MASTER

Better customer retention with a multi-class customer churn prediction model that incorporates past customer churn reasons to drive proactively marketing actions

A unique customer churn prediction model that incorporates in-house information about past churn reasons in the dataset

van Beers, Joelle J.M.

Award date:
2022

Master thesis

Better customer retention with a multi-class customer churn prediction model that incorporates past customer churn reasons to drive proactively marketing actions.

A unique customer churn prediction model that incorporates in-house information about past churn reasons in the dataset.

Eindhoven University of technology

A master thesis for Innovation Management

Author: Joelle van Beers

First supervisor: Prof. Dr. Ed J. Nijssen

Second supervisor: Dr. Ir. Remco Dijkman

Third supervisor: Dr. Jeroen J.L. Schepers

Company supervisor: Sepp Haans

Date: 5-12-2022

Preface

Dear reader,

The thesis you will read is the result of my graduation project as part of the master's program Innovation Management at the Eindhoven University of Technology. This study was conducted at the company Freshheads, located in Tilburg. As my student life ends with this as my final project, I can reflect on a pleasant and educational experience at the university. Obtaining my master's degree was made possible by several people to whom I would like to express my appreciation.

First, I want to show my gratitude to my mentor and first supervisor, Ed Nijssen. Thank you for your guidance and help. Your quick responses and thoughtful feedback allowed me to move forward quickly and adequately. In addition, your metaphors provided a new perspective on specific topics. Besides my gratitude for my first supervisor, I would like to thank my second supervisor; Remco Dijkman as well. Thank you for providing feedback; your data science perspective on my study was really valuable and gave me new insights.

Furthermore, I want to thank Freshheads for trusting me and offering me a graduation project. In particular, I want to thank Sepp Haans since he has been my supervisor all this time. Even though the project took longer, you always supported and encouraged me. Also, my gratitude goes to Monique and Joost, who are working at Fressheads. Without your help, my thesis would have been less succesful. I was fortunate to learn a lot from both of you.

Lastly, I would like to thank those who were not practically involved in this project but who supported me in other ways. Especially my family, friends, and roommates, which I could rely on all the time. Thank you!

Joelle

December 2022

Abstract

Organizations have customer retention as a high priority since customer acquisition is much more expensive than customer retention. Many proactive retention campaigns are deployed in which customers with a high probability of churn, predicted by a prediction model, are targeted with incentives in the hope that these customers will be influenced to retain. This research expands the existing literature on customer churn prediction. Whereas much literature focuses primarily on improving the performance of prediction models, this study focuses on incorporating in-house feedback data on the reason for churn. This study treats churn reason as the dependent variable in a multi-class customer churn prediction model. The output of this model predicts not only who will churn but also why someone will churn. This creates opportunities for more targeted proactive marketing actions by knowing the reason for churn. In addition, this makes it easier to define segments and distinguish who the persuadable customers are. The results show that adding feedback data about past customer churn reasons is a valuable addition to customer churn prediction models and that XGBoost can be used as a classifier for this. This study can serve as a step-by-step approach to using in-house feedback data to define a multi-class customer churn reason prediction model.

Keywords: Customer churn prediction, customer retention management, multi-class prediction, machine learning, customer churn reasons, proactive retention campaign.

Executive summary

Introduction – One of the main challenges for organizations is to identify profitable customers, satisfy the customers and retain the current customer base. A common approach to managing customer retention is running proactive retention campaigns. Proactive retention means that the company takes action to solve in advance the problem that generates churn (Tamaddoni et al., 2017). The most common strategy to proactively manage customer churn is to identify customers who are most likely to churn and target these customers in proactive retention campaigns (Tamaddoni et al., 2017). The identification of customers with a high probability of churn is known as customer churn prediction. A customer churn prediction model is a tool that organizations use to prevent customers from reaching the button to cancel their subscriptions.

From a marketing perspective, the literature argues that to develop successful proactive retention campaigns, it is also necessary to build on why customers are no longer satisfied with the product or service. After all, when the cause of dissatisfaction is understood, an organization can differentiate appropriate retention efforts for each customer (Kumar et al., 2015). However, the current literature on customer churn prediction focuses mainly on improving customer churn prediction models. While there is rarely explored and demonstrated potential advantages of using customer churn prediction as a marketing strategy (Ahn et al., 2006; Bhattacharyya & Dash, 2021; Tamaddoni et al., 2017).

Therefore, this study uses prior feedback data about customer churn reasons as the dependent variable in a multi-class customer churn prediction model. It investigates whether it is possible to predict not only who is going to churn but also why someone is going to churn. The model's output presents new opportunities for marketing managers to target potential churners in a more segmented and personalized way for a proactive retention campaign. The main research question of this study is:

"How to design a customer churn prediction model that is competent in predicting customer churn reasons using AI and prior feedback data about customer churn reasons?"

Theoretical background – To answer the main research question, literature about customer retention management, customer churn reasons, and customer churn prediction is inspected. During the literature exploration, multiple frameworks for customer retention management were identified, which all relied on the same approach. The framework is typically two-tiered; one tier is identifying customers with a high probability of churn, usually done by customer churn prediction, and the second tier is persuading the potential churners to stay (Jahromi et al.,2014).

The first tier is done by customer churn prediction. This is the process of calculating the churn probability for each customer in the database using a predictive model based on past information and prior behavior (Jahromi et al., 2014). It aims to identify early churn signals and recognize customers with an increased likelihood to churn (Vafeiadis et al., 2015). Nowadays, this is usually done with machine learning models. Boosting has been successfully applied to customer churn prediction cases in different study fields. Boosting models promise good customer churn prediction results. The most popular boosting algorithm is AdaBoost and XGBoost, with the decision tree as the weak classifier.

It is of great importance that the data collected for the customer churn prediction model is relevant to the learning goal, such that the data explains the label or the response to the observation (Nwanganga & Chapple, 2020). Besides basic variables such as customer demographics and customer behavior, RFM (Recency, Frequency, and Monetary) variables have been proven to play an undeniable role in predicting customer churn (Buckinx & van den Poel, 2005; Jahromi et al., 2014).

A confusion matrix is used to calculate multiple evaluation metrics for measuring the performance of the models. The most popular evaluation metric is accuracy. Precision, recall, and F1 score are also good measurements. The Area Under the Curve (AUC) is a more graphically visible evaluation measurement. The AUC, also known as the area under the Receiver Operating Characteristic (ROC) curve, is used for ranking the performance of data mining models.

Following the literature, an overarching conclusion can be drawn that, in general, customers churn for price-related reasons or because the product and/or (customer) service does not satisfy their expectations.

Methodology - The prior feedback data about customer churn prediction could be divided into four categories of churn reasons; price-related issues, service-related issues, covid, and no longer teaching. This data resulted in a categorical label used as the dependent variable in a multi-class prediction model. The different categories of churn reasons were used for segmentation to distinguish the persuadable customer from those who were not, allowing a better targeting strategy for the proactive retention campaign. The step-by-step process charter in Figure 1 summarizes the complete methodology of this study. Nevertheless, the study starts with a binary classification to be able to conclude that a customer churn prediction with this dataset is possible.

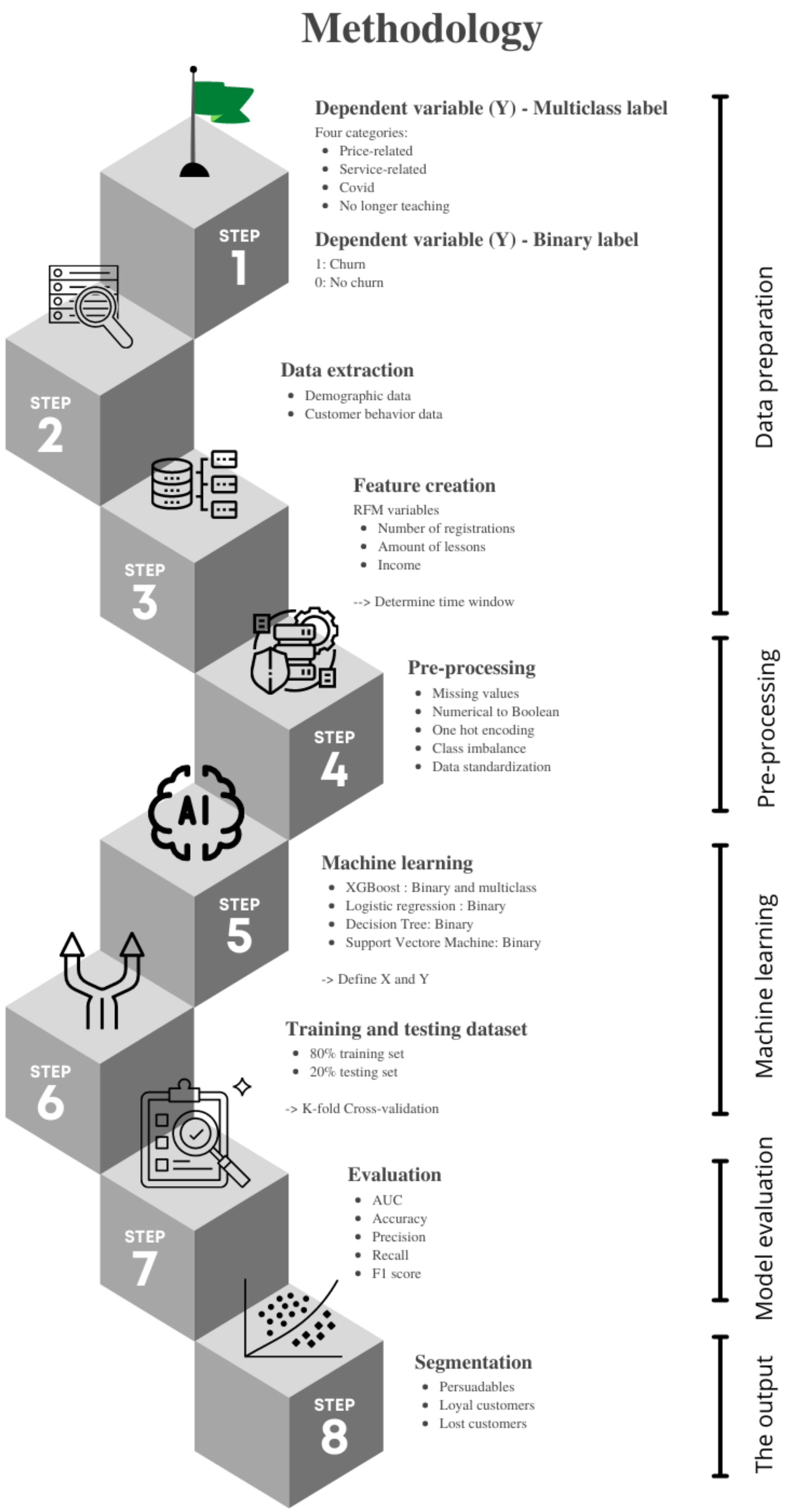

Figure 1: Summary methodology

Findings and managerial implications - This study provides insights into 1.) The development of a customer churn prediction model that can predict customer churn reasons, and 2.) How can the prediction of customer churn reasons be used for proactive marketing actions, i.e., proactive retention campaigns?

The results showed that XGBoost is the best-performing model for customer churn prediction. For the development of the customer churn prediction models, important data variables are; demographic data, customer behavioral data, and RFM variables. RFM variables play a significant role in this study as they are the most correlated with the dependent variable and the most important features within the model.

The binary classification model performs well, with an AUC score of 90% and an accuracy of 82%. This indicates that the data used in this study is credible for customer churn prediction. The performance of the multi-class prediction for customer churn reason is lousy. This most likely has to do with the small amount of data about churn reasons. This model cannot be considered as truth yet and needs further development. Although, it shows that it is possible to predict churn reasons with the study's approach. Organizations should keep updating the model with new data, as it is expected to continue improving.

The indicated churn reasons allow a new segmentation. Within this segmentation, the persuadable customer is distinguished from the non-persuadable. Where persuadable customers can be targeted in the proactive retention campaign, this segmentation has yet to be tested, which makes it difficult to conclude whether this segmentation successfully reduces the churn rate. Therefore, the organization is advised to start experiments to test the segmentation and the developed model.

