Enhancing Robustness of Uplift Models used for Churn Prevention against Local Disturbances

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften

(Dr.-Ing.)

von der KIT-Fakultät für
Wirtschaftswissenschaften
am Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Math. oec. Frank Oechsle

Referent: Prof. Dr. Thomas Setzer
Korreferent: Prof. Dr. York Sure-Vetter
Tag der mündlichen Prüfung: 05.12.2022

Karlsruhe, 2022

Für Romy und Papa

List of included papers

Paper A

F. Oechsle, T. Setzer, and S. M. Blanc, “On the assumptions of true lift models for churn prevention”, in Multikonferenz Wirtschaftsinformatik (MKWI)2016, Technische Universität Ilmenau, 09.-11. März 2016, Band 2. Hrsg.: V. Nissen, pp. 1233–1244, Universitätsverlag Ilmenau, 2016

Paper B

F. Oechsle and D. W. Schönleber, “Towards more robust uplift modeling for churn prevention in the presence of negatively correlated estimation errors”, in Proceedings of the 53rd Hawaii International Conference on System Sciences, pp. 1562–1569, 2020.

Paper C (in submission)

F. Oechsle, "Increasing the robustness of uplift modeling using additional splits and diversified leaf select", submission to Journal of Marketing Analytics (ISSN: 2050-3318 (print); ISSN: 2050-3326 (electronic); Journal no.: 41270)

Contents

Part I.

Overview

Chapter 1.

Introduction

1.1. Motivation and Research Goal

The amount of generated digital data rises year over year according to an exponential growth (Jan et al., 2019). In parallel the availability of suitable hardware for processing data steadily increases. Computing power and storage capacity improve rapidly (Foerster-Metz et al., 2018). Simultaneously the legal framework gets updated. After more than 20 years of no changes in regulations as of May 25, 2018 the European so called General Data Protection Regulation went into effect. This regulation replaces the former Data Protection Directive of 1995. It was passed by the European Parliament as a contemporary legal framework appropriate to the digital age. To summarize there is an environment, which boosts the development and importance of analytical customer relationship management in the business-to-consumer segment even further.

On the other hand a quite large toolbox is available for data driven decision-making concerning several business-to-consumer use cases. There are established methods for tackling diverse challenges along the customer lifecycle especially in industries with contractual relations. It begins with the acquisition of new customers, leads to cross and upselling or customer experience and finally to churn management. At the majority of these central moments the necessary tools are available for practice, especially prediction modeling techniques. In sales for example there are custom response models for the prediction of purchasing probabilities. Oftentimes the more elaborated uplift modeling approaches are already applied. But when it comes to churn uplift concepts either do not reliably achieve the desired impact.

So far no best practice for customer churn prevention measures has evolved based on predicted churn probabilities, neither via response models nor via uplift models. Frequently churn prevention campaigns even increase churn (Radcliffe, 2007b; Ascarza, 2018), which is exactly the access point for this thesis. The goal of the thesis is correspondingly to point out what the root causes are for this observation. In a second step it aims on developing solutions for robust, reliably churn decreasing, churn prevention actions for practice.

1.2. Business Context Churn

Churn broadly means the loss of customers, which can be the non-appearence of anew purchases or the churn announcements of consisting contracts (Pejić Bach et al., 2021). For the first variant no contractual relation is needed. The second variant is the classic churn use case in the subscription economy.

As attested by typical empirical churn ratios, churn is a relatively seldom event. In the telecommunications sector for instance those quotas are in the region of up to 3.5% per month (Verbraken et al., 2014). 3.5% per month in a short term predictive modeling perspective represent a challenging imbalance of the sample and a significantly reduced random hit rate of customers who are likely to churn. But in the long term this circumstance is more than sufficient to make churn a commercially important topic for companies, since over the course of one year this ratio per month leads to up to 40% customers who are likely to leave.

Churn management covers measures to reduce customer churn. Apart from actions that serve the general customer satisfaction, there are two churn management disciplines. On the one hand churn prevention and on the other hand customer retention. Prevention means a prophylactic avoidance of churn, while retention stands for winning back customers after cancellation or at least cancellation announcement. Prevention thereby is the methodically more challenging part, since the cancellation has to be anticipated and respectively predicted. In the retention case the customer triggers the start of the action. The thesis focuses on the prevention case.

1.3. Literature Review

There are several research papers which lay the foundation for uplift modeling concepts while they are considering the underlying probability estimation problem of two not simultaneously observable events (Chickering and Heckerman, 2000; Hansotia and Rukstales, 2002; Lo, 2002). Within those papers the work of Lo (2002) particularly attracted interest. His "True Lift Model" is the starting point for this thesis. Regardless of the detailed concept, it is a matter of estimating a difference between two probabilities. Those concepts can essentially be classified into two categories: 1.) methods that estimate the target uplift directly within one single prediction model and 2.) approaches using two distinct models to separately estimate the likelihoods which are then combined to uplift. The direct approach is the superior one (Radcliffe and Surry, 2011; Zhao et al., 2017). While the theoretical concepts may be very clear and straight forward, in practice track records are missing. Current research takes account of this fact and addresses towards uplift modeling first and foremost disturbance, noise, estimation errors or similar priorities arising from practical application. A main difficulty of research on that is the non-availability of publicly accessible real world datasets. This thesis examines the causes for uncertainty like disturbance, noise and others. It models them and evaluates the impact depending on several customer selection methods for churn prevention campaigns on a real world dataset. The thesis therefore works exactly on the current challenges of uplift modeling research. Those obstacles are accentuated by further barriers, which as described come along with the field of churn.

Chapter 2.

Summary of the included papers

Table 2.1.: Connection of included papers

	Paper A	**Paper B**	**Paper C**
Business context	Churn	Churn	Churn
Uplift modeling	direct & indirect	direct	direct
Used dataset	simulated (300k obs.)	simulated (225k obs.)	real world (64k obs.)
Estimation of Uplift	none	geometrically designed	via decision tree
Feature space	none	2 dimensions	7 dimensions
Noise & disturbance	in general	local spatial errors (circular, 3 radii)	local spatial errors (circular, 8 radii)
Selection methods	none	5 (excl. additional splits)	6 (incl. additional splits)
Simulation	none	Monte Carlo (250 runs per radius)	Monte Carlo (1000 runs per radius)

2.1. Connection

The overlap of the three listed papers mainly consists of the underlying business challenge. All of them are triggered and motivated by the nonexistent best practice in the churn management subdiscipline prevention. As distinguished from the second subdiscipline retention there is no sample solution available for churn prevention via predictive modeling. Namely it exists no proven method which reliably accounts for achieving the business goal of churn minimization. That applies to both uplift modeling techniques and especially response modeling approaches. The three papers highlight fundamental assumptions on which current concepts of uplift modeling are based on (esp. paper A), debate alternative ideas based on decision trees (esp. paper B) and show by using real world data that those novel approaches in error-prone scenarios are more promising (esp. paper C).

Paper A builds the foundation through basal thoughts towards the inherent assumptions of uplift modeling approaches for the considerations following in paper B and paper C. So far there is no specific focus on either direct or indirect uplift modeling as common concepts. It rather provides a general view on potential root causes for estimation errors, without an indeed operated uplift estimation.

Paper B deals with the suggestions of paper A and transfers them to an approach containing local spatial estimation errors, circular with three different radii. It undertakes a quantitative analysis via Monte Carlo simulation on a synthetically generated dataset with two dimensions and 225.000 observations. Therefore alternative distance respecting selection methods are introduced and first indicators are carved out, showing that those novel methods could be able to dominate the classic method. Preliminary the used uplift estimations are constructed by geometric design.

Lastly paper C follows up on paper B and enhances it with another five radii and an extra selection method which uses additional splits. It particularly includes an evaluation on a 64.000 observations containing real world dataset by means of a Monte Carlo simulation with four times the number of runs. Moreover the employed uplift estimations are engineered by decision tree on this real world dataset using seven features.

2.2. Paper A: On the assumptions of true lift models for churn prevention

This research focuses on missing success stories of churn prevention in a very basic manner as it fathoms and challenges the assumptions underlying the current uplift modeling approaches. The issue of estimating the treatment effect of a customer relationship management (CRM) campaign (particularly churn) is constituted with the help of Lo (2002) and his "True Lift Model", while simultaneously the concept itself is introduced. Additionally the typical nomenclature is specified (CRM, churn, prevention, retention) and the existence of two only asynchronously observable events (customer reaction when treated, customer reaction when not treated, uplift as their difference) is identified as nonsolvable constraint which complicates the estimation of the two corresponding probabilities.

Finally four groups of customers are defined according to their churn probability characteristics, following the thoughts of Radcliffe and Simpson (2008). That is the Persuadables which should be addressed, the Sleeping Dogs who are not to be addressed, the Lost Causes and the Sure Things. The Persuadables are the customers who can be convinced by the campaign or significantly reduce their churn probabilities with treatment. In contrast to that the campaign addressing the Sleeping Dogs provokes churn, that is their churn probabilities significantly increase when treated. Referring to this the paper also depicts that wrong decisions most likely come along with highly detrimental consequences since every addressed Sleeping Dog implies a potential churn increase.

The paper mentions both known approaches of uplift modeling, which are the direct estimation of uplift as well as the separate calculation of two models and a subtraction of probabilities afterwards. In doing so it falls into place, that the academic foundation is comprehensible and the selection method with respect to the True Lift concept is theoretically straightforward. But the adjacent considerations of the paper indicate that the underlying assumptions could be not met in real world business scenarios, which is described consolidatingly as a nonobservance of the established bias-variance tradeoff. The thereby emerging and ignored uncertainty is characterized as the decisive argument for the non-existent success stories.

Lastly it is concluded that future approaches for uplift modeling and particularly for churn prevention measures should factor in the contempt of the bias-variance tradeoff and the interconnected disturbance. The paper builds the foundation for further investigations regarding advanced uplift models. At the same time it provides a first stimulus towards game-changing events through local spatial errors by referencing Manahan (2005) and his implicit suggestion to incorporate offer attractiveness as predictor. Paper B, as described in the next subsection, then seizes this impulse.

2.3. Paper B: Towards more robust uplift modeling for churn prevention in the presence of negatively correlated estimation errors

Paper B approaches the importance of churn management and highlights churn prevention procedures through a subscription business perspective. It accentuates that despite of the tremendous growth of the subscription economy, reliable uplift modeling concepts for churn prevention are still missing. In principle the research as well takes the "True Lift Model" of Lo (2002) as a basis and defines uplift as the difference of the churn probabilities dependent on the (non-) participation in a corresponding campaign. Backed up with sources (Radcliffe and Surry, 2011; Zhao et al., 2017), it declares the direct prediction as the superior method and thus disregards the indirect prediction via two separate models. The paper is inspired by recent research regarding estimation errors and primarily paper A. Besides the per se existing issue of predicting two not at once observable events, inconsistent estimations elicited by game-changing events are mentioned as major reason for futile churn prevention uplift models. In the worst case scenarios they even boost churn. Cited examples of game-changing events that affect particular customer segments, are price increases, product migrations or tariff launches of competitors.

Emanating from a two-dimensional feature space those events are assumed to be local and spatially bounded in a circular shape distinguished by an error seed and an error radius. The impact of a specific event fades with cumulative distance to the error seed until it vanishes once the distance excels the error radius. Accordingly in a next step spatial neighbourhood of customers in the feature space is depicted as fraught with risk when it comes to customer selection. Concretely it is described that oftentimes similar customers in terms of feature values, which defines spatial neighbourhood, collectively are selected assuming that they possess appropriate churn probability uplifts. A later occurrence of a spatial error could then be devastating whether it locates in the selected neighbourhood. As the paper explains in an adverse setting plenty of customers who are affected by negatively correlated estimation errors would take part in the corresponding churn prevention campaign.

In the evaluation chapter of the research the painted situation is quantified with the help of a Monte Carlo simulation. Therefore the two-dimensional feature space is artificially subdivided in nine rectangles, and consequently in leaves of a decision tree, once regularly once randomly. Afterwards values $\Delta_i \in (0;1]$ for $i = 1,2,3$ and $\Delta_i \in$ [-1; 0] for $i = 4, 5, .., 9$ are randomly assigned to the rectangles and accepted to be the correct estimations of uplifts belonging to the corresponding leaves and the contained customers. Next several distance respecting selection methods are introduced for challenging the classic selection approach which concentrates exclusively on uplift and ignores distance in the feature space. In a completive step the discussed local spatial and more precisely circular errors are created in order to retroactively bias the correct and thus applied for selection uplift estimations. The performance of the different selection approaches is evaluated based on 250 runs per error radius embedded in a Monte Carlo simulation.

With regard to the results of the carried out simulations it becomes apparent in this research that in an as per description disadvantageous scenario and decision trees with regular leaves it can be promising to use selection methods which trade off uplift against distance in the feature space. It may be profitable not to follow the classic method depending on the error radius. Practically this is documented by lowered probabilities for ending up with churn increasing churn prevention campaigns. Encouraged by this insight the paper finally motivates to think about additional splits man-made integrated in the underlying decision tree with the aim to gain more distance amongst the selected customers. Indeed this is just what is implemented in paper C on a real world dataset.

2.4. Paper C: Increasing the robustness of uplift modeling using additional splits and diversified leaf select

Paper C addresses the subject via the current Covid crisis and the thereby furthermore taking place growth acceleration of the eCommerce and subscription business. The considered business sector is again churn, whereat churn explicitly is defined as not only cancellation of subscriptions but also not continuing online purchases. Accordingly the general increment of subscriptions and online purchases in sum is mentioned as a factor which increases the importance of churn management and especially churn prevention. Consistent with papers A and B, this paper has to state that notwithstanding the current macroeconomic trend there still is no robust uplift modeling solution for churn prevention. Similarly the conventional impediments like two not at once observable rare events or additional churn provoking errors are listed.

For the basic setup the concept of paper B is adopted. That is a) usage of True Lift approach, b) focus on direct uplift modeling, c) intervention in a previously generated decision tree which is assumed to be able to correctly estimate churn uplifts, d) occurrence of local spatial and circular errors of same architecture characterized by error seed, error radius and distance of the individual customer to the error seed and e) concluding evaluation via Monte Carlo simulation. This procedure is motivated by the latest research on uplift modeling which predominantly examines the robustness of estimations depending on disturbance and in particular by paper B.

Compared to paper B this research and therefore a substantial portion of its contribution consists of a) the upgrading of the selection methods with a new diversifying portfolio approach by means of additional splits, b) a more extensive Monte Carlo Simulation (8 radii, 1000 runs pro radius), c) a rising number of dimensions (from two to seven) spanning the feature space and d) the usage of the publicly available real world Hillstrom dataset. This very dataset upfront is fitted to a churn scenario and enables the paper to implement an uplift estimation per decision tree. Coming from this estimation the known methodology of paper B is adapted by employing the noted modifications. Especially the missing research on the basis of real world datasets is depicted as a contemplated central problem in the literature towards uplift modeling.

The present study concludes that depending on the magnitude of bias and disturbance situations exist in which it is necessary or at least recommendable not to comply with the classic selection approach of uplift modeling. At the same time it is carved out that the diversifying portfolio approach including additional splits produces desired outcomes. This circumstance is quantified through lower spread of average uplift per selected customer, less failures (churn aggravating churn prevention campaigns) and less grave failures. Though the classic selection method in the contained evaluations permanently achieves the highest expected uplift $E[\Delta]$, the add split method partly demands a marginal risk premium in terms of $E[\Delta]$ for the obtained robustness and planning certainty. The research suggests to apply this knowledge gain in practice and perceives that it is a valuable instrument for particularly eCommerce and subscription business use cases.

