Master Thesis

BA – Digital Business

Title report:

An explorative study towards the feasibility of uplift modeling within a direct marketing setting and a web-based setting

Author	Maurice Munting
Student number	S2288508
Faculty	BMS
Master & specialization track	Master Business Administration, Digital Business University of Twente
Course (code)	Master thesis (201500102)
First supervisor	Dr. Agata Leszkiewicz
Second supervisor	Dr. Efthymios Constantinides
Domain of interest	Digital Marketing
Faculty	BMS/ ETM
Date	16-12-2020

UNIVERSITY OF TWENTE.

Executive summary

Uplift modeling has emerged over the last few years as the advanced predictive machine learning method, which aims to overcome the limitations and shortcomings of previous traditional predictive models. These shortcomings have particularly been identified in the domain of direct marketing. Uplift modeling has challenged the response paradigm of traditional predictive response models that have been used on the basis of historical behavioral customer data to predict the conditional probability that a customer would respond to a direct communicated promotion (campaign). When applied in direct marketing, uplift modeling aims to model the actual incremental gain in responses in relation to a marketing intervention or stimulus, by means of estimating the causal treatment effect for individuals. Whereas response models are based on the traditional classification of base responders and customers (and non-responders) who are likely to respond to the marketing communication, regardless of receiving a direct promotional offer. Previous literature has extensively compared the performance of response models and uplift models and has shown that uplift models have the enhanced ability to capture incremental or true responders. Additionally, previous research has mainly focused on the development and comparison of the predictive performance of uplift algorithms in order to further improve its practical use in the direct marketing setting. This is because uplift modeling is different from traditional classification settings since the data is collected in a randomized trial. As a result, uplift models estimate the uplift as the difference in predicted conversion/response probabilities within the treatment and control groups. Hence, from a direct marketing and business perspective, it optimizes future targeting and conversion of direct communications through the identification of those customers with the highest incremental response and/or return on marketing investments.

Instead of comparing the predictive performance of uplift models, this report has conducted explorative research on the practical application and feasibility of uplift modeling in the broader use than optimizing direct marketing communication such as campaigns. Historically, post and email campaigns have been a frequently used channel of direct communications, whereas nowadays websites are becoming a unique and predominant channel of direct communications and sales. The increased volume of online shopping behavior in combination with the endless browsing and navigational pathways offered by E-business has resulted in continuous interactions between firms and customers and several possibilities and challenges for marketers to capture a stream of behavioral information to establish (personalized) product recommendations or advertisements for conversion optimization purposes. However, how these visitors respond in relation to these types of personalized web-stimuli or interventions and whether these interventions truly achieve incremental conversion has not been discovered yet. Moreover, it is currently unknown whether uplift modeling, in a similar fashion as for campaigns, can be applied in a web-based setting for conversion optimization and what potential influential factors, differences or challenges could determine its feasibility. As a result, this exploratory research report is focused on the central question: How can uplift modeling be used for future targeting optimization and ultimately conversion optimization within a web-based setting using a randomized trial setting?

In order to answer this question, this report has extensively reviewed current uplift literature and has practically and conceptually compared the feasibility of uplift within an email-based campaign setting and a web-based setting. The conceptual comparison of both settings as well as the practical analysis has indicated the business value of designing and modeling uplift experiments, but also its predictive sensitivity in relation to specific business environments and characteristics. Although uplift models in the campaign setting were generally able to group and rank subjects with the highest incremental treatment effect (uplift) in the first decile(s), no down lift or negative uplift was observed when higher factions or deciles (e.g. lower uplift) of the ranked customers were being selected. Moreover, the uplift model(s) were not fully capable to identify customers as persuadable

customers, sure-things or do-not-disturbs or lost-causes, which could be explained through the environment in which the uplift experiment was designed and the type of treatment (campaign) used. The latter three mentioned types of classified customers generally have a higher chance of occurring when targeting large fractions of the customer base, which often results in a decreased profitability (depending on costs of treating and targeting customers). Besides this, the practical and conceptual analysis in this report regarding the feasibility and application of uplift in a web-based setting has also shown its dependency on specific requirements and circumstances. The most influential challenge within the design and data collection process of uplift experiments within a web-based setting consists of the collection of counterfactual data on an individual and anonymous basis. This challenge has shown to be mainly applicable within the web-based setting rather than in the campaign setting since customers in the latter setting often give direct permission to be contacted and to be exposed to a direct promotion (treatment). In a web-based setting on the contrary, visitors can theoretically be exposed to the treatment/web-stimuli without a direct permission requirement, other than the (indirect) permission with regards to the acceptance of cookies and privacy constraints. Additionally, the report has shown how different behavioral data collection methods can be employed to enhance the capability of capturing anonymized individual behavior in terms of clickstream data and demographics, which can subsequently be used as valuable input variables required for uplift models. Most of these approaches however require businesses to have advanced analytical online behavioral data collection and storage tools in place and require investments of resources and time for manually logging customers' online (clickstream) behavior.

An answer to the central question within this research report is given based on the mentioned highlighted certain challenges, solutions and implications in this report regarding the feasibility question of uplift modeling in a dynamic web-based environment. It can be concluded that uplift modeling in both of the examined settings within this report offers a lot of business value, especially in comparison to the application of traditional response models within direct marketing. However, the feasibility of uplift and the extent of its business value is highly dependent on specific circumstances such as the costs/investments and the potential prohibitive profits involved in the conventional and/or initial process of large-scale customer targeting. Additionally, the feasibility of uplift within the web-based setting depends on the capability of businesses to collect behavioral and dynamic customer information and requires extensive preprocessing and data mining activities in order to gain data usable for modeling purposes. These advanced and/or even manual behavioral data collection methods are especially required in web-based scenarios that deal with unregistered or anonymous web-users to achieve large and balanced response levels among treated and non-treated web-visitors. Therefore, this research report has shown that the design and feasibility of uplift to be more effective in a controlled web-environment and circumstances where decision-makers have control over a certain (web-based) treatment whose manipulation is expected to cause a significant behavioral change. Hence, the answer to the central question of this report is that uplift modeling within a web-based setting is mainly feasible and applicable under the circumstance of a controlled web environment, in which for example visitors are required to register or to log in when vising the website. Ultimately, uplift in a digital setting can be viewed as a valuable analytical method that aims to estimate the personal treatment effect of certain web-based interventions. Since this exploratory oriented research report was limited in its ability to demonstrate and evaluate its practicality, future research should further empirically investigate the business value of uplift in a web-based setting. Finally, this advanced machine learning topic in marketing is very promising from a business and research perspective, since it is also closely related to the developments within the landscape of marketing analytics and personalized or digital marketing.

Table of contents

Executive summary ... 2
Introduction ... 5
Aim of research & research question ... 7
Review of uplift modeling literature ... 8
Conventional response models vs uplift modeling in direct marketing ... 8
The design process and objective of uplift modeling ... 11
Uplift algorithms & Strategies ... 12
Digital marketing through web-based conversion optimization ... 15
Evaluating the quality and performance of uplift models ... 17
Research design & methodology ... 20
Conceptual comparison ... 21
Description of data set 1 for practical analysis (campaign setting) ... 26
Description data set 2 for practical analysis (web-based setting) ... 26
Procedure of analysis ... 27
Analysis of data set 1 ... 28
Exploratory data analysis. ... 28
Fitting the models ... 29
Empirical setting 1: Results of data set 1 ... 31
Two-model estimator ... 31
Random Forest ... 31
Response Variable transformation for uplift (RVTU) ... 32
Final model assessment & conclusion ... 33
Model profiling ... 35
Empirical setting 2: Practical analysis of the web-based setting ... 36
Business value of uplift experiments in online (web-based) marketing ... 38
Conclusion ... 39
Practical implications ... 40
Limitations and future research ... 40
References ... 42
Appendix ... 45

Introduction

Direct and digital marketing efforts of an organization play a central role in the domain of customer acquisition, customer retention and increasing sales. Due to the ever-growing digitalization and increase of technologies, organizations can send marketing promotions directly to customers via diverse communication channels such as email, posts and websites. Additionally, Ascarza (2018) states that firms nowadays have an enhanced ability to gather data in order to gain insights about customers, due to new developments in data gathering, analyzing techniques and tools in the domain of digital field experimentation. The development of analytical methods results in marketing decision models that support all stages of a customer lifecycle (Gubela et al., 2019). The primary goal of most direct marketing efforts is to trigger some specific change in customer behavior, which is mainly measured via a type of return on investment calculation (Radcliffe & Surry, 2011). Traditionally, businesses often use marketing campaigns as a form of direct marketing effort or intervention. Traditionally, direct marketing captured historical customer base information that was subsequently used for future designing and targeting of campaigns to customers. These campaigns are used to promote products to potential customers by contacting them via a direct channel of communication such as mail or telephone (Lai et al., 2006). The goal of these direct communications can be described as the extra stimulation of the desired customer behavior or interaction with the organization. Therefore, direct and digital marketing promotions become an important marketing tool aiming to acquire new customers, retain existing customers and/or improve sales.

However, direct marketing costs are one of the biggest contributors to overall marketing costs and therefore require marketing communications to be adequately evaluated and targeted (Kondareddy, Argwal, & Shekhar, 2016). Traditionally, firms use predictive models to identify customers who are most likely to respond to direct marketing communications such as campaigns. However, an implicit assumption is that all purchases or other favorable responses (e.g. visit, conversion) are generated by direct contact (Lai et al., 2006). Moreover, it is unknown and often not considered how a customer would have behaved and responded without the direct communicated promotion. Furthermore, direct marketing literature offers evidence that direct communication such as email campaigns can disconcert certain customers or prospects (Rzepakowski & Jaroszewicz, 2012b). Available methods that identify these types of customers are scarce since most traditional methods tend to solely focus on response prediction (Gubela et al., 2019). As a result, organizations and researchers become increasingly interested in new data mining approaches to develop predictive tools, which aim to enhance the performance and the associated cost and ROI of direct marketing communications (Diemert & Renaudin, 2018). This new stream of predictive modeling is described as uplift modeling which is mainly applicable in the domain of medicines and (direct) marketing. Uplift modeling aims at identifying and selecting the subset of individual customers who are likely to change their behavior in response to a (marketing) action or intervention (Rzepakowski & Jaroszewicz, 2012a). Controversially, it also identifies the subset of customers that should be removed from future direct targeting or that should be targeted in different ways. Consequently, the goal of uplift is equivalent to modeling the differential (causal) effect of a marketing incentive on customer behavior that should result in a favorable response. Uplift modeling can enable firms given limited resources and multiple options for designing direct marketing communications (treatments) to determine which course of action (decision) is best to undertake in terms of achieving the highest incremental response and return on marketing investments. Therefore, uplift modeling in the setting of direct marketing communications tends to optimize treatment assignment of direct communications and the associated time and costs spent, since their goal is to target those only customers who are most likely to be responsive to the campaign.

Besides the objective and value of uplift to optimize the targeting process of direct communications such as email campaigns, websites can be seen as a unique channel of direct communication which poses unique challenges for marketers. Digitalization has modified the online behavior of customers in terms of their endless browsing behavior, continuous price checking through online competitors and the various navigational pathways that are being offered to customers (Hernandez et al., 2017; Kwan et al., 2005). This has resulted in multiple challenges for online marketers to attract and retain customers, establish loyalty and increase conversions. Several studies have suggested that web or UX design can become a significant driver of online conversion, and exposure to irrelevant marketing or web stimuli could be caused by a lack of understanding of users shopping intent, which can result in low conversion rates (Ding et al., 2015; Mokryn et al., 2019). Moreover, marketers and E-commerce businesses have more opportunities to (continuously) interact and capture user's interests and behavior instead of collecting historical transactional data used for more traditional direct channels of communications such as email campaigns. However, as of yet, most websites offer customized information and or product recommendations based on the user's historic information and preferences and provide (mass) customization by means of A/B testing approaches, assuming that the user's preferences are static. Similar to campaign management within traditional direct marketing, the thought process of mass customization relies on the assumption that every conversion is generated by contacting or exposing the web visitor or customer to a certain (web-based) stimulus. Additionally, it assumes that websites are able to recognize their visitors through either registration requirements or advanced behavioral tracking. As of yet, only a few websites use some sort of behavioral targeting in order to establish interactive web-stimuli such as product recommendations, personalized advertisements and chatbots. Furthermore, even fewer websites know how users could change their intentions over time or in a response to a certain web-based stimulus. Uplift modeling is based on the premise that individuals (subjects) can show significant heterogeneity in response to a certain stimulus or treatment (Guelman, 2014). Therefore, uplift modeling has the potential to model the differential (causal) effect of a web-based marketing incentive on customer behavior, by emphasizing the individual dynamic attributes of these visitors. Ultimately, uplift modeling could, similar to the direct marketing setting, enable marketers and E-business to determine which course of action (decision regarding a web-stimuli) is best to undertake in terms of achieving the highest incremental response for purposes of for example conversion optimization.

Several approaches for uplift modeling have been studied in the literature within the domain of medicines and marketing. More specifically, previous literature compared the performance of response models and uplift models and has shown that uplift models are more accurate and impactful compared to standard response models within the domain of direct marketing (Gubela et al., 2019; Rzepakowski & Jaroszewicz, 2012c). However, uplift modeling remains a novel analytical tool since it incorporates aspects of data mining, which are continuously being developed due to improved machine learning algorithms, modeling strategies and processing capabilities of technological tools and software. As a result, existing literature mainly focuses on the establishment and comparison of different algorithms and modeling strategies to further improve the predictive accuracy of uplift models. Additionally, most uplift models are based on artificially created data or simulation data, whereas only a few studies base their models on real-time data (Kondareddy, Argwal, & Shekhar, 2016). Besides this, the application of uplift modeling within current literature is mainly limited to the optimization of direct marketing communications such as campaigns to increase sales, conversions or retentions (Ascarza, 2018; Diemert & Renaudin, 2018; Rzepakowski & Jaroszewicz, 2012b). As of yet, it remains unclear whether uplift modeling is a valuable tool for applications beyond campaign and targeting optimization. For instance, the question remains unclear how uplift modeling can be used as an analytical tool for web optimization and ultimately conversion optimization of websites and E-business.

As a result, this explorative oriented research report will address this gap by examining and comparison of two empirical settings: campaigns and websites. This is achieved by studying one public dataset from a traditional uplift and campaign optimization perspective and by studying one web-based data set which is collected on the website of the University of Twente. Moreover, this report focuses on demonstrating the application and feasibility of uplift modeling within a digital web-based environment and also within a real-world application. The following section of this report covers and summarizes the central question and aim of this research. In the second section of this report, a literature review on uplift modeling will be given. More specifically, this section will describe the theoretical background and perception of uplift modeling in comparison to traditional modeling approaches used in marketing and describe how developments within digital marketing have changed the usage and perception of these models. The third section covers the research design, describing the two empirical settings in which the feasibility and applicability of uplift modeling will be explored by means of conceptual and practical comparison of the two settings. The subsequent section describes the practical analysis of uplift modeling within a campaign and web-based setting through the analysis of two datasets. Finally, conclusions, implications, limitations and opportunities for future research are drawn within the last section of this report regarding the feasibility and business value of uplift modeling from the perspective of a campaign and web-based setting.

Aim of research & research question

This explorative research will explore the usage and feasibility of uplift modeling beyond the traditional direct marketing campaign setting and purposes of campaign optimization. By demonstrating the potential of uplift modeling within a traditional digital marketing setting, this research report aims to also explore the potential feasibility of uplift modeling in a web-based setting that is characterized by more dynamic information such as clickstream data and anonymous web-visits. Therefore, the following central question will be answered within this report: How can uplift modeling be used for future targeting optimization and ultimately conversion optimization within a web-based setting using a randomized trial setting?

Review of uplift modeling literature

In this section, related studies and previous literature will be presented on uplift modeling. Firstly, an introduction and motivation for the application of uplift modeling in relation to conventional predictive modeling techniques will be given from the (direct) marketing perspective. Secondly, an overview will be given on the current state-of-the-art of machine learning algorithms and strategies for uplift modeling. Furthermore, the developments of predictive models in digital (web-based) marketing will be stated and the relation of uplift modeling and the development of behavioral and analytical marketing trials to capture, analyze and personalize offerings to customers will be discussed. Finally, the criteria for evaluating the quality and the predictive performance of uplift models are discussed, to get an increased understanding of uplift models and the interpretation of the modeled results.

Conventional response models vs uplift modeling in direct marketing

Predictive modeling techniques supporting decision-making in digital marketing have been applied over the past years in order to predict customer behavior. More specifically, predictive modeling applications include forecasting usage of social media, predicting sales, predicting customer churn, visits, or conversations based on advertisements or promotional campaigns (Ascarza, 2018; Gubela et al., 2020). Within the field of predictive marketing models, two categories can be distinguished, traditional propensity models and uplift models. Table 1 provides a summary of these models used in previous literature. The goal of these models is to allocate marketing resources efficiently, identify effective channels in multi-channel advertising and most importantly increase marketing efforts (campaign or advertisements) effectivity and profit through well-informed targeting decisions (Gubela et al., 2019). Firstly, propensity models incorporate several types of models frequently used in marketing and sales such as penetration or lookalike models, purchase models and response models. Penetration or lookalike models and purchase models are different from response models since the former models mainly emphasize the predation of customer behavior on the basis of historical customer information. These models are used under the assumption that individuals with similar historical attributes will act similarly to those who already have shown favorable behavior in the past. Whereas response models aim to predict and target customers who have shown a favorable behavior in the past (response) in relation to a marketing promotion or effort (Radcliffe & Surry, 2011). Similarly, Rzepakowski and Jaroszewicz (2012) state that response models, in comparison to the other type of propensity models, are based on (sampled) customers who have been subjected to an organizational marketing effort or promotion in the past and their objective is to estimate the probability of a favorable outcome (response) after the marketing action. Therefore, these models attempt to predict (future) customer behavior in relation to an organizational marketing stimulus in order to optimize future targeting.

