Academic Platform Journal of Engineering and Science 8-3, 450-455, 2020

Academic Platform Journal of Engineering and Science

journal homepage: http://apjes.com/

The Effect of the Length of the Customer Event History and the Staying Power of the Predictive Models in the Customer Churn Prediction: Case Study of Migros Sanal Market

*[1]Birol Yüceoğlu,

[1]Migros T.A.Ş., Turkey, byuceoglu@migros.com.tr,

Research Paper Arrival Date: 08.08.2019 Accepted Date: 10.03.2020

Abstract

The customer churn prediction problem is studied for various sectors under several aspects. In this study, we consider the effect of the length of the customer event history and the staying power of the predictive models for the churn prediction problem of a leading online fast-moving consumer goods retailer in Turkey. These are important aspects of the churn prediction models as they help decision makers to determine the optimal length of the past data for predicting the customer churn as well as lifespan of the predictive models. We find that the length of the customer event history logarithmically increases the predictive power of models, validating findings in the literature in the newspaper subscription sector. Regarding the staying power of the predictive models, we conclude that the models in online fast-moving consumer goods retailing has a slightly longer lifespan that the models discussed in the literature for an Internet service provider and an insurance company.

Keywords: The customer churn, Machine learning, Supervised learning, Length of customer event history, Staying power.

1. INTRODUCTION AND RELATED LITERATURE

The customer churn is an important topic in marketing and customer relationship management (CRM) in various sectors. The main reasons for the importance of the customer churn is increasingly strong competition and the monetary effects of losing customers. In addition to lost sales, churning customers provide bad word-of-mouth for the company. They can create negative experiences and have the potential to convince other loyal customers to churn. Furthermore, finding new customers is five times costlier than keeping the existing customers, Rust and Zahorik [15]. It is therefore, essential to predict the customer churn for preventing losing customers that are predicted to churn in the future.

The customer churn prediction problem has been studied for various sectors, including financial services, insurance, telecommunications and retailing. Glady et al. [7] predict the customer churn for a Belgian financial services company using customer lifetime value analysis. Advances in the data analytics allows various approaches to be used in the customer churn prediction problem. Lee et al. [10] use time-series classification to predict the customer churn for a telecommunications company. Kim et al. [9] use social networks to observe propagation patterns of the customer churn in customers' networks. In terms of conventional data analytics approaches, Lemmens and Croux [11] use bagging and boosting to predict the customer churn for a telecommunications company. Note that these examples are from contractual sectors, where switching to a competitor is more difficult or in some cases costlier, either due to the cost of breaking a contract or the cost of the effort of breaking a contract. Predicting the customer churn in retailing provides an additional challenge as it is required to define customers and churning customers in this non-contractual setting. In a non-contractual setting, such as retailing, customers have lower switching costs. Therefore, it is easier for the customer to switch part of their purchases to a competitor. This event is called partial defection. Usually, partial defection leads to total defection as the customers switch more and more of their purchases to the competitors. Buckinx and Van den Poel [3] study customer attrition in FMCG retail, which requires defining customers and churning customers. They use RFM (recency, frequency, monetary) variables to predict the customer churn. The predictors that they use are also employed in this analysis. To decide on their customer base, they use a five months period of observation. The customers are included in the model if they are more frequent, i.e. their mean frequency is above average, and regular, i.e. the coefficient of variation for their interpurchase time is below average. They are labeled as churners if they do not satisfy this criterion in the next five months. In a recent paper, Martinez et al. [12] predict customers' future repurchase behavior in a non-contractual B2B setting. They use approaches like Lasso logistic regression, extreme learning machine and gradient boosting to predict customers' future repurchase behavior. In this B2B setting, predicting the

* Corresponding author: Birol Yüceoğlu: Migros T.A.Ş. Atatürk Mahallesi Turgut Özal Bulvarı No 7 34758 Ataşehir, İstanbul, Türkiye, byuceoglu@migros.com.tr, +90 216 579 3278

Doi: 10.21541/apjes.603809

customer churn becomes especially important for planning purposes.

