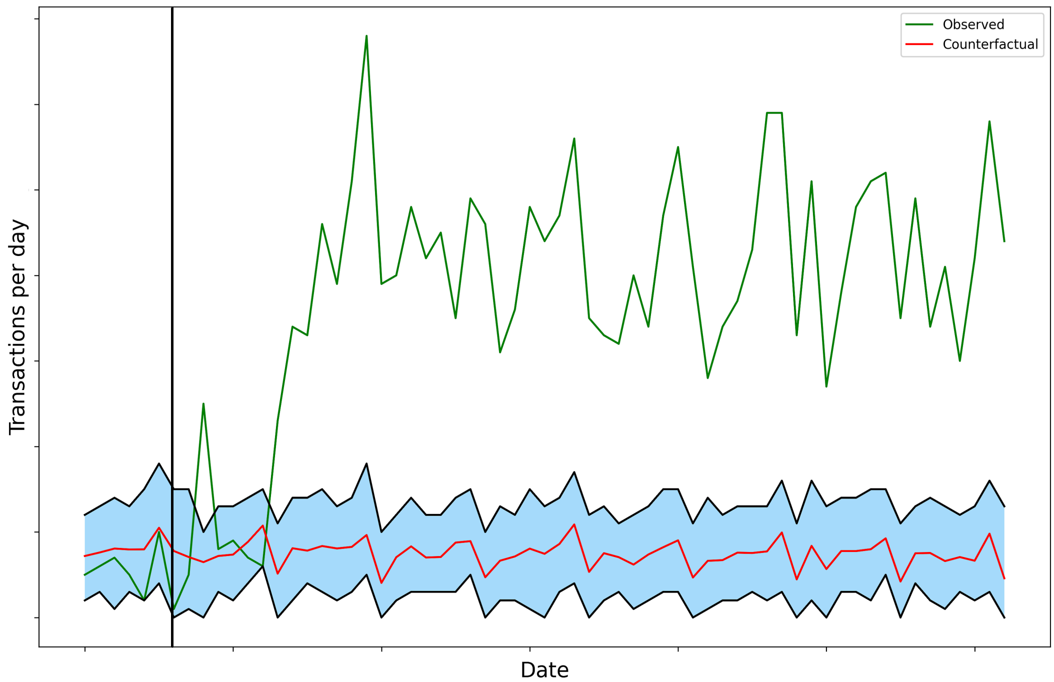


Figure 1.3: Illustration of how the counterfactual can be used to analyze a price change.

The remainder of this document is organized as follows. In Chapter 2 we present an overview of the existing literature on churn and counterfactual. In

Chapte 3 we specifies the methodology of this study. In Chapter 4 we presents the results and the description of the data set. In Chapter 5 we formulates our conclusions and future research.

2
Literature Review

Our research focuses on detecting and assessing the impact of an intervention that may alter the retailer's transactional profile. We argue that this transactional assessment is the underlying issue of many revenue management problems, such as, pricing strategies, churn detection and retention campaigns, credit and fraud risks. In this section, we review the relevant literature regarding the aforementioned topics.

In a churn context, customer relationship management (CRM) is extremely important for retailer retention, as the cost of acquiring a customer is greater than the cost of retaining a customer (MIN et al., 2016). Thus, churn prediction models are crucial to support retention campaigns and pricing decisions. Various machine learning algorithms are commonly used to predict churn. The most used are logistic regression (JAHROMI; STAKHOVYCH; EWING, 2014; MATUSZELAŃSKI; KOPCZEWSKA, 2022), decision trees (HÖPPNER et al., 2020), artificial neural networks (CAIGNY et al., 2020; KUMAR; YADAV, 2020), deep learning models (KIM et al., 2017), linear models (KARMAKAR et al., 2022) and boostings (LEMMENS; GUPTA, 2020; BEEHARRY; FOKONE, 2022; MILOŠEVIĆ; ŽIVIĆ; ANDJELKOVIĆ, 2017).

Most prediction models are supervised classification techniques highly dependent on a clear and assertive definition of the dependent variable ("churn" or "not-churn"). This is the case for the contractual churn, i.e., whenever the service interruption is tied to a termination of a contract. However, in many non-contractual cases, it is not possible to observe churn in a clear and straightforward manner. Since non-contractual churn might be partial and continuously ranged, the definition of churn is usually done in a subjective (ad-hoc) manner (BUCKINX; POEL, 2005; PERIŠIĆ; JUNG; PAHOR, 2022; MATUSZELAŃSKI; KOPCZEWSKA, 2022), but can also be tied to a probabilistic repurchase model (CHOU et al., 2022).

In our context, churn is interpreted as a causal-effect of a non-observed intervention usually related to competitors offering cheaper/better services. Similarly to the Synthetic Control (SC) method (ABADIE; DIAMOND; HAINMUELLER, 2010; ABADIE; DIAMOND; HAINMUELLER, 2015) and the artificial counterfactual (ArCo) method (MASINI; MEDEIROS, 2021), our approach focus on a single unit under intervention and no explicit control group. Our methodology significant differ from SC and ArCo since we use non-parametric probabilistic regression and address the case of a latent (non-observable) intervention. For further details on causal inference on observational data, we refer to Yao et al. (2021). We also refer to Györfi et al. (2002) for details on the theory of non-parametric regression and to Stuart (2010) for its application to causal inference.

2.1
Extended literature on churn

CRM's mission is to enhance and deepen consumer relationships so that a customer's lifetime value to a brand can be maximized (PEPPERS; ROGERS; DORF, 1999). To thrive with CRM, businesses must match offerings and campaigns to prospects consumers, or handle the consumer life cycle intelligently (EDELSTEIN, 2000). In addition, long-standing customers consume more and still refer to the company's products and services (DAWES; SWAILES, 1999). From this standpoint, minimizing churn of customers is essential and, consequently, it is necessary to develop churn detection models and apply them in business. These models aim to identify customers who are about to abandon the service offered so that managers can define some strategy to retain them and manage the wear and tear of the relationship. In a competitive environment, customer retention is essential, in addition to attracting new customers.

Churn detection should be treated differently depending on the type of business being evaluated, as churn characteristics for a postpaid and prepaid telecommunication services are different. In the first case, the service is linked to a contract; churn is total and specific for each product; the end of contracting the service is the occurrence of churn in its clearest form; churn is determined at the exact moment when the customer breaks the relationship. In the second case, which is the focus of our research, the service is not tied to a contract; the cost of the customer to change the service provider or supplier is low or zero; churn is not directly observed, as the customer does not warn, and can be partial or continuous. In this situation the definition of churn is subjective (MATUSZELAŃSKI; KOPCZEWSKA, 2022) and given that, analysts need to take into account what churn represents in the area under study. In this configuration, it is possible for a customer to return after a period of inactivity (JAHROMI; STAKHOVYCH; EWING, 2014).

Due to the ease of data availability and the data storage capacity, arising from the advances in information technology (FADER; HARDIE, 2009) and cloud services (JIANG et al., 2021), it is necessary to extract knowledge and support decision making. For this, state-of-the-art machine learning methods are used to harness information from a large amount of data, such as logistic regression (JAHROMI; STAKHOVYCH; EWING, 2014; MATUSZE-

LAŃSKI; KOPCZEWSKA, 2022), decision trees (HÖPPNER et al., 2020), artificial neural networks (CAIGNY et al., 2020; KUMAR; YADAV, 2020), deep learning models (KIM et al., 2017), linear models (KARMAKAR et al., 2022), and boostings (LEMMENS; GUPTA, 2020; BEEHARRY; FOKONE, 2022; MILOŠEVIĆ; ŽIVIĆ; ANDJELKOVIĆ, 2017).

When working with non-contractual settings, where consumers have the ability to alter their buying behaviour without consulting the firm, model construction for customer's profile prediction gets more difficult. The churn of these users directly impacts revenue management. This involves managerial decisions to implement pricing policy, credit and retention campaigns in a merchant acquirer, in our case study. These interventions, made by the company or by the competition, have an impact on revenue and need to be measured through data analysis.

In the non-contractual churn context, there are three possible states as illustrated in Figure 2.1: the active customer base, soft churns (relaxed churn) and hard churns (tight churn). Active customers are those who are satisfied with the services provided and consume regularly, that is, they are the loyal customers. Soft churns are the customer base that will leave the company and those that have significantly reduced the use of services (MANDIĆ; KRALJEVIĆ; BOBAN, 2018). Hard churn represents the window after the last transaction performed (WANG, 2018). The objective is to identify the state of pre-hard churn, so that there is still time to act to retain the customer, allowing the monitoring of the processes that may lead to churn (LORIA; MARCONI, 2021). If companies are not accurate in identifying churns, there may be an incentive for a fictitious churn, resulting in a loss of revenue.

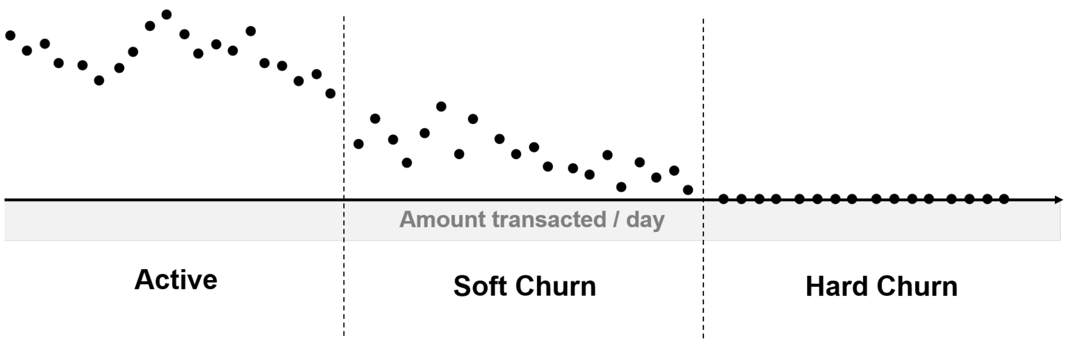


Figure 2.1: Possible states in the non-contractual churn context.

Churn prediction models are often based on RFM (recency, frequency, monetary) variables (BUCKINX; POEL, 2005). We can also see some characteristics related to customer behavior, customer perceptions, customer demographics, macroenvironment, and customer-company interaction in relation to

RFM (CAIGNY et al., 2020). To this day, RFM is used by marketers to analyze consumer behavior and thus segment the customer base. Usually we starting defining a score for each of the following three dimensions: - How recent was the last purchase? - How often do you buy? - What is the value of purchases? The RFM variables are used in different stages of the non-contractual churn forecast:

– **Sample selection**: authors use RFM-related statistics as a criterion to define the studied sample;

– **Churn definition**: RFM variables are used as the basis for models construction and analyzes to generate a churn indicator variable in non-contractual arrangements;

– **Choice of Variables / Model**: RFM variables are identified as the main predictors of classification models for non-contractual churn prediction.

