UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DYNAMIC UPLIFT MODELING

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

GOWTHAM TALLURU
Norman, Oklahoma
2017

DYNAMIC UPLIFT MODELING

A THESIS APPROVED FOR THE
GALLOGLY COLLEGE OF ENGINEERING

BY

Dr. Charles Nicholson, Chair

Dr. Randa Shehab

Dr. Ziho Kang

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Charles Nicholson for great support and suggestions during my Masters. I would like to thank him for providing an opportunity to work as a teaching assistant for his Spreadsheet Based Decision Support Systems class. I greatly appreciate Dr. Nicholson's friendly and always welcoming nature for any kind of technical discussion. His classes Intelligent Data Analytics and Advanced Analytics and Metaheuristics enhanced my analytics knowledge and greatly helped me in job interviews. I would like to thank Dr. Randa Shehab and Dr. Ziho Kang for their valuable suggestions to my research. I would like to thank Nicola for all kinds of administrative support and reminding me about each and every deadline. I would also like to acknowledge the support of Cheryl and Kristi for supporting my Teaching Assistant duties.

I express my deepest gratitude to my parents for their love and efforts to make me a successful individual. Last but not the least, I would like to thank my friends Aditya, Eiman, Japneet, Pratiksha, Pritesh, Saket, Saurabh, Shipra for making my two years of Masters memorable. I would also like to thank my research group friends Alex, Nicole, Paul, Pauline, Sai and Vera for their valuable suggestions.

Table of Contents

List of Tables

List of Figures

Abstract

In this thesis, a new approach to Uplift modeling which considers time dependent behavior of the customers is analyzed. Uplift modeling attempts to measure the impact of a treatment on an entity in a controlled experiment. While the overall incremental effect can be measured indirectly (i.e., the average performance of a treatment group over a statistically equivalent control group), the entity-specific performance cannot be determined. It has applications in business, insurance, banking, personalized medicine, and other fields. Direct marketing, a multi-billion dollar field in the US alone, is a key area in which uplift modeling is studied and can have a significant financial impact. In direct marketing, the entities studied are customers and the treatments are various direct-to-consumer promotions delivered through mail, email, social media, etc. Simulated customer and campaign datasets which reflects the naturally observed trends are used to analyze the effectiveness of various modelling approaches.

Research on Uplift modeling specific to above mentioned fields started in the beginning of 21^{st} century even though the idea of Uplift is present before that. Researchers have introduced a wide range of uplift modeling approaches. These approaches broadly include two model approach, additive model approach and unified modeling approach. But all of the research until now has considered this as a static problem, modeled at a single instance of time.

The method introduced in this work considers modeling uplift in a dynamic environment and simulates the periodic purchasing behavior of the customer. In contrast to static uplift models, the uplift in the purchase probability of the customers considered in this problem is dependent on time as well as customer's previous purchases and offers

received. In addition, the model will not have direct access to all the parameters effecting customer actions, but it has to learn them with time. The effectiveness of various modeling approaches, two model approach, additive model approach and unified modeling approach is analyzed in this work for dynamic uplift modeling. Appropriate modifications are made to these methods for adapting them to the longitudinal paradigm. The results obtained from these models are compared to the model with zero treatment and random treatment.

This study demonstrates significant potential for both researches and retail companies for thinking about the problem of uplift longitudinally. Retail companies can use the methodology used for data generation for matching the customer purchase data available with them. The model built from there can be used both to design direct marketing campaigns as well as to predict future purchases.

1. Introduction

1.1 Purpose and Significance of the Study

The United States is the largest advertising market in the world with an overall expenditure of more than 190 Billion USD per year (Zenith 2017). Besides United States other countries having significantly large advertising industry include China, Japan, United Kingdom and Germany (expenditure more than 20 Billion USD per year) accounting for a global advertisement expenditure of around 559 Billion USD (Zenith 2017). **Figure 1** shows the scale of advertising expenditure made by largest ad markets in the year of 2016. It can be observed that the amount of money spent on advertising in United States itself is higher that the money spent altogether in next five largest ad markets. This explains the importance and quantity of beneficiaries of research related to marketing industry.

The major media that are contributing to the advertising industry are television, desktop internet, mobile internet, newspapers, magazines, radio, outdoor and cinema. For many years television has been the major advertising media representing more than 35% of advertising expenditure globally. However, with the increase in the presence of high speed internet, smartphones and applications like Facebook, Snapchat and Instagram, mobile and desktop advertising is cutting into the share of television advertisement. It has been estimated that by 2019 internet advertisement will surpass television advertisement making mobile advertisement the major player in advertisement.

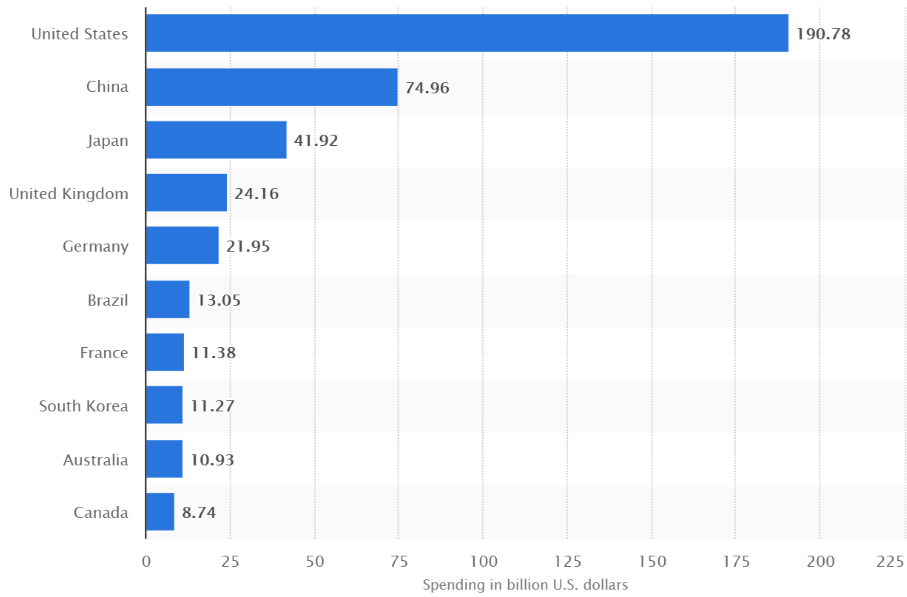


Figure 1: World's largest ad markets advertising expenditure in 2016 (Billion USD) (Zenith 2016)

This change in trend will significantly increase the importance of one of the primary marketing strategies, that is direct marketing. Direct marketing is a branch of marketing in which ads and offers are strategically directed to individual customers via media like email, mail, Facebook pages, and Instagram etc. Retail companies cannot send these ads/offers to all the customers in their database at all times considering the cost and other potential negative effects associated with over advertisement. This behooves the retail companies to have a strategy or model to efficiently select the customers at each marketing phase. Retail companies generally use customer demographic data, previous browsing and purchasing history to select customers for direct marketing.

One of the early techniques used for customer selection in direct marketing campaigns is known as response modeling. A response model predicts whether a customer will respond to an offer if provided one. One of the drawbacks with this model is it will not account for lost revenue due to sending offers to customers who will buy even without an offer. For example, a retail company will lose revenue if it sends a promotional discount to a customer who will buy the product even without offer. If the actual cost of product is $700 (assume a laptop), and the offer is a $100 discount, the loss in revenue accounts for 15% of the revenue.

1.2 Introduction to Uplift modeling

A class of marketing models that address this drawback is known as Uplift modeling. Uplift models attempt to identify customers who will buy *if and only if* they receive a promotional offer. **Figure 2** shows the flowchart depicting the work flow of an Uplift model. Customers present in the database of the company are assigned to two groups: a control group and a treatment group. The assignment of control or treatment to a customer is often made randomly such that there is no statistical difference between the groups with respect to customer attributes. While other approaches for assigning treatment and control groups exist, the primary idea is that through statistical analysis the company can isolate the impact of a given marketing effort. Customers in the treatment group will receive an offer while customers in the control group will not receive any offer. The response of the customers in both the groups will be recorded and used to estimate change in purchase probability of any new customer due to receiving an offer.

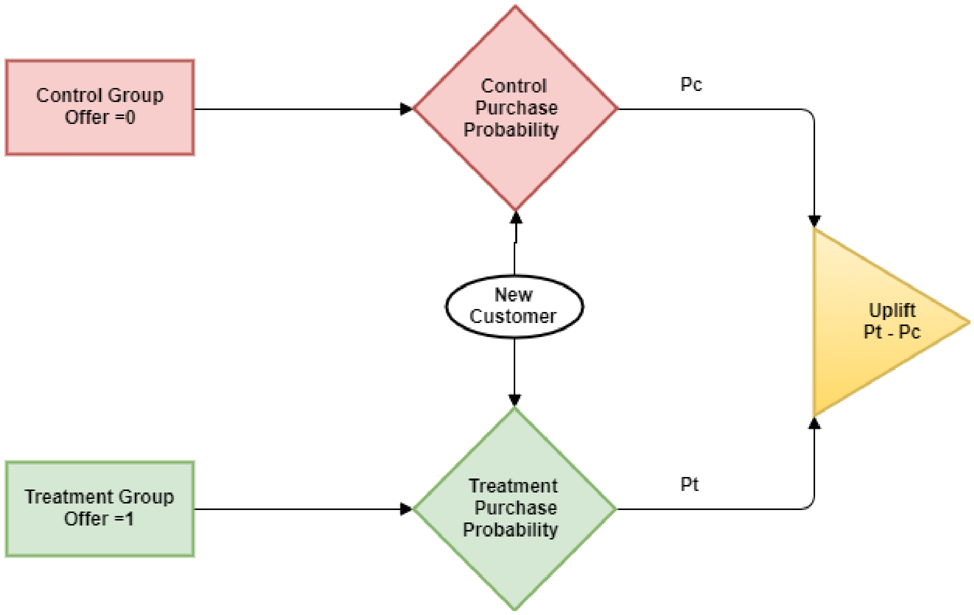


Figure 2: Flowchart of generic uplift model

Customers can be classified into four groups based on their response to direct marketing (Chickering and Heckerman 2000). **Figure 3** shows these four groups which are as follows

i. Customers who always buy the product irrespective of the offer (Always buy).

ii. Customers who never buy the product irrespective of the offer (Never buy).

iii. Customers who buy the product if and only if they get an offer (Persuadables).

iv. Customers who will stop buying product if they get an offer (Anti- Persuadables).

		Purchase without offer	
		1	0
Purchase with offer	1	Always Buy	Persuadables
	0	Never Buy	Anti Persuadables

Figure 3: Classification of Customers based on their response to direct marketing

The company will lose the offer amount if they send an offer to *always buying* customer whereas they will lose advertising cost and effort if they send it to a *never buying* customer. The consequences of sending offers to *anti-persuadable* customers are even worse. Companies will lose prospective purchases by sending offers to these customers. The objective of any uplift model is to identify as many *persuadable* customers as possible in the campaign group and to avoid customers belonging to other three groups.

A basic method of building an uplift model to identify persuadable customers based on their static attributes is as follows.

i. Customers with a similar set of attributes at a single point in time (e.g., income, age, city, gender, etc.) are separated into control and treatment groups and a specific offer will be sent to the customers in the treatment group. Let A_s denote this vector of “static” customer attributes.

ii. Probability of purchase in the control group $P_c(A_s)$ and probability of purchase in treatment group $P_t(A_s)$ are calculated as a function of customer static attributes.

iii. Uplift is the computed as the difference between $P_t(A_s)$ and $P_c(A_s)$. It is the increase in the probability of purchase of a customer because of receiving an offer.

$P_t(A_s)$, $P_c(A_s)$ and Uplift can be calculated using various predictive modeling techniques on purchase data collected from control and treatment group customers. Early Uplift models in the literature concentrated on explaining the importance of modeling uplift rather than response rate of marketing campaigns. Recent Uplift models concentrated on using various predictive modeling techniques like decision trees, logistic regression, Artificial neural networks, SVMs, Ensemble models to identify the most appropriate customer characteristics A_s in order to maximize the difference between $P_t(A_s)$ and $P_c(A_s)$. But this earlier research has not concentrated on capturing dynamically changing behavior of the customers due to receiving repeated offers and making multiple purchases. In this research, the longitudinal nature of the uplift problem is addressed instead of addressing it as a static instance. The time dimension of repeated offers to the same customer reveals new issues.

Example:

i. Customers who frequently receive offers from a firm may not get too excited about a new offer.

ii. Customers who have received significant offers in the past won't buy a new product without an offer.

iii. Customers who buy a product once in 4 years may buy it earlier if they find a good offer on it.

In this thesis I introduce the concept of *Dynamic uplift modeling* which emphasizes the time-dependent behavior of customers in longitudinal marketing efforts. Most of the research till now was about estimating uplift at a single instance and time-dependent behavior of the customer is not yet addressed. Dynamic uplift modeling however models the uplift probability as a function of both static as well as dynamic attributes of customers. Dynamic attributes of customers, in turn, will be a function of static attributes and previous marketing actions. Let A_d denote the set of dynamic customer attributes. The general framework of dynamic uplift modeling follows:

i. Probability of purchase in control group = $P_c(A_s,\ A_d)$

ii. Probability of purchase in treatment group = $P_t(A_s, A_d)$

Where $A_d = f(A_s\ , g(I_p, I_o))$

iii. Dynamic Uplift (Ψ) = $Pt(A_s, A_d) - Pc(A_s, A_d)$

Dynamic uplift captures the changing trend in uplift probability because of the purchases customers made and marketing actions customers are exposed to over time.

1.3 Objectives

The objectives and work flow of this study are as follows:

1. Generating artificial data representing static attributes of customers.
    a. Defining functions to calculate the static probability of purchase with an offer and without an offer for customers.
    b. Defining functions for calculating dynamic attributes of customers based on purchase cycle time, previous purchases and offers received.
2. Running the model through a fixed time period and generating campaign data based on dynamic control and treatment probabilities.
3. Building predictive models based on static attributes of the customers to predict which customers should be targeted at what time to optimize the number of successful offers. The model will not have access to dynamic attributes of the customers and it has to learn it from the data.

1.4 Outlay of the thesis

This thesis is divided into five chapters. Chapter 2 discusses previous works related to this study and gives an introduction to current work. Previous Uplift models developed by Radcliffe and Surry (1999), Chickering and Heckerman (2000), Hansotia and Rukstales (2002), Radcliffe (2007), Radcliffe and Surry (2011), Rzepakowski and Jaroszewicz (2011), Guelman et al. (2012), Zaniewicz and Jaroszewicz (2013), Soltys et al. (2015), Kondareddy et al (2016), Zhao et al. (2017) are discussed in first section of this chapter. Various modeling approaches followed by these authors were classified into three broad categories and their advantages and disadvantages are discussed in broader level. Also at the end current approach followed in this research is explained.

Chapter 3 discusses methodology followed for generating artificial data. This chapter explains the procedures followed in generating static customer attributes data frame, control and treatment weights, control and treatment odds and probabilities. Later methodology used for incorporating dynamic effects is discussed. At the end methodology used for running the model with time is discussed.

Chapter 4 discusses uplift modeling in a dynamic environment. It also discusses modifications required for conventional uplift models and evaluation metrics to model uplift in an environment changing with time. Chapter 5 summarizes the work done in this thesis followed by conclusions and recommendations for future work.

