Predicting reach to find persuadable customers: improving uplift models for churn prevention

Théo Verhelst[1][0000−0002−2047−5690], Jeevan Shrestha[2], Denis Mercier[2], Jean-Christophe Dewitte[2], and Gianluca Bontempi[1][0000−0001−8621−316X]

[1] Machine Learning Group, Université Libre de Bruxelles, Brussels, Belgium
{Theo.Verhelst,Gianluca.Bontempi}@ulb.be
[2] Data Science Team, Orange Belgium, Brussels, Belgium
{Jeevan.Shrestha,Denis1.Mercier,Jean-Christophe.Dewitte}@orange.com

Abstract. Customer churn is a major concern for large companies (notably telcos), even in a big data world. Customer retention campaigns are routinely used to prevent churn, but targeting the right customers on the basis of their historical profile is a difficult task. Companies usually have recourse to two data-driven approaches: churn prediction and uplift modeling. In churn prediction, customers are selected on the basis of their propensity to churn in a near future. In uplift modeling, only customers reacting positively to the campaign are considered. Though uplift is better suited to maximize the efficiency of the retention campaign because of its causal aspect, it suffers from several estimation issues. To improve the uplift accuracy, this paper proposes to leverage historical data about the reachability of customers during a campaign. We suggest several strategies to incorporate reach information in uplift models, and we show that most of them outperform the classical churn and uplift models. This is a promising perspective for churn prevention in the telecommunication sector, where uplift modeling has failed so far to provide a significant advantage over non-causal approaches.

Keywords: Causal Inference, Churn Prediction, Uplift Modeling

1 Introduction

The telecommunication market is saturated, and companies need to invest in customer relationship management to keep their competitive edge. It is common knowledge that preventing churn is less expensive than attracting new customers [11]. The classical strategy for churn prevention consists in ranking customers according to their churn risk and offering the most probable to leave an incentive to remain (e.g. a promotional offer). Predicting churn is a difficult problem, involving large class imbalance, high dimension, latent information, low class separability, and large quantities of data. A wide variety of machine learning models have been applied to this problem in the literature [22, 30, 19, 13, 21, 26].

The pipeline for a typical customer retention campaign is outlined in Figure 1. First, a predictive model is trained on historical data from past campaigns. Then,

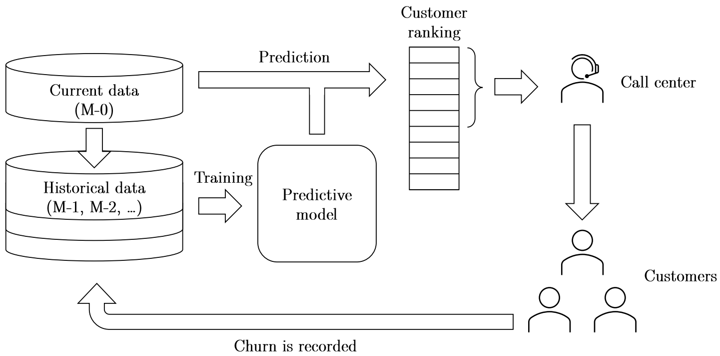

Fig. 1. Overview of the pipeline for customer retention.

this model predicts a score for each customer and ranks them accordingly. The list of customers with the highest scores is randomly split in a target and a control group, and the target group is sent to a call center. The call center contacts each of them individually, and the reaction of the customer is recorded and added to the historical data set for training future models.

The customers' ranking is provided by a predictive model estimating the probability of churn. This approach, however, disregards the causal aspect of the problem. Targeting high-risk customers is not necessarily the best strategy: for instance, some customers slightly less inclined to churn could be far more receptive to retention offers, and focusing the campaign on these customers could be more effective. This idea is exploited by uplift models, which estimate the causal effect of the campaign on an individual customer, rather than the risk of churn [10]. A wide variety of uplift models has been developed in the literature [17, 14, 1, 29, 8].