Among the most used predictor variables in churn studies, only predictors related to customer behavior and demographic data have been used in non-contractual arrangements, as shown in Table 2.1. The areas of research activity are listed in Table 2.1, as well as the definition of churn for each research.

Table 2.1: Predictors of churn and churn definition in non-contractual arrangements in previous research.

Reference	Behavioral background	Demography	Area	Churn definition
Buckinx & Poel (2005)	X	X	FMCG Retail	5 months
Jahromi, Stakhovych & Ewing (2014)	X		Online retail (B2B)	6 months
Hopmann & Thede (2005)	X		Retail (B2B)	12 months
Martínez et al. (2020)	X	X	Retail (B2B)	1 month
Glady, Baesens & Croux (2009) (2009)	X		Retail Bank	Dynamic
Jahromi et al. (2010)	X		Telecommunication	7 days
Babkin & Goldberg (2017)	X		Online retail	1 day
Yoon, Koehler & Ghobarah (2010)	X	X	Online Ads	181 days
Calciu (2009)	X		Cruiseship & Catalogue sales	Dynamic
Lee et al. (2018)	X		Games	5 weeks
Kim et al. (2017)	X		Casual Games	Dynamic
Milošević, Živić & Andjelković (2017)	X		Games	14 days
Our study	X		Merchant Acquirers	7 days

In addition to arbitrary rules based on RFM, churn forecasting models in non-contractual arrangements can be made through clusters analysis and by the customer's life value (CLV). The first is according to the consumption standard / RFM, where a different churn is defined for each cluster or a model for each cluster (JAHROMI et al., 2010; MATUSZELAŃSKI; KOPCZEWSKA, 2022). The latter, instead of looking to the past to detect if the customer has changed, CLV focuses on the future, to estimate whether the relationship will remain profitable, by summing up future cash flows (GLADY; BAESENS; CROUX, 2009; IMANI et al., 2021). The term net present value has also been used to indicate the customer's value, as noted by Calciu (2009).

Churn is defined in our context as the result of an unobserved intervention, generally connected to competitors providing superior services. Our approach, like the SC method and the ArCo, focuses on a single unit under intervention with no stated control group. Our technique differs significantly from those of SC and ArCo because we employ non-parametric probabilistic regression and handle the scenario of an unobservable intervention. The counterfactual is a method used to assess the effects of an activity on some entity. The counterfactual is hypothetical. It represents the retailers' behavior if the intervention had not taken place. In this approach, we may assess behavior in the absence of intervention or utilize counterfactual explanations to obtain insights from supervised machine learning models that did not provide the desired result. The counterfactual can be applied in churn, generating a constant monitoring of the customer at the end, being able to have a faster response to the competitor's attack. Counterfactual can also help with pricing, by estimating the price response function, calculating potential demand and price elasticity. It can also be used as a reference for determining default, as well as being used as an attribute in a Credit or Behavioral Score model. In this research, we focus in intervention by the competition. We will see more details of counterfactual and causal inference in the next section.

2.2 Extended literature on counterfactual and causal inference

A frequent purpose for counterfactuals is credit risk prediction through machine learning models. People whose credit applications have been denied should understand why they were denied, either to better comprehend the decision-making process or to assess their active possibilities for changing the outcome. These explanations are provided via counterfactuals. In our case, we would look at the customer's behavior if the competition had not intervened or if economic factors had not influenced the customer's behavior, inducing churn. When the customer disconnects from his control group, we would track his churn. Counterfactuals explain why a certain outcome was not accomplished, offer logical grounds to protest to an unjust conclusion, and offer recommendations on how the intended prediction could be attained in the future (WACHTER; MITTELSTADT; RUSSELL, 2017). The counterfactual involves estimating the effects on a specific consolidated variable by comparing an aggregated entity that has had some type of intervention to other aggregated entities that have not been affected by the same occurrence (control group). In our particular instance the control group is made up of retailers who did not get any treatment during the research period.

Many of the machine learning techniques produce useful results, but they are difficult, if not impossible, to comprehend for stakeholders. Counterfactual explanations are a good way for explaining specific predictions of a model. Through the development of a control group, counterfactual facilitates this understanding by comparing what happens with the entity being assessed to what would have happened if there had been no intervention in this entity. The counterfactual gives an explanation to the applicant, as well as input on how stakeholders might behave in the future to achieve the desired result. Through the use of counterfactual, the explanations can also assist machine learning model developers in identifying, detecting, and fixing flaws and other performance concerns. Counterfactual explanations do not provide a clear response to the "why" of a decision. They provide choices for reaching the desired result (VERMA; DICKERSON; HINES, 2020). In this study, the counterfactual is treated as learning from observational data. This type of application can be used to determine ideal actions as well as the effect of causal treatment (KÜNZEL et al., 2019; BERTSIMAS; KALLUS, 2020).

The proposed methodology of applying a counterfactual to measure impacts in a control group is broad and covers several areas of action (CHAMON; GARCIA; SOUZA, 2017; ALBADVI; VARASTEH, 2010; DASGUPTA; MASON, 2020; BERTSIMAS et al., 2017). Chamon, Garcia & Souza (2017) applied a synthetic control to estimate the impact on the level and volatility of the exchange rate, through the measures of the Central Bank of Brazil. Albadvi & Varasteh (2010) ran simulations to assess the causal effect of the decriminalization of prostitution on sexually transmitted infections in Rhode Island. Dasgupta & Mason (2020) studied the effect of interest rate ceilings on bankruptcy. In the area of individualized care for diabetes, Bertsimas et al. (2017) used KNN to estimate the counterfactuals.

The counterfactual somewhat resembles the synthetic control (SC) pioneered by Abadie & Gardeazabal (2003) and Abadie, Diamond & Hainmueller (2010), and ArCo by Carvalho, Masini & Medeiros (2018). Despite the fact that both the ArCo and the SC procedures create a counterfactual as a function of observable factors from a group of peers, the two approaches have major differences. First, as Ferman, Pinto & Possebom (2020) point out, the SC method generates the counterfactual using a convex combination of peers, which biases the estimator. The ArCo solution was devised by Carvalho, Masini & Medeiros

(2018) and is a general, perhaps nonlinear function. Even in the case of linearity, the method does not impose any limitations on the parameters (CARVALHO; MASINI; MEDEIROS, 2018). Our technique differs significantly from those of SC and ArCo because we employ non-parametric probabilistic regression and handle the scenario of an unobservable intervention.

Viviano & Bradic (2019) provide the first set of criteria that may be used to build valid synthetic control tests using predictions from a range of machine learning techniques, such as Random Forests and Neural Networks. Coate et al. (2021) applied machine learning techniques to assist in the execution of a synthetic control model. Their model was used to analyze counties within the United States that showed a vote shift from the majority of Democratic voter share to Republican between the 2012 and 2016 election cycles. Souza (2019) used machine learning algorithms to accurately predict counterfactuals, which can then be used to estimate a distribution of treatment effects, in a large energy efficiency program in the US.

Causality may be inferred in two ways: through a controlled and random test or through observational data. The first randomly assigns participants to a treatment group or a control group, and a single difference between the control and treatment groups is a variable to be inferred. There is no control over treatments or subjects in observational data, and we don't know why a subject chose a certain treatment. The challenge is to estimate the subject's outcome if he were treated to a different intervention. The presence of confounders, which are factors that impact both treatment assignment and result, can lead to the spurious effect or even Simpson's paradox, causing issues in estimating the counterfactual. According to Yao et al. (2021), there are numerous approaches to dealing with confounders issues: re-weighting methods, stratification methods, matching methods, tree-based methods, representation based methods, multi-task methods, and meta-learning methods. We employed matching methods to find retailers with similars transaction profile to determine the impact of a particular treatment on unit.

The goal of matching approaches is to balance the distribution of variables in the treatment and control groups. As a result, we attempt to decrease the bias in evaluating the treatment impact (STUART, 2010). See Yao et al. (2021) for more details about the approaches.

A parametric estimate, regardless of the amount of data, cannot match the data if the enforced format differs significantly from the data. The parametric estimate eventually limits the function's fit to the data. Non-parametric estimation is used in our work to estimate a function that fits the data without applying any assumptions about the data. Kernel estimation is a type of non-parametric estimation.

In our study Nadaraya-Watson estimator (NW) is used for sample weighting technique, as the kernel approach. Kernel matching is a type of non-parametric matching that creates the counterfactual result by using weighted averages of observations in the control group (STUART, 2010). We expand this approach to accommodate unobserved intervention since we don't know when the intervention occurs when dealing with non-contractual churn. We developed a process to preserve the control group by implementing a synthetic control (ABADIE; DIAMOND; HAINMUELLER, 2010; ABADIE; DIAMOND; HAINMUELLER, 2015). In their weight adjustment technique, Abadie, Diamond & Hainmueller (2015) employed linear regression, while we used non-parametric regression for point and prediction range estimates.

In our work, we take a data-driven approach to dealing with a wide range of decision questions in operations research and management science (OR/MS), geared towards churn but extendable to other applications. Bertsimas & Kallus (2020) and Bertsimas & Kallus (2016) adopt this strategy. Bertsimas & Kallus (2016) explores the topic of optimal price prescription, learning from past demand and other information in the context of revenue management and pricing, keeping in mind that demand data is observational. The counterfactual application's greatest value is in the data, not the model. In Chapter 3 we will discuss about the methodology of our proposed framework to construct a non-parametric probabilistic counterfactual (NPPC) for a retailer's transactional potential based solely on the transactional data of validated peers.

3
Methodology

In this chapter, we develop a complete framework to construct a non-parametric probabilistic counterfactual (NPPC) of a retailer solely based on the transactional data of validated peers. We have divided the chapter into two parts, the first being related to the background knowledge of the literature and the last to the proposed methodology.

3.1
Background knowledge

A generalized function that describes data is denoted as $Y_i = \hat{y}\left(X_i\right) + \epsilon_i$, where the intended error value should be minimized. The estimation of $\hat{y}(x)$ may thus be measured using the mean of Y_i. As a result, we have:

$$\hat{y}_n(x) = \sum_{i=1}^{n} W_{n,i}(x) \cdot Y_i \tag{3-1}$$

where the weights $W_{n,i}(x)$ are smaller the farther X_i is from x.

