Revenue Uplift Modeling

Conference Paper · December 2017

Revenue Uplift Modeling

Completed Research Paper

Robin Gubela
Chair of Information Systems
Humboldt University of Berlin
Spandauer Str. 1, 10178 Berlin
robin.gubela@gmx.de

Stefan Lessmann
Chair of Information Systems
Humboldt University of Berlin
Spandauer Str. 1, 10178 Berlin
stefan.lessmann@hu-berlin.de

Johannes Haupt
Chair of Information Systems
Humboldt University of Berlin
Spandauer Str. 1, 10178 Berlin
johannes.haupt@hu-berlin.de

Annika Baumann
Chair of Information Systems
Humboldt University of Berlin
Spandauer Str. 1, 10178 Berlin
annika.baumann@hu-berlin.de

Tillmann Radmer
Chair of Information Systems
Humboldt University of Berlin
Spandauer Str. 1, 10178 Berlin
tillmann.radmer@gmail.com

Fabian Gebert
Data Science Department
Akanoo GmbH
Mittelweg 121, 20148 Hamburg
fabian@akanoo.com

Abstract

The measurement of the effectiveness of a marketing campaign is a challenging task. Whereas established approaches do not consider causality, uplift models take into account which customers display some behavior because of the marketing action and model this target as differential response. The paper categorizes existing approaches toward uplift modeling collected from different fields into a conceptual taxonomy to establish the state-of-the-art and proposes a novel approach named revenue uplift modeling. Contrary to existing approaches, which model incremental response, revenue uplift models predict the incremental revenue with the goal to maximize the gain per marketing incentive for heterogeneous customers. An experiment based on a large real-world dataset of e-commerce shops across several industries provides a benchmark on the choice of machine learning methods to implement the identified uplift modeling approaches and demonstrates the effectiveness of the revenue uplift model in a real-world e-commerce environment.

Keywords: Data mining, Decision analysis, Customer targeting, Marketing campaign planning

Introduction

Advertisements are omnipresent. A recent study of media use and advertisement exposure points out that the typical U.S. adult encounters a total of about 153 advertisements each day (Media Dynamics, Inc., 2014). Accordingly, advertising investment is substantial. In 2015 alone, about 161 billion USD were spent on digital advertising across all Internet-connected devices worldwide (eMarketer, 2016). To ensure accountability of investments and allocate marketing resources efficiently, it is important to measure the effectiveness of advertisement and more generally marketing communication. This remains a challenging undertaking. In particular, for a marketing stimulus to be judged effective, it should lead a customer to perform an intended action (e.g., purchase a product, download an app, sign-up for a newsletter, etc.). Data on individual customers, ad exposure, and customer conversion is often available, especially in online marketing. However, a co-occurrence of customer behavior and ad exposure is insufficient to conclude that the ad caused the observed customer action. Establishing such causal link represents a major obstacle in measuring marketing effectiveness (Rzepakowski and Jaroszewicz, 2012b).

A large body of literature examines data-driven models for customer targeting in offline settings, e.g. catalog marketing, and e-commerce, e.g. real-time couponing. Literature surveys in customer relationship management (CRM) (Ngai et al., 2009) and specific CRM tasks such as churn modeling (Verbeke et al., 2012) or direct marketing (Bose and Xi, 2009) illustrate the popularity of supervised machine learning methods to develop targeting models. Using data from a past campaign including explanatory variables (e.g., customer characteristics) and a response variable (e.g., whether a customer has churned or bought an item from a sales catalog), a learning method estimates a functional relationship between the response and explanatory variables. The estimated model facilitates predicting the value of the response from the explanatory variables (e.g., for novel customers). The uplift modeling community calls this approach response modeling because the model learns to recognize customers that have responded in the past (Radcliffe and Surry, 1999). Although widely used in the literature, the response modeling approach is flawed in that it disregards causality. For example, a customer may receive a special offer and buy the advertised product subsequently, but she may have bought the same product without the discount (Radcliffe, 2007). Uplift models overcome this inadequacy through predicting differential response; that is whether the customer buys because of the offer (Kane et al., 2014). Therefore, the uplift concept quantifies the true effectiveness of a campaign (Lo, 2002).

Uplift models support marketing managers in campaign planning and targeting marketing communication to customers who would not convert without the incentive (Soltys et al., 2015). This implies that an uplift model aims at estimating a causal link between a marketing action (e.g., offering a customer a special deal) and customer behavior (e.g., accepting the offer). Estimating the change in customer behavior that results from a solicitation, uplift models are especially suitable to support targeting decisions in campaign planning and increase campaign profitability (Siegel, 2011; Radcliffe and Surry, 2011; Larsen, 2010). An analysis of the literature on uplift models in marketing reveals that existing approaches focus on conversion and churn modeling, the goal of which is to win novel customers and prevent customer defection, respectively (Park and Park, 2016; Verbeke et al, 2012). In this regard, the strategic marketing objective behind current models is market share. In terms of the underlying uplift modeling methodology, conversion and retention models predict a dichotomous response variable using classification methods.

The paper extends previous literature through introducing revenue uplift modeling. A revenue uplift model predicts the incremental revenue that results from targeting a customer with a marketing message. In many applications, customers differ in their spending (Jacobs et al., 2016). Modeling revenue uplift accounts for this type of heterogeneity, which a conversion model is unable to accommodate. Therefore, revenue uplift modeling reflects the value-based idea of CRM (Reinartz and Kumar, 2003). Considering the focus of prior work on conversion uplift for customer acquisition and retention, revenue uplift modeling is also a relevant addition in that it provides an approach to target marketing campaigns that aim at increasing customer spending such as cross-/up-selling campaigns (Netessine et al., 2006).

In summary, the paper makes three contributions. First, existing approaches toward uplift modeling are categorized to sketch the field and highlight conceptual differences. This is useful since uplift modeling is still a niche topic in the academic literature. Second, a novel modeling strategy is proposed to predict revenue uplift. Targeting marketing communication so as to maximize revenue uplift is especially suitable for campaigns that aim at growing existing customers (e.g., cross-/up-selling). In that sense, the new approach naturally complements existing solutions for conversion and retention uplift modeling, which are geared toward customer acquisition and preventing customer defection, respectively. Third, a comprehensive empirical evaluation is carried out to demonstrate the effectiveness of the new uplift model in a real-world e-commerce environment. In addition to assessing alternative uplift modeling strategies, the experiment also provides original insights into the comparative performance of alternative machine learning methods for classification and regression to implement uplift models.

The results of the experiment confirm the effectiveness of the proposed approach. For the large e-commerce data set employed in the study, which comprises campaign results and actual sales from several e-shops across different industries, the new revenue uplift model provides the largest increase in incremental revenue and outperforms the benchmarks considered in the study. Although the model's uplift estimate is not unbiased, it's bias is somewhat lower compared to revenue models from challenger approaches because of the unique modification of the target variable. Furthermore, previous (conversion) uplift models are found ineffective in that they fail to outperform a simple response modeling approach. These results provide strong evidence that revenue uplift modeling is a useful technique to target marketing communication to responsive customers.

The paper is organized as follows: The next section introduces uplift modeling fundamentals before relevant prior work is revised. Subsequent sections elaborate on the proposed methodology and the experimental design. Afterwards, empirical results for conversion and revenue uplift modeling are reported, integrated, and discussed. The paper then concludes with a summary and outlook to future research.

Uplift Modeling Fundamentals and Process Model

The philosophy of an uplift-based targeting approach is that marketing communication should concentrate on customers who are influenced by the campaign (Rzepakowski and Jaroszewicz, 2012a). Rather than predicting customers' response probability and soliciting likely responders, as done in response/churn modeling (Chen et al., 2015; Neslin et al., 2006), the targeting decision should be based on the change in customers' likelihood to respond due to being targeted. These customers are called *Persuadables* in the literature and constitute the only group worth a marketing investment (Kane et al., 2014).

Identifying the treatment effect requires information on the response of individuals who have not received the treatment. Since each individual cannot be simultaneously treated and not-treated, the treatment effect is identified using the outcome observed in a control group. Therefore, an experimental setting with randomized treatment and control group is a prerequisite to develop an uplift model. This may be seen as a disadvantage compared to response modeling. However, in marketing and especially online marketing obtaining control group information is relatively straightforward. In particular, A/B testing is a popular approach to perform random experiments in e-commerce. For example, a website owner may randomly assign visitors to different groups each of which get to see a different version of the homepage, hoping that the random assignment facilitates causal statements as to the effectiveness of the different page versions.

To the best of our knowledge, the literature on uplift models relies exclusively on this approach of treatment-control group comparisons to establish causality. However, it is important to note that A/B tests may fail to implement a statistically sound random experiment, especially in high-dimensional settings, which may invalidate conclusions on causal links, and may be impractical in large-scale settings where a vast number of tests are performed in parallel (Kohavi et al., 2013). For consistency with previous literature on uplift modeling, we focus on randomized trials in the form of A/B tests as vehicle to establish causal relationships. Evaluating other causal inference procedures such as, for example propensity scores or instrumental variables (Imbens, 2004) for uplift modeling is a fruitful area of future research but beyond the scope of this paper.

A/B tests are used to estimate the marginal performance increase due to a marketing incentive, the uplift, but also facilitate the training of models based on this metric (e.g. Radcliffe and Surry, 2011). In Figure 1, we summarize the concept of uplift with the four-fields target matrix from Kane et al. (2014). Response models distinguish between responders and non-responders (left and right column) irrespective of the actual effect of treatment. The goal of uplift models is to use the information on the control population to also account for variation in response rate dependent on whether the marketing incentive was received. In other words, uplift models identify likely treatment responders (upper right), who respond specifically due to the marketing incentive and would not respond otherwise.