The customer churn in a multi-channel setting is important for different purposes as well. Hernant and Rosengren [8] study a Swedish grocery retailer that starts to operate an online store in addition to the physical stores. They find that the customers who use the physical store and start to also buy online are more profitable than the single channel customers or the online customers who start to buy from physical stores. Thus, predicting and preventing the customer churn is important for not losing the most profitable customers in a multi-channel setting.

In addition to predicting the customer churn, other aspects of the churn prediction problem have also been studied. Ballings and Van den Poel [1] study the effect of the length of customer event history. The authors create predictive models using historical customer data from one to sixteen years for a newspaper company. They conclude that the predictive performance of models increases as the length of the event history increases. However, this increase is logarithmic and after five years, adding past data has only marginal effect on the performance of the predictive models. They leave further study of the topic for other sectors as an open question. We make a similar observation and assert that the observations of authors are valid also in online FMCG retailing and that six months is a good choice for the length of the event history of the customers to predict the customer churn. This finding has the potential to impact the customer churn prediction models in the retail industry in terms of determining the length of the period of observation.

Risselada et al. [14] study the staying power of various models for churn prediction. The authors study the staying power of various models for the churn prediction problem. After building a model, they use the same model for predicting upcoming periods and compare the results obtained in terms of top-decile lift and Gini coefficient. They use decision trees and logistic regression (with and without bagging) and report their results on two data sets from an Internet service provider and an insurance company. They conclude that the predictive power of the models for the Internet service provider decreases two periods after the period of the estimation. For the insurance case, the predictive quality decreases one period after the estimation period. They leave further study of the topic for other sectors as an open question. We observe that the staying power of our prediction model is slightly longer. We assert that our model can predict two periods after the period of estimation with a reasonable success. However, moving further to three periods marginally decreases the performances of the models. In other words, it is less crucial to retrain the existing models in the retail industry as the staying power of the predictive models is longer.

Studying these aspects of the customer churn prediction models is important for decision makers on finding the optimal length of the customer event history, balancing the performance of the models and their data requirements, and on deciding the frequency for updating the predictive models. In this non-contractual setting, deciding on the length of the customer event history also allows us to define customers to be included in the analysis. This study focuses on these aspects of the customer churn prediction for the online store of Migros Turkey, one of the leading fast-moving consumer goods (FMCG) retailers in Turkey. The online store, Sanal Market, is founded at 1997 and is one of the pioneers of online FMCG retailing in Turkey. Currently, Sanal Market operates in more than 20 cities with more than 20.000 products. Even though Sanal Market can still be considered as an alternative sales channel for Migros Turkey, the sector is growing much more rapidly than conventional retailing and with stronger competition, customer retention measures are essential for Migros Turkey to protect and grow its online customer base.

The rest of this paper is organized as follows: In Section 2, we describe the methodology, the approaches, and the time frame for the study. In Section 3, we investigate the effect of the length of the customer event history and conclude that the length of the customer event history logarithmically increases the performance of the predictive models in online FMCG retailing as stated by Ballings and Van den Poel [1] in the newspaper subscription sector. In Section 4, we discuss the staying power of the predictive models. We find that the models in the online FMCG retailing have a slightly longer lifespan that the models for sectors discussed by Risselada et al. [14]. In Section 5, we present our concluding remarks.

2. METHODOLOGY

We use supervised learning to predict the customer churn and compare the results obtained by three approaches: logistic regression, random forests, gradient boosting. In the supervised learning, ensemble methods are frequently used in classification problems. Breiman [2] introduces the random forests, an ensemble technique based on growing many decision trees with bootstrapping and selecting feature subsets and combining their results. As a result, the final model has some randomness and the approach is known to be averse to over-fitting. For this reason, the approach is frequently used in classification problems. The gradient boosting, introduced by Friedman [5,6] is an iterative boosting algorithm, where weak learners are trained so that the errors from the previous iterations are corrected. The approach becomes more and more popular because of several open-source packages employing the algorithm, such as Scikit-Learn, LightGBM, and Xgboost. Logistic regression, Cox [4], is a statistical technique based on linear regression models that is developed for categorical dependent variables. The approach is frequently used as a baseline approach in the classification problems. This study uses Scikit-Learn machine learning library, implemented for Python, Pedregosa et al. [13], where all three approaches are present.