In regression models of a parametric function is specified order to capture the relationships between an outcome variable y and a set of regressors. A basic example is linear regression, which implies that the $x = \left(x^{(1)}, \ldots, x^{(d)}\right)^T$ components are linearly combined, i.e.,

$$\hat{y}\left(x^{(1)}, \ldots, x^{(d)}\right) = a_0 + \sum_{i=1}^{d} a_i x^{(i)}, \quad \left(\left(x^{(1)}, \ldots, x^{(d)}\right)^T \in \mathcal{R}^d\right) \tag{3-2}$$

The unknown parameters are estimated using least squares, which chooses the estimates that best matches the data while reducing error:

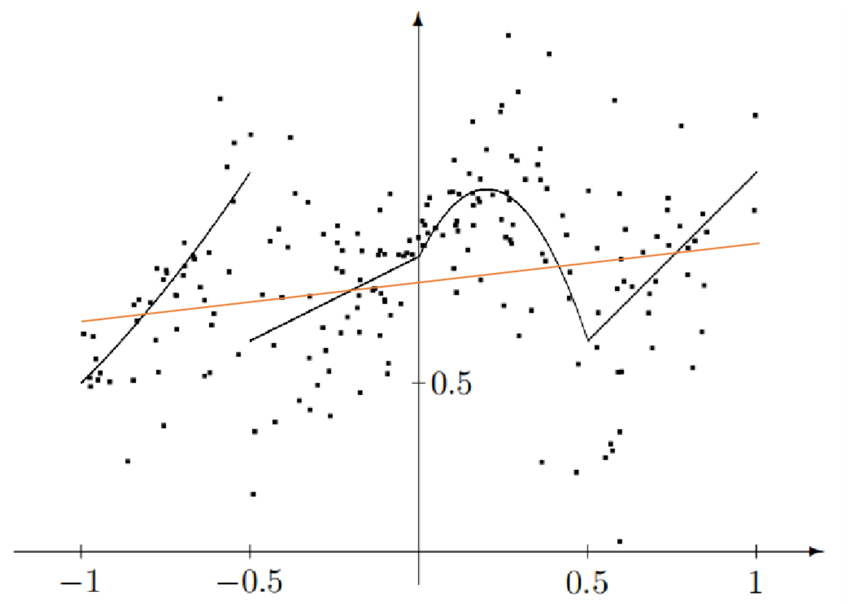
$$\left(\hat{a}_0, \ldots, \hat{a}_d\right) = \arg \min_{a_0, \ldots, a_d \in \mathcal{R}^d} \left\{ \frac{1}{n} \sum_{j=1}^{n} \left| Y_j - a_0 - \sum_{i=1}^{d} a_i X_j^{(i)} \right|^2 \right\} \tag{3-3}$$

where $X_j^{(i)}$ represents the i th component of X_j, and $z = \arg\min_{x \in D} f(x)$ is an abbreviation for $z \in D$ and $f(z) = \min_{x \in D} f(x)$.

The fundamental drawback of parametric estimate is that if the defined function does not provides a good approximation to the relationship between y and $x's$, the fit will be poor. It may be simple to create a function that best matches the data for univariate issues, but this is not the case for multivariate problems. Györfi et al. (2002) used simulated data to demonstrate this problem in his book on the theory of non-parametric regression. This example is composed of $n = 200$ points such that X is standard normal confined to $[-1,1]$, i.e., the density of X is proportional to the standard normal density on $[-1,1]$ and zero elsewhere. The regression function is polynomial in pieces.

$$\hat{y}(x) = \begin{cases} (x+2)^2/2 & \text{if } -1 \leq x < -0.5 \\ x/2 + 0.875 & \text{if } -0.5 \leq x < 0 \\ -5(x-0.2)^2 + 1.075 & \text{if } 0 < x \leq 0.5 \\ x + 0.125 & \text{if } 0.5 \leq x < 1 \end{cases} \tag{3-4}$$

the conditional distribution of $Y - \hat{y}(X)$, given X, is normal, with a mean of zero and a standard deviation of $\sigma(X) = 0.2 - 0.1\cos(2\pi X)$. Figure 3.1 depicts the data points, the regression function, and a linear estimate based on the simulated data.



Source: adapted from Györfi et al. (2002).

Figure 3.1: Data points, regression function (black lines), and linear regression estimate (orange line).

Györfi et al. (2002) cites the local averaging as one of four paradigms associated with non-parametric regression. The Nadaraya-Watson kernel estimate and the k-nearest neighbor (KNN) estimate are two examples of this. KNN selects the X_i's closest to x based on their distance and calculates $m(x)$ using the average of the neighbors discovered, along with their associated Y_i's and it is defined as:

$$\hat{y}_n(x) = \frac{1}{k}\sum_{i=1}^{k} Y_{(i)}(x) \tag{3-5}$$

in this case, the weight $W_{n,i}(x)$ equals $1/k$ if X_i is one of k nearest neighbors of x, and 0 otherwise.

We conduct the averaging location in our study by locating the closest neighbors, calculating their weights, and then taking the average to get the forecast of the assessed unit. The mean of the data is a constant that best matches the data in terms of squared errors.

$$\hat{y}_n(x) = \arg\min_{c \in \mathcal{R}} \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)(Y_i - c)^2 \tag{3-6}$$

In summary, we have a weighted problem if we use the weighted average. If this is correct, we can extend the Eq. 3-6 and instead of using a constant c, we can use a linear function or any other function:

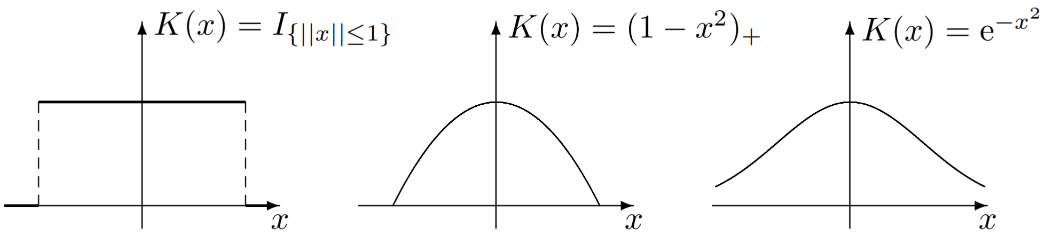
$$\{\hat{a}_k(x)\}_{k=1}^{l} = \arg\min_{\{a_k\}_{k=1}^{l}} \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\left(Y_i - g\left(X_i, \{a_k\}_{k=1}^{l}\right)\right)^2 \tag{3-7}$$

where $g\left(\cdot, \{a_k\}_{k=1}^{l}\right) : \mathcal{R}^d \to \mathcal{R}$ is a function that depends on the $\{a_k\}_{k=1}^{l}$ parameters. Using a local least squares criteria, select values for these parameters for each $x \in \mathcal{R}^d$. This is the local modeling paradigm from Györfi et al. (2002). If $g(x, \{c\}) = c\left(x \in \mathcal{R}^d\right)$ is chosen, the Nadaraya-Watson kernel estimate is obtained:

$$\hat{y}_n(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)} \tag{3-8}$$

where $h > 0$ is a bandwidth.

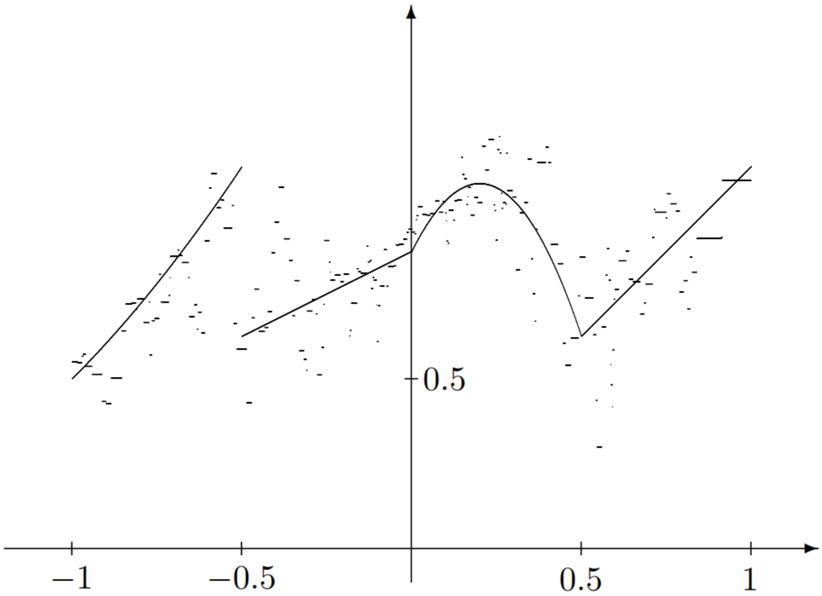
The univariate kernel sensitivity in h is quite great. In our study, we used a multivariate kernel, which is more complicated, and the inverse distance, where $k = 1/d$, which handles this situation well and eliminates the need for h. Györfi et al. (2002) presents numerous classic kernels, such as the naïve kernel ($K(x) = \mathcal{I}_{||x|| \leq 1}$) and the Epanechnikov kernel ($K(x) = (1 - ||x||^2)_+$), for the simulated data in Eq. 3-4, as seen in Figure 3.2.



Source: Györfi et al. (2002).

Figure 3.2: Examples for univariate kernels.

Figure 3.3 depicts an estimation of the KNN based on Eq. 3-5 for the simulated data from Eq. 3-4, as well as the L_2 error.
where we reorder the data according to the distances value of $||X_i - x||$. If we get a tie, we declare i closer if $i < j$.



Source: Györfi et al. (2002).

Figure 3.3: Undersmoothing: $k = 3$, L_2 error= 0.011703.

If $k \to \infty, k/n \to 0$, then the KNN regression function estimate is weakly consistent for all (X,Y) distributions where ties occur with probability zero and $EY^2 < \infty$. This is demonstrated by Stone's Theorem (STONE, 1977). This may be proved in a broad sense for both a kernel estimate that include the Nadaraya-Watson kernel employed in our research. More theoretical information may be found in Györfi et al. (2002).

3.2 Proposed methodology

We have developed a complete framework to build a non-parametric probabilistic counterfactual (NPPC) of a merchant exclusively based on transactional data of validated peers. The proposed framework aims at assessing the transactional potential by considering observable (e.g., price renegotiation) or non-observable (e.g., competitors lowering prices) interventions. To do so, we develop an adaptive rolling horizon scheme to dynamically construct a validated synthetic control group used as predictors in a non-parametric probabilistic kernel regression.

Let us conceptually define a verified training sample as a time window where the target retailer is free of interventions. When applying the NPPC, we assume that the retailer of interest did not undergo any intervention in the first training sample $\mathcal{S}$ of the rolling horizon. Additionally, we argue that the construction and validation of the synthetic control group $\mathcal{C}$ must safeguard the counterfactual estimate from non-observed interventions in the control group.