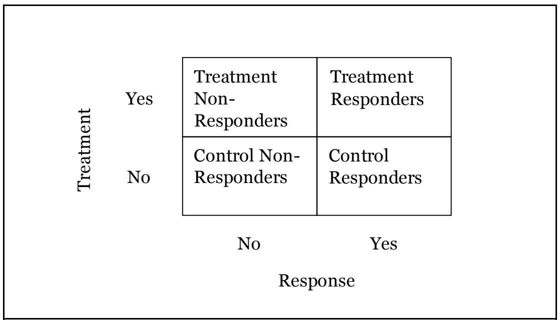

Figure 1. The Four-Fold Target Matrix

To formalize the methodological difference between uplift and response modeling, let $\boldsymbol{X_i} = (X_1, \dots, X_n) \in \mathbb{R}^{\mathrm{n}}$ be a vector of characteristics (i.e., explanatory variables) of customer i, and let $Y_i \in \{0,1\}$ be a binary response variable, for example whether customer *i* bought a product in a previous campaign. Uplift models build on the concept of A/B testing, meaning that customers are divided into two groups: treatment and control (Kohavi et al., 2009). Let $T_i \in \{0,1\}$ be an indicator variable of the group membership of customer i, with $T_i = 0$ and $T_i = 1$ indicating membership to the control and treatment group, respectively. Then, with $P(Y_i|\boldsymbol{X_i}, T_i = 1)$ and $P(Y_i|\boldsymbol{X_i}, T_i = 0)$ denoting customer-level probabilities in the corresponding groups, traditional response models predict the conditional probability $P(Y_i|\boldsymbol{X_i}, T_i = 1)$, whereas an uplift model predicts the change in behavior resulting from a treatment $P(Y_i|\boldsymbol{X_i}, T_i = 1) - P(Y_i|\boldsymbol{X_i}, T_i = 0)$. In marketing, the treatment can be an advertisement, direct mail, or some other marketing action. Many supervised learning methods are available to estimate conditional response $P(Y_i|\boldsymbol{X_i})$ (Hastie et al., 2009).

An intuitive approach to develop an uplift model involves estimating two models to predict $P(Y_i|\boldsymbol{X_i}, T_i = 1)$ and $P(Y_i|\boldsymbol{X_i}, T_i = 0)$, respectively. Campaign planners can then calculate the uplift for individual customers as the difference between these models' predictions and target customers in the order of their estimated uplift. This approach is known as the two-model or indirect approach (e.g. Lo and Pachamanova, 2015). The indirect approach embodies the objective to maximize responders in the treatment group while minimizing control group responders but suffers important limitations. First, estimating two models increases computational costs. Second and more importantly, the distribution of the difference of the probabilities is often different from the distribution of the respective probabilities, which causes bias and poor model performance (Chickering and Heckerman, 2000; Rzepakowski and Jaroszewicz, 2012b).

The shortcomings of the indirect approach led to the development of improved uplift modeling regimes. Furthermore, the distinction of treatment and control group customers has implications for all stages of the model development process. To systematize related work in the field and identify the contribution of the paper, the uplift modeling process for marketing (UMPM) is introduced in Figure 2. The process model is based on the well-known KDD process (Fayyad et al., 1996) and ex

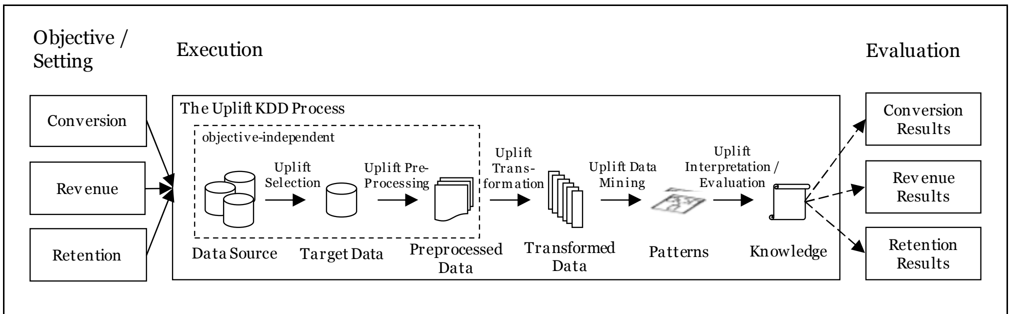

Figure 2. Uplift Modeling Process for Marketing

Prior work on uplift modeling in marketing focuses predominantly on conversion uplift. Only two studies examine retention uplift (Guelman, 2014; Siegel, 2013). Revenue uplift modeling has not been considered at all but is introduced here. The UMPM strives to raise the awareness of different modeling objectives in campaign planning. To that end, the UMPM distinguishes three stages: (1) selecting a suitable business objective out of conversion, revenue, or retention modeling for a specific campaign, (2) pursuing the chosen objective to gain insight (along the stages of the UMPM), and finally (3) evaluating results in recognition of the campaign objective to identify and target the truly responsive customers with the next marketing campaign.

All campaign goals in Figure 2 imply a profit objective. Winning new customers with conversion modeling increases revenues, even if the magnitude of the increase is not the focus of attention. Cross-/up-selling campaigns and campaigns aiming at customer growth in general maximize revenue directly, while campaigns to prevent customer attrition sustain future revenues. Clearly, none of the objectives and underlying uplift modeling strategies is generally preferable. Rather, the point of the UMPM is to stress that campaign planners who use uplift models to support targeting decisions should choose a definition of uplift that best matches the campaign objective and then develop a corresponding model. For example, when customer spending varies substantially and there is a small fraction of high value customers, a revenue uplift model will recommend a smaller campaign size than a conversion uplift model, which maximizes incremental sales. The smaller and more focused campaign is likely to be more profitable because it avoids the costs of soliciting low-value customers. This view is supported by the empirical results of this study.

The UMPM has been designed to provide maximal flexibility in the choice of the objective based on the respective specific marketing situations of campaign planners. Therefore, while the goal of the next campaign could be conversion-/retention-related, a rather value-related aim could be operationalized in another post-initiative (or vice versa). This single-campaign focus is typically not supported by other revenue-based models such as the customer lifetime value (CLV) which pre-empt decisions due to long-term strategies. Furthermore, CLV models are typically considered if long-term contractual agreements result from the desired action, which is rather the case for insurance or banking products/services than for fast moving consumer goods in e-retail. This is not least the case because of the accumulated value a customer generates if being locked-into a long-run agreement. In contrast, the buyer-seller relationship in e-commerce is typically rather transactional, which is why CLV models are rarely applied in this field. One might also argue that the focus on long-run customer relationships, as embodied in CLV models, is more geared toward tactic/strategic marketing management, whereas uplift models with their short-term campaign planning objectives (see Figure 2) are a tool for operational marketing planning. For example, measuring the causal influence of a marketing action on customer-level CLV is a complex undertaking., because changes in long-term strategic performance indicators like customer-level CLV and customer equity, respectively, can only be observed in the longer run where a multitude of external factors will simultaneously affect these indicators, leading to serious modeling issues with respect to endogeneity.

Figure 2 indicates that the selection of a campaign objective has methodological implications. Multiple stages in the model development process depend on the objective. Most importantly, the response variable Y_i is dichotomous in conversion and retention modeling (success/failure to convert/retain customer) and continuous in revenue modeling (purchase amount). Accordingly, conversion/retention uplift models require classification methods to estimate conditional response $P(Y_i|\boldsymbol{X_i})$ whereas revenue uplift models use regression methods (Hastie et al., 2009). Subsequent parts of the paper will further detail objective-specific modeling implications.

Related Literature

The review of prior work is organized along the stages of the UMPM (Figure 2). In general, specific modeling challenges arise in uplift modeling due to the estimation of causal effects. For example, the distinction of customers into treatment and control group affects data selection (Kane et al., 2014) as well as preparatory activities including the handling of missing values, outliers and feature selection (Hua, 2016; Yong, 2015; Larsen, 2010; Hansen and Bowers, 2008). It also affects the evaluation of uplift models, which often grounds on a comparison between model predictions for treatment and control group customers (Nassif et al., 2013; Radcliffe and Surry, 2011; Radcliffe, 2007). Data transformation is important for uplift modeling because a suitable transformation of the explanatory variables or the response facilitates predicting uplift using standard learning methods (e.g. Tian et al., 2014; Lo, 2002).

An alternative strategy is to modify existing learning methods. In the spirit of the KDD process, an algorithmic modification exemplifies uplift data mining, which represents the prevailing approach in prior work. Corresponding studies strive to estimate uplift directly using tree-based algorithms with adapted splitting and pruning criteria (e.g. Hansotia and Rukstales, 2002), ensembles of uplift decision trees (Guelman et al., 2015), artificial neural networks (Manahan, 2005), k-nearest neighbours (Larsen, 2010) and support vector machines (Jaroszewicz and Zaniewicz, 2016; Zaniewicz and Jaroszewicz, 2013).

The paper focuses on uplift transformation. Compared to uplift data mining, approaches for response and covariate transformation are generic. As will be detailed below, they facilitate an implementation of the modeling methodology using conventional machine learning methods. Given that this is the first paper to study revenue uplift modeling, it is useful to compare a broad set of different regression methods. Such comparison can identify methods that work well for revenue uplift. Future work could then develop modification of these methods to approach the revenue uplift modeling problem directly. In contrast, it seems less suitable to start the journey into revenue uplift modeling with a modification of one regression method, arbitrarily chosen from a vast space of alternative methods (Hastie et al., 2009).