However, the added value of uplift modeling over churn prediction has been seldom assessed empirically. While it is clear that uplift is less biased than churn for estimating causal effects, the gain in performance is debated and context-dependent [5, 7, 27]. In settings such as customer retention, characterized by non-linearity, low class separability, and high dimensionality, the theoretical advantages of uplift might be insufficient to outweigh its drawbacks with respect to the usual strategy of churn prediction.

In this article, we suggest leveraging information about the reaction of the customer to the campaign to improve uplift estimation. In the marketing domain, *reach* denotes the proportion of the population exposed to the campaign, more specifically for advertisement campaigns [6]. In this article, we define *reach* as the reaction of the customer to the attempted call, that is, whether or not the customer picked up the phone and had a conversation with the phone operator. This variable is potentially informative about customer behavior, and, as a result, could improve the estimation of customer uplift. It is important to note that reach is only known after the campaign. Thus, it cannot be simply added as input to the model as an additional feature. We have to devise a dedicated approach to incorporate it into the learning process. In this sense, reach serves as an inductive bias for the uplift model, rather than an additional predictive feature. This paper shows that an uplift model, properly adapted to account for this new source of information, provides a significant improvement over the state-of-the-art.

The main contributions of this paper are:

- The proposal of 4 original strategies to incorporate reach in uplift models.
- An assessment of these strategies on a real-world data set from our industrial partner Orange Belgium, a major telecom company in Belgium.
- A significant improvement of uplift estimation, clearly outperforming state-of-the-art uplift models and the classical churn prediction approach.

The rest of this paper is divided as follows. In Section 2, we define basic notions in churn prediction and uplift modeling. In Section 3, we present reach modeling and various strategies to improve uplift estimation. In Section 4, we evaluate these strategies against several baselines, and we present our results in Section 5. We discuss our findings and suggest future work in Section 6.

2 Churn prediction and uplift modeling

In what follows, uppercase letters denote random variables, bold font denotes sets, and lowercase letters denote realizations of random variables. Causal inference notions are formalized using Pearl's notation [23]: an intervention fixing a variable T to a value t is noted $\mathrm{do}(T = t)$, and a random variable Y in a system under such an intervention is noted Y_t. For example, Y_0 is the churn indicator when the customer is in the control group ($T = 0$), whereas Y_1 is the churn indicator for the target group. We also denote customer features by a set of variables $\boldsymbol{X}$, with a realisation $\boldsymbol{x}$. Finally, R is the reach indicator ($R = 1$ for reached customer, $R = 0$ otherwise).

Let us first formalize in probabilistic terms the two main approaches for selecting customers in a retention campaign: churn prediction and uplift modeling. Churn prediction estimates the probability $P(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$ that a customer churns ($Y = 1$) given the customer descriptive features $\boldsymbol{x}$. Typical examples of descriptive features are tariff plan, metadata on calls and messages, mobile data usage, invoice amount, customer hardware, etc. Conventional supervised learning models can be used to predict churn [16, 20, 25, 24]. An extensive review of machine learning for churn prediction is given in [15]. The main drawback of this approach is the absence of causal insight: in fact, there is no indication that the campaign will be most effective on customers with a high probability of churn. The causal perspective is instead adopted by uplift modeling.

Uplift modeling estimates the causal effect of the campaign on the customer's probability of churn. To estimate this effect, it considers two scenarios: the intervention case $\mathrm{do}(T = 1)$ (i.e. the customer is offered an incentive) vs the control case $\mathrm{do}(T = 0)$ (i.e. the customer is not contacted). The uplift is the difference in the probability of churn between these two scenarios. For a set of descriptive features $\boldsymbol{X} = \boldsymbol{x}$, it is

$$U(\boldsymbol{x}) = P(Y_0 = 1 \mid \boldsymbol{X} = \boldsymbol{x}) - P(Y_1 = 1 \mid \boldsymbol{X} = \boldsymbol{x}). \quad (1)$$

Note that, unlike probabilities, uplift can be negative. A negative uplift indicates that the customer is more likely to churn when contacted by the call center. An uplift model is trained on historical data from one or more past campaigns with a randomized group assignment (target or control). The reaction of the customer (e.g. stay or churn) is then monitored for a fixed period of time, typically some months. The group assignment and customer churn records can then be used to update the historical data set, and subsequently train a new uplift model. Several approaches exist to estimate uplift, either using one or more predictive models [17, 14] or estimating uplift directly [1, 29, 8]. For a review of state-of-the-art uplift models, we refer the reader to [10].

