MASTER

A prescriptive system for strategic resource allocation to reduce customer churn

Pladdet, F.M.B.

Award date:
2019

Faculty of Industrial Engineering and Innovation Sciences

A Prescriptive System for Strategic Resource Allocation to reduce Customer Churn

F.M.B. Pladdet BSc., 1017392

First supervisor: Dr. R.M. Dijkman, TU/e Information Systems
Second supervisor: Dr. A.M. Wilbik, TU/e Information Systems
External supervisor: S. Broersen PhD, Deloitte Analytics and Information Management

Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Operations Management and Logistics
of Eindhoven University of Technology, April 2019

TU/e, School of Industrial Engineering
Series Master Theses Operations Management and Logistics

Subject headings:
Prescriptive Analytics, Resource Allocation, Customer Lifetime Value, Churn Prediction

Abstract

This research presents a prescriptive analytics study focused on optimal resource allocation in a sales context. Sales agents visit customers in order to prevent customers churn. This study proposes a system which selects which customers to visit in a non-subscriptual customer-company relation. A system is developed which uses predictive modelling to forecast a customer's churn probability, the effect of a visit on that churn probability and the expected future value of the customer. The system proposed in this study prioritizes for visits that lead to the highest increase in the sum of the expected future value from all current customers, customer lifetime equity (CLE). The result of a case study shows that compared to a baseline without visits, the CLE can be increased by 3.2 percent when using the system. The actual visits, as done by the case company, resulted in an increase of the CLE by 0.6 percent when compared to the baseline without visits. This indicates that value can be generated by using the prescriptive system. Furthermore, the result of the system was achieved with 78 percent less visits when compared to the actual number of visits, showing that the system also increases efficiency. This study shows that prescriptive resource allocation can be beneficial in a non-subscriptual context and increases efficiency.

Executive summary

There are multiple decisions that have to be made within a sales department. All these decisions can have direct influence on the chance of making a sale, the height of the revenue and keeping or losing a client. Knowing how to make these decisions in an optimal way can be beneficial for a company and create a competitive advantage. Where these decisions previously had to be made by humans, the introduction of big data changed this decision process. With the use of prescriptive analytics, optimal decisions can be made based on historical data.

This research proposes a prescriptive system which prescribes what short term actions sales agents need to undertake to prevent customers from leaving. The system maximizes the value of existing customers by selecting the actions which generate the highest value based on the effect of the action and the potential value of the customer. The system uses a churn model per customer cluster and predicts all customers' churn probability with and without a visit by a salesagent. The action effect is combined with the future value of all individual customers resulting in a list of optimal visits as shown in figure 1. The list resulting from the system prescribes which customers need to be visited in the coming month to maximize future value.

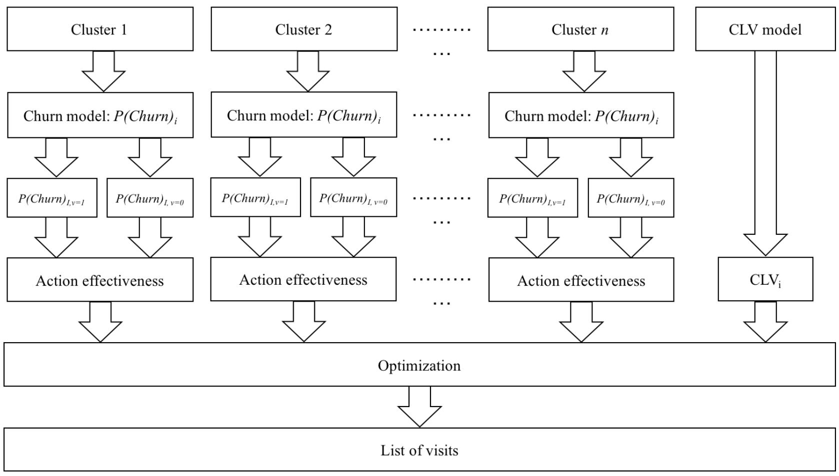


Figure 1: Prescriptive system

The system was evaluated using a case study with the customer data of the small scale outlets in the Netherlands. This showed that by using the visits as proposed by the system, the sum of all future customer value (CLE) can be increased by 3.2 percent when compared to the situation without visits.

Compared to the results of the actual visits as done in the tested period the CLE was increased by 2.6 percent while decreasing the number of visits by 78 percent.

Table 1: Evaluation of the prescriptive system

Situation	CLE	Number of Visits	Return of visits	Increase
Without visits	€74,663,533	0	-	-
Using actual visits	€75,124,699	85	€437,694	0.6%
Using prescriptive system	€77,045,032	19	€2,381,498	3.2%

The system can directly be used as a support system which identifies what customers to visit in the coming month. New customers can be assigned to the cluster most applicable, and use the models to predict the churn probability. The customer can then be included in the visit selection.

Figure 2 shows a possible setup of the system in the visits selection process. The prescriptive system serves as a support system which identifies beneficial visits and is combined with expert knowledge from sales agents. The customers which are selected by both the prescriptive system and the sales agents can have a high priority and the customers selected by only the sales agent or only by the prescriptive selection can have a medium priority.

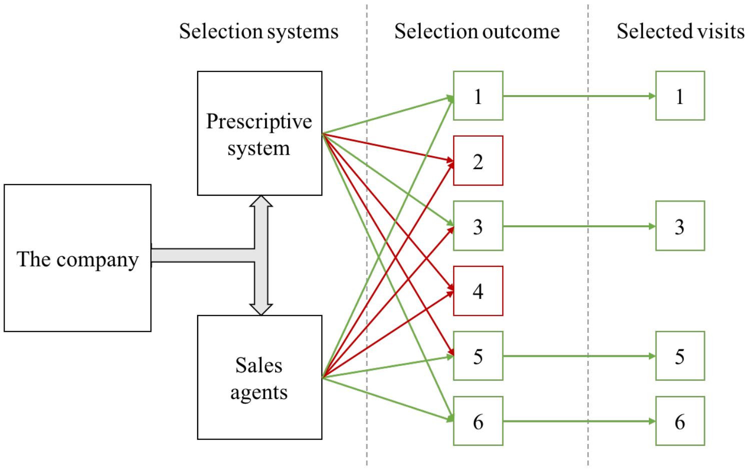


Figure 2: Application of the system

Since the prescriptive system is a system with static clusters and pretrained models, it is advisable to monitor the performance of the churn model for the newly predicted month. The performance of the churn model will change due to customers churning and the acquisition of new customer, because this will reduce the relevance of the used trainset. Therefore, the churn model should revaluated regularly.

A limitation of the system is that the churn model is able to correctly classify 57 percent of all churn observations and 38 percent of the observations classified as churn is an actual churn observations. It is therefore not directly usable as full replacement for manual visit planning, but can serve as a support system. Improving the model could make this possible in the future.

Acknowledgements

Before you lies my master thesis, the result of seven months of hard work. In September 2018, I started my thesis internship at Deloitte's Analytics and Information Management department which became my first contact with a working life. It made me realise that I'm ready to make the jump and I'm looking forward to bring the things I have learned into practice. This project is my last step as a student and is made possible with the help of people I would like to thank.

First I want to thank my mentor Remco Dijkman for his help and patience throughout the project. I enjoyed our meetings and really appreciate the time you have put into mentoring and pushing me to get my thesis to the highest level. I also want to thank Anna Wilbik for her thoughts and feedback on the technical parts of my thesis.

Secondly, I want to thank my Deloitte mentor Sjors Broersen for his feedback, suggestion and linking me to the right people. Other Deloitte colleagues I want to thank are Thomas Breeuwsma for being my thesis buddy, Michiel Blomaard for the many enjoyable coffee breaks during the week and all other colleagues that made me feel at home at Deloitte.

Thirdly, I want to thank my girlfriend Jolijne, for her never ending support and helping me push through the hard moments. I also want to thank my parents for supporting me throughout my time as a student and allowing me to choose my own path.

I enjoyed working on this project with the help of inspiring people from the university and Deloitte. I look back on a valuable experience and I am proud of the result. I wish you much pleasure reading my thesis.

Mathijs Pladdet
April 7, 2019

‘Without big data analytics, companies are blind and deaf, wandering around like a deer on a freeway.’

Geoffrey Moore

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter introduces the subject and motivation behind this study. This is followed by an explanation of the research goals and the research design of this thesis.

1.1 Motivation

Almost all companies rely on some sort of sales to attract revenue. In the business to business sales, this sales process is usually centered around sales agents who manage accounts and visit business clients. There are multiple decisions that have to be made within a sales department. What client to contact, how to contact a client, when to contact a client, who to assign to what client and what to offer a client. All these decisions can have direct influence on the chance of making a sale, the height of the revenue and keeping or losing a client. Knowing how to make these decisions in an optimal way can be beneficial for a company and create a competitive advantage.

Where these decisions previously had to be made based on domain knowledge, on human intuition or on a static intervals, the introduction of big data changed this decision process. With the use of big data analytics, decisions can be made based on historical data. The big data analytics can be categorized in three tiers, the descriptive, predictive and prescriptive tier. Descriptive analytics aims to describe what happened in the past, predictive analytics tries to predict what will happen in the future and prescriptive analytics aims to optimize decision-making based on predictions. The potential for sales processes lies in this last tier, where a system prescribes what decisions to make.

The goal of this research is to create such a prescriptive system which prescribes what short term actions sales agents need to undertake to prevent customers from leaving. The system will maximize the value of existing customers by selecting the actions which generate the highest value based on the effect of the action and the potential value of the customer. More specific, the churn probability of a specific customer and lifetime value of this customer will be predicted in order to calculate the expected value that will be lost when a customer leaves. This will be complemented by a forecast of the effect of sales agent actions on reducing churn. These three elements are used to optimize the actions of the sales agents in a given period to maximize customer value of existing customers.

Where other literature focuses on using prescriptive analytics for keeping subscription customers, will this study focus on using prescriptive analytics for resource allocation in a non-subscriptual B2B setting.

1.2 State of the art

Previous research on prescriptive analytics in sales processes has focused on optimal assignment of sales agent to new opportunities (Baier et al., 2012), optimizing the size of teams for the current new opportunities (Kawas, Squillante, Subramanian, & Varshney, 2013) and on creating the optimal sales teams (Von Bischhoffshausen, Paatsch, Reuter, Satzger, & Fromm, 2015). The aim of these researches was to optimize for a maximized chance of a sale coming from a new opportunity.

However, as Fü, Matzler, and Faullant (2006) state, attracting a new client can be five to six times as expensive as keeping a current client. This shows the relevance of not only looking at new opportunities but also to look at keeping current customers. The chance that a customer will leave in the next period is called the churn rate or churn probability. Predicting churn has been an active subject in the academic world (Rubinstein, 1985; Peppers & Rogers, 1995; Zhao, Li, Li, Liu, & Ren, 2005; Khodabandehlou & Zivari Rahman, 2017). These predictions are used to decide which customer to prioritize and to gain insight in the risk of loosing customers.