Literature

Uplift Taxonomy

The uplift transformation framework (Figure 3) formally introduces and contextualizes revenue uplift modeling in the data transformation stage of the UMPM. The tree provides marketing analysts two options, a transformation of the input space (i.e., covariates) or the output space (i.e., the response variable). Response transformation can be further distinguished in terms of the underlying modeling objective.

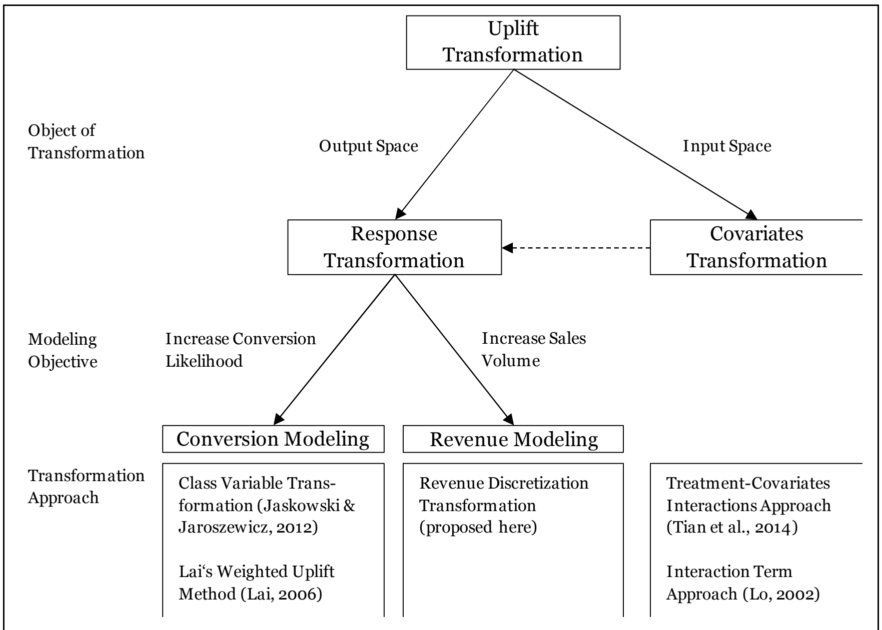

Figure 3. Uplift Transformation Framework

If the objective is to increase conversion or retention rates, the response is a binary indicator variable which equals one if a customer has shown the focal behavior (has converted/churned) and zero otherwise. Response models rely exclusively on this information. Uplift models for conversion also predict a binary response variable but alter the group definition to model incremental conversions. The underlying learning methods are the same as those used in response modeling (e.g., logistic regression, neural networks, etc.). The two main transformation approaches are the class variable transformation (CVT) (Jaskowski and Jaroszewicz, 2012) and Lai's weighted uplift method (LWUM) (Lai, 2006). The paper focuses on the latter approach because recent benchmarking results indicate that it often outperforms alternative techniques (Kane et al., 2014).

Targeting models for revenue uplift transforms an originally continuous response variable (here, the revenue per customer) using information on whether customers were part of the treatment or control group. Depending on the specific transformation strategy, the new response can be continuous or binary. Drawing inspiration from previous work concerning the advantages of classification over regression models in direct marketing (Bodapati and Gupta, 2004), the proposed response discretization approach (RDT) produces a binary modeling target. However, an intermediate step in the novel approach, which grounds on Jaroszewicz (2016), delivers a continuous transformed response variable, which offers an alternative route to develop a revenue uplift model. A methodological difference between the two approaches is that RDT works with classification methods whereas the alternative relies on regression methods.

The literature proposes two approaches for covariate transformation; the interaction term method (ITM) (Lo, 2002) and the treatment-covariates interactions approach (TCIA) (Tian et al., 2014). Conceptually, both approaches are similar and differ only in the scaling of the response and normalization of the explanatory variables. In view of this, the empirical analysis includes the more recent TCIA approach.

Note that covariate transformation can be combined with response transformation. Thus, there are four options to build uplift models using covariate transformation. Either models are built on the untransformed conversion variable (conversion response modeling with modified covariates), the untransformed revenue variable (revenue response modeling with modified covariates), the transformed conversion variable (conversion uplift modeling with modified covariates) or the transformed revenue variable (revenue uplift modeling with modified covariates). The two latter options are illustrated with the dotted arrow between the covariates transformation and response transformation boxes in Figure 3.

Underlying Approaches Detailed

Conversion Response Transformation

The LWUM approach transforms the response variable so as to facilitate the use of conventional classification models to predict conversion uplift. Let $z_{\mathrm{i},c}$ be the binary transformed response of customer i, with c identifying the campaign objective (i.e., conversion). The response $z_{\mathrm{i},c}$ equals one for treatment group customers who convert and control group customers who do not convert. Both states represent a success (Lai, 2006). In all other cases, $z_{\mathrm{i},c}$ is set to zero. Formally, this logic is captured in:

$$\mathrm{z_{i,c}} = \begin{cases} 1 & \text{if } \mathrm{T_i} = 1 \cap \mathrm{Y_{i,c}} = 1 \cup \mathrm{T_i} = 0 \cap \mathrm{Y_{i,c}} = 0 \\ 0 & \text{otherwise} \end{cases}$$

with the transformed response $z_{\mathrm{i},c} \in \{0,1\}$. Recall that $T_i \in \{0,1\}$ is an indicator variable for control/treatment group, $Y_{i,c} \in \{0,1\}$ the original response variable, which captures the status of customer i (no conversion/conversion), and $i = 1, \dots, N$ indexed customers in a past campaign of size N and $\boldsymbol{X_i} = (X_1, \dots, X_n) \in \mathbb{R}^n$ is a vector of covariates. With LWUM, uplift is defined as:

$$Uplift_i^{Lai} = P\big(z_{i,c} = 1 \big| \boldsymbol{X_i}\big) \cdot \mathrm{w_{pos}} - P\big(z_{i,c} = 0 \big| \boldsymbol{X_i}\big) \cdot w_{neg}$$

where $\mathrm{w_{pos}}$ and $\mathrm{w_{neg}}$ are weighting parameters determined by the ratio of positive or negative cases in the data, respectively. For $\mathrm{w_{pos}} = w_{neg} = 1$, this approach reduces to the CVT introduced by Jaskowski and Jaroszewicz (2012).

Revenue Response Transformation

From an analytical point of view, the key feature that distinguishes conversion and revenue uplift is the target variable: Instead of transforming the (binary) conversion variable $Y_{i,c}$, the (continuous) revenue variable $Y_{i,r}$ is subject to transformation. In particular, let $Y_{i,r} \in \mathbb{R}$ be the original response revenue variable capturing sales revenue of customer i, with r once again indicating the primary objective of a campaign.

The proposed RDT approach for revenue uplift modeling is based on the concept to discretize a continuous response in order to decrease the bias due to incorrect model specification and increase prediction accuracy (Bodapati and Gupta, 2004). Although the authors consider a response modeling setting, their finding appears relevant for uplift modeling as well. When correctness of a model's specification cannot be ensured, which is often the case in real-world data due to factors such as omitted variables, the resulting bias in OLS estimation can be reduced through a discretization of the target variable at the expense of an increase in variance. In large sample sizes, the importance of variance, however, diminishes. Bodapati and Gupta (2004) gain this insight in simulation experiments with a maximum of 20,000 observations. Much larger sample sizes occur when targeting marketing communication in online environments and/or running campaigns to increase sales in e-commerce.

The logic behind value discretization can be illustrated with the sales situation of a book club (Bodapati and Gupta, 2004). Instead of predicting the annual number of books for all customers individually, the managerial challenge is to predict whether this number exceeds a pre-defined threshold. The actual task is then to determine the value of a discretizing function, $d(y)$, which the authors define as:

$$\mathrm{d(y)} = \begin{cases} 0 & \text{if } \mathrm{y} \in (0, \mathrm{y_{threshold}}] \\ 1 & \text{if } \mathrm{y} \in (\mathrm{y_{threshold}}, \infty) \end{cases}$$

with $y_{threshold}$ as the set value of the absolute number of books in the example. Supervised classification models facilitate estimation of this function (Bodapati and Gupta, 2004).