Since propensity and response models have traditionally been applied within the domain of direct marketing, it is important to describe this area. Direct marketing is described as the process of acquiring and retaining customers by providing a framework for three activities: analysis of individual customers, strategy formulation and implementation such that customers respond directly to communications (Hasouneh & Alqeed, 2010). Direct marketing is often interchangeably used with database marketing. These terms can be distinguished by their objective in which database marketing emphasizes the analysis of historical customer information in combination with future strategy forming (Hasouneh & Alqeed, 2010). Hence, historical transactional and demographical data and data from previous communications are collected and stored in a database, where each record is associated with a number of individual characteristics. Direct marketing subsequently uses this information for the implementation and targeting of certain marketing programs or interventions, which promote products or services by means of contacting (potential) customers via a direct channel of communication such as email, posts and telephone (Lai et al., 2006). Hansotia & Rukstales (2002) stated that direct marketing communications are more effective when they are based on the customer's historical purchasing behavior. Once historical information of a previous campaign (including response variable and customer characteristics) have been captured, predictive models are established on the basis of supervised machine learning, which aims to estimate the conditional probability that a customer would respond to a communicated promotion (Lai et al., 2006). Hence, these models are based on supervised classification algorithms, which estimate the functional relationship between a binary class label (e.g. response or no response) and a set of variables that characterize customers (Gubela et al., 2019). Afterward, the model is used to rank the whole customer base in the campaign by their estimated probability of responding.

However, the usage of these types of predictive models has been highly criticized when the goal is to incorporate the actual change in customer behavior in relation to a marketing intervention. While the traditional response models provide some insights that a customer has been influenced or was at least aware of the marketing promotion, it does not guarantee to discover incremental favorable customer behavior (Radcliffe & Surry, 2011). Lai et al. (2006) refer to this as the response-paradigm, because response models predominately target voluntary buyers since these types of customers are most likely to achieve the highest response rate on direct marketing communications. Whereas the decision of voluntary buyers to respond to a marketing campaign could be influenced by other factors such as word-of-mouth instead of the actual effect of the direct marketing communication itself. Therefore, customers are mainly being targeted who have a high chance of responding positively to marketing communications, without explicitly considering the likelihood that these customers would have reacted in the absence of the promotion. Similarly, Konareddy et al. (2016) has shown that traditional models do not account for the likelihood of response irrespective of the marketing action. This results in a lack of causality which is crucial for measuring the true impact of direct marketing actions or interventions (Gubela et al., 2020). Moreover, it's unclear whether the actual marketing effort was beneficial for a given individual based on the individual's characteristics (Sołtys & Jaroszewicz, 2015). As a result, traditional predictive response modeling techniques are useful for predicting response or behavior but are limited in terms of sophisticated customer targeting optimization in the areas of conversion or retention.

A solution to the application of predictive modeling within the discipline of marketing is uplift modeling. Similar to the traditional predictive models, uplift considers a direct-marketing setting. However instead of estimating the probability that customers will respond (purchase or visit), uplift modeling attempts to model the difference between conditional class probabilities in a treated group of customers and a control group (Sołtys & Jaroszewicz, 2015). Uplift models estimate the differential change (impact) in the response behavior of customers in relation to a certain marketing action or intervention (Gubela et al., 2020). Therefore, uplift models focus on customizing treatment assignments to individuals by considering the causal link between a marketing intervention and customer response. Furthermore, uplift modeling emphasizes the estimation of heterogeneous treatment effects. An understanding of heterogeneous treatment effects is important since uplift modeling considers the treatment effectiveness to vary with characteristics of individuals and subsequently aims to discover the optimal treatment assignment rule (Olaya et al., 2020). As a result, uplift modeling attempts to correctly predict the optimal treatment for a given individual subject in relation to the subject's characteristics. To summarize, conventional response models focus on estimating the probability that customers will purchase or respond if customers/prospects are subjected to a specific treatment. Whereas uplift models attempt to identify incremental responses through estimating the increase in response (conversion) probability if we treat them over the corresponding probability if we do not treat them (Radcliffe & Surry, 2011). This implies that uplift can be used for targeting customers whose response rate increases due to the marketing intervention, while response models predominantly indicate responsive customers who may or may not be affected by the intervention. As a result, uplift could enable firms and marketers to identify and to select the subset of individuals who are likely to change their behavior in response to a (marketing) action or intervention.

	Type of model used	Objective	Randomized trial setting	Contribution
(Mokryn et al., 2019)	Propensity/ response model incorporating conventional supervised machine learning	Prediction of purchase intention	No	Visitors in session information and the recent trendiness of products clicked in that session have modeled using traditional classification to be able to estimate the purchase intent of occasional visitors.
(Van Den Poel & Buckinx, 2005)	Propensity/ response model	Predicting online purchase behavior through clickstream information	No	General and detailed clickstream behavior enhanced the predictive performance of purchase behavior
(Lai et al., 2006)	Usage of traditional response model vs uplift models	Identifying and targeting only those customers who purchase decision can be positively influenced, i.e. buyers who are non-voluntary	Yes	Uplift modeling is a practical solution for marketers to achieve influential & incremental marketing, instead of responsive marketing where certain individuals will make a purchase in absence of a direct communication or contact.
(Rzepakowski & Jaroszewicz, 2012)	Traditional response models vs uplift models	Predicting the next web- visit of a customer in relation to a marketing campaign	Yes	Uplift models have shown to be more effective when predicting new visits or purchases since response models do not distinguish between spontaneous visitors/buyers and new visitors/buyers.
(Radcliffe & Surry, 2011)	Response models vs uplift models	Statistical analysis and comparison of the modeling perception of different modeling approaches	Yes	Response models cannot be fundamentally used for sophisticated incremental response prediction and the term response (deliberately) loaded and incorporates the unmistakable connotation of causality
(**Gubela et al., 2019)**	Uplift modeling	Demonstration of the stat-of-art of different uplift algorithms and modeling strategies	Yes	Several uplift modeling techniques differ in their modeling approach and predictive accuracy, uplift models are able to pattern the causal effect of a marketing incentive on customer behavior

Table 1: Summary of previous studies on predictive response modeling (& uplift modeling)

The design process and objective of uplift modeling

As described in the earlier section, the difference between response models and uplift models is associated with a different modeling approach. The research design within uplift modeling is important in order to incorporate the causal effect of the treatment for estimating the actual incremental gain in relation to a marketing intervention (Sołtys & Jaroszewicz, 2015). Additionally, response models within a traditional classification setting aim at achieving a high predictive accuracy on a given data set and focus solely on class probabilities. Whereas uplift focuses on the change in class probabilities caused by a specific action or treatment (Rzepakowski & Jaroszewicz, 2012a). In order to establish an uplift model and to interpret uplift predictions as causal, data should be collected in a counterfactual way, for example in a randomized control trial (A/B test) (Diemert & Renaudin, 2018). Three elements are required to be present in the data: a set of variables representing (pretreatment) characteristics of individuals, a decision variable indicating the assignment to either treatment or control group and the corresponding outcome (response). For instance, a random sample of customers/prospects should be split into two subsamples, a treatment sample (T) and a control sample C. A treatment in terms of a marketing intervention will be given to members of T. In the case of a binary classification problem, we denote that the dependent variable Y = {0, 1} and assume that 1 is the favorable outcome such as a purchase, page visit, or conversion. An uplift model fits the following equation:

$$Uplift = P(Y = 1 \mid x; T) - P(Y = 1 \mid x; C)$$

Where the model considers the difference in probability between two potential outcomes P(Y = 1 | x; T) and P(Y = 1) | x; C), if the subject characterized by a vector of variable x is treated or not, respectively (Diemert & Renaudin, 2018; Radcliffe & Surry, 2011). Therefore, the conditional treatment effect (uplift) is modeled by subtraction of the probability of two outcome states in relation to being assigned to the treatment or control group.

Besides the different required modeling approaches, uplift models can be distinguished from response models regarding the prediction objective. Uplift models can be used to classify different types of customers or prospects. For instance, Shaar et al. (2016) and Lai et al. (2006) have stated that uplift can be both positive and desirable or negative and undesirable (Table 2).

	Respond	**No respond**
Treatment	Persuables (positive)	Lost causes (negative)
Control	Do-not-disturbs or sure-things (negative)	Persuables (positive)

Table 2: Contingency table of classifying and targeting customer through uplift modeling

According to table 2, customers (subjects) who did respond favorably in relation to a marketing intervention/treatment or did not respond due to not being subjected to the marketing intervention, can be classified as positive uplift. Gubela et al. (2019) define these customers as persuadable who buy if being treated and refrain from buying otherwise. Controversy, customers can be classified as do-not-disturbs or sure-things if they respond while not being subjected to a marketing intervention and even as lost causes if they do not respond whilst being subjected to a treatment (Olaya et al., 2020b). According to Shaar et al. (2016), this is described as a negative uplift or down lift, because it can be argued that those customers have already decided not to perform the desired action under any circumstance. Direct marketing efforts such as promotional email campaigns which predominantly aim at achieving the highest response (response modeling) might tend to put off subsets of customers and might include subsets of customers that already decided to respond regardless of the intervention (Rzepakowski & Jaroszewicz, 2012b). The latter mentioned subset of customers are often referred to as voluntary buyers, sure-things or as unnecessary marketing costs, who base their purchase decision on other factors than the direct marketing campaign itself (Lai et al., 2006). For instance, other motives such as word-of-mouth could have resulted in the intention to respond or buy. As a result, in order to maximize the effectiveness of direct marketing efforts (campaigns), while making efficient use of marketing resources, marketers should distinguish and target the customers based on a broader classification than solely responders and non-responders. This is due to the possibility of targeting individuals who would have responded regardless of the marketing intervention and individuals who did not respond, resulting in potential unnecessary costs. As a consequence, campaigns or other marketing interventions can be considered successful if it succeeds in enhancing the response rate of the treatment group compared with the response rate of the control group.

Uplift algorithms & Strategies

The classical machine learning algorithms, as described in the appendix of this report, are generally not well suited regarding their direct applicability in randomized trials. Conventional supervised machine learning cannot be (directly) used in these scenarios in order to model the incremental impact of a (marketing) treatment. A traditional machine learning algorithm predicts the result after the action and does not incorporate its causal impact (Rudas, & Jaroszewicz, 2018). Moreover, machine learning algorithms for the purpose of uplift modeling can predominantly be distinguished from traditional algorithms through the problem of causal inference (Sołtys & Jaroszewicz, 2015). This problem is resembled by the fact that for every subject only one outcome is observed under a certain treatment condition. Therefore, the response or value of the response under the treatment alternative (counterfactual response) is unobserved (Guelman, 2014). Consequently, it is impractical to predict whether the action was beneficial for a given individual. Additionally, this means that the predictions of uplift models in comparison to the predictions of traditional machine learning models cannot be assessed at the level of individuals. This is different from classification, where the true class of an individual subject is known within the training data set, which can thereafter be compared with the predicted value (Sołtys & Jaroszewicz, 2015). As a result, uplift algorithms can be seen as adapted traditional algorithms, where data is modeled on a treated training and control data set in order to predict a label or value.

Previous work on uplift modeling algorithms can be distinguished into two streams. The first and common approach is to build an uplift model based on two separate classifiers. This is also known as the indirect estimation method of modeling uplift (Guelman, 2014). The most used method within this approach is described as the double classifier or two-model method which uses two separate probabilistic models, one which fits on the treatment group and predicts the probability Pt (Y = 1| X), while the second fits the control group and predicts Pc (Y = 1| X) (Diemert & Renaudin, 2018). Afterward, the conditional treatment effect (uplift) can be modeled by the subtraction of both probability models. The advantages of this approach emerge from its simplicity and the possibility to use any method of classification (Ja´skowski & Jaroszewicz, 2012). However, previous research has shown some disadvantages using the double classifier approach. Radcliffe and Surry (2011), have stated the risk that both models can heavily focus on predicting the class probabilities themselves, instead of attempting to predict the actual difference between class probabilities in the treatment and control group. This is due to the fact that both classification models are built independently and separately on the treatment and control group. This could cause an increased risk of overemphasizing or underemphasizing predictor variables that are directly related to the uplift (Devriendt et al., 2018). Soltys and Jaroszewicz (2015) argue that this problem can be overcome when the training data is large enough to estimate the conditional class probabilities in the treatment and control group or under the circumstance that incremental gain is correlated with the class variable. Additionally, ensemble methods could be seen as an effective method for improving the accuracy of the double classifier.

The second approach to uplift modeling is based on a single classifier that directly models the difference in conditional probabilities between the treatment and control group. Guelman (2014) describes these models as direct estimation methods for modeling uplift. Previous research has shown a preference for this approach, where uplift models are based on modified versions of machine learning methods, such as classification and regression trees to model the uplift (Radcliffe & Surry, 2011; Rzepakowski & Jaroszewicz, 2012b). For instance, Radcliff and Surry (2011) introduced a decision tree algorithm used for classifying buyers into non-buyers based on a direct marketing campaign. While traditional decision tree algorithms consider the class or labeling variable as the splitting attribute, the modified decision tree for uplift focuses on the treatment and control group in relation to the class variable (Gubela et al., 2019). Therefore, a modified algorithm was used as a splitting criterion for growing the tree, in a statistical way of maximizing the difference between treatment and control success (response) probabilities. Additionally, Rzepakowski and Jaroszewicz (2010) have introduced an uplift decision tree, which is associated with modern tree induction algorithms focusing on theoretical measures. The splitting criteria consist of divergence statistics in order to model the amount of information that a test gives about the difference between treatment and control class probabilities (Radcliffe & Surry, 2011; Rzepakowski & Jaroszewicz, 2012b). According to Guelman (2014), this splitting criteria aims to maximize the distance in the class distributions of the response variable between treatment and control groups. Moreover, tree-based approaches for modeling uplift are popular due to their flexibility as non-parametric modeling tools. These models do not make any assumptions about the functional form of the data. This results in flexible models like decision trees that can be used for tuning several parameters to improve predictive performance.

Although tree-based approaches for modeling uplift are extremely flexible, a key disadvantage has been captured within the literature of uplift modeling. This problem consists of building a single tree uplift model. Decision trees suffer from high variance, due to the hierarchical nature of the spitting process (Radcliffe & Surry, 2011; Sołtys & Jaroszewicz, 2015). A small change in the data or the effect of an error in the top split advances down to all the splits below within the tree. Moreover, single trees can be unstable and poor in performance due to the amount of variance of predictor variables, which can lead to different trees when splitting the data into training parts. Guelman (2014) states that the instability of trees is even higher for uplift cases, due to the nature of the dataset being resembled by treatment heterogeneity effect. The variance of decision trees can be reduced by methods such as bagging and random forest in order to create trees with higher stability. Both methods are based on a similar idea of creating multiple copies of the original training data set using bootstrapped samples (James et al., 2013). In bagging, a separate tree is fitted on each copy followed by combining all and averaging all trees to a single model. Whereas when building these trees with random forest, each time a spit in the tree building process is considered, a random sample of m predictors is chosen (James et al., 2013). This prevents the chance of biases and errors when there for example is a very dominant predictor variable that would always be on the top split. Therefore, random forest is useful in reducing the variance and instability of building a tree by creating a sequence of de-correlated trees resulting in an averaged tree that is more reliable than a single tree (Guelman, 2014).

Besides the stream of literature on traditional algorithm modification, a stream that focuses on uplift modeling strategies can be distinguished. According to Gubela et al. (2019), this can be viewed as the strategic process of embedding conventional machine learning algorithms into an overall modeling framework or design that facilitates uplift prediction. These strategies focus on data transformation and overcoming challenges faced by the application of conventional machine learning algorithms for uplift purposes. Lo (2002) has introduced a modeling strategy based on the transformation of explanatory variables in the dataset to facilitate direct uplift modeling. This approach is based on a dummy variable representing the treatment and control group, which is multiplied with the entire input dataset (Gubela et al., 2020). This results in an interaction term between each predictor and treatment, which is then used in a standard logistic regression setting. Similar to this, Jaskowski and Jaroszewicz (2012) introduced a class variable transformation method aiming to convert a conventional classification model such as logistic regression into an uplift model. This method was based on the modification of the output space (response variable). Within this approach, a new dependent variable was created incorporating the original binary response variable combined with a variable referring to whether the individual was treated or not (Ja´skowski & Jaroszewicz, 2012). Both previously mentioned modeling strategies are different from typical linear or logistic regression models. More specifically, they incorporate a direct modeling approach instead of building a model on two groups separately, followed by the subtraction of coefficients of both groups to produce a single model (Shaar et al., 2016). As a result, a single classifier is built which directly models the difference in success probabilities in the treatment and control group and thus resembles the direct approach of modeling uplift.