In this study, we decide to restrict the customer base to focus on loyal and regular customers. Upon discussions with the business units, we decided to focus on customers having transactions in two different months in the last six months. These customers are considered as more regular and loyal, so their churn is more important to the company. Note that this also prevents us from obtaining artificially good results. Without such a restriction, new customers or irregular customers can also be included in the analysis. However, they are more prone to churn, which can disrupt the patterns in the data set and inflate the number of churners. This also may have the effect of falsely improving the results by only predicting the churn of irregular or new customers correctly. For the customer churn, we use two definitions, again based on the discussions with the business units. In the short term total defection, we predict customers who are not going to use the online store in the next month. In the long term total defection, we predict customers who are not going to use the online store in the next three months. We only consider the total defection scenarios as they are more valuable for the business units. The results of this study are also valid for the partial churn scenario. We use the transaction data from different time periods in our experiments. Our empirical study consists of two parts:

- Length of Customer Event History: In this part, we study the effect of the length of customer event history using the transaction data from 1 September 2015 to 30 November 2016. We use the data from September, 2016 to label the customers in the short term total defection scenario and the last three months (September, 2016 - November, 2016) to label the customers in the long term total defection scenario. Then, we use the past customer transaction data from one month (August, 2016) to 12 months (September, 2015 - August, 2016) by increasing the time interval by one month for each experiment. In all these experiments, 70% training - 30% testing split is used for those customers with the corresponding time interval.

- Staying Power: In this part, we examine the staying power of our predictive models. We build our predictive models using the data from 1 June 2015 to 29 February 2016. Again, December, 2015 is used for labelling the customer for the short term total defection scenario and the last three months (December, 2015 - February, 2016) are used for labelling the customer for the long term total defection scenario, and the data is split into two as 70% training and 30% testing. Then, we use this model to predict the churn for seven periods. The first period, labelled *t*, covers between 1 September 2015 to 31 May 2016. We use the same model to predict the customer churn for the subsequent periods from 1 October 2015 to 30 June 2016 (for period *t+1*), until 1 March 2016 to 30 November 2016 (for period *t+6*).

3. THE EFFECT OF THE LENGTH OF THE CUSTOMER EVENT HISTORY

In this section, we answer a question from the literature about the effect of the length of the customer event history. The length of the customer event history is especially important in determining the data requirements of churn prediction models and can shed some light on how models evolve in the presence of the historic data. Furthermore, storing and processing data require resources and determining the optimal length of customer event history is also important in terms of efficiency. Determining the optimal length of the customer event history also gives insight into the behaviour of the customers and can be used to determine the customer base.

Ballings and Van den Poel [1] observe that the length of customer event history logarithmically increases the performance of the prediction in newspaper subscription. In their experiments, they use data belonging to 16 years to predict the customer churn or a newspaper subscription. They observe that the data belonging to the most recent five years performs almost as well as their best prediction. They state in Section 6:

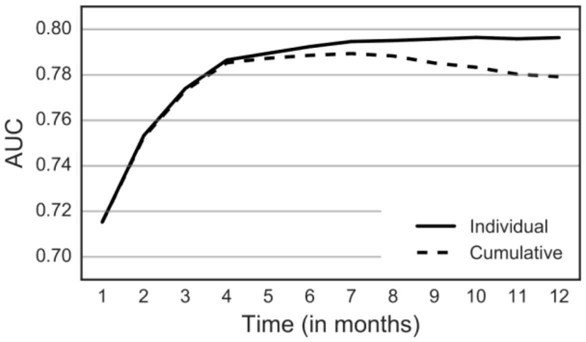
> *"Although we are confident about our results, it is unclear whether they can be generalized over a wider range of subscription services. Hence, it would be interesting to validate our findings on customer databases in other industries."*

