Literature review

# afef

## 0.1 Customer churn dfddddddddddddddddddsdafdb dfd efef ef ef ef ef efe fef efew few few fwef e few fwefwefwe fef effwe f few f

Customer Relationship Management is defined as a process, in which the business manages its interactions with the customers using data integration from various sources and data analysis (Bardicchia 2020). Oliveira (2012) specifies 4 areas in which CRM approaches can be of use and what questions do they aim to answer:

* Customer identification (acquisition) - who can be a potential customer?
* Customer attraction - how can one make this person a customer?
* Customer development - how can one make a customer more profitable?
* Customer retention - how can one make the customer stay with the company?

The last one is the main focus of this study.

Chapter 1

Improving the loyalty of the customer base is profitable to the company. This has its source in multiple factors, the most important one being the cost of acquisition. Multiple studies have shown that retaining customers costs less than attracting new ones (Dick and Basu 1994; Gefen 2002; Buckinx and Poel 2005). Moreover, there is some evidence that loyal customers are less sensitive to the competitor’s actions regarding price changes (Achrol and Kotler 1999; Choi et al. 2006).

There are 2 basic approaches for the company to deal with customer churn. The first one is an “untargeted” approach. The company seeks to improve its product quality and relies on mass advertising to reduce the churn. The other way is a “targeted” approach - the company tries to address aim their marketing campaigns at the customers that are more likely to churn (Burez and Poel 2007). This approach can be divided further, by how the targeted customers are chosen. The company can target only those that have already decided to resign from a further relationship. For example, in contractual settings, this can mean canceling the subscription or breaching the contract. The other way to approach the churn problem is to try to predict, which customers are likely to churn soon. This has the advantage of having lower cost, as the customers that are about to leave are likely to have high demands from the last-minute deal proposed to them (Tamaddoni Jahromi et al. 2010).

As pointed out by Tamaddoni Jahromi et al. (2010) in their literature review, most of the studies concerning churn prediction were done in contractual settings. In other words, churn was defined as the client resigning from using the company’s services by canceling the subscription or breaching the contract. Such a way to specify the churn is different from the businesses in which the customer doesn’t have to inform the company about resigning.

One problem that arises in the non-contractual setting is the definition of churn. As there is no clear moment that the customer decides not to use the company’s services anymore, it has to be specified by the researcher based on the goals that one has to achieve from the churn analysis. Oliveira (2012) defined partial churners as the customers not making new purchases in the retail shop for the next 3 months. A different approach was used by Buckinx and Poel (2005). All the customers that had the frequency of purchases below average were treated as “churners” since these customers were shown to provide little value to the company.

### 0.1.1 Customer churn prediction

If the company can successfully predict, which customers are most likely to leave, it can target them with a retention-focused campaign. Contrary to targeting all of the customers with such a campaign, focusing on the customers that are most likely to leave leads to a reduction of the cost of the campaign.

Churn prediction fits well with the framework of classification, as the variable that one would like to predict is binary (churn-no churn). However, not only such binary prediction is valuable for later retention campaign efforts. As noted by Wai-Ho Au, Chan, and Xin Yao (2003), equally important is that the machine learning model can predict the likelihood of the customer leaving. After such prediction, the customers can be ranked from the most to the least likely to churn.

This has two benefits. First, the company can decide what percentage of the customers to target in the retention campaign and is not bound by how many customers the model will predict as potential churners. Second, the company can decide how strong the targeting should be based on the likelihood to leave. For example, based on cost-benefit analysis of various targeting approaches, one could decide that for the top 10% of the most “risky” customers the company should offer big discounts for the next purchase, while for the top 30% - only send an encouraging email.

The churn prediction task can be decomposed into 2 main important aspects that one has to tackle. First is the decision about a specific Machine Learning model that gives the best performance. The second is deciding on the model formula - in other words, deciding about which variables should be included in the model and what should be the form of the relationship.

### 0.1.2 Machine Learning models for churn prediction

In previous churn prediction studies multiple machine learning algorithms for prediction were tested out (for an overview see Verbeke et al. (2011)). The two most widely used techniques are Logistic Regression (LR) and Decision Trees. An important feature of both of them is that they are relatively simple, and because of that the way they make predictions can be assessed by a qualified expert (Paruelo and Tomasel 1997). However, these two methods often give sub-optimal results compared to more advanced and recent approaches like Neural Networks or Random Forests (Murthy 1998; Oliveira 2012). Moreover, this was shown not only in the case of churn prediction setting but also in more general benchmarks that used multiple datasets and comparison metrics (Caruana and Niculescu-Mizil 2006).