Figure 3.4 summarizes one step of the rolling horizon scheme. Let us denote y_t as the number of daily transactions of the target retailer at time $t \in \mathcal{T}$. Let also $x_{i,t}$ be the number of transactions at time t for the potential peer $i \in \mathcal{I}$. Given synthetic control group $\mathcal{C} \subseteq \mathcal{I}$ and a training sample $\mathcal{S} \subseteq \mathcal{T}$.

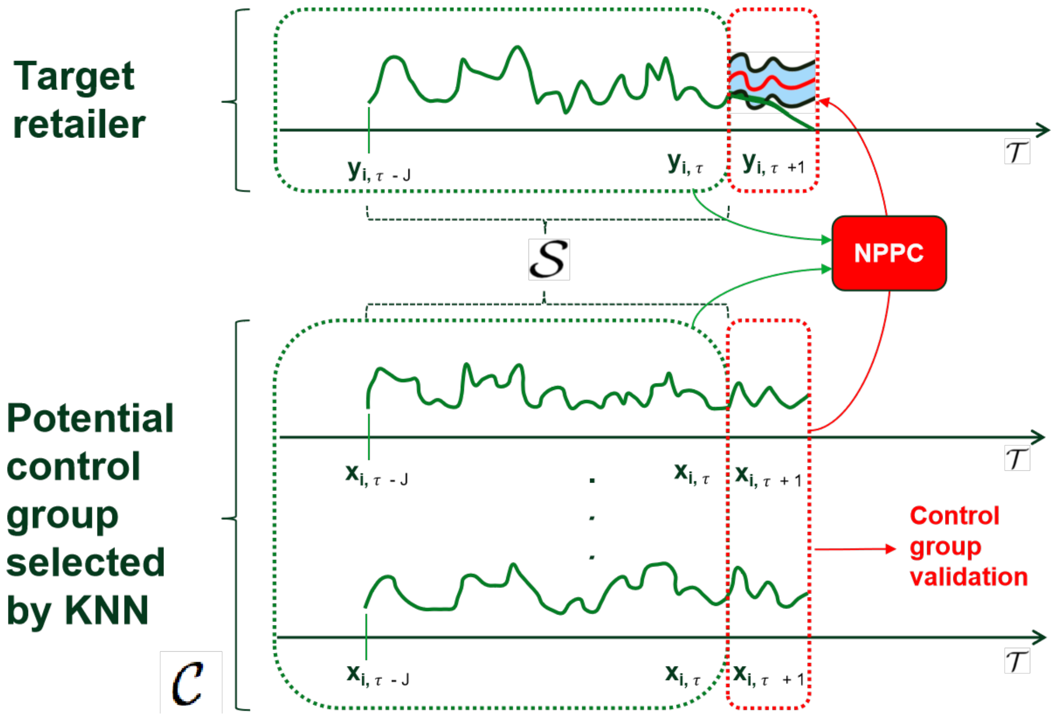


Figure 3.4: Framework for the counterfactual estimation.

Assuming that the retailer under evaluation receives no intervention in the first step, the control group will not receive either, but in a rolling horizon scheme, this control group must be validated at each step so that it can be used to create the non-parametric probabilistic counterfactual (NPPC). We were able to measure potential retailers behaviors and, as a result, determine whether or not the retailer received intervention by consistently violating the prediction interval. In Figure 3.5, we summarize the proposed framework to construct the NPPC of a retailer. Then, we depict the main steps of the framework in the following subsections.

The first step is to build the control group after we have verified training samples that are free of intervention.

3.2.1 Construction of the Control Group

We choose a control group that is identical to the retailer of interest except for one detail: its exposure to intervention. There should be no intervention in the control group at any time.

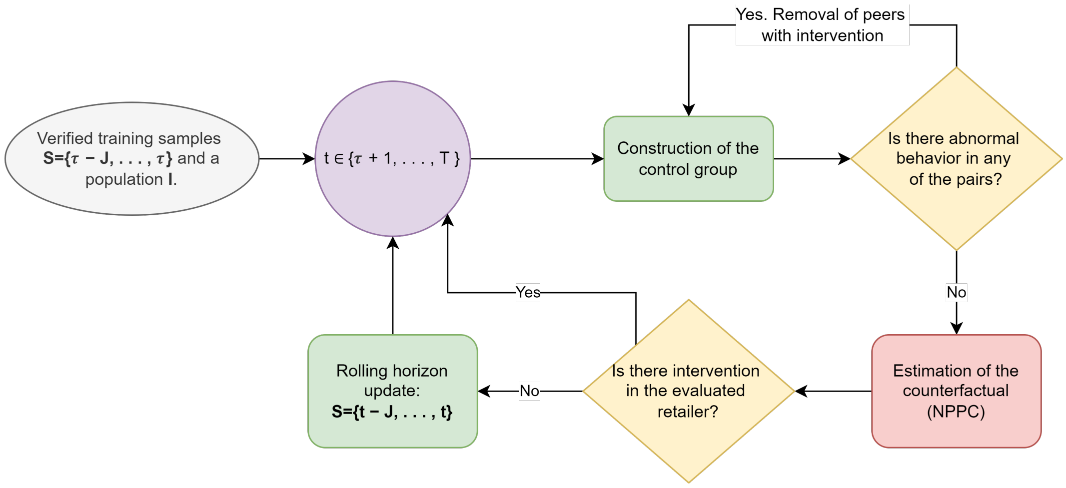


Figure 3.5: Flowchart of the proposed methodology for the NPPC.

Following Stuart (2010), we argue that matching methods reduce estimation bias of the causal effect. In particular, we use the KNN with relatively small number of neighbors that resembles the (intervention-free) behavior of the target retailer. We denote by

$$N(k;\mathcal{S},\mathcal{I}) = \min_{\mathcal{C}\subset\mathcal{I}} \left\{ \sum_{i\in\mathcal{C}} \sum_{\tau\in\mathcal{S}} (y_\tau - x_{i,\tau})^2 \,\middle|\, |\mathcal{C}| = k \right\} \tag{3-9}$$

the KNN selection using the last verified training sample $\mathcal{S}$ and a population $\mathcal{I}$. We argue that this selection is likely to discard churned peers or other abnormal behaviors. The control group is chosen based on the time series most similar to the retailer being evaluated. To illustrate the selection, we confront in Figure 3.6 the target training data with its nearest neighbors.

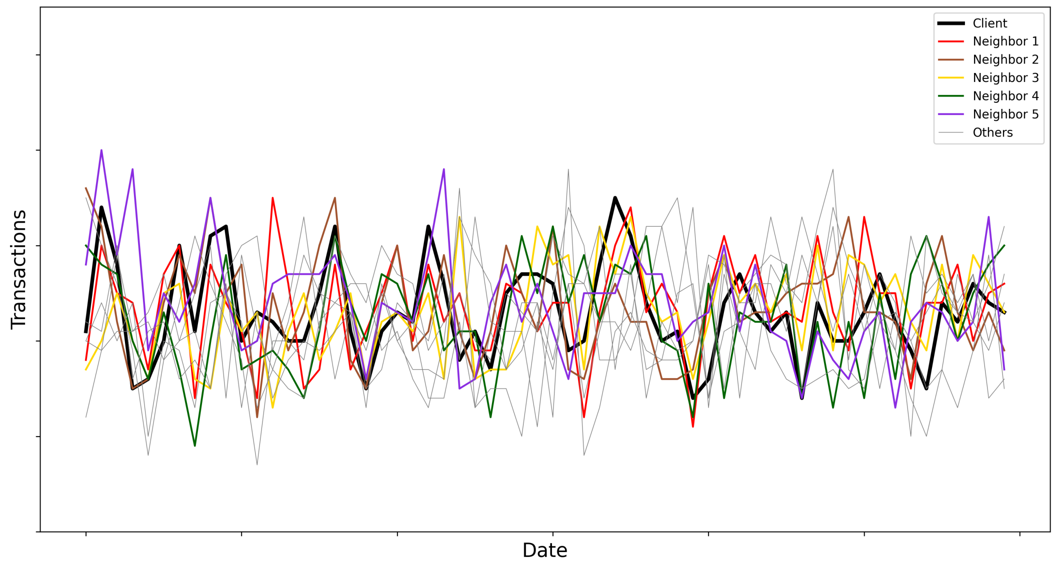


Figure 3.6: Example of identification of the closest neighbor within a 60-day period.

If we put more retailers in this view, we'd see curves getting further and further away. Next step is to validate this control group, removing abnormal behavior.

3.2.2 Control Group Validation

Validation occurs in the following manner: the control group can intervene in the rolling horizon, requiring a validation procedure. The counterfactual would readapt to the intervention if this procedure did not exist and the retailer being evaluated received an intervention.

We propose a validation procedure for the synthetic control group (SCG) that aims to exclude pairs with abnormal behavior when compared to the verified training sample, because the construction of the NPPC should mitigate the effect of unobservable interventions, even at post-training dates. We define the distance between the pairs using a given control group $\mathcal{C}$ and a given date $t \in \mathcal{T}$ by the Euclidean distance:

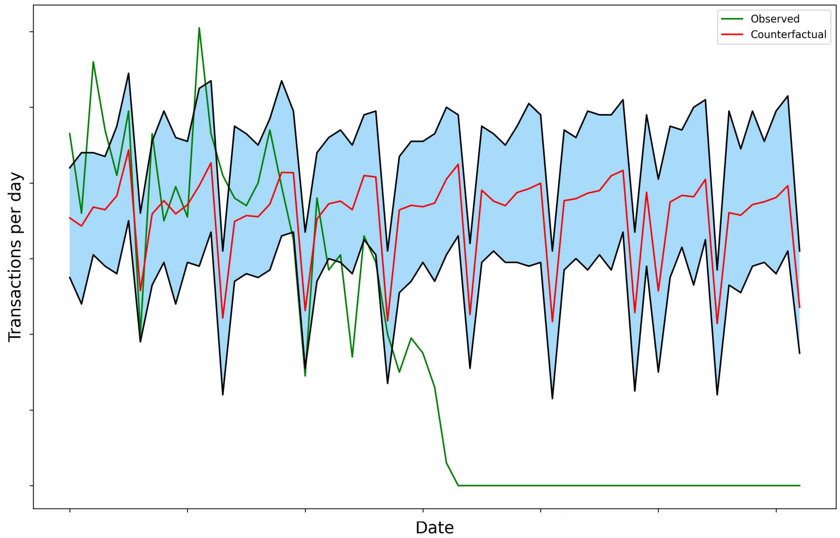
$$d_{i,t} = \left[\sum_{j \in \mathcal{C}} (x_{i,t} - x_{j,t})^2\right]^{0.5} \tag{3-10}$$

this distance can be thought of as a reference point for validating the synthetic control group. In a given verified training sample $\mathcal{S}$, the maximum distance is a measure of the distance between a given customer i and its group. Now, we use the last verified training sample $\mathcal{S}$ to compute maximum distance