In the last two decades a shift can be seen in what methods are used to predict churn probability. Where Murtaugh, Burns, and Schuster (1999) and Bolton (1998) predict the churn by using statistical analysis like the Kaplan-Meier survival estimate or Cox proportional hazard regression, more recent literature focuses on using data-mining techniques. These are techniques like artificial neural networks (Buckinx & Van Den Poel, 2005), decision trees (Coussement & De Bock, 2013), support vector machines (Zhao et al., 2005) and k-nearest neighbor (Keramati et al., 2014). Khodabandehlou and Zivari Rahman (2017) compare these different machine learning techniques and conclude that artificial neural network has the highest accuracy for churn prediction and show the relevance of using boosting methods for further improvement of churn prediction models.

Other than the shift towards data mining techniques, the churn prediction field of study does not shift its focus towards a different setting. The research remains mostly focused on churn prediction in a subscription setting like insurance companies (Hendrikse, 2017; Huigevoort, 2015; Srigopal, 2018) or other companies providing a service (Gür Ali & Aritürk, 2014; N. Lu, Lin, Lu, & Zhang, 2014). One of the reasons for this is that this a setting with a clearly defined churn moment. However by mostly focusing on this setting, non-subscriptiual settings like B2B retail settings without contracts are ignored.

The step towards using this churn prediction in a prescriptive way, especially in the B2B non-subscriptual setting, is still in an immature phase. There has been some research on this topic like the research of E. Lee et al. (2018) or the research of Srigopal (2018), but more progress can be made. Therefore, this research will focus on the use of prescriptive analytics for keeping current clients in a B2B non-subscriptual setting.

1.3 Research goal

As described in the previous sections, the goal of this research to create a prescriptive resource allocation system which maximizes future value by predicting which actions most increase the probability of a high-profit customer staying as a customer. The research goal is formulated as follows:

The goal of this research is to create a prescriptive resource allocation system for sales agents which indicates what actions to undertake in order to maximize the total future value of existing customers in a B2B non-contractual setting.

A prescriptive resource allocation system consists of both forecast models and an optimization model. The optimization model needs input about the generated value of actions to be able to determine if an action is most beneficial for the total expected value. The expected value of a customer is calculated using two parts as shown in equation 1.1. The first part of the right hand side of the equation shows the probability that a customer will remain customer including the effect of an action. The second part of the right hand side is the future value of this customer.

$$Expected\ Value = [1 - Churn\ Probability + Action\ Effect] \cdot Customer\ Lifetime\ Value \tag{1.1}$$

The action effect has a positive effect on the expected value and by predicting which actions generate the most value, the optimal actions can be selected to maximize expected future value. Based on equation 1.1, the four subgoals are formulated to create the resource allocation system. Subgoal one, two and three are forecast models and serve as input for the optimization model as formulated in subgoal four.

For determining the expected loss in value and the value of decreasing the churn chance for a specific customer the following subgoals are formulated:

1. Model customer churn probability (CR) in a B2B non-contractual setting
2. Model customer lifetime value (CLV) in a B2B non-contractual setting

For determining the effectiveness of specific actions on the churn probability of a specific customer the following subgoal is formulated:

3. Model effect of salesagent actions (AE) on customer churn probability

The optimization model should lead to the optimal trade-off between costs of specific actions and the benefit of these actions for the total future value of all customers. For creating this optimization model which determines what actions to execute in order to maximize future value over all customers, the following subgoal is formulated:

4. Create resource allocation model (RA) with sales agent actions as decision variables

1.4 Research Design

To achieve the research goal, all subgoals need to be studied after which a method is chosen and applied. All subgoals follow the same procedure which starts with a literature search resulting in an overview of used methods, followed by a decision on which model to use and ends with a modelling phase. The prescriptive system will be formed by joining these elements as illustrated in figure 1.1

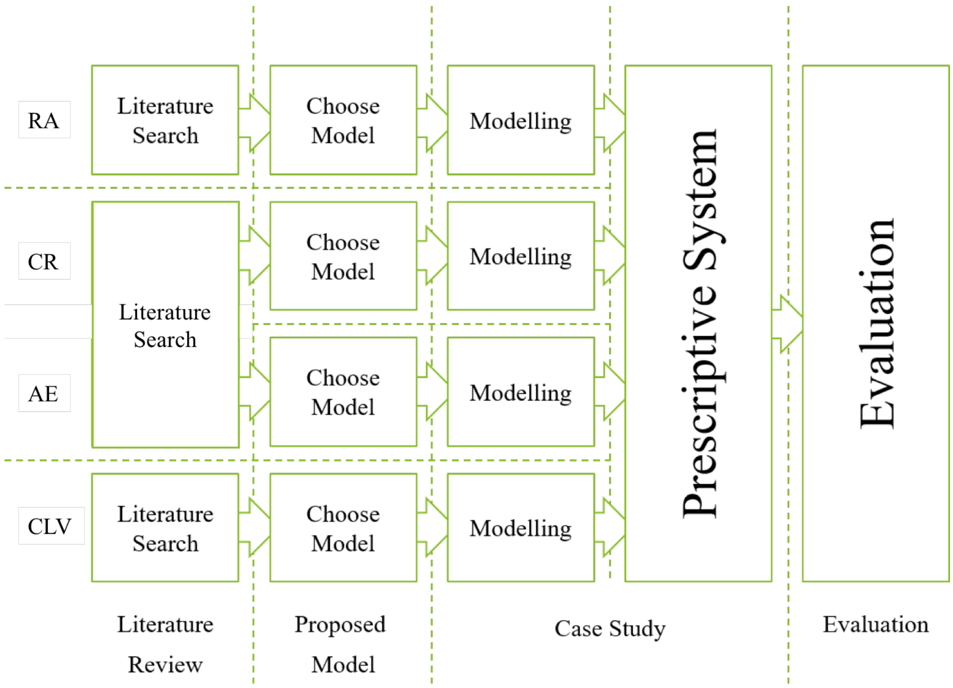


Figure 1.1: Research phases

The literature review will be based on recent academic publications and form an overview of methods that are used for for optimizing resource allocation and methods for calculating CR and AE and CLV. These methods will be evaluated on usefulness for a B2B non-contractual setting and the available data. The chosen methods from the literature review will be used to formulate the proposed system.

The proposed system will be evaluated using a case study. For this the data of a producer of chemicals will be used, which sells their products in a non-subscriptual B2B relation. The provided data is based on the availability of the data and what the company was legally and for competitive reasons allowed to share. The different variables are selected in consultation with company experts and their expectation of churn probability affecting variables.

The case study focuses on the Small Scale Outlet (SSO) customer group, which consists of 1500 business to business clients like small stores with a total of 181,000 orders. The scope is limited to this group, since these shops have higher responsiveness to actions than larger hardware stores according to company experts and are visited by sales agents throughout the year. Furthermore, the scope is limited to the Dutch market because the data of the Dutch market is considered by company experts to be reliable and complete.

The company provided the following data:

- Daily transaction data per customer per order per brand for all Small Scale Outlets (SSO) in the Netherlands from 2015-01 to 2018-11
  - Customer code, Customer name, Brand, Order code, Date, Quantity in liter, Revenue, Discounts given, Contribution margin
- Complete Customer Relationship Management (CRM) data for all SSOs
  - Customer code, Customer name, Visits, Demographics, Assigned sales agent

For the case study the Cross-Industry Standard Process for Data Mining cycle will be used as shown in figure 1.2. This CRISP-DM cycle is used in multiple studies (Caetano, Cortez, & Laureano, 2015; Sen & Ucar, 2012; Moro & Laureano, 2011) for case studies and sets a framework for applying data mining to projects. It consists of the phases Business Understanding, Data Understanding, Data preparation, Modelling, Evaluation and Deployment. The phases are discussed in chapter four, five and six.

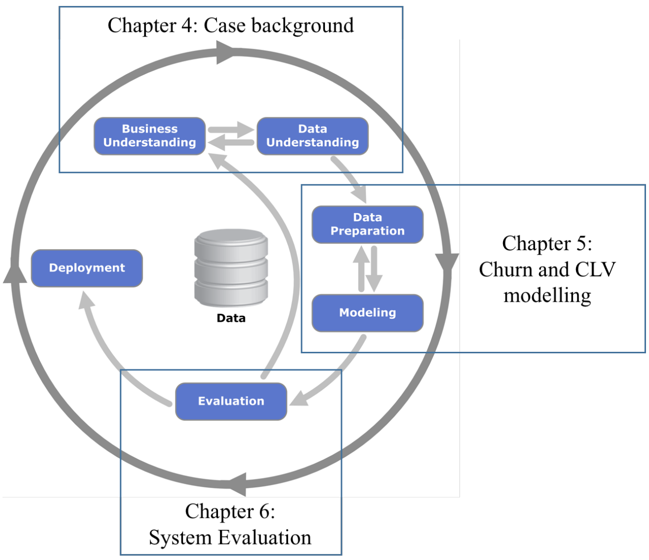


Figure 1.2: CRISP-DM Cycle

The evaluation of the model is performed by calculating the expected future value of the customers in the dataset when sales agents assign the actual amount of resources to specific actions for specific customers. This is then compared with the expected future value when using the actions of the prescriptive system. A higher expected future value when using the prescriptive system indicates that an advantage can be achieved when using the proposed system.

Note that the deployment phase of the CRISP-DM cycle is not explicitly discussed. This report will be the first step towards deployment.

The next chapters will follow the structure as shown in figure 1.1 and figure 1.2. First the relevant literature will be discussed to determine usable methods. This is followed by proposition for a prescriptive resource allocation system, after which this system is evaluated using a case study with the CRISP-DM cycle. This thesis will end with a conclusion, applications and a thorough discussion of the results.

Chapter 2

Literature Review

The aim of this chapter is to provide a literature review on the relevant subjects as explained in chapter 1.4. The methods from this chapter will be used as input for the proposed model in chapter 3. First an overview of the status quo of prescriptive analytics will be given followed by an overview of prescriptive analytics in sales processes and methods for churn calculation. This chapter will end with current methods for customer lifetime value calculation. For the section 2.1 and 2.2 a review protocol is used, which can be found in appendix A. Section 2.3 and 2.4 are based on literature reviews of Khodabandehlou and Zivari Rahman (2017) and Esmaeiligookeh, Tarokh, and Tarokh (2013) respectively.

2.1 Prescriptive Analytics

This section describes the status quo of prescriptive analytics and starts by introducing the concept. This is followed by how prescriptive analytics can be applied and the different methods for prescriptive analytics. This section ends with a conclusion on the states quo of prescriptive analytics.

2.1.1 General background

Prescriptive analytics is considered to be the third and last tier of business analytics (Evans & Lindner, 2012; Soltanpoor & Sellis, 2016; Pladdet, 2018). It uses different techniques to determine what actions to undertake to achieve an optimal outcome. There are two other tiers of business analytics which are descriptive analytics and predictive analytics. Figure 2.1 shows all tiers of business analytics and the underlying question they tries to answer.