2. Review of Literature

Uplift Modeling is a Causal Inference as well as a Machine Learning problem (Soltys et al 2015). It is an inference problem because uplift has to be estimated between two instances that are mutually exclusive. Same customer cannot simultaneously receive and not receive an offer. So, Uplift has to be estimated from similar customers in control group and treatment group and this is done over a group of customers rather than a single customer. Uplift modeling is also a machine learning problem because it uses various predictive modeling and cross-validation techniques to train models on data collected from marketing campaigns.

Researchers used various combinations of inference, predictive modeling and performance metrics to produce better uplift models for various scenarios. In general, the approaches can be classified into two classes. One is two model approach where two different models are made for control group and treatment group. Second is a unified model for both control and treatment groups which uses modifications to predictive modeling techniques to predict uplift. There is another approach called additive model which looks similar to unified model but behaves like a two model approach. In this chapter, a few of previous works carried out by the different researchers on uplift modeling will be reviewed.

2.1 Previous Uplift Models

As per Rubin (1974) in a framework of N customers $Y_i(1)$ is person i's outcome when he receives an offer and $Y_i(0)$ is the same persons outcome when he doesn't receive an offer. The causal effect, τ_i on a customer is denoted by

$$\tau_i = Y_i(1) - Y_i(0) \quad (2\text{-}1)$$

But this causal effect cannot be measured on a single customer because he/she cannot receive and not receive an offer simultaneously. So, this causal effect is measured as Conditional Average Treatment Effect (CATE) over a subgroup of the population.

$$\text{CATE: } \tau_i(X_i) = E[Y_i(1)|X_i] - E[Y_i(0)|X_i] \tag{2-2}$$

Where X_i is the feature vector of the customer.

Radcliffe and Surry (1999) proposed one of the early uplift models which explained the importance of uplift modeling in contrast to a model predicting response for a treatment. Two decision trees were used to model the response rate of customers in control group and treatment group based on their features. Then uplift or incremental response is calculated by taking the difference of these two models. The inferences made from this two-model method is compared with the inferences made from the model predicting the response of a treatment group. The results explained that the customer group having highest response may not be a great candidate for direct marketing.

Chickering and Heckerman (2000) explained uplift in terms of expected revenue rather than in terms of response rate. If a population contains N_a Always buying customers, N_p Persuadable customers, $N_{\bar{p}}$ Anti Persuadables and $N_{\bar{a}}$ Never buying customers then the expected revenue from a customer who received an offer is

$$-c + \frac{N_a + N_p}{N} * r_s \tag{2-3}$$

Similarly expected revenue from a customer who didn’t receive an offer is

$$\frac{N_a + N_{\bar{a}}}{N} * r_u \tag{2-4}$$

Where c is the cost of the advertisement and r_u and r_s are actual revenue and discounted revenue from a purchase. Therefore, lift in profit can be defined as

$$\tau_{revenue} = -c + \frac{N_a + N_p}{N} * r_s - \frac{N_a + N_{\bar{a}}}{N} * r_u \tag{2-5}$$

Here $\frac{N_a+N_p}{N}$ is $E[Y_i(1)|X_i]$ and $\frac{N_a+N_{\bar{a}}}{N}$ is $E[Y_i(0)|X_i]$. These two probabilities are estimated from a single decision tree by forcing the final split in the tree on marketing action (0,1). Even though it uses single decision tree for prediction, its approach is similar to two model approach. The draw back with this model is that higher level splits in the decision tree were not customized to elicit the difference between expected revenue from control and treatment groups.

Hansotia and Rukstales (2002) also explained a similar uplift model. This model is an additive uplift model in which logistic regression was used to predict control and treatment response probabilities. A regression model is used on top of them to smooth the responses. Hansotia and Rukstales (2002) used a single logistic regression model in which both control group data and treatment group data are collectively used for training. In addition to the default attributes of the customer a binary indicator variable is added to each instance to indicate whether the instance belongs to the treatment or control group. They also developed a decision tree approach with modified splitting rules on both control and treatment datasets to predict uplift. Lo (2002) took a similar approach to building a unified model using logistic regression. Even though additive models look like a single model, their functioning will be similar to Uplift models based on two model approach.

Radcliffe (2007) defined a metric to quantitatively measure the performance of an Uplift model. The theory of Gains chart and Gini coefficient is extended to uplift modeling for developing a quantitative measure. **Figure 4** shows a traditional gains chart between the number of customers targeted and number of purchases. The red line in the figure represents a perfect model whereas blue line represents a random model. The green line in between represents the response of customers based on some predictive model.

The Gini coefficient is the ratio of the area above diagonal for actual curve to the corresponding area of the perfect curve.

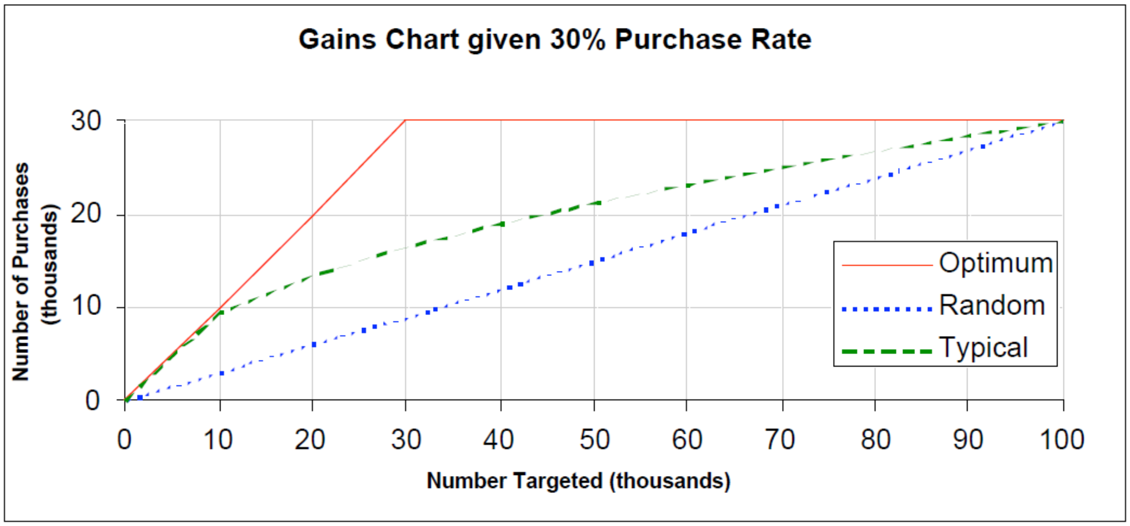


Figure 4: Gains chart representing the behavior of response rate (Radcliffe 2007)

So, a perfect response model will have a Gini index of 1 and response model with no predictive power will have a Gini index of 0.

Radcliffe extended gains chart and Gini index to quantitatively measure the performance of Uplift model. **Figure 5** shows the gains chart for Uplift (Qini Curve) between the number of people targeted and number of incremental purchases. The drop in the number of incremental purchases at the end of the curve is due to targeting anti-persuadable customers. The red curve represents the perfect uplift model where the number of incremental purchases reaches the maximum. The blue curve represents a merely random model with no predictive power. Green and purple curves represent curves which perform somewhere between perfect and random model. Similar to Gini index, Qini value Q is defined for Uplift models which is nothing but the ratio of areas above diagonal for actual model and perfect model. Another metric q_0 is defined as the ratio of the areas corresponding to these curves before the Qini curve starts turning down. The same theory can be extended to uplift modeling targeting increment in profit instead of purchase probability.

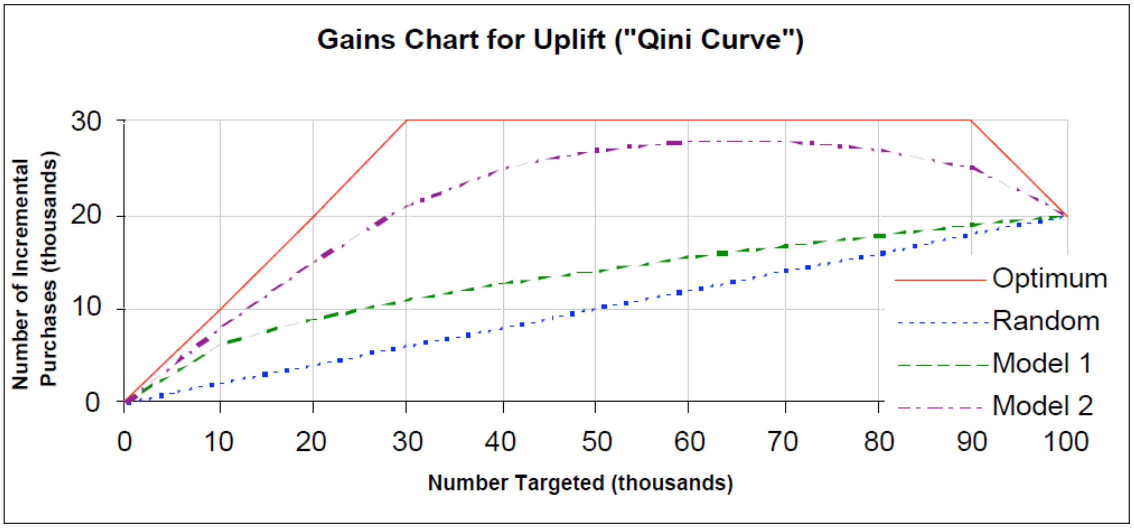


Figure 5: Qini Curve representing the behavior of Uplift model (Radcliffe 2007)

Radcliffe and Surry (2011) explained a few additional metrics like Validated qini, Monotonicity of incremental gains, Maximum impact, Impact at cutoff, Tight validation and Range of Predictions to quantitatively infer the effectiveness of the uplift model. They explained theoretically why two model approach will not be effective in modeling uplift when the uplift signal is relatively small when compared to the purchase proportion in control and treatment groups. **Error! Reference source not found.Figure 6** shows the distribution of purchase probability in control group (Blue, in the middle), treatment group (Red, at far left) and Uplift (Black, at far right) with respect to two customer attributes x and y. The three plots on the top are purely based on equation (2-6) whereas three plots on the bottom are along with sampling error. The underlying equations used by Radcliffe and Surry for generation these visualizations are

$$P^T = U[0, x) + \frac{U[0,y)}{10} + 3\ ,\ \ P^C = U[0, x) \quad (2\text{-}6)$$

$$Uplift = P^T - P^C = \frac{U[0,y)}{10} + 3 \quad (2\text{-}7)$$

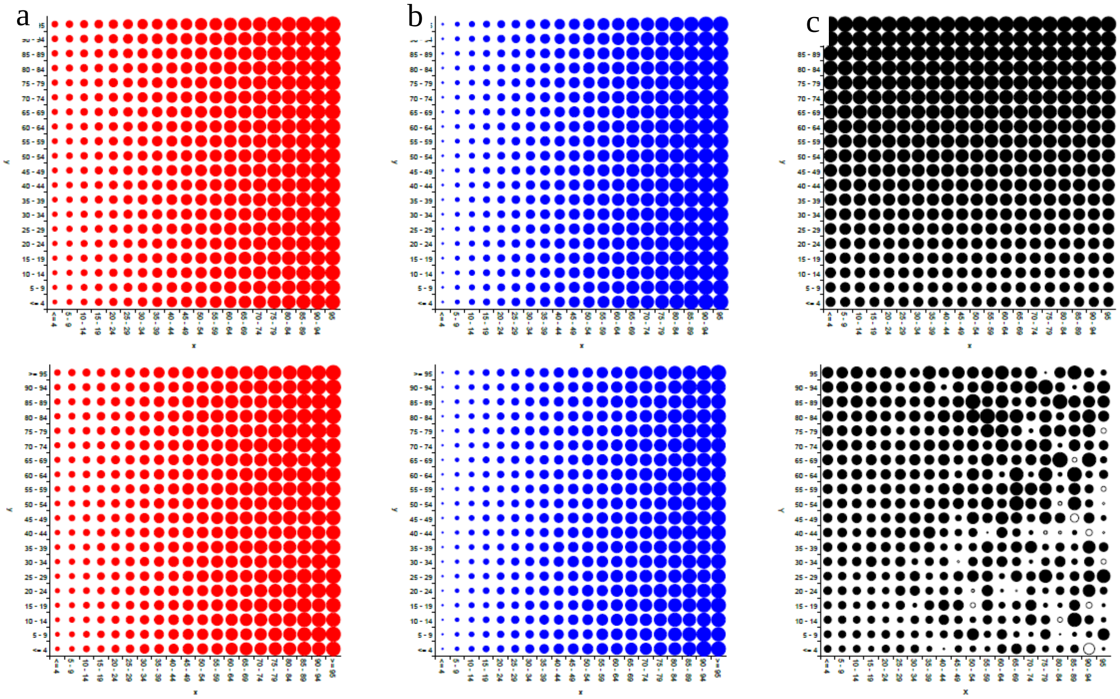


Figure 6: Modeling and Visualizing Uplift as a function of two variables x & y (Radcliffe & Surry 2011)

According to equation (2-7), Uplift should have correlation only with y. But Uplift models based on two model approach could not pick this trend which emphasizes the necessity of unified Uplift models with customized splitting criteria in case of weak signal.

After illustrating the drawbacks of two model approach, Radcliffe and Surry explained their unified approach which uses Significance based splitting, Variance based pruning, Bagging and pessimistic qini-based variable selection. The metric used by Radcliffe and Surry for measuring the effectiveness of split is

$$t^2\{\gamma_{TR}\} = \frac{(n-4)(U_R - U_L)^2}{C_{44} * SSE} \tag{2-8}$$

$$C_{44} = \frac{1}{N_{TR}} + \frac{1}{N_{TL}} + \frac{1}{N_{CR}} + \frac{1}{N_{CL}} \tag{2-9}$$

$$SSE = \sum_{i\epsilon\{T,C\}} \sum_{j\epsilon\{L,R\}} N_{ij} p_{ij} (1 - p_{ij}) \tag{2-10}$$

Uplift trees developed using the above metrics are pruned based on Variance. In this technique k decision tree models are built on resampled data and nodes having variance above certain cutoff will be pruned. It is also obvious that the root node will have zero variance and variance will gradually increase as we go to the leaf nodes. This uplift decision tree is pruned up to the depth of nodes having moderate variance. On top of this bagging is also used to achieve additional model stability.