The RDT approach proposed in this paper combines the idea of revenue uplift modeling with the target design from conversion modeling in a multi-layer transformation scheme. The revenue variable $Y_{i,r}$is first transformed to obtain $z_{\mathrm{i},r}$ (Jaroszewicz, 2016) and then this variable is discretized to receive $z_{\mathrm{i,rg}}$. More formally, the two-step transformation corresponds to:

$$\mathrm{z_{i,r}} = \begin{cases} +\mathrm{Y_{i,r}} & \text{if } \mathrm{T_i} = 1 \cap \mathrm{Y_{i,r}} > 0 \cup \mathrm{T_i} = 1 \cap \mathrm{Y_{i,c}} = 1 \\ -\mathrm{Y_{i,r}} & \text{if } \mathrm{T_i} = 0 \cap \mathrm{Y_{i,r}} > 0 \cup \mathrm{T_i} = 0 \cap \mathrm{Y_{i,c}} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

with $z_{\mathrm{i},r} \in \mathbb{R}$ as the transformed revenue that captures additional information from the group membership indicator. In particular, $z_{\mathrm{i},r}$ is equal to the original sales revenue for treatment group customers who made a purchase, equal to the negative sales revenue for control group customers who made a purchase, and zero otherwise. This transformation produces a novel response variable for direct uplift modeling. A single regression model suffices to predict $z_{\mathrm{i},r}$ which itself possesses all necessary information for uplift predictions. For RDT, however, $z_{\mathrm{i},r}$ is only an intermediate step. Rather than predicting $z_{\mathrm{i},r}$ with regression methods, a discretization procedure on $z_{\mathrm{i},r}$ facilitates use of classification methods and, more importantly, has the option to capitalize on the advantages of value discretization (Bodapati and Gupta, 2004).

$$\mathrm{z_{i,rg}} = \begin{cases} 0 & \text{if } \mathrm{z_r} \in (-\infty, 0] \\ 1 & \text{if } \mathrm{z_r} \in (0, \infty) \end{cases}$$

where $z_{\mathrm{i,rg}} \in \{0,1\}$. The key differentiating factors to Bodapati and Gupta's (2004) discretization are that the response variable has been pre-transformed and that negative numbers are captured in $z_{\mathrm{i,rg}}$, because customers who converted without having received a certain treatment are included. This points out that in $z_{\mathrm{i,rg}}$ information related to the treatment and control group is provided which underlines its characteristic of reflecting change in behavior because of having received a treatment.

The reason why the threshold has been set to zero is related to the objective in the context of uplift modeling. A "failure" is defined by $z_{\mathrm{i,rg}} = 0$. Customers who display the behavior intended by the marketer but without having received the treatment ($z_{\mathrm{i},r} = -Y_{i,r}$) and customers from both treatment and control group with zero purchases ($z_{\mathrm{i},r} = 0$) belong to this category. In contrast, "success" is related to customers who have purchased a product with the causal connection to the campaign treatment ($z_{\mathrm{i},r} = +Y_{i,r}$). This group is the only one that fulfills the condition $0 < z_{\mathrm{i},r} < \infty$ since the price of a product always starts at one cent and is never infinite. Compared to other approaches such as CVT and LWUM, RDT defines "success" differently in terms of the four-fields target matrix presented in Figure 1. In this regard, the only group to target depicts the treatment responders and not, in addition, control non-responders.

Covariates Transformation

Covariates transformation deals with the transformation of the input space. In case of TCIA, a dummy variable $T_i^* \in \{-1; +1\}$ is created. Its value depends on whether the customer has been in the treatment or control group. Then, T_i^* is multiplied with each of the n covariates to determine the interaction term, i.e. $T_i^* * X_i^*$ where X_i^* is modified using a mean centering procedure (Long, 1994). This additional term is taken into account when building uplift models. Following the idea by Lo (2002), the general design to model uplift is $E(Y_i|X_i) = f(T_i, X_\mathrm{i}, T_i * X_i)$ which can be further substituted into $E(Y_i|X_i, T_i = 1) - E(Y_i|X_i, T_i = 0)$. In the next step, TCIA takes each element from the input space and transforms it using $Z_i = T_i^* X_i^*/2$ which is used to predict the response based upon the modified covariates Z_i.

Experimental Design

Data and Experimental Setting

The experimental setting is based on a real-time targeting process in e-commerce. When customers visit the website of an e-commerce shop, a subset of selected customers receive an e-coupon at some point during their session with a discount of 10% off the final basket value. Each targeted customer receives a unique coupon code which needs to be used during the check-out process in the basket to activate the discount. Coupons are commonly used in digital marketing to simulate conversion and generate additional sales (e.g., Khajehzadeh et al. 2014).

While clicking through the website, a visitor is randomly assigned to either being scored by a random process or by a model, i.e. all customers are subject to pre-screening determining if they are eligible for the coupon campaign. Only those customers are further considered who have a high likelihood of responding to the coupon. In the next step, customers having a high likelihood of responding are randomly assigned to the treatment or control group. Customers in the treatment group receive a coupon, those in the control group do not receive a coupon. This process provides the treatment and control setting required for uplift modelling (Figure 4). The filtering stage, which identifies customers with a high likelihood of response, creates a selection bias towards more likely purchasers in the overall group resulting in a quasi-experiment.

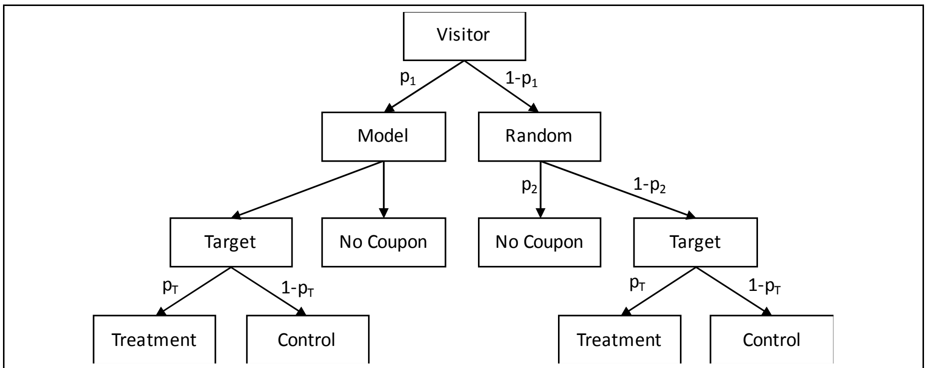

Figure 4. Treatment / Control Group Assignment Process for the Dataset

A partner from industry provided the real-world data which is based on twenty-five different e- shops. There are 3,051,990 observations per variable and 62 variables. Each observation represents an individual customer session. The variables mainly capture customer-specific information such as key areas of the websites visited, including related length of time information. The data also covers the group membership indicator, shop-ID, time stamp together with information on (raw) conversions and basket values.

Table 1 summarizes (i) the fraction of visitors in the treatment and control group, (ii) how many visitors of each group have made a purchase, and (iii) the overall uplift on the dataset based on the group differences in conversion rates. From the table's last column, it can be concluded that the overall uplift on the real-world dataset for the experiments is low. This suggests that the specific coupon offer is not particularly effective in increasing conversion behavior. However, this does not affect the suitability of the data since the focus of the paper is on uplift modeling strategies and thus the relative gain in conversions/revenues due to an improved targeting strategy.

Since the primary objective of the paper is to introduce revenue uplift modeling and the novel RDT approach in particular, it is interesting to examine the consequences of the steps in RDT on the revenue distribution. This analysis is shown in Figure 5. Besides the group membership distribution (left panel), the distribution of the baseline revenue response Y_r is presented (upper right). Moreover, the two smaller plots on the bottom highlight the distributions of the transformed revenue response without discretization (z_r) and after discretization (z_{rg}), respectively. Note that the analysis is based on a sub-sample of the whole data set (~ 420,000 observations) which is representative to the data used in the empirical study of this paper.

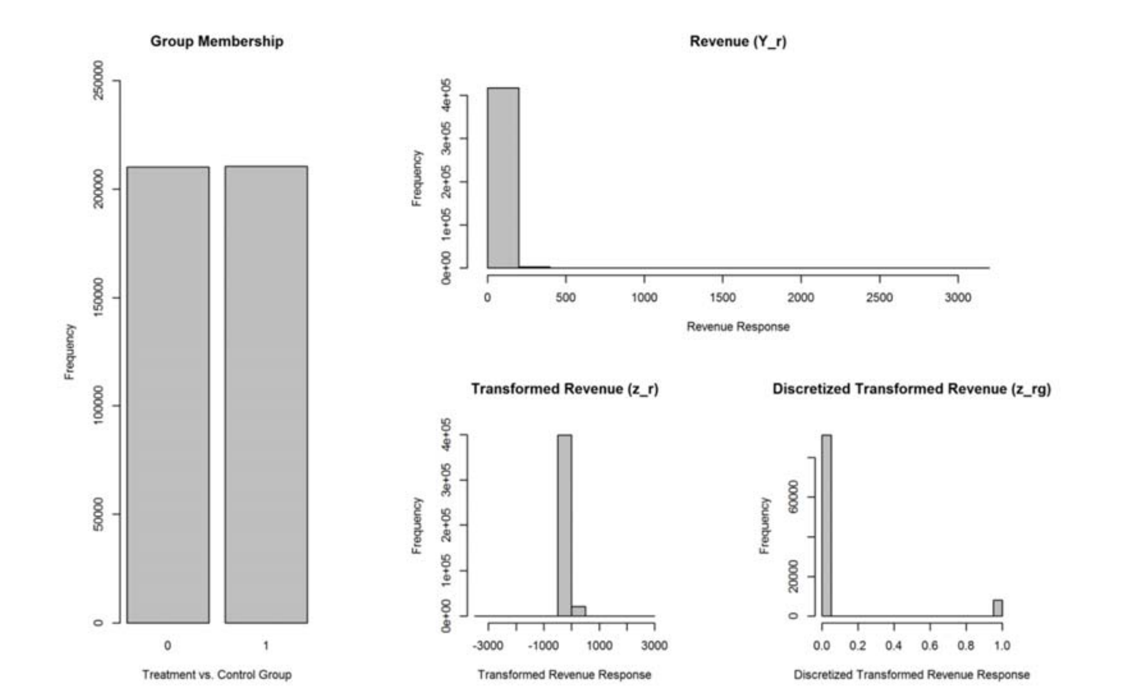

Figure 5. Implications on Distributions of Revenue Response Transformations

The horizontal axes for both the Y_r and z_r plots extend to show the most extreme observed values, even if their frequency is too low to be displayed without scaling. Thus, for Y_r the minimum for revenue is 0€ and its maximum is around 3,000€ and for z_r the minimal value is -3,000€ and the maximal value 3,000€ (due to the transformation logic of the first step of RDT). It is noteworthy that there are few cases where a treatment led to a purchase of a high-priced product. The same is true for control group customers who have purchased a product starting in the price category of 1,000€ (i.e., generating high revenue hereby). This is because the high values (with and without inverted sign) occur so rarely. Since the comparably most frequent values of z_r are either negative or zero, the discretized transformed response z_{rg} is mainly zero. Only in few cases, i.e. when $z_r \in (0, \infty)$, it holds that $z_{\text{rg}} = 1$. This is visualized in the bottom right chart.