3 Reach modeling

While uplift modeling is theoretically unbiased for maximizing campaign efficiency, there is some evidence in the literature that it suffers from estimation issues [7, 27]. This aspect can be so relevant as to cancel the benefits related to its causal design. Nevertheless, there is an additional piece of information that can be used to improve uplift estimation: the reaction of the customer to the call. More specifically, some customers will not pick up the phone, will hang up immediately, or more generally will not respond positively to the call. This information, automatically recorded by the call center, is a strong marker of customer receptivity. In email and online advertisement, a similar notion exists, under the name of click-through-rate [28] or response rate [12]. Although response models have been developed to improve direct marketing [2, 12, 9], current literature on uplift modeling ignores this information during the learning process. Expert knowledge in the telecom sector indicates that customers who do not pick up the phone or hang up immediately should be avoided because targeting them can increase their propensity to churn. We denote with $R = 1$ *reached* customers, i.e. customers who picked up the phone and had a dialogue with the phone operator. Otherwise, the customer is deemed *unreached* ($R = 0$). We present three ways to integrate reach information to improve uplift estimation. The four resulting equations are summarized in Table 1.

Reach probability as a feature The first approach (called `R-feature`) consists in building a predictive model of reach from historical data, and integrating the reach probability $\hat{r}$ among the input features of the uplift model. Note that we cannot directly plug the reach indicator as an input feature, since such information is not available before the campaign. This approach consists in learning the function $U(\boldsymbol{x}) = P(Y_0 = 1 \mid \boldsymbol{x}, \hat{r}) - P(Y_1 = 1 \mid \boldsymbol{x}, \hat{r})$.

Decomposition of probability The second approach (`R-decomp`) is based on the decomposition of the probability of churn with respect to the reach:

$$\begin{aligned} U(\boldsymbol{x}) &= P(Y_0 = 1 \mid \boldsymbol{x}) - P(Y_1 = 0 \mid \boldsymbol{x}) && (2) \\ &= P(Y_0 = 1 \mid \boldsymbol{x}) - P(R_1 = 0 \mid \boldsymbol{x}) P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 0) \\ &\qquad\qquad - P(R_1 = 1 \mid \boldsymbol{x}) P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1) && (3) \\ &= P(Y_0 = 1 \mid \boldsymbol{x}) - P(R_1 = 0 \mid \boldsymbol{x}) P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 0) \end{aligned}$$

$$
\begin{aligned}
&\qquad - [1 - P(R_1 = 0 \mid \boldsymbol{x})]P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1) && (4)\\
&= P(Y_0 = 1 \mid \boldsymbol{x}) - P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1)\\
&+ P(R_1 = 0 \mid \boldsymbol{x})\left[P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1) - P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 0)\right]. && (5)
\end{aligned}
$$

The last equation contains 5 terms but can be estimated with two uplift models and a simple classifier. The first two terms, $P(Y_0 = 1 \mid \boldsymbol{x}) - P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1)$, can be estimated with a uplift model by restricting the target group to reached customers. The third term, $P(R_1 = 1 \mid \boldsymbol{x})$, can be estimated by a predictive model of reach. The last two terms between brackets, $P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1) - P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 0)$, can also be returned by an uplift model, but using the reach indicator R instead of T as the treatment indicator for the model.