List of table and figures

Table 1: Confusion matrix 20
Table 2: Segmentation matrix uplift modeling........ 23
Table 3: Performance of machine learning methods for binary classification 36
Table 4: Confusion matrix XGBoost classifier 37
Table 5: Performance of multiclass XGBoost classifier 40
Table 6: Confusion matrix XGBoost multiclass classifier 40
Table 7: Output deliverable........ 42
Table 8: Raw dataset 52
Table 9: Dataset with created features 54
Table 10: Correlations of dependent (Y) variable "churn" and "churn reason" 58
Table 11: Confusion matrix DT 59
Table 12: Confusion matrix SVM 59
Table 13: Confustion matrix LR 59
Table 14: Confusion matrix KNN 59
Table 15: Confusion matrix multi-class (ignore NaN)........ 60

Figure 1: Summary methodology........ 6
Figure 2: Visual of practical relevance of this study 14
Figure 3: Conceptual framework for designing retention campaigns. 15
Figure 5: Amount of lessons per period 27
Figure 4: Amount of registrations per period 27
Figure 6: Amount of income per period 27
Figure 7: Chosen time window 28
Figure 8: Segmentation matrix 32
Figure 9: Step by step chart methodology........ 34
Figure 10: ROC Curve XGBoost binary classifier........ 37
Figure 11: Output table customer churn prediction........ 38
Figure 12: Output table multiclass customer churn reason prediction 41
Figure 13: Segmentation matrix from test set 43

Table of contents

Preface 2

Abstract 3

Executive summary 4

List of table and figures 8

1. Introduction 11

1.1 Research background 11

1.2 problem statement 12

1.3 Scientific relevance 12

1.4 Practical relevance 13

2. Theoretical background 14

2.1 Customer retention management 14

2.2 Customer churn reasons 16

2.3 Customer churn prediction 16

2.3.1 Data selection 19

2.3.2 Evaluation of customer churn prediction models 20

2.4 Profit-based loss function and Uplift modeling 21

3. Methodology 24

3.1 data preparation 24

3.1.1 churn reasons as dependent variable 24

3.1.2 Data extraction 25

3.1.3 Feature creation 26

3.2 Pre-processing 28

3.2.1 Missing values 28

3.2.2 Numerical to Boolean 28

3.2.3 One hot encoding 29

3.2.4 Class imbalance 29

3.2.5 Data standardization 29

3.3 Machine learning 29

3.3.1 Modeling techniques 29

3.3.2 Definition X and Y 30

3.3.3 Training dataset and testing dataset 30

3.3.4 Hyperparameter optimization 30

3.3.5 Cross-validation 31

3.4 Model evaluation 31

3.5 The output 31

3.6 The workflow 33
4. Results 35
4.1 Results binary customer churn prediction model 35
4.2 Results multiclass customer churn prediction model 38
5. Conclusion 43
5.1 Key findings 43
5.2 Implementation advice 45
6. Discussion 46
6.1 limitations 46
6.2 Future research 47
7. References 48
Appendices 51
Appendix A 51
Appendix B 53
Appendix C 55
Appendix D 57
Appendix E 59
Appendix F 60

1. Introduction

1.1 Research background

Customer Relationship Management (CRM) is a widely recognized business approach (Vafeiadis et al., 2015). The challenge is to identify profitable customers, satisfy the customers and retain the current customer base. Customer retention is defined as the customer continuing to transact with the firm (Ascarza, Neslin, et al., 2018). This is important for companies since the cost of customer acquisition can be five to six times higher than the cost of customer retention (Vafeiadis et al., 2015). This stresses the importance for companies to pay attention to the opportunities that arise to retain the current customer base.

Not surprisingly, customer retention management is one of the main objectives within the CRM domain (Mehrabioun & Mahdizadeh, 2021). A common approach to managing customer retention is running reactive and proactive retention campaigns. Reactive retention means the company waits for the customer to churn and then tries to ‘’win back’’ the customer. Proactive retention means that the company takes action to solve in advance the issue that generates churn (Tamaddoni et al., 2017). This study focuses on the last strategy, namely proactively retaining customers. Where churn is defined as the moment a customer decides to stop being a customer.

The most common strategy to proactively manage customer churn is to identify customers who are most likely to churn and target these customers in proactive retention campaigns (Tamaddoni et al., 2017). The identification of customers with a high probability of churn is known as customer churn prediction. A customer churn prediction model is a tool that organizations use to prevent customers from reaching the button to cancel their subscriptions. In the last decade, much literature has been published on developing such churn prediction models and their performance. Previously, these predictions were mainly made with statistical analyses. However, with the rapid influx of data and the development of artificial intelligence, data-mining methods (Machine Learning techniques) are now commonly applied to predict customer churn (Gür Ali & Aritürk, 2014). These machine learning models are trained to identify signals in customer behavior data that indicate customers are likely to churn, even long before they know it themselves.

From a marketing perspective, the literature argues that to develop successful proactive retention campaigns, it is also necessary to build on why customers are no longer satisfied with the product or service. After all, when the cause of dissatisfaction is understood, an organization can differentiate appropriate retention incentives for each customer (Kumar et al., 2015). However, the current literature on customer churn prediction focuses mainly on the data mining part, which means the attention is on improving the performance of the machine learning models. While there is rarely explored and demonstrated potential advantages of using customer churn prediction as a marketing strategy (Ahn et al., 2006; Bhattacharyya & Dash, 2021; Tamaddoni et al., 2017). Nevertheless, many organizations request feedback data when customers leave, and this data can provide insights into why a customer was no longer satisfied and decided to churn. However, this valuable data is currently not utilized in customer churn prediction models when developing a proactive retention campaign that targets these potential churners.

Therefore, this study uses prior feedback data about customer churn reasons as the dependent variable in a customer churn prediction model. It investigates whether it is possible to predict not only who is going to churn but also why someone is going to churn. As the output of the model presents new opportunities for marketing managers to target potential churners in a more segmented and personalized way for a proactive retention campaign.

1.2 problem statement

Momoyoga, a Software-as-a-Service (SaaS) company, experiences high churn rates, jeopardizing its existence. Solving this requires a proper proactive retention management strategy. This is achievable when it is clarified which customers are most likely to churn and the reason why they are going to churn. This research focused on the development of a customer churn prediction model that can be used as a segmentation tool to retain customers proactively.

To be state of the art, this research will use artificial intelligence (AI) and, thus, a machine learning algorithm.

The main research question of this study is:

How to design a customer churn prediction model that is competent in predicting customer churn reasons using AI and prior feedback data about customer churn reasons?

Related to the main question, the following sub-questions are formulated.

1. Which data is useful for customer churn (reason) prediction?
2. What are possible reasons for churn?
3. Which machine learning models are best for predicting customer churn (reason)?
4. Which evaluation metrics can be used to evaluate the prediction model?
5. How can a prediction model for customer churn reason be used for marketing actions?

1.3 Scientific relevance

The rationale behind customer churn prediction is to send targeted marketing actions to potential churners to retain the current customer base. The fact that marketing actions need to be sent is because these customers are not satisfied anymore. However, what is remarkable is that the customer churn prediction literature does not focus on the fact that customers are not satisfied anymore. Instead, the literature mainly focuses on improving the various machine learning models to generate the best possible predictive model instead of thinking about why a customer might leave (Ahn et al., 2006;

Bhattacharyya & Dash, 2021; Tamaddoni et al., 2017). Nowadays, machine learning models have advanced to the point where they are relatively easy to use and have generally been proven to work. Therefore, this study is not primarily focused on improving a customer churn model, but it explores the possible opportunities to incorporate data in the model that could say something about why a customer is going to churn. This means that this study will contribute to the customer churn prediction literature as a customer churn prediction model with a unique addition.

Also, little research has been performed on customer churn prediction in a SaaS industry, especially in Business-to-Business (B2B). This could be because SaaS is relatively new, and data is challenging to obtain. Both studies by Jahromi et al. (2014) and (Frank & Pittges, n.d.) concluded that B2B context, generally, has received less attention in the literature on customer churn prediction models. Besides this, most customer churn prediction literature is based on telecommunication data, as this data is easy to obtain. This study will use data from a B2B SaaS company to develop the customer churn prediction model. Whichever might suggest new findings in the customer churn prediction literature.

1.4 Practical relevance

It is valuable for organizations to know which customers have a high risk of churn. This is valuable since it saves much money, given that customer acquisition is far more expensive than customer retention. In addition, once companies know which customers have a high probability of churn, they can execute targeted marketing campaigns to retain the customer.

In customer churn prediction, customer data is analyzed and put through a machine learning model that then provides a list as output that includes all customers at high risk of churn. This list is passed to the retention team, and the high-risk churn customers are targeted in a proactive marketing campaign. However, there is no information about why this customer is no longer satisfied with the product/service. This causes much experimentation with variations in incentives which can cost a lot of time and money. Whereas if it were known why someone is no longer satisfied, it is much easier to decide if someone is persuadable to retain. Nevertheless, it also allows differentiating in incentives which might be even more successful in retaining the customer.

Moreover, many organizations have collected feedback from previously churned customers that say why this customer chose to churn. It is a missed opportunity that this valuable data is not included in machine learning models, given the development and capabilities of data mining today.

This study practically contributes by providing an output list that includes a customer segmentation that enables more targeted and personalized marketing actions. The output of the developed model in this study provides a list of potential churners, but on top of that, also the highest probability of churn reason for a specific customer, which allows more targeted and personalized incentives to be sent to high-probability churners.

Figure 2 shows a minimalistic representation of the practical relevance of this study and the deliverable.

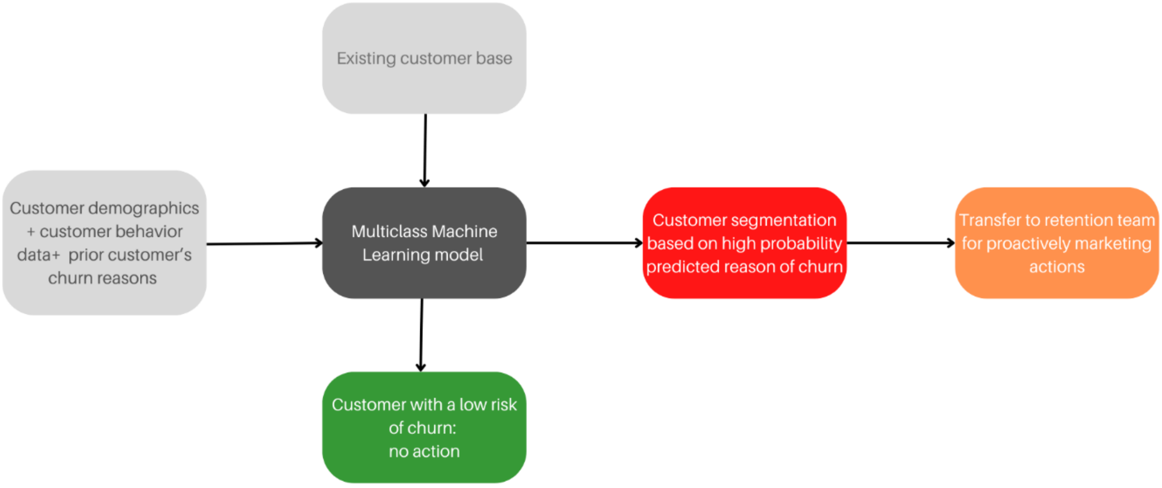

Figure 2: Visual of practical relevance of this study

One point that needs to be outlined is that even though customer churn prediction has been widely researched, the results may still differ from company to company as they deal with different motives of churn and data. This highlights that every research in customer churn prediction is unique, especially in deciding which stored data to include in the models. Therefore, to get optimal exploitation, each organization must carefully analyze the data.

2. Theoretical background

The theoretical background summarizes relevant articles found in the customer churn prediction literature. To discuss the relevant articles, this chapter is divided into four main topics: Customer retention management, customer churn reasons, customer churn prediction, and profit-based loss functions and uplift modeling. In addition, this chapter helps to answer the sub-research questions.

2.1 Customer retention management

Companies should make their utmost effort to satisfy their existing customers to improve customer retention. This has all to do with the fact that acquiring a new customer is anywhere from 5 to 25 times more expensive than retaining an existing customer (Shabankareh et al., 2022; Vafeiadis et al., 2015). As a result, customer retention management within organizations is essential, and it can be highly profitable when done correctly. Customer retention management consists of developing techniques that enable firms to keep their profitable customers and aims to increase customer loyalty (Mehrabioun & Mahdizadeh, 2021). Nevertheless, it helps minimize the potential damage that churn can cause.

According to Burez & van den Poel (2007) there are two approaches to managing customer churn: untargeted and targeted. An untargeted approach relies on mass advertising to the whole customer base. The targeted approach relies on identifying customers who are likely to churn, providing them with an incentive, or customizing a service plan to stay, also called a treatment. Subsequently, the targeted approach can also be subdivided into two different types. The first type is a reactive approach; here, a company waits until a customer churn and then offers the customer an incentive. The second type is a proactive approach; here, the company is proactively targeting customers based on their likelihood to churn in the future. As mentioned in the introduction chapter, this study focuses on the second type: proactively retaining customers.

Frameworks for customer retention management were developed to identify the potential churners that can be targeted by proactive marketing campaigns aiming to retain them (Blattberg et al., 2008; Tamaddoni et al., 2017). Figure 3: Conceptual framework for designing retention campaignshows such a framework (Ascarza et al., 2016).

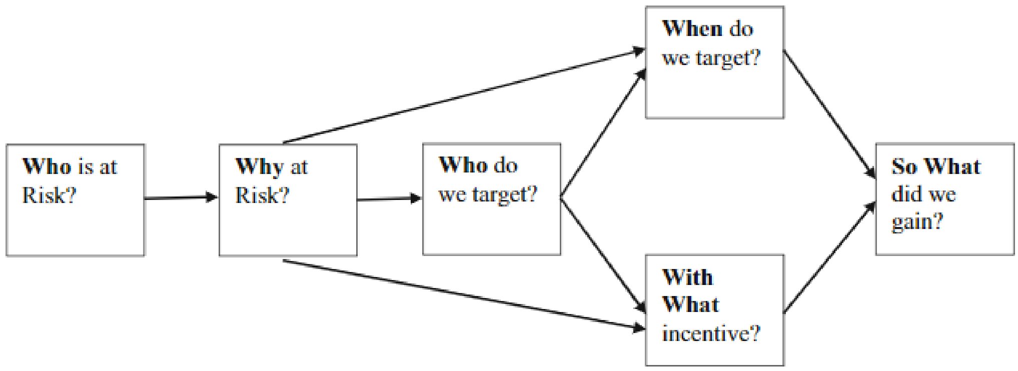

Figure 3: Conceptual framework for designing retention campaigns.

The different frameworks found in the literature are fundamentally convergent: An organization must first identify customers at risk of not being retained. Second, diagnose why each customer is at risk. Third, decide which customers to target, and then decide when to target these customers and with what treatment. The final step is to implement the proactive campaign and evaluate it. The framework is typically two-tiered (Jahromi et al., 2014). One tier is the identification of customers with a high probability of churn usually done by customer churn prediction, and the second tier is the task of persuading the potential churners to stay. The latter is usually done by proactively sending incentives to the customers, i.e., a proactive retention campaign.