Table 1. Uplift on Dataset

Group	Share on Dataset	Number of Observations	Number of Converters	Conversion Rate	Uplift
Treatment	74.9%	2,285,835	175,791	7.69%	0.22%
Control	25.1%	766,155	57,285	7.47%	
Total	100%	3,051,990	233,076		

Base Learners

Alternative uplift modeling approaches are implemented using supervised learning methods. Table 2 lists the methods that have been considered in the experiments. The selection of methods includes well-established individual learners (e.g. logistic regression and tree-based learners) and ensemble algorithms (e.g. random forest and gradient boosting). Interested readers find a comprehensive description of these methods in Hastie et al. (2009). In addition, Table 2 includes some methods that have recently shown promising results, especially in medical and biological informatics research (e.g. Extremely Randomized Trees, compare Nattee et al., 2017; Soltaninejad et al., 2016; Gotz et al., 2014) or seem to be often overlooked despite their advantages (e.g. Theil-Sen Regression, see Fernandes and Leblanc, 2005).

Many learning methods exhibit meta-parameters to adapt an algorithm to a particular data set (Hastie et al., 2009). Such parameters are tuned using grid-search, for which candidate parameter values have been obtained from literature (Baumann et al., 2015).

Validation Strategy

The whole dataset has been partitioned into a training set (~40%), a meta-parameter optimization set (~30%) and a validation set (~30%). In a first step, the models of all approaches are built on the training set and tested on the parameter optimization set to identify the optimal parameter configuration for the respective models. In a second step, the best models are trained on the training and parameter optimization set together (covering 70% of the whole dataset) and tested on the validation sample.

Performance Measures

Measures to assess predictive models are based on a comparison of actual and predicted outcomes for every individual unit of observation (Hastie et al., 2009). In uplift modeling, however, such comparison is impossible since no customer can receive and not receive a treatment at the same time (Radcliffe and Surry, 2011; Radcliffe, 2007). This phenomenon is known as the fundamental problem of causal inference (Holland, 1986). To evaluate uplift models, Qini curves and the corresponding Qini values have been developed. They can be considered an extension of cumulative gain charts and the corresponding Gini coefficient, which facilitate an assessment of response models (Radcliffe, 2007). Gain analysis assesses models in terms of cumulative increase of responses that follow from a model-based compared to a random targeting. For the standard lift metric, *gain* is defined as the number of conversions or the value of these conversions for response and revenue models, respectively, while the uplift metric considers the *incremental* or relative gain as compared to the control group.

The performance of uplift models is visualized using Qini curves by plotting the incremental gain against the percentage of the population that is targeted. Incremental gain is determined by, first, ordering the population by their model score and segmenting customers into groups with decreasing predicted response probability. Second, the incremental gain within each segment is calculated as the difference between responders (or revenue) in the treatment group and control group adjusted for the size of the groups.

Table 2. Base Learners

Conversion Modeling	Revenue Modeling
Logistic Regression (LogR)	Linear Regression (LinR)
Calibrated Linear Support Vector Machine (SVM)	Ridge Regression (Ridge)
k-Nearest-Neighbors (KNN)	Lasso Lars Regression (LL)
Naïve Bayes (NB)	Stochastic Gradient Descent for Regression (SGDR)
Stochastic Gradient Descent for Classification (SGDC)	Theil-Sen Regression (TS)
Random Forest for Classification (RFC)	Random Forest for Regression (RFR)
Calibrated Random Forest for Classification (RFC-C)	Extremely Randomized Trees (ERT)
Extremely Randomized Trees (ERT)	
Gradient Boosting for Classification (GBC)	

The Qini coefficient provides a single number of model performance, which is useful to compare alternative models. To calculate the Qini coefficient, the Qini curve of a model is compared to a random model (Radcliffe and Surry, 2011). The performance line of the latter starts in the coordinate system's origin and ends up in $(N,\ n)$ with N as the population size and n as the total incremental number of purchases (conversion modeling) or total incremental revenue (revenue modeling) if everyone is targeted instead of a certain subpopulation (Radcliffe, 2007). The random model poses a useful baseline that relevant models need to outperform to generate value. The Qini values Q is defined as the area between the model gain curve and the random model (diagonal line). It can be understood as an absolute measure of incremental gain. For clarity, we denote the Qini values for the two modeling objectives, i.e. incremental number of purchases for conversion modeling and incremental revenue for revenue modeling, by Q_c and Q_r respectively.

A limitation of Q may be seen in the fact that different parts of gain/Qini curve carry different relevance to marketing practice. Campaigns are typically target to a small fraction of customers. Thus, the gain of a model for smaller targeting fractions is particularly important. Ling and Li (1998) proposed a weighting procedure to account for this issue in response modeling. We adopt their approach for uplift modeling. In formal terms, let Q_{wc} and Q_{wr} be the weighted scores across deciles of a certain model for conversion and revenue modeling, respectively.

Then, $\mathrm{Q_{wc}} = \frac{(0.9*\mathrm{Q_{1,c}}+0.8*\mathrm{Q_{2,c}}+\cdots+0.1*\mathrm{Q_{9,c}})}{\sum_i \mathrm{Q_{i,c}}}$ and $\mathrm{Q_{wr}} = \frac{(0.9*\mathrm{Q_{1,r}}+0.8*\mathrm{Q_{2,r}}+\cdots+0.1*\mathrm{Q_{9,r}})}{\sum_i \mathrm{Q_{i,r}}}$ with c and r indicating conversion and revenue, respectively, and $i = (0, 1, \ldots, 9)$ representing a decile index.

The following chapters present the experiments using the above performance measures. These are (i) Qini curves and Qini values Q_c and Q_r, (ii) their weighted versions Q_{wc} and Q_{wr} and (iii) incremental revenue.

Conversion Modeling

In terms of conversion modeling, we consider LWUM, TCIA and response modeling. Response modeling serves as benchmark that disregards the uplift philosophy. In total, 316 different classification models have been developed per approach, using the learning methods outlined above. Accordingly, a total of 948 classifiers are compared in the experiment. Some models have returned comparably biased probabilities. To address this problem, probability calibration based on Platt Scaling (Platt, 1999) and isotonic regression have been used for certain linear support vector machines and random forests, respectively. Figure 6 depicts model performance in terms of Qini curves per uplift modeling approach. The legend in each plot also provides Qini values.

Figure 6 indicates that most of the uplift models succeed in outperforming the naïve benchmark, which is represented by the diagonal line. However, uplift models built on top of a Naïve Bayes classifier deviate from this pattern and typically perform weaker than the naïve benchmark. In this sense, Figure 6 warrants the conclusion that Naïve Bayes is not a suitable approach for this type of learning problem and should be avoided. Although its performance is better than that of Naïve Bayes, stochastic gradient descent for classification appears to be another candidate learner which proofs inadequate for the focal prediction task. The corresponding Qini curve falls sometimes below the naïve benchmark and never exceeds it with substantial margin. On the other hand, tree-based ensemble classifiers are among the best classifiers and show consistently good results across all uplift modeling approaches. The same applies to the KNN classifier, which is always among the top three methods per approach. This result is surprising in that KNN is a rather simple classifier.

A positive result shown in Figure 6 is that several of the considered uplift models display a steep increase in performance within the first decile. It is common practice in marketing to target only a small subset of the customer base with a campaign. Therefore, the degree to which a model delivers high uplift in the first decile (i.e., succeeds in identifying a small subset of highly responsive customers) is of paramount importance for campaign planning practice.