We study the effect of the length of customer event history by predicting the customer churn using the data belonging to 12 months. The set of features we use include the transaction amounts (total, product categories), the number of transactions and transaction days, the discount percentages per product categories, the number of distinct categories purchased from, and the private label purchases. We create 25 features for each month and use two types of analysis in terms of how we include the features in our analysis.

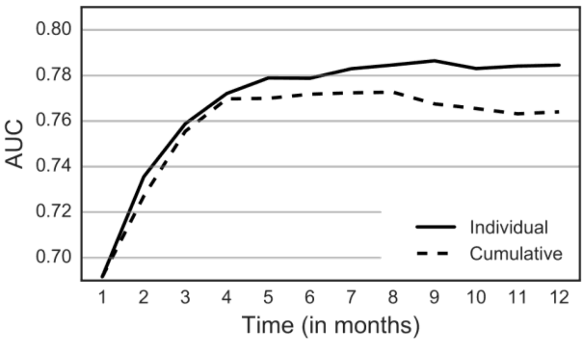
Our performance measure is area under curve. In cumulative analysis, we use 25 features and whenever we include the data from previous months we use aggregate values. In the individual analysis, we use 25 features for each month included in the analysis, up to 300 features for 12 months. We build 12 models to study the effect of the length of the customer event history.

We observe that the performance of our predictions increases logarithmically with respect to the length of the customer event history. In the cumulative analysis after six or seven months, the performance slightly decreases in some cases. In the individual analysis, the performance of the predictive model does not decrease with respect to the length of the customer event history. In most of the cases, the performance after six or seven months is quite close to the best model. The results for short term total defection are presented in Figure 1 and the results for long term total defection are presented in Figure 2. Our results are like Ballings and Van den Poel

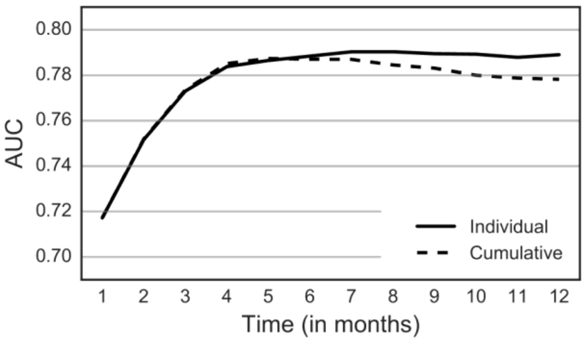
[1] and answer the generalization question from online FMCG retailing perspective. We find that the length of the customer event history logarithmically increases the performance of predictive models and data belonging to the most recent six or seven months is enough to predict the customer churn. The length of the customer event history can also be used to determine the customer base of the company for the churn prediction problem.



(a) Gradient boosting



(b) Random forest



(c) Logistic regression

Figure 1. The results for the effect of the length of the customer event history for short term total defection. The performance of the models increases only marginally after six or seven months.

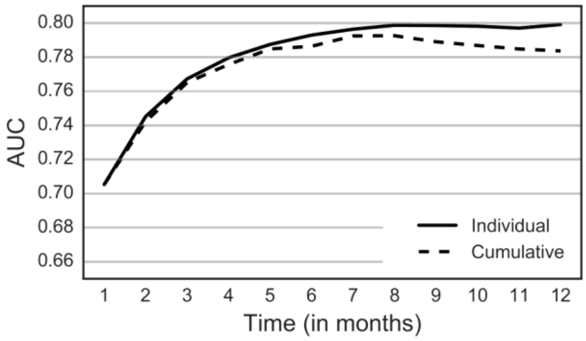
4. THE STAYING POWER OF THE PREDICTIVE MODELS