Rzepakowski and Jaroszewicz (2012) presented a unified decision tree uplift model with multiple treatments and customized criteria for splitting. The primary metric for splitting a unified uplift decision tree is

$$\Delta\Delta\, P(A) = \left| \left(P^T(y_0 \,|a_0) - P^C(y_0 \,|a_0)\right) - \left(P^T(y_0 \,|a_1) - P^C(y_0 \,|a_1)\right) \right| \tag{2-5}$$

The above equation represents the uplift attained by splitting combined control and treatment data based on a categorical predictor y_0 such that $y_0 = a_0$ and $y_0 = a_1$. The strategy of unified uplift tree is to select splits that maximize $\Delta\Delta\, P(A)$. In addition to $\Delta\Delta\, P(A)$ Rzepakowski and Jaroszewicz used measures of divergence between class distributions such as Kullback-Leibler divergence, squared Euclidean distance and chi-squared divergence to direct the splits in the Uplift decision tree. The above divergences between two distributions $Q = (q_1, q_2, \dots \dots, q_n)$ and $P = (p_1, p_2, \dots \dots, p_n)$ were defined as

$$KL(P:Q) = \sum_i p_i \log \frac{p_i}{q_i} \tag{2-6}$$

$$E(P:Q) = \sum_i (p_i - q_i)^2 \tag{2-7}$$

$$\chi^2(P:Q) = \sum_i \frac{(p_i - q_i)^2}{q_i} \tag{2-8}$$

So the idea of Rzepakowski and Jaroszewicz's model is to maximize

$$D_{gain}(A) = D(P^T(Y):P^C(Y)\big|A) - D(P^T(Y):P^C(Y)) \tag{2-9}$$

At each split where D can take any of the above-explained divergence measures. The expression (2-9) represents the increment in divergence in the leaf nodes from parent node by splitting on predictor A. This is similar to Gini gain and entropy gain in conventional decision trees. Methodology to prune the tree based on control and treatment probabilities at each split is also explained. **Figure 7** shows comparative analysis of various metrics used for Uplift modeling by Rzepakowski and Jaroszewicz. It can be observed that two model approach had lowest predictive power followed by ΔΔP model whereas unified tree models with KL and Euclidian divergence metrics had higher predictive power. Same concept is further extended to multiple treatment uplift modeling.

Guelman et al. (2012) extended Rzepakowski and Jaroszewicz's idea of building unified uplift decision tree based on measures of divergence to Random Forest uplift model. The predicted uplift will be an average of the prediction from each tree in the Random forest. This approach provided additional tuning parameters to the model such as number of random predictors at each split and number of trees for optimizing bias and variance. Guelman et al. also explained about the advantage of measuring the relative importance of predictors while using Random Forest Uplift model. This is done by measuring the average increment in divergence based on the splits connected to each predictor.

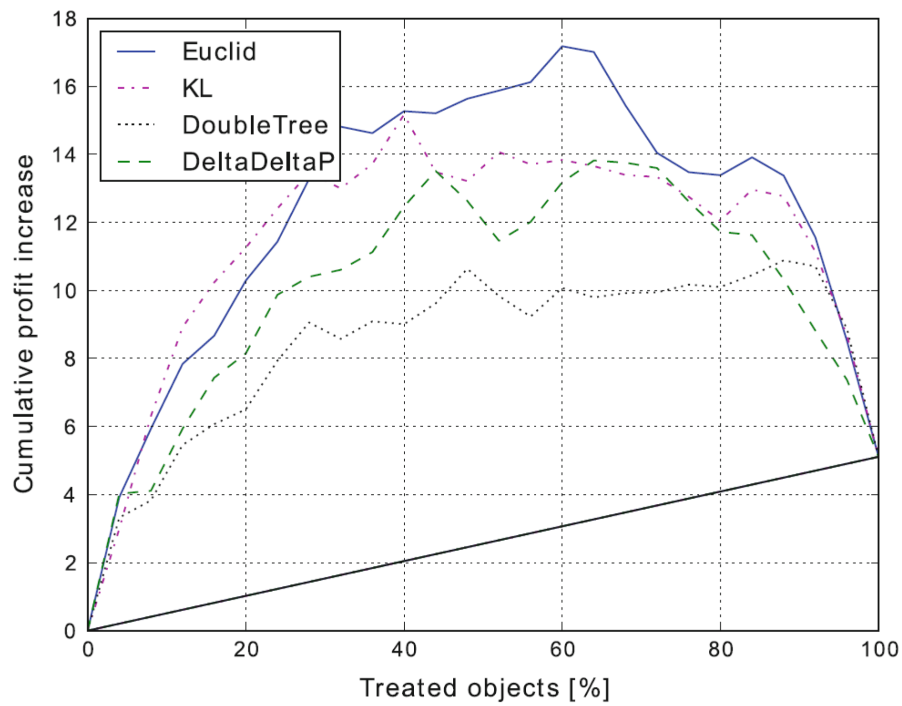


Figure 7: Comparison of various metrics used for Uplift modeling (Rzepakowski & Jaroszewicz 2011)

Zaniewicz and Jaroszewicz (2013) explained Uplift Support Vector Machines for uplift modeling by slightly modifying it into a prediction problem. The Uplift model was defined as

$$M(x): R^m \rightarrow \{-1,0,1\} \tag{2-10}$$

which assigns +1 to persuadable customers and -1 to anti persuadable customers and 0 to always buying and never buying customers. This is done by defining two hyperplanes $H_1: < w, x > -b_1 = 0$ $H_2: < w, x > -b_2 = 0$. Persuadable customers will fall to the right of both the hyperplanes, anti-persuadables will fall to the left of both hyperplanes whereas always buying and never buying customers will fall between two hyper plane

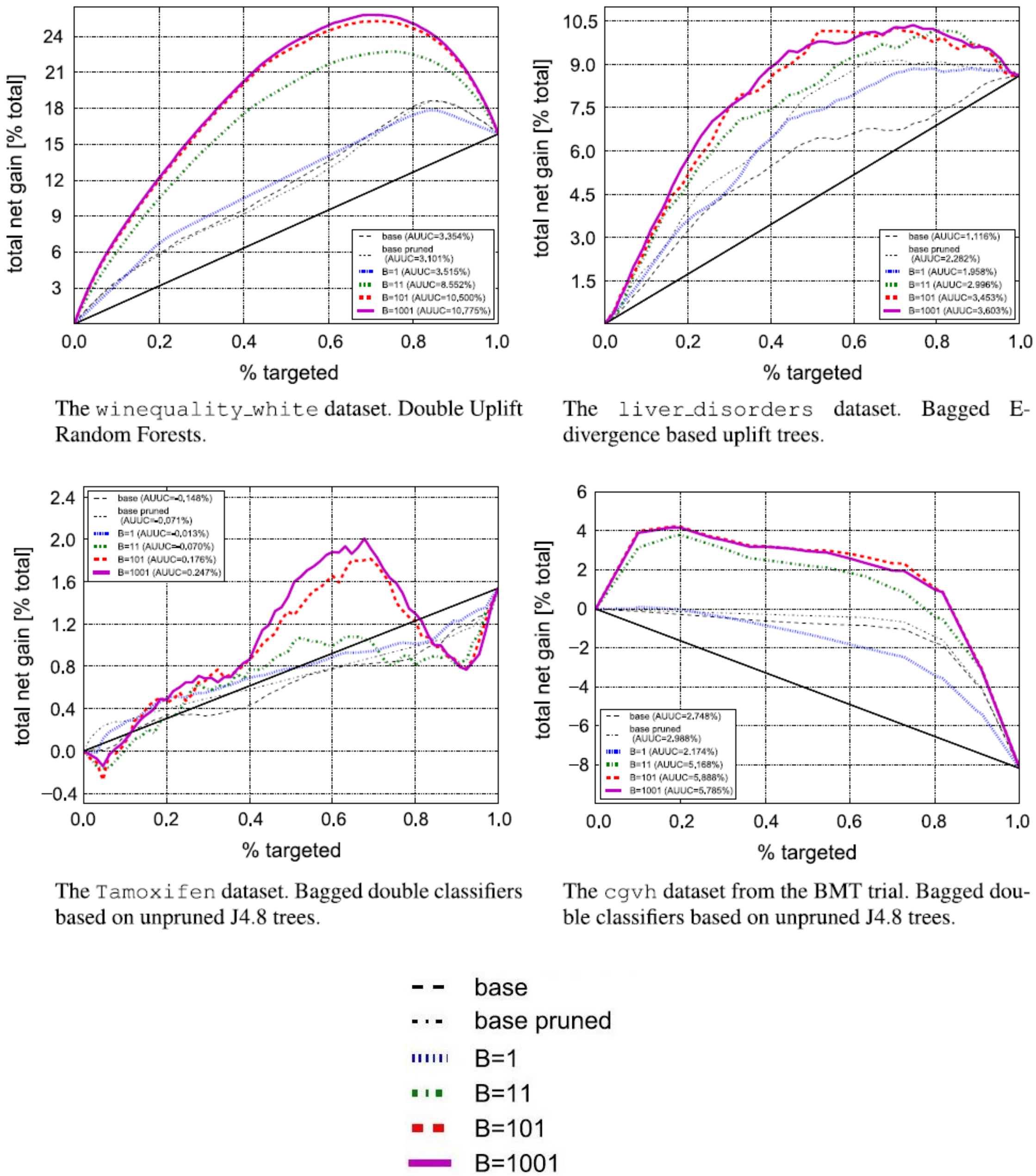
$$M(x) = \begin{cases} +1 & \text{if } (w, x) > b_1 \text{ and } > b_2 \\ 0 & \text{if } (w, x) \leq b_1 \text{ and } > b_2 \\ -1 & \text{if} (w, x) \leq b_1 \text{ and } \leq b_2 \end{cases} \tag{2-11}$$

Soltys et al. (2015) further extended the work of Radcliffe and Surry (2011) and Guelman et al. (2012) by experimental evaluation of ensemble models like Bagged trees and Random Forest. They also explained that the intrinsic nature of the Uplift problem makes it most suitable for ensemble modeling. By this time, a few standard datasets were available to benchmark proposed new models. Soltys et al. (2015) used standard data sets to evaluate the performance of pruned and unpruned divergence based uplift trees, double classifier uplift models, Bagged Uplift trees, Bagged double trees, Uplift Random Forests, and Double Uplift Random Forests. Repeated cross-validation (128 times) is used on data sets (80% training and 20% test) and the average of the result is used for evaluating the models.

From **Figure 8** Soltys et al. (2015) show that the performance of all kinds of Uplift models ranging from double decision tree to unified divergence based uplift trees can be improved by ensembling. It can also be observed that the quality of Uplift increases up to a certain number of trees and reaches optimum. As seen in other machine learning models Uplift Random Forest performed better than Uplift bagged trees. In this case, the difference between the quality of these two ensemble trees is small.

Kondareddy et al (2016) proposed a two-step method in which first step will use decision tree explained by Rzepakowski and Jaroszewicz (2011) and the second step will constitute a logistic regression model.

Zhao et al. (2017) introduced new methodology called Contextual Treatment Selection for addressing Uplift modeling with multiple treatments. They also introduced modified uplift curve to the measure the performance of uplift models.



The `winequality_white` dataset. Double Uplift Random Forests.

The `liver_disorders` dataset. Bagged E-divergence based uplift trees.

The `Tamoxifen` dataset. Bagged double classifiers based on unpruned J4.8 trees.

The `cgvh` dataset from the BMT trial. Bagged double classifiers based on unpruned J4.8 trees.

Figure 8: Effect of Ensembling (B No. of trees) on Uplift model performance (Soltys et al. 2015)

The Zhao et al. (2017) methodology involves tree construction with new splitting criteria and additional ensemble to avoid overfitting on training data. The effectiveness of a split is measured by

$$\Delta\mu(s) = P\{X\epsilon\phi_l|X\epsilon\phi\} \max_{t_l=0,1\ldots,k} E[Y|X\epsilon\, \phi_l\,, T = t_l] +$$

$$P\{X\epsilon\phi_r|X\epsilon\phi\} \max_{t_l=0,1\ldots,k} E[Y|X\epsilon\, \phi_l\,, T = t_l] -$$

$$\max_{t_l=0,1\ldots,k} E[Y|X\epsilon\, \phi\,, T = t] \qquad (2\text{-}12)$$

Each node in the tree will have a feature space ϕ based on the previous splits. The split value and treatment will be selected in such a way that maximizes $\Delta\mu(s)$. The tree building will continue until leaf nodes reaches minimum number of samples or there are no splits with positive $\Delta\mu(s)$.

Zhao et al. (2017) used artificially generated data to validate their method. The feature space was 50 dimensional and uniformly distributed between [0,10]. The effect of each treatment is modeled as shown in equation (2-13). f(X) represents the probability of purchase without treatment and $U(x)$ in each case represents the uplift in probability of purchase with each treatment. a^i, b^i_j, c^i_j are random constants.

$$Y = \begin{cases} f(X) + U[0, \alpha X_1] + \epsilon & \text{if } T = 1 \\ f(X) + U[0, \alpha X_2] + \epsilon & \text{if } T = 2 \\ f(X) + U[0, \alpha X_3] + \epsilon & \text{if } T = 3 \\ f(X) + U[0, \alpha X_4] + \epsilon & \text{if } T = 4 \end{cases} \qquad (2\text{-}13)$$

$$f(x_1, \ldots\ldots, x_{50}) = \sum_{i=1}^{50} a^i.\exp\{-b^i_1\left|x_1 - c^i_1\right| - \cdots - -b^i_{50}\left|x_{50} - c^i_{50}\right|\} \qquad (2\text{-}14)$$

2.2 Dynamic Uplift model

The uplift models discussed in the previous section consisted of two model approach, unified model approach, models with modified evaluation functions and uplift models for multiple treatments. But all of them dealt with the problem of uplift at a single instance of time based on a set of static customer attributes. Also, all the parameters directly affecting the purchase behavior of the customer in case of control and treatment are included in the model.

Research in this thesis will address uplift modeling in a dynamic environment which simulates the periodic purchasing behavior of the customer. The control and treatment probability of purchase will be a function of static attributes of the customers as well as dynamic attributes of the customer which change with time. All the parameters affecting the control and treatment probabilities will not be included in the prediction functions. Instead dynamic uplift model has to learn these features with time. The models discussed in the previous section explained metrics for measuring the performance of uplift models at a static instance. In this research modifications required to extend this model to a dynamic scenario will be analyzed.

3. Artificial data Generation

3.1 Static Data Model

Artificial datasets pertaining to customers, control probability weights and treatment probability weights, campaign data are generated using random number generating functions and appropriate transformations. The approach followed in this research for generating artificial data is similar to the approach followed in Zhao et al. (2017), Parret (2016) and Cham (2013). Parameters like number of customers, number of attributes per customer, number of time steps considered in the study, percentage of customers targeted in each phase of marketing are adjusted to make the model behave similar to naturally observed patterns in customer behavior.

