



HAL
open science

Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach

Arno de Caigny, Kristof Coussement, Wouter Verbeke, Khaoula Idbenjra, Minh Phan

► To cite this version:

Arno de Caigny, Kristof Coussement, Wouter Verbeke, Khaoula Idbenjra, Minh Phan. Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach. *Industrial Marketing Management*, 2021, 99, pp.28-39. 10.1016/j.indmarman.2021.10.001 . hal-03599615

HAL Id: hal-03599615

<https://hal.science/hal-03599615v1>

Submitted on 16 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

**UPLIFT MODELING AND ITS IMPLICATIONS FOR B2B CUSTOMER CHURN
PREDICTION: A SEGMENTATION-BASED MODELING APPROACH**

Arno De Caigny¹, Kristof Coussement¹, Wouter Verbeke², Khaoula Idbenjra¹, Minh
Phan¹

¹ IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie
Management, F-59000 Lille, France

² Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000, Leuven,
Belgium

E-mail addresses: a.de-caigny@ieseg.fr (Arno De Caigny); k.coussement@ieseg.fr (Kristof
Coussement); wouter.verbeke@kuleuven.be (Wouter Verbeke); k.idbenjra@ieseg.fr (Khaoula
Idbenjra); m.phan@ieseg.fr (Minh Phan)

Corresponding author: Kristof Coussement, +33 3 20 54 58 92, k.coussement@ieseg.fr

UPLIFT MODELING AND ITS IMPLICATIONS FOR B2B CUSTOMER CHURN PREDICTION: A SEGMENTATION-BASED MODELING APPROACH

Abstract

Business-to-business (B2B) customer retention relies heavily on analytics and predictive modeling to support decision making. Given this, we introduce uplift modeling as a relevant prescriptive analytics tool. In particular, the uplift logit leaf model offers a segmentation-based algorithm that combines predictive performance with interpretability. Applied to a real-world data set of 6,432 customers of a European software provider, the uplift logit leaf model achieves superior performance relative to three other popular uplift models in our study. The accessibility of output gained from the uplift logit leaf model also is showcased with a case study, which reveals relevant managerial insights. This new tool thus delivers novel insights in the form of customized, global, and segment-level visualizations that are especially pertinent to industrial marketing settings. Overall, the findings affirm the viability of uplift modeling for improving decisions related to B2B customer retention management.

Keywords: customer retention, churn prediction, uplift modeling, segmentation-based modeling, interpretability, visualization

1. Introduction

Customer retention involves continuous transactions with firms (Ascarza et al., 2018), which is important because attracting new customers can cost five to six times more than retaining existing ones (Colgate & Danaher, 2000). Such considerations are especially relevant in business-to-business (B2B) settings, in which 68% of executive-level managers believe customers are becoming less loyal (Michels & Dullweber, 2014), the average transaction value per customer is high and the pool of customers who account for bigger portions of sales is often smaller than in B2C (Rauyrueen et al., 2011). Thus, a loss of even one customer can cause significant losses of future sales revenue. Because each customer is important (Stevens, 2005), the proactive identification of those customers at risk of churning is crucial. For example, Ascarza et al. (2018) propose a framework of questions to help firms design effective retention strategies, according to “Who is at risk and why?”, “Whom should we target and with what incentives?”, “When do we target?”, and “What can we gain?” Building on this framework, we introduce a segmentation-based uplift modeling architecture (“Whom do we target?”) that can go beyond churn predictions (“Who is at risk?”) to enhance customer retention efforts. The raise of big data in B2B settings enables this analytical approach of customer retention management (Paschen et al., 2019). Furthermore, we acknowledge the benefits of using artificial intelligence (AI) techniques in B2B settings and specify how they can deliver decision support for retention managers (Jabbar et al., 2020) and inform targeted B2B marketing campaigns (Liu, 2020).

Data-driven algorithmic churn detection approaches for B2B customer retention management have received relatively little prior academic attention (Martínez-López & Casillas, 2013; Tamaddoni Jahromi et al., 2014; Wiersema, 2013; Yu et al., 2011). These data-driven algorithmic approaches mainly involve *predictive analytics* that underlie traditional churn prediction models, designed to identify customers at risk of ending their relationship with the firm (Tamaddoni Jahromi et al., 2014). They rely on measures of customers' past behavior to rank them, from most to least likely to churn, so that a firm can target some retention effort toward those most likely to leave. But traditional churn prediction models are not fully aligned with key business objectives, including maximizing the return on investments in retention campaigns, because they predict only a single outcome: churning versus non-churning (Devriendt, Berrevoets, et al., 2020). Furthermore, they cannot address the vast challenges associated with B2B customer relationship management, as result from the complexity of communication (Michels & Dullweber, 2014), the use of “just-the-facts” approaches instead of emotional or entertaining marketing (Russo et al., 2016), and personalization and customization demands (Harrison-Walker & Neeley, 2004). If a churn prediction model predicts a customer is about to end its relationship with the focal firm, it cannot reveal whether this customer might be persuaded to change its mind; such assessments require uplift or net-effect churn models, as part of *prescriptive analytics*. The goal of prescriptive analytics is to indicate what has to be done to obtain a desired outcome. Prescriptive analytics is related to descriptive and predictive analytics, but the focus is on actionability rather than describing or predicting. Uplift modeling is prescriptive in the sense that, unlike traditional churn prediction models that answer “Who is at risk?”, these techniques seek to estimate the net effect of retention campaigns to answer the actionable question of “Whom do we target?”. That is, the goal is to identify customers about to churn but that might be retained by a targeted campaign, to optimize the selection of customers to target, according to an estimate of the positive change in customer reactions to different retention actions. In practice, uplift models use randomized experiments and divide customers into treatment and control groups, which reveals the heterogeneity in customers’ responses to retention offers. Then the uplift modeling process applies the results of randomized experiments to the entire customer base.

We gather a real-life uplift churn data set of 6,432 customers—1,399 of which received a retention campaign and 5,033 that did not—of a European B2B software provider and thereby seek to extend B2B churn prediction literature in three ways. First, though various traditional churn prediction approaches have been discussed in B2B contexts (Chen et al., 2015; Gordini & Veglio, 2017; Hopmann & Thede, 2005; Tamaddoni Jahromi et al., 2014), none of them include the diverse effect of a retention intervention on customers’ churn behavior. A proper selection of customers for a retention campaign in B2B is of crucial importance for the return on marketing investment (ROMI) because incentivizing B2B customers is often costly. Recent B2C research, however, has shown that incorporating customers in a retention campaign based on their churn probability ranking is less effective than ranking customers based on their sensitivity to the intervention (Devriendt, Berrevoets,

et al., 2020). Indeed, there might be customers with high risk of churning for whom a retention campaign has no impact, and thus should not be targeted (Ascarza, 2018). Hence, uplift modeling is an important decision-making tool for customer retention management. We address this research gap and introduce uplift modeling to a B2B customer churn setting. Second, recent data science research describes modeling approaches that split the customer base into various segments, then apply segment-specific prediction models (e.g. De Caigny et al., 2018). Such segmented prediction approaches provide more granular, segment-specific insights into the drivers of the prediction problem, with no sacrifice of predictive performance. Along these lines, we introduce a new hybrid, segmented, uplift modeling approach, called the uplift logit leaf model (uplift LLM), which combines uplift decision trees (uplift DT) to segment the customer base with uplift logistic regression (uplift LR) models to deliver segment-specific predictions and insights. To gauge its performance, we also compare the performance of ~~benchmark~~ our proposed uplift LLM against uplift DT (Rzepakowski & Jaroszewicz, 2012a), uplift LR (Lo, 2002; Radcliffe & Surry, 1999), and uplift random forests (uplift RF) (Guelman et al., 2012). Third, we respond to calls for insights into predictive modeling algorithms that combine data science outcomes with human decision makers (Paschen et al., 2019). In particular, interpretable data science literature is keen to provide algorithmic insights to marketing decision makers in the form of visualizations (e.g. Hanssens & Pauwels, 2016; Vellido, 2019; Zhang & Zhu, 2018), but prior uplift models do not explicitly aim to deliver direct insights about which customer metrics drive their predictions. Therefore, we contribute to both uplift modeling and interpretable data science literature by proposing new uplift visualizations at the segment level.

In the next section, we review customer churn prediction in the B2B context and introduce uplift churn modeling to B2B settings. Section 3 outlines the uplift LLM and its building blocks, and then in Section 4, we present the experimental design. Section 5 zooms in on the prediction results of the uplift LLM, relative to other considered uplift algorithms, and provides a visualization to reveal key drivers, which managers can use to improve their decision making. Finally, we conclude with some limitations and areas for further research in Section 6.

2. Customer retention management

2.1. Customer churn prediction in a B2B context

Customer retention management in B2B settings is distinct from business-to-consumer (B2C) contexts, for two main reasons. First, a group of people (rather than a single consumer) makes purchase decisions. In industrial settings, those decisions rely on cost and budget considerations, and the budget owner often makes the final call (Wind & Webster, 1972). Also, Williams et al. (2011) conclude that price sensitivity is the most notable difference between churners and stayers in B2B services. A common practice in contractual settings is to offer a discount at renewal time to encourage potential churners to renew their contract (Lemmens & Gupta, 2020). Targeted campaigns offering a discount towards churners, being more price sensitive (Williams et al., 2011), could thus be successful

in B2B. Second, companies in B2B contexts often make larger purchases with greater transactional value (Rauyruen et al., 2011). As a result, the customer churn dynamics in B2B settings can differ from those in B2C settings, suggesting the need to develop analytical models specifically tailored to B2B settings. The need for customer churn prediction is high when the nature of B2B contexts with larger purchase amounts and more transactions is considered (Rauyruen & Miller, 2007; Tamaddoni Jahromi et al., 2014). Although extensive literature details customer churn prediction in B2C contexts (e.g. De Caigny et al., 2018; Verbeke et al., 2012), parallel research on the topic in B2B settings is relatively scarce, as the list in Table 1 suggests. This summary overview also indicates two research gaps.