Bounds on uplift In marketing, there is empirical evidence that non-reached customers tend to have a negative uplift. Not reaching a customer has thus a doubly detrimental effect: the resources of the call center are wasted, and the customer is more likely to churn than if no call had been made. This domain knowledge may be translated into an inequality $P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 0) \geq P(Y_0 = 1 \mid \boldsymbol{x})$. We derive the third approach (`R-upper`) using this assumption and the decomposition in Equation (3):

$$
\begin{aligned}
U(\boldsymbol{x}) &= P(Y_0 = 1 \mid \boldsymbol{x}) - P(Y_1 = 1 \mid \boldsymbol{x})\\
&\leq (1 - P(R_1 = 0 \mid \boldsymbol{x}))P(Y_0 = 1 \mid \boldsymbol{x})\\
&\quad - P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1)P(R_1 = 1 \mid \boldsymbol{x})\\
&= P(R_1 = 1 \mid \boldsymbol{x})\left[P(Y_0 = 1 \mid \boldsymbol{x}) - P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1)\right]. && (6)
\end{aligned}
$$

Equation (6) requires two models: a simple predictive model of the reach variable (using only the target group), and an uplift model where the target group has been restricted to reached customers.

A symmetrical reasoning may lead to the hypothesis that a reached customer is less likely to churn than if not contacted: $P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 1) \leq P(Y_0 = 1 \mid \boldsymbol{x})$. From such assumption and (3), we derive a lower bound:

$$
U(\boldsymbol{x}) \geq P(R_1 = 0 \mid \boldsymbol{x})\left[P(Y_0 = 1 \mid \boldsymbol{x}) - P(Y_1 = 1 \mid \boldsymbol{x}, R_1 = 0)\right]. \qquad (7)
$$

Equation (7) is similar to Equation (6) but it requires the probability of not being reached, and the target group of the uplift model's training set is restricted to non-reached customers. This approach is named `R-lower`. Note that, among all methods presented in this section, `R-upper` and `R-lower` are the only biased estimators of uplift (since they estimate a bound instead). `R-feature` and `R-decomp` both estimate uplift, although they differ in the way they incorporate reach information.

4 Experiment

This experimental session benchmarks the approaches of Section 3 against several baselines:

- `Uplift`: An uplift model with no information about reach.

Approach	Equation
`R-feature`	$P(Y_0 = 1 \mid \hat{r}) - P(Y_1 = 1 \mid \hat{r})$
`R-decomp`	$P(Y_0 = 1) - P(Y_1 = 1 \mid R_1 = 1)+$ $P(R_1 = 0) \cdot [P(Y_1 = 1 \mid R_1 = 1) - P(Y_1 = 1 \mid R_1 = 0)]$
`R-upper`	$P(R_1 = 1)\,[P(Y_0 = 1) - P(Y_1 = 1 \mid R_1 = 1)]$
`R-lower`	$P(R_1 = 0)\,[P(Y_0 = 1) - P(Y_1 = 1 \mid R_1 = 0)]$

Table 1. Summary of the approaches used to integrate reach in uplift modeling. The conditioning on $\boldsymbol{x}$ is implicit in every term.

- `ML approach`: A classical churn prediction model[3] returning $P(Y = 1 \mid \boldsymbol{x})$.
- `R-target`: Using the estimated probability of reach as a score, that is, $P(R = 1 \mid \boldsymbol{x})$.

Since the first two baselines are state-of-the-art strategies, it is important to check whether incorporating reach information outperforms those approaches. The baseline `R-target` is introduced to check whether the reach alone may be used to find persuadable customers. Based on previous experiments [27], we used the X-learner algorithm [17] to build uplift models, and random forests [3] to learn churn and reach predictive models. The unbalancedness between churners and non-churners is addressed with the EasyEnsemble strategy [18], averaging models trained on positive instances (churners) with models trained on equally-sized sampled subsets of negative instances (non-churners).

The dataset is provided by our industrial partner Orange Belgium and relates to a series of customer retention campaigns in 2020, spanning over 3 months. A monthly dataset concerns about 4000 customers, for a total of 11896 samples. Each campaign includes a control group of about 1000 customers (for a total of 2886 control samples, 24.3% of the total), and a target group whose size depends on the load of the call center. Customer churn is monitored up to two months following the call. The churn rate in the control group is 3.6%, and 3.4% in the target group. The reach rate is 44.1% in the target group. Additional details cannot be disclosed for evident confidentiality reasons.