Recently, the XGBoost algorithm (Chen et al. 2015) has been gaining popularity in multiple prediction tasks. XGBoost’s main strengths are the ability to infer non-linear relationships from the data, and relative speed, which allows the researcher to try out multiple hyperparameters and decide on the best ones (Chen et al. 2015). Because of that, it is considered a go-to standard for machine learning challenges, and very often solutions based on it achieve the best results in various competitions and benchmarks (hcho3 2020). In the context of churn prediction, XGBoost was used by Gregory (2018). It achieved superior performance compared to other techniques, specifically Logistic Regression and Random Forests.

### 0.1.3 Variables used in previous churn prediction studies

Previous churn prediction studies used a variety of variables to include in the model formulation. Buckinx and Van den Poel (2005) divided them into 3 broad categories - behavioral, demographic, and perception.

The first category of variables tries to describe how the customer has interacted with the company before. Typical features belonging to this category are recency, frequency, and monetary value, which constitute the basis of the RFM customer segmentation framework. These features are used in multiple studies (Oliveira 2012; Bhattacharya 1998), and typically accompany more complex variables that are the main focus of particular studies. Besides these basic features, some studies focus on other areas of customer behavior. For instance, dummies indicating particular products that the customer has bought in previous purchases were shown to be a useful predictor for churn prediction (Buckinx and Van den Poel 2005; Athanassopoulos 2000).

Some of the variables studied previously are available only in some of the domains. For example, one possibility of analyzing customer behavior in the context of e-commerce shopping is analyzing the customer behavior while interacting with the company’s website. Koehn, Lessmann, and Schaal (2020) analyzed such click-stream sequential data using Recurrent Neural Networks. They found that such information can serve as a good predictor of customer churn. Similar studies were performed by Berger and Kompan (2019) and Yu et al. (2011), however, the clickstream data was aggregated and preprocessed manually, instead of using a sequential modeling approach.

The second category of features used in churn prediction constitutes of demographic variables about the customer, such as age, gender or address. Such variables were shown to be good predictors of customer churn in multiple studies (for an overview see Verbeke et al. (2011)). However, the availability of such predictors to use in modeling is very often limited for multiple reasons. In non-contractual settings, customers don’t have to always provide such data to the company. Moreover, usage of such personal data can be in some cases considered unethical, and lead to predictions biased against particular age or gender.

Another way to include demographics data in the churn prediction model was shown by Zhao et al. (2005). They successfully used the census data obtained from the statistical office for particular regions that the customer is residing in.

The last category of variables used for churn prediction specified by Buckinx and Van den Poel (2005) is customer perception about the company. According to Kracklauer, Passenheim, and Seifert (2001), customer satisfaction is the most important factor driving customer retention. However, although such features could have potentially high predictive power, they are usually hard to observe and quantify meaningfully. The most widely used approach is asking the customers for direct feedback using questionnaires or providing a way to post a review on the purchase. This kind of feedback can be obtained in different forms, one of them being textual reviews. A couple of previous studies were aimed at extracting meaningful features from such reviews using different text mining methods. De Caigny et al. (2020) have used text embedding approach, while Suryadi (2020) - simple tf-idf technique. In both studies, the results using such methods were superior compared to the models without including such information.

## 0.2 Explainable Artificial Intelligence efefe e fe fe f efe f efe fe fef wef fwe fwe fwe f wef wefwe few f wef we

### 0.2.1 Introduction

While deciding on the type of Machine Learning algorithm, one usually faces the explainability-performance trade-off (Nanayakkara et al. 2018). More flexible models, like bagging, boosting or neural networks, very often present superior performance to less flexible approaches. On the other hand, their predictions cannot be explained as easily as in the case of for example Decision Trees or Linear Regression.

Explainable Artificial Intelligence (XAI) is a set of tools aimed at explaining predictions of these highly flexible models. This area started gaining popularity among Machine Learning researchers to somehow transfer the advantages of simple models to the approaches that provide superior performance.