The literature argues that to minimize customer churn and build customer loyalty successfully, the customer retention team needs to focus on understanding the causes of customer churn (Ascarza, Netzer, et al., 2018; Bhattacharyya & Dash, 2021; H. S. Kim & Yoon, 2004). In fact, the churn reason reflects what the customer values (Kumar et al., 2015). Churn can be caused by multiple problems such as technological developments, economic issues, qualitative factors, service type and coverage, and even a bad experience of encountering the telephone center employees of an organization (Shabankareh et al., 2022). Therefore, successful customer churn management is built on why customers churn to develop and target appropriate retention efforts to customers.

2.2 Customer churn reasons

According to Kumar et al. (2015) and Roschk & Gelbrich (2014) there are three main reasons for customer churn. The first reason is a monetary failure; this represents a financial loss by the customer, for example, billing issues or competitors offering a better price. The second is service failure; this is an issue in service delivery. The last churn reason is lack of attention; this can happen when organizations handle customers unfriendly, impatient, or uncaringly (Gustafsson et al., 2005; Roschk & Gelbrich, 2014).

Besides this, Burez & van den Poel (2007) also differentiated in churn reasons, namely: commercial churners and financial churners. Commercial churners do not renew their contract due to different reasons, including poor services and product problems, and financial churners are those churners who need to stop the subscription because they find it too expensive. Here one sees overlap with the reasons given by Kumar et al. (2015) and Roschk & Gelbrich (2014). From this literary background, an overarching conclusion can be drawn that, in general, customers churn for price-related reasons or because the product and/or (customer) service does not satisfy their expectations.

2.3 Customer churn prediction

The first tier in the customer retention management framework is identifying customers at high risk of churn. As mentioned, this is usually done by customer churn prediction, which consists of a data mining process and, to be more specific, using machine learning models.

Customer churn prediction is the process of calculating the churn probability for each customer in the database using a predictive model based on past information and prior behavior (Jahromi et al., 2014). It aims to identify early churn signals and recognize customers with an increased likelihood to churn (Vafeiadis et al., 2015). Customer churn prediction enables organizations to focus their efforts on customers who are genuinely at risk to churn. In addition, it saves money that would be wasted in providing incentives to customers who do not need them, i.e., the untargeted approach (Neslin et al., 2006).

The SAS institute defines data mining as ‘’the process of selecting, exploring, and modeling large amounts of data to uncover previously unknown data patterns for business advantage’’ (SAS Institute, 2017). Machine learning or deep learning is the data mining technique for customer churn prediction. Customer churn prediction models can be constructed using different machine learning techniques, each with its strengths and limitations. The study by Vafeiadis et al. (2015) compared several popular machine learning algorithms proposed to tackle customer churn prediction. The five most popular machine learning methods will be outlined:

1. Artificial Neural Network (ANN)

   Using ANN is a popular approach to address complex problems, such as the customer churn prediction problem. A popular supervised model is a Multi-Layer Perceptron trained with a variation of the Back-Propagation algorithm (BPN). Research has shown that ANN performs better than Decision Trees (DT), and it outperforms Logistic Regression for customer churn prediction.
2. Support Vector Machines (SVM)

   SVM are Supervised learning models with associated learning algorithms that analyze data and recognize patterns; it is used for classification and regression analysis. It is a technique based on structural risk minimization. In customer churn prediction, SVM outperforms DT and sometimes ANN, but it depends on the collected data.
3. Decision Trees (DT)

   DTs are sets of decisions that generate classification rules for a specific dataset. The tree structures have leaves representing class labels and branches representing conjunctions of features that lead to those class labels. In customer churn prediction, the accuracy of a DT can be high, but again, it highly depends on the collected data. DT does not perform outstandingly in capturing complex and non-linear relationships between the attributes.
4. Naïve Bayes (NB)

   This simple probabilistic classifier is based on applying Bayes' theorem with strong independence assumptions. An NB classifier assumes that the presence or absence of a particular feature of a class is unrelated to the presence or absence of another feature. It achieves great results on the customers' churn prediction problem for the wireless telecommunication industry.
5. Regression analysis-logistic regression analysis (LR)

   Regression analysis is a statistical process for estimating the relationship among variables. When speaking of customer churn, regression analysis is not widely used. However, LR is used since it is a type of probabilistic statistical classification model—also used to produce a binary prediction of, for example, the customer churn variable, which depends on one or more predictor variables. In the churn prediction problem, LR is usually used with great performance, but a proper data cleanup on the initial data is necessary

Results from a customer churn modeling tournament organized by the Teradata Center at Duke University showed that LR and DT were the most common estimation techniques (Neslin et al., 2006).

The study by Neslin et al. (2006) concluded that the method that is chosen for customer churn prediction is of great importance. Proactive retention campaigns can have a significant potential advantage compared to the untargeted approach. However, these campaigns can be wasteful if churn predictions are not accurate. If churn predictions are not accurate, the organization spends, for example, money on customers who have stayed anyway. Therefore, it is essential to predict customer churn as accurately as possible (Burez & van den Poel, 2009). The higher the predictive accuracy, the higher the additional profits can be. It is essential to use multiple methods and techniques in combination to build the model in order to achieve the highest accuracy (Neslin et al., 2006).

Boosting is a technique that is used to improve classification performance by combining decisions from many classification models. Boosting has been successfully applied to customer churn prediction cases in different study fields. The most popular boosting algorithm is AdaBoost and XGBoost, with DT as the weak classifier. Boosting has shown performance improvement within several machine learning algorithms and therefore supports the advantage of the application of boosting techniques (Robert C Blattberg et al., 2008; Vafeiadis et al., 2015). Besides this, Le et al. (2022) argue that XGBoost can learn from its mistakes, fine-tuning extensive hyperparameters, scaling imbalanced data, and processing null values. Furthermore, XGBoost has been proven to boost weak learning in both classification and regression problems. In the training process, a new tree is added at every iteration. Where every iteration fixes prior tree mistakes and residuals, after all these iterations, the algorithm combines the previous trees to generate the final prediction (Le et al., 2022). Overall, XGBoost promises good customer churn prediction results.

Choosing a machine learning model must be done adequately to obtain good performance. There are a few common characteristics to pay attention to when choosing a suitable model. Machine learning techniques fit into two categories: Supervised learning algorithms and unsupervised learning algorithms. Supervised learning algorithms learn patterns based on labeled examples of past data. Unsupervised learning algorithms seek to uncover patterns without the assistance of labeled data.

Supervised learning is the most commonly used category of machine learning algorithms. The purpose of supervised learning is to use an existing dataset to generate a model that makes predictions about future, unlabelled data. The input of supervised learning is a training dataset with a specific label. For customer churn prediction, this is primarily a binary label that explains if a customer has churned or not; 0 (no churn) and 1 (churn). The algorithm uses the training data to develop a model as its output (Nwanganga & Chapple, 2020). Unsupervised learning techniques work differently, it develops models based on unlabelled training datasets. Instead of providing a method for assigning labels to input based on historical data, unsupervised techniques allow discovering hidden patterns in the dataset (Nwanganga & Chapple, 2020). Unsupervised learning is usually used with a dataset that, for example, includes images and textual data.

Besides the categorization of the type of data, it is also possible to categorize the algorithms based on what they learn. Three major types of knowledge can be learned from data. The first one is classification techniques. Classification techniques use supervised machine learning to predict a categorical response. The model's output is a non-numerical label / categorical variable in this case.

The second technique is the regression technique. These are used in supervised learning to predict continuous responses. The output of the model is a numeric value. The last technique is the similarity learning technique. It is used in an unsupervised dataset and helps to identify common patterns (Nwanganga & Chapple, 2020).

2.3.1 Data selection

The data type and amount of the data are often dependent on the machine learning problem and the selected algorithm. However, it is of great importance that the data collected is relevant to the learning goal, such that the data explains the label or the response of the observation (Nwanganga & Chapple, 2020).

However, besides basic variables such as customer id and demographics, there is a popular data collection approach that needs to be outlined. This is the RFM – Recency, Frequency, and Monetary – model. RFM variables have been proven to play an undeniable role in predicting customer churn (Buckinx & van den Poel, 2005; Jahromi et al., 2014). This model is often used when companies try to find whom to target in their campaigns. Recency stands for the elapsed time since the last purchase. The more recent a customer's purchase is, the more likely that the customer is active. The second variable, frequency, represents how often the customer has purchased in a certain time period. The frequency of purchases made by a customer can be a measure of customer churn likelihood in the future (Kumar & Reinartz, n.d.). The third variable, monetary value, represents the value of previous purchases. The monetary value of a customer's past purchases can be an indicator to predict future behavior (Jahromi et al., 2014). Besides the RFM variables, the L – length of the relationship is sometimes added to the model (Nwanganga & Chapple, 2020).

Referring back to the fact that it is crucial to have data relevant to the learning goal, sector-specific information should also be collected outside the RFM variables to have a successful prediction. Neslin et al. (2006) suggest including variables linked to previous marketing efforts. In particular, data on previous targeted offers to reduce churn would allow targeting based on response to churn reduction efforts rather than targeting on who is likely to churn.

The amount of the data is related to the selected algorithm, some algorithms perform well with smaller datasets, and some algorithms need large amounts of data to perform. This is also in contrast with the outcomes of the customer churn prediction tournament that was organized by the Teradata Center at Duke University; here, 82% of the participants used fewer than 80 predictors, where most of them even below 40 predictors. However, the other 18% used more than 140 predictors (Neslin et al., 2006). It stresses the importance of making careful decisions according to the case in which the prediction model is applied

When considering which data to select for the customer churn prediction model. It is also important to stress out for which time period you want to predict. Because not all available data is continuously, for example looking at the RFM variables. Frequency of purchases should be emphasized on a specific time period in order to get a meaningful value.

Following Gattermann-Itschert & Thonemann (2021) most churn prediction studies consider a single time window to predict customer churn. However, this can result in potentially missing relevant information from the data and it does not capture that conditions and churn motives can change over time. They suggest to use multiple time slices, where each time slice considers data from a specific time window. This means that features are computed relative to a reference point within that window. The multiple time slices are then combined to one dataset. This study concluded that a multi-slicing approach, performs better compared to a single-slice approach (Gattermann-Itschert & Thonemann, 2021).

2.3.2 Evaluation of customer churn prediction models

In supervised learning, the effectiveness of a machine learning model can be evaluated based on the number of errors that it makes. Most of the time, data scientists refer to the percentage of correct predictions, known as accuracy, when dealing with a classification problem. When dealing with regression problems, the difference between the values predicted by the algorithm and the actual values is evaluated. For unsupervised learning, it is more challenging to evaluate the model since there is no set ‘’right’’ or ‘’wrong’’ answer. The effectiveness lies in the value of its insight (Nwanganga & Chapple, 2020).

There are two types of errors. First, false positive errors; occur when the model labels an observation as predicted positive when it is, in reality, negative. These errors are also known as type 1 errors. False negative errors; occur when the model labels an observation as predicted negative when it is, in reality, positive. The false negative errors are also known as type 2 errors. Likewise, we label correctly predicted observations as true positives or true negatives. A simple way of gathering the numbers of true positives and true negatives is by constructing a confusion matrix. A confusion matrix (Table 1) shows the numbers of true positives, true negatives, false positives, and false negatives. The total number of items in all four quadrants equals the number of test cases (Baldi et al., 2000).

	Actually positive	**Actually negative**
Predicted positive	True positives (TP)	False positives (FP)
Predicted negative	False negatives (FN)	True negatives (TN)

Table 1: Confusion matrix

From the confusion matrix, multiple evaluation measures can be derived. The most popular evaluation metric for a classification task is accuracy. This is the number of correctly predicted samples divided by the total number of samples. This means that the accuracy gets higher if a model has more correct predictions. It goes from 0 (no correct predictions) to 1 (all predictions are correct). Although this metric shows a good interpretation of the model performance, it does not differentiate between the different effects of the results (Hackeling, 2014). Such as, if a customer is predicted as a non-churner but is actually a churner, the customer is going to churn while this could be prevented. The other way around yields for if a customer is predicted as a churner but is actually a non-churner, but this does not make as much impact. Therefore, other metrics are usually derived from the confusion matrix.

One is precision, the number of true positives divided by the sum of true positives and false positives. This equation indicates a low precision when there are relatively many false positives. Another measure is recall, also called sensitivity or True Positive Rate. The recall is calculated by dividing the number of true positives by the number of true positives and false negatives. Well-fitted models result in high recall scores. The last measure is the F1 score, which is calculated by conveying the balance between precision and recall. Higher F1 scores indicate better model performance (Baldi et al., 2000; Nwanganga & Chapple, 2020).

The Area Under the Curve (AUC) is a more graphically visible evaluation measurement. The AUC, also known as the area under the Receiver Operating Characteristic (ROC) curve, is used for ranking the performance of data mining models. The measures mentioned above are good evaluators but use one specific cutpoint. However, in many instances, predictors are encountered on a continuous or ordinal scale. In that case, it is better to assess the performance over the range of possible cutpoints, and the ROC curve achieves this. It includes a probabilistic framework which means it includes all possible decision thresholds (Mandrekar, 2010).

It is easy to interpret, as its area equals the probability that a randomly chosen positive example (a churning customer) has a higher probability of being positive than a randomly chosen negative example (a non-churner). The AUC value lies between 0.5 and 1, with 1 being an excellent classifier. However, it can be misleading as it gives equal weight to each value in the full range, even though only a limited range may be of the problem's interest (Kim & Moon, 2012).

2.4 Profit-based loss function and Uplift modeling

While classic retention management strategies focus on targeting customers with the highest probability of churn, more recent approaches focus on targeting customers based on their expected profit potential.

One of these approaches is the profit-based loss function outlined by Lemmens & Gupta (2020). The function ensures that customers are ranked based on the impact of the intervention after accounting for the cost of the intervention and that the model minimizes the cost of prediction errors by penalizing customers based on their expected profit lift. Lemmens & Gupta (2020) have proven with two field experiments that taking the profit-loss function into account leads to significantly more profitable campaigns than competing models.