First, Schmittlein et al. (1987) distinguish non-contractual from contractual customer relationships. The contractual relationships in B2B are often characterized by the other party's requirements, and are thus relatively stable and long-lasting (Ford, 2004). Contractual settings typically have a two-clock nature (Ascarza & Hardie, 2013). This means that usage of the services and the renewal happen on a different time scale. For example, the usage of software service happens daily, while the contract renewal decision is yearly. This allows to predict churn at the renewal point based on observable usage patterns while under contract, and permits companies to develop a clean retention management strategy. If companies succeed in improving retention rates, profits increase significantly (Degbey, 2015). In our study, we focus on this contractual setting, but Table 1 indicates that existing studies of customer churn detection in B2B settings only focus on non-contractual settings. Second, in terms of uses of predictive or prescriptive analytics frameworks, we find mainly the former. Prescriptive analytics offer suggestions of what to do to optimize some outcome of interest, rather than predicting what is likely to happen and leaving the decision (or optimization) up to users. In methodological data science terminology, this classification reflects the difference between uplift or net-effect modeling (*prescriptive analytics*) and traditional customer churn prediction (*predictive analytics*) (Lo, 2002; Radcliffe & Surry, 1999). As Table 1 shows, existing studies in a B2B customer churn context focus on predictive analytics, such as when Tamaddoni Jahromi et al. (2014) compare multiple churn prediction models (i.e. random classifier, decision trees, logistic regression and boosting) applied to 11,021 business customers of an Australian online retailer. They confirm that boosting is the best prediction model, such that this model-driven approach to churn prediction and retention strategy development outperforms commonly used managerial heuristics. Chen et al. (2015) use a data set of 69,170 business customers of a Taiwanese logistics company to identify the most valuable businesses, as well as the riskiest customers, across four churn prediction models (i.e. decision trees, logistic regression, neural networks and support vector machines). They argue that decision trees perform best. In their effort to identify the benefits of churn prediction modeling, Gordini and Veglio (2017) study an Italian online retailer with 80,000 business customers and assert that their proposed support vector machines model outperforms logistic regression, neural networks, and support vector machines. Finally, Barfar et al. (2017) show that both

rationality and bounded rationality have significant roles in predicting churn, according to a data set of 100,000 small- and medium-sized companies. We add to these collected insights by investigating the benefit of using uplift modeling, as an AI-driven decision support method that can move beyond predictive analytics. Our findings address questions about the standard practice of selecting customers for a retention campaign solely on the basis of their likelihood to churn, in that customers who are most likely to churn are not necessarily those that will respond positively to retention campaigns. Therefore, we investigate whether, as some recent insights in customer retention management suggest (e.g. Ascarza, 2018; Devriendt, Berrevoets, et al., 2020), companies need to focus on identifying observed heterogeneity in responses to retention campaigns, regardless of churn risk.

Study	Industry	Number of customers	B2B Context		Research Type	
			Non-Contractual	Contractual	Predictive	Prescriptive
(Tamaddoni Jahromi et al., 2014)	Retailing	11,021	X		X	
(Chen et al., 2015)	Logistics	69,170	X		X	
(Gordini & Veglio, 2017)	E-commerce	80,000	X		X	
(Barfar et al., 2017)	Services	100,000	X		X	
Our study	IT services	6,432		X		X

Table 1: Studies of customer retention analytics in B2B settings

2.2. Uplift modeling for customer retention management

Uplift modeling in a customer retention management context aims to estimate the net difference in churn probability that results from a targeted retention campaign action. A prerequisite for uplift modeling is data about two sets of customers, including their characteristics and churn behavior. The first set consists of customers who have not been targeted with a retention campaign (i.e., control group C), and the second set includes those who have received it (i.e., treatment group T). Uplift modeling incorporates the treatment decision variable (to treat a customer or not) into the model specification. Thus, consider a customer i , with observed characteristics \mathbf{X}_i (e.g., firmographics, behavioral variables), for which Y_i denotes observed churn behavior, so that 1 equal churned and 0 otherwise. Then let T_i denote whether customer i is treated with a retention campaign, so 1 equal treated and 0 otherwise. In a traditional churn prediction context, the churn probability of customer i would be calculated, given its characteristics: $P[Y_i | \mathbf{X}_i]$. Treatment is not considered, so $T_i = 0$ for every customer i , and $P[Y_i | \mathbf{X}_i] = P[Y_i | \mathbf{X}_i, T_i = 0]$. Uplift modeling instead includes the treatment and uplift (U) for a customer i , defined as:

$$U_i = P[Y_i | \mathbf{X}_i, T_i = 0] - P[Y_i | \mathbf{X}_i, T_i = 1]. \quad (1)$$

In a customer retention management context, the customer base can be classified along two dimensions in uplift modeling, as functions of whether customers churn when targeted and whether they churn when they are not targeted. Figure 1: Conceptual uplift table in customer retention management Figure 1 depicts this theoretical classification, which results in four customer quadrants. First, *persuadables (I)* must be targeted. For these customers, the retention offer is effective, so they will churn if they are not targeted but convert into non-churners in response to the retention campaign. Second, *sure things (II)* stay whether they are targeted by the retention campaign or not. They should not be targeted; the costly incentive does not change the final outcome. Third, *do not disturb (III)* only churn if targeted. Recent empirical evidence suggests that a retention campaign can increase churn among some customers by lowering their inertia and prompting them to switch subscriptions, as well as by increasing the salience of past usage patterns among potential churners (Ascarza et al., 2016). Fourth, *lost causes (IV)* churn independently of receiving a retention offer and should not be targeted.

		Customers churn when targeted? (treatment group)	
		No	Yes
Customers churn when not targeted? (control group)	Yes	(I) Persuadables	(IV) Lost causes
	No	(II) Sure things	(III) Do not disturbs

Figure 1: Conceptual uplift table in customer retention management

Considering the scarce research into uplift modeling for customer retention management, Table 2 offers an overview of applications, methods, and interpretability aspects available in broader marketing literature. It suggests three main conclusions. First, uplift modeling for customer retention management is nonexistent in B2B and only recently introduced in B2C retention management literature. Radcliffe (2007) present a case on how uplift modeling can reduce churn in a mobile phone company. He concluded that around 30% of the customers with the highest uplift should be targeted to reduce the churn rate from 9% to 7.8% percent. Next, Guelman et al. (2012) applies their uplift RF algorithm on a retention case for an insurance company. Also Ascarza et al. (2018) empirically demonstrate the superiority of uplift modeling by applying it to select which customers to target in two real-world B2C field experiments, one for a wireless provider and another for a special interest membership organization. The results indicate that in both cases, targeting customers based on uplift is more effective than targeting them solely according to churn probabilities. Devriendt et al. (2018)

also apply uplift modeling, to a data set of 200,903 customers in the financial services industry, and achieve significantly better results than random targeting if 60% of the customer base is targeted. They propose that the fraction of customers that needs to be targeted should be optimized to obtain the best results. Finally, Lemmens and Gupta (2020) propose a profit-based loss function to predict the financial impact of a retention intervention for every customer. Based on the results of two real-world field experiments, they conclude that their approach leads to more profitable retention campaigns than traditional uplift models.

Second, various AI-based methods for uplift modeling offer adjustments of classical predictive modeling methods such as logistic regression (e.g. Lo, 2002; Radcliffe, 2007), decision trees (e.g. Athey & Imbens, 2016; Rzepakowski & Jaroszewicz, 2012a; Zhao et al., 2017), **decision** forests (e.g. Athey et al., 2019; Guelman et al., 2012), and support vector machines (e.g. Zaniewicz & Jaroszewicz, 2017). The uplift model types listed in Table 2 show only most popular uplift techniques in the broader marketing domain, and do not present an exhaustive overview of uplift modeling strategies. Moreover, for every model type listed, several variations might exist, but we only discuss those directly related to the uplift LLM in our methodology section. We kindly refer you to previous benchmarking literature for more elaborate uplift modeling overviews (Devriendt et al., 2018; Gubela et al., 2019). According to Table 2, uplift LR, uplift DT, and uplift RF are the most popular uplift methods that yield the best predictive performance across contexts. Existing data science research (e.g. De Caigny et al., 2018) also suggests that uplift modeling methods can be applied to full customer data sets all at once, or else to segments separately, resulting in a segmented uplift modeling approach. In that case, various segments with distinct uplift drivers can be identified, resulting in better targeted and actionable retention campaigns. However, we know of no uplift modeling studies that adopt segmented approaches and build uplift models according to a priori segments found in the customer base. To address this gap, we propose a segmentation-based uplift modeling approach (as detailed in Section 3).

Third, interpretability is a key topic in data science and several analytical approaches have been proposed for explaining model predictions (Ribeiro et al., 2016). Nevertheless, interpretability is difficult to measure as it cannot be described in a pure mathematical formula, and comes down to the degree to which humans can interpret and understand the cause of a decision (Miller, 2019). Analytical approaches that aim to explain model predictions vary in scope and flexibility. The scope indicates on what level explanations are given which ranges from global (i.e. across all customers) to local (i.e. on a more granular level). Flexibility reflects the way interpretability is achieved and can be either intrinsic or model-agnostic (Molnar, 2019). Also for customer retention management and beyond, interpretability of uplift models is a key requirement (e.g. Verbeke et al., 2011). The uplift modeling outcomes must be able to inform the decision-making strategy adopted by the customer retention manager and reveal various drivers that can determine the success of retention campaigns.

Interpretability also indicates drivers that might steer personalization efforts. The findings in Table 2 distinguish between *global* and *local* interpretability (Molnar, 2019). The former seeks insights into the overall uplift model’s predictions, over all customers for all possible variables’ values. The latter instead supports interpretations at a more granular level, typically at customer level which is the lowest level of granularity. We also include segment level which is an important decision level in customer retention management (Maldonado et al., 2021). Most previous uplift modeling research focuses on global interpretability (e.g. Ascarza, 2018; Guelman et al., 2012), but decision makers in B2B contexts often require local interpretability to manage customer retention on a more granular level (Tamaddoni Jahromi et al., 2014). For example, Gubela and Lessmann (2020) acknowledge the importance of looking at customer subsets to evaluate business outcomes. They propose to rank customers based on their predicted individual treatment effects and only consider the top deciles for further discussion. Therefore, together with the new segmented model, we propose a visualization framework that satisfies local interpretability requirements, as detailed further in Section 5. The importance of inherently interpretable models for decision making is acknowledged in recent research (Rudin, 2019). In line with recent advances in customer retention management, uplift LLM is inherently interpretable and thus does not require model-agnostic techniques such as partial dependence plots, depicting the functional relationship between one or a few independent variables on the dependent variable, or variable importance to derive interpretable output.