Results are evaluated in terms of uplift curve [10], which estimates the causal effect of the campaign for different numbers of customers. The uplift curve measures the difference in probability of churn between customers in the target and control groups. For a given predictive model f, and a threshold τ over the score provided by f, the uplift curve is defined as

$$\text{Uplift}(\tau) = P(Y_0 = 1 \mid f(\boldsymbol{X}) > \tau) - P(Y_1 = 1 \mid f(\boldsymbol{X}) > \tau). \tag{8}$$

This quantity is estimated empirically by subtracting the proportion of churners in the control and target groups, restricted to the customers with a score above the threshold. The uplift curve then is obtained by varying the threshold over all possible values.

In order to obtain a measure of the performance variability, we created 50 independent random splits of the data set into training and test sets, in propor-

[3] Note that `ML` stands for *maximum likelihood* of churn.

tion 80%/20%. Each of these splits is used to train each model, and we report the area under the uplift curve on the test set, averaged over the 50 runs.

We also evaluated several variations of the 4 approaches listed in Table 1. But, since they did not provide any significant improvement, we did not include them in the results. These variations are: i) the average of `R-lower` and `R-upper`, ii) the product of the reach and uplift model predictions, and iii) the average of the reach and uplift models prediction.

5 Results

Approach	AUUC
`R-feature`	0.857 ($\pm$0.547)
`R-decomp`	0.584 ($\pm$0.549)
`R-upper`	0.427 ($\pm$0.507)
`R-lower`	0.674 ($\pm$0.575)
`Uplift`	0.541 ($\pm$0.509)
`ML approach`	0.604 ($\pm$0.621)
`R-target`	0.247 ($\pm$0.397)

Table 2. Area under the uplift curve (AUUC), averaged over 50 runs. The confidence interval is one standard deviation. The best approach is underlined.

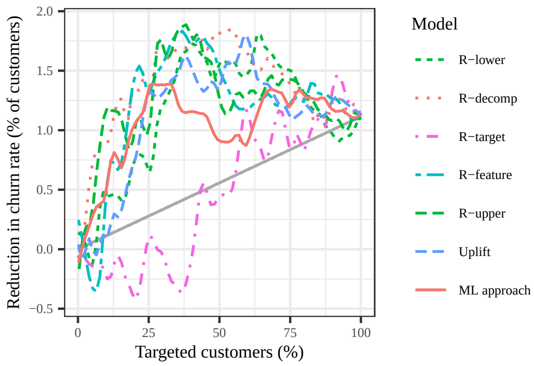

Fig. 2. Uplift curves for the first run of the experiment.

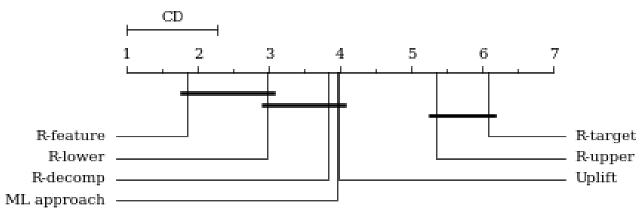

Fig. 3. Average ranking of the different approaches, with a line grouping approaches which do not have a significant rank difference. The critical mean rank difference is CD = 1.24, based on a Friedman-Nemenyi test with $p = 0.05$.

Table 2 reports the average area under the uplift curve (AUUC) over 50 runs while the uplift curves of the first run are in Figure 2. A Friedman-Nemenyi test of rank [4] is reported on figure 3, which indicates the mean rank of each approach over the 50 runs. A method is considered significantly better if the mean rank difference is larger than CD = 1.24, based on a p-value of $p = 0.05$. The best performing model, in terms of area under the uplift curve and standard deviation, is `R-feature`. It is significantly better than all other models, except for `R-lower`. Among the approaches integrating reach, `R-decomp` and `R-lower` perform similarly, while `R-upper` is not able to outperform the baselines. The two baselines `Uplift` and `ML approach` have similar performances, and, as expected, `R-target` performs quite poorly.