Model Parameter	Initial Value
Number of Customers	1000
Attributes per Customer	30
Number of Time Steps	300
Mean of Customer Attributes	6
Standard deviation of Customer Attributes	2
Purchase probability cutoff	0.5

Table 1: Model Parameters and their Initial Value

Initially, customer data frame is created with 1000 customers and with each customer having 30 attributes. These attributes are generated using Gaussian random number generator with a mean of 6 and standard deviation of 2. This results in most of the customers attributes being on a scale of 2 to 10. Similarly, weight matrix is created to generate control and treatment odds. Each treatment weight vector corresponds to each type of offer customers may receive in marketing campaign. Treatments may have positive, neutral or negative impact on customers based on customer attributes.

$$\mathrm{M}_{(\mathrm{i,j})} = \mathrm{N}(6,2) \tag{3-1}$$

$$\mathrm{C}_{\mathrm{j}} = \mathrm{N}(0,0.05) \tag{3-2}$$

$$\mathrm{T}_{(\mathrm{j},1)} = \begin{cases} \mathrm{C}_{\mathrm{j}} + \mathrm{N}(0.01, 0.01) & 1 \leq \mathrm{j} \leq 10 \\ \mathrm{C}_{\mathrm{j}} & 11 \leq \mathrm{j} \leq 20 \\ \mathrm{C}_{\mathrm{j}} - \mathrm{N}(0.01, 0.01) & 21 \leq \mathrm{j} \leq 30 \end{cases} \tag{3-3}$$

$$\mathrm{T}_{(\mathrm{j},2)} = \begin{cases} \mathrm{C}_{\mathrm{j}} & 1 \leq \mathrm{j} \leq 10 \\ \mathrm{C}_{\mathrm{j}} + \mathrm{N}(0.01, 0.01) & 11 \leq \mathrm{j} \leq 20 \\ \mathrm{C}_{\mathrm{j}} - \mathrm{N}(0.01, 0.01) & 21 \leq \mathrm{j} \leq 30 \end{cases} \tag{3-4}$$

$$\mathrm{T}_{(\mathrm{j},3)} = \begin{cases} \mathrm{C}_{\mathrm{j}} + \mathrm{N}(0.01, 0.01) & 1 \leq \mathrm{j} \leq 10 \\ \mathrm{C}_{\mathrm{j}} - \mathrm{N}(0.01, 0.01) & 11 \leq \mathrm{j} \leq 20 \\ \mathrm{C}_{\mathrm{j}} & 21 \leq \mathrm{j} \leq 30 \end{cases} \tag{3-5}$$

$$\mathrm{W}_{(\mathrm{j,k})} = [\mathrm{C}_{\mathrm{j}}, \mathrm{T}_{(\mathrm{j},1)}, \mathrm{T}_{(\mathrm{j},2)}, \mathrm{T}_{(\mathrm{j},3)}] \tag{3-6}$$

$$\mathrm{SPO}_{(\mathrm{i,k})} = \mathrm{M}_{(\mathrm{i,j})} \mathrm{~X~} \mathrm{W}_{(\mathrm{j,k})} \tag{3-7}$$

$$\mathrm{SPP}_{(\mathrm{i,k})} = \frac{\mathrm{e}^{\mathrm{SPO}}}{1+\mathrm{e}^{\mathrm{SPO}}} \tag{3-8}$$

In equations 3-1 to 3-8, i will be equal to number of customers in the study and j will be equal to number of attributes per customer. k will be equal to one more than number of treatments (Offers) considered in the study. M represents customer matrix with dimension (i x j) and W represents weight matrix of dimension (j x 4). C_{j} and T_{j} represents control and treatment weight vectors of length 30. SPO stands for static purchase odds while SPP stands for static purchase probability. The first column in SPP represents static control purchase probability while remaining three rows represent treatment probability corresponding to each offer. From equation (3-3) it can be understood that customer attributes 1 to 10 will positively impact purchase probability with treatment-1 and customer attributes 21 to 30 will negatively impact the same. Customer attributes 11 to 20 will have neutral impact on purchase probability due to treatment-1. Similarly, other treatments impact purchase probability of customers based on other set of customer attributes. Different treatments will have different kind of impact on the customers and the objective of the modeling phase is to map treatments to customers at each marketing phase such that the uplift in purchase probability is maximized. **Figure 9** shows the slight shift in purchase probability of customers with treatment-1. This difference is not uniform for all the customers. But it will wary from customer to customer based on their attributes. The distribution of Treatment-1 weights is not normal because it’s a difference between two normal distributions. This static control and treatment probabilities are used to calculate dynamic control and treatment probabilities based on customer attributes effecting dynamic behavior.

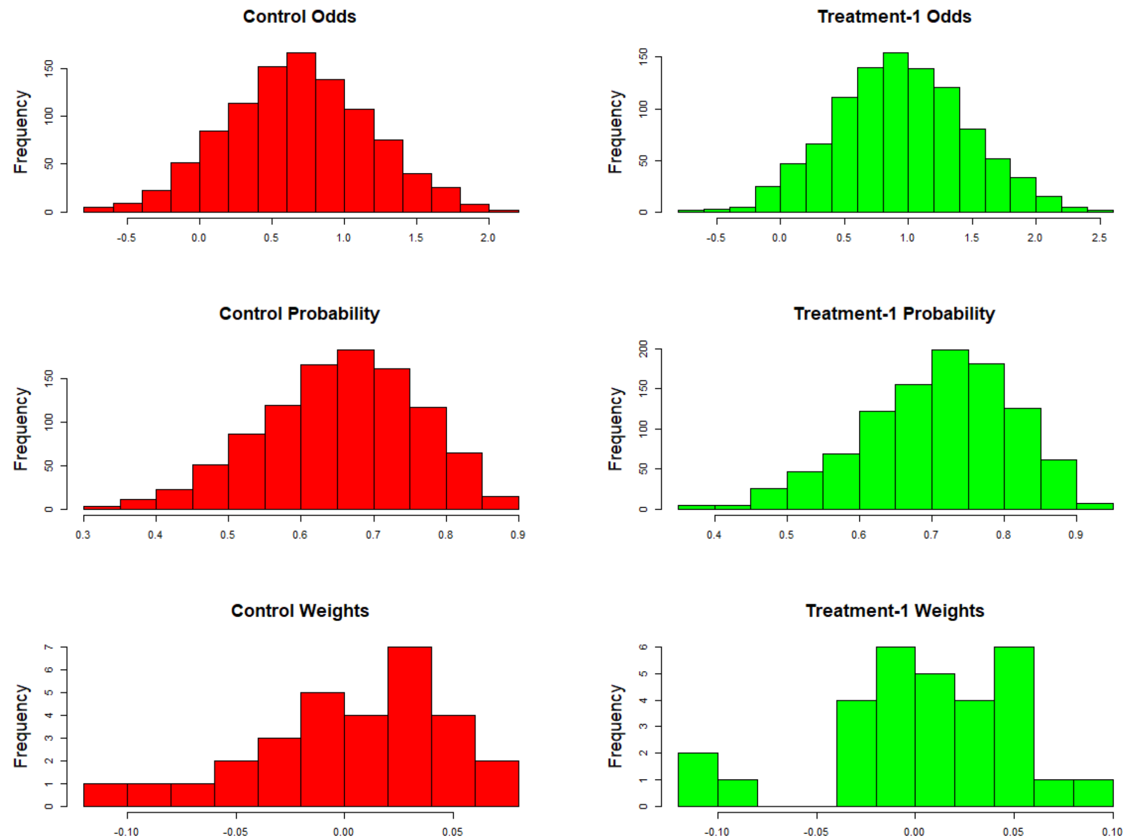


Figure 9: Distribution of Static Odds, Probability, and Weights related to Control and Treatment-1

3.2 Dynamic Data Model

Loyalty score and Addition score were defined as a function of static customer attributes to simulate the time-dependent changes in the behavior of the customers. Loyalty score of a customer will increase with an increase in number of purchases. Addiction score of a customer will increase with an increase in number of offers he/she receives. Both these dynamic parameters will depreciate with time. Loyalty score will increase the customer's control purchase probability whereas addiction score will decrease the treatment purchase probability. It can be observed in the real world that, as the customers buy more satisfactory products from a company, they will become more loyal to the company. Loyal customers may buy products even without offers. Equation (3-9) shows how loyalty score of each customer will increase with each purchase. $\mathrm{LS_{i,t}}$ represents the loyalty score of customer i at time t. $\mathrm{IP_{i,t}}$ is an indicator variable indicating purchase made by customer i at time t. Customers having higher attribute-1 will become more loyal with less number of purchases and vice versa.

$$\mathrm{LS_{i,t}} = \begin{cases} \mathrm{LS_{i,t-1}} + \mathrm{M_{i,1}} & \text{if } \mathrm{IP_{i,t}} = 1 \\ \mathrm{LS_{i,t-1}} & \text{if } \mathrm{IP_{i,t}} = 0 \end{cases} \tag{3-9}$$

In contrary to loyalty score, addiction score will decrease the treatment purchase probability. As a customer receives more and more offers, he/she will become less sensitive to offers. In this model, increase in addiction score with the number of offers received is dependent on attribute-2 of customers. Equation (3-10) shows how addiction score of each customer will be updated with respect to number of offers received. $\mathrm{AS_{i,t}}$ represents addiction score of customer i at time t. $\mathrm{IO_{i,t}}$ is an indicator variable indicating offer received by customer i at time t. Customers with high attribute-2 will represent customers who quickly lose interest in repeated promotional offers.

$$\mathrm{AS_{i,t}} = \begin{cases} \mathrm{AS_{i,t-1}} + \mathrm{M_{i,2}} & \text{if } \mathrm{IO_{i,t}} = 1 \\ \mathrm{AS_{i,t-1}} & \text{if } \mathrm{IO_{i,t}} = 0 \end{cases} \tag{3-10}$$

Both loyalty score and addiction score are set to zero after they reach a certain threshold. This represents a secondary level cycle in the purchase behavior of customers.

In addition to this, each customer will have a purchase cycle time which is also a function of static customer attributes. In this model each customer purchase cycle time will depend on their 5th and 10th attribute. Customers having higher values for these attributes will have longer purchase cycle time and vice versa. **Figure 10** shows control and treatment probabilities of customer-20 and customer-19 with time. Sum of attribute-

5 and attribute-10 of customer-20 and customer-19 is equal to 18 and 9 respectively. Based on these attributes, it can be observed that these two customers have different purchase cycle times. Time Factor, TF in equation (3-11) can be tuned to uniformly increase the purchase cycle time (PCT) for all the customers.

$$PCT_i = (M_{i,5} + M_{i,10}) * TF \tag{3-11}$$

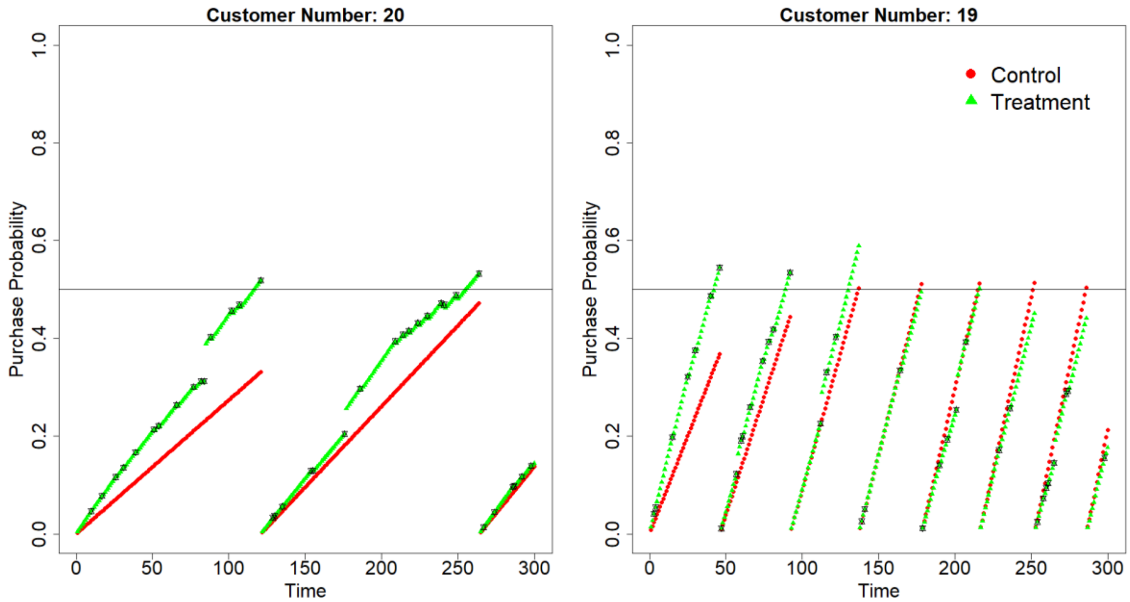


Figure 10: Dynamic purchase probability trend of two customers with different purchase cycle time

Dynamic control probability will be the product of static control probability ($SPP_{i,1}$), factor representing purchase cycle time and factor representing increment due to loyalty score. Equation (3-12) and (3-13) show the equations for calculating dynamic control probability. $A_c = 1.01$ and $B_c = 6$ are constants chosen by trial and error such that Dynamic control probability (DCP) is constrained between 0 and 1. RPT_i represents the recent purchase time of the customer.

$$DCP_{i,t} = SPP_{i,1} * \frac{(t - RPT_i)}{PCT_i} * LF_{i,t} \tag{3-12}$$

$$LF_{i,t} = \frac{A_c^{B_c * LS_{i,t}}}{1 + A_c^{B_c * LS_{i,t}}} \tag{3-13}$$

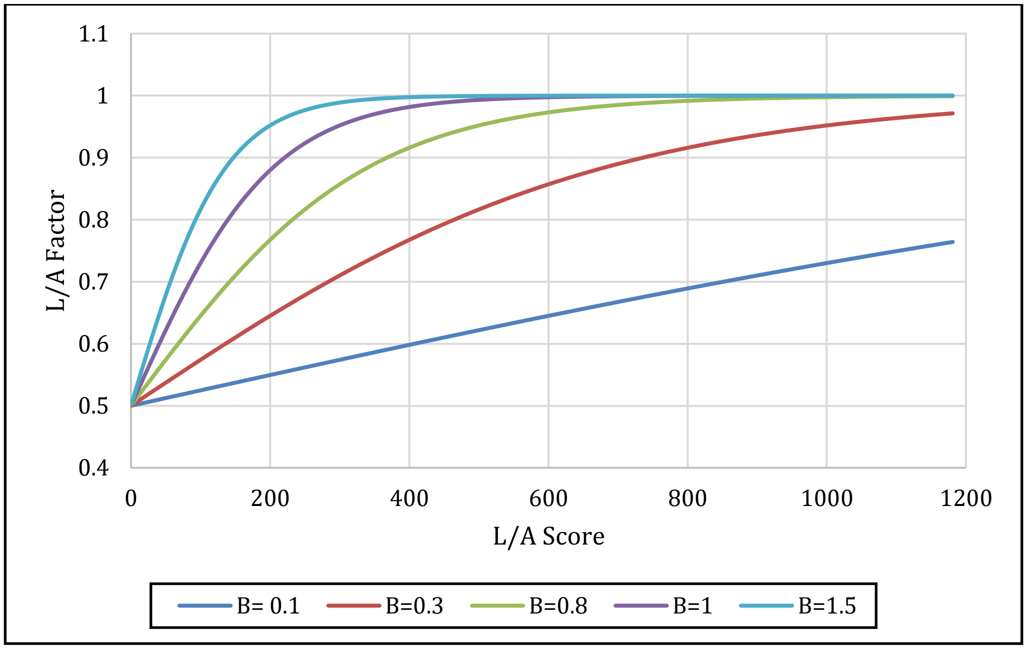


Figure 11: Relationship btw L/A Score and L/A Factor for different values of B

Similarly, dynamic treatment probability will be the product of static treatment probability, factor representing purchase cycle time and factor representing decrement due to addiction score. Equation (3-14) and (3-15) show the method of calculating dynamic control probability. $A_t = 1.01$ and $B_t = 1$ are constants chosen by trial and error such that Dynamic treatment probability (DTP) is constrained between 0 and 1. **Figure 11** shows how the relationship between Loyalty/Addiction score and Loyalty/ Addiction factor can be manipulated by changing the values of B. Based on the equations defined, the model is run for $noOfTimeSteps = 300$ and number of purchases made and number of offers received by each customer were recorded. Also, at each time step, data pertaining to customers who made purchases and who received offers is stored in a data frame and used for making uplift model for next time step.

$$DTP_{i,t} = SPP_{i,2} * \frac{(t-RPT_i)}{PCT_i} * AF_{i,t} \tag{3-14}$$

$$AF_{i,t} = \frac{A_t^{B_t * LS_{i,t}}}{1+A_t^{B_t * LS_{i,t}}} \tag{3-15}$$

3.3 Customer Segmentation based on Dynamic Uplift

In section 2.1, customers were segmented into four groups namely always buying, never buying, persuadables and anti- persuadables. These four customer segments can be further explained with respect to dynamic model and in addition to that customers can be further characterized by their sensitivity to offers and satisfactory purchases. The control and treatment probability trends of these customers are graphically explained with respect to time.