To summarize, various machine learning algorithms have been used to build uplift models in a randomized trial setting, which is different from traditional classification settings (Table 3). Uplift models are commonly established through modified supervised machine learning methods such as logistic regression, decision tree, random forest, and support vector machines. Moreover, uplift modeling techniques can be distinguished into indirect or direct estimation models (Diemert & Renaudin, 2018; Gubela et al., 2019; Guelman, 2014). A key requirement for direct uplift models consists of using a modified version of a conventional supervised machine learning algorithm or to apply a transformation strategy on the dataset. The latter mentioned transformation strategies embed conventional machine learning algorithms into a framework to facilitate uplift prediction in a randomized trial. Whereas indirect uplift estimation models are able to apply traditional machine learning methods. Indirect estimation molding methods create uplift models in an intuitive two-stage procedure to predict. However, previous studies have preferred the predictive performance and accuracy of direct models over indirect models (Gubela et al., 2019; Radcliffe & Surry, 2011; Sołtys & Jaroszewicz, 2015). Therefore, the selection of one of the two uplift modeling approaches is based on the trade-off between simplicity and accuracy. With regards to previous work on uplift modeling, it can be stated that most studies have focused on the application of specific modified algorithms or modeling strategies to facilitate uplift. More specifically, most studies focus on the development and/or comparison of the performance of different modified algorithms and transformation strategies to facilitate uplift modeling. Whereas only few studies have demonstrated the application of these techniques on real-time data and the feasibility of uplift within specific contexts such as a web-based environment. As a result, this study will demonstrate the application and feasibility of three earlier mentioned uplift techniques in a direct marketing setting and subsequently explore their feasibility in broader use.

Uplift modeling streams	**Stream of uplift modeling**	**Advantage/Disadvantage of the approach**	**Main conclusion**
Direct estimation of modeling uplift	Radcliffe and Surry (2011) ; (Rzepakowski and Jaroszewicz (2012b)	Modified (single) tree-based algorithms, which apply theoretical/informational measures as splitting criteria to directly estimate the difference in conditional probabilities between	Modified tree-based algorithms are used, following a splitting criterion for growing the tree, in a statistical way of maximizing the difference between treatment and control success (response) probabilities.

		the treatment and control group (uplift).	
	Guelman (2014)	Modeling multiple trees (random forest) to reduce the variance of single tree growing methods in order to achieve enhanced stability	Random forest is useful in reducing the variance and instability of building a tree by creating a sequence of de-correlated trees resulting in an averaged tree that is generally more reliable than a single tree
Indirect estimation of modeling uplift	Rzepakowski and Jaroszewicz (2012)	Simplistic and intuitive approach and the possibility to use any method of classification	Conditional treatment effect (uplift) can be modeled in an intuitive manner by subtraction of both probability models
	Radcliffe & Surry (2011)	Risk of over-emphasizing the prediction of the class probability itself, instead of estimating the actual difference (incremental) change between class probabilities of treatment and control group,	This disadvantage is caused due to the fact that two independent classification models are separately built. Risk can be overcome by using the direct estimation method or if the data set is large enough
Advanced transformation strategy for traditional supervised machine learning methods	Lo (2002)	Traditional classification methods such as logistic regression can be applied for purpose of uplift modeling through the transformation of explanatory variables in the dataset to facilitate direct uplift modeling	Uplift is modeled using conventional classification methods and in a direct manner.
	Jaskowski and Jaroszewicz (2012)	Enables the usage of traditional classification methods through the creation of a new dependent incorporating the original binary response variable combined with a variable referring to whether the individual was treated or not	This method is referred to as the response variable transformation method for uplift (RVTU) and can be applied in a similar way as conventional classification methods.

Table 3: Summary of state-of-the-art uplift modeling streams

Digital marketing through web-based conversion optimization

In the current digitized market, the goal of attracting sufficient online traffic is vital to the success of online businesses. Online marketing aims to produce conversion and ultimately purchase or subscription, which can only be achieved and optimized through understanding customer behavior and needs. According to Kwan, Fong and Wang (2005), changing online behavior and changing patterns of surfers' access to e-commerce sites poses challenges and opportunities for internet marketing. Low conversion rates in online shopping and conversion optimization are widely recognized challenges for e-commerce sites and digital business (Mokryn, Bogina, & Kuflik, 2019). E-businesses must be devised to provide customers preferred customizations and preferred traversal patterns leading from product awareness and exploration to commitment of a purchase or conversion (Kwan et al., 2005). Moreover, marketers and E-commerce businesses have more opportunities to (continuously) interact and continuously capture user's interests and behavior instead of rationally collecting historical transactional data. This has resulted in behavioral targeting, which aims to customize messages to individual customers based on their specific and dynamic shopping interests and demographic characteristics (Dwyer, 2017). Additionally, Hernandez et al. (2017) refer to behavioral targeting as the generic name of online technologies that collect, organize clickstream data and require the usage of machine learning algorithms to uncover browsing patterns in order to match/target online marketing interventions such as ads to individual customers. Behavioral targeting embeds a tag or identifier within a consumer's browser by means of a cookie that is used to track (individual) browsing behavior (Dwyer, 2017). Predictive customer-oriented models can then be built in order to personalize and recommend products, services, advertisements and other marketing communications.

Recent models in the literature consider dynamic behavioral patterns and machine learning techniques on web-based data such as clickstream data to predict individual intent and behavior (Hee et al., 2016; Kwan et al., 2005; Mokryn et al., 2019). Van den Poel and Buckinx (2005) have shown that general clickstream behavior at the level of visit, detailed clickstream information, customer demographics and historical purchase behavior are key features of predicting visitor's conversion commitment. Especially clickstream behavior and dynamic attributes of visitors such as frequency of visit, organic visit, time from the last visit, in-session dynamics like dwell time and time spent viewing a page are strong indicators of conversion intention (Mokryn et al., 2019; Van Den Poel & Buckinx, 2005). Visitors can then be characterized by these dynamics and additional historical attributes (if available). Conversion prediction can be made in these settings for either current or future visits of customers or even for future targeting purposes. For instance, Bhatnagar et al. (2016) has shown that the duration of the (first) visits generally is a strong predictor for the possibility of a next visit. However, predicting the intention of web-visitors and the probability of conversion becomes more challenging in situations of first-time or anonymous visits, where no (historical) information over the visitor exists. This is still an area for researchers and E-business that is not yet fully leveraged in terms of identifying behavioral patterns and the potential of machine learning techniques to aid in predicting conversion optimization.

The rise of analyzing behavior patterns and in-session dynamics for e-business and e-marketing is mainly caused due to the potential value of machine learning. This one of the advanced analytical techniques businesses can apply in order to transform customer data into valuable insights to establish effective marketing decision-making. Davenport & Ronanki (2018) refer to Machine learning as the algorithmic ability of computers to use data and create models that learn on a part of the collected data and uses the created algorithmic model to make predictions based on new data. Several researchers capture the function of machine learning techniques in marketing as the modeling opportunity to aid personalized engagement in online and offline marketing (Kietzmann & Treen, 2018; Kumar et al., 2019). It can be seen as the facilitating technological tool to enhance the customer's relation and conversion through personalized web-based offerings. Uplift modeling is based around this idea, in which modified machine learning algorithms are used to optimize personalized treatment and to potentially personalize the experience of users (Olaya et al., 2020b). For instance, Guealman (2014) refers to uplifting modeling as the modeling technique to solve the personalized treatment problem aiming to personalize the choice of treatment that maximizes the probability of a desirable outcome for an individual. As a result, uplift modeling focuses on predicting the personal treatment effects and thus aims to customize treatment assignments to individuals, by considering the causal link between a marketing intervention and customer response. This description of uplift modeling is associated with the task of causal inference, which stems from the causal literature (Rubin, 1974). Furthermore, Olaya et al. (2020) has argued that uplift modeling is mainly applicable and valuable in circumstances where decision-makers have control over one or multiple actions or variables, whose manipulation is expected to cause a change in the customer's behavior. Therefore, Gubela et al. (2019) recommended uplift for purposes of maximizing intentions of customers to buy a certain particular product by focusing on the shop-based journey of the customer during a certain time span (i.e., from entering to leaving the shop). Moreover, customers browsing behavior when buying products is seen as a valuable opportunity for modeling incremental sales, since products examined but not purchased could be recommended or offered the next day

(Hansotia & Rukstales, 2002). Therefore, the application of uplift modeling seems extremely valuable for experimental trials and semi-controllable business situations, where decision-maker dispose of several actions to take in order to improve sales, conversion or retention.

The usage of randomized trials for personal treatment selection or uplift modeling has been emphasized in order to ultimately achieve optimized business and marketing related decision-making. Several studies and similar to the approach in this study, link uplift modeling with principles of A/B testing (Ascarza, 2018; Devriendt et al., 2018). A/B testing is often used as a web-development and UX tool where changes and variations in lay-out or design are tested on two or more groups of visitors in order to test and discover the optimal design (Devriendt et al., 2018). Both approaches are based on the idea of assigning one or multiple treatments to groups of subjects (customers) for the purpose of optimizing marketing-decision-making. The observed behavior between the treated and control group allows uplift modeling to predict the incremental impact of the treatment on the individual customer level and on the basis of the individual customer characteristics (Devriendt et al., 2018). Whereas, A/B testing emphasizes decision making at the higher superficial customer base level, with the main goal of assessing the performance of the treatment itself. Consequently, A/B testing for Web or UX-development is based on optimizing and selecting the web-based treatment which achieves the highest overall response or conversion for the whole population, without considering individual characteristics and responses of subjects in relation to the treatment. For instance, Ascarza (2018) has shown that it is better to target customers with proactive churn programs based on their (individual) sensitivity to the program instead of identifying and targeting the customers with the highest risk of churning. This highlights the opportunity for businesses and especially marketers to look beyond estimating the performance of the interventions themselves. Instead of this, emphasis should be put on leveraging the potential heterogeneity in the response of individual customers in relation to the assignment or exposure of a treatment, which can subsequently be used for future targeting personalization and effective marketing decision-making. Ultimately, it can be concluded that the difference between uplift modeling in comparison with traditional A/B and multivariate testing lies in customizing treatment assignments for individuals.

Evaluating the quality and performance of uplift models

Uplift modeling is a machine learning technique that is slightly different from traditional predictive modeling. The difference is connected with the performance evaluation metrics of traditional predictive models and uplift models. Model assessment of predictive models such as regression and classification (trees) rely on the estimation of the prediction error of the model (Guelman, 2014). This is accomplished by comparing the predicted response value obtained by the model for a given individual subject is close to the actual response value for that subject or observation (Gubela et al., 2019). However in uplift models, the actual outcome of a subject that is estimated is unobserved, due to the un-observability of a subject being in both treatment and non-treatment (control) states. Therefore, the error of the model regarding the difference in predicted and actual outcomes cannot be observed and assessed at the individual level (Devriendt et al., 2020). Similarly, Guelman (2014) defines this as the problem of assessing the causal effect of the treatment for a single subject, which is the difference between an observed outcome and its counterfactual. This difference cannot be calculated due to the problem of causal inference stated by Holland and Rubin (1989), which is the limited observability of only one outcome or response under possible treatment alternatives. Instead of assessing the error of the model on an individual level, uplift models rely on comparing similar groups of observation on the difference in outcome for different treatment states (Devriendt et al., 2018). Consequently, the best approximation to predict and evaluate the individual treatment effect (ITE), consists of evaluating the subpopulation treatment effect (Guelman, 2014). Therefore, a decile-based evaluation approach is used based on equivalent segments of the population that are expected by the model to experience the same as a similar net effect (uplift) of the treatment.

One of the most used ways of visualizing uplift and assessing model quality, consist of using uplift charts. This is a visualization of the uplift score of each individual of the test set obtained by the model. Since it is not possible to develop a quality measure based on comparing the actual observed outcome and the counterfactual for individual subjects, the scores are ordered from high to low and binned together in deciles. Each decile includes individuals who are either treated or not treated. The incremental response (uplift) of a decile is calculated by subtracting the response rate of the control group from the response rate of the treatment group (Devriendt et al., 2018). A successful model is able to rank the responders of the treatment group (persuadables) high in the first deciles, whereas responders of the control group (do-not-disturbs) are being ranked as low in the last deciles. Consequently, from a theoretical and ideal perspective, an uplift chart should look like the example in the left-hand side of Figure 1. However, several practitioners of uplift modeling have shown uplift models are generally not stable regarding the performance of the model in different or special circumstances (Kane et al., 2014; Radcliffe & Surry, 2011). Moreover, uplift modeling is situational based and slight changes in circumstances or control variables will change the results. As a result, the uplift chart shown on the right side of Figure 1 is a more realistic chart and gives a clear overview of the difference in the ability to rank (favorable) response in relation to the treatment across similar groups of individual customers.

Although visualization of an uplift model is beneficial for interpreting the results, it is not the optimal way of assessing and comparing the quality and performance of several uplift models. The most used quality measure for comparing uplift models consist of the Qini (gains) curve and coefficient. The Qini curve represents a two-dimensional presentation of model performance and is a generalization of Gini coefficient and is related to the Receiver Operating Characteristic (ROC) curve that is often used for assessing conventional machine learning models (Guelman, 2014; Radcliffe & Surry, 2011). According to Devriendt et al. (2018), the Qini curve functions similarly to a gains curve and is plotted in Figure 2. The blue curved line is based on the cumulative difference in response rate between treatment and control test data set as a function of the selected segments of individuals/ customers as ranked by the performance uplift model from high to low uplift. Whereas the plotted grey line is resembling the incremental gains that are being achieved by random targeting. The Qini metric is the actual Q ratio of the uplift curve and is thus defined as the area between the Qini curve and a random targeting line (R. Gubela et al., 2019). This relative number can be interpreted as the additional proportion of favorable responders or persuadables out of the total population (Devriendt et al., 2020). Although the curve is often increasing, a decrease of the curve shows that additional targeting fails to capture customers classified as persuadables. Moreover, it will even bend below the random targeting curve if it includes individuals that are characterized as do-not-disturb. This will result in a negative uplift or down lift. To conclude, the Qini metric and curve indicate the performance of the model and the incremental gains or losses (uplift) that can be achieved above random targeting of the whole customer base.

$$Qini = Gini_{Treatment} - Gini_{Control}$$

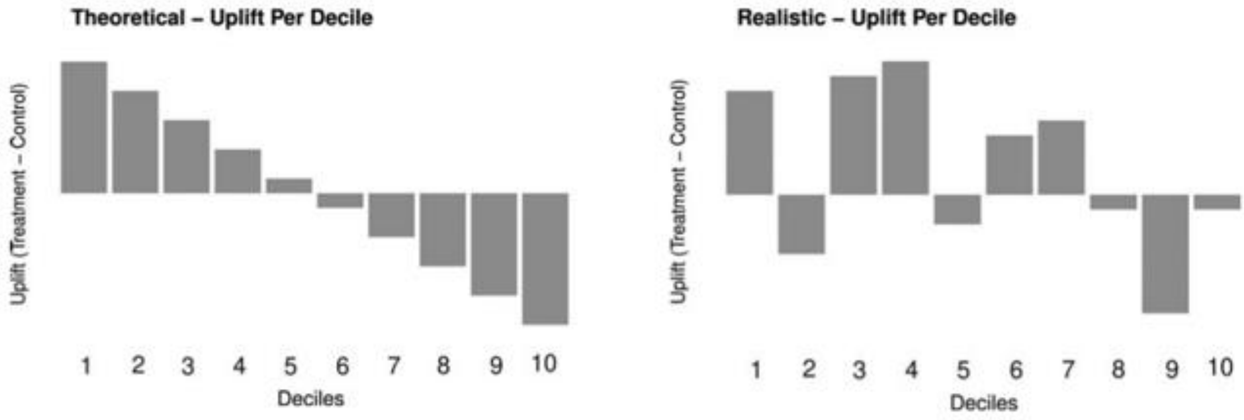


Figure 1: Visualization of theoretical quality vs realistic quality Uplift. Retrieved from: Devriendt, F., Moldovan, D., & Verbeke, W. (2018). A Literature Survey and Experimental Evaluation of the State-of-the-Art in Uplift Modeling: A Stepping Stone Toward the Development of Prescriptive Analytics. In Big Data (Vol. 6, Issue 1, pp. 13–41). Mary Ann Liebert Inc.

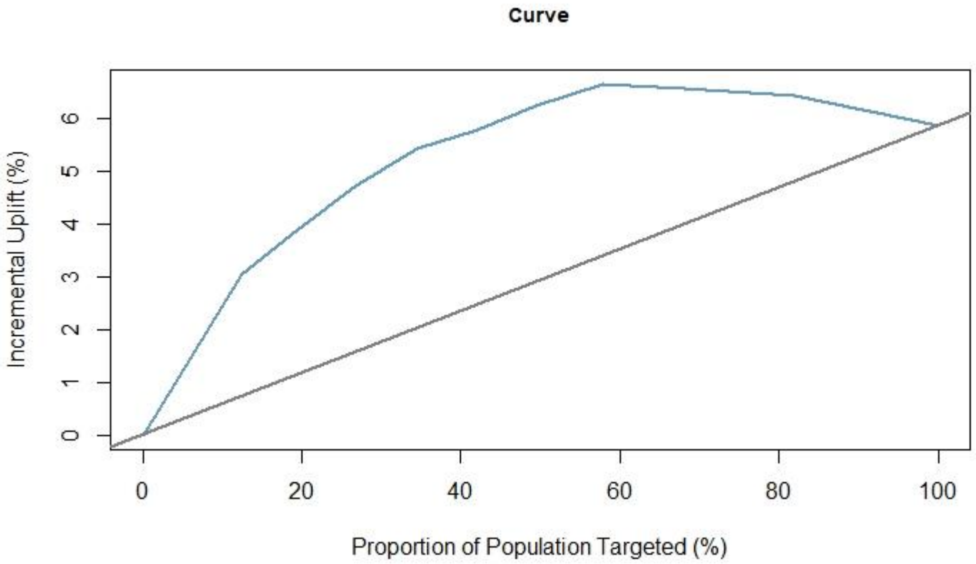


Figure 2: Visualization of Qini curve similar to regular gains curve

Research design & methodology

Previous sections have shown that uplift models within the domain of direct marketing are generally built within specific experimental settings, in which a product, discount or retention related marketing campaigns could act as an intervention (treatment). Besides building a successful uplift model for campaign and targeting purposes, this study aims to explore the application of uplift models for conversion optimization within web-based and dynamic sessions. For instance, it remains unclear whether uplift modeling can be used for conversion optimization in situations of more dynamic and anonymous web-visits by means of personalized web-based treatments such as web page layout, call-to-action (CTA) button, adds-on or pop-ups. In order to discover how uplift modeling can be applied in broader terms than solely for the purpose of campaign and targeting optimization, this report will address and compare two different settings in terms of the appropriateness for uplift modeling. The first empirical setting considers the approach of modeling the incremental effect of an email campaign aiming to enhance future targeting for conversions or retention purposes. Whereas the second empirical setting is focused around a web-based environment, characterized as a more dynamic environment that deals with first-time visits or anonymous web-visitors. Moreover, these two settings will be analyzed and compared on an explorative basis based on the data requirements and appropriateness of building uplift models (Figure 3). In the following sections, a more detailed description of the empirical setting and datasets will be given.