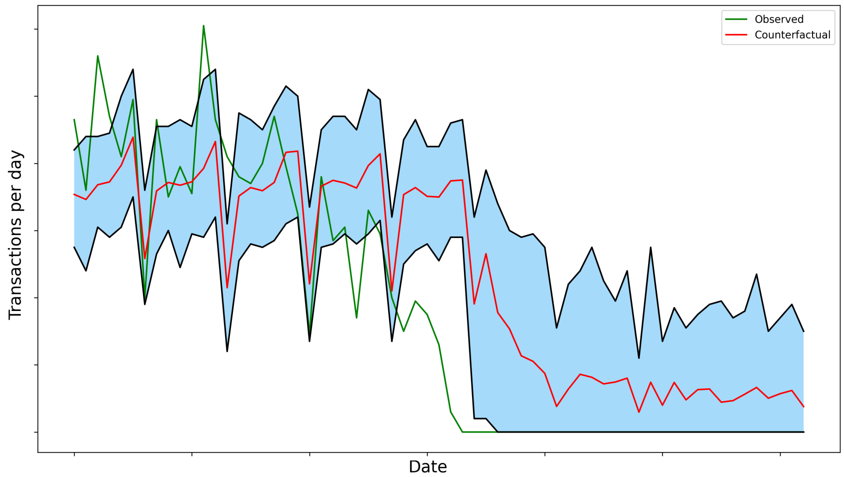
$$D_i(\mathcal{S}) = \max_{\tau \in \mathcal{S}} \{d_{i,\tau}\}, \tag{3-11}$$

a reference threshold to validate the SCG. Indeed, the maximum distance is a measure of how far a given retailer i was from the group in a given verified training sample. For a given post-training dates $t > \max_{\tau \in \mathcal{S}}$, we argue that if $d_{i,t} > D_i(\mathcal{S})$, then the peer behavior might be considered abnormal, and should be excluded from the control group. This simple heuristic procedure has shown to be very effective in our computational studies. In addition to this rule, customers who go seven days without transacting are removed from the control group.

Figures 3.7-(a) and (b) from the same retailer show the effectiveness of this procedure in keeping the control group free of interventions. If this procedure did not exist, the counterfactual would adapt to the new reality of the evaluated retailer.



(a) NPPC application with the validation procedure.



(b) NPPC application without the validation procedure.

Figure 3.7: Particular case of a retailer with and without a validation procedure.

3.2.3 Non-parametric Probabilistic Counterfactual

Given that the retailer under consideration has a validated control group, we can estimate his NPPC for time $t \in \mathcal{T}$. We define the non-parametric probabilistic counterfactual as:

$$\hat{F}_t(z; \mathcal{C}, \mathcal{S}) = \sum_{i \in \mathcal{C}} \left(\frac{K_{\mathcal{S}}(i)}{\sum_{c \in \mathcal{C}} K_{\mathcal{S}}(c)} \right) \mathbb{1}(x_{i,t} \leq z), \tag{3-12}$$

as the Nadaraya-Watson (local averaging) estimate for the cumulative probability distribution of the target retailer, where $\mathbb{1}(x_{i,t} \leq z) = 1$, if $x_{i,t} \leq z$, and $\mathbb{1}(x_{i,t} \leq z) = 0$, otherwise. We consider a inverse distance kernel

$$K_{\mathcal{S}}(i) = \left[\sum_{\tau \in \mathcal{S}} (y_\tau - x_{i,\tau})^2\right]^{-0.5}, \tag{3-13}$$

due to its simplicity, interpretability and relatively fast computation. Note that we can interpret the NPPC as a discrete distribution, whereby a potential outcome $x_{i,t}$ has the probability of $K_{\mathcal{S}}(i)/\sum_{c \in \mathcal{C}} K_{\mathcal{S}}(c)$. Thus, we use these outcomes and associate probabilities to directly construct quantiles for a probabilistic assessment of the retailer's behavior, see for instance Figure 1.1, 1.2 and 1.3.

Moreover, we argue that this estimate is a conditional distribution since it reflects the characteristics of the chosen synthetic control group $\mathcal{C}$ and training sample $\mathcal{S}$. For causal-effect assessment, we argue that the training sample $\mathcal{S}$ must be verified and synthetic control group $\mathcal{C}$ constructed and then validated.

3.2.4
Intervention check at the retailer

If the retailer's observed value falls within the prediction interval, we can conclude that there was no intervention at time $t \in \mathcal{T}$ and update the rolling horizon accordingly. If there is a violation of the prediction interval, we keep the current validation horizon in which the control group was validated and look one step ahead. If the retailer of interest remains outside the prediction interval, it is a good indication that he has been influenced as illustrated in the Figure 1.2 and Figure 1.3. This is the non-parametric probabilistic counterfactual framework.

Now, we summarize our framework on a pseudo-algorithm that highlights the main steps to construct the NPPC.

Algorithm 1 Construction, validation and verification of the counterfactual

Input: A verified training sample $\mathcal{S} = \{\tau - J, \dots, \tau\}$ and a population $\mathcal{I}$.
Output: The counterfactual probability distribution $\hat{F}_t$, $\forall t \in \{\tau + 1, \dots, T\}$.
for $t \in \{\tau + 1, \dots, T\}$ **do**
  $\mathcal{C} \leftarrow N(k; \mathcal{S}, \mathcal{I})$ ▷ Construction of the control group.
  $\mathcal{A} \leftarrow \{i \in \mathcal{C} \,|\, d_{i,t} > \mathcal{D}_i(\mathcal{S})\}$ ▷ Subgroup with abnormal behavior.
  $\mathcal{V} \leftarrow \mathcal{C} \setminus \mathcal{A}$ ▷ Validated control group.
  if $|\mathcal{V}| \leq LB$ **then**
    $\mathcal{I} \leftarrow \mathcal{I} \setminus \mathcal{A}$ ▷ Remove abnormal peers from the population.
    goto 4
  $\hat{F}_t \leftarrow F_t(\cdot\,; \mathcal{V}, \mathcal{S})$
  if $\hat{F}_t(y_{i,t}) \in [\frac{p_0}{2}, 1 - \frac{p_0}{2}]$ **then** ▷ Verification procedure
    $\mathcal{S} \leftarrow \{t - J, \dots, t\}$ ▷ Update verified training sample.

where, $1 - p_0$ is the confidence level obtained with the non-parametric probabilistic counterfactual.

4
Results

The data for our case study was obtained from a merchant acquirer, one of the largest in Brazil. We use anonymized transactional time series (number of transactions per day) to construct the potential outcome of each retailer of interest using the proposed NPPC. The objective of the case study is to assess the accuracy of the NPPC whereas the retailer interest is free of any intervention. This way we can measure the prediction error with a mitigated causal-effects of observable or non-observable interventions. For instance in Figure 1.1, the difference between the counterfactual and observations can be interpreted and measured as prediction errors. On the other hand, in Figure 1.3, the difference between the counterfactual and observations is associated with the causal effect of the observable intervention of a price renegotiation. Similarly, in Figure 1.2, the difference reflects a churn behavior due to a non-observable intervention (e.g., competitors lowering prices). Next, we describe the sample selection to mitigate causal-effects and measure accuracy and probabilistic nowcasting point power of the proposed NPPC.

4.1
Sample Selection

To achieve the research objectives, the first stage of this study involved the basic process of extracting raw data, processing and grouping the data. The data set was extracted directly from the company's enterprise resource planning (ERP) software using Microsoft SQL Server (DATE; DARWEN, 1987). We used Python (ROSSUM; JR, 1995) for descriptive, statistical analysis and to construct the NPPC. Some open-source packages of this language were used, such as pandas (MCKINNEY et al., 2010), numpy (HARRIS et al., 2020), scikit-learn (PEDREGOSA et al., 2011), scipy (VIRTANEN et al., 2020) and matplotlib (HUNTER, 2007). We will describe the data set in more detail below.

The goal is to design a case study to measure the prediction power of the NPPC. For that we must mitigate the causal-effect of observable and non-observable interventions by properly selection a sample evaluated. Essentially we perform a backtest on past observations of selected retailers whose "future" is already known. We use forward-looking information to create *ad hoc* filters in order to regulate the intervention and evaluate its effect. This is accomplished by choosing samples based on particular criteria:

1. We choose active retailers who have been using the company's services for at least six months, starting from the first evaluation date, in $t \in \mathcal{T}$, and who have active status on the last evaluation date of the framework, in $t \in \mathcal{T}$. This rule is implemented to exclude hard ("observable") churn from the case study;

2. We exclude retailers who stayed idle for seven days in a row during the time window $\mathcal{T}$, since we focus on retailers with relatively high transactional frequency. Therefore we either exclude low-frequency or churned retailers.;

3. We only select retailers with an average transaction per transaction day of three or more, in $t \in \mathcal{T}$, because retailers with fewer and/or erratic transactions are frequently mistaken with churn retailers.

4. We also eliminate retailers who had a Z-score greater than 2 standard deviations with a confidence level of 95% for three consecutive days in $t \in \mathcal{T}$. This filter works as a simple outlier detection to exclude causal effect of other non-observed interventions;

Through the use of these criteria, we were able to limit the usage of counterfactuals to around 17% of the users from a pool of around 560,000 retailers. Figure 4.1 depicts the original number of consumers in the first selection and the number of retailers who stayed in the final evaluation.

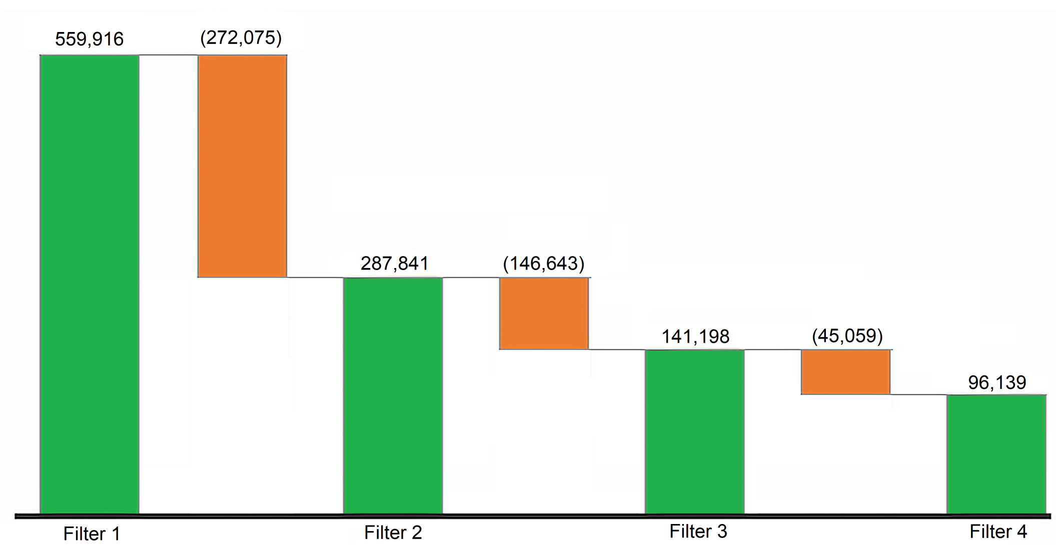


Figure 4.1: The amount of samples in each filter is indicated via a waterfall chart.