Chapter 3.

Conclusion

3.1. Summary of Contributions

The thesis questions the assumptions of current uplift modeling approaches and hence reveals weaknesses, which can arise in practice. It formalizes the problem of noise, disturbance, estimation errors or dependency on behaviour of local neighbours as mentioned by numerous research papers (Dasgupta et al., 2008; Kusuma et al., 2013; Droftina et al., 2015a,b; Athey et al., 2015; Lo and Pachamanova, 2015; Oechsle et al., 2016; Athey and Imbens, 2016; Zhao et al., 2017; Rößler et al., 2021) via its spatial error concept. This research introduces new selection methods, which deviate from the classical approach. They do not only consider expected churn probabilities but also the spatial distance in the feature space as decision criterion. The superior selection method with incorporated additional splits can be interpreted as an intervention in the construction of the underlying decision tree. Its dominance in settings sufficiently affected by errors, is clearly determined on Hillstroms real world dataset. The results show a reduced likelihood of churn increasing churn prevention campaigns in relevant error-prone situations when not sticking to the classic method, especially with the additional split variant. Moreover the less frequent failures are of lower extent and the risk premium is acceptable. The thesis thus supplies a tool for practicioners to develop their churn prevention measures to more robustness. Obviously users from other industries can adapt the insights as well.

3.2. Critical Appraisal and Outlook

An arguable assumption of this thesis and respectively of the three included papers is the concept of radially spreading local errors, following a cosine behaviour. While other research supports the idea of locally bounded errors (Dasgupta et al., 2008; Kusuma et al., 2013; Droftina et al., 2015a,b), the cosine behaviour is disputable. This manner is not inviolably deduced but follows the intuition that the error impact should fade from the seed to the edge. However it can be said that most probably the spatial confinement is way more important than the actual way of diffusion within the error area.

Another downside associated with the error construction concept is that the errors are simulated in retrospect. Even though this is then applied to a real world dataset, the generation of errors is simulated. To compensate this a real world dataset with observations at all necessary time points would be desirable. This aspect is one part of the frequently mentioned need for real world data and supports its relevance for future research. The used data sets can obviously be more extensive than the 64k sample used by Hillstrom, as indicated by Diemert et al. (2018), or originate directly from the subscription business.

Concerning the used selection methods the application of a new metric regarding the Euclidean distance should be considered to evaluate if comparable or even better results can be achieved. Even though using Euclidean distance was successful in the current thesis, there are no arguments against considering different metrics in future research.

Finally this research solely focuses on problem solving by using decision trees. Neither random forests as a combination of several decision trees nor completely different machine learning methods are evaluated. This is not mandatory since the goal is reached by the presented approach, but it should still be used as connection point for potential future successes in the area of uplift modeling.

Bibliography

Ahn, J.-H., Han, S.-P., and Lee, Y.-S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the korean mobile telecommunications service industry. *Telecommunications policy*, 30(10-11):552–568.

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1):80–98.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. In *Proceedings of the National Academy of Sciences*, volume 113, pages 7353–7360. National Academy of Sciences.

Athey, S., Imbens, G., and Ramachandra, V. (2015). Machine learning methods for estimating heterogeneous causal effects.

billwerk GmbH (2019). Subscription based services.

Chen, T., Fenyo, K., Yang, S., and Zhang, J. (2018). Thinking inside the subscription box: New research on e-commerce consumers. *McKinsey & Company*, pages 1–9.

Chickering, D. M. and Heckerman, D. (2000). A decision theoretic approach to targeted advertising, uai'00 proceedings of the 16th conference on uncertainty in artificial intelligence. *San Francisco, CA, USA: Morgan Kaufmann Publishers Inc*, 82:88.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., and Joshi, A. (2008). Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, pages 668–677.

Devriendt, F., Berrevoets, J., and Verbeke, W. (2021). Why you should stop predicting customer churn and start using uplift models. *Information Sciences*, 548:497–515.

Diemert, E., Betlei, A., Renaudin, C., and Amini, M.-R. (2018). A large scale benchmark for uplift modeling. In *KDD*.

Diemert, E., Betlei, A., Renaudin, C., Amini, M.-R., Gregoir, T., and Rahier, T. (2021). A large scale benchmark for individual treatment effect prediction and uplift modeling. *arXiv preprint arXiv:2111.10106*.

Droftina, U., Štular, M., and Košir, A. (2015a). A diffusion model for churn prediction based on sociometric theory. *Advances in Data Analysis and Classification*, 9(3):341–365.

Droftina, U., Štular, M., and Košir, A. (2015b). Predicting influential mobile-subscriber churners using low-level user features. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 56(4):522–534.

Foerster-Metz, U. S., Marquardt, K., Golowko, N., Kompalla, A., and Hell, C. (2018). Digital transformation and its implications on organizational behavior. *Journal of EU Research in Business*, 2018(S 3).

Gallo, A. (2014). The value of keeping the right customers. *Harvard business review*, 29(10):304–309.

Gu, S., Ślusarczyk, B., Hajizada, S., Kovalyova, I., and Sakhbieva, A. (2021). Impact of the covid-19 pandemic on online consumer purchasing behavior. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(6):2263–2281.

Guelman, L., Guillén, M., and Pérez-Marín, A. M. (2015). Uplift random forests. *Cybernetics and Systems*, 46(3-4):230–248.

Hadden, J., Tiwari, A., Roy, R., and Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10):2902–2917.

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

Hillstrom, K. (2008). The minethatdata e-mail analytics and data mining challenge. *MineThatData blog*.

Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., Ali, S., and Jeon, G. (2019). Deep learning in big data analytics: a comparative study. *Computers & Electrical Engineering*, 75:275–287.

Kane, K., Lo, V. S., and Zheng, J. (2014). Mining for the truly responsive customers and prospects using true-lift modeling: Comparison of new and existing methods. *Journal of Marketing Analytics*, 2(4):218–238.

Keramati, A. and Ardabili, S. M. (2011). Churn analysis for an iranian mobile operator. *Telecommunications Policy*, 35(4):344–356.

Kim, H.-S. and Yoon, C.-H. (2004). Determinants of subscriber churn and customer loyalty in the korean mobile telephony market. *Telecommunications policy*, 28(9-10):751–765.

Kusuma, P. D., Radosavljevik, D., Takes, F. W., and van der Putten, P. (2013). Combining customer attribute and social network mining for prepaid mobile churn prediction. In *Proc. the 23rd Annual Belgian Dutch Conference on Machine Learning (BENELEARN)*, pages 50–58. Citeseer.

Lai, L. Y.-T. (2006). *Influential marketing: a new direct marketing strategy addressing the existence of voluntary buyers*. PhD thesis, School of Computing Science-Simon Fraser University.

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):78–86.

Lo, V. S. and Pachamanova, D. A. (2015). From predictive uplift modeling to prescriptive uplift analytics: A practical approach to treatment optimization while accounting for estimation risk. *Journal of Marketing Analytics*, 3(2):79–95.

Lu, J. (2002). Predicting customer churn in the telecommunications industry—-an application of survival analysis modeling using sas. In *SAS User Group International (SUGI27) Online Proceedings*, volume 114.

Manahan, C. (2005). A proportional hazards approach to campaign list selection. *SAS User Group International (SUGI) 30 Proceedings*.

Ngai, E. W., Xiu, L., and Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602.

Oechsle, F. and Schönleber, D. (2020). Towards more robust uplift modeling for churn prevention in the presence of negatively correlated estimation errors. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Oechsle, F., Setzer, T., and Blanc, S. M. (2016). On the assumptions of true lift models for churn prevention. In *Multikonferenz Wirtschaftsinformatik (MKWI) 2016, Technische Universität Ilmenau, 09.-11. März 2016, Band 2. Hrsg.: V. Nissen*, pages 1233–1244. Universitätsverlag Ilmenau.

Pejić Bach, M., Pivar, J., and Jaković, B. (2021). Churn management in telecommunications: Hybrid approach using cluster analysis and decision trees. *Journal of Risk and Financial Management*, 14(11):544.

Radcliffe, N. (2007a). Generating incremental sales: Maximizing the incremental impact of cross-selling, up-selling and deep-selling through uplift modelling. *Stochastic Solutions*, pages 1–10.

Radcliffe, N. (2007b). Using control groups to target on predicted lift: Building and assessing uplift model. *Direct Marketing Analytics Journal*, pages 14–21.

Radcliffe, N. J. (2008). Hillstrom's minethatdata email analytics challenge: An approach using uplift modelling. *Stochastic Solutions Limited*, 1:1–19.

Radcliffe, N. J. and Simpson, R. (2008). Identifying who can be saved and who will be driven away by retention activity. *Journal of Telecommunications Management*, 1(2).

Radcliffe, N. J. and Surry, P. D. (2011). Real-world uplift modelling with significance-based uplift trees. *White Paper TR-2011-1, Stochastic Solutions*, pages 1–33.

Rößler, J., Tilly, R., and Schoder, D. (2021). To treat, or not to treat: Reducing volatility in uplift modeling through weighted ensembles. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, page 1601.

Rzepakowski, P. and Jaroszewicz, S. (2010). Decision trees for uplift modeling. In *2010 IEEE International Conference on Data Mining*, pages 441–450. IEEE.

Rzepakowski, P. and Jaroszewicz, S. (2012). Uplift modeling in direct marketing. *Journal of Telecommunications and Information Technology*, pages 43–50.

Shaar, A., Abdessalem, T., and Segard, O. (2016). Pessimistic uplift modeling. *arXiv preprint arXiv:1603.09738*.

Tamaddoni Jahromi, A., Sepehri, M. M., Teimourpour, B., and Choobdar, S. (2010). Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. *Journal of Strategic Marketing*, 18(7):587–598.

Verbraken, T., Verbeke, W., and Baesens, B. (2014). Profit optimizing customer churn prediction with bayesian network classifiers. *Intelligent Data Analysis*, 18(1):3–24.

Zaniewicz, Ł. and Jaroszewicz, S. (2013). Support vector machines for uplift modeling. In *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 131–138. IEEE.

Zhang, X., Zhu, J., Xu, S., and Wan, Y. (2012). Predicting customer churn through interpersonal influence. *Knowledge-Based Systems*, 28:97–104.

Zhao, Y., Fang, X., and Simchi-Levi, D. (2017). Uplift modeling with multiple treatments and general response types. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 588–596. SIAM.

zuora Inc. (2020). Subscription economy index.

Part II.

Papers A to C

Paper A: On the Assumptions of True Lift Models for Churn Prevention

On the Assumptions of True Lift Models for Churn Prevention

Frank Oechsle[1], Thomas Setzer[2], and Sebastian M. Blanc[2]

[1] 1&1 Internet SE, Karlsruhe, frank.oechsle@1und1.de

[2] Karlsruhe Institute of Technology, Institute of Information Systems and Marketing, thomas.setzer|sebastian.blanc@kit.de

Abstract

Preventing customer churn by subjecting carefully selected customers to customer relationship management activities is of crucial importance to many service industries. A promising selection of customers can be achieved using so called true lift or incremental models, which focus on customers at high churn risk, that are also likely to be persuadable through appropriate campaigns. In comparison to simpler models, true lift modeling however not only requires estimating churn probabilities of untreated customers but also their churn probabilities when treated. We argue that the estimation of the latter probabilities introduces a novel source of uncertainty not considered in state-of-the-art true lift models. In this paper, we assess the consequences of these uncertainties for true lift modeling. We identify assumptions regarding distribution of churn probabilities made by true lift models and argue that these assumptions are most likely not met in any practical setting. As a result, churn prevention campaigns can easily fail and even increase total churn rate, which might provide an explanation for the very few published empirical success stories on true lift models.

1 Introduction

A broad spectrum of scientific literature discusses customer relationship management (CRM) along the lifecycle of a typical customer, including customer identification, attraction, development, churn prevention and retention (Ngai et al. 2009). Prevention as well as retention campaigns aim at reducing the number of contract cancellations. In contrast to retention, which aims at winning the customer back after announcement, prevention precedes the churn announcement and aims at reducing churn announcement probabilities. In the telecommunications sector with annual churn rates estimated at up to 20% and more (Tamaddoni Jahromi et al. 2010) and low marginal costs per customer, churn prevention is of crucial importance. Numerous works studying the determinants of churn risk (for instance Kim and Yoon 2004, Ahn et al. 2006, Keramati and Ardabili 2011, Lu 2002, Zhang et al. 2012) highlight its importance. See Hadden et al. (2007) for a comprehensive overview on churn management.

A key question in customer churn management is which customers to address in CRM activities aiming at churn prevention. Based upon churn probability with and without treatment, customers can be grouped into segments of different relevance, as illustrated in Figure 1. Customers for whom the churn probability can be noticeably reduced by a treatment (referred to as *Persuadables*) are of great interest and should be subject to treatment. Obviously, a treatment can also lead to the opposite effect for other customers labeled *Sleeping Dogs*, who should not be subject to CRM activities. Customers with approximately equal churn probability with and without treatment (*Lost Causes* and *Sure Things*) are not of particular interest for churn prevention activities.

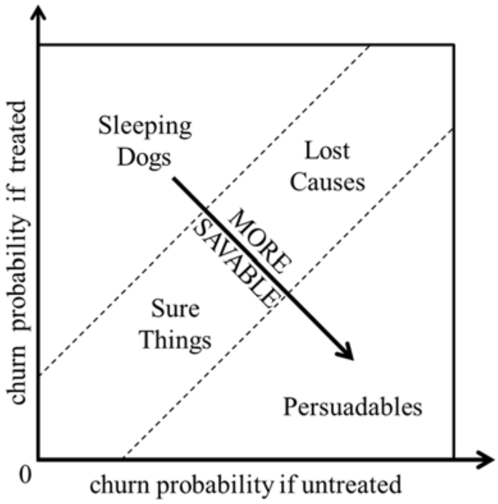

Figure 1: Definition of Customer Segments according to Radcliffe and Simpson (2007)

The aim of true lift models is consequently to select the *Persuadables* for treatment while avoiding to select *Sleeping Dogs*. To discriminate the customers of both segments, it is intuitive to use the difference of the churn probabilities with and without treatment.

In practice, the churn risks of individual customers (with and without treatment) are predicted using data mining techniques. However, while it has been shown that churn probability predictions of untreated customers are quite accurate (like other predictions of probabilities using data mining techniques), targeting the customers who can be positively influenced by a treatment has proved to be much more challenging and only few success stories have been reported on this important topic.

In this paper, we argue that the low number of success stories might be a result of the implicit assumptions of true lift models, which are likely to be violated in practice. In particular, we demonstrate that true lift models assume that (a) the uncertainty in churn probability estimates is equal for all probabilities and that (b) estimates of churn probabilities with and without treatment are equally distributed and uncorrelated.

Both assumptions can easily be violated since customers with high churn probability are much less frequent than customers with low churn probability. In fact, customers with high churn probabilities must be seldom, since simply because they have high churn probabilities, it is likely that they already have churned. As a result, the number of customers available for probability estimation decreases with churn probability ascending, which in turn increases uncertainty in churn probability estimates (as the statistical support for probability estimates decreases).

Because of this uncertainty, a *Sleeping Dog* can easily be misclassified as a *Persuadable*. This is even more likely if churn probabilities are on average larger with than without treatment.