A profit lift represents the net impact of the intervention of a customer. The campaign's profit is the sum of the profit lift of all targeted customers. The expected profit lift of a retention action given the intervention cost is explained in the following equation formulated by Lemmens & Gupta (2020). This approach fits many contexts, especially where organizations seek to target a set of individuals with a specific intervention.

$$E(\pi_i|\delta) = E(CLV_i - \delta|X_i, T_i = 1) - E(CLV_i|X_i, T_i = 0),$$

Where

$E(CLV_i - \delta|X_i, T_i = 1)$

is the net residual customer lifetime value (CLV) of customer *i* if targeted with an offer that costs δ.

And

$E(CLV_i|X_i, T_i = 0),$

Is the (net) residual CLV if customer i is not targeted.

If a customer is targeted, the net residual CLV is the discounted value of the cash flows after the campaign minus the per-customer cost of the retention intervention. Theoretically, the net residual CLV should be estimated over an infinite time horizon, but most companies and academics focus on a specific time period and use a truncated CLV (Lemmens & Gupta, 2020). The expected profit lift can be either negative or positive. It is negative if the cost of the intervention is greater than the residual CLV. This can happen because of waking up so-called "sleeping dogs". The intervention then reminds sleeping dogs of their dissatisfaction with the service, and the risk of churning is then increased (Ascarza et al., 2016).

The profit-based loss function can be used with any estimation technique and more advanced machine learning methods. Using the profit-based loss function in an organization's strategy allows them to assess the individual contribution of each customer to the total profit of a targeted retention campaign. Instead of ranking customers based on their probability of churn, they can be ranked based on their profitability, allowing the company to determine the optimum campaign target size to maximize profitability.

Besides the profit-based loss function, uplift modeling has also become a popular approach. This approach achieves even more improvement by targeting customers that are likely to churn and likely to be retained when targeted in a proactive retention campaign. This means that when high-probability churners are excluded from the campaign since it is known that they are not persuadable, a further increase in profitability can be achieved (Verbeke et al., 2018).

Uplift modeling aims to differentiate between segments of customers based on their response if treated and if not treated with an incentive (Ascarza, 2018; de Caigny et al., 2021; Lemmens & Gupta, 2020). For example, in some situations, it is seen that some customers do not respond when treated with an incentive, whereas they would stay loyal when not treated with an incentive. Table 2 shows the different situations that can happen following the matrices by de Caigny et al. (2021) and Verbeke et al. (2018). Four groups are differentiated in two dimensions, based on response behavior on being treated with an incentive or not.

Treated	**Stay**	Persuadable customers	Loyal customers
	Churn	Lost customers	Sleeping dogs
		Churn	**Stay**
		Not treated	

Table 2: Segmentation matrix uplift modeling.

1. Persuadable customers
   The persuadable customer accepts the offer when treated and will churn when they are not treated. These are the customer who is being sought. These are the customers that must be targeted in the proactive retention campaign and generate more revenue.
2. Sleeping dogs
   These are the opposite of persuadable customers. They will churn when treated with an incentive and stay loyal if they are not treated. It is of great importance to avoid these customers in the target group. It is a problematic segment since these customers have characteristics that might look like they are going to churn, and therefore basic prediction models will predict them as churners. They ‘’forget’’ they have a subscription, hence the name sleeping dogs.
3. Lost customers
   These are the customers who will churn no matter whether they are treated with an incentive or not. Targeting them in a retention campaign generates additional costs.
4. Loyal customers
   These customers will stay no matter whether they are treated with an incentive or not. Therefore, treating loyal customers does not generate additional returns but additional costs as they will accept the incentive, which costs money. This segment can cost more than the lost customers.

Verbeke et al. (2018) described that it could be possible that not all four segments exist. It depends on the customer base. For example, it can be possible that there are no sleeping dogs in a customer base for a particular campaign. In this situation, there is no risk of adversely affecting the customers. On the other hand, it is also possible that there are no persuadable customers. In this case, one should not run a campaign because no additional revenue will be generated. To overcome this, performing a cost-benefit analysis might be a good indicator. Often the fraction of persuadable customers is negligible.

In order to start with uplift modeling, data about whether customers responded to a retention campaign must be available. Preferably data about a campaign that is almost identical to the campaign that is going to be launched. If this data is not available, organizations can start obtaining this data by doing experimentations (Verbeke et al., 2018).

3. Methodology

This research is performed using quantitative analysis methods. Qualitative insights are used to derive meaningful and comprehensible explanations of the methods' contexts. Solution directions are inspired by existing literature, while exploratory data analyses and machine learning models are constructed empirically. This research is of inductive nature, as the research addresses a particular problem in a specific context and covers multiple stages, from data extraction to model deployment.

A data mining approach that is often used is called the CRISP-DM model. The research is conducted according to the CRISP-DM guidelines and execution of machine learning techniques. CRISP-DM is the most widely-used analytic methodology. The model was introduced in 1999. Although some more recent models have been developed, the original CRISP-DM model can still be recognized in recent studies, which remain focused on the traditional paradigm of a sequential list of stages from data to knowledge (Martinez-Plumed et al., 2021).

3.1 data preparation

The data preparation was considered the process of extracting the data from the database. Besides this, the data was cleaned in such a way that it became one extensive dataset instead of multiple datasets. This section contains information about how the prior feedback data is used in this study, which data is extracted from the database, what feature creation was done, and which time window was selected for extracting the data.

3.1.1 churn reasons as dependent variable

Compared to other customer churn prediction models, the uniqueness of this model is that it is able to predict customer churn reason. This is possible by using prior feedback data as a multi-class dependent variable instead of a binary variable which is used by usual customer churn prediction. In this multi-class label categorical data about churn reasons are incorporated. This subsection will explain in more detail how this multi-class variable was created.

The company collected customer feedback data from 749 churned customers. This data is important and valuable to understand why customers are unsatisfied and stop using the service. The feedback form customers had to complete after they had indicated churn could be found in Appendix C.

The form indicated seven churn reasons to choose from; no longer teaching, I chose a different (software) solution, needed more support, technical issues, too expensive, yogis don't enjoy using Momoyoga, and wanted additional features. However, it quickly emerged that with this number of categories, it would not be feasible to predict accurately why a customer will churn. This, along with class imbalance and the fact that there is relatively little data (749), the decision was made to merge the categories into a maximum of four categories.

These categories are based on the churn motives that emerge in the literature, namely price-related churn motives (too expensive) and service-related churn motives (technical issues, wanted additional features, yogi's don't enjoy Momoyoga, and I chose a different (software) solution). It was chosen to add "I chose a different (software) solution" to the service-related churn motive, but with the awareness that this could also be a price-related churn motive.

In addition, a relatively high percentage of customers left because they no longer teach yoga classes. This has become a separate category because these customers are assumed to be not persuadable to retain. They have no yoga studio anymore and, thus, no reason to use the service. When the data was captured, Covid greatly impacted the customer churn rate. Therefore, an additional category Covid was created for the customers that indicated Covid as a reason for churn. This leads to the following four categories.

- Price-related churn motives
- Service-related churn motives
- Covid
- No longer teaching

3.1.2 Data extraction

Knowing how the dependent variable (the multi-class label) was created, the independent variables could be selected and prepared for the model.

Momoyoga has a subscription-based business model where clients pay a fixed amount per month or annually (annually is cheaper). If customers pay annually, cancellation is not possible within these 12 months. However, customers that pay per month can always cancel their subscription. The subscription can be canceled via a simple button and feedback form in the application. For customers, there are no additional costs besides the subscription costs. Half (54%) of customers pay monthly, while the remaining (46%) pay annually.

For this research, data were extracted from the data warehouse (Grafana) of Momoyoga. The data source contains information about all customers that Momoyoga has ever had. They hold a pervasive database with demographic and customer behavior data. SQL, Excel, and Python were used for data preparation.

This study considered all customers (churned and active) from 01-12-2014 to 20-03-2022. A dataset, in this study called the base dataset, was constructed out of the database that contains demographic and behavioral information about all customers (current and churned) from their first activation date until 20-02-2022 or their churn date. This information was transferred into explanatory variables. These variables and an explanation can be found in Appendix A. There are two main types of data: structured and unstructured. Structured data is usually composed of numbers or words. Unstructured data is usually composed of everything else, including texts, images, videos, speech/audio, etc. (Rajendran, 2022). This research uses mainly structured data, especially the datatypes: numerical and categorical. The variables which include a date are seen as numerical timestamps.

3.1.3 Feature creation

From the raw dataset, additional features could be created from multiple features that could relate to customer churn. As a result, the following features were created:

- Days before activated: These are the number of days between the moment the account was created and the first moment they used the service.
- Days before payment: The number of days between account activation and first payment.
- Days active: The number of days a customer has been active, looking at the last activity and the date the account was created.

As the literature study stated, RFM variables are suspected to be good predictors of customer churn. The database was explored to determine if RFM variables could be added to the dataset. For example, through logical thinking, one assumption was made that the number of classes (Frequency) given by a client and the number of yogis that have registered per class might say something about whether a client will churn. With the assumption that if clients offer fewer classes and fewer yogis show up, a client is more likely to churn. The same yields for the amount of income (Monetary) for each yoga studio; less income could say something about customer churn. The Recency variables were already included in the dataset (Column: Updated). The new data is not directly applicable to the dataset created in the steps above. Therefore, three "new" features have been created. The following features were calculated and created:

- Number of registrations per lesson: These are the number of people registered for a specific class.
- Amount of lessons: This is the sum of the number of classes a yoga studio provides.
- Income: This is the income generated by the yoga studio. This revenue comes from the yogis.

It was essential to select a time period to add the created features to the base dataset. Through descriptive analyses of the data from the generated features, it can be concluded that about five to six months before the churn date, a significant decrease can be seen in the number of lessons provided and the number of registrations. Surprisingly, the chart of income feature is different compared to the others. It shows a decrease occurring four months prior to the churn date. See Figures 4,5, and 6.

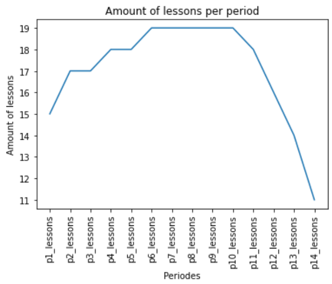

Figure 5: Amount of lessons per period

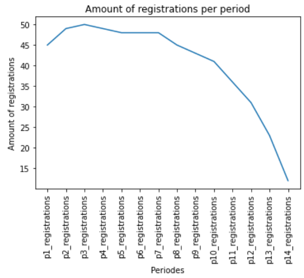

Figure 4: Amount of registrations per period

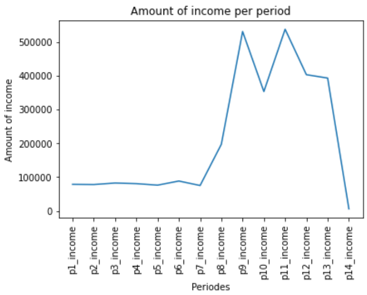

Figure 6: Amount of income per period

In close consultation with the business experts from the company and with the insights of the descriptive statistics in mind, a three-month period was chosen. How the time period was chosen mainly depended on some intuition about the product or service. A period of three months was chosen to still allow for a significant period of time prior to the churn date, but not too short or too long to be able to respond with a marketing action adequately. With this chosen time period, the model's output will answer the question of how likely a customer will churn in the upcoming three months.

For customers who have already been churned, data from the last three months of their subscription are picked because this period tells something about customer behavior right before the churn date. A random three-month period is picked for the remaining customers who have not been churned. This is randomly picked since these customers do not have a specific churn date. Therefore the goal is to check how a customer that never churned behaves in a casual three-month period.

These three months (Pt) were then divided into multiple time slices, in this case, three months (P1, P2, and P3). P1 is the first month of the three months, P2 is the second month, and P3 is the third month. The choice to break down the time period into individual months was made because it allows for a more detailed examination of how a given customer behaves at a specific time before the churn date.

The algorithm is then able to find relationships between these different periods. Because the time window is split into three periods, it also allows calculating the difference between two time periods. These differences are also added as additional variables in the dataset.

Figure 7 shows the time window graphically.

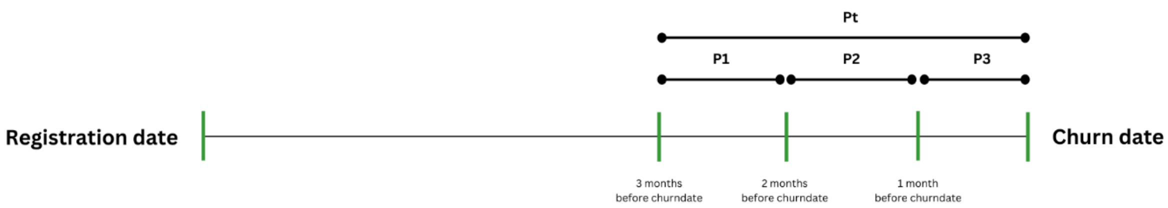

Figure 7: Chosen time window

The different datasets can converge together using the merge function in python; this function links ‘’Client_id’’ together, allowing one dataset to be created. The dataset was expanded with the created features; an overview can be found in Appendix B.

3.2 Pre-processing

After the data preparation stage, all data was extracted and the data set was complete. However, the raw data needed to be suitable for a machine learning model. This was done via data pre-processing. It needed to handle missing values, categorical data and class imbalance.

3.2.1 Missing values

The number of missing values per column has been assessed. Most of the missing values in the dataset make sense; for example, not all customers in the database have a churn date. Although there are some values that need to be clarified. There are 77 missing values for ‘’amount_of_lessons’’, 14 missing values for ‘’amount_of_teachers’’, 311 missing values for ‘’first_payment’’, 269 missing values for ‘’city’’’, 79 missing values for ‘’payment_frequency’’ and 1 missing value for ‘’first_activated’’.