The proposed methodology is validated using two metrics: 1. A hypothesis test is used to evaluate the model, and 2. An accuracy metric is used to compare the nowcasting error to a naïve forecasting. The latter will be discussed in detail in the next section.

4.2 Nowcasting point assessment

In terms of the accuracy metric, we want to compare the nowcasting point of the proposed methodology with a naive approach. We consider the median of the NPPC as a point nowcasting point and the the average transaction level as the naive forecast. The naive forecast varies with the rolling horizon since, for each time step, we compute the average of the corresponding training sample. As an error metric, we consider Median Relative Absolute Error (MdRAE) since it is more robust than outliers as other error metrics (e.g, MAE, RMSE). Formally this is given by

$$\text{MdRAE} = \text{Median}_{t=1,\dots,T}\left(\frac{|y_{i,t} - \widehat{y}_{i,t}|}{|y_{i,t} - \bar{y}_i|}\right) \tag{4-1}$$

where $\bar{y}_i$ is the in-sample average of the transactional of the store of interest, $y_{i,t}$ is the merchant transactional and $\widehat{y}_{i,t}$ is the NPPC median, here considered as a nowcasting point.

The difference between the observed transactional of the interest retailer and the methodology's nowcasting represents the methodology's measure of error, whereas the denominator is the difference between the observed transactional and the in-sample average of the merchant's transactional and represents the naive forecasting error.

The median of the ratio between the absolute error of our nowcasting point and the absolute error of a benchmark model is calculated using the Median Relative Absolute Error (MdRAE). The metric is 1 if our model's nowcasting point equals the benchmark's forecast. If the benchmarks' forecasts are more accurate than ours, the outcome will be greater than 1. If ours is superior, it is less than 1. Division by zero is an issue for MdRAE. To alleviate this problem, we follow Armstrong & Collopy (1992) recommendation to exclude extreme values. We summarize the results displayed in Table 4.1, with the tiers being ordered in an ascending way, that is, Tier A are the small retailers and, Tier D, the longe ones. In our experiment, we consider a training window of $J = 60$ has 60 days of transactions, and the counterfactual is applied in $T = 63$ days.

In Table 4.1, Tier A is the group of consumers with the lowest amount of transactions, Tier B has intermediate features, Tier C are retailers with a profile above average, and Tier D is the company's key retailers; the Retailers column shows how many consumers were analyzed for each Tier. Minimum

Table 4.1: Summary of distribution of the MdRAE application, to check if the counterfactual is better than the benchmark.

Tier	Retailers	Minimum	Q25	Mean	Median	Q75	Q90	Maximum
A	20,815	0.11	0.77	0.87	0.91	1.01	1.09	4.31
B	52,169	0.04	0.67	0.79	0.85	0.97	1.04	3.52
C	19,626	0.03	0.53	0.73	0.74	0.91	1.01	2.57
D	3,529	0.03	0.48	0.65	0.65	0.83	0.97	4.14

denotes the lowest recorded MdRAE value per Tier; the first, second, and third quartiles of the MdRAE distribution, per tier, are represented as Q25, median, and Q75, respectively. Mean is the average of the MdRAE distribution per tier; Q90 is the 90th percentile of the MdRAE distribution per tier; and Maximum is the highest observed MdRAE, per tier. According to the results of this table, the suggested model outperformed the naive model for the vast majority of retailers, with 72% for Tier A, 82% for Tier B, 89% for Tier C and 92% for Tier D.

A graphical visualization, through a boxplot, of the result of estimating counterfactual through our methodologies is shown in Figure 4.2, related to the accuracy metric that checks whether the counterfactual is better than the benchmark (naive model). It was observed that, regardless of the retailer's tier, the proposed model outperforms the naive model in 81.63% of the samples, given the position of the third quartile, in Figure 4.2. As the retailer has a higher revenue profile, the model has better accuracy. This is explained by the number of transactions for each retailer profile, as a small retailer has few daily transactions and a variation of one or two units represents a lot, whereas for larger retailers this becomes imperceptible.

4.3 Probabilistic nowcasting point assessment

The objective of this section is to assess the probabilistic validity of the prediction interval. For that, we use the Kupiec's statistical test to asses the proportion of failures test for value-at-risk (VaR) backtesting (POF) (KUPIEC et al., 1995). This test is extensively used in evaluate the financial risk measure Value-at-Risk (VaR) on a backtesting scheme. To do so, calculate $\hat{p}$, which is the number of times the observed value falls beyond the nowcasting point range. Let us define the indicator function for

$$I_t = \begin{cases} 0, & \text{if } \hat{F}_t(y_t) \in [\alpha, (1-\alpha/2)] \\ 1, & \text{otherwise} \end{cases}$$

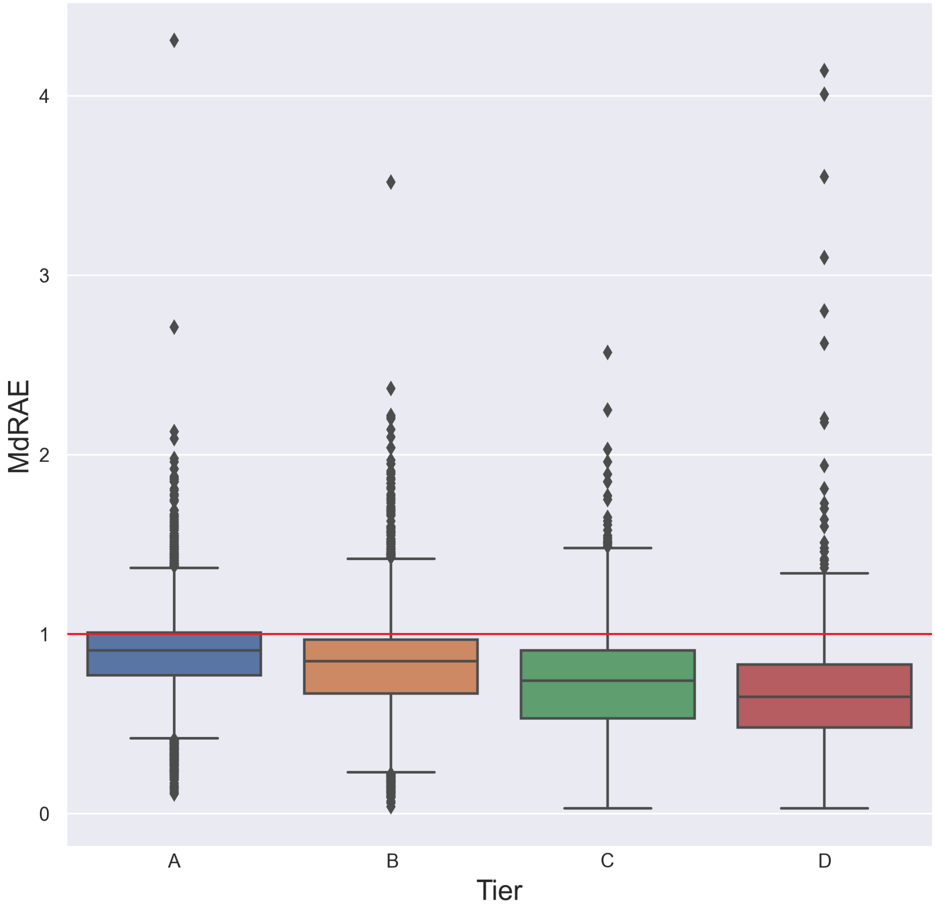


Figure 4.2: MdRAE boxplot for each tier. MdRAE = 1 is indicated by the red line.

for each $t = 1, \ldots, T$ and given a significance level of 10 percent. Then, we define the average estimator

$$\hat{p} = \frac{1}{T} \sum_{t=1}^{T} I_t$$

and its $100(1-\alpha)$ test acceptance interval

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \tag{4-2}$$

The test range in hypothesis testing is centered on the null hypothesis value. We used a 95 % confidence level for model evaluation. In a summary, we perform the hypothesis test as follows:

$$\begin{aligned} \mathrm{H}_0 &: p = p_0 \\ \mathrm{H}_1 &: p \neq p_0 \end{aligned} \tag{4-3}$$

where the null hypothesis denotes a well-fitted model.

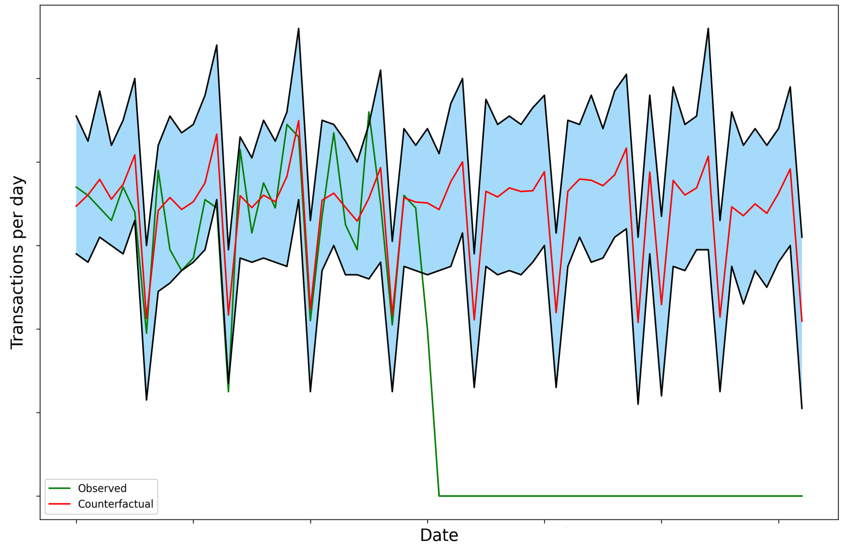
The probabilistic prediction metric, which tests the hypothesis to verify the fit of the model, 69.46% of the samples fitted the model well, with a confidence level of 95%. The results of the hypothesis test are shown in Table 4.2.

Table 4.2: Summary of the hypothesis test, where the null hypothesis denotes a well-fitted model.