The remainder of the paper is structured as follows. We first review related work on true lift modeling. We then identify the implicit assumptions underlying true lift models in Section 3. These assumptions are then compared to conditions in real world applications in Section 4, and effects of violating the assumptions are discussed in Section 5. We finally conclude and debate our results as well as their implications for future true lift modeling efforts.

2 Related Work

The concept of true lift was first introduced by Lo (2002), who defined the true lift as a novel measure of campaign effectiveness. The true lift is based on the idea that a method of selecting customers for treatment in a marketing campaign should not only increase the probability of a desired outcome (e.g. a sale or a prevented churn) but must also outperform a random selection. This is illustrated in Table 1, where campaign results for a segmentation suggested by a model are presented in the *Model* row. In contrast, results for a random selection are displayed in the *Random* row. In both cases, the *Treatment* (*Control*) column indicates results for the customers that are (not) subject to the campaign. Consequently four groups with different cumulative responses (denoted A, B, C and D) exist. A customer selection method should increase the increment A-B, i.e., lead to better results for treated compared to untreated customers, and simultaneously outperform a random selection. Overall, true lift models are aimed at maximizing (A-B)-(C-D), whereas classic response models focus on maximizing A-C.

	Treatment	Control	Increment
Model	A	B	A-B
Random	C	D	C-D
Delta	A-C	B-D	(A-B)-(C-D)

Table 1: Definition of True Lift, following Lo (2002)

The concept of the true lift was motivated by Chickering and Heckerman (2000), who first not only modeled the expected response of a treatment. For an advertisement campaign for MSN subscriptions with 110,000 customers, two separate models for the expected profit with and without sending of a mail were built. The approach however only slightly outperformed an off-the shelf response model. In contrast, Hansotia and Rukstales (2002) found that increments of the response probabilities could be predicted with good accuracy in a direct marketing campaign with 282,277 customers of a major retailer.

In a simulation study, Lo (2002) showed that a standard response model merely marginally outperforms a true lift model regarding the response rate rank order. True lift models however clearly perform best with respect to the true lift. For example the top decile of the response (true lift) model generates a treatment response rate of 0.93 (approximately 0.7) whereas, in the same decile, the response rate difference to the untreated group is roughly 0.3 (0.41).

For the wireless telecommunications industry, Manahan (2005) aimed to reduce customer loss by contacting selected customers with a contract renewal offer. The model (a logistic model with cubic splines) did however not perform well, which the authors attributed to missing predictors, such as regarding the attractiveness of the offer.

Overall, empirical results for true lift models are assorted and did not robustly lead to satisfying results across different works. This result is quite surprising in the light of the intuitive and promising theoretical foundation of true lift models. Understanding the issues with true lift models leading to decreased performance is consequently of great importance.

Improving the predictive accuracy of the applied prediction models is an obvious starting point for increasing the performance of true lift models. For instance Rzepakowski and Jaroszewicz (2010) proposed technical modifications to decision trees for better performance in uplift modeling, which are found to be beneficial for the selection of patients for medical treatments.

Furthermore, Radcliffe and Surry (2011) noted that previous works predicted the treated and untreated probabilities separately and proposed to directly predict the difference between the probabilities in one single model. Based on case studies illustrating the effectiveness of the approach, the authors recommend this model as the superior approach. Rzepakowski and Jaroszewicz (2012) applied this approach in an email campaign with 64,000 customers using a special tree-based classifier. The model outperformed classic response models as well as common uplift models. Similarly, Zaniewicz and Jaroszewicz (2013) surpassed other uplift models (with decision trees and standard support vector machines) using a support vector machine in a medical scenario.

In summary, customers are selected in a way to optimize the difference A-B (see Table 1) in all models. The difference between probabilities with and without treatment is consequently a reasonable basis for the selection. However, empirical results are mixed when the probabilities are predicted separately. While technical improvements and directly predicting the difference of probabilities in one single model increased performance, the surprisingly small performance increase of basic models is still unexplained.

In this paper, we investigate the assumptions underlying prevalent true lift models, which were implicitly assumed to be satisfied to date, as a cause of the astonishingly low empirical performance of true lift models. The new insights provide sound guidance for future research on true lift models, which afterwards can be applied with higher performance in practice.

3 Assumptions of Current True Lift Models

Reconsidering the illustration of Radcliffe and Simpson (2007) in Figure 1, the common selection method using the difference between probabilities, starts choosing customers in the lower-right corner (*p_untreated* = 1 and *p_treated* = 0), where the increment (*p_untreated* - *p_treated*) is maximal. If customers with lower increments (denoted Δ) are also chosen, the lines separating selected from not selected customers are displayed in Figure 2. All lines are parallel to the bisecting line of the first quadrant (which corresponds to $\Delta = 0$). The angle bisector is, from a probabilistic perspective using the expected value, the barrier where selection begins to make sense, because churn rates can be expected to decrease when treated for higher (and thus positive) values of Δ.

The line representing the border of the *Persuadables*-triangle in the illustration of Radcliffe and Simpson (2007) is one of the parallel lines. The line indicates the optimal threshold for selection in terms of Radcliffe and Simpson, if prevention comes along with costs (transaction, offer, etc.). Otherwise it would be optimal to contact every customer below the bisecting line of the first quadrant.

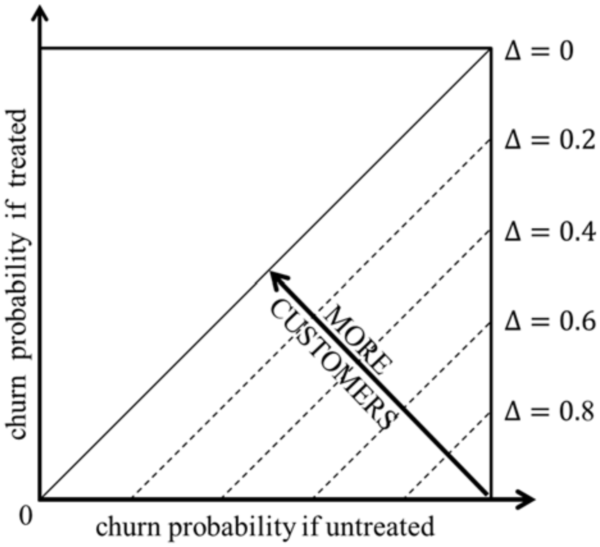

Figure 2: Graphical Illustration of Selection by Delta

Obviously, the true churn probabilities of customers are unknown and selection would otherwise be trivial. The probabilities are consequently predicted using data mining techniques.

In order to ensure that the selection by delta minimizes expected churn, the expected values of the probability estimates must be equal to the correct values, i.e., *EV_treated = p_treated* and *EV_untreated = p_untreated* (Assumption 1).

To fulfill the bias-variance trade-off of statistical learning theory, the uncertainty in the churn probability estimates must furthermore be equal for all probabilities (Assumption 2). Statistical learning theory clearly indicates that ignoring the variance is most likely not optimal.

For the purpose of achieving Assumption 2, the uncertainty in the predicted churn probability must be independent of the probability itself (Assumption 2a). In addition, every combination of churn probability of treated and corresponding churn probability of untreated customers exists and is of the same frequency (Assumption 2b) to ensure equal support for all estimates.

Lastly, the likelihood of wrong decisions because of estimation errors must not exceed the likelihood of beneficial decisions to ensure optimality of selection by the difference of probabilities. For this reason, treated and untreated churn probabilities must be unrelated, in particular churn probabilities with treatment must not be systematically larger than untreated probabilities (Assumption 3).

Overall, while the selection by difference of treated and untreated churn probabilities is, from a theoretical point of view, clearly the optimal criterion to minimize the expected number of churning customers, several assumptions have to be satisfied to ensure this optimality. In the next section, we compare these assumptions to conditions in real-world applications to determine which of them are likely to be satisfied or violated.

4 True Lift Assumptions in Practice

The considerations of the previous section, together with their implications for customer targeting according to the expected value theory, did not result in a substantial number of empirical success stories with delta-based true lift models so far. We argue that a major issue with such approaches is that the implicit requirements regarding distributions of probabilities and their relations will most likely be violated in practical settings, which will now be discussed in detail.

First, reconsidering Figure 1, the fundament for targeting customers are the relations of '*churn probability if untreated'* and *'churn probability if treated'*, while the uncertainty in probability estimates is not explicitly considered. State-of-the art data mining and predictive modeling techniques are however able to quite reliably predict the correct value on average. The expected value of predictions can therefore be assumed to be equal to the correct probabilities. Assumption 1 is consequently most likely satisfied and true lift models can accordingly also be applied to the estimated instead of the correct churn probabilities.

Before we discuss the other assumptions of true lift models, we first derive realistic assumptions regarding the distribution of churn probabilities amongst customers as a basis for the in-depth discussion.

Churn probabilities cannot be expected to be uniformly distributed between zero and one, neither for untreated customers, nor for treated customers. For instance, a uniform distribution of churn probabilities of untreated customers would mean an average annual churn rate of approximately 50% – an extraordinary high percentage value even in "churn intensive" industries such as telecommunications. In general, empirical observations indicate that we are likely to find more customers with low churn probabilities (those who remain loyal), while many customers with high churn probabilities already left the company.

More realistic distributions of churn probabilities with and without treatment are presented in Figure 3. The plot on the left-hand side shows churn probability distributions for 300,000 customers with and without treatment, following different beta distributions. The corresponding parameters in this case are *alpha_untreated* = 1, *beta_untreated* = 32, *alpha_treated* = 2 (= 2 * *alpha_untreated*), *beta_treated* = 31 (= *beta_untreated* - *alpha_untreated*).

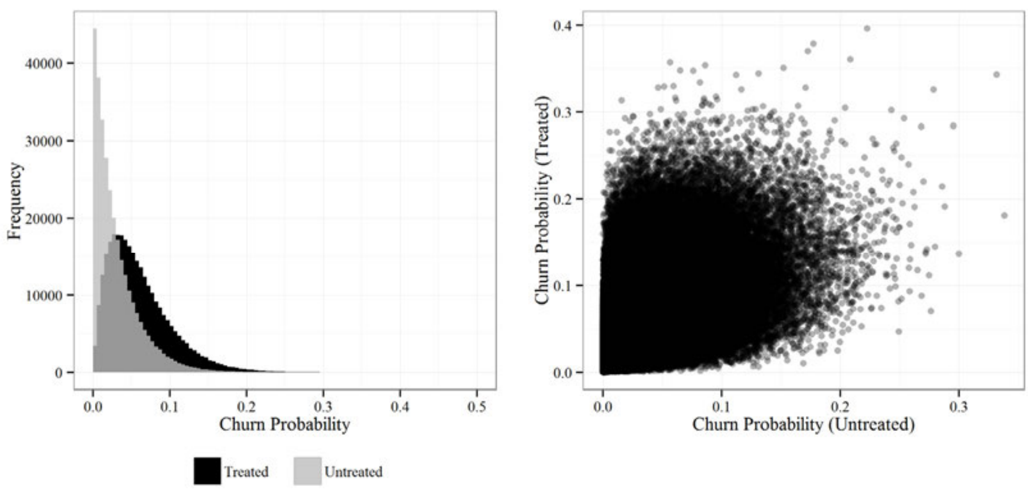

Figure 3: Challenging Distribution of Churn Probabilities

The average churn probability when treated clearly exceeds the average churn probability without treatment for the chosen distribution parameters. It is often observed in practice that subjecting a customer to a treatment, for instance a churn prevention telephone call, actually on average increases the churn probability. This finding can be explained easily since customers are actively reminded of their contract as well as the contract runtime and will, with high probability, re-evaluate the contract as a result of the churn prevention activity. Furthermore, if campaigns on average decreased probabilities, churn prevention would be comparatively easy. In contrast, we have a more challenging scenario where a random selection would actually increase churn. We can consequently directly reject the appropriateness of Assumption 3.

It is additionally intuitive to suppose a correlation between a customer's churn probabilities when treated and when untreated. For our considerations, we assume a rather moderate correlation of 0.5. This yields to the bivariate distribution illustrated on the right-hand side in Figure 3.

We can now analyze the delta-based customer selection criterion in the introduced setting in order to assess the validity of the other assumptions.

Clearly, the most interesting customers for churn prevention activities, i.e., customers with high churn probability when untreated but a probability approaching zero otherwise, are in the lowest lower-right corner and have the highest delta values.

However, we observe that – while we see many customers with low delta values – we hardly find any customers with large deltas. Hence, high deltas are estimated using relatively few data points and, thus, the churn probability reduction has increasingly lower statistical support. Consequently, the delta criterion prefers customers in regions where the statistical support is comparatively low. This coherence becomes even more tangible when rasterizing the two-dimensional space of probabilities. For this purpose we first subdivide the data display in the right part of Figure 3 into squares of side length 0.005, and count the observations belonging to a square (sub-segment). We then compute the differences between the probabilities (treated and untreated) of the particular segment centers and respectively assign the resulting value as the concerning square's delta. Figure 4 shows the resulting boxplot depicting the distributions of support depending on the delta. The figure illustrates the decay of support per sub-segment with increasing delta. While support is in many cases high for low delta values, support is very low for deltas larger 0.05.

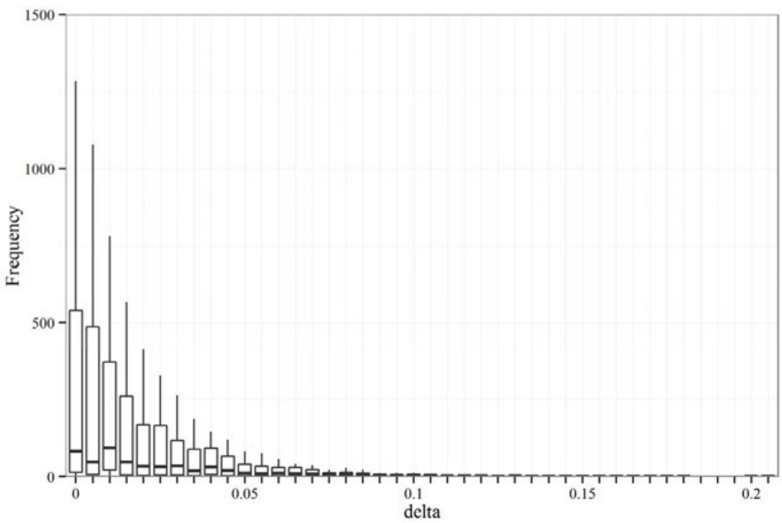

Figure 4: Decreasing Sample Size with increasing Delta

As a consequence, we can conclude that Assumption 2b is most likely not met. The support for statistical inference strongly varies for diverse (combinations of) probabilities. This in turn leads to differing variance of estimates, since estimations with high support exhibit low variance while decreasing support substantially increases variance. This is further illustrated in Figure 5, where the distributions of estimated probabilities are depicted for three segments with sample sizes 10, 100 and 1,000 as well as an average churn rate (our proxy for the mean churn probability in a segment) of p=0.1. Clearly, the smaller the sample size, the higher the variance of estimates.