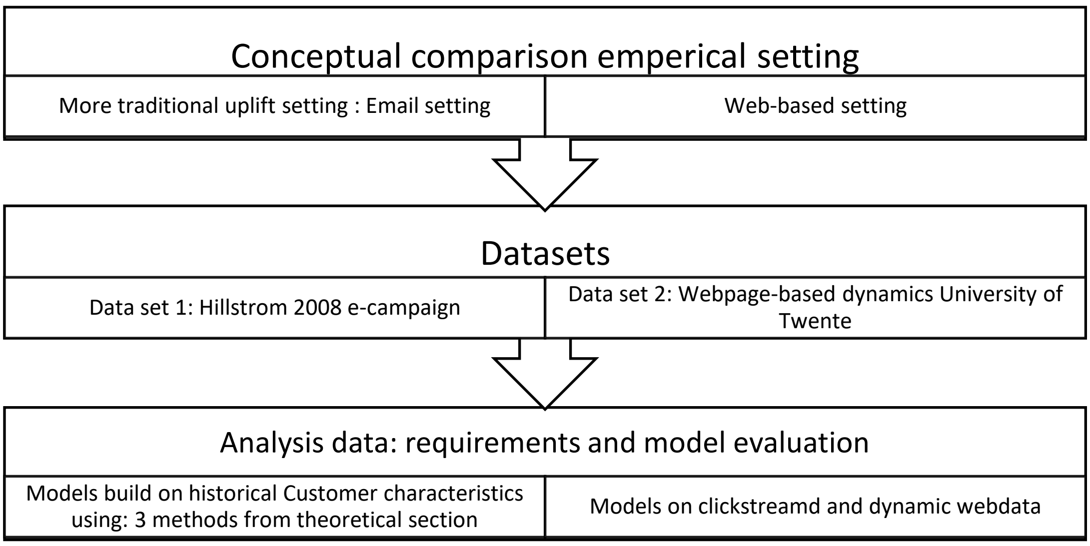


Figure 3: Overview of research design

Conceptual comparison

In order to explore the feasibility and practicability of uplift modeling in two different (direct) marketing settings such as campaigns and websites, it is important to conceptual investigate both settings. Table 4 below addresses some key differences and challenges that distinguish the application of different types of predictive models in these two empirical settings. It also shows how this study compares to relevant other prior research studies on designing and using predictive marketing models (e.g. uplift) in these two settings. One of the key differences between both settings relies on the controllability and accessibility of conducting an experiment for the purpose of uplift modeling. In an emailing setting, it is possible to control that the subjects only receive the treatment once within the experimental period. Whereas in a web-based setting, web-visitors can visit the website (and be exposed to treatment) whenever. For instance, potential interactions with the web-based system could influence the subsequent design or ad exposure, based on the reaction of the user (Diemert & Renaudin, 2018). Therefore, collecting dynamic web-data is more difficult to control and requires more technical conditions within the process of experimental data collection. As a result, Diemert and Renaudin (2018) mention the importance of collecting data for uplift experiments at the first interaction or to log user's variables at the start of the test and observe the reward. Gubela (2019) argues that the appropriate approach is dependent on the characteristics of the web production system and the type of response variable or type of conversions such as website visits, clicks. Similarly, Mokryn et al. (2019) have shown that either tracking registered visitors or logging visitors' online behavior such as browsing sequences (clickstream data) enables marketers and researchers to model action and to predict purchase intent in an early stage. Also, time-specificity and accessibility distinguish both settings. This is because an email campaign as a treatment can be (automatically) scheduled and the response can be observed given a certain period of time. In contrast, the current generation of digital and web-based production systems are characterized by users who are targeted dynamically based on their observed interactions over time (Diemert & Renaudin, 2018). Ultimately, conducting an uplift experiment within a web-based can be seen as a more continuously accessible setting that requires, a trade-off to be made on either considering only the first interaction of a user during a randomized trial or to log the user variables at the start of the test and observe the reward/behavior during the experimental period.

Besides this, the design of an uplift experiment in both settings shares the requirement of permission from the customer or web-visitor, but rather in distinctive ways. In the email setting, permission is required in order to contact the customer in the first place. Email marketing is seen as a permission-based approach within the direct marketing domain, which is based on the assumption that a customer who has given permission to be contacted is a more loyal, profitable customer and should be retained (Hasouneh & Alqeed, 2010). Conversely, the chance of a favorable response and potential positive uplift effect is assumed to be immensely smaller under the circumstance of contacting customers or prospects that did not give permission to be approached or treated by email. Whereas the web-based setting permission is not directly required for contacting (treating) the web-visitor, but rather to use cookie technology or registration requirements to track and collect behavioral information and the response of an individual web-visitor (Hansotia & Rukstales, 2002). This difference is also associated with the difference in data input that is being collected and used within uplift modeling. Collecting data and modeling uplift within an email setting is based on known customer information such as demographic and historical purchase information and responses of previous interventions (campaigns). This information is used to estimate the uplift and is the starting point for marketers and researchers to design and optimize future campaigns based on the estimated uplift. Whereas a web-based setting often does not feature rich historical customer or prospect information and rather deals with unregistered visitors and dynamic online behavior. Each visitor is characterized by general dynamics in terms of clickstream information and in-session dynamics such as dwell time and page visiting time, which subsequently can be used for conversion prediction for either the current or next visit (Gubela et al., 2019). In conclusion, uplift modeling within an email setting aims to model the incremental response in relation to the observed treatment heterogeneity effect that considers the treatment effectiveness to vary with historical characteristics and previous responses of individuals. Whereas the web-based setting considers the treatment effectiveness to vary with historical information (if available), dynamic behavioral and in-session characteristics of individual web-users.

Another important aspect when conducting an experiment for uplift purposes consists of addressing the (potential) marketing costs of both settings. Addressing these costs is not necessarily required for conversion modeling that strives at maximizing incremental sales or visits (conversions), but it is rather important for a revenue uplift model that aims at maximizing incremental revenues (Gubela et al., 2020). The objective of the latter approach of modeling uplift is only applicable within the email setting and within a web-based setting, which deals with registered or identified visitors. This is because revenue uplift modeling is based on historical customer information and often considers a coupon that is either sent by mail or email campaign. Revenue uplift models assume that customers exhibit heterogeneity in their (historical) spending and aim to model the incremental campaign profit (Gubela et al., 2020). Hanstosia and Rukstakles (2002) have developed a set of incremental break-even rules that can be integrated within uplift machine learning algorithms, which focus on estimating the expected profit subtracted by the cost of the campaign (couponing discount). Also, break-even rules could be integrated, which emphasize a contractual (response) setting by focusing on Net present value, customer lifetime or Long-term value (LTV). Within this break-even rule, the expected LTV under treatment should exceed LTV under control. Additionally, Gubela et al. (2020) make a further distinction in marketing costs by dividing the costs of experimental campaigns into contact costs and cost of incentive. Contact costs occur whenever a customer has been treated and depend on the number of targeted people. These costs can vary from zero or near-zero to several euros (automated e-couponing vs outbound campaigns and call campaigns). Cost of incentive covers the cost that is associated with the incentive of the campaign (treatment) such as the coupon discount which is offered to persuade customers (Gubela et al., 2020). Incentive costs can be either absolute or relative and depend on the customer’s actual response and potentially the shopping basket size. Additionally, the classification of being a persuadable used in uplift modeling prediction is dependent on the campaign discount (Devriendt et al., 2020). For example, a customer could be lost cause when offered a 5% discount and a persuadable when offered 10%. The costs in a web-based setting however, depend on the type of treatment. For instance, couponing as treatment is also possible within a web-based setting for non-register or unidentified customers with similar costs as in the campaign setting. Treatment such as a modified web-layout, UX design, CTA design, adds-on or pop-ups screens recommending a product will have zero or near non-zero costs except the potential time and resources spent on capturing (logging) visitor’s online behavior. In conclusion, there are some differences in costs when designing an uplift experiment within a campaign or web-based setting, which mainly depends on the type of treatment. Furthermore, the type of costs depends on whether the objective of uplift modeling is to identify customers for whom the marketing treatment had a sufficiently large impact either on response likelihood (conversion modeling) or on the resulting profit (revenue modeling) in order for the marketing investment to be justified.

The final distinction that can be made between email and web-based setting, which is also a summarization of the above-mentioned differences and similarities, consists of the key objective when designing experiments and building uplift models. Conversion or revenue uplift modeling in an email-based setting requires permission and uses rich historical information to optimize future targeting to ultimately maximize sales, profits, or customer retention. A web-based setting often does not feature rich historical customer or prospect information and often deals with unregistered visitors/customers. Uplift models in the latter setting are more focused on predicting the expected conversion in terms of subscription for email, purchases or specific page visits. Therefore, it could be argued that uplift modeling in campaigns setting focuses on improving customer retention, reviving and retaining customers within their existing customer base and ultimately increasing customer lifetime value. Whereas the prominent objective within a web-based setting is more focused on conversion optimization, customer acquisition (registration) and cross-selling. Additionally, conducting an uplift experiment in a web-based setting could also be based on modeling the incremental effect of different web-lay-outs for UX optimization purposes and personalized or interactive web-based stimuli such as pop-up ads, which could enhance the shopping experience itself. Although the differences in designing and employing uplift experiments in both settings, they share and fulfill the purpose of not only selecting customers for marketing programs but also aim to optimally match customers to different treatments. To conclude, uplift modeling in a campaign setting is besides profit optimization useful for future campaign targeting and retaining and the revival of customers. In contrast, uplift modeling in a web-based setting can enhance acquisition in the first place, which subsequently could be used for campaign optimization and profit optimization.

Summary	**Modeling approach**	**Empirical setting**	**Dynamic or historical data input?**	**Objectives/business value**	**Contribution**
Diemert and Renaudin (2018)	Uplift	Campaign	Both	Predicting conversion	Emphasized requirements and sanity checks for conducting and collecting large experimental data for uplift purposes on the current web-based production systems
Mokryn, Veronika and Kuflik (2019)	Traditional classification (logistic regression)	Web-based	Both	Predicting purchase intent	Logging visitors online behavior such as browsing sequences (click-stream data) enable marketers and researchers to model action and to predict purchase intent in an early stage
Gubela, Bequé, Gebert and Lessmann (2018)	Uplift	Campaign including a discount	Both	Predicting conversion	Compares model performances in terms of business value and advantages of uplift models that emphasize conversion prediction in an E-commerce setting using a campaign (covering a coupon) as treatment.
Gubela, lessmann, and Jaroszewicz (2018)	Uplift	Campaign including a coupon discount	Historical	Maximizing incremental revenues	Introduced new target variable transformation to enable revenue uplift modeling, which in contrast to conversion uplift modeling, directly accounts for the heterogeneity in customer spending and targets customers to maximize the difference between revenue and campaign costs from couponing.
Hansotia and Rukstales (2002)	Uplift	Campaign	Both	Optimizing campaign design and targeting process, so that the return on direct marketing investment exceeds the firms' hurdle rate for these investments	Distinguishes design rules of uplift models that model incremental visit/conversion from a customer who would not be likely to visit in the absence of the promotion and the incremental spending on the same visit. Both models could improve promotional communications, form a program analysis perspective as well as program enhancement activity
Ascarza (2018)	Uplift	Campaign	Historical customer base information	Optimizing retention through sophisticated targeting	Targeting customers with proactive churn programs based on their sensitivity to the program is more effective instead of identifying and targeting the customers with the highest risk of churning.
Hasouneh and Alqeed (2010)	Qualitative study	Both	Both	Examines the role response data from direct email campaigns in relation to the development of a loyal customer relationship	Analyzing interaction or response data from e-mail direct marketing campaigns provides new insight into the development and retainment of a long-lasting customer relationship. Tracking click-through activities enables monitoring the relationship development in and between monetary transactions and reflects the level of interest towards marketing programs, in particular loyalty programs.

Van Den Poel et al. (2005)	Logit modeling (response)	Web-based	Both	Prediction of Purchase intent	Out of different dynamic data and variable input categories, detailed clickstream variables are the most important ones in classifying customers according to their purchase probability

Table 4: Overview conceptual comparison of previously designed predictive models within a campaign and web-based setting

Description of data set 1 for practical analysis (campaign setting)

In order to practically investigate the feasibility of uplift models across two different settings, a description should be given of the dataset and experimental procedure. The first empirical setting of the two cases considers the traditional direct marketing setting regarding an email campaign of Hillstrom 2008, which is an internet-based retailer. This publicly available marketing campaign dataset contains information about 64,000 customers who last purchased a product within twelve months and will hereafter be referred to as data set 1. The individual subjects within data set 1 were subjected to test a promotional email campaign (treatment) and were randomly assigned to one of three groups and. As a result, 1/3 of subjects have been treated with an email campaign concerning men's clothing and 1/3 of the subjects have been targeted with an email campaign concerning women's clothing. The remaining 1/3 of the subjects have not been treated at all. The experimental results were collected over a period of two weeks time following the campaign. The details of the dataset are shown in Table 11 of the appendix representing respectively the marketing data and (historical) customer attributes. The treatment allocation is represented by the variable segment.

Furthermore, the dataset includes two possible target variables: visits and conversion. In our application, we focus on a single treatment instead of a multi-treatment effect. Therefore, the analysis is simplified by restricting the treatment variable from three categories to a binary variable. Hence, the treatment variable resembles whether the subject received an email on women's merchandise (treated group) or if the subject was not targeted by the email campaign (control group). Additionally, we mainly focus on the target variable visit instead of conversion due to the similarity of this dependent variable with the dependent variable of our second empirical setting. These modifications or restrictions do not harm the analysis nor will bias the results. According to Ruda et al. (2018), the distribution of the predictors in the treated and controlled groups should be identical in order for uplift models to have a causal interpretation. This can practically be achieved through complete randomization and assigning cases to both groups, where the assignment is random and independent of the predictors. Dataset 1 consists of complete randomized treatment allocation, where the subjects were randomly assigned to the treatment and control group. This will be checked in the subsequent sections within this report.

Description data set 2 for practical analysis (web-based setting)

In order to investigate the feasibility of uplift modeling in a digital and web-based environment, data has been collected on the web-production system (website) of the University of Twente. The data collection follows a randomized trial (A/B test) setup and considers the perspective of more anonymous based dynamic web surfers and –visitors. In this case, there is less historical customer information available and the available information is rather based on in-session dynamics and click-stream oriented data. Data has been collected on the first and unique interaction or visit of the web-surfer on the webpage of the University regarding the topic of Master's tuition fees. Specifically, click-stream information and dynamic behavior have been captured regarding the navigational pathways within the web-production system and response information has been captured regarding the CTA conversion from web visitors on the topic of Master's tuition fees. The experiment has been set up via the Google Optimize tool and the data will be collected anonymously through the usage of Hotjar. Hotjar is a web-analytical tool that captures behavioral and clickstream data on websites. This tool was chosen since it is able to capture general demographics and generic clickstream data (information at the level of the session) on an individual basis, whereas other analytical data collected tools often capture data on an aggregated basis. The randomized trial A/B test setup incorporates a treatment (A/B) consisting of a modified design of the Master's tuition fees webpage. More specifically, the treatment will consist of a changed Call to Action (CTA) button. Furthermore, website visitors visiting the webpage of the changed CTA will randomly be assigned to either the treatment group (webpage with the changed CTA) or to the control group (webpage with the original CTA button). The predictor variables consist of characteristics and certain dynamic attributes as shown in Table 12 in the appendix. The dependent or target variable of interest is based on a binary variable such as clicked on CTA, which indicated whether the web-visitor clicked on the CTA button and visited the next page. An overview of the variables within this dataset, hereafter mentioned as data set 2, is provided through Table 12 within the appendix.

Procedure of analysis

In order to achieve consistent results that allow a comparison of the feasibility and performance of uplift modeling across two empirical settings, a general procedure will be followed. This procedure will be applied to both datasets and is shown in Figure 4 and Figure 5 of the appendix. Both data sets were based on a complete randomized experiment including a single binary treatment with a randomized split between treatment and control group. The procedure starts with the data cleaning and preprocessing process including dealing with missing values, attaching the correct labels to the variables and preparation of the dataset for uplift modeling purposes. This is followed by explorative data analysis on the dataset. Lastly, the model will be fitted on a training set and validated on a test set. The analysis will be conducted using the programming language R. The packages that are being used are the Dplyr, Uplift and tools4uplift packages. The first package is used for general cleaning and transformation of data (if necessary). The latter two packages are created by Guelman (2016) and Belbahri et al. (2019) and include modified functions of existing machine learning packages to train, test and assess uplift models on their performance. The uplift package of Guelman (2016) focuses on direct uplift approaches based on tree-based and non-parametric classifiers such as Random forests. Whereas the tools4uplift package emphasizes indirect uplift modeling through parametric classifiers such as logistic regression. Both packages are applied in this study to split data in a training and test data set fitting the requirements of uplift and for fitting, assessing and profiling of the uplift models.