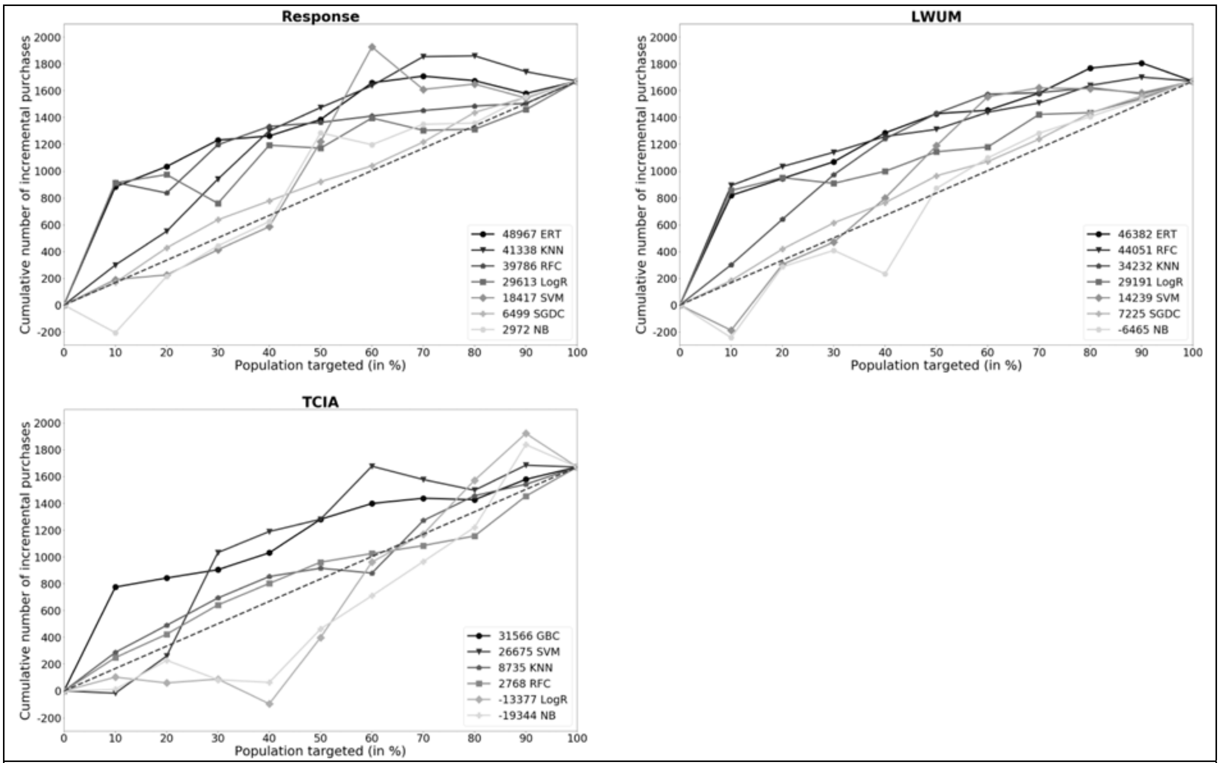

Figure 6. Best Models Per Approaches for Conversion Modeling

To simplify comparisons of alternative uplift modeling approaches to each other, Table 3 reports the per-decile-uplift for each approach and classifier. In addition, the second to last and last columns provide the weighted average for conversion Q_{wc} and the rank of an approach-classifier combination across all candidates in Table 3, respectively. Table 3 reveals that the overall best approach in the comparison is a response model with underlying ERT classifier (Q_{wc} = 5,598). This is a stunning result, suggesting that none of the uplift models outperforms a simple response model. Although the latter ignores the critical point that only persuadable customers are worth targeting, the incremental conversion of the response model exceeds that of the uplift approaches, which are deliberately designed to maximize incremental response. In this sense, the results of Table 3 put the merit of conversion uplift modeling very much into perspective.

The second-best approach in the comparison is LWUM developed on top of a random forest classifier ($Q_{wc} = 5{,}365$), followed by another implementation of this approach using the ERT classifier ($Q_{wc} = 5{,}316$). The other uplift approach, TCIA, performs much worse and proofs inferior to Lai's approach. LWUM was the overall best approach in a recent uplift modeling benchmark (Kane et al., 2014). In this sense, superiority over TCIA, which we observe, is consistent with prior work. However, the performance of the response modeling approach remains the key finding from the conversion modeling experiment. Delivering the largest incremental gain in conversions across all but the ninths decile, which is barely relevant for marketing practice, response modeling can well be considered a dominant approach for the employed data. This sets a hard benchmark for the revenue uplift experiment using the same data, which is presented in the next section.

Table 3. Performance Statistics of all Approaches and Models for Conversion Modeling											
Approach	Model	Dec. 1	Dec. 2	Dec. 3	Dec. 4	Dec. 5	Dec. 6	Dec. 7	Dec. 8	Dec. 9	Q_{wc}	Rank
Response	ERT	881	**1034**	**1231**	1261	1384	1656	1708	1673	1579	**5598**	**1**
	RFC	**916**	836	1197	**1333**	1363	1411	1451	1485	1505	5259	4
	LogR	913	973	759	1193	1171	1396	1304	1312	1459	4790	5
	KNN	299	550	940	1301	**1474**	1637	**1854**	**1860**	1742	4641	8
	SVM	190	225	413	585	1219	**1926**	1609	1649	1546	3339	11
	SGDC	168	427	636	777	921	1039	1216	1437	1551	3088	14
	NB	-206	214	442	623	1285	1197	1349	1358	1555	2622	17
LWUM	RFC	893	1034	1140	1258	1311	1439	1508	1639	1701	5365	2
	ERT	818	944	1070	1285	1428	1454	1582	1769	1807	5316	3
	LogR	855	951	909	998	1143	1180	1422	1436	1543	4676	7
	KNN	301	640	973	1241	1430	1574	1584	1622	1578	4510	9
	SGDC	185	422	606	763	981	1117	1263	1423	1559	3143	13
	SVM	-189	303	470	798	1189	1555	1621	1613	1584	3064	16
	NB	-243	286	408	232	871	1097	1282	1407	1533	2129	18
TCIA	GBC	775	843	905	1031	1281	1399	1438	1427	1579	4698	6
	SVM	-17	261	1033	1190	1281	1677	1578	1498	1685	3883	10
	KNN	288	490	694	854	916	880	1271	1457	1543	3286	12
	RFC	249	424	643	802	960	1025	1084	1157	1453	3087	15
	LogR	103	60	88	-96	399	962	1171	1573	**1922**	1587	19
	NB	11	229	84	64	463	710	965	1221	1838	1523	20

Revenue Modeling

The proposed RDT is deployed as candidate for revenue modeling. To demonstrate its merits, it has been tested against TCIA and the benchmark of response revenue modeling.

Due to the nature of the RDT approach, i.e., the revenue response is a binary target variable after discretization, the models that have been presented for conversion modeling have been considered for predictions with this approach as well. Hence, next to 506 regression learners for response modeling and TCIA each, additional 316 classifiers have been considered on the RDT approach, making a sum of 1,328 models for the revenue experiment.

This section compares the performance of stated revenue models using the described performance measures. As before from the huge model library, only those models are considered that have passed parameter optimization with greatest success, i.e. each base learner's best model. Although the subsequent Qini curves visualize the per-decile performance of the underlying models as in conversion modeling, recall that the Qini value Q_r differs to Q_c in that it identifies the value instead of the number of incremental purchases. Figure 7 illustrates this value as a function of the respective population's fraction for the (i) revenue response transformation (top-left), (ii) response benchmark approach (top-right) and (iii) covariates transformation for revenue modeling (bottom-left). The legends display the model values of Q_r.

The performance of the shown models is summarized in Table 4 which reflects the results of the best models per decile and ranks them according to their weighted average for revenue. As before, for a theoretical fixed budget setting, the best approach and model combination is emphasized per decile.

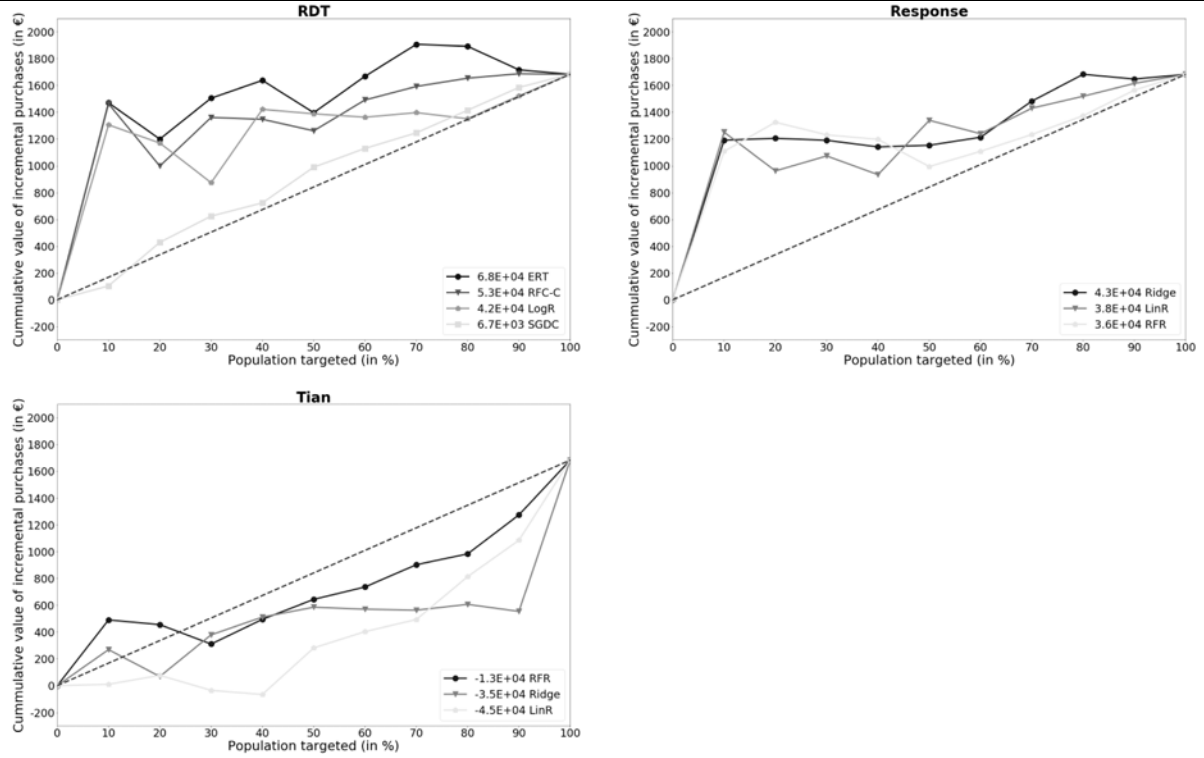

Figure 7. Best Models per Approach for Revenue Modeling

Figure 7 and Table 4 clarify that the RDT approach outperforms response modeling and uplift covariates transformation on almost all deciles. At the whole, the best model on this approach, extremely randomized trees, ranks highest with $Q_{wr} = 6{,}806$€. Remarkably, when further comparing the performance of this model with all other models for each decile separately, it outperforms on eight out of nine deciles in total. Another interesting observation is that in terms of Q_{wr}, all models of RDT rank better than all response models which, in turn, dominate all transformation-related models (with one exception pointing out to SGDC). This further enhances the reliability of our claim that the proposed approach maximizes value not just occasionally in terms of a single model. TCIA not just performs worse compared to the other approaches; for the majority of deciles it not even complies with random targeting.