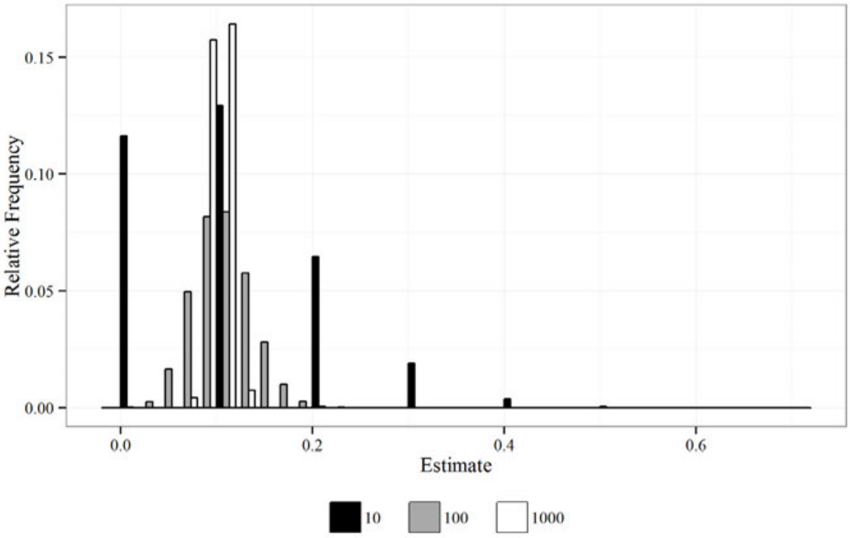

Figure 5: Probability of wrong Decisions dependent on Support

Another issue arises, from a statistical point of view, when preferring customers with higher untreated churn probabilities over those with lower untreated churn probabilities, even when the mean delta and the support are identical in both segments. This is illustrated in Figure 6.

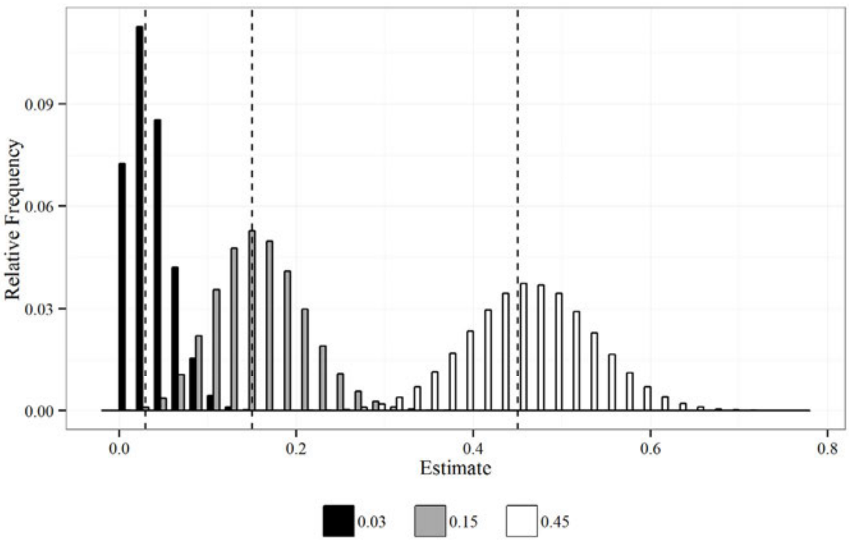

Figure 6: Probability of wrong Decisions depending on Churn Probability

Figure 6 shows the distribution of the estimated churn probabilities for treated customers in three different segments with sample size n = 50 but different original churn probabilities p = 0.03, 0.15, 0.45 (the dashed vertical lines in the plot). The figure displays a simple and well-known statistical truth: the higher the mean churn ratio (the probability parameter in a Bernoulli distribution), the wider the distribution spreads. In our case, the implication is, that the risk of actually increasing the churn ration by a treatment, equals the share of the cumulated distributions exceeding the original probabilities (the dashed lines), increases with p.

Obviously, Assumption 2a is in practice also violated, plainly because of basic statistical properties of estimators for ratios or probabilities.

Overall, several assumptions of true lift models are most likely not satisfied in practice. In particular, the basic bias-variance trade-off of statistical learning is not explicitly considered in the models, which can again lead to decreased model performance.

5 Impact of Violated Assumptions

After our in-depth analysis of the assumptions of true lift models in practice, we now discuss the consequences of the violation of these assumptions.

The effect of the skewed probability distribution regarding the uncertainty of probability estimates is definitely of particular interest. Supervised analytical procedures usually determine churn probabilities by grouping similar customers (customers with similar attribute values according to a predefined distance metric) and then computing the ratio of observed churners and non-churners in the individual groups. However, as the customer density decreases with increasing churn probability, for higher probability values the number of similar customers grouped together – and thus the support for the ratio-based probability estimates – declines, and therefore leads to increasingly unstable probability estimates. We will now debate why the resulting probability-specific uncertainties are critical in true lift modeling.

The churn probability of a customer follows a particular binomial distribution when treated, and probably another binomial distribution if untreated (excluding the consideration of priors as used in Bayesian approaches, which are out of scope of this paper), with the number of customers in the group and the group-specific churn rate (of treated customers only, of untreated customers only) as parameters.

The selection via the delta criterion prefers customers with high initial churn probabilities in regions with poor statistical support, as detailed in the previous sections. As a corollary, the two statistical effects of increasing uncertainty for lower sample size and for increased probabilities apply at the same time for the customers who are most likely selected. The total effect is depicted in figure 7. The figure shows the distribution of estimated churn probabilities (treated versus untreated) for two segments that are promising in general, as the mean churn probability of treated customers respectively is much smaller than the one of untreated customers. On the left-hand side, with n=1000, *p_untreated*=0.03 and *p_treated*=0.02, the overlap of both distributions is relatively small, while it is large on the right hand side for the distributions with n=20, *p_untreated*=0.3 and *p_treated*=0.2, indicating that the risk of wrong targeting is much higher in the second segment.

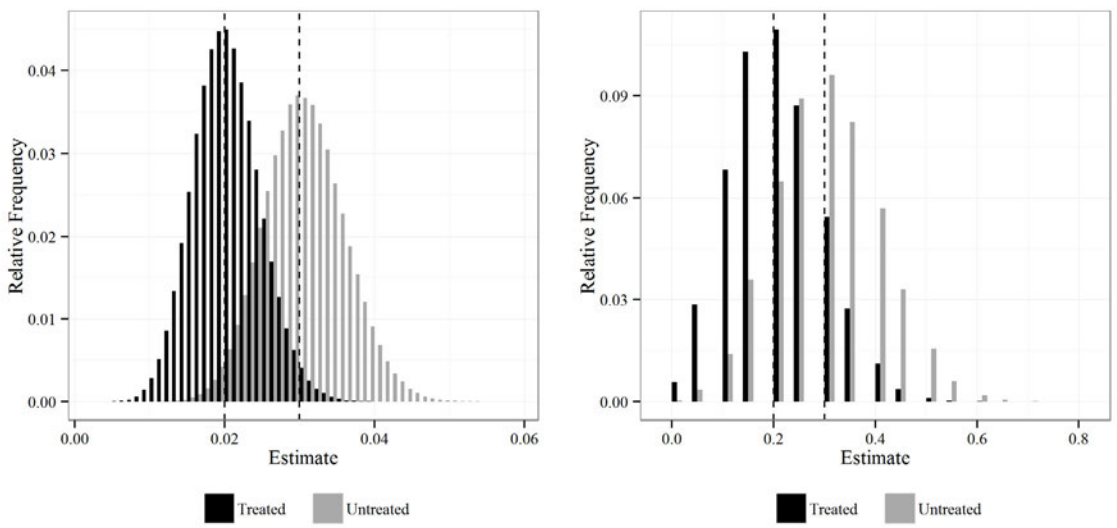

Figure 7: Probability of wrong Decisions depending on Support and Churn Probability

Evidently, making wrong decisions should be avoided whenever possible. In the challenging case of churn prevention with a higher overall churn probability when customers are randomly treated, wrong decision-making is particularly critical due to the fact that we have to assume a correlation between the probabilities, and the opportunities for improvement are seldom compared to the more frequently appearing hazards. Thus wrong decisions usually are likely to get penalized, all the more as only a small portion of a few percent of the customer base is subject to a prevention campaign.

6 Summary, Conclusion and Outlook

In summary, targeting customers with current state-of-the-art true lift models ignores the bias-variance trade-off and can lead to very poor customer selection. Above all, the uncertainty in the prediction of churn probabilities and its dependence on the probability itself is not considered, as the implicit assumption of the selection criteria in true lift models is a uniform distribution of churn probabilities as well as a statistical independence of probabilities of customers with and without treatment.

In this paper we argue that these assumptions are hardly met in practical settings, since probabilities cannot be expected to be distributed uniformly between 0 and 100% (nor in any other intervals) – thus leads to varying support for statistical churn probability estimation procedures. Moreover, original churn probabilities are likely to be correlated with those after treating customers. In fact, it is much more appropriate to assume right-skewed, long-tailed probability distributions in almost any setting of practical relevance.

As a consequence, the probability estimates for lower churn probabilities will be far more stable than estimates of high churn probability. In other words, the probability or risk of choosing wrong customers and increasing churn rates is higher when targeting customers with high initial churn probability. We argue that this uncertainty needs to be appropriately considered in customer targeting models for churn prevention, although this requires more complex approaches than basic true lift models.

Furthermore, correlated churn probabilities (treated / untreated) in combination with a right-tailed distribution of basic (untreated) churn probabilities, increases the likelihood of targeting the wrong customers, when using the assumed decay in churn probability caused by treatment (the delta) as selection criterion. Overall, the risk of wrong decisions is high with current true lift modeling approaches, which can easily lead to an increased churn rate when conducting a churn prevention campaign. This result most likely explains the low number of success stories published to date.

We conclude that other approaches that appropriately consider the uncertainty in probability estimates are required to construct successful churn prevention models. The aim of this paper is to emphasize the problem with current methods from a statistical perspective in order to provide the basis for future work on this important issue of customer targeting in churn prevention campaigns and in other applications.

7 Literature

Ahn JH, Han SP, Lee YS (2006) Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry. Telecommunications Policy 30(10):552–568

Chickering DM, Heckerman D (2000) A Decision Theoretic Approach to Targeted Advertising. In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, p 82–88

Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer Assisted Customer Churn Management: State-of-the-Art and Future Trends. Computers & Operations Research 34(10):2902–2917

Hansotia B, Rukstales B (2002) Incremental Value Modeling. Journal of Interactive Marketing 16(3):35–46

Keramati A, Ardabili SMS (2011) Churn Analysis for an Iranian Mobile Operator. Telecommunications Policy 35(4):344–356

Kim HS, Yoon CH (2004) Determinants of Subscriber Churn and Customer Loyalty in the Korean Mobile Telephony Market. Telecommunications Policy 28(9):751–765

Lo VSY (2002) The True Lift Model - A Novel Data Mining Approach to Response Modelling in Database Marketing. ACM SIGKDD Explorations Newsletter 4(2):78–86

Lu J (2002) Predicting Customer Churn in the Telecommunications Industry - An Application of Survival Analysis Modeling Using SAS. In: Proceedings of the 27th SAS User Group International Conference

Manahan C (2005) A Proportional Hazards Approach to Campaign List Selection. In: Proceedings of the 30th SAS User Group International Conference

Ngai EWT, Xiu L, Chau DCK (2009) Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification. Expert Systems with Applications 36(2):2592–2602

Radcliffe NJ, Simpson R (2007) Identifying Who Can be Saved and Who Will be Driven Away by Retention Activity - Using Uplift Modelling to Reduce Churn in Mobile Telephony. Journal of Telecommunications Management 1(2):168–176

Radcliffe NJ, Surry PD (2011) Real-World Uplift Modelling with Significance-Based Uplift Trees. Stochastic Solutions Whitepaper.

Rzepakowski P, Jaroszewicz S (2010) Decision Trees for Uplift Modeling. Proceedings of the 10th IEEE International Conference on Data Mining (ICDM), p 441–450

Rzepakowski P, Jaroszewicz S (2012) Uplift Modeling in Direct Marketing. Journal of Telecommunications and Information Technology 2:43–50

Tamaddoni Jahromi A, Sepehri MM, Teimourpour B, Choobdar S (2010) Modeling Customer Churn in a Non-Contractual Setting: The Case of Telecommunications Service Providers. Journal of Strategic Marketing 18(7):587–598

Zaniewicz L, Jaroszewicz S (2013) Support Vector Machines for Uplift Modeling. In: Proceedings of the Data Mining Workshops (ICDMW), 13th IEEE International Conference on Data Mining, p 131–138

Zhang X, Zhu J, Xu S, Wan Y (2012) Predicting Customer Churn Through Interpersonal Influence. Knowledge-Based Systems 28:97–104

Paper B: Towards More Robust Uplift Modeling for Churn Prevention in the Presence of Negatively Correlated Estimation Errors

Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020

Towards More Robust Uplift Modeling for Churn Prevention in the Presence of Negatively Correlated Estimation Errors

Frank Oechsle
esentri AG, Pforzheimer Str. 132
76275 Ettlingen, Germany
frank.oechsle@esentri.com

David W. Schönleber
esentri AG, Pforzheimer Str. 132
76275 Ettlingen, Germany
david.schoenleber@esentri.com

Abstract

The subscription economy is rapidly growing, boosting the importance of churn prevention. However, current true lift models often lead to poor outcomes in churn prevention campaigns. A vital problem seems to lie in instable estimations due to dynamic surrounding parameters such as price increases, product migrations, tariff launches of a competitor, or other events with uncertain consequences. The crucial challenge therefore is to make churn prevention measures more reliable in the presence of game-changing events. In this paper, we assume such events to be spatially finite in feature space, an assumption which leads to particularly bad churn prevention results if the selected customers lump in an affected region of the feature space. We then introduce novel methods which trade off uplift for reduced similarity in feature space when selecting customers for churn prevention campaigns and show that these methods can improve the robustness of uplift modeling.

1. Introduction

Referring to McKinsey's survey of US shoppers "Thinking inside the subscription box: New research on e-commerce consumers" [1], "the subscription e-commerce market has grown by more than 100 percent a year over the past five years. The largest such retailers generated more than $2.6 billion in sales in 2016, up from a mere $57.0 million in 2011". This survey was carried out in the end of the year 2017 and was published in the beginning of the year 2018. A similar development for the German subscription market is described by billwerk in their 2019 published white paper "subscription based services" [2]. They highlight that the revenues of German vendors of subscription-based services since 2015 are exponentially growing by more than 100 percent per year. At any rate, the subscription business is an economy gaining in importance, and is after the big successes in North America now conquering the European market. In consequence, churn management, and with it the subdomain churn prevention, will become of paramount prominence.

Yet, state-of-the-art uplift models often lead to poor outcomes in churn prevention campaigns, like any other common churn prevention approach as well. The crucial question thus is how to do churn prevention in a more reliable way, i.e., in a way that the benefit of a campaign is more probable. A churn managing company basically would like to know how each of their customers will react when being targeted within the scope of a churn prevention campaign such as a phone call with a specific contract renewal offer, in comparison to their behaviour when they are not targeted. At this context "reaction" or "behaviour" means in particular to announce churn or not to announce churn.

The underlying challenge is thus to predict the probability of a customer to announce churn, depending on the participation in a (specific) churn prevention campaign. This probability consists of two different probabilities, namely the probability of churning without being contacted and the probability of churning when being contacted. Even if only one of the probabilities is notably misestimated, the success of the whole churn prevention campaign is in danger.

The issue of estimating these probabilities is further aggravated by the rarity of churn *per se*, which implies that successful churn prevention cases are even rarer. Accordingly, the estimation of the probability of customers that can be successfully prevented from announcing churn when receiving an appropriate measure is both challenging and crucial, since failure provokes the opposite of the aimed target. Thus even partial failure in estimating churn probabilities can lead to increased churn rates, which is eminently adverse since it is much more expensive to acquire new customers than retaining the inventory customers [3]. Consequently, we need an approach that ensures the absence of failure as far as possible while still realizing existing chances of churn reduction.