In this section, we answer another question from the literature about the staying power of the predictive models. Risselada et al. [14] study the predictive qualities of models in the periods after the estimation period for an Internet service provider and an insurance firm. They observe that the quality of a model decreases one or two periods after the period of the estimation. They state in page 207:

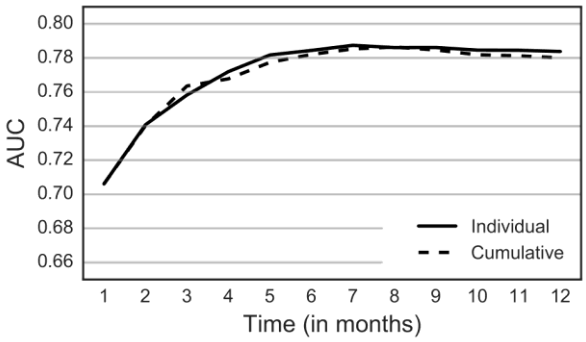
> *"Although we are confident with the results, it remains unclear whether they are generalizable over a broader range of services that the two we studied."*

To observe the staying power of the churn prediction models for the online FMCG retailer, we use the three models mentioned in Section 2, namely gradient boosting, random forests, and logistic regression, for the short term and the long term total defection. After creating a model to predict the customer churn in period *t*, we use the same model without retraining to predict the customer churn in the subsequent six periods, up to the period *t+6*. We use data belonging to 18 months (from June, 2015 to November, 2016) and features mostly related to RFM variables, proved to be effective predictors in the customer churn by Buckinx and Van den Poel [3]. We evaluate the performances of the models using two criteria: AUC and top-decile lift. The set of predictors that we use are presented in Table 1. We use similar variables to Buckinx and Van den Poel [3], but we also include the physical store purchases of the customers as Sanal Market operates in a multi-channel environment. The predictors that we use allows us to measure the effect of the recency, frequency, monetary values of the customers as well as their relationship with the product categories, their basket contents and their trust relationship with the company. We train our model using the transaction data from June, 2015 to February, 2016, with the last three months providing labels for our models. We train the models to predict the customer churn in time period *t* (between September, 2015 and May, 2016) and use the same model to predict the customer churn in six subsequent periods. The performances of the models increase in period *t+1* for the short term total defection (between October, 2015 and June, 2016, see Figure 3(a)) and decrease only slightly in period *t+2* (between November, 2015 and July, 2016). However, after the third period the predictive power of a model decreases significantly. The increase in the predictive performance at period *t+1* is due to the increased churn probability at that period. Our observations hold for both AUC and top-decile lift in both scenarios. For the long term total defection, the models lose their predictive power after two periods, at *t*+2, see Figure 4(a). The results for the short term total defection scenario are presented in Figure 3 and the results for the long term total defection are presented in Figure 4. We assert that the staying power of the churn prediction models for online FMCG retailing is slightly longer than the sectors studied by Risselada et al. [14]. All three approaches used in the prediction behave similarly, showing the robustness of the results in terms of approaches selected.

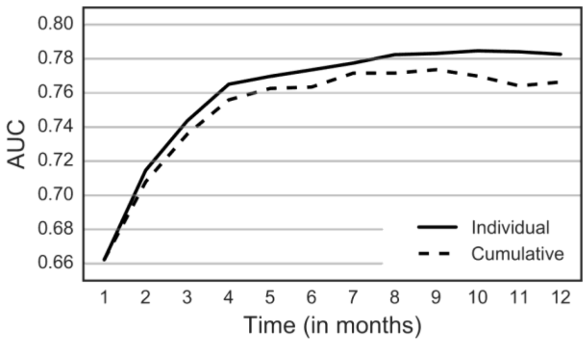
In cases where the customer base is large, and the data preparation and model training are cumbersome, these results suggest that using a trained model for two subsequent time periods does not cause a significant decrease in the predictive performance of the models.