Table 4. Performance Statistics of all Approaches and Models for Revenue Modeling

Approach	Model	Dec. 1	Dec. 2	Dec. 3	Dec. 4	Dec. 5	Dec. 6	Dec. 7	Dec. 8	Dec. 9	Q_{wr}	Rank
Response	Ridge	1191	1205	1191	1141	1153	1213	1483	1683	1647	5563	4
	RFR	1108	**1325**	1231	1197	995	1108	1234	1374	1565	5379	5
	LinR	1251	962	1074	934	1339	1239	1430	1519	1615	5267	6
RDT	ERT	**1470**	1199	**1505**	**1638**	**1396**	**1668**	**1907**	**1892**	**1717**	**6806**	**1**
	RFC-C	1463	998	1361	1347	1261	1491	1592	1654	1687	6081	2
	LogR	1305	1170	874	1421	1388	1363	1397	1351	1522	5656	3
	SGDC	104	429	624	724	990	1130	1246	1414	1585	3069	7
TCIA	RFR	491	456	311	495	644	737	903	984	1274	2533	8
	Ridge	269	69	380	512	586	571	564	607	555	1738	9
	LinR	14	77	-32	-65	282	403	496	813	1087	735	10

Comparison Conversion vs. Revenue Modeling

While the best models have been empirically examined for the conversion and revenue objective separately, the key question now refers to whether revenue uplift modeling provides more value compared to conversion uplift modeling and response modeling. This comparison is carried out in this section to not just demonstrate the superiority of revenue modeling for this type of marketing application and campaign, but to also prove the effectiveness of the proposed approach based on incremental revenue; a performance indicator being widely used in industry. Figure 8 and Table 5 present performances of the identified best model per conversion/revenue approach from the previous analyses.

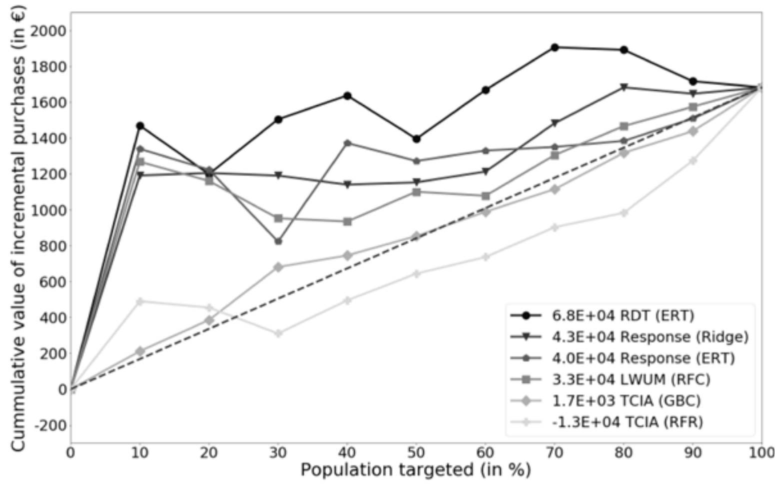

Figure 8. Top Conversion and Revenue Models for Incremental Revenue

From Figure 8 and Table 5 we learn that the extremely randomized trees learner on RDT is overall superior. Across all deciles, this model clearly outperforms (1) response modeling, (2) conversion uplift modeling (i.e., random forest classifier on LWUM and gradient boosting on TCIA) and (3) revenue uplift modeling in the shape of the random forest regressor that predicts with a transformed input space.

It is striking that of the models selected for this analysis, most of the top performers are tree-based and that, among them, ERT seems to be most valuable. Analyzing the best performance for each decile across objectives, the respective ERT models deliver the highest comparable value of incremental purchases on all deciles. The best model per decile is highlighted in bold font in Table 5.

Another argument in favor of the dominance of the RDT approach compared to the others stated stems from literature. Guelman (2014) suggests targeting the top ten percent most likely customers to respond positively to the campaign's treatment, i.e. only the customers from the first decile. Following this advice, RDT enhances incremental revenue comparably greatest with 1,470€. This is about 10% more incremental revenue than the second-best model as stated in Table 5.

Table 5. Incremental Revenue of Best Conversion and Revenue Models

Approach	Objective	Model	Dec. 1	Dec. 2	Dec. 3	Dec. 4	Dec. 5	Dec. 6	Dec. 7	Dec. 8	Dec. 9
Response	Revenue	Ridge	1191	1205	1191	1141	1153	1213	1483	1683	1647
	Conversion	ERT	1340	**1224**	823	1372	1273	1330	1351	1385	1509
LWUM	Conversion	RFC	1268	1161	953	935	1101	1078	1305	1467	1575
RDT	Revenue	ERT	**1470**	1199	**1505**	**1638**	**1396**	**1668**	**1907**	**1892**	**1717**
TCIA	Revenue	RFR	491	456	311	495	644	737	903	984	1274
	Conversion	GBC	212	386	681	746	854	989	1115	1318	1439

The results confirm the contributions made in Bodapati and Gupta (2004) as discretizing revenue to apply classification models is a treasured possession we suggest campaign planners to carry in their toolboxes. According to the results of this paper, this is not just a valid but furthermore an innovative approach for extending the landscape of uplift modeling research and practice.

Conclusion

Empirical results have confirmed the proposed approach to be a valuable tool for revenue uplift modeling. For the data at hand, the parameter-optimized extremely randomized tree algorithm on RDT is most successful in identifying persuadable customers based. In other words, compared to other approaches considered in the comparison, RDT achieves the largest increase in incremental revenue. Although the model's uplift estimate is not unbiased, model building on a discretized (i.e., binary) response implies a smaller bias compared to an unmodified, continuous revenue response (Bodapati and Gupta, 2004).

More generally, the paper has reviewed several uplift modeling approaches and compared their effectiveness against each other and traditional response modeling in a large-scale experiment. Experimental results suggest that uplift modeling does not outperform response modeling in terms of conversion, whereas revenue uplift modeling does add value. Accordingly, the proposed approach complements previous uplift modeling strategies and provides better performance when targeting marketing campaigns the primary goal of which is increasing revenue. Next to the comprehensive empirical study, the paper has developed a formalized uplift modeling process for marketing.

The applicability of the proposed model is not restricted to the online sphere. In fact, the original idea of uplift modeling stems from an offline setting. Many authors have indicated the effectiveness of uplift modeling with physical marketing incentives. These include Guelman et al. (2012, 2015) who sent out information letters and conducted outbound courtesy calls within the insurance industry, Kane et al. (2014) who point to a direct paper mail campaign and Radcliffe (2007) who uses catalogue mails in retail. If the data requirements for uplift modeling are fulfilled (i.e., random assignment of customers to the treatment group and sufficient number of samples), offline retailers such as brick-and-mortar stores can also make use of the approaches and models described here. There also exist situations where online communications take place (e.g., per e-mail), but purchases are undertaken offline (e.g. in brick-and-mortar stores), which build a bridge between online and offline interactions.

As indicated with the UMPM, revenue uplift models should be considered only if the marketing goal is to maximize incremental revenue. By comparing revenue uplift models to conversion uplift models and response models, we empirically confirmed that the former is superior if the campaign goal is revenue maximization. We may thus recommend targeting corresponding campaigns using the modeling approach proposed here. However, if customer acquisition/retention is the primary marketing goal, previous uplift approaches are probably better suited.

In future research, the discretization of the revenue response could be modified in that not a binary variable is induced but a categorical one (i.e., coarsening revenue) converting it into a multi-class classification problem. This would be especially valuable to account for broad or multimodal distributions of customer spending. Furthermore, in terms of the generalizability of RDT, it would be interesting to examine application areas other than e-couponing. A final note for future research is directed to design direct revenue uplift models that are out of the scope of transformation-based modeling architectures.

References

Baumann, A., Lessmann, S., Coussement, K. and De Bock, K. W. 2015. "Maximize What Matters: Predicting Customer Churn with Decision-Centric Ensemble Selection," in *Proceedings of the 23rd European Conference on Information Systems* (ECIS'15), Münster, Germany.

Bodapati. A., and Gupta. S. 2004. "A Direct Approach to Predicting Discretized Response in Target Marketing," *Journal of Marketing Research* (41:1), pp. 73-85.

Bose, I., and Xi, C. 2009. "Quantitative Models for Direct Marketing: A Review from Systems Perspective," *European Journal of Operational Research* (195:1), pp. 1-16.

Chen, Z.-Y., Fan, Z.-P., and Sun, M. 2015. "Behavior-Aware User Response Modeling in Social Media: Learning from Diverse Heterogeneous Data," *European Journal of Operational Research* (241:2), pp. 422-434.