Every tier uses the insights of the previous tier and adds more value by generating new insights. Descriptive analytics focuses on determining what happened in the past until present time. This can be done in the form of dashboards and KPIs and gives insight in how a company is doing. Predictive analytics focuses on the future and uses historical data to forecast what will happen in the future. Prescriptive analytics in general consists of a predictive part in the form of forecast models and an optimization part

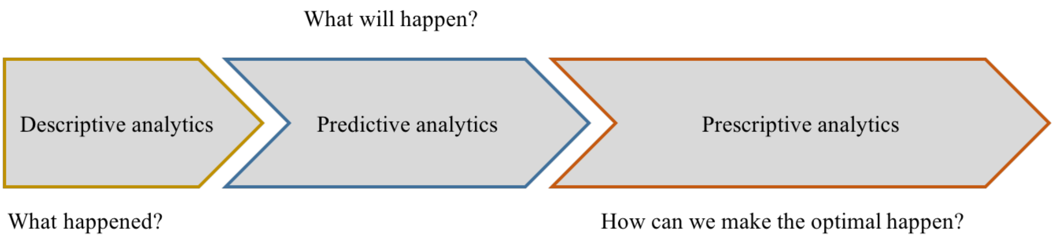


Figure 2.1: Three tiers of data analytics (Evans & Lindner, 2012)

in the form of optimization models. The forecast models predict certain variables which are used as input in the optimization model and the optimization model prescribes a set of actions to undertake which will lead to an optimal situation. This can result in a decision support system, but also in an autonomous decision system.

2.1.2 Application of Prescriptive Analytics

Prescriptive analytics is applied in a variety of industries and for a variety of goals. Figure 2.2, shows the distribution in the found literature resulting from the used review protocol. It can be seen that prescriptive analytics is mostly used in the production industry and mostly for job scheduling and control. A reason for this could be that there are multiple parties and processes involved when producing goods. This leads to more facets that can be optimized and more costs that can be reduced from implementing prescriptive analytics.

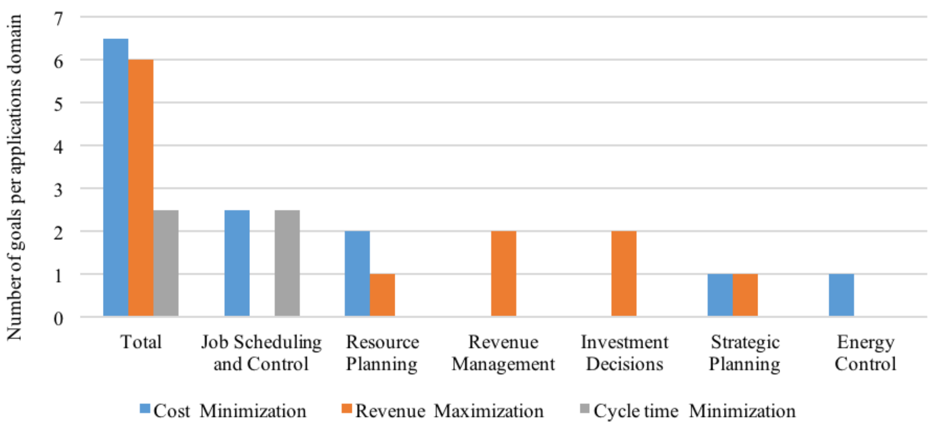


Figure 2.2: Applications of prescriptive analytics

Furthermore, the literature shows that prescriptive analytics is used for six different purposes. Job scheduling and control (Y. Liu et al., 2010; Souza, 2014; Lombardi, Milano, & Bartolini, 2017; Kuo & Lin, 2010; Krumeich, Werth, & Loos, 2016), resource planning (Souza, 2014; Ramezani, Lu, & Hussain, 2013; Von Bischhoffshausen et al., 2015), revenue management (J. Lee, Shi, Wang, Lee, & Kim, 2016; Souza, 2014), investment decisions (Guedes, Pinto, Vale, Sousa, & Sousa, 2013; Lash & Zhao, 2016), strategic planning (Y. Liu et al., 2010; Souza, 2014) and energy control (Zavala, Constantinescu, Krause, & Anitescu, 2009). In these different purposes there are three main goals for which prescriptive analytics is used. These are cost minimization, revenue maximization and cycle time minimization.

2.1.3 Methods for Prescriptive Analytics

The methods for prescriptive analytics consist of a forecast method and an optimization method. These methods depend on the goal and setting in which the prescriptive analytics is used. Table 2.1 shows the combinations of forecast method and optimization method as found in the literature resulting from the review protocol. What can be seen is that neural networks are very commonly used as forecast method and linear programming is a widely used method for optimization.

Table 2.1: Methods for prescriptive analytics

Forecast method:	Optimization method:								
	SO	LP	MoP	PSO	IP	MIP	LS	CP	SMT
Linear Regression		X							X
Logistic Regression		X							
Regularized Regression		X							
Neural Network		X	X	X	X	X	X	X	
Support Vector Machine		X							
Support Vector Regression		X							
Decision Tree								X	X
Gaussian Processes	X	X							
Exponential Smoothing		X							

SO: Stochastic Optimization, LP: Linear Programming, MoP: Multi-objective Programming, PSO: Particle Swarm Optimization, IP: Integer Programming, MIP: Mixed Integer Programming, LS: Local Search, CP: Constraint Programming, SMT: Satisfiability Modulo Theories

The general approach is using formulas with variables that affect the objective function and constraints. The outcome of interest in the optimization for prescriptive analytics is not the optimal value of the objective function $f(x)$, but the values of the decision variable x (Pladdet, 2018). Equation 2.1 shows the formulation of the objective function to extract the optimal outcome of the decision variables.

$$\arg\max_x f(x) \ \ or \ \arg\min_x f(x) \tag{2.1}$$

The optimization model can include multiple constraints which limit the possible values for the decision variables and enable the calculation of an optimal outcome. How these constraints are formulated depends on the optimization method.

2.1.4 Conclusion

Prescriptive analytics can be valuable in a variety of applications. It combines forecasts with optimization techniques to be able to pro-actively manage processes and companies. Both revenue maximization and cost minimization are important goals when applying prescriptive analytics. Based on the number of relevant articles found about prescriptive analytics, it can be stated that the research on prescriptive analytics is still in an immature stage. According to domain experts, companies are still mostly focused on descriptive analytics and some also on predictive analytics. However, prescriptive analytics is were big data can be of its most value and therefore an increase in articles on this subject can be expected. This study will therefore contribute by adding extra research about prescriptive analytics and focus on resource allocation in sales processes.

2.2 Prescriptive analytics in Sales Processes

This section will further elaborate on the use of prescriptive analytics in sales processes. It will start by providing general background and is followed by different prescriptive analytics studies in a sales process context. It will end with a conclusion on the use of prescriptive analytics in sales processes.

2.2.1 General background

As stated in the introduction, there are multiple studies about allocating the optimal amount of resources in a sales context. The focus in these papers is mostly on attracting new customer where the sales agent is assigned to a customers who is most likely to make the sale. Srigopal (2018) describes the use of prescriptive analytics for current customers. The used optimization methods vary between the different papers. Well known methods for resource allocation optimization are integer programming, mixed integer programming, linear programming and multi objective programming.

2.2.2 Optimal allocation of sales teams to customers with Integer programming

Von Bischhoffshausen et al. (2015) use integer programming to formulate the optimal sales teams and allocate these to customers which maximizes revenue. The benefit of this method is that the optimization only takes integer values into account. The disadvantage is that it is therefore also less flexible. A sales-response function was calculated, which predicts the probability of a sales for every customer-sales team combination. The authors created a decision variable for each sales agent-team combination and a decision variable for every team-customer account combination. These decision variables are binary and take the value of 1 when the salesagent is in a specific team and 0 if a team is assigned to a specific customer account. Von Bischhoffshausen et al. (2015) include constraints which assure that a team can only be assigned to a customer account which it is capable to, that only one team is assigned to one customer account, that the maximum workload is not exceeded and that the maximum travel time of the sales agent is not exceeded.

2.2.3 Optimal allocation of salesagents per team with Mixed integer programming

Kawas et al. (2013) describe the use of mixed integer programming for prescriptive resource allocation of the number of sales agents for different categories of sellers to be able to assign the most effective teams. Mixed-integer programming is a variant of integer programming which allows for the combination of integer and continuous decision variables. The benefit of using mixed-integer programming is that it is more flexible than integer programming and it allows for resources to be scheduled partially. This can be required in situations where one resource is allocated to multiple tasks. A disadvantage of using mixed integer programming is that it is computationally demanding and can become very complex (Pladdet, 2018).

The goal of Kawas et al. (2013) is to identify how much agents must be assigned to these different seller categories to maximum revenue. The authors use a sales response function to calculate the estimated revenue potential for combinations which they then use to create an objective function with all these response decision. Kawas et al. (2013) use constraints for the maximum amount of hires and fires and cost to model a more realistic situation.

2.2.4 Resource allocation with Linear programming

Vanderbei (2014) describes the use of linear programming for resource allocation problems. Linear programming is an optimization method which can be applied in a wide range optimization problems (Pladdet, 2018). It can be solved fast and uses linear equations to describe the relations between variables. The disadvantage of this method is that it does not allow non-linear relations and can become a simplified version of the real situation. Vanderbei (2014) combine the costs of the resources with the returns of the resources when used for different application to create an optimal allocation. The returns and costs are directly incorporated in the objective function. The constraints are the amount of resources that are available and possibly the minimum amount of resources that need to be used (Vanderbei, 2014). The latter is the case in the resource allocation for sales processes, because a fixed number of resources is available and agents are not fired or hired.

2.2.5 Resource allocation with Multi objective programming

Ramezani et al. (2013) use multi objective programming to optimize resource allocation in order to increase speed and reduce cost. Multi objective programming is an optimization method which can be used when multiple objectives need to be optimized simultaneously (Lotfi, Hatami-Marbini, Agrell, Aghayi, & Gholami, 2012). The advantage is that a balance can be found between multiple objectives. This can however lead to sub optimal solutions for both objectives. In sales processes, the main driver is revenue, but it is possible to optimize resources for both revenue maximization and cost minimization.

2.2.6 Optimal resource allocation of retention investments

Srigopal (2018) describes the use of prescriptive analytics for allocating resources to different retention actions. This is an example of prescriptive analytics which is focused on existing customers instead of new customers. Srigopal (2018) uses a combination of survival analysis and ranking to determine what resources should be allocated to retention investments. The research uses customer data of an insurance company in a B2B subscription setting. By predicting when a customer will end its subscription and the effect of different actions. An optimal allocation can be determined.

2.2.7 Conclusion

Prescriptive analytics shows beneficial in sales context. The papers of Von Bischhoffshausen et al. (2015) and Kawas et al. (2013) focus on how to allocate resources in order to maximize revenue for new customers. Their aim is to find the combination of sales agent and customer that is most likely to lead to a sale. Vanderbei (2014), Ramezani et al. (2013) and Lotfi et al. (2012) describe the use of prescriptive analytics in a general resource allocation setting, where the minimum amount of resources are used to reach the optimal result. Srigopal (2018) describes the use of prescriptive analytics for existing customers in a subscription setting. This research will also focus on the use of prescriptive analytics for existing customers, but where Srigopal (2018) focused on customers in a subscription setting will this research focus on customers in a non-subscriptual setting. Figure 2.3 shows how the research is positioned in the current literature compared to other literature.