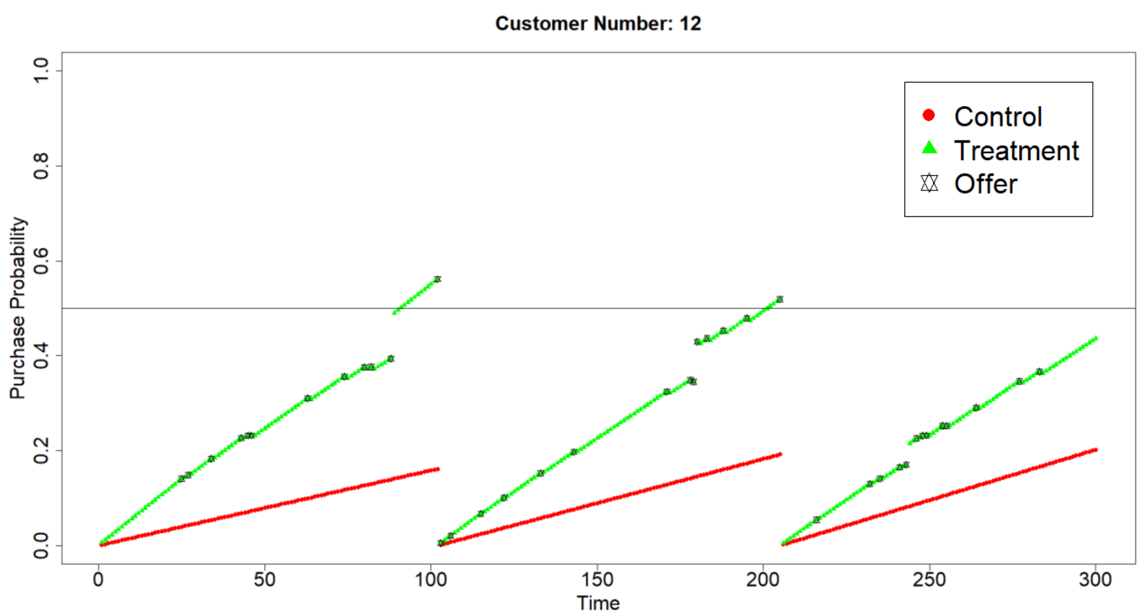


Figure 12: Control and Treatment Probability trend of a Persuadable Customer

Figure 12 shows the dynamic purchase probability of a persuadable customer. The cutoff purchase probability is set at 0.5. The control purchase probability is increasing at a slower pace and may take a longer time to reach cut off purchase probability. But the treatment probability increased at a faster pace and reached purchase cut off probability and made customer purchase twice during the time of study. The offers received by the customer are represented by black stars as shown in the legend. The stars plotted below the purchase probability cutoff line are unsuccessful offers and stars plotted above the purchase probability cutoff line are successful offers. The ratio of these two can be used to measure the efficiency of the campaign and it will be discussed in further section. **Figure 13** shows the dynamic purchase probability of a never buying customer. Both control and treatment probability of the customer never reached purchase cutoff probability in spite of receiving multiple offers. In this case, the vendor will lose the money he spent to send offers to this customer.

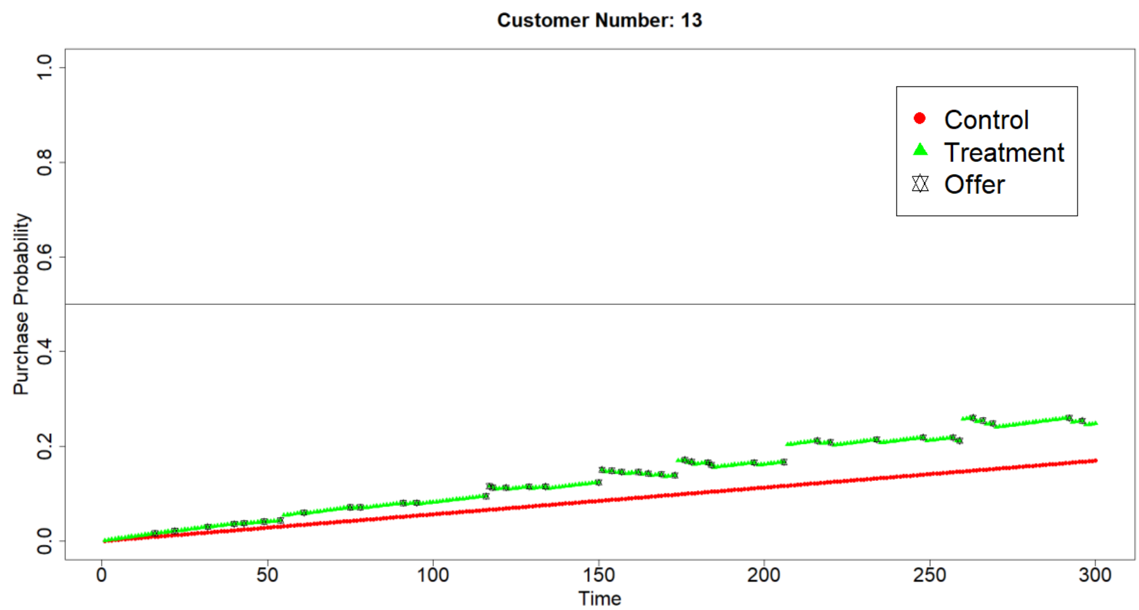


Figure 13: Control and Treatment Probability trend of Never Buying Customer

Figure 14 shows the dynamic purchase probability of an always buying customer. It can be observed that there is no significant difference between control probability and treatment probability. The change in the trends in the difference between these two probabilities is due to the dynamic effect caused by loyalty and addiction scores. The vendor can direct the advertising effort put on these customers to persuadable customers.

Figure 15 shows the dynamic purchase probability of an anti-persuadable customer. The control purchase probability of these customers is higher than the treatment purchase probability. The vendor is losing both advertisement effort and probable future purchases by targeting these customers.

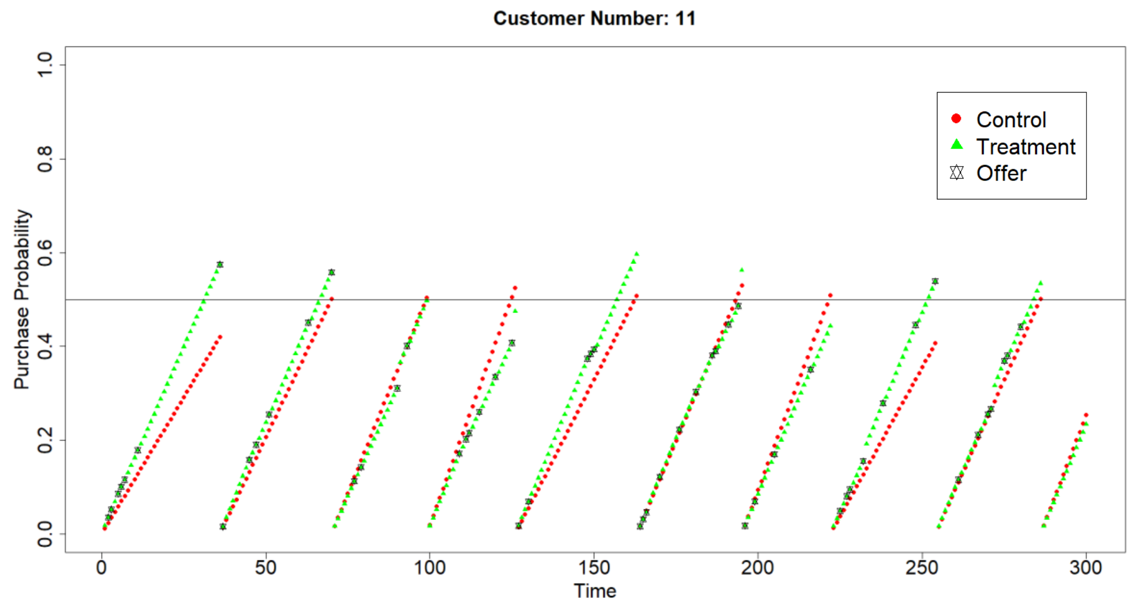


Figure 14: Control and Treatment Probability trend of Always Buying Customer

Even though the model is behaving according to these control and treatment probabilities, the vendor will have access to only data related to offers sent to customers and purchases made by the customers. Only this data can be used to develop uplift models which will be used for targeting customers for further phases of direct marketing. A separate data frame is created for storing this data and it will be updated by the offers received and purchases made by the customers in each session. With time the size of the data frame will become very big and may require significantly large computational effort to build dynamic uplift model. At this stage, random sampling with higher weightage to recent instances can be used for reducing computational effort and time.

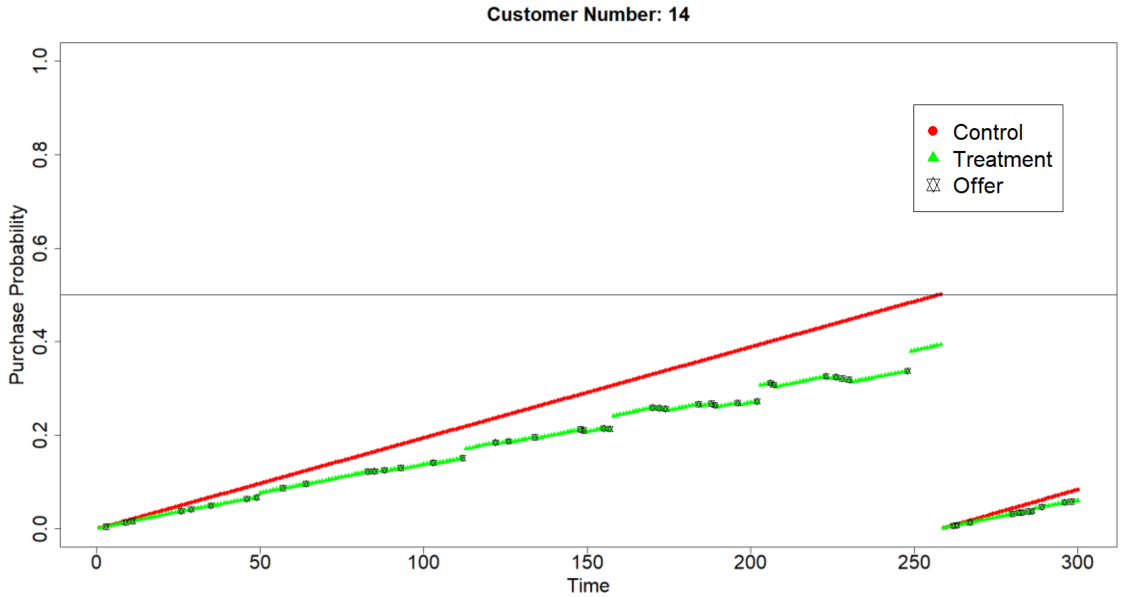


Figure 15: Control and Treatment Probability trend of Anti Persuadable Customer

3.4 Multiple Treatment Scenario

The methodology explained in previous section can be extended to multiple treatment scenario. Based on the customer attributes, different customers will have different kind of effect with different treatments. Three treatment types (1,2,3) is considered in this model. Static and dynamic purchase probability is calculated for each of these treatments. Customer attributes can have positive, neutral or negative impact on each of the treatment as explained in section 3.1. Similarly, these treatment probabilities will get impacted differently by dynamic attributes. Treatment-2 purchase probability can be more sensitive to addiction score than sensitivity of treatment-1 purchase probability. **Figure 16** and **Figure 17** show the dynamic purchase probability trend of two customers customer-20 and customer-14. In contrast to single treatment, in few cases, it may not be possible to unconditionally classify a customer as persuadable, anti-persuadable, always buying or never buying customer. But the following classifications can be made with respect to a particular treatment. It can be observed from Figure 16 that customer-20 is persuadable with respect to treatment-1 and treatment-2 but anti persuadable with respect to treatment-3. It can also be observed that uplift due to treatment-1 is significantly higher than the uplift due to treatment-2. In Figure 17, customer-14 is persuadable only with treatment-2 but not with other two treatments. It can also be observed that customer-14 is highly anti-persuadable with respect to treatment-3.

The objective of uplift modeling with multiple treatments is to map treatments to customers which can produce highest possible uplift in purchase probability. Different treatments will have different costs associated with them and this can make the model optimization as well as machine learning and inference problem. Selecting right treatment for the right customer at the right time will increase the efficiency of the marketing campaign and in turn increase the net revenue of the company.

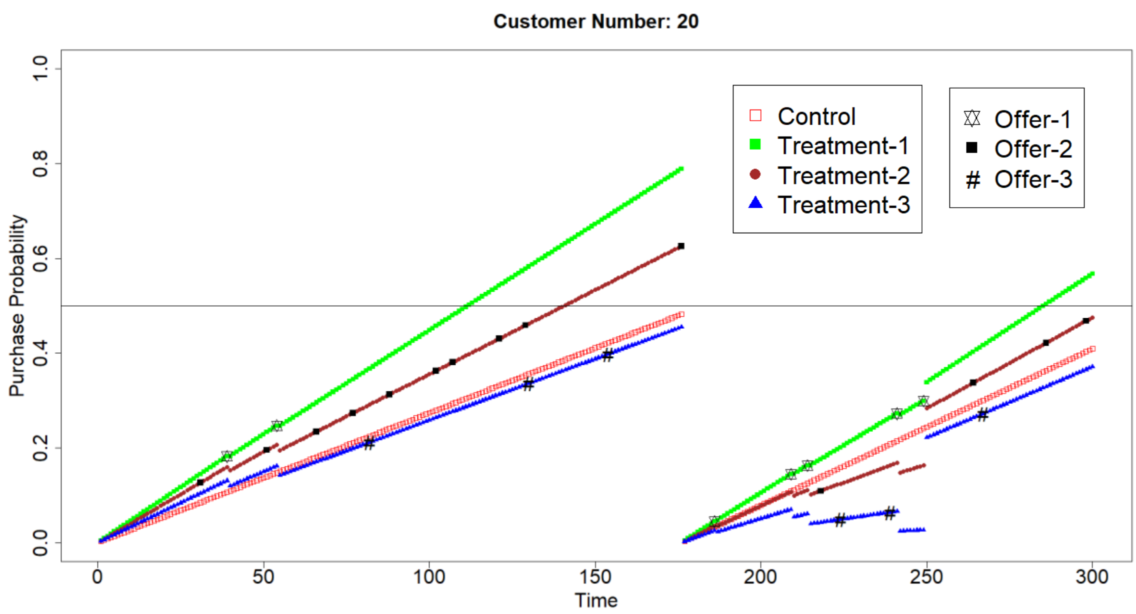


Figure 16: Dynamic Purchase Probability with multiple treatments (Customer-20)