Reference	Context	Uplift Model Type							Segmented Approach?		Interpretability		
		Logistic regression	Decision Tree	Decision Forests	Boosted Trees	SVM	KNN	Other	LLM	No	Yes	Global	Local
(Chickering & Heckerman, 2000)	Advertising		X							X		X	
(Hansotia & Rukstales, 2002)	Purchase	X	X							X		X	
(Radcliffe, 2007)	Purchase & churn	X	X							X		X	
(Rzepakowski & Jaroszewicz, 2012b)	Purchase		X							X		X	
(Guelman et al., 2012)	Churn			X						X		X	
(Kane et al., 2014)	Inquiry/buy & online visit & profit	X								X		X	
(Guelman et al., 2015)	Purchase			X						X		X	
(Sołtys et al., 2015)	Online visit			X						X		X	
(Cao et al., 2017)	Purchase	X								X		X	
(Zaniewicz & Jaroszewicz, 2017)	Online visit					X				X		X	
(Michel et al., 2017)	Purchase		X							X		X	
(Zhao et al., 2017)	Purchase			X	X	X	X			X		X	
(Devriendt et al., 2018)	Purchase, online visit, response & churn	X	X	X						X		X	

(Ascarza, 2018)	Churn			X					X		X	
(Gubela et al., 2019)	Purchase	X		X		X	X	X			X	
(Lemmens & Gupta, 2020)	Churn				X				X		X	
(Gubela et al., 2020)	Purchase & purchase amount	X		X	X	X	X	X	X		X	
Our study	Churn	X	X	X					X	X	X	X

Table 2: Studies of uplift modeling in the marketing domain

3. Uplift logit leaf model and building blocks

The uplift LLM is a two-step algorithm designed for segmented uplift modeling, to provide segment-specific insights. Methodologically, uplift LLM relies on uplift decision trees and uplift logistic regressions, which are complementary (Lo, 2002; Rzepakowski & Jaroszewicz, 2012a). That is, the uplift decision tree can find interaction effects, whereas the uplift logistic regression deals better with linear relationships. Their combination in the uplift LLM algorithm thus should increase overall model performance. Furthermore, the algorithm represents an uplift modeling extension of the LLM, originally developed for traditional customer churn prediction (De Caigny et al., 2018), for which empirical confirmation demonstrates the beneficial effects of combining decision trees with logistic regressions based on a benchmarking study on 14 data sets. De Caigny et al (2018) also discusses how LLM differs from conceptually similar models such as logistic model trees (Landwehr et al., 2005).

Figure 2 presents a conceptual representation of the uplift LLM algorithm and data flows. It requires a set S of customers that receive a retention campaign, as the treatment group, and another group that did not, as the control group. Uplift is defined as a decrease in the churn probability, as a result of receiving the retention campaign. This approach employs the Neyman-Rubin framework to estimate the uplift, i.e. the Individual Treatment Effect (ITE), in terms of potential outcomes (Rubin, 1978; Splawa-Neyman et al., 1990). The ITE is estimated by contrasting the potential outcomes that would be obtained when applying and when not applying the treatment, and which are estimated by the uplift model. The true ITE is unobserved for any instance in real-world applications. However, it can be estimated under assumptions that highly depend on external information. For example, a clear understanding of the assignment mechanism that is employed to apply the treatment to subjects. First, the *Stable Unit Treatment Value (SUTVA)* (Rubin, 1978) assumption states that the potential outcomes for a subject are unaffected by the treatment assigned to other subjects. That is, the churn value of outcome Y for customer i when treatment T is applied will remain the same, regardless of the assignment mechanism or the treatments received by other customers, i.e., no spillover or network effects. Second, the positivity or common support assumption guarantees that the probability of a customer to be treated is positive, i.e., $P(T = 1|x) > 0$. This assumption makes it possible to find suitable matches between instances in the treatment and the control group. Namely, there is overlap on observed variables X between the treated and control subjects. Third, under *ignorability*, also known as unconfoundedness (Rosenbaum & Rubin, 1983) or selection on observables, the potential outcomes are independent of the treatment, conditional on the observed variables (Y_i independent of T given X). Ignorability indicates that one can ignore under what rule subjects are assigned to either treatment group. However, the selection of the variables to condition depends largely on domain knowledge. All customers in S also are divided into $n \in \mathbb{N}_1$

segments (S_1, S_2, \dots, S_n) based on the decision rules of an uplift DT. Subsequently, for every segment, an uplift LR then is fitted, resulting in a segment-dependent uplift probability estimate for every customer. This segmented approach thus reveals segment-specific uplift drivers.

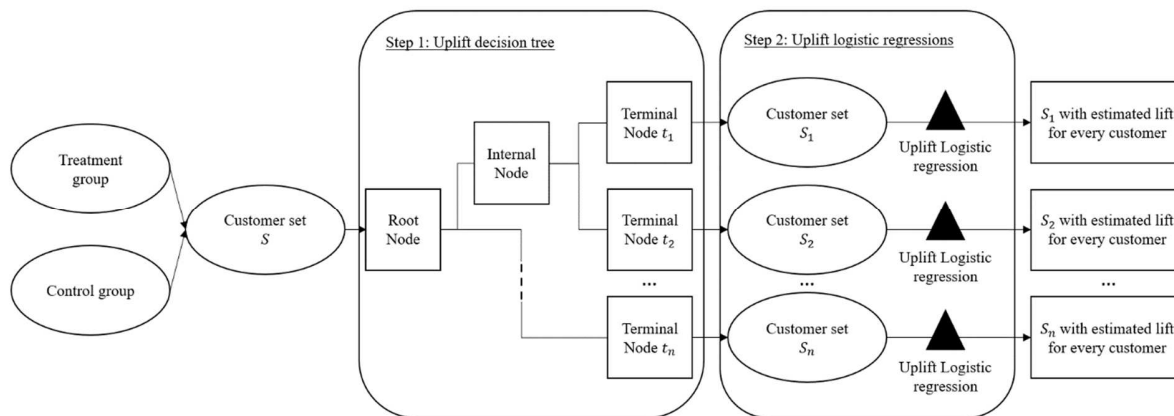


Figure 2: Overview of the uplift LLM algorithm

The first step divides the customer set into various segments using the decision rules of an uplift DT. Decision tree-based approaches rely on adaptations of well-known decision tree induction algorithms (Gubela et al., 2020). Candidate splits are evaluated according to a quality measure that indicates whether a split can maximize the impact of the retention campaign. The best splits then get iteratively selected until a certain stopping criterion is met. Divergence-based splitting criteria for growing uplift trees, as based in information theory, is often used in uplift modeling (Rzepakowski & Jaroszewicz, 2012a). The objective of every split is to maximize the distance between churn distributions in the treatment and control groups in the child nodes. The fraction of churners and non-churners in the treatment groups thus should be as different as possible to the fraction of churners and non-churners in the control group. Divergence measures indicate the difference in churn distribution in both child nodes. Through repeated splitting of nodes, more complex models gradually emerge, but so does the risk of overfitting, in which case the uplift DT is modeling idiosyncratic characteristics of the data set that hinders generalizability. To deal with complexity, and thus the overfitting risk, uplift DT algorithms include a post hoc pruning step, as a complexity reduction technique, which removes parts of the tree that do not provide enough power for the classification of customers. In addition, hyperparameters can control the complexity of the tree according to specified boundaries. For example, a minimal leaf size would limit the minimum number of customers in the nodes, as a stopping criterion.

The second step in the uplift LLM algorithm fits the uplift LR for every segment s . The uplift DT employs a hard-split criterion, meaning that every customer i can only belong to a single segment s . From a modeling perspective, the treatment and control group might be grouped into a single training sample, to

estimate the uplift model (Cao et al., 2017; Lo, 2002). The group origin (i.e., treatment or control) of a customer i is flagged by a treatment dummy variable T_i . The segment specific uplift logistic regression thus accounts for the customer's characteristics X_i , the churn indicator Y_i , the treatment variable T_i , and all possible interactions among the customers' characteristics and treatment X_iT_i . The predicted uplift is the difference between churning when not treated minus the probability of churning when treated. Using Equation 1, uplift for customer i then can be estimated as:

$$U_i = \frac{\exp(\alpha_s + \beta_s' X_i)}{1 + \exp(\alpha_s + \beta_s' X_i)} - \frac{\exp(\alpha_s + \gamma_s + \beta_s' X_i + \delta_s' X_i)}{1 + \exp(\alpha_s + \gamma_s + \beta_s' X_i + \delta_s' X_i)}, \quad (2)$$

where α_s denotes the segment specific intercept, β_s is the segment's main effect of the customer characteristics, and δ_s indicates the segment specific main treatment effect (Lo, 2002).

4. Experimental design

4.1. Data set

The experiment is based on real-world data from a European B2B software company. The company provides process automation software that is offered to customers worldwide as renewable contracts. The data set contains 6,432 observations, and each observation represents a customer contract with the focal firm. For each customer, the independent variables are observed over the complete contract period, but the dependent variable is observed only over the six-month window after a contract renewal date. This contract renewal date varies across customers. If a customer does not to renew its contract after six months, we define it as a churner. The retention campaign targeted 1,399 customers, and a churn rate of 17.58% was observed (see Table 3). For the control group, the churn rate was higher, at 19.29%. The baseline uplift represents the overall effect of the retention campaign on the churn rate, calculated as the difference between the churn rate of the control group minus the churn rate of the treatment group, which equals 1.71%.

	Number of Customers	%	Number of Churners	Churn Rate (%)	Uplift
Treatment	1,399	21.75%	246	17.58%	1.71%
Control	5,033	78.25%	971	19.29%	
Total	6,432	100.00%	1,217	18.92%	

Table 3: Data set characteristics

The customer information produces 144 independent variables, which can be categorized according to the focal sales funnel, which involves the distributor, the reseller, and the customer. Variables at the *distributor* and *reseller* levels refer to the characteristics of specific intermediaries linked to the customer. The category with the most variables is the *customer* level. Representing the

characteristics of the end customer, this wide range of variables can be grouped into four categories: firmographics (e.g., customer industry, size), behavioral (e.g., website visits, campaign data), product usage (e.g., number of times a specific action is performed with the product), and relationship (e.g., length). Table 4 presents an overview of the categories and subcategories of variables considered, as well as an exact count of variables in each (sub)category.