Deleting all rows with missing values was not an option as this would reduce the dataset significantly, and valuable information will be deleted as well. The missing values in column ‘’number_of_lessons’’ means that no lessons were provided in that period. These missing values are thus set to 0. For ‘’amount_of_teachers’’, it is assumed that a missing value means there is only one teacher.

Besides this, numeric missing values are transformed to the median of that specific column. The remainder of the missing values (timestamps or categorical) have been ignored.

Also, XGBoost supports missing values by default and will automatically set it to NaN (Not a Number).

3.2.2 Numerical to Boolean

Some numerical values are stored as numbers but cannot be treated as numbers in the model. This is the situation with referral_id. The numerical value has no meaning here, as the number is an id of another customer. Therefore, it is only essential to know whether another customer has referred someone. Therefore, this variable has been converted from a numerical value to a boolean value. This means that the moment someone is referred, the value 1 is assigned, and when a customer is not referred, the value 0 is assigned.

3.2.3 One hot encoding

One hot encoding is a process of converting categorical data to improve predictions and the classification accuracy of a machine learning algorithm. It creates a new binary feature for each possible category and assigns a value of 1 to the feature of each sample that corresponds to its original category. One hot encoding was applied to the categorical features in the dataset to create a better classifier.

3.2.4 Class imbalance

Class imbalance is a problem that often happens in machine learning classification problems; it occurs when there is a skewness towards the majority class. When a model is trained on an imbalanced dataset, the learning of the model becomes biased towards the class in major. The categorical label with the churn reasons has to deal with imbalanced data. Also, there is a small imbalance between active customers (67%) and churned customers (33%).

Sample weights have been included to handle the class imbalance of this dataset. The idea of sample weights is to weigh the loss computed for different customers differently based on whether they belong to the majority of the minority classes. Thus, samples associated within the minor class, the churned customers, get assigned a higher weight than active customers.

3.2.5 Data standardization

Standardization of data is useful when the data have different scales. Standardization means that the variables are centered at zero, and the variance is standardized at 1. The effect of standardization is to rescale the features with the properties of a standard normal distribution. To standardize the data in this study, the Standardscaler from the skicit.learn library was used. Note that standardization is not required for decision tree classifiers, which means it is also not required for XGBoost.

3.3 Machine learning

With the steps mentioned above, the dataset ended up with 45 columns and 3260 rows. This means that 3260 unique customers dating back to 2014 form a suitable base for developing the prediction model. This dataset was ready for feeding into the prediction models.

3.3.1 Modeling techniques

It has emerged from the literature that it is best to compare multiple models side by side. This provides insight into which model performs best and thus has the highest prediction accuracy. The best-performing model is also used to move forward.

The modeling techniques that are used in this research are Logistic Regression (LR), Decision Tree (DT), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost), given that these models are appointed in the literature as among the most popular. Regarding the literature, XGBoost was the most promising model and this model was relied on the most in this study.

3.3.2 Definition X and Y

The X and Y variables for the machine learning model were determined. The list of variables included in the model is given below.

X = ‘country’, ‘payment_frequency’, ‘pricing_plan_id’, ‘amount_of_lessons’, ‘amount_of_teachers’, ‘referred’, ‘payment_reminders’, ‘max(created)’, ‘days_active’, ‘days_before_activated’, ‘days_before_payment’, ‘pt_lessons’, ‘p1_lessons’, ‘P2_lessons’, ‘p3_lessons’, ‘pt_registrations’, ‘p1_registrations’, ‘p2_registrations’, ‘p3_registrations’, ‘diff_lessons_p1_p2’, ‘diff_lessons_p2_p3’, ‘diff_lessons_p1_p3’, ‘diff_registrations_p1_p2’, ‘diff_registrations_p2_p3’, ‘diff_registrations_p1_p3’, ‘pt_income’, ‘p1_income’, ‘p2_income’, ‘p3_income’, ‘diff_income_p1_p2’, ‘diff_income_p2_p3’, ‘diff_income_p1_p3’

Binary classification:

Y = ‘Churn’

Multiclass classification:

Y = ‘Churn Reason Numeric’

Not all features from the data set are included in the X variables. This is because some were not logical to include or were used to calculate new features. For example, ‘customer_id’ was not included because it is not relevant for the prediction; this variable is purely to check which name belongs to the customer. The same yields for the variable ‘status’; this variable would already explain the dependent variable.

3.3.3 Training dataset and testing dataset

This dataset was randomly split into a training and testing dataset. 80% of the data is used for training, and 20% of the data is used for testing. This distribution is commonly used in the machine learning industry.

3.3.4 Hyperparameter optimization

Hyperparameters are all parameters within a model that can be adjusted arbitrarily by the operator before training the data. The aim is to find a set of hyperparameters with the right combinations that help to gain the best performance. However, if the hyperparameters are not correctly tuned, the estimated model parameters produce suboptimal results. Which means the model will make more errors. Therefore, hyperparameter optimization was applied in this study by using the hyper opt library.

3.3.5 Cross-validation

Since the dataset was split into only one set of training and one set of testing data, the models' performance metrics are highly reliant on those two sets. These two sets are trained and validated only once and may perform differently if trained multiple times on different subsets of the data. Therefore, K-fold cross-validation was used with k=4, k=10, and k=15. In K-fold cross-validation, the split in training and testing is performed K times with a different subset of the data.

3.4 Model evaluation

As this study has to deal with an imbalanced dataset, the evaluation metric which will be relied on the most is the AUC. The goal is to get the AUC as high as possible.

Despite accuracy being seen as the most popular and easiest-to-interpret evaluation metric, this metric is not always very meaningful. This is because of imbalanced datasets. In the case of the dataset in this study, 2/3 of the customers are active customers, and 1/3 are churned customers. If the model predicts a customer as active, this random probability of a good prediction is already 66.7% by default, and vice versa for a churned customer is this 33.3%. Therefore, accuracy can be helpful in some comprehension and simple problems but does not give a reliable result.

The AUC metric utilizes probabilities of class prediction (Zvornicanin, 2022). This allows the ability to evaluate and compare the models more precisely. AUC means Area under the Curve, and with the curve, the ROC curve is meant. The ROC curve shows the relationship between the false-positive and true-positive rates for different probability thresholds of model predictions. A random prediction means a straight line from 0.0 to 1.1 and gives an AUC score of 0.5. A perfect AUC score is 1, and the worst is 0.

3.5 The output

As discussed earlier, the customer churn prediction model can be appropriate as a marketing tool to determine who will be targeted in a proactive retention campaign. For example, where an uplift model or profit-loss function focuses on which customers yield the most profit, this model can focus on sending a (personalized) incentive based on the reason for churn. Besides this, knowing the reason for churn makes it easier to determine if this customer is persuadable to retain.

From the output of the multiclass customer churn prediction model, a few assumptions were made regarding segmentation. Uplift modeling was used as an inspiration. Note that these assumptions depend entirely on the data an organization is dealing with. This can vary from organization to organization and from issue to issue. The following assumptions are based on the data used in this particular study. Thus, the segmentation classification in this study is not generalizable one-to-one with other studies and datasets. On the other hand, the reasoning behind the segmentation can, of course, be applied to other data sets.

The first assumption is that customers with a high churn probability based on the reason "no longer teaching" can be classified as lost customers. This is because they quit their job as a yoga teacher or sell their yoga studio. They can't use the service anymore, and therefore these customers are not persuadable to stay customers.

The second assumption is that customers with a high churn probability in the categories: service-related and price-related issues are the most persuadable customers. This is because they are likely to churn because they are not satisfied with some aspects of the service or the price of the service. When this changes, customers may be satisfied again and stay customers.

The third assumption is that customers classified with churn reason covid are wrong classified and probably will not churn as Covid is no longer an issue. This leads to the following segmentation matrix in Figure 8.

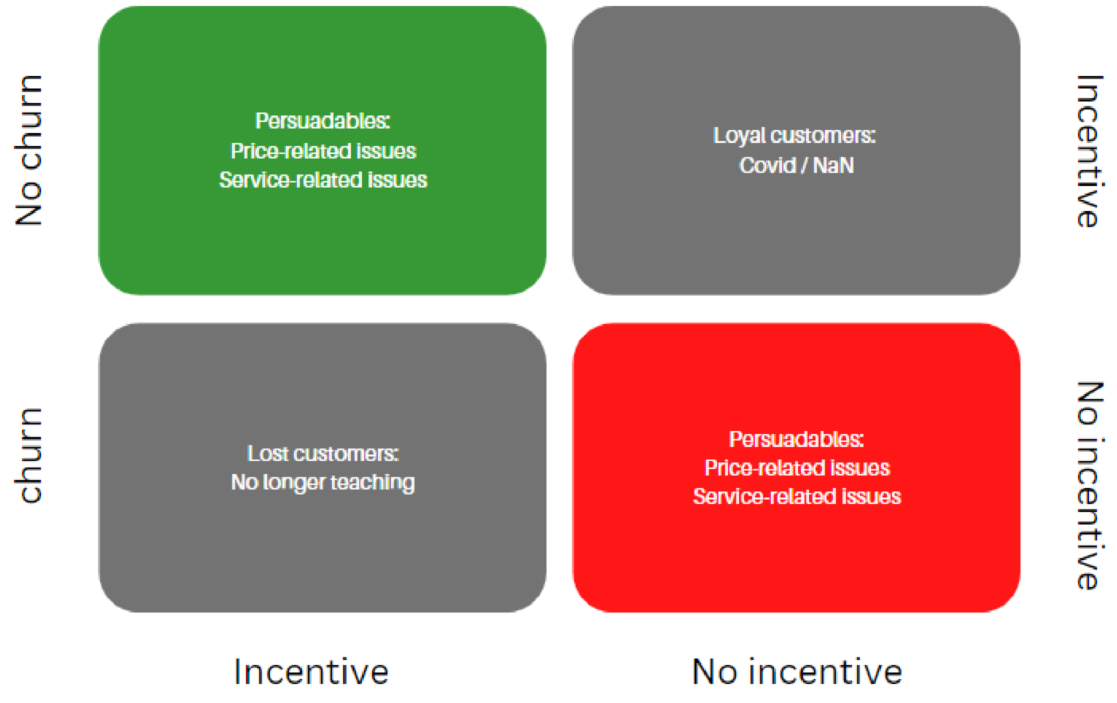

Figure 8: Segmentation matrix

The segmentation matrix shows that for persuadable customers sending an incentive is necessary. If they receive an incentive, they can be retained, but if they do not, there is a big chance that the customer will churn.

For loyal customers, it does not matter whether to receive an incentive; they will stay customers no matter what. The same yields for the lost customers; they will churn nevertheless. These two segmentations only cost money when they are targeted in a proactive retention campaign.

Therefore, it is of great importance to find the customers that are going to churn because of service-related issues or price-related issues. These are the customers to target in a proactive retention campaign.

With this classification, we assume that the prediction model can predict the sleeping dogs as no churners. However, this cannot be granted and will be further elaborated on in the discussion chapter.

3.6 The workflow

When all the steps mentioned above had been carried out and the database was ready to be put through the various modeling techniques mentioned, the first step was a binary classification task. This means that the model was going to predict if someone is going to churn or if someone is not going to churn. This intermediate step was chosen to ensure that the model could accurately predict a churner. After this was proven, the multiclass prediction model predicting the reason for churn was executed. The only difference between these two steps is the dependent (Y) variable. Whereas with multiclass, this is categorical instead of binary. The step-by-step chart in Figure 9 visually depicts the process of the output.

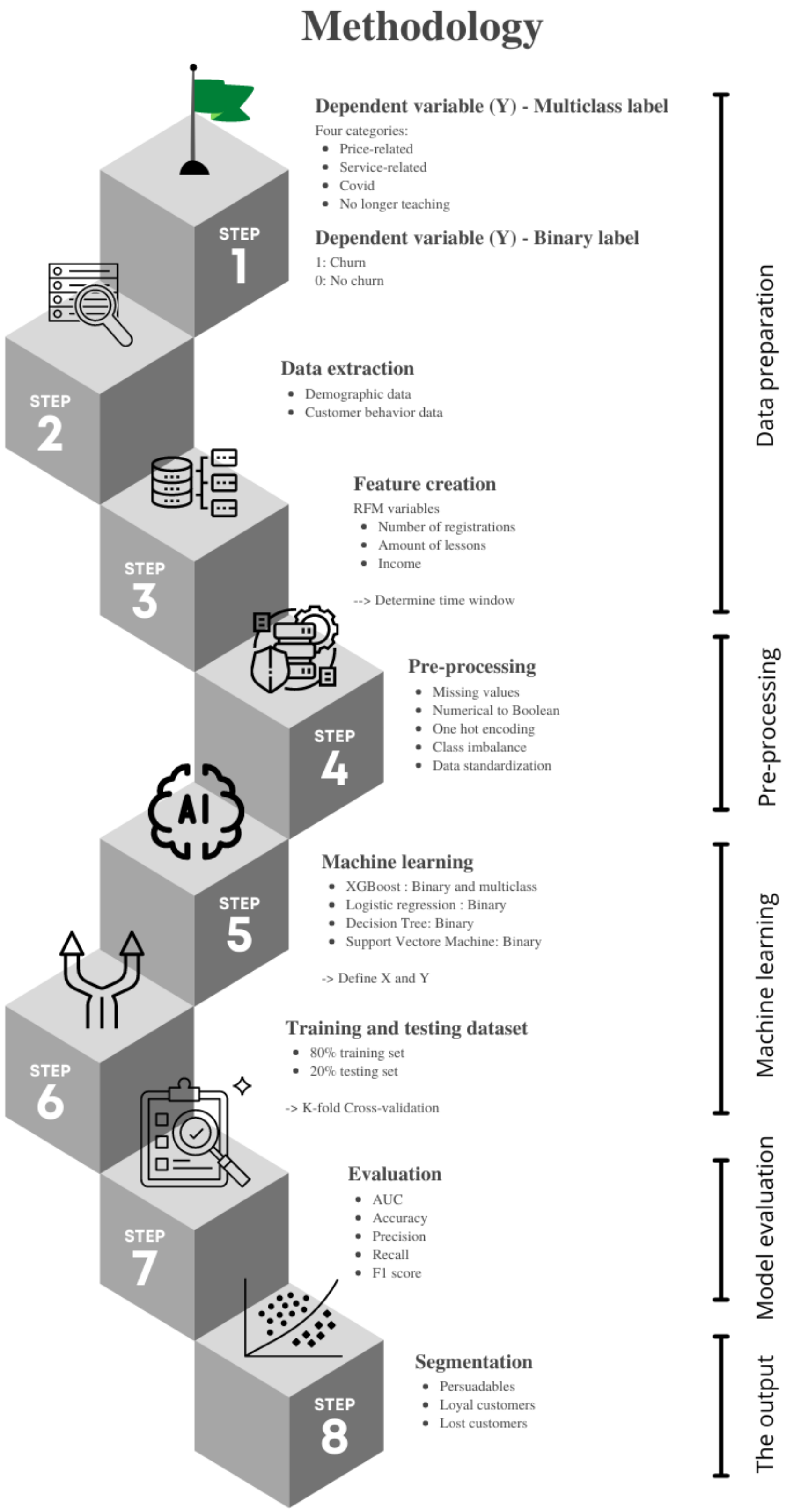

Figure 9: Step by step chart methodology

4. Results

The results in this chapter are mainly based on the output of the machine learning model(s) that is developed as the methodology chapter above describes. After all, the output, and thus results, help to answer the research question: ‘’How to design a customer churn prediction model that is competent in predicting customer churn reasons using AI and prior feedback data about customer churn reasons?’’