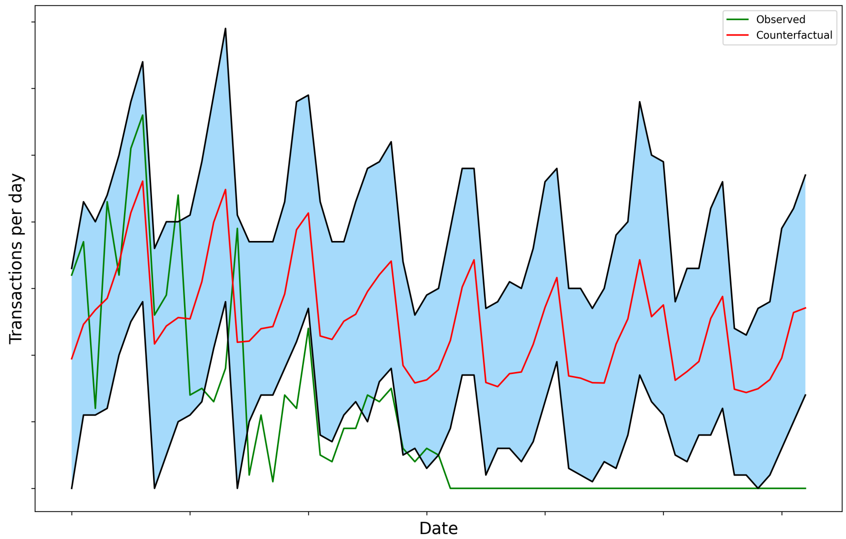
Tier	Retailers	Percent
A	20,815	72.83%
B	52,169	70.04%
C	19,626	65.62%
D	3,529	62.40%

In this case, the lower the transactional profile of the retailer, the better the adjustment of the proposed methodology. This is explained by the sample set of each profile and the amplitude of the classification of each tier, where the greater the profile, the greater the amplitude and, as a result, the fewer similar transactional profiles exist for the construction of the NPPC.

Figures 4.3-(a) and (b) and Figures 4.4-(a) and (b) supplement the initial illustrations by displaying additional examples of observable (price renegotiation) and unobservable (competition) interventions. In some cases, churn occurs quickly, as shown in Figure 4.3-(a) and (b), as opposed to the situation shown in Figure 1.2, but the proposed methodology still detects the intervention well due to the consistent violation of the prediction intervals. Figures 4.4-(a) and (b), on the other hand, show the retailer's transactional potential after a price renegotiation, allowing us to measure the retailer's price elasticity.

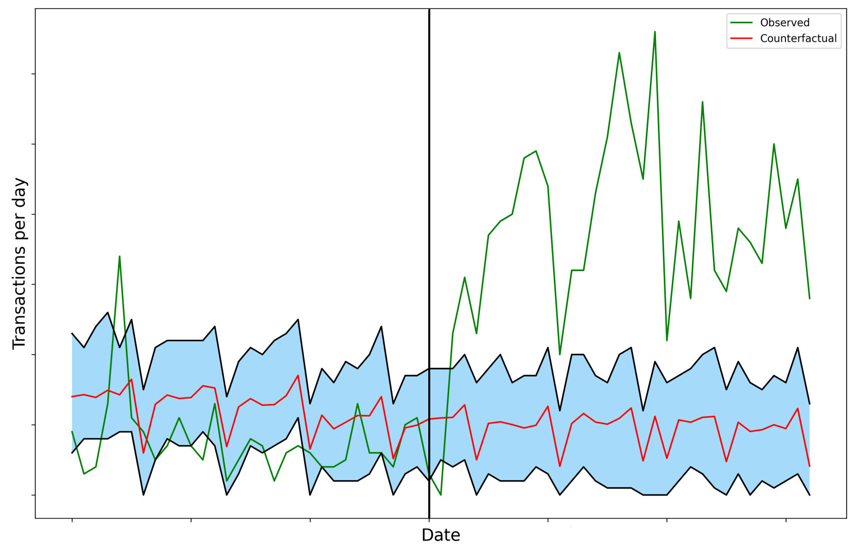


(a) Application of the NPPC for a competitive intervention (fast churn).

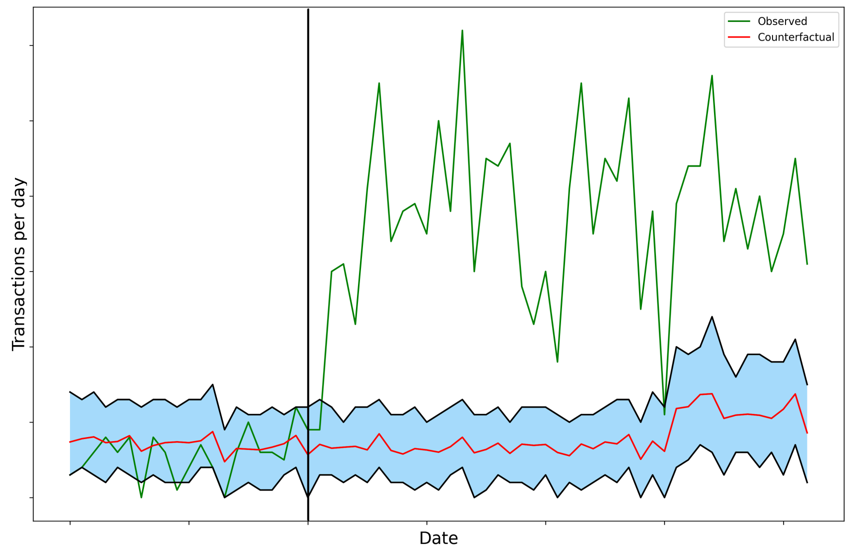


(b) Application of the NPPC for a competitive intervention (slow churn).

Figure 4.3: Application of the NPPC for unobservable interventions (churn).



(a) Application of the NPPC for a price renegotiation intervention.



(b) Application of the NPPC for a price renegotiation intervention.

Figure 4.4: Application of NPPC for observable interventions (price renegotiation). Timing of intervention is indicated by the black line.

5
Conclusion

We developed a complete framework to construct a non-parametric probabilistic counterfactual (NPPC) for a retailer's transactional potential based solely on the transactional data of validated peers. We proposed an adaptive rolling horizon framework to dynamically identify locally abnormal behaviors for a given retailer due to an observable or non-observable intervention. Our approach provided a probabilistic counterfactual range in which the retailer's daily transaction value should be found if no intervention occurs.

The proposed framework starts at a past time window when we assume the retailer is free of interventions. Then, we perform a rolling horizon scheme whereby, for each time window, we perform a 4-step process: (i) the construction of the control group via KNN; (ii) the validation of the control group for out-of-sample observations; (iii) the construction of the NPPC; (iv) the verification of the statistical significance of the causal-effect.

We conducted a case study to determine the proposed framework's accuracy using the MdRAE, and we validated the probabilistic prediction interval using Kupiec's statistical test. We measured the nowcasting error and applied some filters to the base to remove retailers with intervention. We analyzed 96,139 retailers were evaluated, with 20,815 being Tier A, 52,169 being Tier B, 19,626 being Tier C, and 3,529 being Tier D, where A represents the customer with the lowest transactional profile and D represents large retailers. According to the MdRAE, the proposed methodology outperformed a naive model by 72%, 82%, 89%, and 92% for Tiers A to D, respectively. The statistical test, in turn, indicated a well-adjusted model for Tiers A to D at 72.83%, 70.04%, 65.62%, and 62.40%, respectively.

We argue that the proposed framework addresses an underlying issue of many revenue management problems of a retailer. For a merchant acquirer, an accurate assessment of retailers' transactional potential is crucial to support managerial decisions on pricing, credit, and retention campaigns. These decisions can be interpreted as interventions, therefore their causal effects on revenue can be properly estimated. Our work leaves open many interesting directions for future research, such as the optimization of hyperparameters related to the number of days to be considered in the training window $\mathcal{S}$. The number of neighbors to be chosen at each step and the size of the initial control group. Finally, this framework can be used in other applications, particularly, in economics and public policy problems.

6
Bibliography

ABADIE, A.; DIAMOND, A.; HAINMUELLER, J. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. **Journal of the American statistical Association**, Taylor & Francis, v. 105, n. 490, p. 493–505, 2010. Cited 3 times in pages 15, 20, and 22.

ABADIE, A.; DIAMOND, A.; HAINMUELLER, J. Comparative politics and the synthetic control method. **American Journal of Political Science**, Wiley Online Library, v. 59, n. 2, p. 495–510, 2015. Cited 2 times in pages 15 and 22.

ABADIE, A.; GARDEAZABAL, J. The economic costs of conflict: A case study of the basque country. **American economic review**, v. 93, n. 1, p. 113–132, 2003. Cited in page 20.

ALBADVI, A.; VARASTEH, S. Customers churn behavior modeling using decision trees (a case study in non-contractual setting). **Advances in Industrial Engineering**, University of Tehran, v. 44, n. 2, p. 127–139, 2010. Cited in page 20.

ARMSTRONG, J. S.; COLLOPY, F. Error measures for generalizing about forecasting methods: Empirical comparisons. **International journal of forecasting**, Elsevier, v. 8, n. 1, p. 69–80, 1992. Cited in page 34.

BABKIN, A.; GOLDBERG, I. Incorporating time-dependent covariates into bg-nbd model for churn prediction in non-contractual settings. **Available at SSRN 2905307**, 2017. Cited in page 18.

BEEHARRY, Y.; FOKONE, R. T. Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. **Concurrency and Computation: Practice and Experience**, Wiley Online Library, v. 34, n. 4, p. e6627, 2022. Cited 2 times in pages 15 and 17.

BERTSIMAS, D.; KALLUS, N. The power and limits of predictive approaches to observational-data-driven optimization. **arXiv preprint arXiv:1605.02347**, 2016. Cited in page 22.

BERTSIMAS, D.; KALLUS, N. From predictive to prescriptive analytics. **Management Science**, INFORMS, v. 66, n. 3, p. 1025–1044, 2020. Cited 2 times in pages 20 and 22.

BERTSIMAS, D. et al. Personalized diabetes management using electronic medical records. **Diabetes care**, Am Diabetes Assoc, v. 40, n. 2, p. 210–217, 2017. Cited in page 20.

BUCKINX, W.; POEL, D. Van den. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual fmcg retail setting. **European journal of operational research**, Elsevier, v. 164, n. 1, p. 252–268, 2005. Cited 3 times in pages 15, 17, and 18.

CAIGNY, A. D. et al. Incorporating textual information in customer churn prediction models based on a convolutional neural network. **International Journal of Forecasting**, Elsevier, v. 36, n. 4, p. 1563–1578, 2020. Cited 3 times in pages 15, 17, and 18.