URI: https://hdl.handle.net/10125/63931
978-0-9981331-3-3

In this paper, we explore the effect of game-changing events such as tariff launches of a competitor on uplift modeling. We assume these events to be spatially finite in feature space and evaluate different customer selection methods based on decision trees via Monte Carlo simulations, including novel selection methods which trade off uplift against more diversity in feature space and prove to be more robust in the presence of such events.

Note that when we use the term *churn* in this paper, we mean *churn announcement*. Only if a company's efforts in retaining the churn announcing customers are of no avail this results in churn. At this point, churn management is divided into prevention and retention. We clearly concentrate on churn prevention in this paper.

2. Related work

The basic theory underlying Lo's true lift model [4] is quite intuitive and well-defined, but surprisingly does not reliably succeed in the churn context. The definition according to Table 1 is neat and in essence considers the incremental effect a campaign has on the selected customers, whereas in the context of a traditional response model the focus is only on the response after the campaign (treatment) devoid of checking what it would have been without the campaign (control), that is maximizing the difference $A - C$. For instance A, B, C and D could denote the probability of purchase or churn for the corresponding customer segment.

Table 1. Definition of true lift, following Lo [4].

	Treatment	Control	Increment
Model-guided	A	B	A-B
Unguided	C	D	C-D
Difference	A-C	B-D	(A-B)-(C-D)

The true lift approach consequently results in a different selection method compared to the classic response model in that it selects by the delta of the customers churn probabilities when treated (= received the campaign treatment) or untreated (= not received the campaign treatment), that is the uplift. The uplift is calculated as

$$\Delta = p_0 - p_1, \tag{1}$$

where p_0 and p_1 are the churn probabilities without respectively with treatment. The selection method according to the uplift is illustrated in Figure 1. There, the lower-right corner represents the optimal point for selection (highest churn probability if untreated but lowest if treated, i.e. maximal Δ). The lower the value of the Δ selection threshold (e.g. due to higher available budget), the more customers are covered in the campaign treatment.

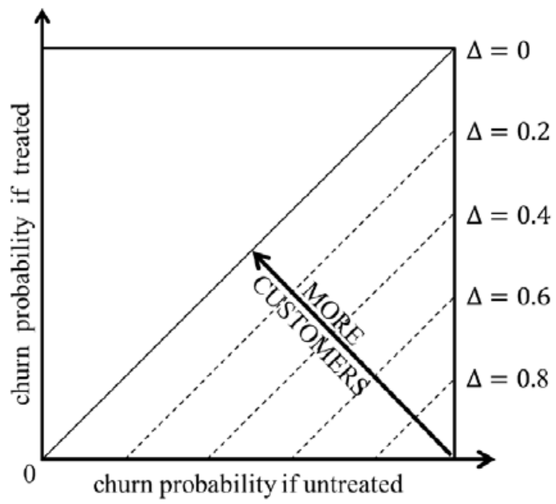

Figure 1. Graphical illustration of selection by delta from Oechsle et al. [5].

Even though the available research relating to uplift modeling is not inexhaustible, there is definitely adequate knowledge about predicting those increments used in Lo's true lift model. Kane et al. [6] depict this in a comprehensive way as well as Guelman et al. [7] do. Hence, the direct prediction of the difference between the probabilities in particular seems to be a mature approach. The leading alternative would be to predict the two probabilities separately, building the uplift afterwards via subtraction.

Nevertheless, empirical results are not satisfying. This picture emerges by for example comparing the corresponding work of Chickering and Heckerman [8], Hansotia and Rukstales [9], Manahan [10], Rzepakowski and Jaroszewicz [11], Radcliffe and Surry [12], Rzepakowski and Jaroszewicz [13] or Zaniewicz and Jaroszewicz [14]. The nonexistent established success stories in practice encourage this point of view.

Consequently, the focus of the most recent studies is uncertainty and estimation errors as a central root cause for the observed phenomenon of missing best practices in terms of churn prevention via uplift modeling. While Lo [15] accounts for the variability in estimates in a marketing context, Oechsle et al. [5] address estimation risks in the subscription business when it comes to churn. Athey and Imbens [16] analyse in general the increasing uncertainty in estimations arising from a smaller sample size. As they deal with decision trees, they propose to use different subsamples for splitting and estimating and try to control their results via confidence intervals. This approach appears not intuitive in so far as it multiplies the initial problem of small leaf size by further splits.

In the remainder of the paper, we pick up this latter stream and investigate the effect of suddenly upcoming estimation errors due to moving environments in the subscription business. We thereby contribute insight concerning the question why churn prevention all too often fails and how it can be done better.

3. Negatively correlated estimation errors due to uncontrollable events

The above mentioned moving environment accumulates dynamic surrounding parameters, which can be on the one hand company-intern changes such as mandatory price increases or product migrations owing to technical improvement, and on the other hand external factors like tariff launches of a competitor or other specific events influencing customer groups in undetermined ways. This fluctuation leads to instable estimations and suboptimal decisions at least. In the worst case, these dynamic parameters can lead to negatively correlated estimation errors within homogenous customer segments.

Consider, for instance, a company whose portfolio includes the cheapest tariff for the entire branch linked with a competitive common service package, and therefore successfully contracted bargain hunters. Consider further that it is suddenly confronted with a competitor launching an even cheaper tariff of adequate quality. In this case it is possible that the just very loyal (low p_0, high p_1) bargain hunters abruptly turn to unfaithful (high p_0, low p_1), change-oriented customers. Let now the prevention campaign process be at a point where this could not be recognized and incorporated any more, then it ends up in probability estimations that still pretend loyal bargain hunters but rather belong to potential churners. In other words underestimated p_0 and overestimated p_1, that is, negatively correlated estimation errors.

Such game-changing events can burst in on a prevention campaign at all times and cannot always be anticipated nor reliably excluded by, for example, smart definition of the target group of the prevention campaign. Their effect can only reasonably be assumed to be spatially finite, i.e. locally bounded with respect to their diffusion in feature space. The most detrimental impact of these events occurs when the customers that seem to be the most promising (and which are therefore selected for treatment) lump in a certain region of feature space which is affected by the upcoming game-changing event in such a way that the overall treatment effect is reversed. Consequently, targeting churn prevention measures at a customer group characterized by their similarity in feature space poses a potential threat to the success of these measures and should hence be penalized or avoided.

In this paper, we evaluate the effect of such game-changing events via Monte Carlo simulations and propose alternative strategies for customer selection, which are more robust to aforementioned events. To this end, we consider a two-dimensional feature space, sketched in Figure 2a). Note that the features are assumed to be normalized, such that the range of the feature values fall within the range $[0, 1]$. This feature space is then divided into rectangles, which is the general concept of decision trees, and afterwards, likewise typical for tree-based methods, each of the rectangles sustains a constant (often between 0 and 1 reflecting a relative frequency) representing the "model" according to the leaf. We further assume a game-changing event to have significant impact on customers at a specific, randomly selected point (E) in feature space as well as a distance-dependent impact on surrounding customers (or users U) that is spatially confined within a distance R from E.

More specifically, assume Δ to be the real and correctly estimated effect of the churn prevention campaign per customer in the absence of the upcoming game-changing event. For a customer with distance r to the center E of the game-changer, we calculate the modified effect Δ' of the churn prevention campaign in the presence of the game-changer according to

$$\Delta' = \begin{cases} \Delta & r > R, \\ \Delta\left[1 - 2\cos\left(\frac{\pi r}{2R}\right)\right], & r \leq R \end{cases} \quad (2)$$

where R is the radius of the circle of influence of the game-changer and r specifies the Euclidean distance of the customer U to the center E of the game-changer. Hence, the effect of a game-changer is maximal for customers close-by while it decreases radially up to a distance R, where its effect vanishes, following the above cosine behaviour. At the center E of the game-changer, the true effect of the churn prevention campaign changes sign. For example, when the real and estimated effect of the prevention campaign would be 0.5 reduction of churn probability, it would switch to -0.5 reduction for a customer lying at the center of the game-changer, but only for the real effect. The estimated effect would still be 0.5.

This modeling is motivated by the fact that Δ is the difference between the churn probability without

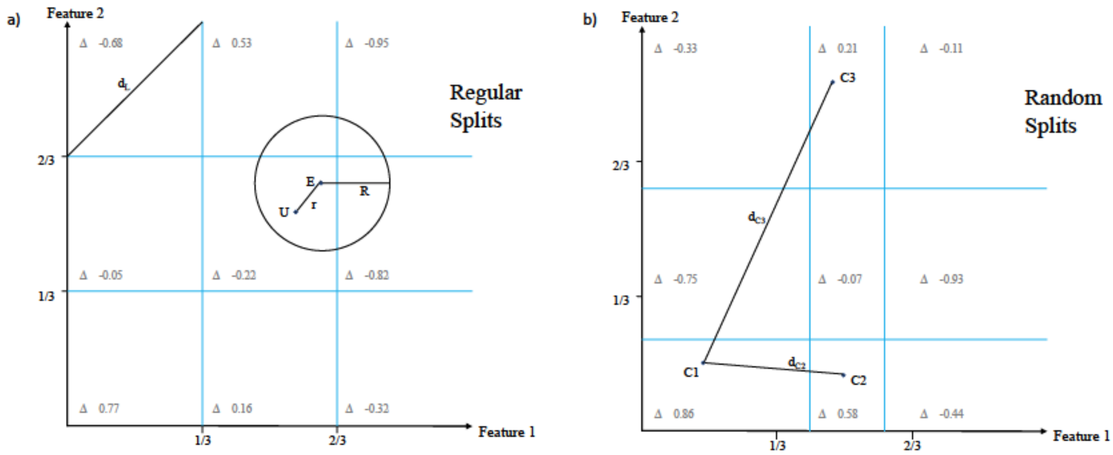

Figure 2. Radial estimation error and linkage via the centroid-method in regular or random decision trees.

treatment p_0 and the churn probability with treatment p_1 [cf. Eq. (1)]. Thus the modeled behaviour can be achieved by overestimating p_0 while underestimating p_1 at the same time and therefore would produce a negatively correlated estimation error. For instance, an initial $\Delta = 0.5$ which changes to $\Delta = -0.5$ in the presence of a game-changing event can be achieved by initially estimating $p_0 = 0.7$, $p_1 = 0.2$ whereas actually $p_0 = 0.2$ and $p_1 = 0.7$ in the presence of a game-changer.

4. Handling similarity

The spatial confinement and construction of the errors as described in the previous section suggests to prefer customers that are not similar to each other, in other words, that are distant in feature space. Since we cannot modify where the individual customer is located in feature space, we focus on handling the similarity of customers by reducing calculated benefits depending on the Euclidean distance between the customers and vary our selection approach using the centroid-method as linkage technique.

A linkage technique is relevant since we use the concept of decision trees in our analysis and thus do not select individual customers but customers aggregated in leaves. For this reason we split the two-dimensional feature space via two continuous splits, vertically and horizontally, resulting in nine rectangles, which cover the feature space and represent nine leaves of a decision tree. In our simulation, we then randomly assign three positive and six negative values to Δ with $|\Delta_i| \in [0, 1]$ for $i = 1, 2, ..., 9$, corresponding to the churn reduction estimated via the decision tree. By dint of the predominantly negative values of Δ we indicate a generally inauspicious base case for churn prevention measures. This is inspired by the absence of track records in churn prevention campaigns. The distances between the cluster centroids that are most relevant in our simulations are the ones between the rectangle (leaf) with the highest Δ and the two other rectangles (leaves) with positive value of Δ. Those two distances d_{C2} (best to second best) and d_{C3} (best to third best) are calculated via the Euclidean distance of the cluster centroids $C1$, $C2$ and $C3$. As $C1$ represents the centroid of the rectangle (leaf) with the highest Δ and $C2$, $C3$ analogical the second and third best leaf, d_{C2} is the Euclidean distance from $C1$ to $C2$ while d_{C3} denotes the Euclidean distance from $C1$ to $C3$. For the calculation of the cluster centroids themselves we use the average of both customer features in the corresponding segment. We furthermore assume the customers to be equally distributed in feature space, and hence these centroids typically do not deviate strongly from the centers of the rectangles (leaves).

The classic way of selection, independent of the business use case (sales, churn, etc.), is to select the best N customers the previously allocated budget is able to fund. It is not seldom that this budget is derived without having an idea about the probabilities and the chances or risks in the specific churn prevention campaign. In this approach, no similarity aspects are taken into account.

In contrast, in this paper, we introduce novel selection methods, which incorporate these similarity aspects by design. We thereby challenge the classic way of selecting customers with several distance-respecting methods. In particular, we introduce the following

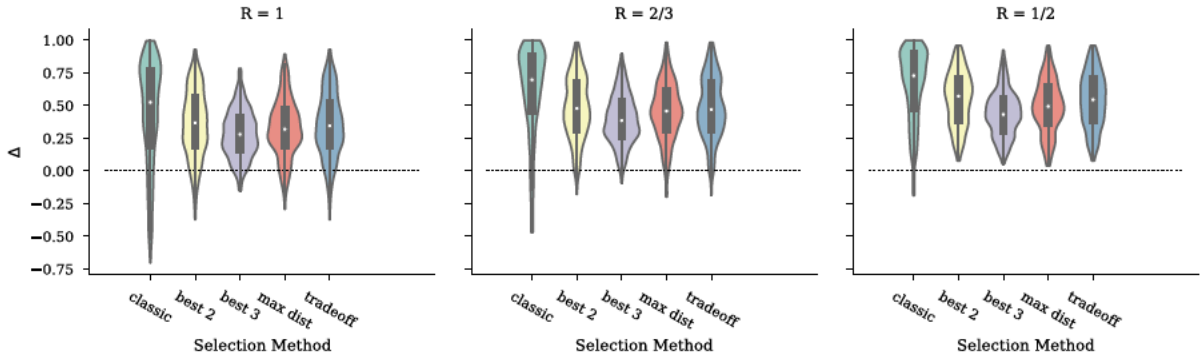

Figure 3. Comparison of selection methods for different radii R (in units of d_L) and regular decision tree splits.

selection methods: *best 2*, *best 3*, *max dist* and *tradeoff*, which are described below. Note that in the now following discussion the leaf represented by the centroid C_i is named leaf i and analogously all variables with index i correspond to leaf i. In addition *best* implies highest value of Δ among all leaves.

best 2 In this method we randomly select $N/2$ customers in the best leaf and $N/2$ customers in the second best leaf. Hence, we automatically trade off uplift against more diversity in feature space.

best 3 Here we randomly select $N/3$ customers in the best leaf, $N/3$ customers in the second best leaf, and $N/3$ customers in the third best leaf. Hence, within the scope of our simulation, we maximally diversify in feature space while trading in even more probability.

max dist In this method we randomly select $N/2$ customers in the best leaf, and $N/2$ customers in the leaf i where the distance to the best leaf is maximal, i.e., $d_i = \max(d_{C2}, d_{C3})$. Thus we prefer distance irrespective of the traded in probability.

tradeoff Here we randomly select $N/2$ customers in the best leaf, and $N/2$ customers in the leaf i which minimizes the quotient

$$q_i = \frac{\Delta_1 - \Delta_i}{d_{Ci}} \tag{3}$$

for $i = 2, 3$, i.e., $q_i = \min(\frac{\Delta_1 - \Delta_2}{d_{C2}}, \frac{\Delta_1 - \Delta_3}{d_{C3}})$. Therefore the focus is simultaneously on both of the parameters.