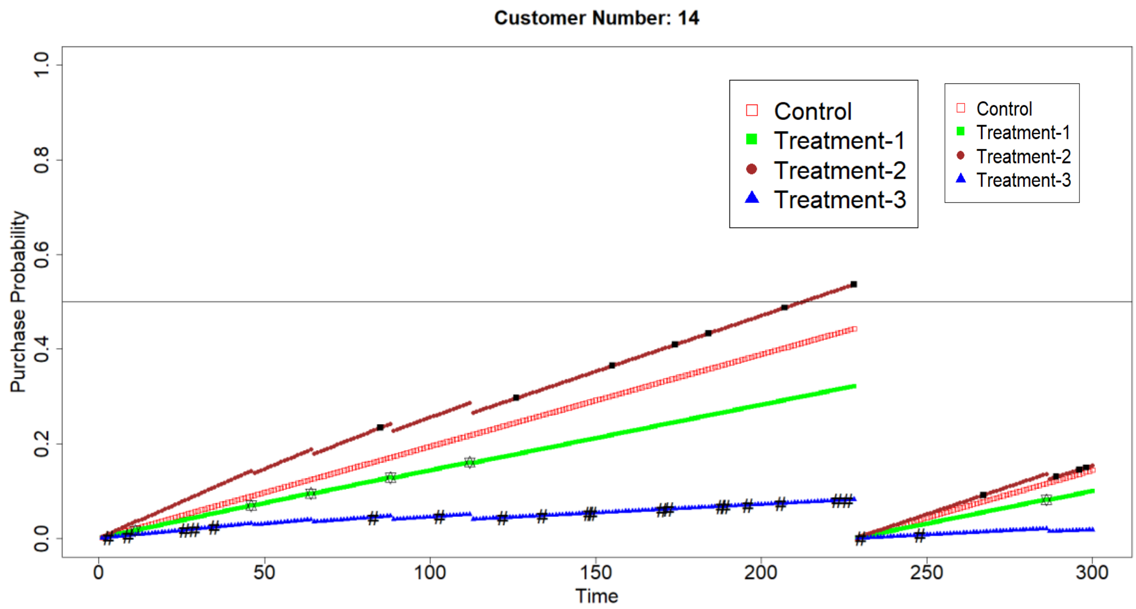


Figure 17: Dynamic Purchase Probability with multiple treatments (Customer-14)

4. Dynamic Uplift Modeling

4.1 Model Assumptions and methodology

The objective of Static uplift modeling discussed in Chapter-2 is to identify which customers should be targeted by direct marketing. But the question dynamic uplift modeling answers is **“Which customers should be targeted at what time by a direct marketing campaign”.** With multiple available treatments, this question can be further modified as “Which customers should be targeted at what time **by what** direct marketing campaign”.

A few assumptions are made for facilitating the approach of dynamic uplift modeling. These assumptions are as follows

i. The offers received by the customers at a certain time shows their impact only at that point of time and won’t be carried forward.

ii. The retail company will have access to the static attributes of the customer but not to dynamic attributes of the customer. With time and data collected from previous marketing campaigns, the dynamic uplift model will learn these parameters.

iii. The static control and treatment purchase probabilities are linearly dependent on the static customer attributes. But the dynamic control and treatment purchase probabilities are not necessarily linearly dependent on static customer attributes.

Figure 18 shows the flow chart of methodology followed in dynamic uplift modeling for single treatment scenario. At each time step, customer data and campaign data are given as input to dynamic uplift model to predict pesuadable customers. A marketing offer will be sent to the customers predicted by the model. Response of these treated customers as well as other customers will be recorded in the campiagn data base and used for improving the model in further time steps. While proceeding forward in time, the dynamic uplift model has to learn the secondary features of the customers which are dependent on static customer attributes. Dynamic uplift model will not have direct access to secondary features of the customers. Various uplift modeling approches like two model approach, additive model approach and unified modeling approach will be tried and evaluated for dynamic uplift model. To evaluate these models, metrics like total number of offers sent to customers, total number of successful offers, percentage of successful offers, and total purchases during the whole campaign period will be recorded.

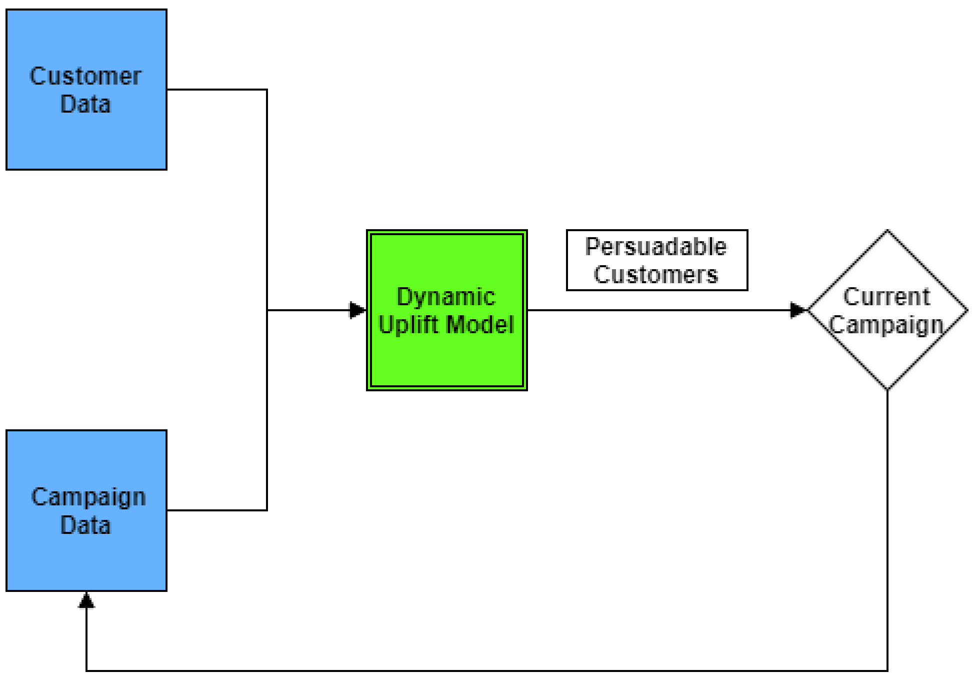


Figure 18: Dynamic Uplift model flow chart

4,4.2 Zero Treatment and Random Treatment Scenario

To evaluate dynamic uplift models in further sections, two scenarios zero treatment and random treatment will be used as a bench mark. In zero treatment scenario, none of the offers will be sent to the customers and all the purchases will be control purchases. In random treatment scenario, offers will be sent to random fraction of customers at each marketing phase/ time step. While comparing a dynamic uplift model with random scenario, it is necessary to consider both cumulative number of purchases and cumulative number of offers sent to the customers. This is because in a customer group with significant number of persuadable customers, cumulative number of purchases will increase with increase in number of offers. This trend may not be monotonous in customer groups with significant number of anti- persuadable customers.

Model is run for zero treatment scenario and random treatment scenarios with different depth of treatment. In this research depth of treatment is defined as the percentage of customers selected for direct marketing campaign. **Table 2** shows the cumulative number of purchases achieved with increasing depth of treatment. Appropriate random treatment scenario will be used to evaluate the uplift model performance based on the cumulative number of offers sent.

Depth of Treatment	0%	2%	5%	10%	20%	30%	40%	50%
Cumulative offers	0	6,000	15,000	30,000	60,000	90,000	120,000	150,000
Cumulative purchases	4,632	4,994	5,187	5,361	5,558	5,648	5,715	5,741

Table 2: No. of Cumulative purchases with increasing depth of treatment

Example: If a dynamic uplift model uses 6000 offers during the marketing period, it cumulative purchases will be compared with random treatment model with 2% depth of treatment (6000 cumulative offers).

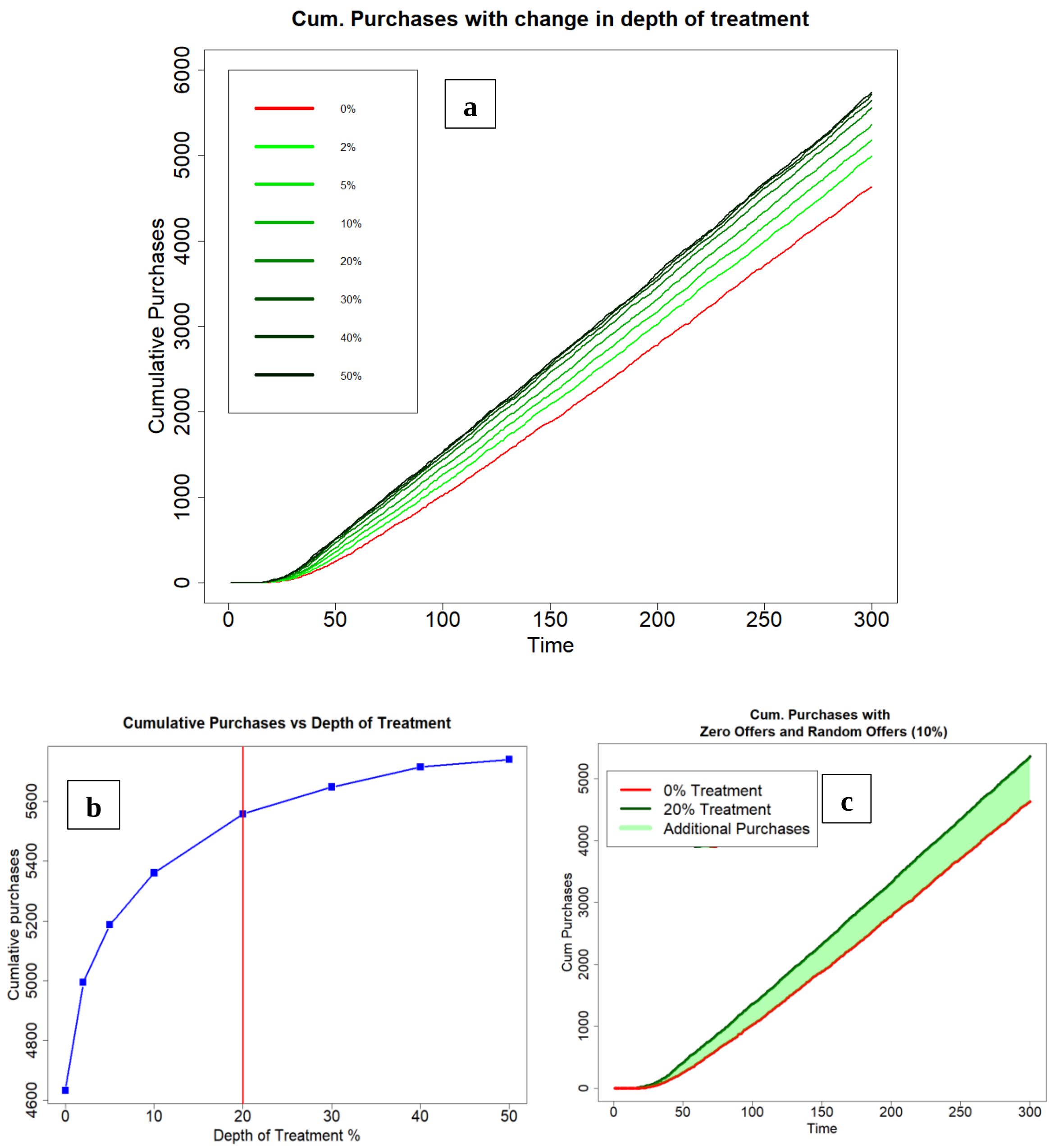


Figure 19: Cumulative Purchases with increasing depth of treatment per phase/time step (Random model)

	Base case without any treatment	Base case with random treatment selection (20%)
Cumulative No. of treatments	0	60,000
Cumulative control Purchases	4,632	2,408
Cumulative Treatment Purchases	0	3,150
Total Purchases	4,632	5,558

Table 3: Summary statistics of Zero Treatment and Random Treatment Scenario

Figure 19a shows the trend in cumulative purchases by all the customers for different depths of treatment. It can be observed that increment in cumulative purchases decreases with increase in depth of treatment. **Figure 19b** shows increment in total purchases with increment in depth of treatment. 20% treatment depth appears to be optimum for random treatment case. **Table 3** and **Figure 19c** shows longitudinal cumulative purchase trend for zero treatment as well as random 20% depth of treatment scenario.

It can be observed that number of control purchases decreased but increment in the number of treatment purchases is higher than the decrement in number of control purchases. In the model defined, sending offers to the customers makes them purchase more number of products in a particular time period. As seen in Figure 12 by making customers select treatment purchases effectively, the purchase frequency of the customers is increasing. In this process a few treatment purchases will replace control purchases, but the overall number of purchases will increase over period of time.

4.3 Logistic Regression Uplift Model

During the initial phase of the campaign, companies won't have data about the purchase trends of the customers. Due to this they are forced to send offers to random customers while simultaneously collecting the purchase data. Even though the marketing campaign will not be efficient in this phase, data collected during this phase can be used for dynamic uplift modeling.

Logistic regression model comes under additive model approach. Each instance of purchase by the customers both with and without promotional offer was recorded in the campaign data base. A single logistic regression model is built using customer attribute data and campaign data after each marketing phase. This model is used to predict the purchase probability of customers with and without treatment at different points of time. The probability of purchase without offer is subtracted from probability of purchase with offer to calculate uplift probability. Initially with limited data, dynamic uplift model will behave only as good as random model. But with increase in the number of instances in campaign data, uplift model will learn the trend in control, treatment as well as uplift probabilities.

Figure 20 shows the dynamic probability plot using logistic uplift model (2% depth of treatment) to select customers for direct marketing at each stage. It can be observed that the model has selected customers only when there is significant difference between treatment probability and control probability. Even though the difference between treatment probability and control probability is changing dynamically based on the offers received and purchases made by the customers, the model has learnt the trend in the data and efficiently selected the persuadable customers. **Figure 21** shows the dynamic probability plot using logistic uplift model with 10% depth of treatment. It can be observed that as the depth of treatment increases, the model starts to select always buying customer instances along with that of persuadables.