CALCIU, M. Deterministic and stochastic customer lifetime value models. evaluating the impact of ignored heterogeneity in non-contractual contexts. **Journal of Targeting, Measurement and Analysis for Marketing**, Springer, v. 17, n. 4, p. 257–271, 2009. Cited 2 times in pages 18 and 19.

CARVALHO, C.; MASINI, R.; MEDEIROS, M. C. Arco: An artificial counterfactual approach for high-dimensional panel time-series data. **Journal of econometrics**, Elsevier, v. 207, n. 2, p. 352–380, 2018. Cited 2 times in pages 20 and 21.

CHAMON, M.; GARCIA, M.; SOUZA, L. Fx interventions in brazil: a synthetic control approach. **Journal of International Economics**, Elsevier, v. 108, p. 157–168, 2017. Cited in page 20.

CHOU, P. et al. Predictive analytics for customer repurchase: Interdisciplinary integration of buy till you die modeling and machine learning. **European Journal of Operational Research**, Elsevier, v. 296, n. 2, p. 635–651, 2022. Cited in page 15.

COATE, J. A. et al. A machine learning method of determining causal inference applied to shifts in voting preferences between 2012-2016. **SMU Data Science Review**, v. 5, n. 1, p. 2, 2021. Cited in page 21.

DASGUPTA, K.; MASON, B. J. The effect of interest rate caps on bankruptcy: Synthetic control evidence from recent payday lending bans. **Journal of Banking & Finance**, Elsevier, v. 119, p. 105917, 2020. Cited in page 20.

DATE, C. J.; DARWEN, H. **A Guide to the SQL Standard**. [S.l.]: Addison-Wesley New York, 1987. v. 3. Cited in page 32.

DAWES, J.; SWAILES, S. Retention sans frontieres: issues for financial service retailers. **International Journal of Bank Marketing**, MCB UP Ltd, 1999. Cited in page 16.

EDELSTEIN, H. Building profitable customer relationships with data mining. In: **Customer Relationship Management**. [S.l.]: Springer, 2000. p. 339–351. Cited in page 16.

FADER, P. S.; HARDIE, B. G. Probability models for customer-base analysis. **Journal of interactive marketing**, Elsevier, v. 23, n. 1, p. 61–69, 2009. Cited in page 16.

FERMAN, B.; PINTO, C.; POSSEBOM, V. Cherry picking with synthetic controls. **Journal of Policy Analysis and Management**, Wiley Online Library, v. 39, n. 2, p. 510–532, 2020. Cited in page 20.

GLADY, N.; BAESENS, B.; CROUX, C. Modeling churn using customer lifetime value. **European Journal of Operational Research**, Elsevier, v. 197, n. 1, p. 402–411, 2009. Cited 2 times in pages 18 and 19.

GYÖRFI, L. et al. **A distribution-free theory of nonparametric regression**. [S.l.]: Springer, 2002. v. 1. Cited 5 times in pages 16, 23, 24, 25, and 26.

HARRIS, C. R. et al. Array programming with NumPy. **Nature**, v. 585, p. 357–362, 2020. Cited in page 32.

HOPMANN, J.; THEDE, A. Applicability of customer churn forecasts in a non-contractual setting. In: **Innovations in classification, data science, and information systems**. [S.l.]: Springer, 2005. p. 330–337. Cited in page 18.

HÖPPNER, S. et al. Profit driven decision trees for churn prediction. **European journal of operational research**, Elsevier, v. 284, n. 3, p. 920–933, 2020. Cited 2 times in pages 15 and 17.

HUNTER, J. D. Matplotlib: A 2d graphics environment. **Computing in science & engineering**, IEEE, v. 9, n. 3, p. 90–95, 2007. Cited in page 32.

IMANI, A. et al. Customer segmentation to identify key customers based on rfm model by using data mining techniques. **International journal of research in industrial engineering, Volume**, 2021. Cited in page 19.

JAHROMI, A. T. et al. Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. **Journal of Strategic Marketing**, Taylor & Francis, v. 18, n. 7, p. 587–598, 2010. Cited in page 18.

JAHROMI, A. T.; STAKHOVYCH, S.; EWING, M. Managing b2b customer churn, retention and profitability. **Industrial Marketing Management**, Elsevier, v. 43, n. 7, p. 1258–1268, 2014. Cited 4 times in pages 15, 16, 17, and 18.

JIANG, Y. et al. Ultra large-scale crowd monitoring system architecture and design issues. **IEEE Internet of Things Journal**, IEEE, v. 8, n. 13, p. 10356–10366, 2021. Cited in page 16.

JUENGERKES, B. E. Fintechs and banks–collaboration is key. **The FinTech book: The financial technology handbook for investors, entrepreneurs and visionaries**, Wiley Online Library, p. 179–182, 2016. Cited in page 10.

KARMAKAR, B. et al. Improved retention analysis in freemium role-playing games by jointly modelling players' motivation, progression and churn. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, Wiley Online Library, 2022. Cited 2 times in pages 15 and 17.

KIM, S. et al. Churn prediction of mobile and online casual games using play log data. **PloS one**, Public Library of Science San Francisco, CA USA, v. 12, n. 7, p. e0180735, 2017. Cited 3 times in pages 15, 17, and 18.

KUMAR, H.; YADAV, R. K. Rule-based customer churn prediction model using artificial neural network based and rough set theory. In: **Soft Computing: Theories and Applications**. [S.l.]: Springer, 2020. p. 97–108. Cited 2 times in pages 15 and 17.

KÜNZEL, S. R. et al. Metalearners for estimating heterogeneous treatment effects using machine learning. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 116, n. 10, p. 4156–4165, 2019. Cited in page 20.

KUPIEC, P. H. et al. **Techniques for verifying the accuracy of risk measurement models**. [S.l.]: Division of Research and Statistics, Division of Monetary Affairs, Federal . . . , 1995. v. 95. Cited in page 35.

LEE, E. et al. Game data mining competition on churn prediction and survival analysis using commercial game log data. **arXiv preprint arXiv:1802.02301**, 2018. Cited in page 18.

LEMMENS, A.; GUPTA, S. Managing churn to maximize profits. **Marketing Science**, INFORMS, v. 39, n. 5, p. 956–973, 2020. Cited 2 times in pages 15 and 17.

LORIA, E.; MARCONI, A. Exploiting limited players' behavioral data to predict churn in gamification. **Electronic Commerce Research and Applications**, Elsevier, v. 47, p. 101057, 2021. Cited in page 17.

MANDIĆ, M.; KRALJEVIĆ, G.; BOBAN, I. Performance comparison of six data mining models for soft churn customer prediction in telecom. **IJEEC- International Journal of Electrical Engineering and Computing**, v. 2, n. 1, 2018. Cited in page 17.

MARTÍNEZ, A. et al. A machine learning framework for customer purchase prediction in the non-contractual setting. **European Journal of Operational Research**, Elsevier, v. 281, n. 3, p. 588–596, 2020. Cited in page 18.

MASINI, R.; MEDEIROS, M. C. Counterfactual analysis with artificial controls: Inference, high dimensions, and nonstationarity. **Journal of the American Statistical Association**, Taylor & Francis, v. 116, n. 536, p. 1773–1788, 2021. Cited in page 15.

MATUSZELAŃSKI, K.; KOPCZEWSKA, K. Customer churn in retail e-commerce business: Spatial and machine learning approach. **Journal of Theoretical and Applied Electronic Commerce Research**, Multidisciplinary Digital Publishing Institute, v. 17, n. 1, p. 165–198, 2022. Cited 4 times in pages 15, 16, 17, and 18.

MCKINNEY, W. et al. Data structures for statistical computing in python. In: AUSTIN, TX. **Proceedings of the 9th Python in Science Conference**. [S.l.], 2010. v. 445, p. 51–56. Cited in page 32.

MILOŠEVIĆ, M.; ŽIVIĆ, N.; ANDJELKOVIĆ, I. Early churn prediction with personalized targeting in mobile social games. **Expert Systems with Applications**, Elsevier, v. 83, p. 326–332, 2017. Cited 3 times in pages 15, 17, and 18.

MIN, S. et al. Customer acquisition and retention spending: An analytical model and empirical investigation in wireless telecommunications markets. **Journal of marketing research**, SAGE Publications Sage CA: Los Angeles, CA, v. 53, n. 5, p. 728–744, 2016. Cited in page 15.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **Journal of machine learning research**, v. 12, n. Oct, p. 2825–2830, 2011. Cited in page 32.

PEPPERS, D.; ROGERS, M.; DORF, B. Is your company ready for one-to-one marketing. **Harvard business review**, v. 77, n. 1, p. 151–160, 1999. Cited in page 16.

PERIŠIĆ, A.; JUNG, D. Š.; PAHOR, M. Churn in the mobile gaming field: Establishing churn definitions and measuring classification similarities. **Expert Systems with Applications**, Elsevier, v. 191, p. 116277, 2022. Cited in page 15.

ROSSUM, G. V.; JR, F. L. D. **Python reference manual**. [S.l.]: Centrum voor Wiskunde en Informatica Amsterdam, 1995. Cited in page 32.

SOUZA, M. Predictive counterfactuals for treatment effect heterogeneity in event studies with staggered adoption. **Available at SSRN 3484635**, 2019. Cited in page 21.

STONE, C. J. Consistent nonparametric regression. **The annals of statistics**, JSTOR, p. 595–620, 1977. Cited in page 26.

STUART, E. A. Matching methods for causal inference: A review and a look forward. **Statistical science: a review journal of the Institute of Mathematical Statistics**, NIH Public Access, v. 25, n. 1, p. 1, 2010. Cited 4 times in pages 16, 21, 22, and 28.

VERMA, S.; DICKERSON, J.; HINES, K. Counterfactual explanations for machine learning: A review. **arXiv preprint arXiv:2010.10596**, 2020. Cited in page 20.

VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. **Nature Methods**, v. 17, p. 261–272, 2020. Cited in page 32.

VIVIANO, D.; BRADIC, J. Synthetic learner: model-free inference on treatments over time. **arXiv preprint arXiv:1904.01490**, 2019. Cited in page 21.

WACHTER, S.; MITTELSTADT, B.; RUSSELL, C. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. **Harv. JL & Tech.**, HeinOnline, v. 31, p. 841, 2017. Cited in page 19.

WANG, Q. **Design the churn analysis on games: a review on techniques for churn analysis**. Tese (Doutorado) — Northeastern University, 2018. Cited in page 17.

YAO, L. et al. A survey on causal inference. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM New York, NY, USA, v. 15, n. 5, p. 1–46, 2021. Cited 2 times in pages 16 and 21.

YOON, S.; KOEHLER, J.; GHOBARAH, A. Prediction of advertiser churn for google adwords. 2010. Cited in page 18.