Category	Subcategory	Number of Variables
Distributor	Firmographics	8
	Products	3
	Partnership level	3
	Relationship with company	2
Reseller	Firmographics	8
	Partnership level	3
	Relationship with company	2
Customer	Firmographics	21
	Behavior	79
	Product usage	7
	Relationship with company	8

Summary statistics of anonymized variables are available upon request.

Table 4: Independent variables

For this study, we implement a 5-fold cross-validation experimental design, in line with previous uplift literature (Devriendt et al., 2018). With it, we compare different modeling techniques and achieve stable, valid experimental results by repeatedly and randomly splitting the data set into training and testing sets (Verbeke et al., 2017). That is, we start by splitting the data set randomly into five equally sized subsets. In each iteration, four subsets constitute the training set to build the uplift models, and one subset is a hold-out test set, designed to validate the generalization performance. Due to the typically imbalanced class distribution, the folds are stratified and randomly sampled, to ensure identical distributions of the churn variable across folds. Furthermore, we require the distribution of control and treatment group observations to be identical across folds, as needed for uplift modeling. Next, the training sets are further divided into training (two-thirds of training data) and validation (one-third of training data) subsets, to support hyperparameter tuning of the various algorithms, as is standard practice in data science (e.g. Devriendt et al., 2018; Verbeke et al., 2012).

We adopt the uplift decision tree algorithm with a Kullback-Leibler divergence splitting criterion (Rzepakowski & Jaroszewicz, 2012a). The uplift logistic regression uses the implementation described in Section 3 (Lo, 2002; Radcliffe & Surry, 1999). Both uplift models are popular and frequently used

(Devriendt et al., 2018; Gubela et al., 2020). They inform the uplift LLM model. We also consider identical hyperparameters to ensure a fair comparison in the experiment, which includes uplift RF as a gold standard in terms of model performance, according to prior modeling studies (e.g. Ascarza, 2018; Devriendt et al., 2018; Guelman et al., 2015). Table 5 lists the considered hyperparameters, optimized for the respective uplift modeling algorithms. For every fold and separately, a grid search over all possible hyperparameter combinations in the validation set indicates the best combination, before we build the model on the complete training data. The stepwise variable selection in the uplift LR uses a general wrapper approach based on the Qini metric. This iterative process for selecting an optimized subset of variables will develop models by adding more variables until no performance gains are achieved. We refer to Devriendt et al. (2018) for a detailed description of Qini based variable selection and the algorithmic procedure.

Algorithm	Hyperparameter	Settings
Uplift DT	Max depth of the tree	[1, 2, 3, 4, 5]
	Minimum number of observations in a node	20
	Minimum number of control observations in terminal node	[5, 10, 20, 50, 75, 100]
	Minimum number of treatment observations in terminal node	[5, 10, 20, 50, 75, 100]
Uplift LR	Stepwise variable selection method	[None, forward-Qini]
Uplift LLM	Max depth of the tree	[1, 2, 3, 4, 5]
	Minimum number of observations in a node	20
	Minimum number of control observations in terminal node	[5, 10, 20, 50, 75, 100]
	Minimum number of treatment observations in terminal node	[5, 10, 20, 50, 75, 100]
	Stepwise variable selection method	[None, forward-Qini]
Uplift RF	Number of variables to be tested in each node	[3, 4, 5]
	Splitting rule	Euclidean distance
	Number of trees	[100, 200, 300]
	Minimum number of observations in a node	20
	Minimum number of control observations in terminal node	[5, 10, 20, 50, 75, 100]
	Minimum number of treatment observations in terminal node	[5, 10, 20, 50, 75, 100]

Table 5: Hyperparameter settings of considered uplift models

4.2. Qini coefficient

The evaluation of uplift models is more challenging than that of predictive models, because the true value of the target variable is not observed, so we cannot know the net effect of the treatment, given that a customer can only belong to the control or treatment group, not both. Therefore, evaluation metrics

in uplift modeling rely on groups rather than individual customers (Devriendt et al., 2018). The Qini metric is a well-established evaluation measure—an extension of the well-known Gini coefficient that assesses goodness of fit (e.g. Devriendt et al., 2018; Radcliffe, 2007).

Specifically, the Qini metric is based on a gains charts for uplift or Qini curve. Figure 3 depicts the Qini curves for random targeting and an uplift model, which plot the cumulative differences in churn rates between control and treatment groups as a function of a selected fraction of customers, ranked from high to low predicted uplift. The dotted line represents the incremental gains achieved by randomly targeting customers. The value at 100% of the population targeted equals total uplift. In this example, the total uplift is 10%. The theoretical maximum uplift equals the churn rate, which would imply that the treatment is 100% effective and converts all churners to non-churners. The Qini metric is derived from the incremental gains curve; it measures the area between the Qini curve of the uplift model (i.e., full line) and the incremental gains achieved through random targeting of customers (i.e., dotted line). The Qini metric can be calculated at different levels and is a good metric for comparing the performance of different models on a given data set, using a single value (Devriendt et al., 2018).

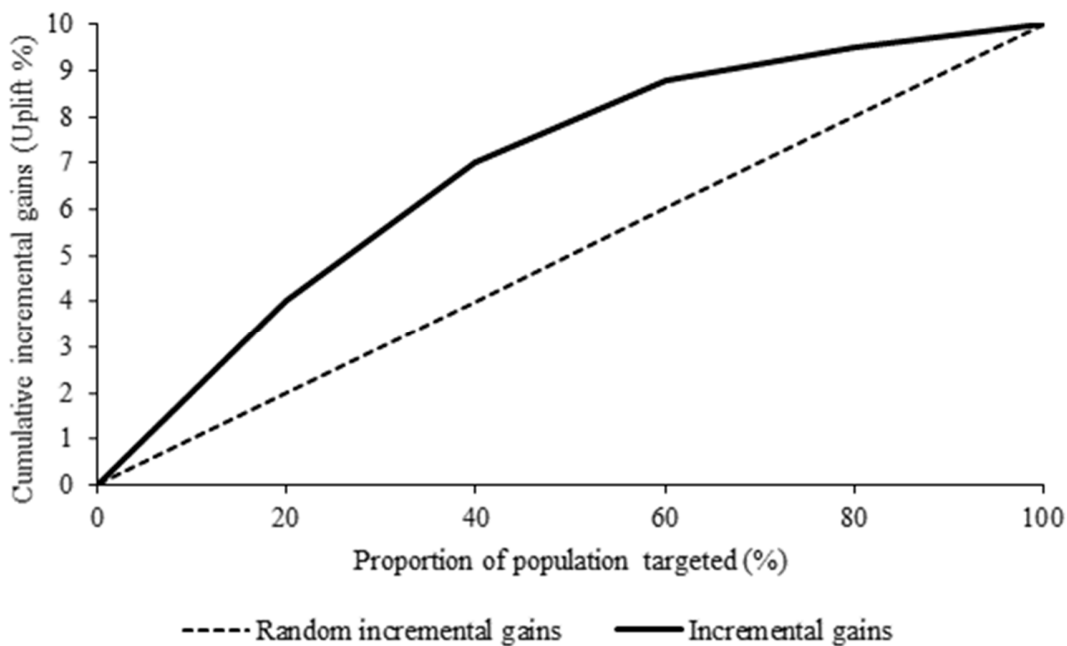


Figure 3: Incremental gains in a Qini curve

5. Results and managerial insights

5.1. Predictive performance

In this section, we report the experimental results and compare the predictive performance of uplift LLM with three popular uplift algorithms. Table 6 details predictive performance in terms of the overall Qini metric, as well as the top 30% and top 10% Qini scores, averaged over the 5-fold CV. Values between brackets indicate standard deviations. The best performing algorithm is underlined for each level of the Qini scores. As Table 6 confirms, uplift LLM performs best at all considered levels, meaning that it performs better than its constituent building blocks, uplift DT and uplift LR, and it also outperforms the state-of-the-art uplift RF algorithm. All differences with uplift LLM and the considered uplift algorithms are significant based on a pairwise comparison using dependent sample t-tests. Other splitting criteria for uplift RF (i.e. Kullback-Leibler, Chi-squared and interaction method) did not give significant improvements over the Euclidean distance splitting criterion.

Figure 4 plots the Qini curves for the uplift LLM and three selected algorithms together with a baseline based on randomly targeting. Note that the results are based on the 5-fold CV (Devriendt et al., 2018). The x-axis refers to the proportion of targeted customers, and the y-axis depicts the average cumulative incremental gains or cumulative uplift, expressed as a proportion of the population, over the different folds. This represents the cumulative difference in churn rate between customers in the control and the treatment condition as a function of the selected fraction of customers, ranked based on their uplift scores. In Figure 4, all algorithms perform better than the baseline, except for uplift LR when targeting 10% of total customers, which actually results in a negative Qini value. The uplift LLM clearly outperforms the selected algorithms at every proportion of customers targeted. All algorithms achieve their optimal uplift value when targeting around 60%–80% of all customers. From a managerial perspective, these results indicate that uplift modeling is a viable strategy for designing effective customer retention campaigns that target the proportion of customers that results in the greatest uplift. Moreover, uplift LLM performs best, so it provides the most appropriate option for ranking customers and maximizing the outcomes of the retention campaign. Uplift LLM has on average a lower number of segments (mean= 4.00, sd.= 3.08) than the uplift DT (mean=6.00, sd.= 4.12). The number of variables in the segment specific uplift LR of the uplift LLM (mean = 10.80, sd.= 6.40) is also on average lower than in uplift LR (mean=37.00, sd.= 4.40).