Variable	Data set 1	Data set 2
Empirical setting	Online E-commerce setting (Hillstrom, 2008)	Online Web-based setting (webpage University of Twente)
Channel used	E-mail	Web-system
Public or Private	Public	Private
No. of observations	42,693	1,105
No. of treatment observations	21,387	549
No. of Control observations	21,306	556
No. of variables	11 (10 used)	11
Response variable (Binary)	Visit	Visit/clicked on CTA button
No. of response	9394 (14.68%)	14 (1.27%)
Treatment to control size ratio	1:1	1:1
Treatment + Respond positive rate	15.14%	1.27%
Control + respond negative rate	10.62%	1.26%
Average treatment effect	4.52%	0.01%
Models used	Two-model estimator, direct tree method (Random forest) and response transformation method	-

Table 5: Overview of empirical settings and data sets used for practical analysis

Analysis of data set 1

The analysis of dataset 1 consists of uplifting modeling approaches as presented in the theoretical and methodological part of this report. Several uplift approaches will be demonstrated and assessed from the empirical setting of an E-mail based campaign stemming from an E-commerce background (Table 5Table 5). The analysis of uplift results within this setting follows the process of exploratory data analysis, training of the model and finally model testing and assessment.

Exploratory data analysis.

In order to successfully build and potentially optimize uplift models, it is necessary to understand the data by doing an explorative analysis. This is especially important when dealing with experimental datasets (Diemert & Renaudin, 2018). For instance, it is important to check whether the treatment has been completely randomly assigned to the individual subjects or if there is any underlying non-random mechanism that could explain the subject to be treated. In the case of a randomized experiment, the treatment and control groups within the data should be approximately similar regarding their distribution of covariates. Therefore, a dataset for uplift purposes requires a balanced distribution, under the acceptance of a small imbalance amongst covariates that could be caused by chance. This is being checked for dataset 1 by using the *Checkbalance* function in R. This function uses a chi-square distribution to test the null hypothesis of the overall balance of covariates against the alternative hypothesis of (too much) lack of balance. The results of the test of balance among the covariates are shown in the table in the appendix. Table 6 below summarizes the overall result of both the original dataset and the training set. Based on the P-value of 0.73, it can be concluded that there is little evidence against the null hypothesis. The same conclusion can be made if we use the overall test of equal distribution of covariates on the partitioned training dataset. Also, based on the table in the appendix, it can be stated that the mean of covariates in both treated and control groups is fairly balanced, with only a small adjusted difference.

Overall test chi-square distribution	Chi-square	P-value
Hillstrom dataset	11.3	0.73
Training dataset	14.4	0.49

Table 6: Test of the overall balance of covariates among control and treatment group data set 1

Besides checking the data for a randomized treatment assignment, it is important to further explore the data and available predictor variables via univariate analysis. For data set 1, a univariate analysis of uplift is conducted in order to display the potential role of each predictor in modeling uplift. An explorative function is used that computes the average value of the response variable for each type of predictor and treatment assignment. This function places continuous variables into binned quartiles with equal size. Table 7 shows the results of three predictor variables and the results of all predictor variables are shown in the table in the Appendix. The first two columns of Table 7 show the number of responses in the control and treatment groups, respectively. Whereas the following two columns show the average response for control and treatment. Finally, the last column reports the difference between the average control and treatment response (uplift). Considering the results below, the predictor variables Women and Men show the highest magnitude of uplift values over the range of the values of these predictors. These variables can be interpreted as potential treatment heterogeneity effects, due to the fact that marketing intervention consisted of an email campaign for women's clothing. Consequently, the campaign persuaded more women than men. In conclusion, the variable indicating if the individual has bought women's or men's merchandise over the past year seems to be an important predictor for the incremental response in website visits.

Variable	Binned variable	N.Treat = 0	N.Treat = 1	Mean.Resp.Treat = 0	Mean.Resp.Treat = 0	Uplift
History	[30,64.9]	3710	3763	0.0863	0.1345	0.0482
	(64.9,158]	3809	3661	0.0848	0.1292	0.0444
	(158,327]	3669	3802	0.1194	0.1539	0.0345
	(327,3.35e+03]	3660	3811	0.1374	0.1910	0.0536
Women	0	6697	6688	0.1011	0.1123	0.0112
	1	8151	8349	0.1113	0.1846	0.0733
Men	0	6666	6852	0.0966	0.1716	0.0750
	1	8182	8185	0.1149	0.1363	0.0215

Table 7: Exploration of the potential important predictor in estimating uplift

Fitting the models

This section describes the fitting procedure of the three uplift models on the training dataset of the Hillstrom (2008) dataset. The training set is partitioned from the original dataset in a different way than normally used in regular machine learning cases. Due to the nature of the dataset including both control and treated subjects, it is necessary to partition the data set into training and testing whilst remaining the same proportion in treatment and response rate. The two tables below show the treatment-response rate, also known as the initial or average observed treatment effect of the Hillstrom email campaign. The initial treatment effect of the campaign for the original Hillstrom data set and training set is almost similar, with respectively 4.52% and 4.57% (shown in Table 5 and Table 8).

Hillstrom data set	R= 0	R= 1
T= 0	89.383%	84.860%
T= 1	10.617%	15.140%

Training data set	R= 0	R= 1
T= 0	89.332%	85.758%
T= 1	10.668%	15.242%

Table 8: Proportion response rate in treatment & control group in the original and training data set

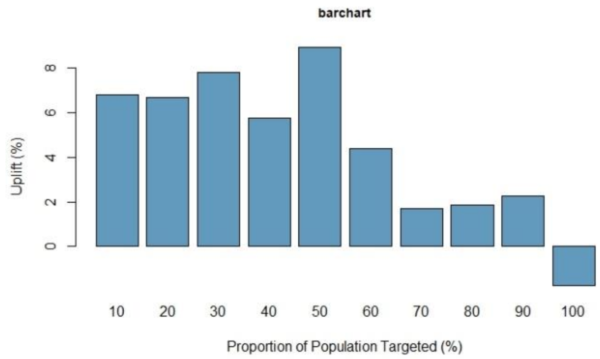
Three alternative uplift models are built on the training data: Two-model estimator, Random Forest and response variable transformation for uplift (RVTU). These models are based on three distinct techniques that were described in the literature section of this report. The random forest approach resembles the tree-based and direct approach to model uplift. The other two models are based on logistic regression. The two-model estimator resembles the most basic, intuitive and indirect approach to uplift modeling. Whereas the RVTU is based on an advanced data transformation technique that aims to prepare the data for the application of any conventional machine learning technique. Through this data transformation, a direct approach modeling uplift can be applied. All three models are fit based on their original variables and values. Although variable selection can improve the predictive performance of these models and reduce the problem of overfitting, no variable selection will be applied within the process of training uplift models. Dataset 1 includes 8 predictor variables, which can be seen as an appropriate number of predictor variables and therefore variable selection is not necessarily required. Additionally, no parameter or model tuning is applied, since the initial goal of this study is not to compare and achieve the optimal performance of these models, but rather to explore and demonstrate their use and to interpret the results.

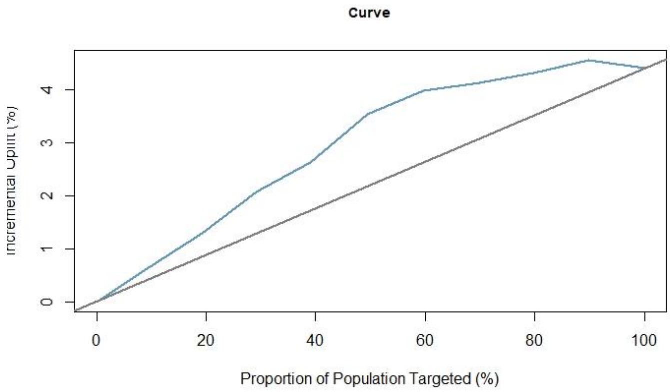
Empirical setting 1: Results of data set 1

In this section, the results are reported of the trained and validated uplift models. After the models have been trained, the models will be validated on the test data set using a prediction method. These predictions contain conditional class probabilities under both treatment and control groups. Next, the predictive performance of these models is assessed by the usage of two functions in R, which cover the performance and Qini measure of the model. The performance is measured by computing the differences of the predicted conditional class probabilities, ranking and grouping these into bins/deciles of an equal number of observations. Lastly, the actual difference in the mean of the response variable will be calculated. The performance of the three alternative techniques will be visualized through a bar chart and through the Qini or gains curve. Afterward, a summarization of all three modeling techniques will be reported based on the Qini measure, top 10 uplift and profiling of subjects based on their estimated uplift

Two-model estimator

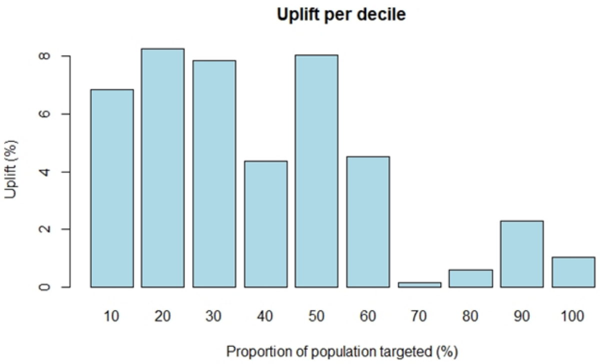
This method of modeling uplift is based on the subtraction of two glm objects, where each object is a fitted logistic regression on the separated non-treated and treated individuals within the data. A summary of the coefficients of the two independent models is shown in the appendix. The bar chart below is based on the table shown in the appendix, which is the output of the performance function in R. When observing the bar chart below, it can be concluded that the model was not fully able to sort the individual in groups from the highest to lowest uplift for the purposes of targeting optimization. Additionally, the bar chart can be interpreted that targeting almost the whole population will decrease the incremental responses and ultimately lead to a slightly negative uplift or down lift. This is also reflected by the drop-off within the Qini curve chart, which happens if 100% of the population is targeted based on the ranked individuals instead of 90%. This suggests that some individuals grouped in the last few deciles are characterized as do-not-disturbs, which might show a negative treatment effect.

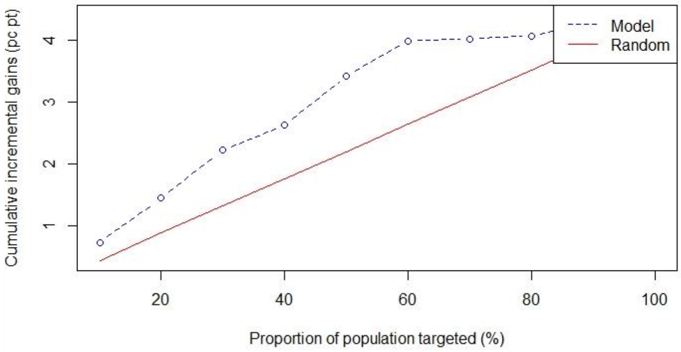




Random Forest

This method of modeling uplift is based on the tree-based approach using the split criteria that is designed for uplift problems or personalized treatment problems. The Random forest algorithm used is different from the traditional, such that an ensemble of B trees are grown, each built on a fraction of the training data including both treatment and control subjects. Also, the splitting criteria are based on the measure of distributional divergence stated by Rzepakowski & Jaroszewicz (2012), which is discussed in the literature section of this report. The distributional divergence in this analysis is based on the squared Euclidean distance, but other alternative measures such as Kullback-Leibler, Chi-Squared divergence and L1-norm divergence could also be applied for model optimization. The uplift random forest model was trained based on the preliminary variables and standard parameters were not changed. The number of trees (randomly) generated in the forest consists of 100. Although random forest greatly reduces the chance of overfitting in comparison to single tree development, there is still a chance for it to occur when the trees in the forest are grown to maximum depth. This holds especially true under the circumstance of a high amount of noise in the data. Therefore, the maximum depth has been limited, which also reduced the computability time within R of producing the Random Forest uplift model. The results of the Uplift random forest model are visually shown with the bar chart and Qini curve below. It can be observed that uplift random forest model was not fully able to ideally sort the individuals based on their incremental treatment effect from high to low. Additionally, both charts illustrate only positive uplift results. The Qini curve implies that when the first 60% of the individuals are targeted, a similar level of uplift can be achieved when targeting the whole population.





Response Variable transformation for uplift (RVTU)

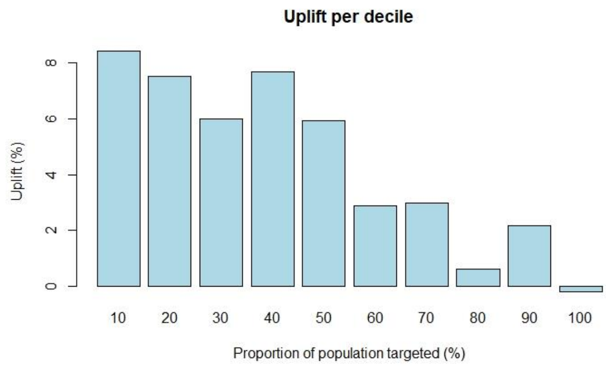
This method of modeling uplift is based on the direct estimation approach, where the (training) dataset is transformed in order for it to be used with any conventional supervised machine learning algorithm. This method was introduced by (Rzepakowski & Jaroszewicz, 2012a) and consists of the transformation of the original binary response variable into a new variable Z. The new variable Z includes the treatment and respond value, see the equation below.

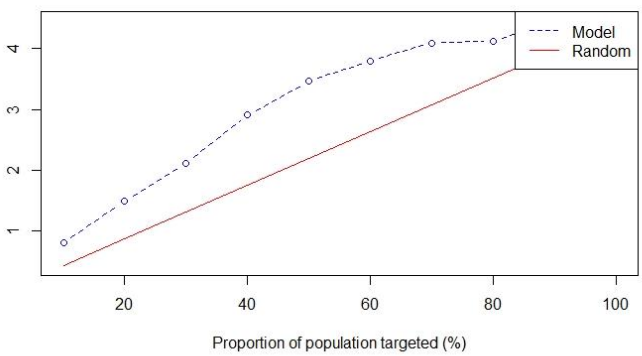
$$Z = \begin{cases} 1 \; if \; t = 1 \; and \; Y = 1 \\ 1 \; if \; t = 0 \; and \; y = 0 \\ 0 \; otherwise \end{cases}$$

Moreover, this new variable is established under the assumption that a subject who both was treated and responded or was not treated (control) and did not respond is more favorable (positive uplift) than a subject that falls within the negative uplift quadrant of Table 2. Hence, if we would have known the outcome of a subject in both the treated and control groups, then Z equals 1 based on the idea that the outcome in the treated group would have been as good as in the control group.

Hence, z equals 1 if we know that for a given subject the outcome would have been at least as good as in the control group. Due to this assumption and transformation of the treatment and response variable into one single response variable, a conventional binary regression can be fitted on to model Z on the baseline covariates. In case of a balanced data set with an equal proportion of control and treated observation, Rzepakowski & Jaroszewicz (2012) have proven that $2 * Prob(z = 1|x) - 1 = Prob(y = 1| \; treated, x) - Prob(y = 1|control, x)$

The variable importance is extracted from the fitted linear model and is shown in the Appendix. The variable importance charts reports that the variables and value combination of Women1, ZipcodeUrban and history are most important in predicting the response variable visits. The uplift chart and gains chart below is the result of a logistic regression (glm object) model fitted on the transformed (training) data. Although the model did not fully sort all individuals in groups from highest to lowest uplift, the pattern of the uplift chart shows that the model can still be interpreted as a fairly good model. This is due to its ability to rank individuals who were treated and responded high in the first decile, whereas responders of the control group are ranked lower in the last decile. This is also reflected by the steepness of the gains curve that models the incremental uplift. Although the uplift is positive for all deciles and the gains curve shows that the model performs better than random targeting, the model does show an improvement in uplift in comparison with targeting the whole population. For instance, an uplift of 3% can be observed from the incremental gains chart when the first 40% of the population is targeted, whereas randomly targeting 40% of the population results in an uplift of 1.75%. However, it can also be observed that even when 80% of the people are targeted, the uplift is as high for this dataset as targeting the full population.





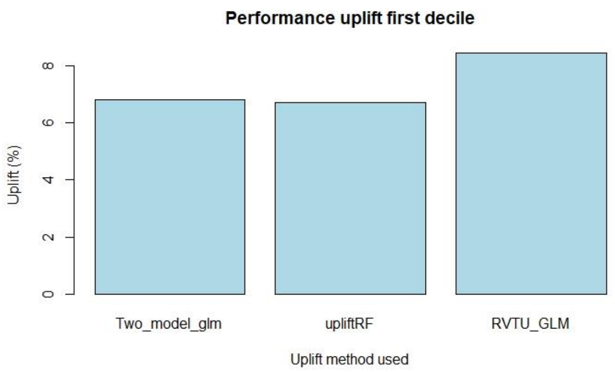
Final model assessment & conclusion

The three alternative uplift techniques used are all close in terms of performance. The Qini metric and the plotted bar charts illustrate the capability of the model to rank the full population according to its uplift. The two-model estimator, which is the most intuitive and naïve method of modeling uplift had surprisingly the highest Qini measure (Table 9). Therefore, it can be stated that the two-model technique achieved the highest overall accuracy in ranking the whole population according to its uplift and therefore performed better in accuracy in comparison to the other two modeling techniques. However, this model also suggested a negative uplift or down lift when targeting the whole population, which is conflicting with the results of the direct uplift modeling techniques. Additionally, the tables in the appendix of the coefficients of the two-model estimate technique in comparison with the RVTU technique are quite different from each other. The coefficients of the two independent linear models showed that several variables were significant in predicting the response visit, while the single linear model from the RVTU method only indicated that the variable women's had a significant positive effect in predating visits. While the two-model estimator is a very intuitive method of modeling uplift and can be handled with conventional supervised machine learning methods, this modeling technique is also known for its naive performance from an uplift perspective in real-world situations (Radcliffe & Surry, 2011). Since the two models are built independently from each other on the control and treatment group, the objective of these models could disregard one another and emphasize the prediction of the class probability of response over the actual prediction of uplift.