4.1 Results binary customer churn prediction model

Let start with the results by outlining the features that correlate with the dependent (Y) variable ‘’churn’’. A correlation shows which variables have a relationship with the dependent variable churn. The entire table with all features and their correlations can be found in Appendix D. Overall, the correlations are not high, as the highest correlation is 0.223. Usually, correlations between 0.5 and 1.0 are seen as strong relationships. Although the correlations in this study are seen as weakly correlated variables, the ones greater than 0.1 are selected and outlined. These are the following ranked from high to low:

- P3_lessons (0.22)
- P3_registrations (0.20)
- Diff_lessons_p1_p3 (0.18)
- Diff_lessons_p2_p2 (0.17)
- Pt_lessons (0.17)
- P2_lessons (0.15)
- Pt_registrations (0.14)
- P2_registrations (0.14)
- Days_active (0.13)
- Waiting_list (0.12)
- Amount_of_teachers (0.10)

Good to mention that correlation is appropriate for examining relationships between numerical data rather than categorical data; therefore, some features are not in the correlation table. This output shows that the RFM variables added during the feature creation step are the highest correlated features with customer churn.

To answer sub-question 3: ‘’ Which machine learning models are best for predicting customer churn (reason)?’’ multiple machine learning (LR, DT, SVM, and XGBoost) models and their performances are evaluated. The results of this comparison are presented in Table 3, which includes all the measurements (AUC, accuracy, recall, precision, F1 score) to measure and compare the performance of the different classification models.

As expected, the results show that XGBoost significantly outperforms the rest of the modeling techniques for this binary classification task. As seen in Table 3, every metric is the best scoring by

XGBoost. The worst-performing model compared to the others is DT. Multiple K-folds (5, 10, and 15) were used to find the best-performing model for each modeling technique. For DT and SVM, this was 10 K-folds, and for LR, it was 15 K-folds. Applying cross-validation for XGBoost did not improve the model; therefore, cross-validation was not used for this model. The XGBoost model has an accuracy of 82% on our test set in predicting whether a customer will churn in three months. Accuracy outcomes above 0.7 are considered a good performing model. The AUC score of 0.9 can be interpreted as a 90% chance that this model will be able to distinguish between the positive value (1), churn, and the negative value (0), no churn. An AUC of 0.5 suggests no discrimination, scores between 0.7 and 0.8 are considered good, 0.8 to 0.9 is considered excellent, and greater than 0.9 is considered outstanding (Mandrekar, 2010).

		LR (k=15)	**DT (k=10)**	**SVM (k=10)**	**XGBoost**
AUC	Rule of thumb: > 0.85 means high classification accuracy	0.84	0.72	0.82	0.90
Accuracy	Rule of thumb: >0.7 is great model performance	0.76	0.75	0.76	0.82
Recall	Higher the better,	0.64	0.68	0.72	0.78
Precision	Higher the better	0.70	0.67	0.66	0.76
F1 score	Higher the better	0.76	0.76	0.73	0.82

Table 3: Performance of machine learning methods for binary classification

Figure 10 shows the ROC curve of the XGBoost classifier, which also shows the AUC score. This plot show clear evidence that the model is not at random because, in that case, the line would have been straight from 0.0 to 1.1.

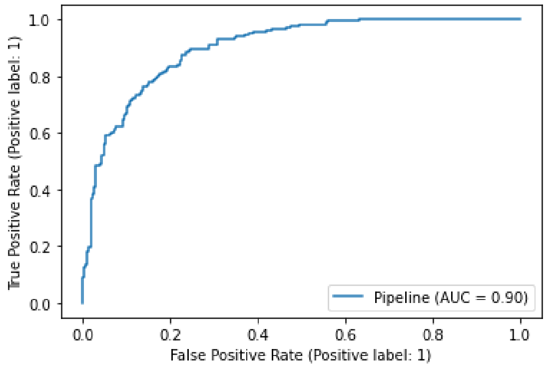

Figure 10: ROC Curve XGBoost binary classifier

As mentioned earlier, some of the evaluation metrics shown in Table 3 are derived from a confusion matrix. As XGBoost is the best-performing model, this confusion matrix is displayed in Table 4. The other confusion matrixes derived from the different modeling techniques can be found in Appendix E. The importance is to get as many correct predictions as possible, thus, true positives (TP) and true Negatives (TN). Fortunately, a large preponderance can be seen in these two green-colored cells. In the other direction, the goal is to have as few incorrectly predicted instances as possible. Hence, these are the false positives (FP) and false negatives (FN).

	Predicted: churn	**Predicted: no churn**
Actual: churn	TP: 196	FN: 55
Actual: no churn	FP: 63	TN: 338

Table 4: Confusion matrix XGBoost classifier

Feature importance refers to a technique that calculates a score for all input features for the model. It represent the ‘’importance’’ of each feature. A higher feature importance score means that the specific feature will have a larger effect on the XGBoost model that is used to predict customer churn. The following ten features are considered as important by the binary prediction model from high importance to low importance.

- Diff_registrations_p1_p3
- Payment_frequency_year
- Diff_lessons_p2_p3
- P3_lessons
- Payment_frequency_month
- P2_registrations
- P1_lessons
- P1_income
- Diff_income_p2_p3
- Payment_reminders

Again, these results show that the RFM variables are important in predicting customer churn. In addition, it also matters how a customer pays, monthly or annually. Lastly, the amount of payment reminders is also an important feature for predicting customer churn.

The ''end'' output of the customer churn prediction with XGBoost classifier is presented in Figure 11.

```
[[0.98535633 0.01464366]
 [0.5895926  0.41040742]
 [0.9923749  0.00762509]
 ...
 [0.19201207 0.8079879 ]
 [0.42341888 0.5765811 ]
 [0.9774377  0.02256235]]
```

Figure 11: Output table customer churn prediction

The first column is the probability for binary value 0, which means no churn, and the second column is the probability for binary value 1, which means churn. The results show that, for example, the first customer (row 1) has a 1.5% chance of churning. While for the 4th row in this output, this customer has an 80.8% chance to churn. The results show the probability of churn in the upcoming three months. This is because of the time window that was chosen before.

The results show that this binary customer churn prediction model performs excellent regarding the rule of thumbs of Mandrekar (2010). This means that the organization possesses appropriate data for customer churn prediction. It is abundantly clear that XGBoost is the best-performing model for this dataset. Using this binary model as a baseline, the next step was to test whether creating a multi-class prediction model with the same data can predict the multi-class label churn reason.

4.2 Results multiclass customer churn prediction model

Also, for the second prediction model, the multiclass customer churn prediction model, the correlations are explored for the dependent (Y) variable ''churn reason''. The correlations are even worse than in step 1, the binary classification. The entire table with correlations can be found in Appendix D. The features with a correlation greater than 0.1 are outlined here from high correlation to low correlation:

- P3_registrations (0.14)
- Diff_lessons_p1_p3 (0.13)
- P3_lessons (0.13)
- Amount_of_lessons (0.11)
- Referral_id (0.10)
- Diff_lessons_p2_p3 (0.10)
- P2_registrations (0.10)

Although fewer features correlate with the dependent variable compared to step 1, the correlating features are almost comparable to the list in step 1. Only ''referral_id'' was not correlating before.

The results of the multi-class prediction model are shown in Table 5. The multi-class prediction model is only performed with XGBoost as the results in step 1, and the literature background showed clear evidence that this model outperforms other machine learning models in customer churn prediction. What is noticeable is that different terms are used for the measures compared to the measurements in Step 1, the binary classification. This is because it concerns a multi-class prediction. An average of the measurements of each class has to be taken. An average can be calculated differently, which is why more measurements are presented in Table 5.

Also, the interpretation of a multi-class prediction differs from that of the binary classification. In multi-class prediction, accuracy is not a good measurement as it is challenging to conclude. The accuracy score of 0.55 means that the model can correctly predict 55% of the time. In a binary classification, it is interpreted that the model is random and therefore considered as a poor model. Although, since there are five different classes to predict, this score is better than it sounds. If the classes are equally balanced, a random prediction means 20%. Observing the accuracy with this interpretation, a score of 55% is not that bad. However, as described earlier, the dataset is imbalanced. XGBoost can handle imbalanced datasets, but it is challenging to say precisely what accuracy means in this case. Therefore, looking at the AUC, precision and recall score in a multi-class prediction is better.

Yang et al. (2021) argue that for multi-class prediction, a simple method to calculate the AUC score is by taking the average AUC of pairwise binary AUC's. It is calculated via the one versus rest method. This means that it compares each class against all other classes at the same time. One class is considered as the "positive" class, and the other classes are considered as the "negative" class. The AUC score of 0.74 reflects how well the classifier predicted each class. The AUC score of 0.74 suggests a 74% chance that the model correctly distinguishes a predicted churn reason from a false predicted churn reason.

The precision score can be interpreted as the model's accuracy for classifying a sample as positive. This means that this model can predict 31% of the time a correct prediction from the positive predictions. The recall score can be interpreted as the ability of the model to detect positive samples. This model can detect 35% of the positive samples.

Evaluation metric	**Output multiclass XGBoost**
Accuracy	0.55
Macro averaged precision : calculate precision for all classes individually and then average them	0.31

Micro averaged precision: calculate class wise true positive and false positive and then use that to calculate overall precision	0.5
Macro averaged recall	0.35
Micro averaged recall	0.5
Macro averaged F1 score	0.28
Micro averaged F1 score	0.5
Averaged AUC score	0.74

Table 5: Performance of multiclass XGBoost classifier

Again, the calculated measurement scores in Table 5 are derived from the confusion matrix. The confusion matrix from the multi-class model is displayed in Table 6. The green-colored cells are the correctly predicted instances. Note that the prediction of the NaN category made the average measurements higher, which indicates that the model performs worse for only the churn reason categories. For example, looking at the category 'price-related,' only 3 out of 65 (4,6%) predictions were predicted correctly. What also stands out is that 106 instances with the predicted reason 'Covid' actually are no churners. A logical explanation for this is that customers were forced to quit due to Covid, but it was not per se that they were not happy with the service. Without Covid, they would not have churned. Nevertheless, the model appears to be great at predicting who will not churn.

However, due to the small amount of data on customer churn reasons it is chosen to keep the NaN values in the dataset. The results from the multi-class model without the NaN category are presented in Appendix F.

		Predicted				
		Covid	**No longer teaching**	**Price-related**	**Service-related**	**NaN**
Actual	**Covid**	10	5	6	1	4
	No longer teaching	18	18	13	15	8
	Price-related	6	3	3	4	0
	Service-related	12	5	3	17	7
	NaN	106	32	40	38	278

Table 6: Confusion matrix XGBoost multiclass classifier

Feature importance refers to a technique that calculates a score for all input features for the model. It represent the ‘’importance’’ of each feature. A higher feature importance score means the specific feature will have a more significant effect on the XGBoost model used to predict the churn reason. The multi-class prediction model considers the following ten features important from high to low importance.

- P3_income
- Country_rou
- Pt_income
- Country_svn
- Country_prt
- P1_lessons
- Diff_registrations_p1_p3
- P1_registrations
- Payment_frequency_year
- Diff_lessons_p1_p2

Notable among these feature importances is that some countries play an important role in predicting customer churn. This could be a coincidental relationship/importance because, for example, a relatively large number of customers from a particular country were churned by chance. However, this could also not be a coincidence, and it indicates, for example, that customers from country "rou" churn more easily than customers from, say, another country. It is difficult to pinpoint an explanation for this at this point in research. Furthermore, RFM variables are again important here, and payment method also reappears.