5. Simulation results

We now benchmark the selection methods introduced above in a Monte Carlo simulation. In this simulation, we consider two different decision tree splits: *regular* and *random*. In the *regular* case illustrated in Fig. 2a), the areas in feature space covered by the leaves are identical, which in practice could result from the constraint of a minimal leaf size. In the *random* case, illustrated in 2b), the decision tree splits are randomly chosen between zero and one, resulting in leaves of varying sizes.

We then randomly and uniformly distribute 225,000 customers in feature space, leading in the regular case to an average of about 25,000 customers per leaf. By adding a customer at the leaf center in each partition i we ensure that independent of the tree splitting there is no empty partition. To evaluate the effect of game-changing events (or negatively correlated estimation errors), we consider three different values for the impact radius R of the circle of influence that is depicted in 2a). At the maximal radius $R = d_L$, the radius coincides with the diagonal of a regular leaf size, and is consequently decreasing for the chosen values of $R = 2/3\, d_L$ and $R = d_L/2$.

For each of the two scenarios, *regular* and *random*, we simulate 250 decision trees and compare the expected Δ values achieved by the different selection methods in the presence of a game changing event. The statistical distributions of the results are depicted in Figs. 3 and 4.

Figure 3 visualizes the simulation results for three different impact radii R and regular decision tree splits. The violin plots depict the distributions of achieved uplift for the selection methods introduced in Section 4 according to the performed simulations. Most notably, while the classic selection method realizes more profitable outcomes overall (as can be seen from

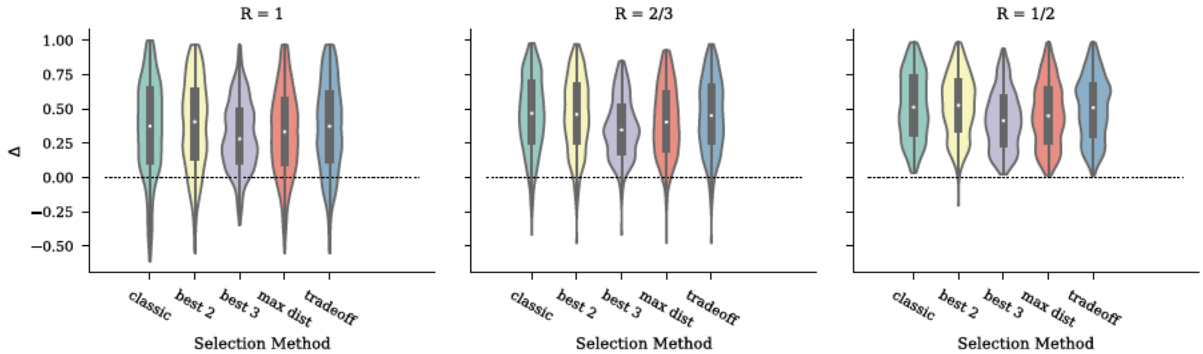

Figure 4. Comparison of selection methods for different radii R (in units of d_L) and random decision tree splits.

the high median value in the miniature boxplot inside the violin plot), this comes at the cost of seldom but significantly adverse negative outcomes. In contrast, more diversifying methods such as *best 2* or *best 3* achieve less favorable results on average, yet also the risk of negative outcomes with high costs are reduced as well. This behavior is most prominent for the largest impact radius $R = d_L$, but persists also for decreasing radii ($R = 2/3\, d_L$ and $R = d_L/2$). The same pattern is found for the *tradeoff* selection, yet this approach is able to reach a comparably low traded-in probability similarly to *best 2* and at the same time reduces failures in a superior way. The last one can be seen more clearly in Table 2 which gives a detailed summary of the simulation results.

For random splits, the benefit from trading off uplift for reduced similarity between customers in feature space disappears, as can be seen from Figure 4. This demonstrates that the architecture of the underlying decision tree in combination with the error rate (frequency and impact) is crucial. We conjecture that in the *random* case the repeatedly strong divergent sizes of leaves and radii of the error diffusion do not ensure a setting where either any churn prevention action is recommendable at all or it is necessary to desist from the classic selection approach.

6. Conclusion and discussion

The results presented in this paper clearly show that there is a notable uncertainty in the efficacy of churn prevention campaigns when spatially occurrent estimation errors are frequent and effectual enough and the general ecosystem is prevention unfriendly (6 out of 9 Δs < 0). Frequent and effectual enough here means that errors happen regularly and that the radius of the error impact has a minimum size relative to the leaf sizes of the decision tree. The latter can be observed in our results by comparing the regular splits scenario versus the random splits scenario. We will analyse the reasons for the blurring effect of random splits on the uplift in more detail in future work.

Despite this problematic setting with large and frequent game changing estimation errors we have illustrated that there are nevertheless methods of trading off uplift for reduced similarity between customers concerning their position in feature space that lead to more robust estimation results in terms of variance. In particular, it is possible in not too toxic ($R \leq d_L$) surroundings to reduce the number of disappointments (churn increasing churn prevention campaigns). To this end, we not only select the customers with the highest Δ ignoring whether they lump in a specific region, but also consider their similarity. Besides some straightforward methods (*best 2*, *best 3*, *max dist*) we also investigated a more elaborated *tradeoff*, namely a method which uses a linearly weighted combination of Δ-difference and Euclidean distance of the cluster centroids for customer selection. Especially this trading off method is promising, since it pays not the highest prices for the targeted risk reduction.

It is clear, however, that we will refine this tradeoff approach in our future work. It should be possible to derive an even superior tradeoff by for example taking a sigmoid relation as a basis, particularly with regard to the assumed cosine behaviour of the error diffusion. In either case, the appropriate rigor for tackling similarity depends on the industry's susceptibility to dynamic changes in customer churn probabilities. It is a parameter that could/should be optimized/learned over time if it could not already be observed in the past.

Another particularly exciting track we will focus on in further research is to directly influence the generation of the decision tree itself. This is to step in the splitting rules as well as to intervene in the pruning of the tree. In this regard it could be clever to perform additional random splits for artificially gaining more distance amongst the leaves, combined by an uplift and distance tradeoff selection method. A novel pruning approach we are currently considering is to aggregate partitions with low similarity respectively high distance in feature space and churn probabilities (or at least uplifts) that are as identical as possible.

We will witness the development of the subscription economy and how it influences the research on uplift modeling in the churn context. The curiosity about dependable solutions for the churn prevention challenge will certainly increase in the face of the current development.

References

[1] T. Chen, K. Fenyo, S. Yang, and J. Zhang, "Thinking inside the subscription box: New research on e-commerce consumers," 2018.

[2] billwerk GmbH, "Subscription based services," 2019.

[3] A. Gallo, "The value of keeping the right customers," 2014.

[4] V. S. Y. Lo, "The true lift model - a novel data mining approach to response modeling in database marketing," *SIGKDD Explorations*, vol. 4, pp. 78–86, 2002.

[5] F. Oechsle, T. Setzer, and S. M. Blanc, "On the assumptions of true lift models for churn prevention," in *Multikonferenz Wirtschaftsinformatik (MKWI) 2016, Technische Universität Ilmenau, 09.-11. März 2016, Band 2. Hrsg.: V. Nissen*, pp. 1233–1244, Universitätsverlag Ilmenau, 2016.

[6] K. Kane, V. Lo, and J. Zheng, "Mining for the truly responsive customers and prospects using true-lift modeling: Comparison of new and existing methods," *Journal of Marketing Analytics*, vol. 2, 12 2014.

[7] L. Guelman, M. Guillen, and A. M. Pérez-Marín, "Uplift random forests," *Cybernetics and Systems*, vol. 46, pp. 230–248, 2015.

[8] D. M. Chickering and D. Heckerman, "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., 2000.

[9] B. Hansotia and B. Rukstales, "Incremental value modeling," *Journal of Interactive Marketing*, vol. 16, no. 3, p. 35, 2002.

[10] C. Manahan, "A proportional hazards approach to campaign list selection," *SAS User Group International (SUGI) 30 Proceedings*, 2005.

[11] P. Rzepakowski and S. Jaroszewicz, "Decision trees for uplift modeling," in *2010 IEEE International Conference on Data Mining*, pp. 441–450, IEEE, 2010.

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

[13] P. Rzepakowski and S. Jaroszewicz, "Uplift modeling in direct marketing," *Journal of Telecommunications and Information Technology*, pp. 43–50, 2012.

[14] L. Zaniewicz and S. Jaroszewicz, "Support vector machines for uplift modeling," in *2013 IEEE 13th International Conference on Data Mining Workshops*, pp. 131–138, IEEE, 2013.

[15] V. S. Lo and D. A. Pachamanova, "From predictive uplift modeling to prescriptive uplift analytics: A practical approach to treatment optimization while accounting for estimation risk," *Journal of Marketing Analytics*, vol. 3, no. 2, pp. 79–95, 2015.

[16] S. Athey and G. Imbens, "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, vol. 113, no. 27, pp. 7353–7360, 2016.

Table 2. Summary of simulation results. $\mathbb{E}[\Delta]$ denotes the expectation value, Successes and Failures correspond to simulation runs with $\Delta > 0$ and $\Delta \leq 0$, respectively.

R/d_L	Splits	Selection Method	$\mathbb{E}[\Delta]$	# Successes	$\mathbb{E}[\Delta]$ Successes	# Failures	$\mathbb{E}[\Delta]$ Failures
1/2	regular	classic	0.66	244	0.67	6	-0.12
1/2	regular	best 2	0.55	250	0.55	0	–
1/2	regular	best 3	0.44	250	0.44	0	–
1/2	regular	max dist	0.49	250	0.49	0	–
1/2	regular	tradeoff	0.54	250	0.54	0	–
1/2	random	classic	0.52	250	0.52	0	–
1/2	random	best 2	0.52	249	0.52	1	-0.20
1/2	random	best 3	0.42	250	0.42	0	–
1/2	random	max dist	0.46	250	0.46	0	–
1/2	random	tradeoff	0.50	250	0.50	0	–
2/3	regular	classic	0.61	230	0.68	20	-0.24
2/3	regular	best 2	0.48	246	0.49	4	-0.07
2/3	regular	best 3	0.40	248	0.40	2	-0.06
2/3	regular	max dist	0.45	244	0.47	6	-0.11
2/3	regular	tradeoff	0.48	247	0.49	3	-0.09
2/3	random	classic	0.46	237	0.49	13	-0.11
2/3	random	best 2	0.45	237	0.48	13	-0.11
2/3	random	best 3	0.36	244	0.37	6	-0.13
2/3	random	max dist	0.41	241	0.43	9	-0.12
2/3	random	tradeoff	0.45	240	0.47	10	-0.13
1	regular	classic	0.44	212	0.57	38	-0.29
1	regular	best 2	0.36	231	0.40	19	-0.12
1	regular	best 3	0.28	239	0.30	11	-0.07
1	regular	max dist	0.33	231	0.36	19	-0.09
1	regular	tradeoff	0.35	235	0.38	15	-0.13
1	random	classic	0.36	220	0.44	30	-0.25
1	random	best 2	0.38	219	0.46	31	-0.16
1	random	best 3	0.31	229	0.35	21	-0.13
1	random	max dist	0.33	215	0.41	35	-0.16
1	random	tradeoff	0.37	216	0.45	34	-0.15

Paper C: Increasing the Robustness of Uplift Modeling Using Additional Splits and Diversified Leaf Select

Increasing the Robustness of Uplift Modeling Using Additional Splits and Diversified Leaf Select

Abstract

While the COVID-19 pandemic negatively affects the world economy in general, the crisis accelerates concurrently the rapidly growing subscription business (Zuora, 2020) and online purchases (Gu et al., 2021). This provokes a steadily increasing demand of reliable measures to prevent customer churn which unchanged is not covered. The research analyses how preventive uplift modeling approaches based on decision trees can be modified. Thereby it aims to reduce the risk of churn increases in scenarios with systematically occurring local estimation errors. Additionally it compares several novel spatial distance and churn likelihood respecting selection methods applied on a real world dataset. In conclusion it is a procedure with incorporated additional and artificial decision tree splits that dominates the results of an appropriate Monte Carlo simulation.

Keywords: churn; prevention; uplift modeling; local errors; decision trees; additional splits

1. Introduction

Pejić Bach et al. (2021, p. 1) define churn as "a situation when customer stops buying products or using services from a company". Regarding the telecommunication industry they correspondingly describe that "churn management aims to minimize the churn using various retention strategies to prevent customers from cancelling subscriptions, such as offering new devices or services". With an even more detailed view one can differentiate between the two churn management disciplines prevention and retention depending on the moment of churn announcement. Prevention combines churn avoiding measures that take place before the customer announces churn while retention means the bunch of actions in the period between churn announcement and expiration of the contract. Companies naturally want preferably narrow churn funnels, which first of all is less churn announcements and therefore less churn. Thus a critical factor for success in the upcoming (subscription) business era will be a strong churn management, as far as possible in a preventive way.

However in practice there is still no trusted concept of reducing churn in a preventive measure. That applies to uplift techniques, which are comparing the customers responses depending on the inclusion in a churn prevention campaign and all the more to response modeling. One reason is the rarity of the event churn in comparison to e.g. purchase, which complicates its prediction. Another challenging aspect is that failures tend to generate additional churn (Radcliffe, 2007b). Failures means false selections in terms of customers would not have churned if they would not have received emails or any other contacting. This results in at least futile churn prevention efforts (Ascarza, 2018).

The paper counteracts those momentous misjudgements of probabilities with a diversifying portfolio approach. This concept by dint of additional and artificial decision tree splits trades in expected churn probability for distance in the feature space. Simultaneously it is able to reduce the risk of churn increasing churn prevention campaigns considerably in a setting with systematically assumed local estimation errors.

The fundamental idea of the line of thought is the true lift model of Lo (2002), which considers the incremental impact of an action towards the target variable, in this case churn, as the guide for decision-making. In order to train a decision tree to estimate churn probability increments as defined by Lo the paper uses and adapts the real world dataset of Kevin Hillstrom (2008) provided in *The minethatdata e-mail analytics and data mining challenge*. Hence it obtains a partition of the feature space in which it randomly incorporates the local errors mentioned above in a next step. Finally it exercises different campaign-selection methods within the framework of a Monte Carlo simulation. The results of this simulation demonstrate the superiority of the portfolio approach in a scenario as described, notably in comparison to the classic approach.

2. Related work

The prediction of uplifts as per Lo is theoretically clear and sufficiently comprehensible (Radcliffe, 2007b; Kane et al., 2014; Guelman et al., 2015). However, with a few mixed exceptions (Manahan, 2005; Radcliffe, 2007b; Devriendt et al., 2021) empirical results as well as best practices and track records in business are not existing in the churn context.

Concerning this matter, Diemert et al. (2018, 2021) quote missing publicly available real world datasets as a fundamental problem for the research on successful usage of uplift models (UM) in general and moreover provide a very large dataset (25M rows, 12 features). Additionally they mention Hillstroms dataset as "the second largest and most popular uplift prediction dataset" (Diemert et al., 2018, p. 3) and note that "in the field of UM a notable exception to private datasets is the Hillstrom study (64,000 samples) collecting the sales results of an e-mail marketing campaign from the 2000's" (Diemert et al., 2021, p. 2). This research will base the simulations on this exact Hillstrom dataset in the remainder of the paper.