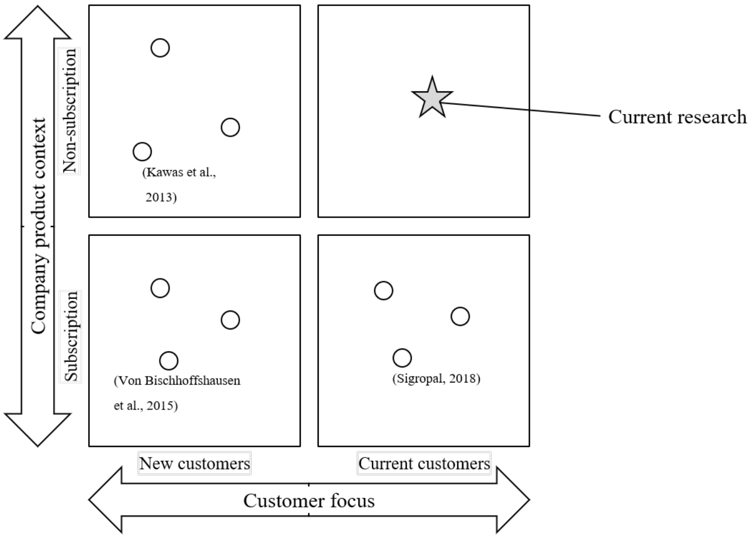


Figure 2.3: Prescriptive analytics research for different customer focus and company product context

The resource allocation optimization problem in this study does not allow for partial resource allocation and can therefore best be modelled using integer programming as described by Von Bischhoffshausen et al. (2015).

2.3 Churn probability calculation

This section provides an overview of the current research on churn probability. It starts by explaining the concept of churn, which is followed by methods for churn probability calculation. This section ends with an conclusion about the usefulness of the different methods in this research.

2.3.1 General background

Churn rate describes the probability that a customer will defect in the next period and is the opposite of the retention rate. This defection can have different forms like customers leaving a company (Huang et al., 2010), customers switching to a competitor (Huang, Kechadi, & Buckley, 2012; N. Lu et al., 2014; Hung, Yen, & Wang, 2006), no use of the service in a given period (Chen, 2016; Coussement & De Bock, 2013), customer changes purchase patterns (Buckinx & Van Den Poel, 2005), no renewal of a product (Coussement & Poel, 2009) and customer with decreasing CLV over time (Glady, Baesens, & Croux, 2009). These reasons can be voluntarily when a customers chooses to churn by itself or involuntarily when a company or other entity decides to churn the customer. This happens for instance when a customer stops paying and the company discontinues a contract.

Churn calculation is used to determine the risk of loosing value and to decide how to allocate resources. For the entire customer base, this can be calculated by using equation 2.2, where inclusion of customers in the customer base is based on the churn definition. By using segmentation, a more specific churn rate can be calculated. However for a customer specific churn rate more advanced techniques need to be used.

$$Churnrate = \frac{customerbase_{t=0} - customerBase_{t=1}}{customerBase_{t=0}} \cdot 100\% \tag{2.2}$$

2.3.2 Methods for Churn rate calculation

Khodabandehlou and Zivari Rahman (2017) present a comparison of different supervised machine learning techniques for customer churn prediction. Table 2.2 shows the methods used in the literature included in this research. From this overview can be concluded that Logistic regression (LR), Support Vector Machine (SVM), Classification and Regression Trees (CART) and Artificial Neural Network (ANN) are dominant methods for churn prediction. In addition, Srigopal (2018) describes the use Cox proportional hazard model in recent literature. The benefits and relevance of the four dominant methods from Khodabandehlou and Zivari Rahman (2017) and Cox proportional hazard model will be further explained.

Logistic Regression

Logistic regression is a popular method for classification and is a variant of linear regression. The benefits of this method are that it uses input variables which effects are directly interpretable from the equation and has an output ranging from 0 to 1 (Pladdet, 2018). The latter is the reason why logistic regression is ideal in binary situations.

Table 2.2 shows the relevance of logistic regression in churn rate calculation. Perlich, Provost, and Simonoff (2004) state that logistic regression is especially relevant for predictions with small training sets, where it provides higher accuracy than methods like decisions trees. The reason for this is, as King and Zeng (2001) state, that the accuracy of logistic regression is sensitive to event rarity. Since customer churn is usually a rare event, logistic regression can be useful for churn prediction in small data sets (Gür Ali & Aritürk, 2014).

Table 2.2: Churn rate methods as discussed in literature of Khodabandehlou and Zivari Rahman (2017)

	Association Mining	SEM	Survival Analysis	LR	ANN	Random Forest	CART	SVM	AdaBoost	K-means	NB	LC	KNN
Ali & Ariturk (2014)			X	X									
Bucking and vd Poel (2004)				X	X(ARD)	X							
Coussement and de Bock (2013)						X	X						
Coussement and Poel (2009)				X		X		X					
Crone et al (2006)					X(MLP)		X	X					
Devi Prasad and Madhavi (2012)							X						
Glady (2009)				X	X(MLP)		X		X				
Guo-en and Weidong (2009)				X	X(BP)		X	X			X		
Hosseini et al (2010)										X			
Huang et al (2010)					X(MLP)		X	X					
Huang et al (2012)				X	X(MLP)		X	X			X	X	
Hung and wang (2004)				X	X(BP)								
Keramati et al. (2014)					X		X	X					X
Li et al. (2011)										X			
Lu et al. (2014)				X					X				
Noyan and Simsek (2014)		X											
Qureshi et al. (2013)				X	X(MLP)		X						
Runge et al. (2014)				X	X		X	X					
Tamaddoni Jahromi et al (2014)				X			X		X				
Tsai and Chen (2010)	X				X(MLP)		X						
Umayaparvathi and Iyakutti (2012)					X		X						
Vafeiadis et al. (2015)				X	X(MLP)	X	X	X			X		
Yu et al. (2011)					X(BP)		X	X					

SEM: Structural Equation Modelling, LR: Logistic Regression, ANN: Artifical Neural Network, ARD: Automatic Relevance Determiniation, MLP: Multi Layer Perceptron, BP: Back propagating, CART: Classification and Regression Trees, SVM: Support Vector Machine, NB: Naive Bayes, LC: Linear Classifiers, KNN: K-Nearest Neighbor

Classification and Regression Trees

Classification and Regression Trees are another popular and commonly used method for classification (Khodabandehlou & Zivari Rahman, 2017). The benefit of this method is that it is able to include missing values and is easily interpreted. A disadvantage is that decision trees do not give the chance of belonging to a class but only what class an observation belongs to (Pladdet, 2018). This can however be overcome by using a combination of decision trees, being called a random forest or by using the number of observations in the end node. Decision trees are well used in churn predictions as can be seen in table 2.2. The reason for using this method are the possibility of using both categorical and numerical data and that it is relatively fast and efficient (Keramati et al., 2014).

Support Vector Machine

Support vector machine (SVM) is a classification method used for non-linear and linear classifications problems (Crone, Lessmann, & Stahlbock, 2006). Advantages of this method are that it can handle overlapping classes and that it is flexible (Pladdet, 2018). A disadvantage is that it can be slow when using large datasets. SVM shows good results for churn prediction (Huang et al., 2010). Khodabandehlou and Zivari Rahman (2017) even conclude that SVM outperforms decision trees in most of the studies. According to Yu, Guo, Guo, and Huang (2011), the reason for this is that SVM can take sample imbalance into account, which is usually the case for churn prediction. This method however does not give a probability of being a churn observation.

Artificial Neural Network

Artificial neural networks (ANN) is a method that uses a chained network of equations in order to produce an output. The advantage of this method is that it is able to extract complex relations between the input variables and the output variables (Pladdet, 2018). Neural networks are trained based on historical data and can easily be adapted when new data becomes available. A disadvantage of neural networks is that it is a blackbox method. This means that effects of variables are not interpretable and that it is unknown how an output is calculated. Table 2.2 shows the popularity of ANN in churn prediction literature of which most researchers use the multi-layer-perceptron variant (Khodabandehlou & Zivari Rahman, 2017). The reason for this is that ANN outperforms decisions trees and in some studies also outperforms support vector machines.

Cox proportional hazard model

In addition to these methods for churn calculation shows Srigopal (2018) the use of the Cox proportional hazard model for churn calculation. The Cox proportional hazard model is a method of survival analysis and creates a survival curve. This is a different approach than the aforementioned methods in that it does not predict the chance of churning in the next period, but it predicts the chance of reaching a specific period. Srigopal (2018) uses this model to determine if an action will have effect on the expected customer lifetime. An advantage of this method is that it is able to create different survival curves for different customers based on customer specific variables. A disadvantage is that that an underlying assumption is that the variables which create the customer specific survival curve do not change over time. It is therefore only possible to use method with actions when these actions occurred at the beginning of a customers' life. However, most visits do not take place at the start of the customerlife and are therefore affecting the expected customerlife from a later point. This makes the effect of the visit biased towards a higher effect when using this model.

2.3.3 Conclusion

The discussed methods for churn forecasting calculate different churn parameters. The first selection criteria for choosing a churn calculation method is that the method needs to forecast a churn probability. The logistic regression and classification trees can predict the probability of belonging to a churn or non-churn class, where the support vector machine and artificial neural network predicts what class an observation belongs to. The Cox proportional hazard model predicts the chance of reaching a specific period and could therefore also be used. The second criteria is that the effect of different retention actions should be directly interpretable from the method by including it as a variable. As discussed does this lead to biased results when using Cox Proportional hazard model. Therefore the logistic regression and decision tree will be used as methods for churn calculation.

2.4 Customer Lifetime Value Calculation Methods

This section provides an overview of the current research in customer lifetime value calculation. First general background information will be provided after which different calculation methods will be discussed. This section will end with a conclusion about the usefulness of the different methods for this study.

2.4.1 General background

Customer lifetime value (CLV) is a measurement of what a customer is worth at the present point in time. It takes into account what the customer is expected to generate in terms of cashflows in the future and the costs that will be made for this customer. CLV is used to determine what customers to focus on (Groeger & Buttle, 2014), to gain insight in what can be expected from customers in terms of profit (Kim, Jung, Suh, & Hwang, 2006) and to recommend certain products (Shih & Liu, 2008).

The basic formula for the present value of future cashflow is shown in equation 2.3. In this equation PV represents the present value, FV represents the future value in period t and r represents the discount rate. The sum of all discounted future values of a specific customer represent the present value of the expected customer.

$$PV = \sum_{t=1}^{\infty} \frac{FV_t}{(1+r)^{t-1}} \tag{2.3}$$

The basic CLV formula is formed by replacing FV_t in equation 2.3 by CM_t for contribution margin in period t and C_t being the costs made for this customer in period t and adding the acquiring costs AQ for that specific customer. This results in equation 2.4

$$CLV = \sum_{t=1}^{\infty} \frac{CM_t - C_t}{(1+r)^{t-1}} - AQ \tag{2.4}$$

The challenge is in how CM_t and C_t are calculated for every individual customer and period.

2.4.2 Methods for Customer Lifetime value calculation

Esmaeiligookeh et al. (2013) conducted a literature review about CLV calculations in which they studied the literature from 1985 to 2012. Table 2.3 shows the different used methods and the formulas of the articles using survival analysis. The calculation methods for CLV calculation in the article of (Esmaeiligookeh et al., 2013) can be categorized in three categories; Survival Analysis (SA), Markov Chain Modelling (MCM) and Recency, Frequency, Monetary modelling (RFM).