Algorithm	Qini	Qini _{top30%}	Qini _{top10%}
Uplift LLM	<u>4.5646</u> (0.0059)	<u>0.5838</u> (0.0005)	<u>0.0649</u> (0.0001)
Uplift DT	1.9344*** (0.1949)	0.2852*** (0.0726)	0.0280*** (0.0189)
Uplift LR	2.1869*** (0.7287)	0.1955*** (0.1825)	-0.0010 *** (0.0257)

Uplift RF	3.2543*** (0.1301)	0.4811*** (0.0185)	0.0558** (0.0020)
-----------	--------------------	--------------------	-------------------

***, **, * indicate significance differences with uplift LLM at the 99%, 95%, 90% confidence level, respectively

Table 6: Qini performance of uplift LLM and selected uplift algorithms

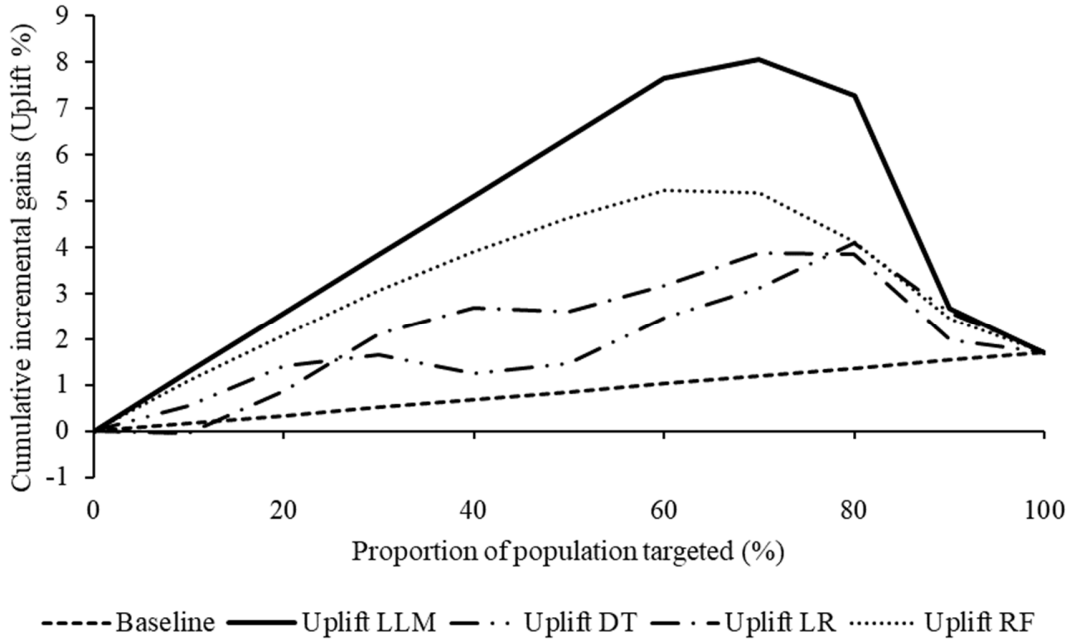


Figure 4: Qini curves for all uplift techniques

5.2. Interpretability of uplift LLM

This section describes how to interpret the uplift LLM algorithm at global and local levels, then use it as a decision-making tool to optimize B2B retention campaigns. The presented model is based on the first fold of the experiment and hyperparameters are optimized according to the procedure described in Section 4.1. The output of the uplift LLM can be interpreted from a *global* perspective, which marketers can use to find an optimal proportion of customers to be considered for targeting. Figure 5 details the incremental gains achieved with the uplift LLM model, whose output ranks customers according to their predicted uplift scores, from highest (left side of x-axis) to lowest (right side of x-axis). Decision makers can use this information to select customers to target with a retention campaign, according to the effectiveness of the retention offer for reducing predicted churn. The greatest impact of the retention campaign appears to occur when targeting around 60% to 70% of the population, which results in an uplift of 6%. Comparing the uplift LLM with uplift DT and Uplift LR, the uplift LLM returns a concise model. Uplift LLM returns 4 segments, while an uplift DT on the same data has 8 leaf nodes. Also, the segment specific uplift LRs have on average only 3 variables, while a global uplift LR

trained on the same data resulted in 34 variables. Figure 6 shows the relative proportion of every segment in different deciles based on ITE. The Customers in the top decile, for example, all belong to segment 3, which is a segment with a very high uplift. Note that also some customers of segment 3 are in decile 10, with the lowest ITE. This is because the second stage uplift LR in the uplift LLM models allows for personalized estimates. Selection of customers to target should thus happen across segments.

From a *local* perspective or segment level, Table 7 also presents the output of the uplift LLM model as a comprehensible decision table that supports customer retention decision making. The customer base gets split into four segments by the uplift DT, according to three rules that we detail subsequently. The first decision rule is based on whether the customer's contact person has a managing role. The company has for around 52% of their customers a direct contact with a person in a managing role. Segments 1 and 2 have higher churn rates in the control group (%churn C column) than do Segments 3 and 4, so a contact with a customer who has managerial responsibilities clearly yields a beneficial effect on churn. Managers are often budget owners who make the final decision about how to spend their budgets, such as by renewing the software subscription. Directly targeting a retention campaign toward the decision maker tends to be more effective, which might explain this beneficial effect.

To assess the effect on uplift, we observe differences over different segments. If the contact person is not the manager, as in Segments 1 and 2, a second decision rule refers to the number of pages visited ($\bar{x} = 22.69$, $sd. = 105.7$, $min. = 0$, $max. = 4,052$), to determine for which customers the retention offer is likely to be effective. Segment 1 has substantially lower uplift than the general value of 1.71%; these customers should not be targeted. This finding also is reflected in the positive treatment effect (i.e., treatment increases churn probability) in the segment-specific uplift LR (Treat $\beta = 0.03$). Segment 2 instead shows a large positive uplift effect, indicating that these customers should be prioritized in the retention campaign. Again, we find confirmation in the uplift LR for this segment, which indicates a negative treatment effect (Treat $\beta = -0.46$), such that the churn probability drops after treatment. A significant, positive interaction effect between the number of campaigns in the previous quarter and treatment also arises (Nbr campaigns LQ:Treat $\beta = 0.04$), so customers who received a lot of marketing communications during the previous quarter are less keen to respond positively to a personalized retention offer. Perhaps they reached saturation with regard to marketing communication from the focal firm (Tellis et al., 2005). Note that different variables across segments might be selected in the uplift LR which signals relevant, segment-specific uplift drivers.

In Segments 3 and 4, the contact person is the manager, but we can distinguish them using a third decision rule, based on duration, or the length of the customer's current contract ($\bar{x} = 374.5$, $sd. = 25.4$, $min. = 38$, $max. = 729$). Targeting customers with a short duration (Segment 3) is beneficial, according to the negative treatment effect in the uplift LR (Treat $\beta = -1.59$). Customers with a long duration (Segment

4) should not be targeted though, because we observe a positive treatment effect ($\text{Treat } \beta = 0.91$). The retention campaign arguably might increase the salience of past usage patterns among potential churners and lead them to reevaluate the contract (Ascarza et al., 2016). To reduce churn in this group, the focal firm might consider other, more effective offers.

To confirm these recommendations for retention targeting, in Figure 7 we present the incremental gains for each segment. The visual analysis of these graphs confirms that Segments 2 and 3 are the most effective to target with retention campaigns. Moreover, these graphs indicate the optimal proportion of customers to target in each segment, which decision makers can leverage to optimize the returns on their retention campaigns. For example, Segment 3 exhibits an uplift of around 9% obtained from targeting about 40% of the customers—significantly greater than the 6% achieved by random selection at the same level. In Figure 8, we present the partial dependence plot between the value per month on the x-axis and the churn probability on the y-axis for Segment 3. The churn probability is clearly lower after the treatment, due to both the treatment effect ($\text{Treat } \beta = -1.59$) and its interaction with value per month ($\text{Value per month:Treat } \beta = -0.01$). Churn is higher for lower valued contracts, whereas this probability falls to null for contracts with the highest value. This finding is in line with business expectations, in that contracts with the highest value generally feature the most loyal customers, for which the importance of the software gradually grows as they build more internal know-how about its uses.