(a) Gradient boosting



(b) Random forest

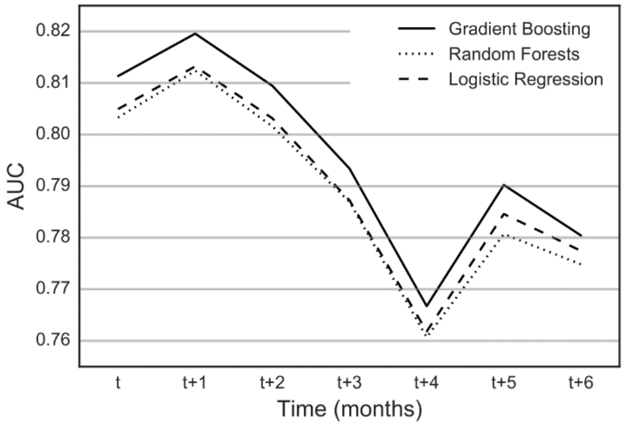


(c) Logistic regression

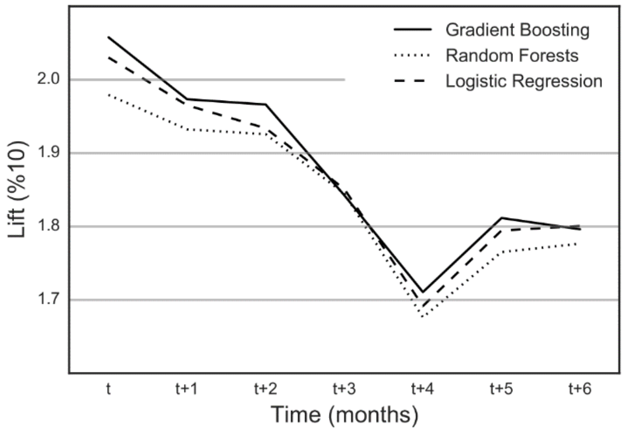
Figure 2. The results for the effect of the length of the customer event history for long term total defection. The performance of the models increases only marginally after six or seven months.

5. CONCLUSION

We have presented a computational study regarding two questions from the churn prediction literature, which are related to the effect of the length of the customer event history and the staying power of the prediction models for FMCG online retailing, respectively. We observe that the length of the customer event history increases the performance of the prediction models logarithmically as observed by Ballings and Van den Poel [1]. Specifically, we see that the performance of the predictive models increases when the customer event history is lengthened to up to six



(a) AUC score of churn prediction model created for time period *t.*

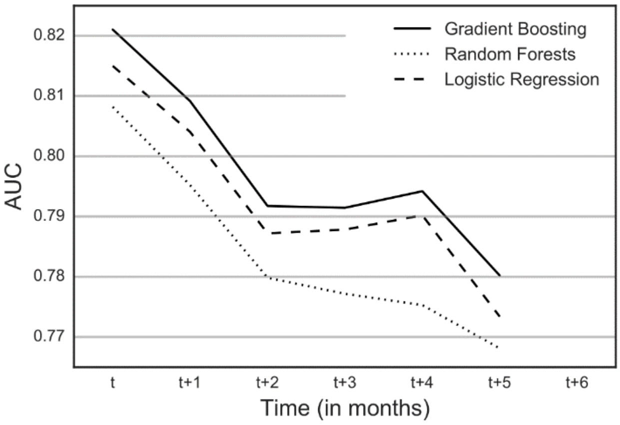


(b) Top-decile lift score of the churn prediction model created for the time period *t.*