Survival Analysis

As can be seen in table 2.3 is survival analysis the most used method for calculating customer lifetime value. The aim of survival analysis is to create a survival curve, which resembles the probability of reaching a specific period. These probabilities are combined with expected contribution margin (J. Lu, 2003), with future cost (Kumar, Moorthy, Aeron, & Janakiraman, 2010) and other variables to create the expected future value of a customer. The benefit of this method is that a survival curve can be calculated with limited data. The KaplanMeijer estimate, as shown in equation 2.5, creates a survival curve using the duration of a customer life and if a customer has churned or not.

$$\hat{S}(t) = \prod_{l:t_l \leq t} \frac{n_l - d_l}{n_l} \tag{2.5}$$

Where t_l is a t where at least one churn event happened, d_l the number of churners at time t_l and n_l the number of customers that are still customer at t_l.

This survival curve can be further tailored to specific customers by creating the survival curve for different customer clusters or by using COX regression. In this method customers have different survival curves based on customer properties (Ozenne, Sørensen, Scheike, & Torp-pedersen, 2017). J. Lu (2003) combines the survival probabilities with monthly margins and a discount rate to calculate the present worth of the expected future contribution margin of a specific customer in equation 2.6. This calculation method is similar to the one from Lehman and Gupta (2003), where the difference is that they calculate the CLV using yearly margins and yearly survival rate.

$$CLV = MM \cdot \sum_{t=1}^{T} \frac{p_t}{(1+\frac{r}{12})^{t-1}} \tag{2.6}$$

Where MM is the average monthly margin, T is the last period until the CLV is calculated, p_t is the probability of surviving until period t and r is the yearly discount rate.

Others extend this model by including the costs made for the customer (Blattberg & Deighton, 1996; Berger & Nasr, 1998). This further specifies the expected value that can be expected from a customer but also requires more forecasting. Hwang, Jung, and Suh (2004) propose combining past profit contribution and potential benefit with survival analysis. They include opportunities for cross-selling at current customers in their potential benefit calculation which further extends the customer valuation. Ahmadi, Taherdoost, and Fakhravar (2011) state that different calculation methods should be used for different buyer seller relationships. They state that simple PV calculation, as in equation 2.3, can be used in situations where risks affecting the customer cashflow are low. For situations where this risk is high, Ahmadi et al. (2011) show that real options analysis is more accurate in predicting CLV than PV based models.

Markov Chain Modelling

Another method for customer lifetime calculation is using Markov chain modelling. In Markov modelling, all possible states of a customer are counted (Pfeifer & Carraway, 2000) and used to form a chain. A possible state in the churn context is staying customer for another period or churning in that period. These possibilities form a chain, where the probability of moving to another state can be determined by using chain theory. Pfeifer and Carraway (2000) use this method and state that the benefits of using Markov modelling are that it is flexible, probabilistic and can be used in decision making. The disadvantage is that the Markov chains can become very complex when multiple states are combined to form the Markov chain. This makes calculating with Markov chains more complex.

Recency, Frequency, Monetary modelling

Recency, Frequency and monetary modelling is a method for choosing customers to target and is mostly applied in marketing studies (Esmaeiligookeh et al., 2013). It uses the recency, which is the time since the last purchase, the frequency, which is the total number of orders of a customer and monetary, which is the average revenue generated by the historical orders of a customer. These variables are used to determine how valuable a customer is. Fader, Hardie, and Lee (2005) use this method where customers are grouped based on these three variables and labelled as in a range from very variable to not very variable. (D. Liu & Shih, 2005) include weights for the different variables to calculate a valuability score. The benefit of this method is that customers can be easily prioritized and that companies can easily diversify their marketing budget based on different target groups. The disadvantage of this method is that it can not be used to determine future value and assumes that customers does not change its behavior (Esmaeiligookeh et al., 2013).

Table 2.3: Methods for calculating customer lifetime value

	SA	MCM	RFM	Formula
Blattberg and Deighton (1996)	X			$CLE = am - A + a \cdot (m - \frac{R}{rr}) \cdot (\frac{rr^n}{1-rr^n})$
Berger and Nasr (1998)	X			$CLV = GC \cdot \sum_{i=0}^{n} \frac{r^i}{(1+d)^i} - M \cdot \sum_{i=0}^{n} \frac{r^{i-1}}{(1+d)^{i-0.5}}$
Colombo and Jiang (1999)			X	
Pfeifer and Carraway (2000)		X		
Lehman and Gupta (2003)	X			$CLV = GC(\frac{r}{1+d-r})$
J. Lu (2003)	X			$CLV = MM \cdot \sum_{i=1}^{T} \frac{p_i}{(1+\frac{d}{12})^{i-1}}$
Rust, Lemon, and Zeithaml (2003)		X		
Fader, Hardie, and Lee (2004)			X	
Hwang et al. (2004)	X			$CLV_i = \sum_{t_i=0}^{N} {}_i\pi_p(t_i)(1+d)^{N_i - t_i} + \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{\pi_f(t_i)+B(t_i)}{(1+d)^{t_i - N_i}}$
Fader et al. (2005)			X	
D. Liu and Shih (2005)		X		
Haenlein, Kaplan, and Beeser (2007)		X		
Yeh, Yang, and Ting (2009)			X	
Kumar et al. (2010)	X			$CLV_i = \sum_{t=1}^{n} \frac{Future\,contribution\,margin_{it} - Future\,cost_{it}}{(1+d)^t}$
Ahmadi et al. (2011)	X			$CLV = \sum_{t=0}^{n} (\frac{max(m \cdot q(S) - S, m*q)}{(1+d)^t} - A_t)$
Cheng, Chiu, Cheng, and Wu (2012)		X		

Variables: CLE : *Customer Lifetime Equity*, a : *acquition rate*, m : *margin on a transaction*, A : *Acquisition cost per customer*, R : *retention cost per customer*, rr : *yearly retention rate*, d : *yearly discount rate*, GC : *contribution margin for each customer*, M : *promotion costs for each customer*, n : *length of the period of cash flow*, π_p : *past profit contribution*, π_f : *future cash flow contribution*

2.4.3 Conclusion

The advantages and disadvantages of Survival Analysis, Markov chain modelling and the RFM model are discussed in this section. All methods are widely used but serve different purposes. The outcome of the customer lifetime value calculation method needed for this research is a monetary value. Therefore the RFM model can not be used. The Survival Analysis and Markov chain modelling both generate probabilities of reaching a specific time period. Of these two methods, the survival analysis is mostly used and therefore this method is preferred. Based on the available variables, the CLV calculation method of J. Lu (2003) as shown in equation 2.6 is used.

Chapter 3

Proposed Prescriptive Resource Allocation System

This chapter will explain the proposed prescriptive resource allocation system, making use of data mining techniques as described in the previous chapter. First the system flow will be presented after which all elements are explained separately. This chapter will end with the benefits of using this system and how the system will be tested in the experimental setup.

3.1 The system setup

The aim of this setup is to create an optimal resource allocation system for sales agents to reduce churn in a business to business non-subscriptual context. Figure 3.1 shows the components of the prescriptive system, which are customer clusters, churn models, action effectiveness determination, a customer lifetime value model and an optimization model. The customers are first clustered based on their customer characteristics, after which a cluster specific churn model is created. These models generate an individual churn probability for all customers in a situation with sales visits and without sales visits. Based on these probabilities the action effectiveness of visits on the churn probability of an individual customer can be determined. This is then combined with the individual's customer lifetime value to generate a list of optimal visits in the optimization model in the coming month. These individual components will be described in section 3.2.

3.2 Individual components of the model

In this section the individual components as shown in figure 3.1 are discussed. The clustering, churn model, action effectiveness determination, customer lifetime value calculation and optimization will be explained using a running example, which will be presented first.

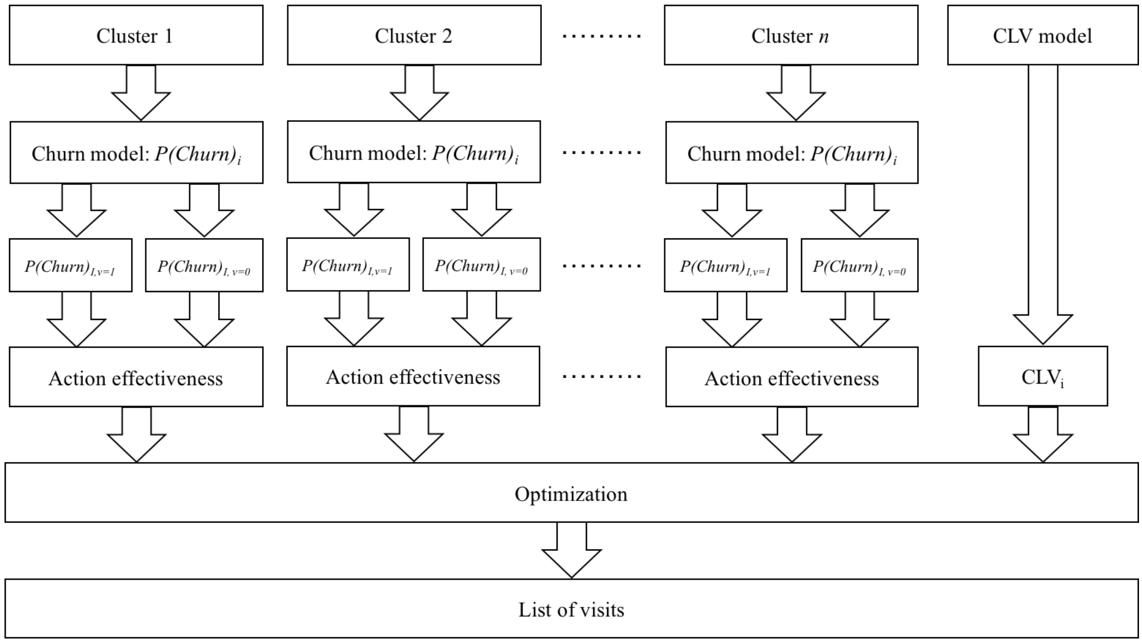


Figure 3.1: Prescriptive system

3.2.1 Running example

The running example used in this section will consist of five example months of individual customers as shown in table 3.1, where each row corresponds with one month of one customer. The months will serve as time t for which the churn probability at the end of $t+1$ will be predicted. For simplicity reason, the customers' ID range from one to five an have been customer for different lengths as shown in the customerlife column. The customers can be frequent buyers or infrequent buyers and differ in the average contribution margin and average number of products per order. Future visit is zero for all customers since this has not yet happened. In reality all months of a customer's life are included in the model and more variables are used.

Table 3.1: Running example customers

ID	Customer life	Frequent buyer	Contribution margin	Number of products	Future Visit
1	2	Yes	100	2	0
2	4	Yes	150	1	0
3	10	No	750	2	0
4	8	Yes	280	6	0
5	5	No	135	3	0

3.2.2 Clustering

The first step is to create relevant clusters for the customers. There will be two clustering methods used, being k-means and Gaussian mixture modelling clustering (GMM). The outcomes of the models using these clusters are compared with each other and with the outcomes of models without clustering after which one method is chosen. The outcome of this clustering step is a cluster weight vector for each customer as shown in equation 3.1.

$$W_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ . \\ . \\ w_{i,n} \end{bmatrix} \tag{3.1}$$

Where i is the customer and 1,..,n are the number of the clusters

For k-means this will be a vector where one weight takes the value one and all other weights take the value zero. For Gaussian mixture modelling clustering the values of all weights will be between 0 and 1 with a total sum of 1. The GMM clustering is especially useful when clusters overlap. To determine the chance that an observation belongs to a certain class, Gaussian mixture modelling uses a combination of Gaussian distributions for the different clusters over the different variables. Based on the position of the Gaussian distribution and the position of the observation, the probability of belonging to a cluster is calculated. The parameter explanation and optimization are described in section 5.1.6.