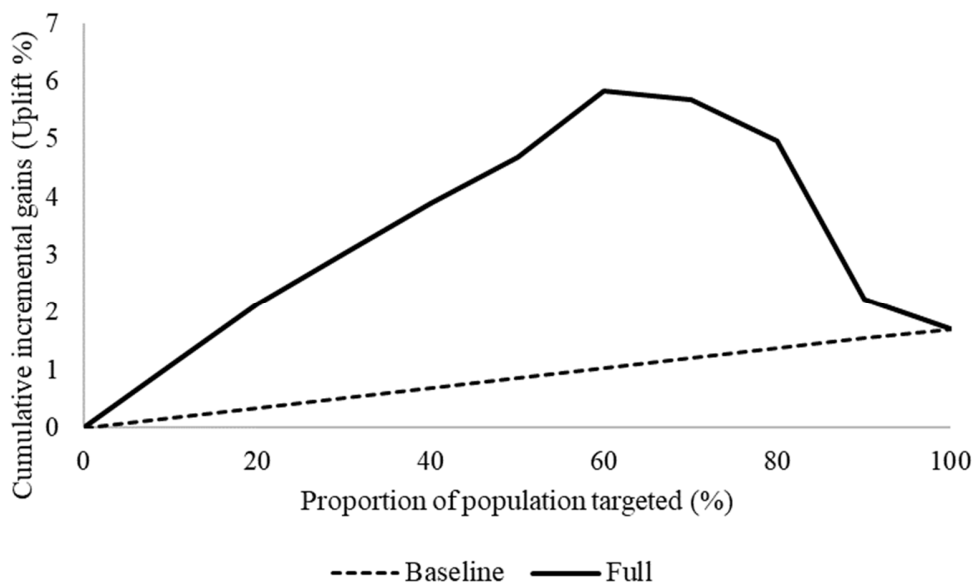


Figure 5: Cumulative incremental gains over all segments

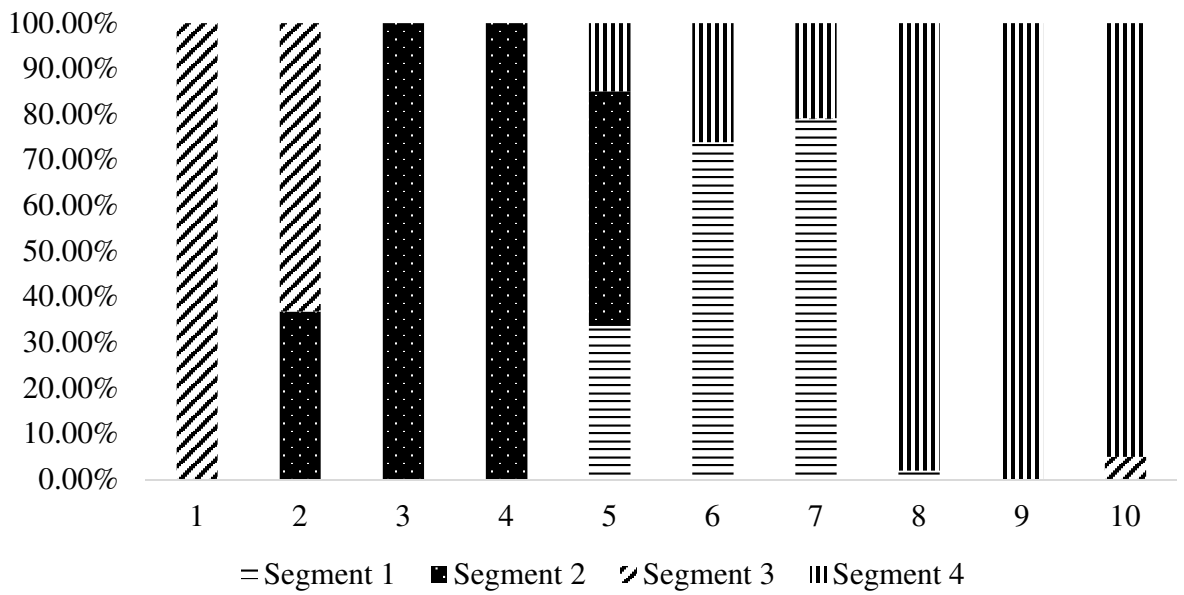


Figure 6: Proportion of customers from a certain segment in every decile (1= top decile) based on ITE

Step 1: uplift DT							Step 2: uplift LR			
Segment	Decision Rules		#obs.	% churn C	% churn T	Uplift segment	Int.	Variable	Treatment	Interaction effect
	Rule 1	Rule2								
1	Jobtitle contact ≠ manager	Nbr pages visited < 10	1,480	24.11%	23.06%	1.05%	(Int.) 1.11***	Activity 1 LQ -0.07***	Treat 0.03***	Activity 1 LQ:Treat 0.02***
2	Jobtitle contact ≠ manager	Nbr pages visited ≥ 10	969	30.19%	14.86%	15.33%	(Int.) 0.64***	Nbr campaigns LQ -0.27***	Treat -0.46***	Nbr campaigns LQ:Treat 0.04***
3	Jobtitle contact = manager	Duration < 382	1,833	15.78%	1.42%	14.36%	(Int.) 1.28***	Value per month -0.01***	Treat -1.59***	Value per month:Treat -0.01***
4	Jobtitle contact = manager	Duration ≥ 382	864	15.64%	25.0%	-9.36%	(Int.) 1.78***	Nbr.employees reseller 0.01***	Treat 0.91***	Nbr.employees reseller:Treat -0.01***

***, **, * indicate significance at the 99%, 95%, 90% confidence level, respectively

Table 7: Decision table for the uplift LLM model

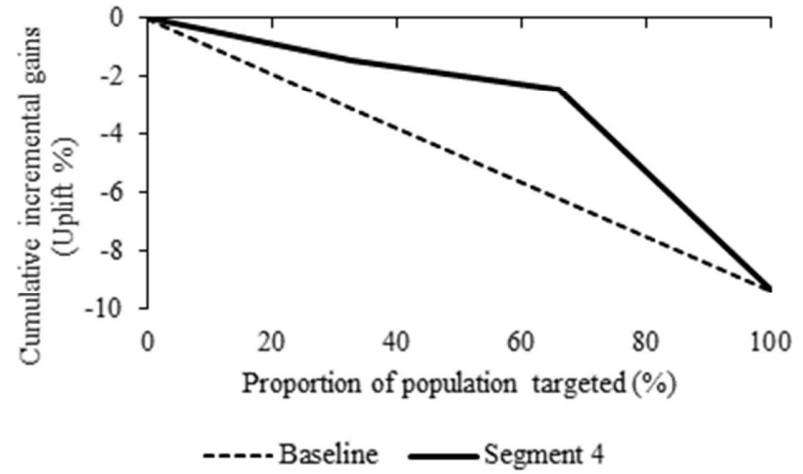
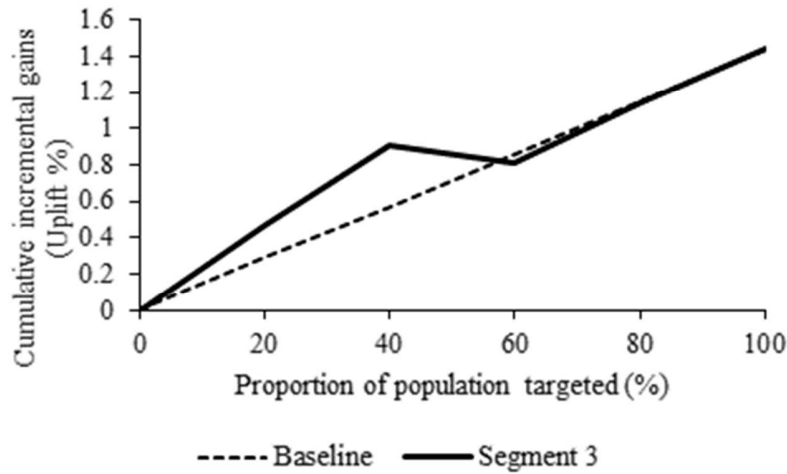
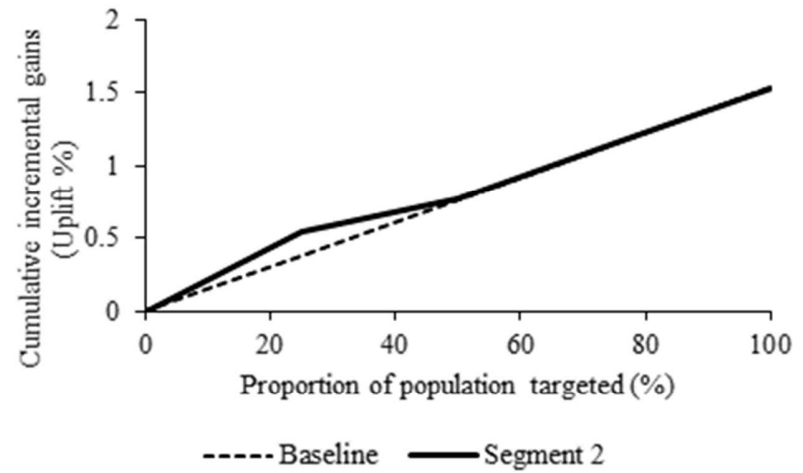
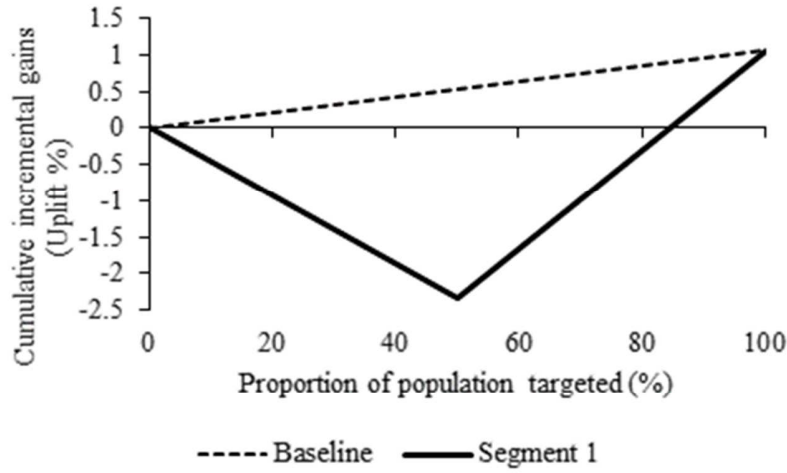


Figure 7: Cumulative incremental gains by segment

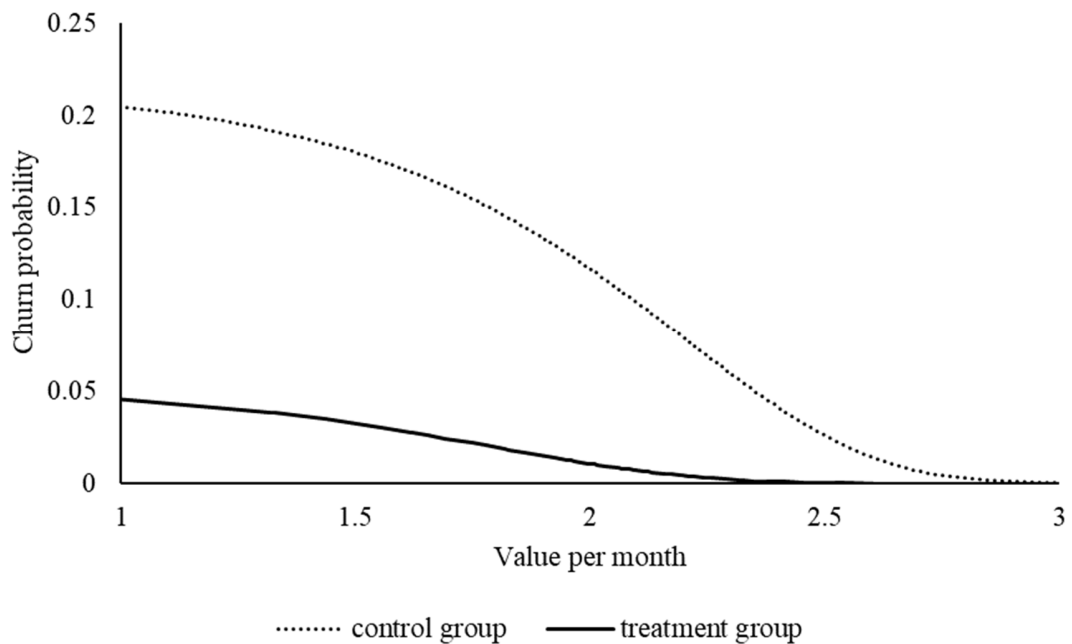


Figure 8: Value per month and churn probability for control and treatment group in segment 3

Note: Value per month is multiplied by a random variable, for confidentiality reasons.