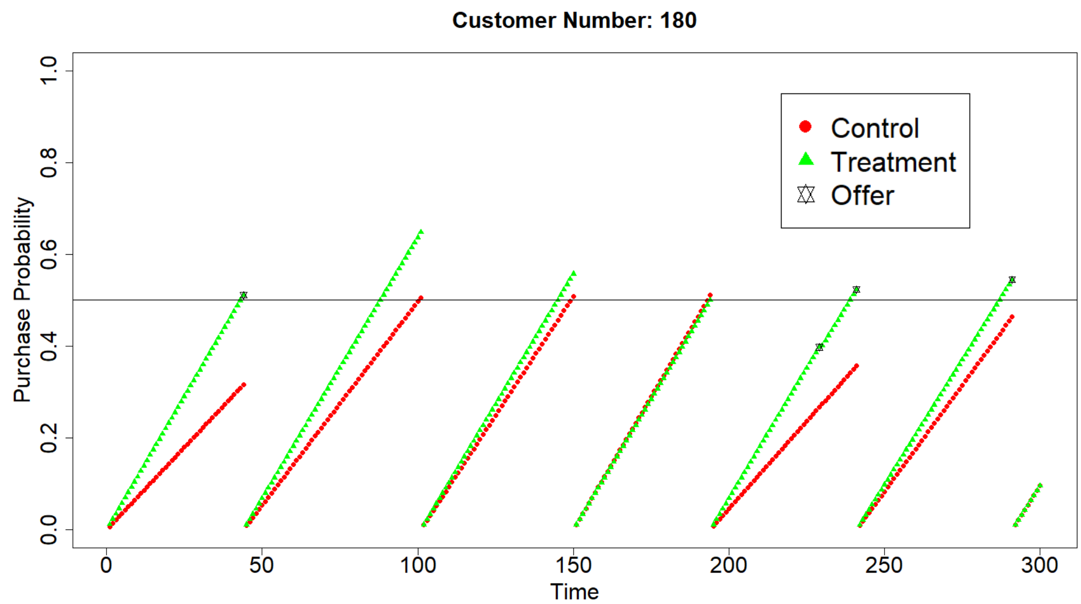


Figure 20: Dynamic Probability Trend with Logistic Uplift Model (2% depth of treatment)

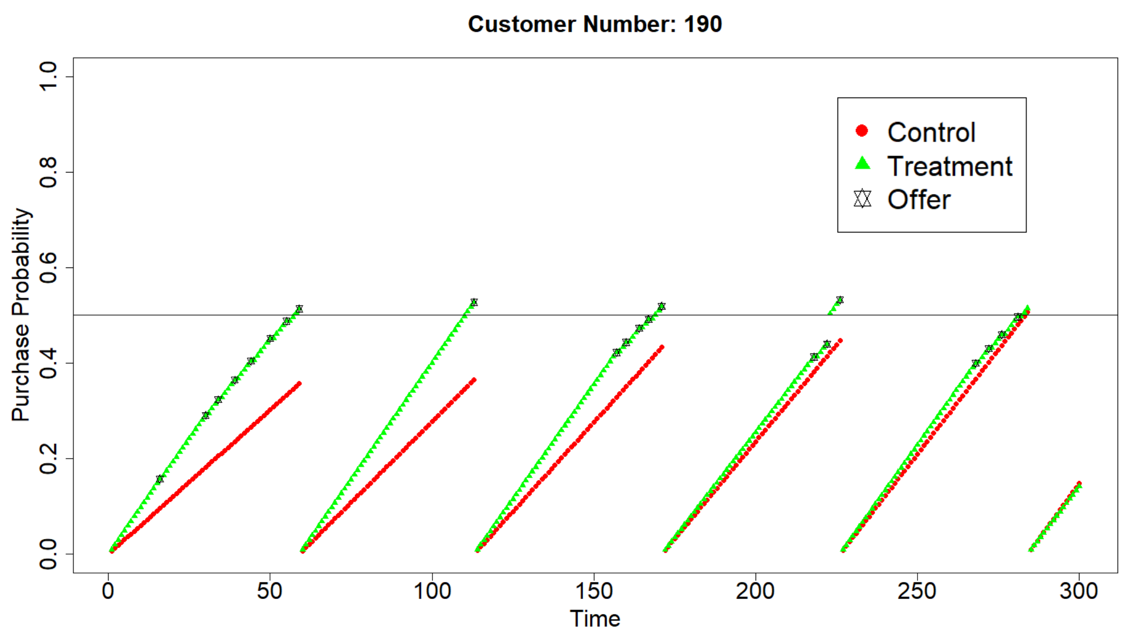


Figure 21: Dynamic Probability Trend with Logistic Uplift Model (10% depth of treatment)

Depth of Treatment per phase/ time step	0%	2%	5%	10%	20%	50%
Cumulative purchases at the end of marketing period (Logistic Regression Uplift model)	4,632	5,173	5,455	5,663	5,763	5,686

Table 4: No. of Cumulative Purchases with increasing depth of treatment. (Logistic Regression Uplift model)

Table 4 shows the total number of cumulative purchases attained for different depths of treatment when logistic regression model is used for shortlisting customers to receive offers at each stage. **Figure 22b** graphically depicts the cumulative number of purchases using random model and logistic regression uplift model. Cumulative number of purchases incresed intially with increase in the depth of treatment the trend reversed at higher depth of treatment (>40%). This reverse in trend could be due to including anti-persuadables customers and also may be due to the negative effects of over advertizement on the customers.

Figure 22a shows the change in cumulative purchases with varying depth of treatment using logistic regression uplift model. It can be observed that the optimum number of total cumulative purchases took place with 10% depth of treatment. **Figure 22c** shows the longitudinal view of cumulative purchase trend for zero treatment, random treatment model and uplift model using logistic regression with 10% depth of treatment. The shaded area in the graph represents increment in cumulative purchases between zero treatment, random teatment (10%) and treatment based on logistic regression uplift model.

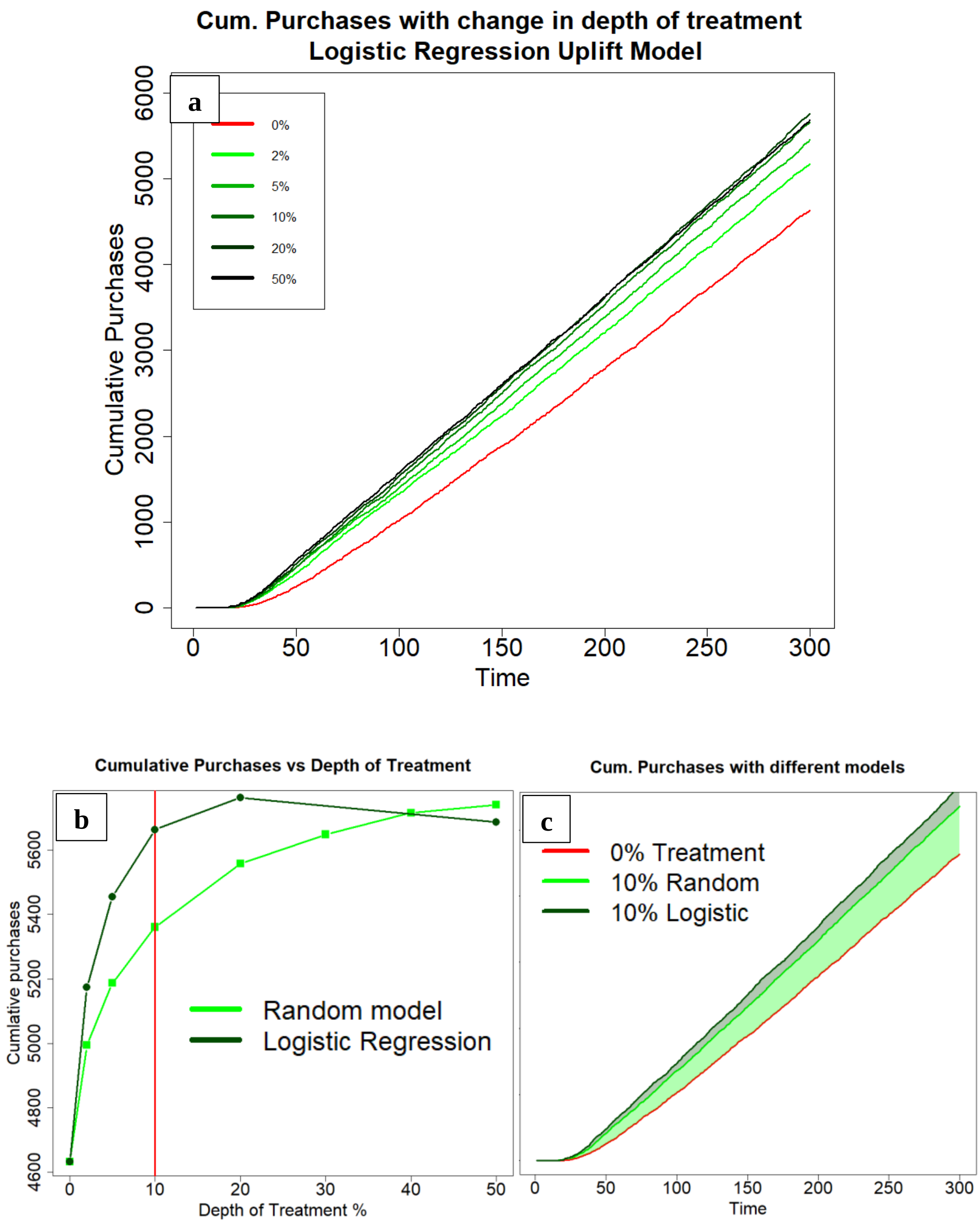


Figure 22: Cumulative Purchases with increasing depth of treatment per phase/time step (Logistic Regression Uplift Model)

4.4 Decision Tree Uplift Model

Decision tree uplift model comes under two model approach. In this approach two different decision trees will be used to predict control purchase probability and treatment purchase probability. The methodology followed for building decision trees is slightly different from logistic regression model. Customer instances having offer are separated into treatment dataset and similarly remaining customer instances are merged into control data set. Treatment purchase probability of new customer instances will be predicted from decision tree built from treatment data. Control purchase probability of new customer instances will be predicted from decision tree built from control data. Control purchase probability is subtracted from treatment purchase probability to estimate uplift probability based on two model decision tree approach. **Figure 23** shows the dynamic probability plot using double decision tree model (2% depth of treatment) to select customers for direct marketing at each stage. With this plot customer selection for treatment looks similar to that of logistic regression but the efficiency with different depth of treatment may differ.

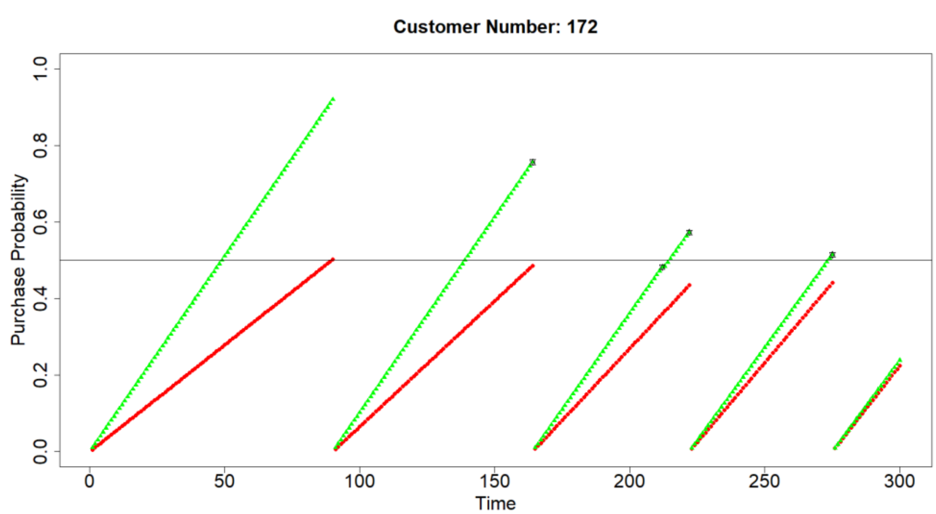


Figure 23: Dynamic Probability Trend with Double Decision Tree Model (2% depth of treatment)

Depth of Treatment per phase/ time step	0%	2%	5%	10%	20%	50%
Cumulative purchases at the end of marketing period (Double Decision Tree Uplift model)	4,632	4,860	4,946	5,123	5,217	5,521

Table 5: No. of Cumulative Purchases with increasing depth of treatment. (Double Decision Tree Uplift model)

Table 5 shows the total number of cumulative purchases attained for different depths of treatment when double decision tree model is used for shortlisting customers to receive offers at each stage. **Figure 24b** graphically depicts the cumulative number of purchases using random model and double decision tree uplift model. The cumulative purchases for different depths of treatment using double decision tree model is surprisingly less than the random model. Even thought this type of observation is not practical, this could be due to improper regularization and decision tree pruning methodology followed.

Figure 24a shows the change in cumulative purchases with varying depth of treatment double decision tree uplift model. It can be observed that there is no clear cut inflextion point in contrast to the Logistic regression uplift model. **Figure 24c** shows the longitudinal view of cumulative purchase trend for zero treatment, random treatment model and uplift model using double decision tree with 10% depth of treatment. It can be observed that the cumulative purchases using double decision tree uplift model is monotonically less than the random model. Decision tree purning is not used while building these control and treatment decision trees at each step considering the fact that model is not going to handle new customer data. This could be the reason behind poor performance of double decision tree uplift model.

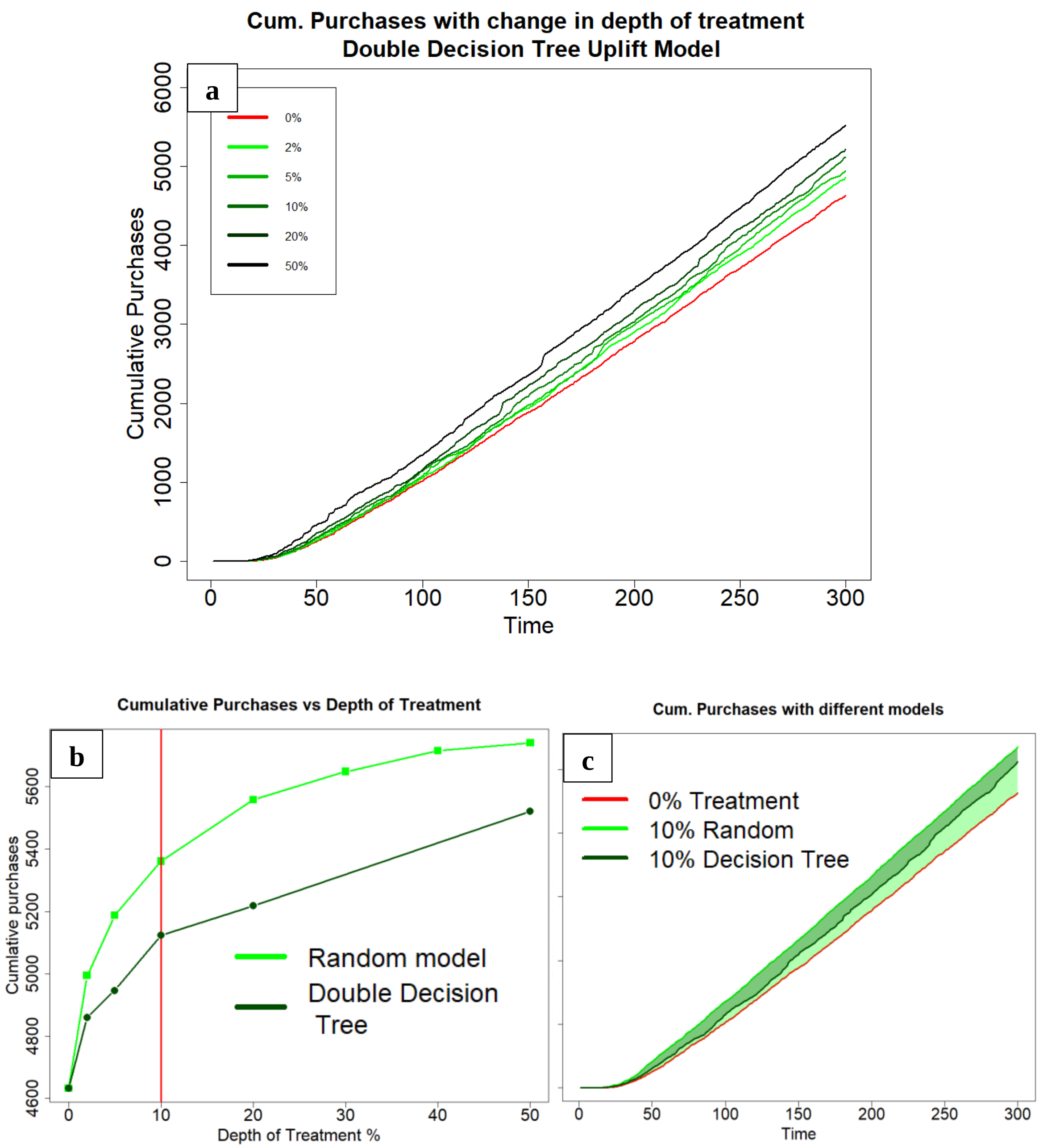


Figure 24: Cumulative Purchases with increasing depth of treatment per phase/time step (Double Decision Tree Uplift Model)

4.5 Unified Uplift Model

In unified modelling approach a decision tree is made on the basis of ΔΔ P. At each split, the decision tree will select those parameters for the split which will result in maximum difference of difference between percentage of observations belonging to control group and treatment group. The stopping conditions for the tree is a minimum value for ΔΔ P. The tree building will go on until the tree cannot find further splits which have ΔΔ P more than minimum ΔΔ P. Another stopping condition that can be imposed on the tree building process is minimum number of observations that can be present in a node. If a branch of the uplift tree reaches that limit the node will not be split further.