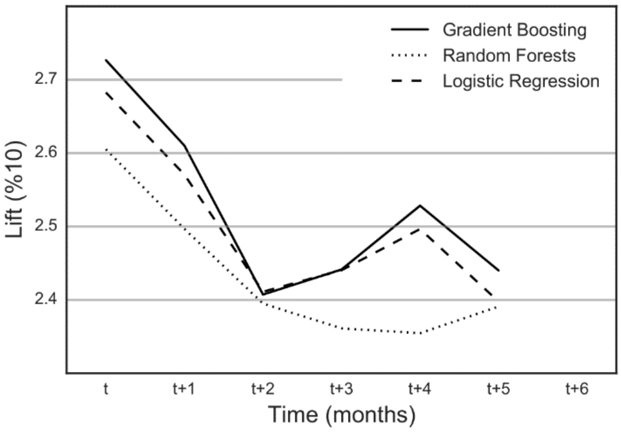
Figure 3. The staying power of the predictive models for the short term total defection scenario.

months. After that, the performance gains are only marginal. In addition to that, we assert that the staying power of a churn prediction model in online retailing is slightly longer than it is in Internet service providing and insurance sectors, Risselada et al. [14]. We observe that the customer churn prediction models can be used for two or three months (based on the definition of the customer churn), which is especially important when the customer base is large and training new models takes considerable time.

The findings have important implications in terms of CRM and data warehousing. Using customer data belonging to the most recent six or seven months can greatly reduce the data storage and processing requirements. Similarly, being able to use the same predictive models for two or three months can provide advantages to various business units in terms of managing predictive models and manageability of CRM actions.



(a) AUC score of churn prediction model created for time period *t.*



(b) Top-decile lift score of the churn prediction model created for the time period *t.*

Figure 4. The staying power of the predictive models for the long term total defection scenario.

ACKNOWLEDGEMENT

This research was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under TEYDEB 1501 program, under grant numbered 3150376.

REFERENCES

[1] M. Ballings and D. Van den Poel, 2012. Customer event history for churn prediction: How long is long enough? Expert Syst Appl, 39(18), 13517-13522.

[2] L. Breiman, 2001. Random forests. Mach Learn, 45(1), 5-32.

[3] W. Buckinx and D. Van den Poel, 2005. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. Eur J Oper Res, 164(1):252-268.

[4] D. R. Cox, 1958. The regression analysis of binary sequences (with discussion). J R Stat Soc Series B Stat Methodol, 20, 215-242.

[5] J. H. Friedman, 2001. Greedy function approximation: A gradient boosting machine. Ann Stat, 29(5), 1189-1232.

[6] J. H. Friedman, 2002. Stochastic gradient boosting. Comput Stat Data Anal, 38(4), 367-378.

[7] N. Glady, B. Baesens, and C. Croux, 2009. Modeling churn using customer lifetime value. Eur J Oper Res, 197(1), 402-411.

[8] M. Hernant and S. Rosengren, 2017. Now what? Evaluating the sales effects of introducing an online store, J Retail Consum Serv, Volume 39, 305-313.

[9] K. Kim, C.-H. Jun, and J. Lee, 2014. Improved churn prediction in telecommunication industry by analyzing a large network. Expert Syst Appl, 41(15), 6575-6584.

[10] Y.-H. Lee, C.-P. Wei, T.-H. Cheng, and C.-T. Yang, 2012. Nearest-neighbor-based approach to timeseries classification. Decis Support Syst, 53(1), 207-217.

[11] A. Lemmens and C. Croux, 2006. Bagging and boosting classification trees to predict churn. J Mark Res, 43(2), 276-286.

[12] A. Martínez, C. Schmuck, S. Pereverzyev, C. Pirker, and M. Haltmeier, 2018. A machine learning framework for customer purchase prediction in the non-contractual setting. Euro J Oper Res, https://doi.org/10.1016/j.ejor.2018.04.034.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011. Scikit-learn: Machine learning in Python. J Mach Learn Res, 12, 2825-2830.

[14] H. Risselada, P. C. Verhoef, and T. H. Bijmolt, 2010. Staying power of churn prediction models. J Interact Market, 24(3), 198-208.

[15] R. T. Rust and A. J. Zahorik, 1993. Customer satisfaction, customer retention, and market share. J Retailing, 69(2), 193-215.