6. Discussion

6.1 Conclusion

AI is booming and will add \$13 trillion to the global economy over the next decade (Fountain et al., 2019). The technologies that enable AI are advancing rapidly and becoming increasingly affordable, which makes AI a value tool in the arsenal of all marketers (Lamm, 2020). Indeed, academics and practitioners realize the benefits of using AI tools for B2B marketing, reinforced by the abundance of available big customer data. For instance, Zhang et al. (2020) found that big data analytics intelligence has a positive impact on firm performance through improved mass-customization capabilities in B2B markets. Extant B2B customer retention management literature focused mainly on customer churn prediction modeling as a popular AI tool (Barfar et al., 2017; Chen et al., 2015; Gordini & Veglio, 2017; Tamaddoni Jahromi et al., 2014), employed to identify customers at risk of ending their relationship with the firm. However, recent literature has demonstrated that companies should switch from predictive to prescriptive analytics; in other words, B2B firms should not target customers solely on the basis of their propensity to churn but rather should estimate the net effect of retention campaigns (e.g. Ascarza, 2018).

In this regard, this study highlights the beneficial impact for both academics and practitioners of an effective, accurate, data-driven uplift model to select customers to include in a retention campaign as to improve its effectiveness. In other words, the ROMI is maximized by targeting only these risky customers for whom the retention campaign is effective, i.e. these customers will churn if not targeted, but continue the relationship with the company as a response to the retention offer of the company. We introduce uplift modeling in a B2B customer churn setting and apply it to a real-life uplift churn data set with 6,432 customers of a European software provider. We further propose an interpretable, segmented uplift modeling approach, i.e. the uplift LLM, and showcase its superiority in terms of predictive performance and interpretability against three well-established uplift algorithms. Therefore, our research study offers important contributes to several research streams. First, we contribute to the B2B customer retention management literature. Customer retention management is important in B2B because customers often have larger purchase amounts and more transactions (Rauyruen et al., 2011). Our extensive literature review reveals that customer retention management is often looked at from a B2C perspective, while B2B settings receive only limited attention. Further, although various papers looked at the impact of analytics and big data in B2B settings (e.g. Farrokhi et al., 2020; Kumar et al., 2020), only a few notable exceptions zoom into traditional customer churn prediction modeling as an AI tool to improve customer retention management (Gordini & Veglio, 2017; Tamaddoni Jahromi et al., 2014). This study suggests the integration of uplift churn modeling as a viable AI-based tool to limit customer attrition in B2B contexts, as uplift modeling has proven merits for customer retention management in B2C settings (e.g. Ascarza, 2018; Devriendt et al., 2018).

Second, our study extends the uplift modelling literature stream. From a methodological perspective, we introduce uplift LLM, a new uplift model algorithm which combines comprehensibility with predictive performance. We compare the performance of the proposed approach against three conceptually related, popular uplift algorithms -uplift DT (Rzepakowski & Jaroszewicz, 2012a), uplift LR (Lo, 2002), and uplift RF (Guelman et al., 2012) -and demonstrate the superior performance of the uplift LLM algorithm over all three algorithms. Thus, scholars and practitioners in the data science field should consider segmentation-based uplift modelling approaches for their real-world uplift prediction efforts.

Third, we contribute to the interpretable data science field. Over the years, the AI domain has shifted focus from the pure algorithmic development and implementation to delivering deep marketing insights needed for making sound marketing decisions. The new uplift LLM as a segmented prediction approach that provides segment-specific insights contributes to this trend. In line with recent literature on interpretable data science (Hansen et al., 2013; Vellido, 2019), we provide new insights in the form of customized, developed visualizations at the global and segment levels, which are relevant for B2B decision making.

6.2 Implications for practice

Our research study suggests several managerial implications. First, marketing managers are encouraged to optimize their retention campaigns with the outcomes of uplift churn models instead of traditional churn prediction models. Traditional churn prediction models do not fully align with the business objective of maximizing the ROMI, i.e. how to maximally reduce customer churn with a given retention effort. Traditional churn prediction models focus on the gross outcome, i.e. whether a customer will churn, independent of the impact of the retention campaign. This might be problematic as there will be customers predicted to churn by the traditional churn prediction model that cannot be retained with the retention campaign. These customers should be excluded from the retention campaign as contacting them is a waste of resources. The solution to this issue is provided by uplift churn models that focus on whether a customer has the intention to churn and will be retained when targeted with the retention campaign. In sum, targeting persuadable customers prescribed by an uplift model will not only reduce churn, but will do so with lower marketing expenditure. Our results suggest that uplift modeling is a viable strategy for B2B firms to fight against customer attrition.

Second, analytics teams should be stimulated to use segmented uplift prediction models as our study proves that these algorithms are not performing worse than their one-model-fits-all-approach counterparts like uplift DT (Rzepakowski & Jaroszewicz, 2012a), uplift LR (Lo, 2002), and uplift RF (Guelman et al., 2012). Extant marketing literature has shown that segmentation is a valuable tool in a variety of contexts, as several groups of customers show different response behavior. For instance, Sood & Kumar (2017) show that a time-based customer segmentation combined with a profitability-based perspective result in different product adoption behaviors. Our study builds further on this idea by showcasing that the uplift LLM approach can generate a data-driven segmentation (step 1) for which each segment has specific churn drivers (step 2).

Third, B2B firms should realize that managing AI tools for customer retention management is not merely a pure black-box methodological issue. Marketing managers need transparent and reliable insights offered by AI tools to steer decisions on which customer is at risk of leaving the firm. Our study showcases outstanding interpretability of the uplift LLM on the global and local perspective. It allows marketers to optimize their retention campaigns by finding the optimal proportion of customers to target, i.e. where the cumulative incremental gains are the highest (global perspective). Further, the uplift LLM algorithm outputs a comprehensible decision table that on the one hand gives insight into the decision rules needed for defining the segments and delivers insight into segment-specific uplift drivers on the other hand (local perspective). This outcome is in line with previous research that found that managing AI might not be viewed as merely a technical issue. Extant research (e.g. Duan et al., 2019; Dwivedi et al., 2021) found that AI technologies should be designed to seamlessly integrate with human being's

capabilities and knowledge in a certain domain. Therefore, AI-explainability is needed for making sound marketing decisions, and optimizing the collaboration between human and AI is needed. Our uplift LLM approach is designed to complement the algorithmic with the interpretability side of AI, and therefore finding a common goal with marketing managers in fighting attrition. In discussing the comprehensible output of the uplift LLM, we show how it can help decision makers design better customer retention management initiatives. The segmented approach uncovers significant differences across segments, and we propose a strategy to target the most fruitful segments.

6.3 Limitations and Future Research Direction

This study also has some limitations that hold valuable avenues for further research. First, the current study leverages a unique, real-world data set in a B2B setting, with more than 6,000 customers. Ideally, the results would be validated with several data sets, but company collaboration to share this type of data is not obvious, because companies realize that these customer data are of strategic importance. Similar to other studies in B2B contexts (Gordini & Veglio, 2017; Tamaddoni Jahromi et al., 2014), we thus rely on a single data set that we review thoroughly. To confirm our results, further research might gather multiple data sets across different B2B contexts. Additionally, a simulation study might help to further explore several impacting factors of uplift LLM in a controlled environment. Second, our study focuses on a B2B setting in which uplift modeling has not been established. It would, however, be interesting to validate the uplift LLM model in a B2C setting or on other applications such as customer acquisition modeling or offline marketing promotions. Third, the uplift LLM demonstrates the merits of a segmented approach to uplift modeling. We use an uplift DT with the Kullback-Leibler divergence splitting criterion to define segments; other uplift DT splitting criteria and intelligent ways to segment the customer base might be tested too. **Further research could for example experiment with uplift specific splitting criteria or unsupervised approaches.** The uplift LLM algorithm could also be adjusted based on learning to rank approaches, originated in the information retrieval domain, to learn the ranking of customers more explicitly (Devriendt, Van Belle, et al., 2020). Fourth, our model unveils a segment for which the uplift is negative, and the retention campaign increases its churn probability. This insight could be leveraged to design multiple, segment-specific incentives to combat customer churn. To incorporate multiple incentives, our single-treatment approach would need to be extended to a multi-treatment setting, which is beyond the scope of the current study.

7. References

- Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*. <https://doi.org/10.1509/jmr.16.0163>
- Ascarza, E., & Hardie, B. G. S. (2013). A Joint Model of Usage and Churn in Contractual Settings.

- Marketing Science*, 32(February), 570–590. <https://doi.org/10.1287/mksc.2013.0786>
- Ascarza, E., Iyengar, R., & Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*. <https://doi.org/10.1509/jmr.13.0483>
- Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schriff, R. (2018). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. *Customer Needs and Solutions*. <https://doi.org/10.1007/s40547-017-0080-0>
- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7353–7360. <https://doi.org/10.1073/pnas.1510489113>
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2), 1179–1203. <https://doi.org/10.1214/18-AOS1709>
- Barfar, A., Padmanabhan, B., & Hevner, A. (2017). Applying behavioral economics in predictive analytics for B2B churn: Findings from service quality data. *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2017.06.006>
- Cao, Y., Xu, C., & Gu, H. (2017). Untangle Customers' Incrementality Using Uplift Modeling with a Case Study on Direct Marketing. *MWSUG2017*.
- Chen, K., Hu, Y. H., & Hsieh, Y. C. (2015). Predicting customer churn from valuable B2B customers in the logistics industry: a case study. *Information Systems and E-Business Management*. <https://doi.org/10.1007/s10257-014-0264-1>
- Chickering, D. M., & Heckerman, D. (2000). A Decision Theoretic Approach to Targeted Advertising. *Uncertainty in Artificial Intelligence Proceedings*.
- Colgate, M. R., & Danaher, P. J. (2000). Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. *Journal of the Academy of Marketing Science*, 28(3), 375–387. <https://doi.org/10.1177/0092070300283006>
- De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 1–13. <https://doi.org/10.1016/j.ejor.2018.02.009>
- Degbey, W. Y. (2015). Customer retention: A source of value for serial acquirers. *Industrial Marketing Management*, 46, 11–23. <https://doi.org/10.1016/j.indmarman.2015.01.002>
- Devriendt, F., Berrevoets, J., & Verbeke, W. (2020). Why you should stop predicting customer churn and start using uplift models. *Information Sciences*. <https://doi.org/10.1016/j.ins.2019.12.075>