Radcliffe (2007b, p. 13) uses the same line when he says "performance of uplift models on fabricated test data is often a particularly unreliable indicator of likely performance on real world data. A significant challenge is therefore to find suitable data that can be made publicly available for benchmarking." Not related to this he brings up that "in practice, most of the real difficulties with uplift modeling derive from noise" (Radcliffe, 2007b, p. 13). He describes several reasons for this noise (addition of estimation errors while fitting a difference, considerably unbalanced treated and control population, uplift phenomenon way smaller than absolute outcome rates) and states "a wide variety of methods to control noise, including careful variable selection and binning methodologies, bagging, stratified sampling and k-way cross-validation methods" (Radcliffe, 2007b, p. 13).

Shaar et al. (2016) underline Radcliffe's perception with their statements "uplift models show high sensitivity to noise and disturbance, which leads to unreliable results" (Shaar et al., 2016, p. 1) and "most of real world datasets contains noise and disturbances, specially for uplift modeling, as uplift effects tend to be smaller than the real treatment effect" (Shaar et al., 2016, p. 9). They allow for that with their disturbance effects minimizing approach called Pessimistic Uplift Modeling. Furthermore they show amongst others using Hillstroms dataset "that our approach outperforms the existing approaches, especially in the case of high noise data environment" (Shaar et al., 2016, p. 1). Their procedure is geared to Lai (2006), who wants to maximize the probability that customers belong to the group that shows the desired response when treated or that does not show the desired response when not treated. Furthermore it supplements Lai's method with weights representing the predicted cases proportions of the whole population. Thus Shaar et al. (2016) generate additional certainty on the expected outcomes by incorporating the overall frequency of an event.

The latest research towards uplift modeling mainly focuses on noise, disturbance, uncertainty and estimation errors (Athey et al., 2015; Lo and Pachamanova, 2015; Oechsle et al., 2016; Athey and Imbens, 2016; Zhao et al., 2017; Rößler et al., 2021). Summing up Zhao et al. (2017, p. 8) put it in a nutshell while describing that their contribution is in a first step to "present a way to obtain an unbiased estimate of the expected response under an uplift model which has not been available in the literature".

Whereas aforesaid papers attend to the uplift modeling challenges from a technical and engineering emphasis, Oechsle and Schönleber (2020) examine the problem of unreliable expected outcomes to a greater extent from a business perspective, in this case churn business. They "investigate the effect of suddenly upcoming estimation errors due to moving environments in the subscription business" (Oechsle and Schönleber, 2020, p. 3). As a moving environment they subsume "dynamic surrounding parameters" like "company-intern changes such as mandatory price increases, product migrations owing to technical improvement, tariff launches of competitors or other specific events influencing customer groups in undetermined ways" (Oechsle and Schönleber, 2020, p. 3). They suppose those "game-changing events" to systematically generate estimation errors, which in the uplift and churn context can be very disadvantageous, exceedingly when similar customers, that is local neighbours in the feature space, are selected. Concretely they define circles with radius R around random error seeds E and attribute the users (or customers) U with Euclidean distance r to E an unnoticed change in expected uplift Δ to Δ' appropriate to

$$\Delta' = \begin{cases} \Delta & r > R, \\ \Delta\left[1 - 2\cos\left(\frac{\pi r}{2R}\right)\right], & r \leq R \end{cases} \tag{1}$$

Finally they indicate supported by simulations that it can be beneficial in defective scenarios to use distance regarding customer selection techniques.

The idea of locally occurrent unanticipated changes in churn probabilities is supported by several publications concerning the topic of churn in the neighbourhood of influential churners (Dasgupta et al., 2008; Kusuma et al., 2013; Droftina et al., 2015a,b). For example (Droftina et al., 2015b, p. 1) assert that "highly influential customers deserve special attention, since their churns can also trigger churns of their peers". Correspondingly Kusuma et al. (2013) show on a real world data set that when 50 percent of the peers of users yet churned, those users' churn rate is two times the overall churn rate amongst all users.

This paper picks up the idea of noise and uncertainty typified by spatially specified sources of error and exert it on a real world dataset (Hillstrom), which previously is tailored to a churn scenario. A decision tree is trained on that dataset and it is acted upon the splitting/pruning via novel selection methods targeted to a predefined churn prevention campaign. The introduced methods are meant to regard distance in the feature space, which is well able to be done per decision tree. Besides that established decision trees employed for uplift modeling only use differences of probabilities for splitting, that is particularly they disregard distances, nor do they use pruning (Rzepakowski and Jaroszewicz, 2010). Thus common decision trees, as well as various other procedures, have an issue with locally occurring errors. The research randomly incorporates these errors in a concluding Monte Carlo Simulation and provides evidence for the superiority of its approach. Certainly even a perfectly engineered prediction model experiences problems if the described errors arise after a perfect estimation process. Hence the focus is not to derive the most accurate prediction model, in this case the most sophisticated decision tree, but rather to reliably implement an arbitrary proper decision tree for using the novel selection methods. In the following third chapter the methodology will be described in-depth.

The contribution of the research therefore consists of a) a publicly available uplift analysis on a real world dataset and b) a straight forward feasible and nevertheless promising approach for daily practice c) based on decision trees combined with a distance respecting course of action d) in the rarely considered and eminently fraught with risk uplift modeling field churn, which intensifies some of the general problems uplift modeling has to deal with.

3. Methodology

As seen in the comparing work of Zhao et al. (2017), Oechsle and Schönleber (2020) or Radcliffe and Surry (2011), the direct path is the superior one of the two popular uplift modeling approaches (direct uplift modeling versus two separate models subtracted afterwards). Thus let there be a decision tree with $I \in \mathbb{N}$ leaves for the direct estimation of the uplift

$$\Delta = p_0 - p_1, \tag{2}$$

of a churn prevention campaign whereas p_0 respectively p_1 displays the churn probability without respectively with treatment. Let further Δ_i for $i = 1, 2, .., I$ be the (correctly) estimated and therefore expected uplift for the customers enclosed in leaf i, whereat w.l.o.g. for simplification only positive uplifts Δ_i are assumed. Leaves with estimated negative uplifts would be excluded from the first for every respectable churn prevention campaign. Let in addition C_i be the center of the leaf i consisting of the average values of all features across the customers of the leaf i. Then the distance d_{ij} of two leaves i and j pursuant to an arbitrary metric, e.g. Minkowski, is defined as the distance of their centers C_i and C_j appropriate to this very metric. Also let the best leaf b be defined as the leaf with the highest dedicated uplift

$$\Delta_b = \max_{i=1,..,I} \Delta_i \tag{3}$$

and the contained customers equivalently stand for the best customers in the same vein.

Typically for a churn prevention campaign, as well as for every other uplift campaign, the best customers are selected as far as the allocated budget allows it. That is one ignores distances and absolutely concentrates on uplifts. However, the paper presents selection methods (*best k*, *max dist*, *tradeoff* and *add split*), which take account of distances as well. Some of them are recent (*best k* for $k > 3$ and especially *add split*), while some of them were already introduced by Oechsle and Schönleber (2020). The subsequent listing defines them and distinguishes the classic selection method.

classic selects all the customers in the best leaf b and thus focuses on uplift.

best k randomly selects $1/k$ of the customers in the k best leaves and thus trades off uplift against diversification.

max dist randomly selects half of the customers in the best leaf b, and half of the customers in the leaf i where the distance to leaf b is maximal. Thus it focuses on distance.

tradeoff randomly selects half of the customers in the best leaf b, and half of the customers in the leaf t which is defined via

$$\Delta_t = \min_{i=1,..,I} \frac{\Delta_b - \Delta_i}{d_{bi}} \tag{4}$$

Thus it considers likewise distance and uplift.

add split artificially conducts an additional split in the best leaf b just as in the second best leaf, which respectively bisect the corresponding leaves concerning the quantity of customers. That is it selects half of the customers in the best leaf and half of the customers in the second best leaf with the pairwise highest distance. Thus it considers likewise distance and uplift.

4. Numerical evaluation

The minethatdata e-mail analytics and data mining challenge of Kevin Hillstrom (2008) marks the starting basis for our research. It is inspired by Diemert et al. (2018, 2021), Shaar et al. (2016) and the winning entry of Radcliffe (2008), who approached the exercise via uplift modeling. His underlying thoughts, independent of the won competition, are illustrated in a separate paper (Radcliffe, 2007a), albeit he zooms in on sales instead of churn.

Hillstroms dataset includes the results of an email marketing campaign relating to the customer behaviour in terms of website visits and purchasing. More precisely it contains 64.000 customers who last purchased within twelve months and afterwards were involved in an e-mail test ($2/3$ were randomly chosen to receive an e-mail campaign featuring merchandise, $1/3$ were randomly chosen to not receive an e-mail campaign). During a period of two weeks following the e-mail campaign anew purchases were tracked.

Therefore in the following research *Churn* is defined as *did not buy again in a certain period of time*, which is represented by the binary target variable *conversion*. Its two possible values, 1 for *customer purchased again within two weeks after the email campaign took place* and 0 for *customer did not purchase again within two weeks after the email campaign took place*, provide a churn prediction target as per definition of Pejić Bach et al. (2021) introduced in the first chapter. 578 out of Hillstroms 64.000 customers purchased again within the above mentioned two weeks. This is a conversion rate of 0.9% which fits to the rareness of the prediction target in ordinary churn prevention cases.

Against this background a decision tree has been developed on Hillstrom's dataset. Preparative tasks have been a) engineering of features to result in only dealing with numeric input variables (seven features), b) calculation of z-Scores for standardisation of the predictors and c) explicit exclusion of the information whether a customer was targeted by the e-mail campaign or not. Finally the tree itself was built on a 80/20 training/validation split of the sample.

There is no more model-tuning since the research does not seek for the best predicting model but one reasonable partitioning of the feature space into leaves in order to utilize the selection methods specific to decision trees.

So the feature space of Hillstroms dataset was sectioned into subareas: the leaves of the decision tree. Every single customer, also the 20% in the validation subset, could be assigned to its corresponding leaf. Casually spoken the whole dataset was scored with the on itself derived model. For this purpose the relative frequency of the value 1 of the binary target variable among the customers of the dedicated leaf defines the estimation of the conversion probability per leaf respectively its customers. Vice versa the complementary probability represents the likelihood of the above defined event churn according to the customers in that specific leaf.

To obtain the basic framework for the hereinafter described simulations the differentiation between the customers that received an email and those who did not preliminary was performed. That is the conversion or rather the churn probability grouped by email recipients and non email recipients was computed per leaf. By subtraction of the churn probability with email from the churn probability without email, Δ [cf. Eq. (2)] was generated as the real and correctly estimated effect of the churn prevention campaign per customer, in the absence of noise and uncertainty typified by spatially specified sources of error. For the generation of these errors the simulations adapt the concept of Oechsle and Schönleber (2020), which was previously outlined and discussed [cf. Eq. (1)].

As described above based on Hillstroms dataset a decision tree is engineered, which complies with the requirements of the methodology introduced in the third section. Concretely the tree consists of $I = 9$ leaves with $\Delta_i > 0$ for $i = 1, 2, .., 9$, whereas the uplifts represent the reduction in likelihood of churn (did not buy again) due to the email campaign in Hillstroms scenario. The chosen metric is the Euclidean distance.

In the following passage the selection methods as listed in section 3 are compared by a Monte Carlo simulation predicated on the described decision tree. An additional construction detail is the stipulated minimal leaf size of 4800 customers, which represents 7.5% of the whole dataset and respectively 9.4% of the training dataset. The reason is that this is an in practice imaginable campaign size and the quantitative comparability of the leaf sizes supports the elucidated selection methods.

Eight miscellaneous radii are used for the construction of the circularly occurring errors [cf.. (1)]

as listed in Table 2. The error radius R ranges from zero to two times $d_{\varnothing C_b}$, which is defined as the average Euclidean distance per customer to the center C_b of the best leaf b. While $R = 0$ serves as a baseline without failures, $R = 2d_{\varnothing C_b}$ somehow will mark a break even point when it comes to the economic logic of the prevention campaign.

The research performs 1000 runs per error radius and with it benchmarks six selection methods (classic, best k for $k = 2$, best k for $k = 3$, max dist, tradeoff, add split) by means of the expected Δ values per customer. The underlying decision tree is always the same, while the position of the error seed E randomly alters. Figure 1 visualizes the statistical distributions of the results, explicitly the distributions of the achieved average uplift per selected customer and per employed selection method. Table 1 depicts the averages per selection method (for the eight times 1000 runs) of achieved (and therefore expected) uplift, number of failures and achieved uplift among failures. More precisely $\mathbb{E}[\Delta]$ specifies the average (per 1000 runs) carried out average uplift per selected customer. The runs among the each undertaken 1000 runs overall that produce negative average uplifts per customer are counted as failures. Vice versa the complementary runs are counted as successes, which later will be relevant for the reading of Table 2. $\mathbb{E}[\Delta]$Successes and $\mathbb{E}[\Delta]$Failures consequently denote the respective average of the average uplifts generated by the dedicated successes and accordingly failures.

In Figure 1 it is very striking that the classic selection approach comes along with the highest level of uncertainty. That is the results of the classic selection method are furthest spread as measured by values of Δ. Conversely the alternative methods, second to none *best 3*, generate more dense ranges of outcomes. Particularly, as consolidated can be seen in Table 1, in comparison the classic approach not only most frequently (separate from *best 3*) led to failures, namely negative values of average uplift per customer (Δ), but also induced clearly more grave failures. This circumstance becomes even more apparent in Table 2 whose composition will be explained below.

Table 2 consists of 48 rows (eight radii times six selection methods), which respectively represent the results of the according unique radius and selection method combination in the above described each 1000 runs. To that effect column one and two identify the radii (as a multiple of $d_{\varnothing C_b}$) and the selection methods. $\mathbb{E}[\Delta]$, $\mathbb{E}[\Delta]$Successes and $\mathbb{E}[\Delta]$Failures, and therefore columns three to seven, have already been explained with Table 1. Concluding the column campaign size contains the number of contacted customers per selection method, which due to the simulation construction does not vary within the different runs. The analysis controls for this dimension to ensure comparability of the selection methods.

In the first column, as previously mentioned, the error radius varies from $R = 0$ to $R = 2d_{\varnothing C_b}$. While $R = 0$ constitutes a perfect surrounding with no need to deviate from the classic proceeding, $R = 2d_{\varnothing C_b}$ delivers failures with nearly every second run (469 out of 1000 for the classic method) and thus contests the general idea of preventing churn.

In-between these boundaries the superiority of the classic approach becomes apparent in terms of $\mathbb{E}[\Delta]$. But it is also the approach with the permanently lowest $\mathbb{E}[\Delta]$ Failures and an oftentimes highest number of failures. The alternative selection methods lower these effects. By doing so the add split approach is most suitable since it creates considerably the fewest failures. Additionally these few failures come along with the highest $\mathbb{E}[\Delta]$ Failures. Above all the add split selection demands the lowest risk premium (as measured by $\mathbb{E}[\Delta]$) for the gained robustness in results. In the case of $R/d_{\varnothing C_b}= 7/4$ even none.

5. Conclusion and discussion

The research described in this paper illustrates the in the subscription business well-known challenges with churn prevention campaigns on a real world dataset. It shows with the help of the previously churn-tailored Hillstrom dataset that noise and uncertainty represented by local spatial errors pose a veritable problem, which can economically destroy whole churn campaigns, especially with the classic selection approach. Thereby it naturally plays a decisive role how voluminous relevant arising errors are. Lastly it is demonstrated that there exist distance respecting alternative selection methods that largely give better results, dependent on the emergence of errors in terms of error radius R.