The ‘’final’’ output of the multiclass customer churn prediction with XGBoost is presented in Figure 12.

```
['Covid' 'No longer teaching' 'Price-related' 'Service-related' nan]

[0.00627515 0.00526122 0.0143755  0.07272766 0.9013605 ]
[0.27938104 0.2038716  0.09339023 0.18737306 0.23598406]
[0.00239848 0.00542429 0.00097984 0.21848565 0.77271175]
...
[0.28383803 0.17375512 0.16836384 0.24769087 0.12635218]
[0.26788902 0.21773812 0.14984459 0.16802569 0.19650261]
[0.07601971 0.03154602 0.03057735 0.10752729 0.7543296 ]
```

Figure 12: Output table multiclass customer churn reason prediction

From this output matrix, one can see that customer 1 (row 1) has the highest probability in category NaN, which means no churn. Another example is row 4; this customer has a 28% chance to churn because of Covid, 17,3% on no longer teaching, 16,8% on price-related, 24,7% on service-related, and 12.6% that the customer is not going to churn.

All these results together result in an output deliverable which is a dataset. For now, a threshold for deciding to classify a customer for churn or no churn is set at 50%, but it can be set on any probability. The deliverable will look as follows (Table 7).

			Churn reasons			
Customer_id	**Churn (Threshold 0.5%)**	**Churn probability (binary classifier)**	**Covid**	**No longer teaching**	**Price-related**	**Service-related**
1880	0	1,46%	0.06%	0.05%	1.43%	7,27%
1328	1	51%	27,94%	20,39%	9,33%	18,73%
1886	0	0,76%	0,02%	0,05%	0.00%	21,84%
716	1	80,8%	28,38%	17,37%	16,83%	24,67%
1139	1	57,6%	26,78%	21,77%	14,98%	16,80%
3067	0	2,25%	7,60%	3,15%	3,06%	10,75%

Table 7: Output deliverable

The deliverable shows that the percentages for the predicted churn reasons are pretty close to each other for these instances. This makes it difficult to determine for what reason a customer will probably churn. For example, one can argue that by taking the highest percentage and assuming it is the reason for churn, but it is also possible to set a threshold. For example, a reason for churn must have at least a 40% chance to allow one to "determine" this churn reason as the correct reason.

The segmentations discussed in the methodology chapter can be completed from the output deliverable. This study did this with the test set (652 instances), assuming the highest percentage in the row counting. This leads to the following numbers per churn reason:

- Covid: 68
- No longer teaching: 100
- Price-related: 49
- Service-related: 97
- NaN (no churn): 338

The completed segmentation matrix is shown in Figure 13.

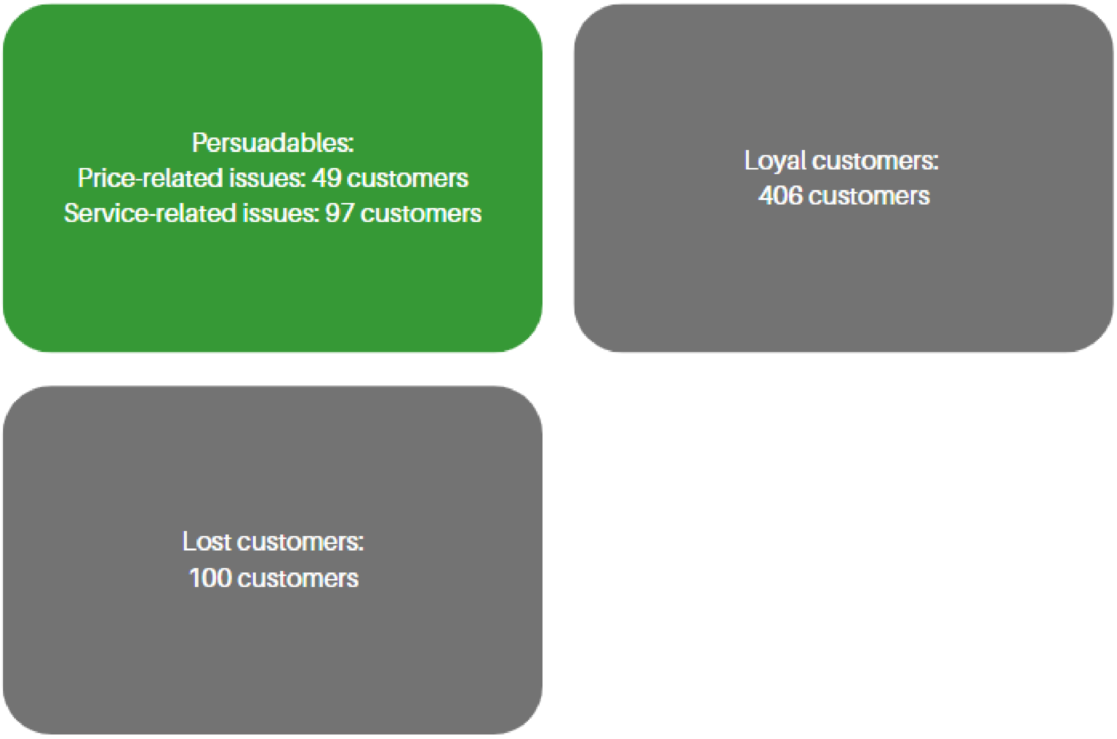

Figure 13: Segmentation matrix from test set

Figure 13 shows that 49 customers have a high probability of churning from the test set because of price-related issues. In addition, there are also 97 customers with a high probability of churning because of service-related issues. These customers are eligible for the proactive retention campaign as they are assumed to be persuadable customers. The remaining customers are not eligible for the proactive retention campaign.

5. Conclusion

Several key findings emerged from the results; these will be discussed in this chapter. The key findings are followed by an implementation advice for the organization.

5.1 Key findings

Referring to the research goal, this study provide insights into 1.) The development of a customer churn prediction model that can predict customer churn reason and 2.) How can the prediction of churn reasons be used for proactive marketing actions i.e. proactive retention campaign.

Both goals are achieved by using a multi-class customer churn prediction model with a categorical dependent variable, including prior customers' churn data, and by proposing a segmentation matrix by deciding whom to target in a proactive retention campaign.

Below, the sub-questions will be briefly discussed one by one as the findings lead to answering the main research question:

Sub-question 1: Which data is useful for customer churn (reason) prediction?
- From the literature, it can be concluded that demographic data, customer behavioral data, and RFM variables are key for predicting customer churn. The results show that a binary classification model with high accuracy and AUC score was obtained by incorporating these data elements. For predicting customer churn reasons, it can be concluded that the RFM variables play a significant role. This is shown by the results on correlations and feature importance.

Sub-question 2: What are possible reasons for churn?
- From the literature and the feedback form, two overarching reasons for customer churn can be identified. These reasons are price-related and service-related. However, for each product or service category, other industry-specific reasons may emerge.

Sub-question 3: Which machine learning models are best for predicting customer churn (reason)?
- For predicting customer churn and consequently customer churn reason, the best model, in this use case, is XGBoost. Overall, adding boosting is a valuable implementation for churn prediction (Vafeiadis et al., 2015).

Sub-question 4: Which evaluation metrics can be used to evaluate the prediction model?
- AUC (ROC curve) and accuracy are the most important evaluation metrics as they give a quick and clear overview of how the model performs. The confusion matrix is important as well since other metrics, such as precision, recall, and F1 score, can be calculated from this matrix. The average of these metrics can be used for a multi-class prediction model.

Sub-question 5: How can a customer churn prediction model be used for marketing actions?
- The literature shows that a customer churn prediction model can be used in multiple ways for marketing actions but is mainly used when creating a proactive retention campaign. It is common to target customers with a high churn probability. Arguments from the literature show that it is essential to not only look at the highest probability but also at the expectation that a customer can be persuaded to retain. In other words, is the customer more valuable than the campaign costs? The predictive model developed in this study makes it possible to segment customers based on the predicted churn reason. As such, the customer churn prediction model can be used to target only the likely persuadable in a proactive marketing campaign.

Putting these insights from the sub-questions together helped to answer this study's main research question: 'How to design a customer churn prediction model in a SaaS context that is competent in predicting customer churn reasons using AI and prior customer churn reasons data?' How the model was designed can be read in Chapter 3: Methodology.

The results of the model show that it is possible to predict customer churn reasons, but with the recommendation that this model cannot yet be assumed to be true. For this, the AUC score and accuracy still need to be higher. On the other hand, the binary classification model has shown that the collected data can accurately predict customer churn. Therefore, the conclusion can be drawn that if data continue to be added to the multiclass prediction model, it will perform better and better. Nevertheless, the model keeps learning from its mistakes.

5.2 Implementation advice

The developed model can apply as a working prototype based on a python script and can be re-runned with new and updated data files. However, as discussed above, the dataset must be updated continuously with new data such that the model perform better than its current performance. Successful implementation of the model will provide a more accurate prediction of churn reasons and, therefore, the ability to execute more targeted marketing actions. This can result in more profit for the organization.

The organization can also choose to use only the binary classification model to anticipate customer churn. This model has a solid performance and can be implemented immediately to target customers with a high probability of churn in a proactive retention campaign. Although the organization must be aware that this also means targeting many customers who are not persuadable.

The organization is advised to conduct experiments to start testing with the multi-class prediction model and segmentation matrix. This is because, through experimentation, the effectiveness of the model and the proactive retention campaign can be measured. This advice is based on the uplift models discussed in Chapter 2: Theoretical background. The results of the experiments should be carefully recorded and added to the database. Generating this data will ensure, on the one hand, that the organization can develop an uplift model in the future and, on the other hand, measure whether the customers are segmented adequately as persuadable by the model. Nevertheless, this will also ensure that sleeping dogs can be better identified.

One of the goals of predicting customer churn reason is to provide opportunities to differentiate incentives according to the reason. For customers with the reason price-related, the assumption is to give a price discount, while for a service-related one, it is wanted to ensure solving the service-related problem. However, this complexity makes performing experiments more complicated.

Multiple experiments and iterations are recommended to investigate different objectives and collect data. Some examples of objectives and experiments could be:

- Does the churn rate decrease when incentives are sent to the persuadable segment? In this experiment, it is crucial to have a control group of a randomly picked customer base that is the same size as the persuadable segment. However, this experiment makes it more challenging to differentiate between incentives.

- Does differentiate incentives on the persuadable yield more profit? In this experiment, the test is only on the persuadable segment. One group receives the same incentive, and one group receives a differentiated one (based on price and service). Next, the results can be tested by looking at the number of accepted incentives and the churn rate.

- Is the predicted reason for churn correct among customers and therefore the segment they are positioned in? Here the organization will ask for dissatisfaction even before the customer has had the opportunity to churn. With this information, the model receives new data on whether the prediction is correct.

6. Discussion

6.1 limitations

Even though the study was conducted with strict attention and to the highest possible standard, several limitations must be mentioned. Critical limitations that should be taken into account are:

- It must be highlighted that the questioning of the feedback form has not been fully thought through. For example, the customer only had one option to indicate a reason for churn, while the customer may have had several reasons to cancel the subscription. For the follow-up, more attention should be focused on identifying the best way to collect customer churn reasons. The suggestion for the short term would be to let customers choose between price-related, service-related, or no longer teaching.

- In addition, relatively few customer churn reasons were known, which resulted in a relatively small dataset. This made predicting churn reasons significantly more difficult than predicting churn on itself. This might be the reason for the low accuracy of the multiclass prediction model. Once more data is obtained, it is also possible to expand the categories of churn reasons and send even more targeted incentives.

- Unfortunately, Covid significantly impacted the organization's churn rate. The dataset included this period. This mainly was captured by asking churned customers if Covid played a role in the cancellation. However, in all likelihood, this does not cover all instances, and there will be some incorrectly predicted instances because of Covid.

- This research has paid little attention to identifying sleeping dogs compared to, for example, an uplift model. At this level of research, it was assumed that the models are so intelligent that these sleeping dogs are considered automatically as no churners.

- Because XGBoost was highly promising and clearly outperformed the rest of the models, there was not much focus on improving the performance of the other models (LR, SVM and DT). More focus and testing on the data preparation and data pre-processing phases could possibly have influenced this. Still, in all likelihood, this would not have outperformed the XGBoost model.

- The study used a time window of 3 months without testing what other time windows would do to the result. For the model to perform better, it could have been that other time windows would have been better.

- Lastly, to generalize this study's insights, it would have been better to conduct the study in multiple organizational settings with different datasets. The model and approach of this study is now only focused on one dataset and one specific use case.

6.2 Future research

Beyond the implementation advice and discussion, there are several exciting perspectives for future research that could complement or improve this research. These are as follows:

- Developing an uplift model that incorporates customer churn reasons. This model would be promising since it can consider the effectiveness of marketing campaigns. The model could predict the reason for churn and whether this person is persuadable with an incentive. As mentioned earlier, for developing such a model, data must already have been collected from a similar proactive campaign.

- The application of this study in other organizations. After all, the reasoning behind the developed model could well be generalizable to other organizations. Most likely in organizations working with a membership-based service or product. They can benefit from this study as it provides a good structure for setting up a prediction model that accounts for customer churn reasons. Furthermore, it can serve as inspiration to start requesting churn reasons via a feedback form.

- To complement this research, more extensive research could be done on what reasons and in what way reasons for churn are best requested from churned customers. In the end, this could improve the prediction of the model. Better marketing actions can be initiated when it is well understood why someone is no longer satisfied and thus churned.

- In addition, it will also be interesting to examine in the future how accuracy improves as there are more instances with a churn reason in the dataset.

- To further develop the model, one could add additional variables to the database. This could be done by, for example, feature engineering or querying new data from the organization's customers. Adding additional RFM variables could also lead to a better prediction.

- Finally, it is relevant to research how experiments can best be executed to test the model that is developed in this study. This could be beneficial since the future research will make sure the experiments will be handled as efficient as possible in order to waist as little money and time as possible.