Depth of Treatment per phase/ time step	0%	2%	5%	10%	20%	50%
Cumulative purchases at the end of marketing period (Unified Uplift model)	4,632	5,438	5,642	5,728	5,745	5,755

Table 6: No. of Cumulative Purchases with increasing depth of treatment. (Unified Uplift Model)

Table 5 shows the total number of cumulative purchases attained for different depths of treatment when unified uplift model is used for shortlisting customers to receive offers at each stage. **Figure 25b** graphically depicts the cumulative number of purchases using random model and unified uplift model. The cumulative purchases for different depths of treatment using unified uplift model is more than the random model. **Figure 25a** shows the change in cumulative purchases with varying depth of treatment for unified uplift model. **Figure 25c** shows the longitudinal view of cumulative purchase trend for zero treatment, random treatment model and unified uplift model 10% depth of treatment.

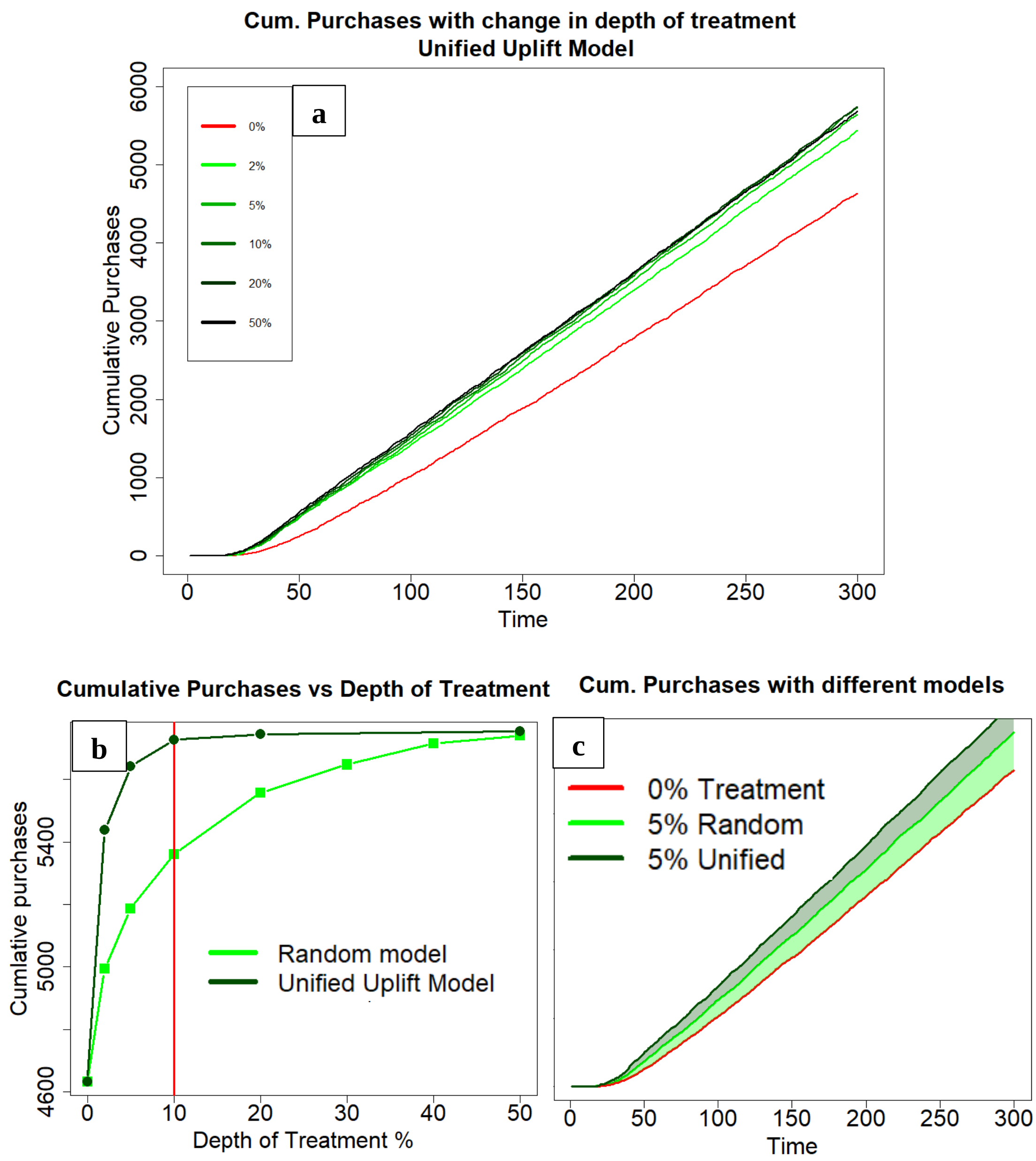


Figure 25: Cumulative Purchases with increasing depth of treatment per phase/time step (Unified Uplift Model)

5. Summary and Conclusions

The following study practically illustrates a methodology for extending the static uplift methodology to a dynamic scenario. In this study customer outcomes are a function of time as well as a few other dynamic attributes which are changing with time. Three different approaches used for static uplift modelling, two model approach, additive model approach and unified modelling approach are extended to dynamic uplift modeling. **Figure 26** shows the total number of cumulative purchases achieved by using each kind of uplift model with varying depth of treatment. It can be observed from the plot that there is a very steep increase in the number of purchases with increase in depth of treatment for unified uplift model. This is due to the inherent tree building structure of the unified uplift model. The double decision tree model performed inferior to random model because of repeatedly targeting same group of customers.

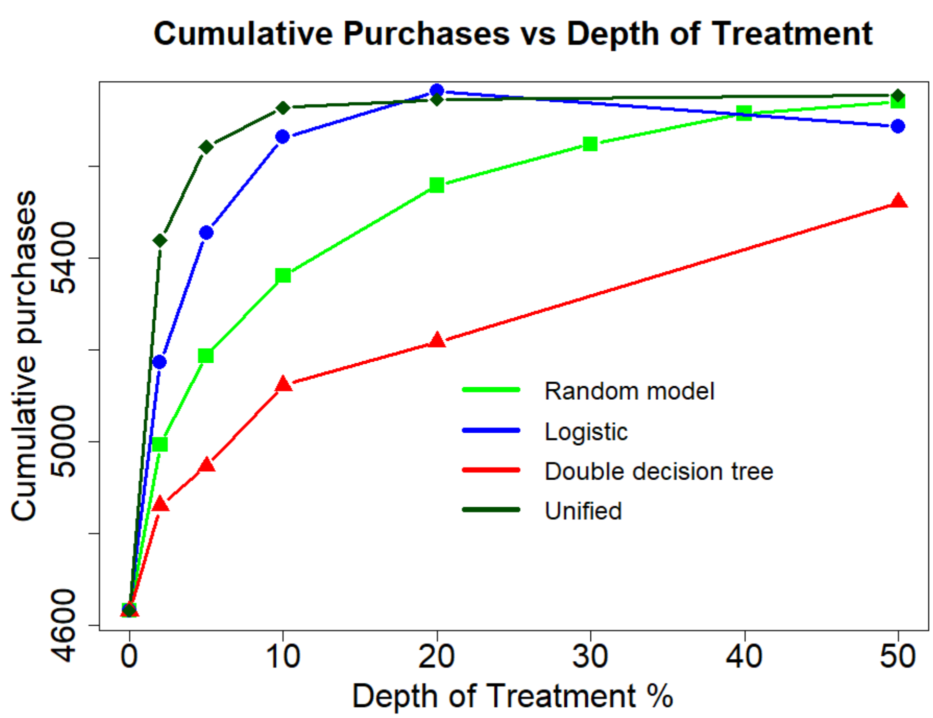


Figure 26: Comparison of three different approaches used in the study

The conclusions from this study are as follows

1. Simulation studies helps in understanding the higher-level behavior of the system based on the behavior of individual entities. The methodology used in this study can be used by retail companies to match the customer purchase data available with them by tuning the hyper parameters. Further tuned model can be used to predict the future purchases.
2. Dynamic Uplift modelling helps in targeting persuadable customers on time dependent basis. This methodology is an extension already available static uplift modeling approaches. In this study dynamic uplift modelling helps in increasing the total number of purchase in a fixed time period.
3. Among the three methodologies used in this study, unified uplift methodology exhibited highest efficiency in achieving maximum purchases with minimum number of promotional offers.

Future works in this area can concentrate on developing customized techniques for dynamic uplift modelling. Also, future works can concentrate on developing quantitative metrics for measuring the efficiency of dynamic uplift modelling.

References

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A., 1984. *Classification and regression trees*. CRC press.

Cham, H.N., 2013. *Propensity score estimation with random forests* (Doctoral dissertation, Arizona State University).

Chickering, D.M. and Heckerman, D., 2000, June. A decision theoretic approach to targeted advertising. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence* (pp. 82-88). Morgan Kaufmann Publishers Inc..

Csiszár, I. and Shields, P.C., 2004. Information theory and statistics: A tutorial. *Foundations and Trends® in Communications and Information Theory*, *1*(4), pp.417-528.

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

Guelman, L., 2014. uplift: Uplift Modeling. *R package version 0.3.5*. *http://CRAN.R-project.org/package=uplift.*

Guelman, L., Guillén, M. and Pérez-Marín, A.M., 2014. A survey of personalized treatment models for pricing strategies in insurance. *insurance: Mathematics and Economics*, *58*, pp.68-76.

Hansotia, B. and Rukstales, B., 2002. Incremental value modeling. *Journal of Interactive Marketing*, *16*(3), pp.35-46.

Kondareddy, S.P., Agrawal, S. and Shekhar, S., 2016, July. Incremental Response Modeling Based on Segmentation Approach Using Uplift Decision Trees. In *Industrial Conference on Data Mining* (pp. 54-63). Springer International Publishing.

Lee, L., 1999, June. Measures of distributional similarity. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 25-32). Association for Computational Linguistics.

Lo, V.S., 2002. The true lift model: a novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter*, *4*(2), pp.78-86.

Manahan, C., 2005. A proportional hazards approach to campaign list selection. In *Proceedings of the thirtieth annual SAS users group international conference (SUGI), Philadelphia, PA.*

Michel, R., Michel, R., Schnakenburg, I., Schnakenburg, I., von Martens, T. and von Martens, T., 2017. Effective customer selection for marketing campaigns based on net scores. *Journal of Research in Interactive Marketing*, *11*(1), pp.2-15.

Perret, O., 2016. A Novel Post-hoc Matching Procedure Using Statistical Learning Methods.

Radcliffe, N.J. and Surry, P.D., 1999. Differential response analysis: Modeling true response by isolating the effect of a single action. *Credit Scoring and Credit Control VI. Edinburgh, Scotland*.

Radcliffe, N.J., 2007. Using control groups to target on predicted lift: Building and assessing uplift models. *Direct Market J Direct Market Assoc Anal Council*, *1*, pp.14-21.

Radcliffe, N.J. and Surry, P.D., 2011. Real-world uplift modeling with significance-based uplift trees. *White Paper TR-2011-1, Stochastic Solutions*.

Rubin, D.B., 2005. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), pp.322-331.

Rzepakowski, P. and Jaroszewicz, S., 2012. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, *32*(2), pp.303-327.

Sołtys, M., Jaroszewicz, S. and Rzepakowski, P., 2015. Ensemble methods for uplift modeling. *Data mining and knowledge discovery*, *29*(6), pp.1531-1559.

Zaniewicz, Ł. and Jaroszewicz, S., 2013. Support vector machines for uplift modeling. *IEEE 1th International Conference on Data Mining Workshops.*

Zhao, Y., Fang, X. and Simchi-Levi, D., 2017, June. Uplift Modeling with Multiple Treatments and General Response Types. In *Proceedings of the 2017 SIAM International Conference on Data Mining* (pp. 588-596). Society for Industrial and Applied Mathematics.

Appendix A: Nomenclature

τ_i	Uplift in Purchase Probability Rubin (1974)
CATE	Conditional Average Treatment Effect
N_a	Number of always buying customers in the study group
N_p	Number of Persuadable customers in the study group
$N_{\bar{p}}$	Number of Anti Persuadable customers in the study group
$N_{\bar{a}}$	Number of Never buying customers in the study group
c	Cost of Advertisement
N	Total Number of customers in the study group
r_u	Revenue from a normal sale
r_s	Net Revenue from a sale with promotional offer
$\tau_{revenue}$	Uplift in Revenue
P^T	Purchase probability with an offer
P^C	Purchase probability without offer
$U[0, x)$	Uniform distribution with bounds 0 and x
N_{TR}	Number of treatment instances in right child node
N_{TL}	Number of treatment instances in left child node
N_{CR}	Number of control instances in right child node
N_{CL}	Number of control instances in left child node
SSE	Sum of squared error
U_R	Uplift in right child node
U_L	Uplift in left child node
$KL(P:Q)$	Kullback-Leibler divergence between two distributions P and Q
$E(P:Q)$	Euclidian divergence between two distributions P and Q
$\chi^2(P:Q)$	Chi-square divergence between two distributions P and Q
$D_{gain}(A)$	Increase in divergence based on split on predictor A
B	Number of trees in ensemble model
ϕ	Feature space
ϕ_l	Feature space of left split

ϕ_r	Feature space of right split
A_t, B_t, A_c, B_c	Constants for converting addiction and loyalty scores into factors ranging between (0,1)
$M_{(i,j)}$	Customer attributes Matrix
C_j	Control Weights Vector
$T_{(j,1)}$	Treatment Weights Vector for Offer-1
$T_{(j,2)}$	Treatment Weights Vector for Offer-2
$T_{(j,3)}$	Treatment Weights Vector for Offer-3
$W_{(j,k)}$	Control and Treatment Weights matrix
$SPO_{(i,1)}$	Static Control Purchase Odds
$SPO_{(i,2)}$	Static Treatment Purchase Odds for Offer-1
$SPO_{(i,3)}$	Static Treatment Purchase Odds for Offer-2
$SPO_{(i,4)}$	Static Treatment Purchase Odds for Offer-3
$SPP_{(i,1)}$	Static Control Purchase Probability
$SPP_{(i,2)}$	Static Treatment Purchase Probability
$LS_{i,t}$	Loyalty Score of Customer i at time t
$AS_{i,t}$	Addiction Score of Customer i at time t
PCT_i	Purchase Cycle Time of Customer I
TF	Time Factor
$DCP_{i,t}$	Dynamic Control Probability of Customer i at time t
RPT_i	Recent Purchase Time of Customer i
$DTP_{i,t}$	Dynamic Treatment Probability of Customer i at time t