Doshi-Velez and Kim (2017) specifies some of the machine learning model’s traits that can accompany typical requirement of achieving the best accuracy:

* fairness - whether the algorithm is biased against a particular gender, age, race, etc.
* robustness - whether the algorithm can provide correct predictions when the parameters change
* trust - whether the final users of the algorithm trust the model’s predictions

Machine learning practitioners when deciding on the methodology to apply have to assess which of the requirements are important in a particular task. For example, in CRM settings the trust in the model’s predictions is way less important than in medical areas, but still can be of crucial for a wide adoption of modeling across the company. On the other hand, sometimes the explainability is important only for the person developing the model, to understand its limitations and be able to improve upon it.

The tools of XAI can help in addressing the aforementioned issues, without losing the usual performance gain from black-box models.

### 0.2.2 Review of XAI methods

XAI methods can be broadly divided into 2 categories - model-level and instance-level. The methods from the first category help in understanding the model’s behavior in general, for all istances. While the second can help understanding model’s reasoning about the particular observation.

For an overview of existing XAI methods, see Biecek and Burzykowski (2021).

* general prediction
* which variables contribute to the prediction(s) and in what way
* how one particular variable influence the prediction(s)
* how well is the model fitting?

### 0.2.3 XAI in marketing

Research on Explainable Artificial Intelligence in Marketing domain is not very developed. (Nie znalazłem ani jednego papera !! ??)[[1]](#footnote-1)

To the best of the author’s knowledge, the only study touching the subject of XAI in the context of marketing is by Rai (2020). In their commentary, they specify potential areas for future research in this field:

* understanding, what are acceptable requirements regarding explainability compared to accuracy in different marketing tasks;
* making AI trustworthy - to understand how the eagerness to use AI system’s predictions grows in the company when various explainability tools are made available to the end-users;
* How model explanations should be presented to various groups of system’s users. For example, a Machine Learning expert is interested in very detailed and complex explanations, while the company’s customer may simply want a 1 sentence summary of what was taken into account while making predictions;

Achrol, Ravi S, and Philip Kotler. 1999. “Marketing in the Network Economy.” *Journal of Marketing* 63 (4\_suppl1): 146–63.

Athanassopoulos, Antreas D. 2000. “Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior.” *Journal of Business Research* 47 (3): 191–207.

———. 2000. “Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior.” *Journal of Business Research* 47 (3): 191–207.

Bardicchia, Marco. 2020. *Digital CRM-Strategies and Emerging Trends: Building Customer Relationship in the Digital Era*.

Berger, P., and M. Kompan. 2019. “User Modeling for Churn Prediction in e-Commerce.” *IEEE Intelligent Systems* 34 (2): 44–52. <https://doi.org/10.1109/MIS.2019.2895788>.

Bhattacharya, CB. 1998. “When Customers Are Members: Customer Retention in Paid Membership Contexts.” *Journal of the Academy of Marketing Science* 26 (1): 31–44.

Biecek, Przemyslaw, and Tomasz Burzykowski. 2021. *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*. CRC Press.

Buckinx, Wouter, and Dirk Van den Poel. 2005. “Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting.” *European Journal of Operational Research* 164 (1): 252–68.

———. 2005. “Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting.” *European Journal of Operational Research* 164 (1): 252–68.

Buckinx, Wouter, and Dirk Van den Poel. 2005. “Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting.” *European Journal of Operational Research* 164 (1): 252–68. https://doi.org/<https://doi.org/10.1016/j.ejor.2003.12.010>.

Burez, Jonathan, and Dirk Van den Poel. 2007. “CRM at a Pay-TV Company: Using Analytical Models to Reduce Customer Attrition by Targeted Marketing for Subscription Services.” *Expert Systems with Applications* 32 (2): 277–88.

Caruana, Rich, and Alexandru Niculescu-Mizil. 2006. “An Empirical Comparison of Supervised Learning Algorithms.” In *Proceedings of the 23rd International Conference on Machine Learning*, 161–68.

Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, and others. 2015. “Xgboost: Extreme Gradient Boosting.” *R Package Version 0.4-2* 1 (4): 1–4.

Choi, Duke Hyun, Chul Min Kim, Sang-Il Kim, and Soung Hie Kim. 2006. “Customer Loyalty and Disloyalty in Internet Retail Stores: Its Antecedents and Its Effect on Customer Price Sensitivity.” *International Journal of Management* 23 (4): 925.