Chickering, D. M., and Heckerman, D. 2000. "A Decision Theoretic Approach to Targeted Advertising," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 82-88.

eMarketer. 2016. "Worldwide Ad Spending: eMarketer's Updated Estimates and Forecast for 2015-2020," (available at http://www.strathcom.com/wp-content/uploads/2016/11/eMarketer_Worldwide_Ad_Spending-eMarketers_Updated_Estimates_and_Forecast_for_20152020.pdf; retrieved May 5, 2017).

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. "The KDD Process for Extracting Useful Knowledge from Volumes of Data," *Communications of the ACM* (39:11), pp. 27-34.

Fernandes, R., and Leblanc, S. G. 2005. "Parametric (Modified Least Squares) and Non-Parametric (Theil–Sen) Linear Regressions for Predicting Biophysical Parameters in the Presence of Measurement Errors," *Remote Sensing of Environment* (95:3), pp. 303-316.

Gotz, M., Weber, C., Blocher, J., Stieltjes, B., Meinzer, H. P., and Maier-Hein. K. 2014. "Extremely Randomized Trees Based Brain Tumor Segmentation." in *Proceedings of BRATS Challenge-MICCAI*.

Guelman, L., Guillén, M., and Pérez-Marín, A. M. 2012. "Random forests for uplift modeling: an insurance customer retention case," in *Modeling and Simulation in Engineering, Economics and Management*, Springer Berlin Heidelberg, pp. 123-133.

Guelman, L. 2014. "Optimal Personalized Treatment Learning Models with Insurance Applications," Doctoral Thesis, Universitat de Barcelona, Barcelona, Spain.

Guelman, L., Guillén, M., and Pérez-Marín. A. M. 2015. "Uplift Random Forests," *Cybernetics and Systems* (46:3-4), pp. 230-248.

Hansen, B. B., and Bowers, J. 2008. "Covariate Balance in Simple. Stratified and Clustered Comparative Studies," *Statistical Science* (23:2), pp. 219-236.

Hansotia, B., and Rukstales, B. 2002. "Incremental Value Modeling," *Journal of Interactive Marketing* (16:3), pp. 35-46.

Hastie, T., Tibshirani, R., and Friedman. J. 2009. *The Elements of Statistical Learning*, Second Edition. New York: Springer.

Holland, P. W. 1986. "Statistics and Causal Inference," *Journal of the American Statistical Association* (81:396), pp. 945-960.

Hua, S. 2016. "What Makes Underwriting and Non-Underwriting Clients of Brokerage Firms Receive Different Recommendations? An Application of Uplift Random Forest Model," *International Journal of Finance and Banking Studies (2147-4486)* (5:3), pp. 42-56.

Imbens, G. W. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics* (86), pp. 4-29.

Jacobs, B. J. D., Donkers, B., and Fok, D. 2016. "Model-Based Purchase Predictions for Large Assortments," *Marketing Science* (35:3), pp. 389-404.

Jaroszewicz, S. 2016. *First Ideas on Revenue Uplift Modeling*, Personal Communication, October 15, 2016.

Jaroszewicz, S., and Zaniewicz, Ł. 2016. "Székely Regularization for Uplift Modeling," in *Challenges in Computational Statistics and Data Mining*, Springer International Publishing, pp. 135-154.

Jaskowski, M., and Jaroszewicz, S. 2012. "Uplift Modeling for Clinical Trial Data," *ICML 2012 Workshop on Clinical Data Analysis*.

Kane, K., Lo, V. S., and 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* (82:4), pp. 218-238.

Khajehzadeh, S., Oppewal, H., and Tojib, D. 2014. "Consumer Responses to Mobile Coupons: The Roles of Shopping Motivation and Regulatory Fit," *Journal of Business Research* (67), pp. 2447-2455.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., and Pohlmann, N. (2013). "Online controlled experiments at large scale," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1168-1176), I. S. Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, J. He, R. L. Grossman & R. Uthurusamy (eds.), Chicago, IL, USA: ACM.

Kohavi, R., Longbotham, R., Sommerfield, D., and Henne, R. M. 2009. "Controlled Experiments on the Web: Survey and Practical Guide," *Data Mining and Knowledge Discovery* (18:1), pp. 140-181.

Lai, L. T. 2006. "Influential Marketing: A New Direct Marketing Strategy Addressing the Existence of Voluntary Buyers," Master of Science Thesis, Simon Fraser University School of Computing Science. Burnaby, BC, Canada.

Larsen, K. 2010. *Net Lift Models*. Slides of a Talk at the A2010 Analytics Conference. September 2-3 2010. Copenhagen. Denmark.

Ling, C. X., and Li, C. 1998. "Data Mining for Direct Marketing: Problems and Solutions," *KDD (Vol. 98)*, pp. 73-79.

Lo, V. S. 2002. "The True Lift Model: A Novel Data Mining Approach to Response Modeling in Database Marketing," *ACM SIGKDD Explorations Newsletter* (4:2), pp. 78-86.

Lo, V. S., and Pachamanova, D. A. 2015. "From Predictive Uplift Modeling to Prescriptive Uplift Analytics: A Practical Approach to Treatment Optimization While Accounting for Estimation Risk," *Journal of Marketing Analytics* (2), pp. 79-95.

Long, J. S. 1994. "Confirmatory Factor Analysis: A Preface to LISREL," in *Factor Analysis and Related Techniques,* M. S. Lewis-Beck (eds.), London: Sage Publication.

Manahan, C. 2005. "A Proportional Hazards Approach to Campaign List Selection," *SAS User Group International (SUGI) 30 Proceedings*.

Media Dynamics, Inc. 2014. *America's Media Usage & Ad Exposure: 1945-2014*, September 21 2014, Nutley, New Jersey.

Nassif, H., Kuusisto, F., Burnside, E. S., and Shavlik, J. W. 2013. "Uplift Modeling with ROC: An SRL Case Study," in *Proceedings of the International Conference on Inductive Logic Programming (ILP'13)*, Rio de Janeiro, Brazil, pp. 40-45.

Nattee, C., Khamsemanan, N., Lawtrakul, L., Toochinda, P., and Hannongbua, S. 2017. "A Novel Prediction Approach for Antimalarial Activities of Trimethoprim, Pyrimethamine, and Cycloguanil Analogues Using Extremely Randomized Trees," *Journal of Molecular Graphics and Modelling* (71), pp. 13-27.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., and Mason, C. H. 2006. "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research* (43:2), pp. 204-211.

Netessine, S., Savin, S., and Xiao, W. 2006. "Revenue Management Through Dynamic Cross Selling in E-Commerce Retailing," *Operations Research* (54:5), pp. 893-913.

Ngai, E. W. T., Xiu, L., and Chau, D. C. K. 2009. "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification," *Expert Systems with Applications* 36(2), pp. 2592-2602.

Park, C. H., and Park, Y.-H. 2016. "Investigating Purchase Conversion by Uncovering Online Visit Patterns," *Marketing Science* (35:6), pp. 894-914.

Platt, J. 1999. "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," *Advances in Large Margin Classifiers* (10:3), pp. 61-74.

Radcliffe, N. J. 2007. "Using Control Groups to Target on Predicted Lift: Building and Assessing Uplift Models," *Direct Marketing Journal, Direct Marketing Association Analytics Council* (1), pp. 14-21.

Radcliffe, N. J., and Surry, P. D. 1999. "Differential Response Analysis: Modeling True Response by Isolating the Effect of a Single Action," *Credit Scoring and Credit Control VI*.

Radcliffe, N. J., and Surry, P. D. 2011. *Real-World Uplift Modelling with Significance-Based Uplift Trees*, White Paper TR-2011-1. Stochastic Solutions.

Reinartz, W. J., and Kumar, V. 2003. "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing* (67:1), pp. 77-99.

Rzepakowski, P., and Jaroszewicz, S. 2012a. "Uplift Modeling in Direct Marketing," *Journal of Telecommunications and Information Technology*, pp. 43-50.

Rzepakowski, P., and Jaroszewicz, S. 2012b. "Decision Trees for Uplift Modeling with Single and Multiple Treatments," *Knowledge and Information Systems* (32:2), pp. 303-327.

Siegel, E. 2011. *Uplift Modeling: Predictive Analytics Can't Optimize Marketing Decisions Without It*, Prediction Impact White Paper sponsored by Pitney Bowes Business Insight.
Siegel, E. 2013. *Predictive analytics: The power to predict who will click, buy, lie, or die*, John Wiley & Sons.
Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., and Ye, X. 2016. "Automated Brain Tumour Detection and Segmentation Using Superpixel-Based Extremely Randomized Trees in FLAIR MRI," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1-21.
Sołtys, M., Jaroszewicz, S., and Rzepakowski, P. 2015. "Ensemble Methods for Uplift Modeling," *Data Mining and Knowledge Discovery* (29:6), pp. 1531-1559.
Tian, L., Alizadeh, A. A., Gentles, A. J., and Tibshirani, R. 2014. "A Simple Method for Estimating Interactions Between a Treatment and a Large Number of Covariates," *Journal of the American Statistical Association* (109:508), pp. 1517-1532.
Verbeke, W., Dejaeger, K., Martens, D., Hur, J., and 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), pp. 211-229.
Yong, F. 2015. "Quantitative Methods for Stratified Medicine," Doctoral Thesis, Harvard University, Graduate School of Arts & Sciences, Cambridge, Massachusetts.
Zaniewicz, Ł., and Jaroszewicz, S. 2013. "Support Vector Machines for Uplift Modeling," *Data Mining Workshops (ICDMW). 2013 IEEE 13th International Conference*, pp. 131-138.