- Devriendt, F., Moldovan, D., & Verbeke, W. (2018). A Literature Survey and Experimental Evaluation of the State-of-the-Art in Uplift Modeling: A Stepping Stone Toward the Development of Prescriptive Analytics. *Big Data*, 6(1), 13–41. <https://doi.org/10.1089/big.2017.0104>
- Devriendt, F., Van Belle, J., Guns, T., & Verbeke, W. (2020). Learning to rank for uplift modeling. *IEEE Transactions on Knowledge and Data Engineering*.
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48(January), 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57(August), 0–1. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Farrokhi, A., Shirazi, F., Hajli, N., & Tajvidi, M. (2020). Using artificial intelligence to detect crisis related to events: Decision making in B2B by artificial intelligence. *Industrial Marketing Management*, 91(September), 257–273. <https://doi.org/10.1016/j.indmarman.2020.09.015>
- Ford, D. (2004). Guest editorial: The IMP group and international marketing. *International Marketing Review*, 21(2). <https://doi.org/10.1108/02651330410531358>
- Gordini, N., & Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*. <https://doi.org/10.1016/j.indmarman.2016.08.003>
- Gubela, R., Bequé, A., Lessmann, S., & Gebert, F. (2019). Conversion Uplift in E-Commerce: A Systematic Benchmark of Modeling Strategies. In *International Journal of Information Technology and Decision Making* (Vol. 18, Issue 3). <https://doi.org/10.1142/S0219622019500172>
- Gubela, R., & Lessmann, S. (2020). Interpretable Multiple Treatment Revenue Uplift Modeling. *Americas Conference on Information Systems*, 0–10.
- Gubela, R., Lessmann, S., & Jaroszewicz, S. (2020). Response transformation and profit decomposition for revenue uplift modeling. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2019.11.030>
- Guelman, L., Guillén, M., & Pérez-Marín, A. M. (2012). Random Forests for Uplift Modeling: An Insurance Customer Retention Case. *Lecture Notes in Business Information Processing*, 115, 123–

- Guelman, L., Guillén, M., & Pérez-Marín, A. M. (2015). A decision support framework to implement optimal personalized marketing interventions. *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2015.01.010>
- Hansen, H., M. Samuelsen, B., & E. Sallis, J. (2013). The moderating effects of need for cognition on drivers of customer loyalty. *European Journal of Marketing*, 47(8), 1157–1176. <https://doi.org/10.1108/03090561311324264>
- Hansotia, B., & Rukstales, B. (2002). Incremental value modeling. *Journal of Interactive Marketing*. <https://doi.org/10.1002/dir.10035>
- Hanssens, D. M., & Pauwels, K. H. (2016). Demonstrating the value of marketing. *Journal of Marketing*, 80(6), 173–190. <https://doi.org/10.1509/jm.15.0417>
- Harrison-Walker, L. J., & Neeley, S. E. (2004). Customer Relationship Building on the Internet in B2B Marketing: A Proposed Typology. *Journal of Marketing Theory and Practice*, 12(1), 19–35. <https://doi.org/10.1080/10696679.2004.11658510>
- Hopmann, J., & Thede, A. (2005). Applicability of Customer Churn Forecasts in a Non-Contractual Setting. In *Innovations in Classification, Data Science, and Information* (pp. 330–337).
- Jabbar, A., Akhtar, P., & Dani, S. (2020). Real-time big data processing for instantaneous marketing decisions: A problematization approach. *Industrial Marketing Management*, 90(September 2019), 558–569.
- Kane, K., Lo, S. Y. V., & Zheng, J. (2014). Mining for the truly responsive customers and prospects using true-lift modeling: Comparison of new and existing methods. *Journal of Marketing Analytics*, 2(4), 218–238.
- Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial Marketing Management*, 90(May), 493–507. <https://doi.org/10.1016/j.indmarman.2019.05.003>
- Lamm, B. (2020). AI Will Change Marketing In Ways You Didn't Expect. *Forbes*.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, 59(1–2), 161–205. <https://doi.org/10.1007/s10994-005-0466-3>
- Lemmens, A., & Gupta, S. (2020). Managing churn to maximize profits. *Marketing Science*, 39(5), 956–973. <https://doi.org/10.1287/mksc.2020.1229>
- Liu, X. (2020). Analyzing the impact of user-generated content on B2B Firms' stock performance: Big data analysis with machine learning methods. *Industrial Marketing Management*, 86(May), 30–39. <https://doi.org/10.1016/j.indmarman.2019.02.021>

- Lo, V. S. Y. (2002). The true lift model: a novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter*.
- Maldonado, S., Domínguez, G., Olaya, D., & Verbeke, W. (2021). Profit-driven churn prediction for the mutual fund industry: A multisegment approach. *Omega (United Kingdom)*, 100. <https://doi.org/10.1016/j.omega.2020.102380>
- Martínez-López, F. J., & Casillas, J. (2013). Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights. *Industrial Marketing Management*, 42(4), 489–495.
- Michel, R., Schnakenburg, I., & von Martens, T. (2017). Effective customer selection for marketing campaigns based on net scores. *Journal of Research in Interactive Marketing*. <https://doi.org/10.1108/JRIM-10-2015-0080>
- Michels, D., & Dullweber, A. (2014). *Do your B2B customers promote your business? - Bain Brief - Bain & Company*.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. In *Artificial Intelligence*. <https://doi.org/10.1016/j.artint.2018.07.007>
- Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Book*.
- Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business and Industrial Marketing*, 34(7), 1410–1419. <https://doi.org/10.1108/JBIM-10-2018-0295>
- Radcliffe, N. J. (2007). Using control groups to target on predicted lift: Building and assessing uplift models. *Direct Market J Direct Market Assoc Anal Council*, 14–21.
- Radcliffe, N. J., & Surry, P. (1999). Differential response analysis: Modeling true responses by isolating the effect of a single action. *Credit Scoring and Credit Control IV*.
- Rauyruen, P., & Miller, K. E. (2007). Relationship quality as a predictor of B2B customer loyalty. *Journal of Business Research*. <https://doi.org/10.1016/j.jbusres.2005.11.006>
- Rauyruen, P., Miller, K. E., & Barrett, N. J. (2011). Relationship quality as predictor of B2B customer loyalty. *WMSCI 2011 - The 15th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings, I*, 1–6.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 34–58.

- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Russo, I., Confente, I., Gligor, D. M., & Autry, C. W. (2016). To be or not to be (loyal): Is there a recipe for customer loyalty in the B2B context? *Journal of Business Research*, 69(2), 888–896. <https://doi.org/10.1016/j.jbusres.2015.07.002>
- Rzepakowski, P., & Jaroszewicz, S. (2012a). Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2), 303–327. <https://doi.org/10.1007/s10115-011-0434-0>
- Rzepakowski, P., & Jaroszewicz, S. (2012b). Uplift modeling in direct marketing. *Journal of Telecommunications and Information Technology*.
- Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting Your Customers: Who-Are They and What Will They Do Next? *Management Science*. <https://doi.org/10.1287/mnsc.33.1.1>
- Sołtys, M., Jaroszewicz, S., & Rzepakowski, P. (2015). Ensemble methods for uplift modeling. *Data Mining and Knowledge Discovery*. <https://doi.org/10.1007/s10618-014-0383-9>
- Sood, A., & Kumar, V. (2017). Analyzing client profitability across diffusion segments for a continuous innovation. *Journal of Marketing Research*, 54(6), 932–951. <https://doi.org/10.1509/jmr.16.0209>
- Splawa-Neyman, J., Dabrowska, D. M., & Speed, T. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 465–472.
- Stevens, R. P. (2005). B-to-B Customer Retention: Seven Strategies for keeping your customers. *B2B Marketing Trends*, 1–13.
- Tamaddoni Jahromi, A., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 43(7), 1258–1268. <https://doi.org/10.1016/j.indmarman.2014.06.016>
- Tellis, G. J. ., Chandy, R. K. ., MacInnis, D., & Thaivanich, P. (2005). Modeling the Microeffects of Television Advertising: Which Ad Works, When, Where, for How Long, and Why? *Practice Prize Reports Marketing Science J. Marketing Res. J. Marketing Res. Marketing Sci. J. Consumer Res.*
- Vellido, A. (2019). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, 32(24), 18069–18083. <https://doi.org/10.1007/s00521-019-04051-w>
- Verbeke, W., Baesens, B., & Bravo, C. (2017). Profit-Driven Business Analytics. In *Profit-Driven Business Analytics*. <https://doi.org/10.1002/9781119443179>
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn

- prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211–229.
- Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3), 2354–2364. <https://doi.org/10.1016/j.eswa.2010.08.023>
- Wiersema, F. (2013). The B2B agenda: The current state of B2B marketing and a look ahead. *Industrial Marketing Management*, 42(4), 470–488.
- Williams, P., Khan, M. S., Ashill, N. J., & Naumann, E. (2011). Customer attitudes of stayers and defectors in B2B services: Are they really different? *Industrial Marketing Management*, 40(5), 805–815. <https://doi.org/10.1016/j.indmarman.2010.12.001>
- Wind, Y., & Webster, F. E. (1972). *Organizational buying behavior* (Prentice-Hall (ed.)).
- Yu, X., Guo, S., Guo, J., & Huang, X. (2011). An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*, 38(3), 1425–1430. <https://doi.org/10.1016/j.eswa.2010.07.049>
- Zaniewicz, Ł., & Jaroszewicz, S. (2017). Lp -Support vector machines for uplift modeling. *Knowledge and Information Systems*. <https://doi.org/10.1007/s10115-017-1040-6>
- Zhang, C., Wang, X., Cui, A. P., & Han, S. (2020). Linking big data analytical intelligence to customer relationship management performance. *Industrial Marketing Management*, 91(16), 483–494. <https://doi.org/10.1016/j.indmarman.2020.10.012>
- Zhang, Q., & Zhu, S. C. (2018). Visual interpretability for deep learning: A survey. *ArXiv*, 19(1423305), 27–39.
- Zhao, Y., Fang, X., & Simchi-Levi, D. (2017). Treatments and General Response Types. *Proceedings of the 2017 SIAM International Conference on Data Mining*, 588–596.