The most remarkable insight finally came from the *add split* selection. This method artificially conducts additional splits in the best leaves before it selects the customers in the thereby arising sub-areas with the pairwise highest Euclidean distance. It directly influences the generation of the decision tree itself, because depending on the interpretation of the dodge it either steps in the splitting rules or it intervenes in the pruning of the tree. By all means the *add split* selection method revealed the most promising results. That implies that there are situations in which it can be beneficial to diverge from common ways of decision tree construction by for example adding supposedly (by the textbook) needless splits. By departing from the concept of expected values this strategy evidently helps reducing abortive churn prevention campaigns.

In less risky scenarios ($R/d_{\varnothing C_b}$<= 1) there is no reason for not choosing the classic selection approach. However, in error-prone settings ($R/d_{\varnothing C_b}$> 1) distance respecting selection approaches based on decision trees are able to outperform the classic way. This appears in the reduced number of churn increasing churn prevention campaigns, as well as in the reduced extent of failures. In only slightly more inconvenient settings ($9/8 <= R/d_{\varnothing C_b} <= 5/4$) it is possible to reduce failures by switching from the classic method respectively even to avoid failures completely by using selection method *add split*. In clearly more inconvenient settings ($11/8 <= R/d_{\varnothing C_b} <= 3/2$) solely *add split* yields a respectable reduction to an acceptable level of uncertainty. In adverse surroundings ($R/d_{\varnothing C_b} >= 7/4$) the distance based methods again outperform the classic approach. Only the rationale of the campaign on the whole is questioned by a failure quota of $1/3$ to $1/2$.

In an overall view the findings can lead to feasible concepts for uplift modeling in general and especially in the churn prevention context, which will be of highest interest for the in all likelihood still growing subscription economy and the e-commerce business. At this juncture the methodology equipes each technically correct evolved decision tree with more reliability in practical applications and thus is a valuable tool for every practitioner.

References

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1):80–98.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. In *Proceedings of the National Academy of Sciences*, volume 113, pages 7353–7360. National Academy of Sciences.

Athey, S., Imbens, G., and Ramachandra, V. (2015). Machine learning methods for estimating heterogeneous causal effects.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., and Joshi, A. (2008). Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, pages 668–677.

Devriendt, F., Berrevoets, J., and Verbeke, W. (2021). Why you should stop predicting customer churn and start using uplift models. *Information Sciences*, 548:497–515.

Diemert, E., Betlei, A., Renaudin, C., and Amini, M.-R. (2018). A large scale benchmark for uplift modeling. In *KDD*.

Diemert, E., Betlei, A., Renaudin, C., Amini, M.-R., Gregoir, T., and Rahier, T. (2021). A large scale benchmark for individual treatment effect prediction and uplift modeling. *arXiv preprint arXiv:2111.10106*.

Droftina, U., Štular, M., and Košir, A. (2015a). A diffusion model for churn prediction based on sociometric theory. *Advances in Data Analysis and Classification*, 9(3):341–365.

Droftina, U., Štular, M., and Košir, A. (2015b). Predicting influential mobile-subscriber churners using low-level user features. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 56(4):522–534.

Gu, S., Ślusarczyk, B., Hajizada, S., Kovalyova, I., and Sakhbieva, A. (2021). Impact of the covid-19 pandemic on online consumer purchasing behavior. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(6):2263–2281.

Guelman, L., Guillén, M., and Pérez-Marín, A. M. (2015). Uplift random forests. *Cybernetics and Systems*, 46(3-4):230–248.

Hillstrom, K. (2008). The minethatdata e-mail analytics and data mining challenge. *MineThatData blog*.

Kane, K., Lo, V. S., and Zheng, J. (2014). Mining for the truly responsive customers and prospects using true-lift modeling: Comparison of new and existing methods. *Journal of Marketing Analytics*, 2(4):218–238.

Kusuma, P. D., Radosavljevik, D., Takes, F. W., and van der Putten, P. (2013). Combining customer attribute and social network mining for prepaid mobile churn prediction. In *Proc. the 23rd Annual Belgian Dutch Conference on Machine Learning (BENELEARN)*, pages 50–58. Citeseer.

Lai, L. Y.-T. (2006). *Influential marketing: a new direct marketing strategy addressing the existence of voluntary buyers*. PhD thesis, School of Computing Science-Simon Fraser University.

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):78–86.

Lo, V. S. and Pachamanova, D. A. (2015). From predictive uplift modeling to prescriptive uplift analytics: A practical approach to treatment optimization while accounting for estimation risk. *Journal of Marketing Analytics*, 3(2):79–95.

Manahan, C. (2005). A proportional hazards approach to campaign list selection. *SAS User Group International (SUGI) 30 Proceedings*.

Oechsle, F. and Schönleber, D. (2020). Towards more robust uplift modeling for churn prevention in the presence of negatively correlated estimation errors. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Oechsle, F., Setzer, T., and Blanc, S. M. (2016). On the assumptions of true lift models for churn prevention. In *Multikonferenz Wirtschaftsinformatik (MKWI) 2016, Technische Universität Ilmenau, 09.-11. März 2016, Band 2. Hrsg.: V. Nissen*, pages 1233–1244. Universitätsverlag Ilmenau.

Pejić Bach, M., Pivar, J., and Jaković, B. (2021). Churn management in telecommunications: Hybrid approach using cluster analysis and decision trees. *Journal of Risk and Financial Management*, 14(11):544.

Radcliffe, N. (2007a). Generating incremental sales: Maximizing the incremental impact of cross-selling, up-selling and deep-selling through uplift modelling. *Stochastic Solutions*, pages 1–10.

Radcliffe, N. (2007b). Using control groups to target on predicted lift: Building and assessing uplift model. *Direct Marketing Analytics Journal*, pages 14–21.

Radcliffe, N. J. (2008). Hillstrom's minethatdata email analytics challenge: An approach using uplift modelling. *Stochastic Solutions Limited*, 1:1–19.

Radcliffe, N. J. and Surry, P. D. (2011). Real-world uplift modelling with significance-based uplift trees. *White Paper TR-2011-1, Stochastic Solutions*, pages 1–33.

Rößler, J., Tilly, R., and Schoder, D. (2021). To treat, or not to treat: Reducing volatility in uplift modeling through weighted ensembles. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, page 1601.

Rzepakowski, P. and Jaroszewicz, S. (2010). Decision trees for uplift modeling. In *2010 IEEE International Conference on Data Mining*, pages 441–450. IEEE.

Shaar, A., Abdessalem, T., and Segard, O. (2016). Pessimistic uplift modeling. *arXiv preprint arXiv:1603.09738*.

Zhao, Y., Fang, X., and Simchi-Levi, D. (2017). Uplift modeling with multiple treatments and general response types. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 588–596. SIAM.

Zuora (2020). Zuora subscribed institute: Subscription economy index.

Table 1. Quintessence of runs with $R > 0$.

Selection Method	$\varnothing\mathbb{E}[\Delta]$	$\varnothing$# Failures	$\varnothing\mathbb{E}[\Delta]$ Failures
classic	0.008	181.7	-0.0039
best 2	0.0063	178.1	-0.0028
best 3	0.0058	190.1	-0.0024
max dist	0.0044	164.9	-0.0017
tradeoff	0.0063	178.9	-0.0029
add split	0.0069	143.4	-0.0022

Table 2. Summary of simulation results.

$R/d_{\varnothing C_b}$	Selection Method	$\mathbb{E}[\Delta]$	# Successes	$\mathbb{E}[\Delta]$ Successes	# Failures	$\mathbb{E}[\Delta]$ Failures	campaign size
0	classic	0.012	1000	0.012	0	–	6055
0	best 2	0.01	1000	0.01	0	–	6008
0	best 3	0.008	1000	0.008	0	–	6964
0	max dist	0.006	1000	0.006	0	–	6894
0	tradeoff	0.01	1000	0.01	0	–	5965
0	add split	0.01	1000	0.01	0	–	6084
1	classic	0.012	999	0.012	1	-0.001	6055
1	best 2	0.009	1000	0.009	0	–	6008
1	best 3	0.007	1000	0.007	0	–	6964
1	max dist	0.006	1000	0.006	0	–	6894
1	tradeoff	0.009	1000	0.009	0	–	5965
1	add split	0.01	1000	0.01	0	–	6084
9/8	classic	0.011	981	0.011	19	-0.001	6055
9/8	best 2	0.008	991	0.008	9	-0.001	6008
9/8	best 3	0.007	991	0.007	9	0.00	6964
9/8	max dist	0.006	997	0.006	3	0.00	6894
9/8	tradeoff	0.008	991	0.008	9	-0.001	5965
9/8	add split	0.009	1000	0.009	0	–	6084
5/4	classic	0.009	928	0.01	72	-0.002	6055
5/4	best 2	0.007	935	0.008	65	-0.001	6008
5/4	best 3	0.006	925	0.007	75	-0.001	6964
5/4	max dist	0.005	955	0.005	45	-0.001	6894
5/4	tradeoff	0.007	932	0.008	68	-0.001	5965
5/4	add split	0.008	999	0.008	1	0.00	6084
11/8	classic	0.008	862	0.01	138	-0.003	6055
11/8	best 2	0.007	859	0.008	141	-0.001	6008
11/8	best 3	0.005	822	0.007	178	-0.001	6964
11/8	max dist	0.005	877	0.005	123	-0.001	6894
11/8	tradeoff	0.007	858	0.008	142	-0.002	5965
11/8	add split	0.007	937	0.008	63	-0.001	6084
3/2	classic	0.007	790	0.01	210	-0.003	6055
3/2	best 2	0.006	770	0.008	230	-0.002	6008
3/2	best 3	0.005	724	0.007	276	-0.002	6964
3/2	max dist	0.004	798	0.005	202	-0.001	6894
3/2	tradeoff	0.006	766	0.008	234	-0.002	5965
3/2	add split	0.006	842	0.008	158	-0.001	6084
7/4	classic	0.005	637	0.01	363	-0.004	6055
7/4	best 2	0.004	635	0.009	365	-0.003	6008
7/4	best 3	0.004	629	0.007	371	-0.003	6964
7/4	max dist	0.003	650	0.006	350	-0.002	6894
7/4	tradeoff	0.004	635	0.009	365	-0.003	5965
7/4	add split	0.005	655	0.008	345	-0.002	6084
2	classic	0.004	531	0.011	469	-0.005	6055
2	best 2	0.003	563	0.009	437	-0.004	6008
2	best 3	0.003	578	0.007	422	-0.003	6964
2	max dist	0.002	569	0.006	431	-0.002	6894
2	tradeoff	0.003	566	0.009	434	-0.004	5965
2	add split	0.003	563	0.009	437	-0.003	6084

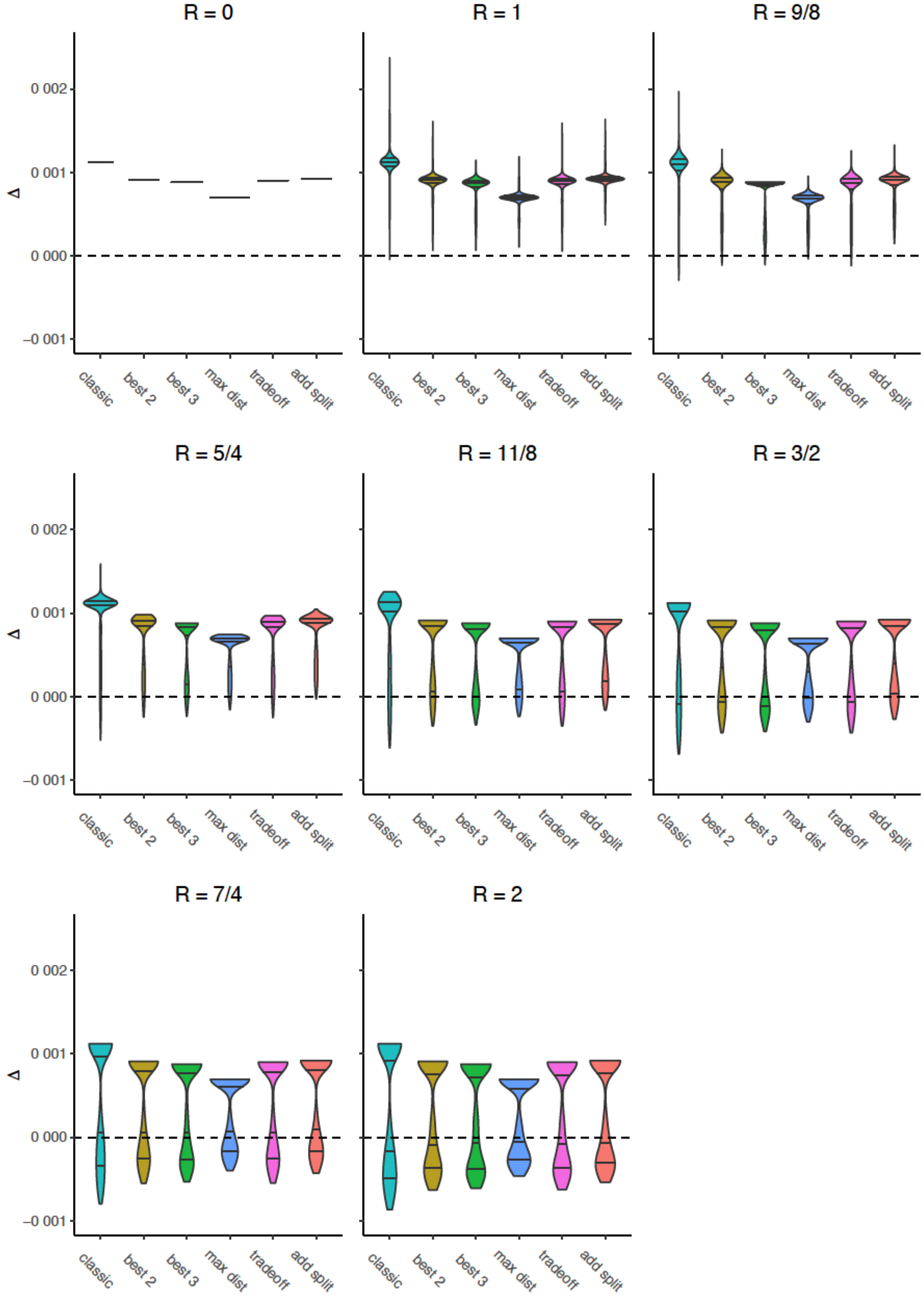

Figure 1. **Comparison of average uplift per selected customer for different selection methods and radii R (in units of $d_{\varnothing C_b}$).**

Frank Oechsle
[redacted]

Eidesstattliche Versicherung

gemäß § 13 Abs. 2 Ziff. 3 der Promotionsordnung des Karlsruher Instituts für Technologie für die KIT-Fakultät für Wirtschaftswissenschaften

1. Bei der eingereichten Dissertation zu dem Thema *Enhancing Robustness of Uplift Models used for Churn Prevention against Local Disturbances* handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich *wie folgt/ bislang nicht** an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

[redacted], den *09.02.2022*

* *Nicht Zutreffendes streichen. Bei Bejahung sind anzugeben: der Titel der andernorts vorgelegten Arbeit, die Hochschule, das Jahr der Vorlage und die Art der Prüfungs- oder Qualifikationsleistung.*