The weight vectors from this step are used to create individual churn models for every customer in section 3.2.3.

Running example

The customers of the running example are clustered based on their customer characteristics. The assumption is that there are two clusters, which in this case cluster 1 for frequent buyers with a high average number of products per order and cluster 2 for infrequent buyers with a low number of products. Because some customers do not explicitly belong in one of these clusters, weights are assigned as if GMM was used. This leads to the weight $w_{ID,1}$ for cluster one and weight $w_{ID,2}$ for cluster per customer as shown in table 3.2.

Table 3.2: Running example after clustering

ID	Frequent buyer	Number of products	$w_{ID,1}$	$w_{ID,2}$
1	Yes	2	0.6	0.4
2	Yes	1	0.5	0.5
3	No	2	0.1	0.9
4	Yes	6	0.8	0.2
5	No	3	0.2	0.8

3.2.3 Churn prediction

The second step is to create a churn prediction model. The churn model predicts the churn chance for individual customers at the end of t+1 or further in the future. The reason for this is that in order to optimally allocate resources, there has to be time to react. Predicting the churn chance at the end of at least one period in advance allows for actions to be taken in that period. This is graphically represented in figure 3.2.

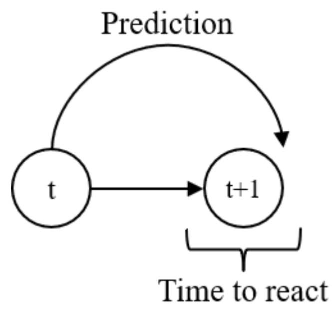


Figure 3.2: Graphical representation of prediction method

The churn prediction model $f(x)$ will be created for each cluster separately. Creating a number of models equal to the number of clusters, C. The individual cluster weights W_i in combination with the cluster models will create individual churn models per customer as shown in equation 3.2 and 3.3.

$$Churn\, probability_{i,t+1} = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ .. \\ .. \\ w_{i,C} \end{bmatrix} \begin{bmatrix} f_1(x_{i,t}) & f_2(x_{i,t}) & .. & .. & f_C(x_{i,t}) \end{bmatrix} \tag{3.2}$$

Or more formally:

$$Churn\, probability_{i,t+1} = \sum_{c=1}^{C} w_{i,c} \cdot f(x_{i,t}) \tag{3.3}$$

Where i is the customer, t is the period and C is the number of clusters.

The used models are variations of logistic regression and decision tree, based on the literature review. The logistic regression generates a chance between 0 and 1 as output and can therefore directly be used to model the churn probability. Classification trees however, are designed to generate the class an observation belongs to, rather than a probability of belonging to a certain class. Therefore, some alterations need to be done to make the classification tree generate a probability. The method described by Provost & Fawcett and Provost (2013) will be used for this, which uses the proportion of the different classes in the set of the end note to calculate the probability. Figure 3.3 shows this calculation.

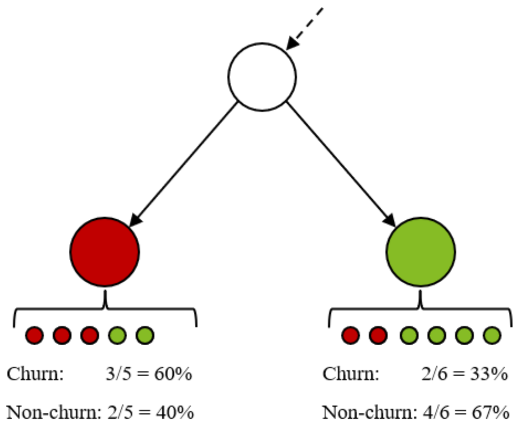


Figure 3.3: Churn probability determination for decision trees

Running example

Since the customers of the running example are clustered in two clusters, there are two churn models created. The churn probability for the individual customers are calculated using both model $f_1(x_{ID,t})$ and $f_2(x_{ID,t})$ and then multiplied by the cluster weights as shown in 3.3. Both models include the future visit variable as variable effecting the churn probability. The churn probability at the end of $t+1$ is calculated.

Table 3.3: Running example after churn prediction

ID	$w_{ID,1}$	$w_{ID,1}$	$f_1(x_{ID,t})$	$f_2(x_{ID,t})$	$Churn\ probability_{ID,t+1}$
1	0.6	0.4	0.4	0.6	0.48
2	0.5	0.5	0.3	0.7	0.50
3	0.1	0.9	0.2	0.4	0.44
4	0.8	0.2	0.9	0.6	0.84
5	0.2	0.8	0.3	0.5	0.46

3.2.4 Action effect calculation

The third step is to determine the effect of an action on the churn probability of a customer. For this, the actions variables in the model for predicting churn. To make the model prescriptive, future action variables should be included which are in the future up until the moment of which the churn probability is predicted. In the situation of figure 3.2 where the churn probability at the end of period $t+1$ is predicted, the visit variable for $t+1$ should be included, which allows for time to react. If the churn probability at the end of $t+2$ would be predicted, then the visit variable for $t+1$ and $t+2$ should be included.

The effect of the actions on the churn probability is determined by predicting the churn probability twice per observation. The first time the churn probability is predicted using the original variable values of the observation, resulting in $Churn\ probability_{i,t+1,a}$. The second time the churn probability is calculated again, but this time the action variable is increased by 1, indicating that a visit has taken place. This leads to $Churn\ probability_{i,t+1,a+1}$, being the churn probability when the action has taken place. The effect of the action on the churn probability is than calculated by using formula 3.4, where i is the customer and a is the action.

$$Action\ effect_{i,t+1,a} = Churn\ probability_{i,t+1,a} - Churn\ probability_{i,t+1,a+1} \tag{3.4}$$

This action effect differs per customer and can differ per period based on what model is used.

Running Example

The churn probability is now calculated again using the same method as before but by using the data in table 3.4, where 1 future visit is added for all customers. The churn probability using this data is called $Churn\ probability_{ID,t+1,a+1}$.

Table 3.4: Running example customers with addition of 1 future visit

ID	Customer life	Frequent buyer	Contribution margin	Number of products	Future Visit
1	2	Yes	100	2	1
2	4	Yes	150	1	1
3	10	No	750	2	1
4	8	Yes	280	6	1
5	5	No	135	3	1

The action effectiveness is calculated using equation 3.4 and shown in table 3.5. The $Churn\ probability_{ID,t+1}$ corresponds to the churn probability as predicted with the original data without the visit.

Table 3.5: Running example after action effectiveness calculation

ID	$Churn\ probability_{ID,t+1}$	$Churn\ probability_{ID,t+1,a+1}$	$Action\ effect_{ID,t+1,a}$
1	0.48	0.40	0.08
2	0.50	0.50	0.00
3	0.44	0.43	0.01
4	0.84	0.69	0.15
5	0.46	0.40	0.06

3.2.5 Customer lifetime value

The last step before optimization is to calculate the expected value that is generated by the action. For this the expected future value of the customer is calculated. The KaplanMeijer estimate will be used for this, which is a widely used method. This estimate calculates the probability that churn moment T is further in the future than time t, $P(T > t)$. This is the probability of reaching a specific customer lifetime. Equation 3.5 is used to calculate the survival rate for every point t, in which all customers' lifetime durations are included. This forms the survival curve.

$$\hat{S}(t) = \prod_{l:t_l<t} \frac{n_l - d_l}{n_i} = P(T > t) \tag{3.5}$$

Where $\hat{S}$ = Survival rate at t, n_l = number of subjects at risk at t and d_l = death events at time t

This probability of reaching period t is updated in every new period by including the information that the subject remained customer until that period. The conditional probability of a customer staying until t given that a customer has stayed until time s can be calculated using Bayes' theorem as shown in equation 3.6.

$$P(T > t\,|T > s) = \frac{P(T > s|T > t) \cdot P(T > t)}{P(T > s)} \tag{3.6}$$

Because the customer lived until s, t must be equal or higher than s. This means that the probability that a customer stays until time s when given that the customer stayed until time t $P(T > s|T > t)$, is equal to 1. Equation 3.7 shows the new equation.

$$P(T > t\,|T > s) = \frac{P(T > t)}{P(T > s)} \tag{3.7}$$

Equation 3.7 can then be rewritten using the KaplanMeijer estimate as shown in equation 3.5 into:

$$P(T > t\,|T > s) = \frac{\hat{S}(t)}{\hat{S}(s)} = p_{i,t} \tag{3.8}$$

The survival curve created by the KaplanMeijer estimate is updated using equation 3.8 by scaling the curve at point s. This creates a curve where the probability of reaching s is one and $s < t$ is scaled based on the probability of reaching s.

The rescaled survival curve of the KaplanMeijer estimate $\hat{S}(t)$ becomes a customer specific probability $p_{i,t}$ for reaching t. This is then used in equation 3.9 to calculate the CLV of a specific customer.

$$CLV_i = CM_i \cdot \sum_{t=CCL_i}^{T} \frac{p_{i,t}}{(1+\frac{r}{12})^{t-1}} \tag{3.9}$$

Where i is the customer, $p_{i,t}$ is the customers probability of reaching period t, CM_i is the average monthly contribution margin of the customer, CCL_i is the current customer lifetime and r is the interest rate.

The T in equation 3.9 can be altered in order to test the robustness of the system. The higher T the higher the CLV_i will be for all customers.

Running example

The remaining customer lifetime value of the running example customers is calculated using the average contribution margin and survival curve.

Table 3.6: Running example customers after CLV calculation

ID	Customer life	Contribution margin	$\sum_{t=CCL_{ID}}^{T} \frac{p_{ID,t}}{(1+\frac{r}{12})^{t-1}}$	CLV
1	2	100	3.8	380
2	4	150	2.4	360
3	10	750	4.2	3300
4	8	280	6.5	1820
5	5	135	4.0	540

3.2.6 Optimization

The optimization model combines the customer lifetime value, churn probability, action effect and the decision variable a_i. It maximizes the customer lifetime equity (CLE), which is the sum of the expected value from all customers.

$$max(CLE) = max \sum_{i=1}^{I} CLV_i \, (1 - (Churn\,probability_{i,t+1} - Action\ effect_{i,t+1\,,a} \cdot a_i)) \tag{3.10}$$

With respect to:

$$\sum_{i=1}^{I} a_i \leq Maximum\ number\ of\ visits \tag{3.11}$$

Constraining for the maximum amount of visits, where the assumption is made that all visits are weighted equally as one visit.

$$\begin{aligned}
&\text{CLV}_i && = CM_i \cdot \textstyle\sum_{t=1}^{T} \frac{p_{i,t}}{(1+\frac{r}{12})^{t-1}} \\
&\text{Churn probability}_{i,t+1} && = \textstyle\sum_{c=1}^{C} w_{i,c} \cdot f_c(x_{i,t}) \\
&\text{Action effect}_{i,t+1\,,a} && = Churn\,probability_{i,t+1,a} - Churn\,probability_{i,t+1,a+1} \\
&\text{a}_i && = \text{0 if no action has taken place, 1 if action has taken place}
\end{aligned}$$

This equation can be optimized by using integer programming where the decision variables a_i determine which action to pursue for what customer in the coming period. In practice this optimization problem will be optimized by calculating the monetary effect of all actions and rank these based on their value. The maximum number of visits are than selected by selecting the ranks 1 to the maximum number of visits. The CLE is than calculated using these visits.