Although the Qini measure is of theoretical importance, from a practical perspective it is often not efficient to target whole populations. This is because the aim of uplift models is to rank the population and select the best subset of individuals for whom the treatment is expected to have an increased effect. For instance, marketing campaigns including a promotional discount or advertisement may only target 20% of the customer base. Therefore, it is more important to know the group of individuals that are ranked and grouped within the 10% or 20% highest decile. Currently, there is no available numeric metric similar to the Qini measure that emphasizes the targeting value or depth. As a result, the figure below shows the first decile consisting of the top 10% ranked individuals with the highest uplift from each uplift modeling technique. By observing the earlier individual bar charts and the figure below, it can be stated that only the response variable transformation technique was able to sort and group the individuals with the highest uplift in the first decile.

Modeling technique	Two-mod	Uplift RF	RVTU
Qini value	0.761	0.077	0.073

Table 9: Qini value of different modeled uplift algorithms



In general, the three uplift bar charts and Qini curves show positive uplift results for every decile/group. The visualizations of all three methods show that the uplift is distributed over the various deciles, instead of being concentrated in the first (few) deciles that resemble the top uplift. Similarly, the Qini curve shows that not one single technique achieved an improved or increased uplift than the overall and initially observed response/uplift from the campaign, which was 4.52%. Moreover, the incremental uplift modeled in the Qini curve does not bend down in cases of targeting a larger fraction of the population. Consequently, this means that no negative uplift or down lift would occur if an individual was being selected from higher fractions of the uplift model for targeting purposes. For instance, no negative uplift or down lift will occur if we target 90% of the customer base instead of only targeting 30%. A possible explanation might be that there are only a few or even no individual subjects who could be indicated as do-not-disturbs in the customer population of data set 1. In conclusion, the established uplift models show no significant improvement nor a significant down lift of targeting individuals from the latter and lower scored deciles, due to lack of negative incremental treatment effects and due to the relatively high initial or average treatment effect of the campaign itself. Consequently, the established uplift models experience difficulties in accurately identifying those individuals who can be viewed as persuadable since there is a lack of a negative treatment effect.

Model profiling

The final step within the process of uplift modeling is to profile individuals (customers) that according to the model have been ranked from highest to lowest uplift in visits in relation to the marketing intervention. The profiling is thus based on the uplift predictions of the model and returns a basic summary of the predictor variables that were used to predict, rank and group the individuals according to their uplift. Table 14 in the appendix illustrates the profile from the RVTU model, based on the 4 most influential variables derived from the variable importance graph of the glm object. The profiling function computes the average of each numeric variable and reports the distribution of each factor within each group/decile. The table shows that top uplift consists of individuals (customers) who have bought women's products in the past year, whilst the latter fraction/ deciles of lower uplift resemble individuals that have bought more men's products. This finding is not that surprising, since the direct marketing treatment or intervention constituted a promotional campaign for women's clothing. Additionally, customers ranked in the top decile (1st group) were indicated with the Zipcode Urban, whilst only a few were indicated as Rural. Lastly, it can be observed that customers ranked in the first three deciles were indicated by a lower amount of dollars spent on their last purchase ranging from $0 -$100 and from $100 - $200. Since our model was not fully accurate in ranking all subjects based on their uplift from high to low, the summary static results of some predictor variables are somewhat unambiguously. However, it can be concluded that profiling provides a general overview of the characteristics of subjects that scored high or low on the treatment effect. This information can subsequently be used for future targeting, campaign and ultimately sales optimization.

Empirical setting 2: Practical analysis of the web-based setting

In this section, the results of the analysis regarding the feasibility of uplift within the second empirical setting are described using the web-based dataset of the University of Twente. Table 10 below shows the proportion in treatment (T) and response rate (R) within data set 2. The table indicates a very low proportion of responders in the control and treatment group of 1.26% and 1.28% respectively. The difference between these two is the average treatment effect (0.02%), which is almost negligible. This can partially be explained by the fact that the overall observed response during this uplift experiment was too low (see Table 5). Partitioning the dataset into training and test, which is required for uplift modeling, would even further reduce the proportion of responders. As a result, this dataset collected is practically not suitable for an empirical quantitative analysis regarding the feasibility and business value of uplift modeling within a web-based setting. Therefore, the practical analysis is conducted in a qualitative manner. In the following section experienced challenges, requirements and solutions are described that could overcome the challenges regarding the appropriate design, application and potential business value of uplift modeling within a web-based setting.

UT web-based data set	Treatment = 0	Treatment = 1
Respond = 0	98.741	98.725
Respond = 1	1.259%	1.275%

Table 10: Proportion response in treatment & control group data set 2

One of the key challenges faced in designing and conducting uplift experiments within a web-based setting includes the permission and capability of tracking individual web-visitors behavior. This challenge is mainly caused by the requirement for uplift modeling experiments to be based on data that is captured at the individual level. This is because uplift modeling is associated with the task of causal inference, in which a treatment in relation to a certain outcome or response is tied to an individual (Sołtys & Jaroszewicz, 2015). Consequently, the objective is to optimize treatment effects based on the premise of treating customers/web-visitors differently. Especially within E-commerce websites, a registration requirement and the usage of cookie technology is required in order to be able to track current-session characteristics and historical information and to identify customers who visit and browse within their sites (Hansotia & Rukstales, 2002). This challenge becomes even more difficult for websites without registration or log-in requirement, due to both technical and privacy restrictions. In these circumstances, visitors are also able to visit the website at any given time. Therefore treatment exposure has to be controlled by either collecting data at the first interaction or by observing the reward between the first and last interaction since multiple interactions with a treatment could impact the behavioral response. This requires the ability to distinguish unique web-visitors from each other. Furthermore, privacy restrictions also enable most web-based behavioral analytical tools to capture data primarily on an anonymous and aggregated customer basis in order to comply with privacy and GDPR ruling (Dwyer, 2009). Therefore, it can be concluded that collecting information regarding individual clickstream data as input variables as well as collecting information on individual responses in relation to a certain online marketing intervention is extremely challenging within a dynamic web-based setting and requires specific tools or methods.

The UT dataset was established by using Hotjar as a behavioral data collection tool in order to overcome the limitation of other tools such as Google Analytics and Google Optimize that capture information on an aggregated basis. Although Hotjar was able to collect anonymous-based clickstream data on an individual basis, the prevalence of the above-mentioned challenges mainly resulted in a low number of overall observations, low (average) response levels as well as both imbalanced and treatment response levels (See Table 5 and 10). This resulted in the impracticability of this data set to be used for uplift modeling purposes. The growing awareness of privacy risks among web users stated by Dwyer (2009) might explain the low overall and average responses since more web-visitors might block and delete their cookies. This is also indicated by the table shown in the appendix, which shows that a higher number of observed experimental sessions (subjects) and conversion or response rates from Google Analytics & Optimize in comparison with the observed data from Hotjar during the same experimental period. The former data is collected through web-based server log data on an aggregated basis whereas the latter one is based on cookie acceptance/permission and includes anonymous data on an individual basis. Moreover, this highlights the importance and influence of visitor's permission and acceptance towards websites to use cookie technology to track a vast amount of behavioral information such as clickstream data and the firm's ability to track visitor's responses in relation to a web-based treatment.

Although the awareness of privacy risks and cookie restrictions limit the ability to track (individual) behavioral data required for uplift experiments, a few alternative methods or workarounds have been mentioned within the literature of online behavioral analytics (OBA). Analytical tools such as Google Analytics and Hotjar apply similar tracking technologies in order to collect behavioral information by means of page tagging or cookie tagging (Hernandez et al., 2017). Besides this technique of behavioral data collection, log-based techniques exist in which behavioral information of web servers is stored in web server logs (Kwan et al., 2005; Van Den Poel & Buckinx, 2005). This approach is based on server data that is generated by the interactions between a person browsing a website and the webserver, which subsequently can be stored into log files and query data (Büchner & Mulvenna, 1998). Therefore, log files include the hidden and valuable users' behavior and can be seen as event logs that for example consist of click events on items through different pathways and purchase events. According to Hernandez et al. (2017), page tagging techniques are more disadvantageous in comparison to log-based techniques of behavioral data collection, due to their dependence on cookie usage and acceptance. Diemert and Renaudin (2018) argue that the challenge within the log-based approach consists of identifying and differentiating unique web-visitors/customers. However, this problem can be overcome by storing the log files into a common log format where two variables such as IP-address and the data of visit can be used to identify unique visitors (Van Den Poel & Buckinx, 2005). Additionally, this approach for behavioral data collection requires appropriate storage space in terms of databases and requires more extensive preprocessing and data mining activities in order to gain data usable for modeling purposes (Kwan et al., 2005). In conclusion, different methods can be used for behavioral data collection that could enhance the capability of tracking users' online behavior in terms of clickstream data and demographics used as input variables required for uplift models. These alternative methods could especially be helpful in web-based scenarios that deal with unregistered or anonymous web-users in order to increase the response levels among treated and non-treated web visitors. Finally, it is important for the success of the design and application of an uplift experiment in the web-based setting to choose a webpage or part of a website that includes a sufficient amount of traffic.

Business value of uplift experiments in online (web-based) marketing

A website is an open system where almost any form of customer behavior is possible. Web-users are offered a wide variety of navigational options, leading into multiple navigational traversal paths to visit a product or information, to buy a product or to register on the website. This results in the challenging task of marketers and researchers to firstly collect and secondly to be able to analyze, model and target this behavior to improve the website structure, to personalize contents, advertise and recommend products. Additionally, marketers and researchers have a widespread of marketing actions or interventions at their disposal in order to further change the observed behavior in a favorable behavior leading to a favorable response (conversion). These two tasks are intertwined with each other. This is because the application of a pop-up screen recommending a product is an example of a marketing intervention and form of direct communication that could lead to a favorable response for some web users, whilst it could result in a negative response for others. Moreover, uplift modeling within a web-based setting can be seen as the modeling approach which uses and combines the premise of behavioral targeting (as input variables) to estimate potential heterogeneous treatment effects of individual web visitors characterized by their dynamic and behavioral attributes. Once, the uplift models are established, the subjects can be grouped and ranked according to their incremental response and profiling can be applied as shown in the practical analysis of empirical setting 1. Profiling in this case would provide an overview of the in-session dynamic and historical (if available) characteristics of subjects that scored high or low on the web-based treatment. This information can subsequently be used for different goals such as future web (UX) optimization, personalization and conversion optimization.

However, in order for E-business and marketers to realize and capture the business value of uplift in a web-based setting, certain costs or investments are required. As shown in the previous section, E-businesses are challenged in terms of their capability of collecting and tracking (individual) behavioral information (e.g. clickstream and response information) required for uplift experiments. Therefore, time and resource investments have to be made regarding the establishment of analytical behavioral tracking tools and/or manual logging techniques whilst also preventing potential privacy risks. These tools or techniques are especially required in a web-based scenario, which deals with unregistered or anonymous web-users to increase the response levels among treated and non-treated web visitors. Whereas these techniques are less indispensable in more controlled web-based environments where visitors are required to log in order to be able to browse certain pages and products. Besides technical and resource-related costs, the costs of the treatment and the conventional process of targeting customers have to be considered. This because uplift models are developed on the premise of targeting a certain (small) fraction of the population, which resembles the subset of customers ranked by the model that is expected to have the highest incremental treatment effect (uplift). Consequently, it would be not efficient and optimal to target larger fractions of the customer base or even the whole customer base since this would discourage the need for uplift. Moreover, from a practical and business perspective, it is more interesting to identify as much uplift as possible in smaller deciles or fractions of the customer base, since the aim of uplift is to discover and to identify what uplift is achieved amongst for the (e.g. 10%) highest-ranked customers. This subsequently is the starting point of uplift for targeting optimization and future targeting of web-based marketing interventions or campaigns. As a result, the discovery of the highest-ranked customers is especially important if a web-stimuli such as an advertisement is used that includes a discount or some sort of direct business or treatment costs, since targeting larger customer fractions may be inhibitory for achieving profit. This depends however on the type of treatment and business environment in which the uplift experiment is designed and applied. Ultimately, a trade-off has to be made regarding the feasibility of uplift in the web-based setting in terms of the costs and benefits involved in designing the experiment and targeting customers.

Conclusion

This report is based on an exploratory research approach regarding the broader feasibility and application of uplift modeling in marketing, by emphasizing two (direct) marketing settings: email campaigns and websites. The findings within this report have answered the central regarding how businesses, marketers and researchers can employ uplift modeling for future targeting optimization and conversion optimization within a web-based setting. First of all, the conceptual comparison of the application and feasibility of uplift modeling in a direct campaign setting and web-based setting has highlighted some key differences and similarities regarding the experimental design process of uplift models. The most influential requirement shared between the two settings consist of the collection of counterfactual data on an individual and anonymous basis since uplift modeling is based on the principle of estimating individual/personalized treatment effects causally. Secondly, an important difference consisted of the permission and accessibility of collecting customer information within a randomized trial. Within the email setting, customers or prospects often give direct permission to be contacted and to be exposed to a direct promotion or communication (treatment). However, in a web-based setting visitors can theoretically be exposed to the treatment/web-stimuli without a direct permission requirement, other than the (indirect) permission with regards to the acceptance of cookies. Additionally, the conducted uplift analysis within the campaign and web-based setting has shown that customer profiling is an important and valuable aspect within uplift modeling since it gives marketers useful insights into the characteristics of subjects in relation to the high or low estimated incremental treatment effect (uplift). This information can subsequently be used for future targeting and conversion optimization to ultimately fulfill business objectives such as retaining customers and improving (cross) sales through direct marketing.

Besides this, the practical and qualitative analysis of the web-based setting has shown how different behavioral data collection methods, following the principle of behavioral targeting, can be employed within an uplift experiment and can result in a different number of experimental observations and responses. Page-tagging and cookie tagging techniques have shown to be disadvantageous techniques, which heavily depend on cookie acceptance of web-visitors to capture individual behavioral information. Consequently, the requirement of capturing a large amount of individual web-based data in combination with the faced data collection constraints has resulted in low overall and balanced response levels, which is inappropriate for the case of uplift modeling. Although other manual programming and log-based methods are less dependent on cookie acceptance, these methods could raise some concerns regarding web-visitors' privacy. Additionally, these methods could be time and resource-consuming because extensive data mining approaches are required such as data pre-processing, data storage and cleaning. With regards to accessibility, uplift experiments within a web-based setting differ from the campaign settings since visitors can be exposed to a treatment at any given time and potential subsequent interactions with the marketing web-based stimuli could influence the response behavior. Webpage and cookie-based tagging techniques have the ability to identify unique customers and thus can identify and capture their first interactions. Whereas log-based methods have to observe the reward between first and last interaction with the web-stimuli. Ultimately, it can be concluded that the feasibility of designing uplift models is more challenging in a dynamic web-based setting in comparison to the design of an uplift experiment within a campaign setting. As a result, uplift requires a controllable web-based environment, where marketers as decision-makers have control over one or multiple (personalized) web-stimuli and have the capability to capture individual behavioral information to be able to model the differential (causal) effect of a web-based marketing incentive on customer behavior.

Practical implications

The conceptual comparison as well as the practical analysis of both settings have pinpointed some interesting findings and implications for businesses and researchers. A key finding consists of the potential specificity of uplift modeling with regards to the predictive performance in practical usage and in relation to characteristics of certain circumstances or environments. For instance, it was striking that data set 1 (Hillstrom, 2008) originally achieved a similar overall uplift (average treatment effect), than any of the modeled uplift algorithms. A potential explanation might be that the customer base of data set 1 consists of only a few individuals who would have negatively responded or would have simply not reacted to a marketing stimulus, resulting in the absence of negative uplift or down lift when targeting a higher fraction of the customer base. This indicates that the design and application of uplift models to be more effective in circumstances or business environments where decision-makers have control over a certain treatment (e.g. campaign or web-stimuli) whose manipulation is expected to cause a significant behavioral change. Moreover, the feasibility of uplift in a web-based setting, as well as in a campaign setting, depends on the trade-off regarding the costs and benefits involved in the design of the experiment and conventional or initial targeting costs of customers. Additionally, uplift modeling within a web-based setting is mainly feasible and applicable under the circumstance of a controlled web environment, in which visitors are required to register/ to log when visiting the website. A more controlled web-based environment will enable marketers to control treatment exposure and will increase the firm's ability to track individual behavioral characteristics and responses required for uplift modeling. Especially, for e-commerce businesses, which often require visitors to be registered in order to browse within their website, the application of uplift modeling seems to be a valuable tool from a practical perspective. Since some of these businesses already apply some sort of behavioral targeting to capture dynamic behavioral information aiming to personalize certain web-based stimuli such as product recommendations or ads, uplift modeling can be used in order to estimate the potential heterogeneous treatment effects of individual web-visitors characterized by their dynamic attributes. Moreover, this results in the key implication for marketers to not only personalize certain web-based offerings such as product recommendations based on historical behavior and customer interest but also to personalize and evaluate the web-stimuli by means of identifying and targeting customers whose response rate increases due to the marketing intervention. Ultimately, this could support marketers and E-commerce businesses in their challenge to optimize UX design and online conversion since it provides an answer to the question of which course of action (decision regarding a personalized web-stimuli) is best to undertake to achieve incremental conversions.