De Caigny, Arno, Kristof Coussement, Koen W. De Bock, and Stefan Lessmann. 2020. “Incorporating Textual Information in Customer Churn Prediction Models Based on a Convolutional Neural Network.” *International Journal of Forecasting* 36 (4): 1563–78. https://doi.org/<https://doi.org/10.1016/j.ijforecast.2019.03.029>.

Dick, Alan S, and Kunal Basu. 1994. “Customer Loyalty: Toward an Integrated Conceptual Framework.” *Journal of the Academy of Marketing Science* 22 (2): 99–113.

Doshi-Velez, Finale, and Been Kim. 2017. “Towards a Rigorous Science of Interpretable Machine Learning.” *arXiv Preprint arXiv:1702.08608*.

Gefen, David. 2002. “Customer Loyalty in e-Commerce.” *Journal of the Association for Information Systems* 3 (1): 2.

Gregory, Bryan. 2018. “Predicting Customer Churn: Extreme Gradient Boosting with Temporal Data.” *arXiv Preprint arXiv:1802.03396*.

———. 2018. “Predicting Customer Churn: Extreme Gradient Boosting with Temporal Data.” *arXiv Preprint arXiv:1802.03396*.

hcho3. 2020. “Awesome XGBoost.” <https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions>.

Koehn, Dennis, Stefan Lessmann, and Markus Schaal. 2020. “Predicting Online Shopping Behaviour from Clickstream Data Using Deep Learning.” *Expert Systems with Applications* 150: 113342.

Kracklauer, Alexander, Olaf Passenheim, and Dirk Seifert. 2001. “Mutual Customer Approach: How Industry and Trade Are Executing Collaborative Customer Relationship Management.” *International Journal of Retail & Distribution Management*.

Murthy, Sreerama K. 1998. “Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey.” *Data Mining and Knowledge Discovery* 2 (4): 345–89.

Nanayakkara, Shane, Sam Fogarty, Michael Tremeer, Kelvin Ross, Brent Richards, Christoph Bergmeir, Sheng Xu, et al. 2018. “Characterising Risk of in-Hospital Mortality Following Cardiac Arrest Using Machine Learning: A Retrospective International Registry Study.” *PLoS Medicine* 15 (11): e1002709.

Oliveira, Vera Lúcia Miguéis. 2012. “Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques.” PhD thesis, Universidade do Porto (Portugal).

Paruelo, JoséM, and Fernando Tomasel. 1997. “Prediction of Functional Characteristics of Ecosystems: A Comparison of Artificial Neural Networks and Regression Models.” *Ecological Modelling* 98 (2-3): 173–86.

Rai, Arun. 2020. “Explainable AI: From Black Box to Glass Box.” *Journal of the Academy of Marketing Science* 48 (1): 137–41.

Suryadi, D. 2020. “Predicting Repurchase Intention Using Textual Features of Online Customer Reviews.” In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, 1–6. <https://doi.org/10.1109/ICDABI51230.2020.9325646>.

Tamaddoni Jahromi, Ali, Mohammad Mehdi Sepehri, Babak Teimourpour, and Sarvenaz Choobdar. 2010. “Modeling Customer Churn in a Non-Contractual Setting: The Case of Telecommunications Service Providers.” *Journal of Strategic Marketing* 18 (7): 587–98.

Verbeke, Wouter, David Martens, Christophe Mues, and Bart Baesens. 2011. “Building Comprehensible Customer Churn Prediction Models with Advanced Rule Induction Techniques.” *Expert Systems with Applications* 38 (3): 2354–64.

Wai-Ho Au, K. C. C. Chan, and Xin Yao. 2003. “A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction.” *IEEE Transactions on Evolutionary Computation* 7 (6): 532–45. <https://doi.org/10.1109/TEVC.2003.819264>.

Yu, Xiaobing, Shunsheng Guo, Jun Guo, and Xiaorong Huang. 2011. “An Extended Support Vector Machine Forecasting Framework for Customer Churn in e-Commerce.” *Expert Systems with Applications* 38 (3): 1425–30.

Zhao, Yu, Bing Li, Xiu Li, Wenhuang Liu, and Shouju Ren. 2005. “Customer Churn Prediction Using Improved One-Class Support Vector Machine.” In *International Conference on Advanced Data Mining and Applications*, 300–306. Springer.

1. Sgdgwegwegwegewweg ijiji [↑](#footnote-ref-1)