Note that, due to the small size of the dataset, the standard deviation of the AUUC is quite high. The data set contains only 11896 samples, 20% of these samples are used in the test set, and the churn rate is only a few percent. This leaves a very limited number of churners in the test set, and thus induces a high variability in the uplift curve between the different runs of the experiment.

6 Conclusion and future work

This paper shows the potential of reach information to improve the estimation of uplift. The superiority of reach models (such as `R-feature`) over conventional churn or uplift models is not surprising, since the information provided by the reach indicator is not available to the baseline methods. However, since reach information is not directly available before the campaign, specific strategies must be used. In these strategies, reach plays more the role of inductive bias than the one of churn predictor.

A potential advantage of this approach is that reach models are relevant to a wider range of use cases than churn prevention. It is common for telecom companies to perform different campaigns using the voice call channel, such as *up-sell* (to propose a better product to the customer), or *cross-sell* (to present additional products). A model of reach can be used in these contexts as well while using the same training data. This is a significant advantage, both in terms of computation time and data volume.

The applicability of this approach is limited by several factors. Firstly, as it is the case for all uplift models, it requires historical data from past retention campaigns. Our approach further requires records on the reaction of the customers to the call. This data might not be readily available for companies with no experience in direct marketing. Secondly, since uplift modeling is a new area of research, only a few uplift datasets are publicly available online. None of these datasets include information about reach. Therefore, it is difficult to assess new approaches exploiting reach information outside the scope of a collaboration with a private company.

We plan to evaluate our approach in future live retention campaigns. Currently, customer retention campaigns are still based on the churn prediction approach, since uplift models have failed so far to provide a significant improvement. This is a unique opportunity to evaluate the added value of our improved uplift model over the classical approach, and going beyond the use of historical data sets. From the perspective of a practitioner, several improvements of the approach can be devised: for example, we considered only the random forest model to predict reach. Other machine learning models might provide better performances. Also, our pipeline addresses class unbalancedness with the Easy Ensemble strategy, and the reach model is included during this step. Since the reach indicator is not as heavily imbalanced as the churn indicator, it might be beneficial to train the reach model separately. Finally, we did not investigate the use of more fine-grained reach information, such as the time of call, or a more detailed description of the customer's reaction. This could potentially further improve uplift estimation. Such detailed information can also be exploited proactively, by calling the customer at a time and a day which maximizes the probability of reach.

References

1. Athey, S., Imbens, G.: Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences **113**(27), 7353–7360 (7 2016). https://doi.org/10.1073/PNAS.1510489113, https://www.pnas.org/content/113/27/7353
2. Bose, I., Chen, X.: Quantitative models for direct marketing: A review from systems perspective. European Journal of Operational Research **195**(1), 1–16 (2009)
3. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
4. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research **7**, 1–30 (2006)
5. Devriendt, F., Berrevoets, J., Verbeke, W.: Why you should stop predicting customer churn and start using uplift models. Information Sciences (2019)
6. Farris, P.W., Bendle, N., Pfeifer, P.E., Reibstein, D.: Marketing metrics: The definitive guide to measuring marketing performance. Pearson Education (2010)
7. Fernández, C., Provost, F.: Causal Classification: Treatment Effect vs. Outcome Prediction. NYU Stern School of Business (2019)
8. Guelman, L., Guillén, M., Pérez-Marín, A.M.: Uplift random forests. Cybernetics and Systems **46**(3-4), 230–248 (2015). https://doi.org/10.1080/01969722.2015.1012892
9. Guido, G., Prete, M.I., Miraglia, S., De Mare, I.: Targeting direct marketing campaigns by neural networks. Journal of Marketing Management **27**(9-10), 992–1006 (2011)
10. Gutierrez, P., Gérardy, J.Y.: Causal Inference and Uplift Modelling: A Review of the Literature. In: Hardgrove, C., Dorard, L., Thompson, K., Douetteau, F. (eds.) Proceedings of The 3rd International Conference on Predictive Applications and APIs. Proceedings of Machine Learning Research, vol. 67, pp. 1–13. PMLR, Microsoft NERD, Boston, USA (1 2016), http://proceedings.mlr.press/v67/gutierrez17a.html
11. Hadden, J., Tiwari, A., Roy, R., Ruta, D.: Computer assisted customer churn management: State-of-the-art and future trends. Computers & Operations Research **34**(10), 2902–2917 (2007)
12. Hansotia, B.J., Rukstales, B.: Direct marketing for multichannel retailers: Issues, challenges and solutions. Journal of Database Marketing & Customer Strategy Management **9**(3), 259–266 (2002)
13. Idris, A., Khan, A.: Ensemble based efficient churn prediction model for telecom. In: Frontiers of Information Technology (FIT), 2014 12th International Conference on. pp. 238–244 (2014). https://doi.org/10.1109/fit.2014.52
14. Jaskowski, M., Jaroszewicz, S.: Uplift modeling for clinical trial data. In: ICML Workshop on Clinical Data Analysis (2012)
15. Kayaalp, F.: Review of Customer Churn Analysis Studies in Telecommunications Industry. Karaelmas Fen ve Mühendislik Dergisi **7**(2), 696–705 (2017). https://doi.org/10.7212/zkufbd.v7i2.875

16. Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., Abbasi, U.: Improved churn prediction in telecommunication industry using data mining techniques. Applied Soft Computing **24**, 994–1012 (2014). https://doi.org/10.1016/j.asoc.2014.08.041
17. Künzel, S.R., Sekhon, J.S., Bickel, P.J., Yu, B.: Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences of the United States of America **116**(10), 4156–4165 (2019). https://doi.org/10.1073/pnas.1804597116
18. Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **39**(2), 539–550 (2009). https://doi.org/10.1109/tsmcb.2008.2007853
19. Mitrović, S., Baesens, B., Lemahieu, W., De Weerdt, J.: On the operational efficiency of different feature types for telco Churn prediction. European Journal of Operational Research **267**(3), 1141–1155 (2018). https://doi.org/10.1016/j.ejor.2017.12.015, https://doi.org/10.1016/j.ejor.2017.12.015
20. Olle, G.D.O., Cai, S.: A hybrid churn prediction model in mobile telecommunication industry. International Journal of e-Education, e-Business, e-Management and e-Learning **4**(1), 55 (2014)
21. Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., Vanthienen, J.: Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. Expert Systems with Applications **85**, 204–220 (2017). https://doi.org/10.1016/j.eswa.2017.05.028
22. Óskarsdóttir, M., Van Calster, T., Baesens, B., Lemahieu, W., Vanthienen, J.: Time series for early churn detection: Using similarity based classification for dynamic networks. Expert Systems with Applications **106**, 55–65 (2018). https://doi.org/10.1016/j.eswa.2018.04.003
23. Pearl, J.: Causality: models, reasoning, and inference. No. 6, Cambridge university press (2009)
24. Umayaparvathi, V., Iyakutti, K.: Attribute selection and Customer Churn Prediction in telecom industry. In: 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE). pp. 84–90. IEEE (2016)
25. Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G., Chatzisavvas, K.C.: A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory **55**, 1–9 (2015). https://doi.org/10.1016/j.simpat.2015.03.003, http://dx.doi.org/10.1016/j.simpat.2015.03.003
26. Verbeke, W., Martens, D., Baesens, B.: Social network analysis for customer churn prediction. Applied Soft Computing **14**, 431–446 (2014). https://doi.org/10.1016/j.asoc.2013.09.017
27. Verhelst, T., Caelen, O., Dewitte, J.C., Bontempi, G.: Does causal reasoning help preventing churn? Under submission (2021)
28. Winer, R.S.: A framework for customer relationship management. California management review **43**(4), 89–105 (2001)
29. Zaniewicz, L., Jaroszewicz, S.: Support vector machines for uplift modeling. In: 2013 IEEE 13th International Conference on Data Mining Workshops. pp. 131–138. IEEE (2013)
30. Zhu, B., Baesens, B., vanden Broucke, S.K.L.M.: An empirical comparison of techniques for the class imbalance problem in churn prediction. Information sciences **408**, 84–99 (2017). https://doi.org/10.1016/j.ins.2017.04.015