Running example

Table 3.7 shows the input for the optimization model from the previous sections. The inputs are used to determine the optimal visits creating optimization problem in equation 3.12

Table 3.7: Running example input for optimization model

ID	$Churn\,chance_{ID,t+1}$	$Action\ effect_{ID,t+1,a}$	CLV	Return of visit	Visit rank
1	0.48	0.08	380	30,4	4
2	0.50	0.00	360	0	5
3	0.44	0.01	3300	33	2
4	0.84	0.15	1820	273	1
5	0.46	0.06	540	32,4	3

$$\begin{aligned} max(CLE) = max((1-(0.48-0.08\cdot a_1))\cdot 380) + (1-(0.50-0.00\cdot a_2))\cdot 360) \\ +(1-(0.44-0.01\cdot a_3))\cdot 3300) + (1-(0.84-0.15\cdot a_4))\cdot 1820) \\ +(1-(0.46-0.06\cdot a_5))\cdot 540)) \end{aligned} \tag{3.12}$$

If we assume a maximum number of two visits, than the optimal visits are visiting customer 4 and customer 3. This leads to a customer lifetime equity of 3114,40 compared to 2811,00 without visits.

3.3 Benefits of the model

The first benefit of using the proposed method is that the system is able to identify the effect of a visit on the churn probability of a specific customer and uses this to prescribe which customers to visit. By clustering the customers, the churn model and action effect can be calculated for customers with similar characteristics. This makes the prediction more accurate and gives insight in how different customers react on visits. The customer specific outcome allows for targeting of customers based on the effectiveness of the actions and customer value on individual level.

Secondly, the system is flexible in that this method can be applied in situations with different kind of retention actions. It can identify the effectiveness of the action individually and the effectiveness of a combination of actions. It is also flexible in the number of customers and new customers can directly be placed in the most relevant cluster and included in the optimization.

The third benefit is that the output of the model consists of a list of customers to visit. This can directly be used as starting point for the planning of sales visits. It can also be used as input for a CRM system where this process is automated.

3.4 Experimental setup

The proposed model will be tested in the case study of chapter 4 using an experimental setup as shown in figure 3.4. The data will be split into a training set and a hold out month. The hold out month serves as a test set and is not used for training. The training data is split into an appropriate number of clusters after which 10 fold cross validation is used for parameter tuning of the churn models. The churn model is then trained per cluster using the optimal parameters. The final model including all clusters is evaluated using the hold out month. The evaluation criteria can be found in section 6.1.1.

Parameter tuning and evaluation of the models will be based on the outcome of 10-fold cross validation, where a model is trained on a trainset and evaluated on a testset. The final evaluation metrics of the model are based on the average of the evaluation metrics over the 10 folds. For the models where oversampling is used, only the trainset will be oversampled to prevent spillover of observations from the trainset to the testset.

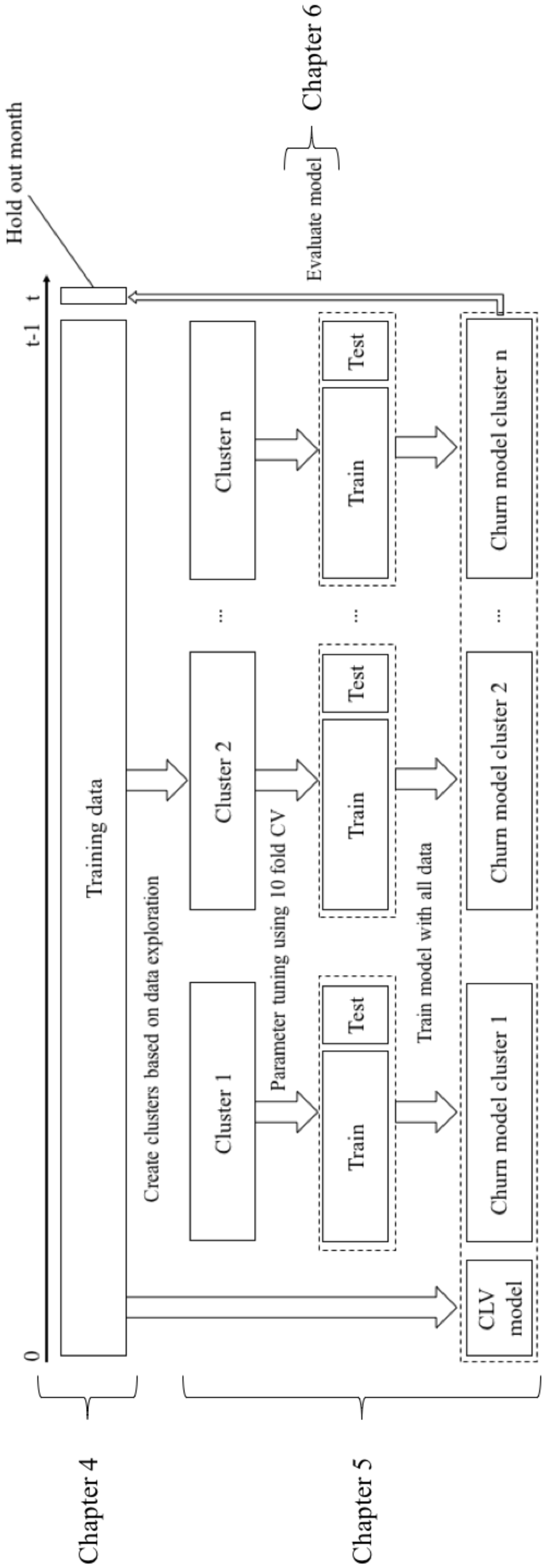


Figure 3.4: Experimental setup

Chapter 4

Case Background

This chapter will describe the background of the case. It starts with the business understanding section, where background is given about the business context and objectives. This is followed by the data understanding section, which describes the dataset and explores the different variables.

4.1 Business Understanding

The aim of this section is to describe the context and goals of the case study. This section consists of the three subsections: business objectives, scope, target variable and data mining goals.

4.1.1 Business Objectives

The company in this case study operates in the chemicals market and sells to vendors, professional users and production companies. The company is a conglomerate with different brands of both consumer and professional products. It operates through business to business (B2B) channels and uses sales agents to attract new sales and manage existing accounts.

This case study focuses on the sales department of the company. In the existing situation, as shown in figure 4.1, the sales agents decide how to allocate their time over different customers themselves. For making this decision, the company offers the sales agents a customer relationship management system (CRM) which includes demographic information about the clients, the order history and actions undertaken by sales agents. At this point, all information in the CRM system is descriptive information based on historical data. The company believes it can increase the total customer value when sales agents resources are allocated more optimal.

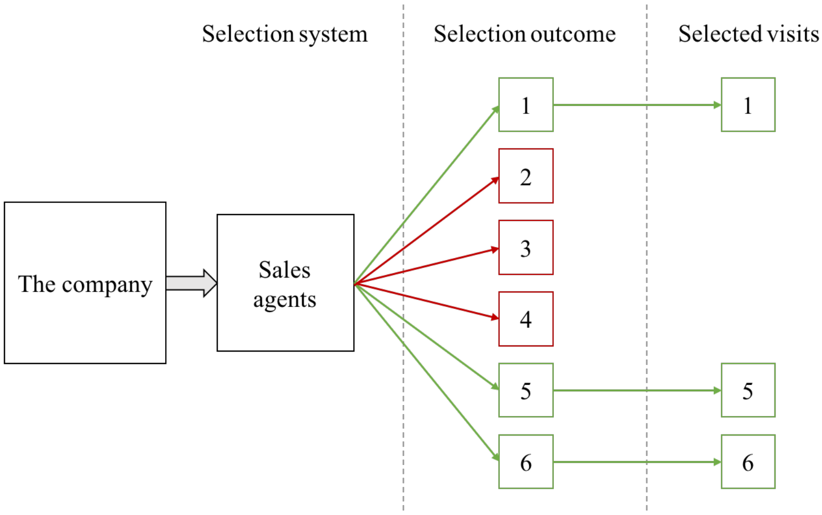


Figure 4.1: Current selection of customers to visit

Besides increasing total customer value based on this model, the company also seeks an actionable outcome. The business objective with its success criteria are presented in table 4.1.

Table 4.1: Business objectives and succes criteria

	Objective	Succes Criteria
1)	Optimize resource allocation of sales agents	1a) Increased total customer value
		1b) Outcome is actionable

4.1.2 Scope

This case study will serve as an exploratory study and be an use case for further use of prescriptive resource allocation. Therefore the decision is made to solely focus on the Small Service Outlets (SSO) located in the Netherlands. The SSOs are small paint and hardware stores which are visited by sales agents throughout the year. No contracts or price agreements are made with these customers and products are bought by order. The customers can receive discounts on product price and can be part of loyalty program gold, diamond or platinum. This case study will only focus on the use of prescriptive resource allocation for keeping existing customers and maximizing the total customer value of these customers.

4.1.3 Target variable

Since the focus in this case study lies on maximizing the total customer value of existing customers, the target value is customer churn. There are no contracts or subscriptions between the customer and the company and therefore there is no explicit moment of customer churn. The churn definition is therefore indirect and needs to be formulated. There are two churn definitions used in this case study: static churn and dynamic churn. The reason for this is that these churn definitions both use activity as their criterion for being churned of which the data is available. The reason for choosing two definitions is to determine the effect of different churn definitions.

A customer is churned according to the static churn definition when that customer did not order for a specific period of time. This period can be n number of months and is an equal period for all customers. This is shown in equation 4.1.

$$Static\,churn_{i,t} = \begin{cases} 0 & if\,Active\,in\,period\,(t,\,t+n) \\ 1 & if\,Not\,active\,in\,period\,(t,\,t+n) \end{cases} \tag{4.1}$$

Where i is the customer, t is the time period and n the number of months

A customer is churned according to the dynamic churn definition when a customer did not order for a period of time specific to this customer. This is based on the median of the order intervals OI_i of this customer and changes over time when the order interval changes. The order interval is calculated by determining the period between consecutive orders. By using the order interval, dynamic churn makes a distinction between customers which order order frequently and those that do not order frequently. The median of the order interval is used because it is less affected by outlier order intervals and gives a more realistic representation of a typical order interval of a customer. The median order interval is multiplied by a factor f in order to get to a period per customer. This is shown in equation 4.2.

$$Dynamic\,churn_{i,t} = \begin{cases} 0 & if\,Active\,in\,period\,(t,\,t+f(OI_i)) \\ 1 & if\,Not\,active\,in\,period\,(t,\,t+f(OI_i)) \end{cases} \tag{4.2}$$

Where i is the customer, t is the time period, f the multiply factor and $M(OI_i)$ is the median order interval of customer i

4.1.4 Data Mining Goals

The company seeks to use a system which supports the sales agents in their decision on how to allocate their time over different existing customers and what actions to use for specific customers in a given period. The objective for using this system is to allocate sales resources for maximization of future customer value. To achieve this, predictive models for the churn probability, customer value and action effectiveness need to be created and combined with an optimization model. The data mining goals are presented in table 4.2.