7. References

Ahn, J. H., Han, S. P., & 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. https://doi.org/10.1016/j.telpol.2006.09.006

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, *55*(1), 80–98. https://doi.org/10.1509/jmr.16.0163

Ascarza, E., Iyengar, R., & Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, *53*(1), 46–60. https://doi.org/10.1509/jmr.13.0483

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schrift, R. (2018). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. *Customer Needs and Solutions*, *5*(1–2), 65–81. https://doi.org/10.1007/s40547-017-0080-0

Ascarza, E., Netzer, O., & Hardie, B. G. S. (2018). Some customers would rather leave without saying goodbye. *Marketing Science*, *37*(1), 54–77. https://doi.org/10.1287/mksc.2017.1057

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. In *BIOINFORMATICS REVIEW* (Vol. 16, Issue 5).

Bhattacharyya, J., & Dash, M. K. (2021). Investigation of customer churn insights and intelligence from social media: a netnographic research. *Online Information Review*, *45*(1), 174–206. https://doi.org/10.1108/OIR-02-2020-0048

Buckinx, W., & van den Poel, D. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*(1), 252–268. https://doi.org/10.1016/j.ejor.2003.12.010

Burez, J., & van den Poel, D. (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), 277–288. https://doi.org/10.1016/j.eswa.2005.11.037

Burez, J., & van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, *36*(3 PART 1), 4626–4636. https://doi.org/10.1016/j.eswa.2008.05.027

de Caigny, A., Coussement, K., Verbeke, W., Idbenjra, K., & Phan, M. (2021). Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach. *Industrial Marketing Management*, *99*, 28–39. https://doi.org/10.1016/j.indmarman.2021.10.001

Frank, B., & Pittges, J. (n.d.). *Analyzing Customer Churn in the Software as a Service (SaaS) Industry*.

Gattermann-Itschert, T., & Thonemann, U. W. (2021). How training on multiple time slices improves performance in churn prediction. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2021.05.035

Gür Ali, Ö., & Aritürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications*, *41*(17), 7889–7903. https://doi.org/10.1016/j.eswa.2014.06.018

Gustafsson, A., Johnson, M. D., Roos, I., Maynard, D., & Ross, S. M. (2005). *The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention*.

Hackeling, Gavin. (2014). *Mastering Machine Learning with scikit-learn.* Packt Publishing.

Jahromi, A. T., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, *43*(7), 1258–1268. https://doi.org/10.1016/j.indmarman.2014.06.016

Kim, H. S., & Yoon, C. H. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, *28*(9–10), 751–765. https://doi.org/10.1016/j.telpol.2004.05.013

Kim, Y. S., & Moon, S. (2012). Measuring the success of retention management models built on churn probability, retention probability, and expected yearly revenues. *Expert Systems with Applications*, *39*(14), 11718–11727. https://doi.org/10.1016/j.eswa.2012.04.048

Kumar, V., Bhagwat, Y., & Zhang, X. (2015). Regaining "lost" customers: The predictive power of first-lifetime behavior, the reason for defection, and the nature of the win-back offer. *Journal of Marketing*, *79*(4), 34–55. https://doi.org/10.1509/jm.14.0107

Kumar, V., & Reinartz, W. (n.d.). *H i g h H i g h L o w Springer Texts in Business and Economics Customer Relationship Management Concept, Strategy, and Tools Third Edition*. http://www.springer.com/series/10099

Le, T. T. H., Oktian, Y. E., & Kim, H. (2022). XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems. *Sustainability (Switzerland)*, *14*(14). https://doi.org/10.3390/su14148707

Lemmens, A., & Gupta, S. (2020). Managing churn to maximize profits. *Marketing Science*, *39*(5), 956–973. https://doi.org/10.1287/mksc.2020.1229

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. In *Journal of Thoracic Oncology* • (Vol. 5, Issue 9).

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *33*(8), 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

Mehrabioun, M., & Mahdizadeh, B. M. (2021). Customer retention management: A complementary use of data mining and soft systems methodology. *Human Systems Management*, *40*(6), 897–916. https://doi.org/10.3233/HSM-201075

Neslin, S., Gupta, S., Kamakura, W., Lu, J., & Mason, C. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. *Journal of Marketing Research*, 204–211.

Nwanganga, Fred., & Chapple, Mike. (2020). *Practical Machine Learning in R*. John Wiley & Sons, Incorporated.

Rajendran, A. (2022, January 5). *Types of data in Machine Learning*. Towards AI.

Robert C Blattberg, Pyŏng-do Kim, & Scott A Neslin. (2008). *Database Marketing*. Springer.

Roschk, H., & Gelbrich, K. (2014). Identifying Appropriate Compensation Types for Service Failures: A Meta-Analytic and Experimental Analysis. *Journal of Service Research*, *17*(2), 195–211. https://doi.org/10.1177/1094670513507486

SAS Institute. (2017, August 30). *Introduction to SEMMA*. SAS Institute

Shabankareh, M. J., Shabankareh, M. A., Nazarian, A., Ranjbaran, A., & Seyyedamiri, N. (2022). A Stacking-Based Data Mining Solution to Customer Churn Prediction. *Journal of Relationship Marketing*, *21*(2), 124–147. https://doi.org/10.1080/15332667.2021.1889743

Tamaddoni, A., Stakhovych, S., & Ewing, M. (2017). The impact of personalised incentives on the profitability of customer retention campaigns. *Journal of Marketing Management*, *33*(5–6), 327–347. https://doi.org/10.1080/0267257X.2017.1295094

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1–9. https://doi.org/10.1016/j.simpat.2015.03.003

Verbeke, W., Baesens, B., & Bravo, C. (2018). *Profit-Driven Business Analytics*. https://onlinelibrary.wiley.com/doi/

Yang, Z., Xu, Q., Bao, S., Cao, X., & Huang, Q. (2021). *Learning with Multiclass AUC: Theory and Algorithms*. http://arxiv.org/abs/2107.13171

Zvornicanin, E. (2022, November 6). *Accuracy vs AUC in Machine Learning*. https://www.baeldung.com/cs/ml-accuracy-vs-auc

Appendices

Appendix A

Column	Type	Explanation
Client_id	Numerical	Customer identification
Created	Numerical - Timestamp	Date that account is created
Updated	Numerical – Timestamp	Last date that the customer was active
City	Categorical - Nominal	City the customer lives
Country	Categorical – Nominal	Country where the customer lives
Profile_register	Numerical - timestamp	Date that the customers' profile is made
Waiting_list	Categorical – Nominal (binary)	Customers have the opportunity to create a waiting list if classes are fully booked. They can choose 0 if they don't want a waiting list and 1 if they would like to make use of a waiting list.
Profile_lang	Categorical - Nominal	Language they use
Cancelled	Numerical - Timestamp	NULL if they did not churn and if the customer churned it shows the date.
Payment_frequency	Categorical - Ordinal	Month if they pay per month or year if they pay annually
Payment_reminders	Numerical	Amount of times they received a reminder to pay for their subscription.
Referral_id	Numerical	It is possible to refer a friend. This is the customer id from the customer that referred the 'new' customer
First_activated	Numerical – Timestamp	Date of account activated
First_payment	Numerical - Timestamp	Date of first payment received
Status	Categorical - Nominal	Active – this is an active member Deleted – this is a churned customer Suspended – Temporarily closed account
Pricing_plan_id	Categorical - Nominal	The pricing plan that is assigned to the customer
amount_of_lessons	Numerical	The number of lessons given by a customer in total over the entire period
max(created)	Numerical	When a class is created

Amount_of_teachers	Numerical	Amount of teachers on one customer account
Comment	Text	All manually entered comments regarding contact with customer

Table 8: Raw dataset

Appendix B

Column	**Type**	**Explanation**
Days before activated	Numerical	Amount of days between the moment the account was created and the first moment they used the service.
Days before payment	Numerical	Amount of days between account activation and first payment.
Days active	Numerical	Amount of days a customer has been active, looking at the last activity and the date the account was created.
Pt_lessons	Numerical	The sum of the number of lessons over the three-month period
P1_lessons	Numerical	The sum of the number of lessons over the first month
P2_lessons	Numerical	The sum of the number of lessons over the second month
P3_lessons	Numerical	The sum of the number of lessons over the third month
Diff_lessons_p1_p2	Numerical	The difference of the number of lessons between the first month and second month
Diff_lessons_p2_p3	Numerical	The difference of the number of lessons between the second month and the third month
Diff_lessons_p1_p3	Numerical	The difference of the number of lessons between the first month and the third month
Pt_registrations	Numerical	The sum of the number of registrations over the three-month period
P1_registrations	Numerical	The sum of the number of registrations over the first month
P2_registrations	Numerical	The sum of the number of registrations over the second month
P3_registrations	Numerical	The sum of the number of registrations over the third month

Diff_registrations_p1_p2	Numerical	The difference of the number of registrations between the first month and the second month
Diff_registrations_p2_p3	Numerical	The difference of the number of registrations between the second month and the third month
Diff_registrations_p1_p3	Numerical	The difference of the number of registrations between the first month and the third month
Pt_income	Numerical	The total income of the three-month period
P1_income	Numerical	The total income of the first month
P2_income	Numerical	The total income of the second month
P3_income	Numerical	The total income of the second month
Diff_income_p1_p2	Numerical	The difference in income between the first month and second month
Diff_income_p2_p3	Numerical	The difference in income between the second and third month
Diff_income_p1_p3	Numerical	The difference in income between the first and third month

Table 9: Dataset with created features

Appendix C

1. Studio name (open)
2. Studio email address (Open)
3. Did Covid-19 or the lockdowns in your country contribute to you deciding not to continue with Momoyoga?

   * yes

   * No
4. Country (open)
5. Please share your reason for not moving forward with Momoyoga.

   * No longer teaching

   * I chose a different (software) solution

   * Needed more support

   * Technical issues

   * Too expensive

   * Yogis don't enjoy using Momoyoga

   * Wanted additional features

   * Other
6. If you selected ''I chose a different software solution'', would you like to list which one here? (open)
7. If you selected ''other'', would you like to share your reason here?
8. Do you have any recommendations for features you would like us to include in future releases?

   * Admin-related features

   * App-related features

   * Calendar integrations

   * Class-related features

   * Custom emails

   * Custom registration form

   * Custom time zones

   * Design-related features

   * More languages

   * Marketing-related features

   * Notifications

   * Other integrations (FB/IG/One Pass)

   * Payment options

   * Schedule-related features

* Website integrations
* Other

9. Would you like to share specific features with us? Feedback from you will help drive software improvements. (open)
10. If the specific features you listed above are added to Momoyoga, would you like to be notified by email?
    * Yes
    * No
11. Any last additional comments on how we can better serve future customers. (open)

Appendix D

Features	Correlation ‘’churn’’	Correlation ‘’churn reason’’
Client_id	0.054	0.086
Churn	1.0	0.64
Pt_lessons	0.17	0.097
P1_lessons	0.116	0.055
P2_lessons	0.152	0.089
P3_lessons	0.223	0.134
Diff_lessons_p1_p2	0.095	0.088
Diff_lessons_p2_p3	0.172	0.107
Diff_lessons_p1_p3	0.184	0.134
Pt_registrations	0.147	0.099
P1_registrations	0.097	0.064
P2_registrations	0.146	0.102
P3_registrations	0.209	0.140
Diff_registrations_p1_p2	0.045	0.044
Diff_registrations_p2_p3	0.126	0.070
Diff_registrations_p1_p3	0.096	0.067
Pt_income	0.018	0.014
P1_income	0.009	0.010
P2_income	0.013	0.012
P3_income	0.031	0.020
Diff_income_p1_p2	0.022	0.013
Diff_income_p2_p3	0.024	0.008
Diff_income_p1_p3	0.040	0.017
Profile_register	0.045	0.026
Waiting_list	0.128	0.067
Payment_reminders	0.042	0.034
Referral_id	0.050	0.102
Credits	0.044	0.029
Active_campaign_id	0.046	0.017
Pricing_plan_id	0.020	0.011
Amount_of_lessons	0.212	0.115
Amount_of_teachers	0.100	0.068
Referred	0.008	0.004

Days_before_activated	0.030	0.027
Days_before_payment	0.019	0.015
Days_active	0.130	0.018

Table 10: Correlations of dependent (Y) variable "churn" and "churn reason"

Appendix E

	Predicted: churn	Predicted: no churn
Actual: churn	TP: 173	FN: 78
Actual: no churn	FP: 78	TN: 323

Table 11: Confusion matrix DT

	Predicted: churn	Predicted: no churn
Actual: churn	TP: 170	FN: 81
Actual: no churn	FP: 98	TN: 303

Table 12: Confusion matrix SVM

	Predicted: churn	Predicted: no churn
Actual: churn	TP: 169	FN: 82
Actual: no churn	FP: 87	TN: 314

Table 13: Confustion matrix LR

	Predicted: churn	Predicted: no churn
Actual: churn	TP: 161	FN: 90
Actual: no churn	FP: 80	TN: 321

Table 14: Confusion matrix KNN

Appendix F

This appendix includes the multi-class prediction output evaluation metrics if the non-churners and NaN category were ignored. One can argue that this would give a more realistic result, but due to the small amount of data, it is chosen to work further with a dataset where the NaN category is involved. Also, that model would eliminate the first step of the binary classification as it also predicts if someone is not going to churn

Evaluation metric	**Output multiclass XGBoost**
Accuracy	0.43
Macro averaged precision : calculate precision for all classes individually and then average them	0.38
Micro averaged precision: calculate class wise true positive and false positive and then use that to calculate overall precision	0.43
Macro averaged recall	0.39
Micro averaged recall	0.43
Macro averaged F1 score	0.37
Micro averaged F1 score	0.5
Averaged AUC score	0.66

Confusion matrix:

		Predicted			
		Covid	**No longer teaching**	**Price-related**	**Service-related**
Actual	**Covid**	4	3	7	5
	No longer teaching	14	31	10	13
	Price-related	3	6	8	4
	Service-related	6	9	6	21

Table 15: Confusion matrix multi-class (ignore NaN)