Limitations and future research

This section will describe the limitations of this study and recommendations for future research. The limitations mainly relate to the practical analysis of the uplift models within both of the described empirical settings. First of all, a limitation of this study applies to the data collected within data set 2 for the purpose of demonstrating and evaluating the feasibility of uplift modeling within the web-based setting. Due to the constraint of available advanced behavioral data collection tools, storage tools and specific characteristics of the website used for the randomized trial (A/B test), a data set was established that was characterized by a low overall response and unbalanced response levels. These characteristics resulted in an impractical dataset for the purpose of modeling uplift and the reliable interpretation of the modeled results. Therefore, this research was not able to empirically test the feasibility and to quantify the potential business value of uplift modeling within a practical web-based setting. Moreover, future research should emphasize the current unknown effect in behavioral response (conversion) under the circumstance of exposing web-visitors to certain personalized web-based-stimuli such as recommendations pop-up-notifications and personalized ads. Therefore, future research should attempt to quantitatively and empirically investigate the business value of uplift for marketers and businesses in a (controlled) web-based environment.

Furthermore, previous direct marketing and uplift literature has shown that individuals can respond differently and even negatively in relation to a certain direct communication (treatment). Therefore, it could be argued that in a web-based setting, there is even a higher chance to discover and model negative uplift or down lift, which is caused by targeting individuals who will change their behavior in a negative way whilst being exposed to a certain web-based treatment. This could especially hold true since customers within an online setting are often confronted with personalized web-based-stimuli, which is often seen as a more intrusive and privacy challenging marketing intervention due to the required usage of online behavioral targeting methods. Consequently, the potential distribution of customers being classified as do-not-disturbs and lost-causes (negative uplift) versus a persuadable (positive uplift) is generally more unknown. Therefore, it is seemingly interesting for future research to empirically investigate the causal treatment effect of these types of (personalized) web-based treatments, which often incorporate some sort of (intrusive) behavioral targeting.

Besides this, another limitation of this study could apply to the practical analysis of data set 1 covering the empirical setting of the application of uplift modeling within campaigns. First of all, a limitation could relate to the selection process of uplift modeling techniques and the usage of several predictive uplift algorithms in a simplistic manner, without for example considering any model or parameter tuning to enhance its predictive performance. This simplistic approach was chosen since the scope and objective of this research was not primarily focused on improving and comparing the predictive accuracy of these models, but rather to demonstrate and to evaluate the feasibility and potential business value of these models. Therefore, future research could perform additional research on the effect of parameter tuning on the predictive performance of several uplift modeling techniques in a practical direct-marketing setting.

Another side note to this research is related to the data quality and usage of the publicly available data set (Hillstrom, 2008). Since the researcher was not aware of the data collection process and origin of the data in data set 1, it may be the case that the quality or interpretation of the data was negatively influenced. Therefore, a recommendation for future research is to re-examine and to compare the results of data set 1, covering the email campaign setting. To summarize, this research served mainly as exploratory research emphasizing the feasibility, application and resulting business value of uplift models within two distinct direct marketing settings such as direct email campaigns and websites. The former setting was to a degree examined by previous studies. Whereas the application and business value of uplift modeling within a web-based setting has previously not yet been investigated. Although the explorative and oriented nature of this research report has highlighted several important challenges, influential factors, and solutions or implications regarding the feasibility of the design and application of uplift models in two distinct direct marketing settings, it was not able to empirically evaluate the application and observe its practical value in the web-based setting. Therefore, it is recommended for future research to further investigate the empirical analysis regarding the feasibility and practical value of uplift modeling within a dynamic web-based setting.

References

Ascarza, E. V. A. (2018). Retention Futility : Targeting High-Risk Customers Might Be Ineffective. *Journal of Marketing Research*, *55*(1), 80–98. https://doi.org/10.1509/jmr.16.0163

Belbahri, M., Gandouet, O., Murua, A., & Nia, V. H. (2019). R package Tools for Uplift Modeling. Retrieved from: https://cran.r-project.org/web/packages/tools4uplift/tools4uplift.pdf

Büchner, A. G., & Mulvenna, M. D. (1998). Discovering internet marketing intelligence through online analytical web usage mining. *ACM Sigmod Record*, *27*(4), 54-61. doi*:https://doi.org/10.1145/306101.306124*

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. In *Source: MIS Quarterly* (Vol. 36, Issue 4). https://doi.org/10.2307/41703503

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard business review, 96(1), 108-116. Retrieved from: https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

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. In *Big Data* (Vol. 6, Issue 1, pp. 13–41). Mary Ann Liebert Inc. https://doi.org/10.1089/big.2017.0104

Ding, A. W., Li, S., & Chatterjee, P. (2015). Learning User Real-Time Intent for Optimal Dynamic Web Page Transformation. *Information Systems Research*, *26*(2), 339–359. https://doi.org/10.1287/isre.2015.0568

Diemert, E., & Renaudin, C. (2018). A Large Scale Benchmark for Uplift Modeling. *Proceedings OfAdKDD & TargetAd (ADKDD'18)*, 603–621. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*, 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gubela, R., Bequé, A., Gebert, F., & Lessmann, S. (2019). Conversion uplift in e-commerce: A systematic benchmark of modeling strategies. *International Journal of Information Managementnternational Journal of Information Technology and Decision Making*, *18*(3), 747–791. https://doi.org/10.1142/S0219622019500172

Gubela, R. M., Lessmann, S., & Jaroszewicz, S. (2020). Response transformation and profit decomposition for revenue uplift modeling. *European Journal of Operational Research*, *283*(2), 647–661. https://doi.org/10.1016/j.ejor.2019.11.030

Guelman, L. (2014). Optimal personalized treatment learning models with insurance applications. Retrieved from: http://diposit.ub.edu/dspace/bitstream/2445/65123/1/Leo%20Guelman_PhD_THESIS.pdf

Guelman. L (2016). Uplift modeling R package. Retrieved from: https://cran.r-project.org/web/packages/uplift/uplift.pdf

Hansotia, B., & Rukstales, B. (2002). Direct marketing for multichannel retailers: Issues, challenges and solutions. *Journal of Database Marketing & Customer Strategy Management*, *9*(3), 259–266. https://doi.org/10.1057/palgrave.jdm.3240007

Hasouneh, A. B. I., & Alqeed, M. A. (2010). Measuring the Effectiveness of E-mail Direct Marketing in Building Customer Relationship. *International Journal of Marketing Studies*, *2*(1). https://doi.org/10.5539/ijms.v2n1p48

Hee, C., Hee Park, C., & Curtis Johnson, S. (2016). Investigating Purchase Conversion by Uncovering Online Visit Patterns Investigating Purchase Conversion by Uncovering Online Visit Patterns Young-Hoon Park. *Marketing Science*, *35*(6), 894–914. https://doi.org/10.1287/mksc.2016.0990

Hernandez, S., Alvarez, P., Fabra, J., & Ezpeleta, J. (2017). Analysis of users' behavior in structured e-commerce websites. *IEEE Access*, *5*, 11941-11958. doi: 10.1109/ACCESS.2017.2707600.

Hillstrom, K. Helping CEOs Understand How Customers Interact With Advertising, Products, Brands, and Channels. Retrieved from: https://blog.minethatdata.com/.

Ja´skowski, M., & Jaroszewicz, S. (2012). *Uplift modeling for clinical trial data*. https://pdfs.semanticscholar.org/6021/f9e1860548e59d7b9bfaca5684bd40f0fbc2.pdf

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *Springer Texts in Statistics An Introduction to Statistical Learning*. Springer Science+Business Media New York. https://doi.org/10.1007/978-1-4614-7138-7

Kane, K., Lo, V. S. Y., & 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. https://doi.org/10.1057/jma.2014.18

Kietzmann, J., & Treen, E. R. (2018). Artificial Intelligence in Advertising: How Marketers Can Leverage Artificial Intelligence Along the Consumer Journey Article. *Journal of Advertising Research*. https://doi.org/10.2501/JAR-2018-035

Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing. *California Management Review*, *61*(4), 135–155. https://doi.org/10.1177/0008125619859317

Kwan, I. S. Y., Fong, J., & Wong, H. K. (2005). An e-customer behavior model with online analytical mining for internet marketing planning. *Decision Support Systems*, *1*(41), 189–204. https://doi.org/10.1016/j.dss.2004.11.012

Kondareddy, S. P., Agrawal, S., & Shekhar, S. (2016). Incremental response modeling based on segmentation approach using uplift decision trees. In *Industrial Conference on Data Mining* (pp. 54-63). Springer, Cham.

Lai, Y.-T., Wang, K., Ling, D., Shi, H., & Zhang, J. (2006). Direct Marketing When There Are Voluntary Buyers. *Direct Marketing When There Are Voluntary Buyers*, 1–6. https://doi.org/10.1109/ICDM.2006.54

Lo, V. S. Y. (2002). The true lift model: a novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter*, *4*(2), 78–86. https://doi.org/10.1145/772862.772872

Mishra, N., & Silakari, D. (2012). Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges. *International Journal of Computer Science and Information Technologies*, *3*(3), 4434–4438. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.301.7387&rep=rep1&type=pdf

Mokryn, O., Bogina, V., & Kuflik, T. (2019). Will this session end with a purchase? Inferring current purchase intent of anonymous visitors. *Electronic Commerce Research and Applications*, *34*. https://doi.org/10.1016/j.elerap.2019.100836

Olaya, D., Kristof Coussement, ·, & Verbeke, · Wouter. (2020). A survey and benchmarking study of multitreatment uplift modeling. *Data Mining and Knowledge Discovery*, *34*, 273–308. https://doi.org/10.1007/s10618-019-00670-y

Pant, B., Pant, K., & Pardasani, K. R. (2009). Decision Tree Classifier for Classification of Plant and Animal Micro RNA's. In *CCIS* (Vol. 51). https://doi.org/10.1007/978-3-642-04962-0_51

Radcliffe, N. J., & Surry, P. D. (2011). *Real-World Uplift Modelling with Significance-Based Uplift Trees*.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Ruda, K., & Jaroszewicz, S. (2018). Linear regression for uplift modeling. *Data Mining and Knowledge Discovery*, *32*(5), 1275–1305. https://doi.org/10.1007/s10618-018-0576-8

Rzepakowski, P., & Jaroszewicz, S. (2012a). Decision trees for uplift modeling with single and multiple treatments. *Knowl Inf Syst*, *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*, *2*, 43–50.

Rzepakowski, P., & Jaroszewicz, S. (2012c). 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

Shaar, A., Abdessalem, T., & Segard, O. (2016). Pessimistic Uplift Modeling. *In 22nd SIGKDD Conference on Knowledge Discovery and Data Mining (ACM SIGKDD)*, 9.

Sołtys, M., & Jaroszewicz, S. (2015). Ensemble methods for uplift modeling. *Data Mining and Knowledge Discovery*, *29*(6), 1531–1559. https://doi.org/10.1007/s10618-014-0383-9

Vafeiadis, T., Diamantaras, K., Sarigiannidis, G., & Chatzisavvas, Kc. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1–9. https://doi.org/10.1016/j.simpat.2015.03.003

Van Den Poel, D., & Buckinx, W. (2005). Interfaces with Other Disciplines Predicting online-purchasing behaviour. *European Journal of Operational Research*, *166*(2), 557–575. https://doi.org/10.1016/j.ejor.2004.04.022

Zaniewicz, Ł., & Szymon Jaroszewicz, ·. (2017). L p-Support vector machines for uplift modeling. *Knowledge and Information Systems*, *53*, 269–296. https://doi.org/10.1007/s10115-017-1040-6

Appendix

Predictive Analytics & Machine learning
Predictive analytics is a branch of data mining aiming to predict a certain occurrence or probability based on existing data. It incorporates statistical and machine learning techniques in order to make predictions about future outcomes, probabilities, or events (Mishra & Silakari, 2012). Predictive analytics can be used in many different business fields such as finance, marketing and logistics. For instance, the application varies from predicting failure of engines based on big data stream from sensors, to predicting customers next moves based on their behavior or historical characteristics (Gandomi & Haider, 2015). In relation to the latter application Chen et al. (2012) and Lo (2002) have shown the increased interest and application of predictive analytics or models in the field of marketing. Predictive models actively are used focusing on cross-selling, campaign management, customer acquisition and retention or churn management.

As stated above, predictive analytics is associated with machine learning and statistical learning techniques. In general statistical and machine learning problems can be distinguished into two categories: unsupervised and supervised. The first technique involves the establishment of a model for predicting or estimating and output based on one or more inputs (James et al., 2013). In this case data is modeled from training data aiming to discover patterns within the data used to predict a class label or value based on a set of parameters (Mishra & Silakari, 2012). While unsupervised machine learning aims to learn from relationships and structures within data due to the fact these models are based on inputs but not on outputs (James et al., 2013). Additionally, within both learning methods, different techniques can be applied in relation to solving two types of problems: classification problems and regressions problems (Mishra & Silakari, 2012). The type of techniques used is associated with the type of data and the type of output variable or the target variable that is being predicted.

There are many different algorithms available for the construction of predictive models. Algorithms vary enormously in their structure, function, and parameters. For classification problems, different classification techniques can be used such as Decision trees, Naïve Bayes, Support Vector Machines and Random Forest (Vafeiadis et al., 2015). These techniques constitute one or several machine learning algorithms, which can be used for model construction. Decision trees can be applied for both regression and classification problems (James et al., 2013). Decision trees represent a flow-chart-like tree structure. Within this tree structure each internal node indicates a test on an attribute, each branch represents the outcome of the test and each leaf node holds the class label (Pant et al., 2009). The growing process of a tree starts with the whole population at the root of the tree and is followed up by the evaluation of a large number of candidate splits using a selected quality measure (Radcliffe & Surry, 2011). The tree is further grown by considering a number of splits for each predictor variable. The best split is then chosen for each and this process will be repeated until a termination criteria is met or until the tree is fully grown (Radcliffe & Surry, 2011). According to James et al. (2013), tree-based methods involve stratifying or segmenting the predictor space into a number of smaller regions. As a result, tree methods are commonly used for their simplicity, intuitive power and ability to divide a large collection of records into smaller sets of records. However, the accuracy and predictive performance of decision trees is debatable, especially with regards to complex and non-linear relationships between attributes (Vafeiadis et al., 2015).

An important aspect of the creation of any (supervised) machine learning model, such as a decision tree, consists of variable section. This also holds true for uplift models. Variable selection overcomes the potential problem of overfitting and avoids the correlation between predictor variables (Radcliffe & Surry, 2011). According to Guelman (2014) overfitting can be defined as a model that overemphasizes the learning patterns caused by existing noise in the data which may not recur in future or validation samples. This could decrease the predictive of accuracy of the model on newly seen test data. Overfitting has an increased chance of happening in circumstances of a large predictor space (James et al., 2013). Additionally, variable section is important in order to reduce the complexity and multicollinearity of the model and increases the stability of the model (Radcliffe & Surry, 2011). Therefore, depending on the type of model that is being built, removing variables can result in an improved model quality with an increased predictive power. For instance, the creation of a single decision tree has been described as a greedy and sensitive model in terms of stability, when a variable gets removed. For instance, removing a variable that was used for splitting the tree at a certain level, the subsequent levels of the tree could change resulting in a tree with more levels and a higher model quality. Besides variable selection, other tools can be applied to increase the model quality and predictive performance. Several researchers state that predictive accuracy can be increased through the usage of several tree ensemble methods instead of creating a single tree based on the pre-specified selection of variables (James et al., 2013; Sołtys & Jaroszewicz, 2015).

	Variable names	Definition
Treatment & target variable	**Segment**	E-mail campaign the customer received
	Visited	Binary indicator whether customer visited website in the following two weeks
	Conversion	Binary indicator whether customer purchased clothing in the following two weeks
Customer attributes	**Recency**	Months since last purchase
	History_segment	Categorization of dollars spent in the past years
	History	Actual value of dollars spent
	Mens	Binary indicator if customer purchased men's clothing in the past year = 1
	Womens	Binary indicator if customer purchased women's clothing in the past year = 1
	Zip_code	Classification of zipcode as urban, suburban or rural
	Newbie	Binary indicator, 1 = new customer
	Channel	Description of channel used to buy in the past

Table 11: Marketing target variables & customer attributes Hillstrom email campaign. Retrieved from: https://blog.mindethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html.

	Variables randomized trial	Definition
Treatment & target variables	**Treated webdesign (CTA design)**	Binary variable indicating whether web-visitor was exposed to the original (control) or modified CTA button design
	Conversion CTA button	Binary variable indicating whether web-visitor clicked on CTA button & visited the next page of interest.
Dynamic attributes	**Referrer URL**	Categorization of URL page visitor was before start of data collection (google, within web system of UT, no referrer available, website partner University)
	Device	Categorization of device used
	Browser	Categorization of browser used
	Operating system	Categorization of operating system used
	Country	Categorization of country
	EU	Dichotomy variable based on whether country is within EU = 1
	Action count	Continuous variable indicting amount of time clicked (excluding rage or spam clicks)
	Number of page visited	Continuous variable indicating number of page visited during first unique session
	Visiting duration	Continuous variable indicating duration of visit during first unique session

Table 12: Target & predictor variables data set 2 web-based setting (University of Twente)

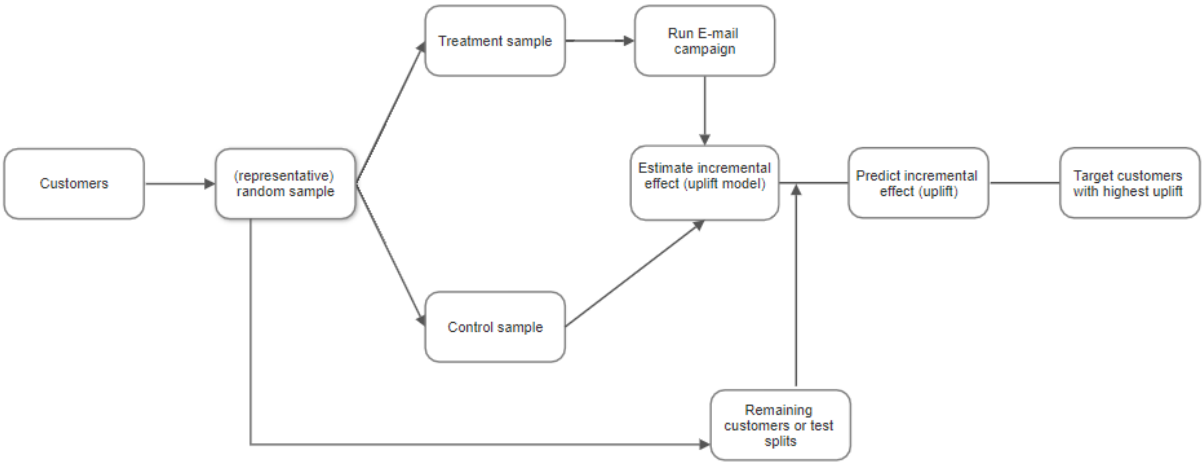


Figure 4: Uplift modeling procedure within the web-based setting

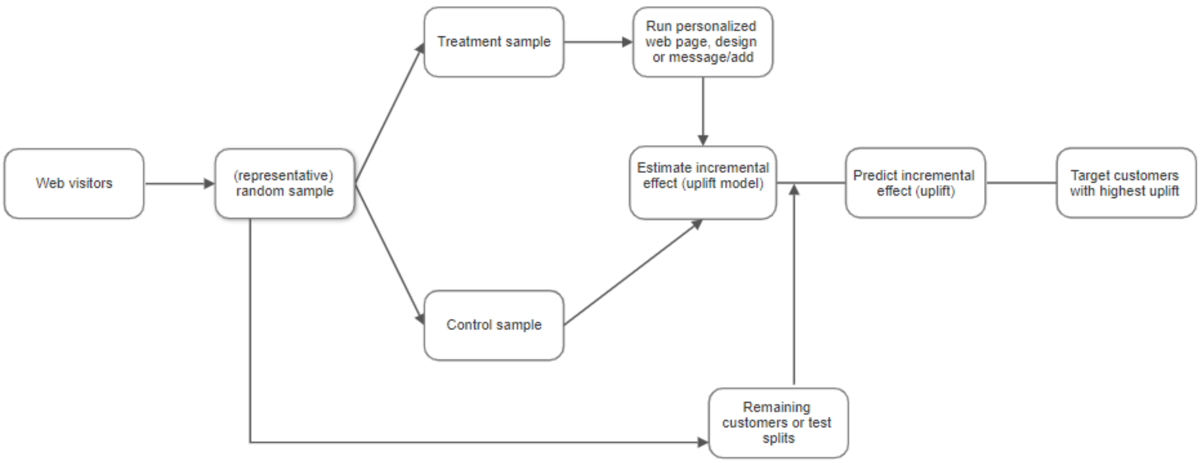


Figure 5: Uplift modeling procedure for the web-based setting

	Observed data Google Analytics & Optimize			Observed data Hotjar		
	Experiment sessions	Experiment conversions (response)		Experiment sessions	Experiment conversions (response)	
Control	3,536	125	3.54%	556	7	1.26%
Treated	3,399	103	3.03%	549	7	1.27%
Total	6,935	228	3.29%	1105	14	1.27%

Table 13: Comparison of observation and percentage response (conversion) across different advanced behavioral data collection tools

Variable name	T = 0	T= 1	Adj.diff	Z	P
recency	5.763	5.750	-0.013	-0.318	0.750
history_segment1) $0 - $100	0.360	0.358	-0.002	-0.396	0.692
history_segment2) $100 - $200	0.226	0.218	-0.008	-1.656	0.098
history_segment3) $200 - $350	0.189	0.193	0.004	0.880	0.379
history_segment4) $350 - $500	0.100	0.102	0.002	0.441	0.659
history_segment5) $500 - $750	0.076	0.080	0.004	1.232	0.218
history_segment6) $750 - $1,000	0.030	0.028	-0.002	-0.914	0.361
history_segment7) $1,000 +	0.018	0.021	0.003	1.601	0.109
history	238.829	244.638	5.809	1.974	0.048
mens0	0.449	0.456	0.007	1.168	0.243
mens1	0.551	0.544	-0.007	-1.168	0.243
womens0	0.451	0.445	-0.006	-1.089	0.276
womens1	0.549	0.555	0.006	1.089	0.276
zip_codeRural	0.148	0.149	0.000	0.114	0.910
zip_codeSurburban	0.448	0.450	0.002	0.362	0.718
zip_codeUrban	0.404	0.401	-0.003	-0.449	0.653
newbie0	0.497	0.496	0.000	-0.079	0.937
newbie1	0.503	0.504	0.000	0.079	0.937
channelPhone	0.437	0.439	0.002	0.388	0.698
channelWeb	0.441	0.437	-0.004	-0.665	0.506
channelMultichannel	0.122	0.124	0.002	0.418	0.676

Table: test of balance among the covariates

Coefficients treated group glm object Two-mode est.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.9226953	0.1292127	-14.880	< 2e-16	***
recency	-0.0549718	0.0081383	-6.755	1.43e-11	***
history_segment2) $100 - $200	-0.0505907	0.0874848	-0.578	0.563	
history_segment3) $200 - $350	0.1579997	0.1186288	1.332	0.183	
history_segment4) $350 - $500	0.1996816	0.1747938	1.142	0.253	
history_segment5) $500 - $750	0.3638075	0.2566475	1.418	0.156	
history_segment6) $750 - $1,000	0.6844314	0.3596078	1.903	0.057	.
history_segment7) $1,000 +	0.5933779	0.5364184	1.106	0.269	
history	0.0001870	0.0004043	0.463	0.644	
mens1	0.4964056	0.0891250	5.570	2.55e-08	***
womens1	0.4468516	0.0887559	5.035	4.79e-07	***
zip_codeSurburban	-0.4684783	0.0729232	-6.424	1.33e-10	***
zip_codeUrban	-0.5559896	0.0749290	-7.420	1.17e-13	***
newbie1	-0.8173787	0.0663431	-12.320	< 2e-16	***
channelWeb	0.3624854	0.0590385	6.140	8.26e-10	***
channelMultichannel	0.1178394	0.0916848	1.285	0.199	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Coefficients control group glm object Two-mode est.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.877e+00	1.139e-01	-16.489	< 2e-16	***
recency	-3.771e-02	6.903e-03	-5.463	4.69e-08	***
history_segment2) $100 - $200	-6.517e-02	7.428e-02	-0.877	0.3803	
history_segment3) $200 - $350	-3.351e-04	1.031e-01	-0.003	0.9974	
history_segment4) $350 - $500	2.072e-01	1.541e-01	1.345	0.1787	
history_segment5) $500 - $750	4.313e-01	2.213e-01	1.949	0.0513	.
history_segment6) $750 - $1,000	4.681e-01	3.161e-01	1.481	0.1386	
history_segment7) $1,000 +	5.546e-01	4.789e-01	1.158	0.2468	
history	2.742e-05	3.556e-04	0.077	0.9385	
mens1	3.005e-01	7.671e-02	3.917	8.96e-05	***
womens1	8.010e-01	7.959e-02	10.064	< 2e-16	***
zip_codeSurburban	-2.724e-01	6.541e-02	-4.164	3.12e-05	***
zip_codeUrban	-2.891e-01	6.662e-02	-4.340	1.43e-05	***
newbie1	-5.266e-01	5.369e-02	-9.808	< 2e-16	***
channelWeb	2.826e-01	4.973e-02	5.683	1.32e-08	***
channelMultichannel	2.302e-02	7.982e-02	0.288	0.7730	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Cum_per	T_Y1	T_Total_N	C_Y1	C_Total_N	Incremental_Y1	Inc_uplift	Uplift
0.1	120	619	84	667	42.04498	0.6621256	0.06792403
0.2	234	1222	166	1339	82.50485	1.2992890	0.06703092
0.3	338	1850	223	1992	130.89659	2.0613636	0.07831566
0.4	443	2480	294	2643	167.13167	2.6319948	0.05760369
0.5	558	3143	346	3261	224.52009	3.5357494	0.08931160
0.6	648	3787	407	3897	252.48832	3.9761941	0.04383960
0.7	691	4436	438	4529	261.99404	4.1258904	0.01720515
0.8	754	5061	492	5185	273.76625	4.3112795	0.01848293
0.9	845	5703	568	5824	288.80082	4.5480445	0.02280871
1.0	946	6350	678	6458	279.33849	4.3990314	-0.0173964

Table: Performance table two-model estimator technique for modeling uplift

Coefficients glm object RVTU method

```
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -0.0268413  0.0622407  -0.431  0.66629
recency                          0.0001248  0.0034141   0.037  0.97083
history_segment2) $100 - $200    0.0394712  0.0367094   1.075  0.28227
history_segment3) $200 - $350   -0.0078305  0.0556326  -0.141  0.88806
history_segment4) $350 - $500    0.0710346  0.0859468   0.826  0.40852
history_segment5) $500 - $750    0.1274430  0.1230930   1.035  0.30051
history_segment6) $750 - $1,000  0.1975907  0.1788049   1.105  0.26913
history_segment7) $1,000 +       0.2188401  0.2686639   0.815  0.41533
history                         -0.0002535  0.0002029  -1.250  0.21141
mens1                            0.0287494  0.0443459   0.648  0.51679
womens1                          0.1356584  0.0443909   3.056  0.00224 **
zip_codeSurburban                0.0251551  0.0346961   0.725  0.46845
zip_codeUrban                    0.0444375  0.0351955   1.263  0.20674
newbie1                          0.0174476  0.0254169   0.686  0.49243
channelWeb                       0.0159005  0.0247452   0.643  0.52050
channelMultichannel             -0.0027408  0.0418521  -0.065  0.94779
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

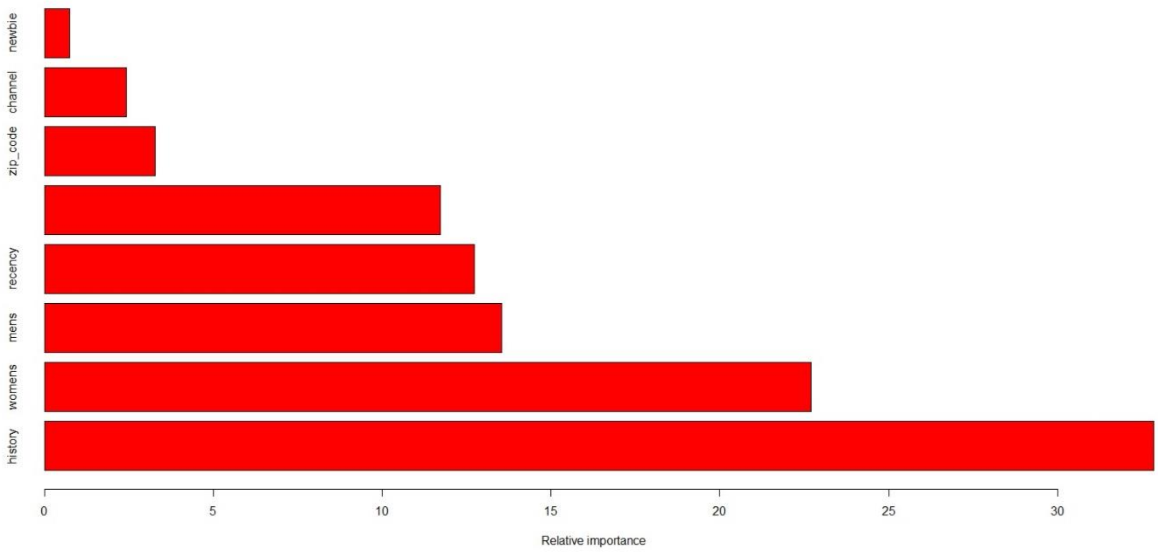


Figure: Variable importance uplift random forest model

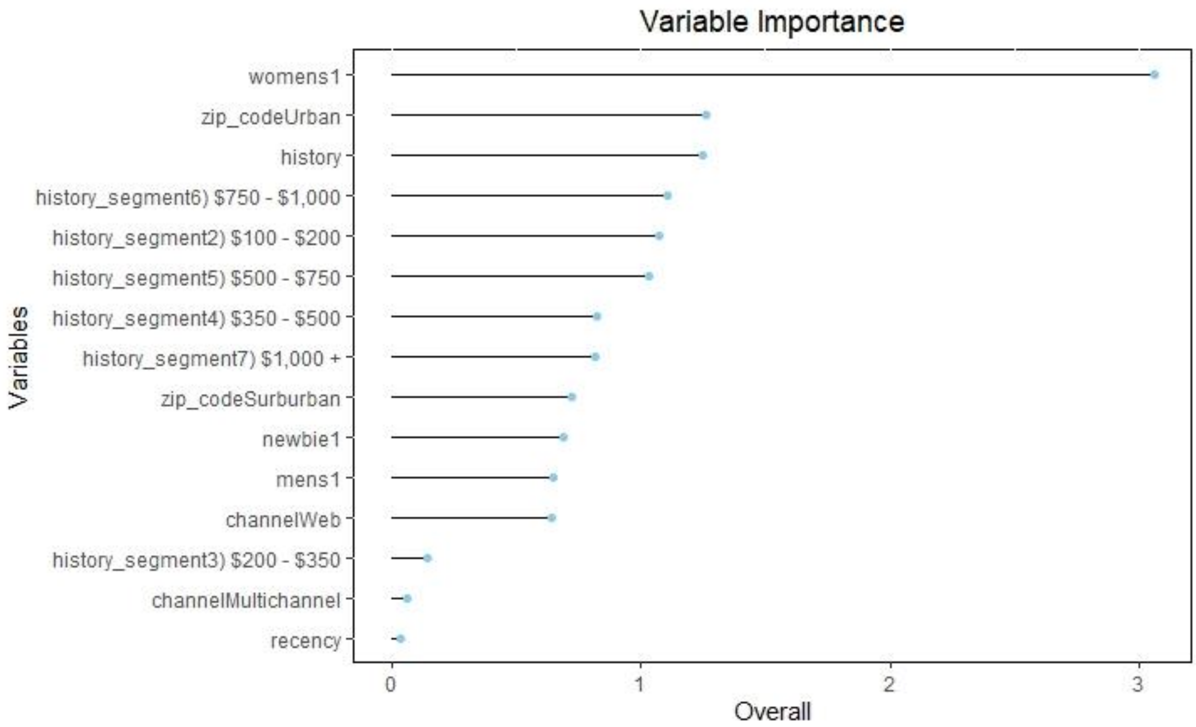


Figure: Variable importance logistic model RVTU technique

Variable	(dummy) categories		1	2	3	4	5	6	7	8	9	10	All
			Groups/Deciles/fractions of customers from data set 1										
		n	1281	1281	1281	1280	1281	1281	1282	1279	1281	1281	12808
	RTVTU_PTE_pred	Avg.	0.543	0.538	0.535	0.530	0.524	0.516	0.512	0.508	0.503	0.492	0.520
History	history	Avg.	239.645	213.098	214.559	264.326	313.732	201.529	168.868	175.484	248.665	376.222	241.616
history_segment	1) $0 - $100	Pctn.	27.40	44.11	51.99	39.77	10.30	24.82	45.63	57.54	44.34	6.79	35.27
	2) $100 - $200	Pctn.	46.21	31.30	18.74	12.73	10.62	38.49	34.56	21.34	13.51	0.78	22.83
	3) $200 - $350	Pctn.	0.00	0.86	4.61	15.31	51.52	27.56	8.74	4.30	11.32	69.48	19.37
	4) $350 - $500	Pctn.	3.28	7.49	9.13	18.28	18.66	2.73	2.26	7.66	18.66	13.35	10.15
	5) $500 - $750	Pctn.	14.44	11.55	11.01	8.83	4.61	3.75	6.01	5.24	7.96	2.65	7.60

	6) $750 - $1,000	Pctn.	8.27	3.75	3.20	2.50	0.94	1.33	1.95	2.74	1.95	0.86	2.75
	7) $1,000 +	Pctn.	0.39	0.94	1.33	2.58	3.36	1.33	0.86	1.17	2.26	6.09	2.03
zip_code	Rural	Pctn.	0.39	2.89	11.71	26.95	22.64	8.12	5.62	13.76	28.88	26.54	14.75
	Surburban	Pctn.	22.17	49.96	62.14	51.56	37.55	30.99	47.89	58.80	49.96	45.67	45.67
	Urban	Pctn.	77.44	47.15	26.15	21.48	39.81	60.89	46.49	27.44	21.16	27.79	39.58
womens	0	Pctn.	0.00	0.00	0.00	0.00	9.45	68.54	90.33	94.76	98.20	97.42	45.87
	1	Pctn.	100.00	100.00	100.00	100.00	90.55	31.46	9.67	5.24	1.80	2.58	54.13

Table 14: Summary of statistics and distribution of 4 variables of data set 1, used for profiling characteristics of subjects who were grouped and ranked from highest to lowews uplift