Table 4.2: Data mining goals

	Goals
1)	A churn probability model on specific customer level
2)	A customer value model on specific customer level
3)	An action effectiveness model on specific customer level
4)	A resource allocation model which optimizes for customer value

4.2 Data Understanding

The data understanding section consists of four subsections, being: initial data, data description, data exploration and data quality. The aim of this section is to present and understand the data. The four subsections will all be discussed separately.

4.2.1 Initial Data

The initial data of the company consists of sales data, demographic data and action data until the 30th of October 2018.

4.2.2 Data Description

The sales data from the company is directly exported from SAP on the 1st of November 2018 and its attributes are shown in table 4.3. It consists data of 1393 customers over a total timerange from 02-02-2015 until 30-10-2018. The reason for this timerange is that the company changed systems at the beginning of 2015 and previous data was no longer stored. The data is stored in rows, where every row corresponds to an order of products from a specific brand for a specific order for a specific customer. All customers have an unique identifier *Customer number*, which is identical in all datasets. *Brand name* corresponds to the 22 different brands being offered by the company, where the name of the brands is replaced by a corresponding number. The *Document number* corresponds with the order number. Since products from different brands can be ordered on one order number, this order number is not unique over the dataset. It does however correspond to one unique order. *Calender day* corresponds to the day the order was placed in a YYYYMMDD format. The *Quantity* is the amount of liters of the product that was ordered for a specific order. The *Total revenue* is the sum of the revenue generated by selling the quantity of a certain product for a specific order in euro. This revenue is calculated after discounts has been deducted. The *Discounts* correspond with the amount of discount which is given. This is is displayed as a negative value. *Contribution margin* is the profit generated from selling the quantity of the specific product in the order. This attribute is calculated in euro.

As can be seen in table 4.3, the *Revenue*, *Discounts* and *Contributionmargin* show both positive and negative values. Only negative values are expected for *Discounts*, where only positive values are expected

Table 4.3: Sales data attributes

Attributes	Explanation	Example	Range/Set
Customer number	Unique identifier	12345678	
Customer name	Name of the customer	Jansen B.V.	
Brand name	Brand name of the product	17	[1, 22]
Document number	Order number	98765432	
Calender day	Date in YYYYMMDD	20181205	[20150102, 20181030]
Quantity	Quantity of product in liters	5,4	[0.0, 31313]
Total revenue	Revenue after discount in euro	120	[-23, 182905]
Discounts	Discount in euro	-20	[-424523, 6458]
Contribution margin	Contribution margin in euro	30	[-11085, 121560]

for *Total revenue* and *Contribution margin*. The reason that these values exist is that return orders are included, where products are returned and the company has to pay back the amount which is payed for the products. These orders will be excluded from the dataset.

The demographic data of the customers is exported from the CRM system of the company and its attributes are presented in table 4.4. *Customer number* and *Customer name* correspond to the same unique identifiers as in the sales data. The *Address* attribute indicates the location of the store of the customer and *Subtype* indicates if the customer is of subtype 1 or subtype 2. The *Loyalty program* attribute indicates if the customer is part of loyalty program Gold, Diamond, Platinum or of no loyalty program. No missing values are present in this dataset.

Table 4.4: Demographic data attributes

Attributes	Explanation	Example	Range/Set
Customer number	Unique identifier	12345678	
Customer name	Name of the customer	Jansen B.V.	
Address	Address of the customer	Lismortel 62 5612 AR Eindhoven	
Subtype	Subtype of customer	1	{1, 2}
Loyalty program	Part of loyalty program type	Gold	{Gold, Diamond, Platinum, None}

The visit data is also exported from the CRM system of the company and its attributes are presented in table 4.5. The *Customer number* and *Customer name* are the unique identifiers corresponding with the unique identifiers in the other datasets. The *Visit status* indicates if the visit is either completed, planned, in progress or cancelled. The *Start event* and *End event* attributes are a timestamp of the start and end of the visit. As can be seen in table 4.5, the start event and end event show values in the future. Since this research is focused on the timerange from 02-01-2015 until 31-10-2018, the visits with timestamps outside this period will be excluded. Furthermore, since some visits are planned, in progress or cancelled, it is unclear if these visits did take place. Therefore, these visits are also excluded.

Table 4.5: Visit data attributes

Attributes	Explanation	Example	Range/Set
Customer number	Unique identifier	12345678	
Customer name	Name of the customer	Jansen B.V.	
Visit status		Completed	{Completed, Planned, In progress, Cancelled}
Start event	Timestamp of start visit	05-12-2018 12:00	[27-08-2015 06:00, 29-09-2029 11:00]
End event	Timestamp of end visit	05-12-2018 13:00	[27-08-2015 07:00, 29-09-2029 12:00]

4.2.3 Customer pool

The customer pool consist of 1393 customers which differ in customer lifetime, amount of orders per year and total order size per year. 15 of these customers do not show any orders and are therefore excluded. The customers have different starts and ends of their customerlife. 496 customers have the start and end of their customerlife within the period 01-2015 to 10-2018. Of these customers the whole customerlife is captured within the dataset. For 464 customers the start of the customerlife is captured within the dataset but the end of the customerlife is not captured. These customers are customers who did not churn at the end of the dataset. 315 customers are active throughout the period of the dataset. It is unclear when these customers started their customerlife and when this will end. For the remaining 103 customers there is only an end of the customerlife visible in the dataset. The assumption is made that customers who ordered in the first two months of the dataset are old customers. Table 4.6 shows the number of customers per group for static churn.

Table 4.6: Start and end event within dataset - Static churn

Event within dataset	Number of customers
Start - End	496
Start - No end	464
No start - No end	315
No start - End	103

Table 4.7 shows the number of customers, revenue and number of orders for the four years indexed for 2015. The type of customers are percentages of the total number of customers. This table is created on a overview level where all data is aggregated per year. What can be seen is that the number of ordering customers, which are the customers that have ordered at least one time in the specific year, show a decreasing trend from 2015 to 2017 and remains stable in 2018. Customers are considered a new customer when the customer did not order in years previous to the specific year. This percentage is high in 2015 due to the start of the dataset, but stable for 2016 to 2018. The current customers are customer that at least ordered in the year before the specific year. After an increase in 2016, the percentage of current customers decreased in 2017 and 2018. The reactivated customers are customers that ordered previously but not in the year directly before the year of interest. These customers were active, stopped being active and became active again. Due to the definition, 2015 and 2016 do not have reactivated customers. 2017 and 2018 shows a small percentage of reactivated customers.

Table 4.7: Ordering customers and revenue per year indexed for 2015

	2015	2016	2017	2018*
Ordering customers	100	87	78	77
New customers	22%	7%	8%	9%
Current customers	78%	93%	90%	88%
Reactivated customers	0	0	2%	3%
Total revenue in Euro	100	102	111	10
Average yearly revenue in Euro	100	102	111	100
Total number of orders	100	97	90	84
Average yearly number of orders	100	112	117	108

The total revenue shows an increasing trend over the years. It has to be noted that the revenue for 2018 is only measured until the end of October and is therefore lower than the actual revenue. The average yearly revenue also shows an increase over the years. This means that the trend is that customers spend more on their total yearly orders. The total number of orders has decreased over the years. This is mostly due to the decrease in ordering customers, since the average yearly number of orders shows an increase over the years.

4.2.4 Data Exploration

To facilitate data exploration, the salesdata, customerdata and visitdata is converted to monthly data. Therefore, the attributes monthly revenue, monthly discounts, monthly contribution margin, monthly number of products were created. The average of these attributes were taken to generate a general overview per customer. The order interval (OI) representing the time between two orders was calculated by measuring the interval between all orders in days and taking the median.

The churn rate will be explored first after which multiple variables are explored. Company experts indicated that their expectation was that the effects of visits differ for customers with different order frequency, order size, the yearly expenditure at the company and the stability of the behavior. Therefore, the initial attributes of which the effect on churn will be explored are $Order\ interval$, $Revenue$, $Number\ of\ products$, $Yearly\ revenue$ and stability of behavior. In addition the number of visits and $Discount$, are included, since these are action variable. The attribute's distribution for churners and non churners will be explored on a high level using boxplots and a lower level using histograms.

Churn

The first variable of interest is the churn rate. For this a variable will be added showing if a customer has churned according to the $Static\ churn$ and $Dynamic\ churn$ definition. These variables are calculated using equation 4.1 and equation 4.2. A customer is considered a churned customer if it has churned between 02-01-2015 and 31-10-2018 according to the used definition. Figure 4.2 shows what percentage of the customers have churned subject to different n for static churn and different f for dynamic churn. The n for static churn corresponds to the number of months a customer needs to be inactive to be considered a churned customer. The f for dynamic churn corresponds to the number of order intervals a customer needs to be inactive to be considered a churned customer.

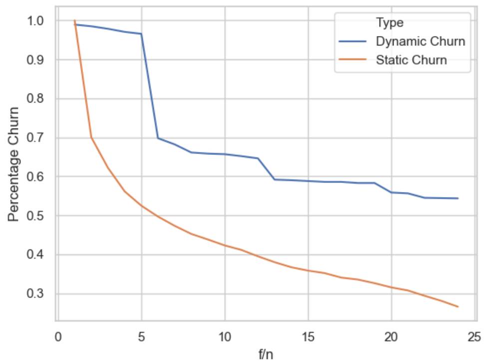


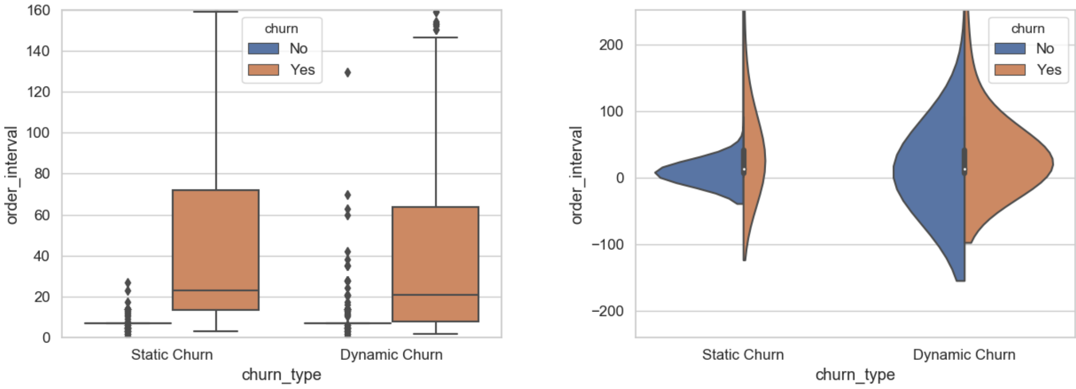
Figure 4.2: Percentage of churn subject to a different period of n months or different factor f

What can be seen from figure 4.2 is that static churn is a less strict definition than dynamic churn. The reason for this is that customers with a high order interval have a longer period in which they have to be inactive. The elbow point was was used to determine n and f. This is $n = 2$ for the static churn, making the period of inactivity before being churned 2 months. For dynamic churn this is $f = 6$, making the period of inactivity before being churned 6 times the typical order interval. These parameters result in a percentage churn of 70 percent of the population for both definitions.

Order interval

Figure 4.3a shows the distribution of the median $Order\ interval$ for non-churners and churners using both the static churn definition and the dynamic churn definition. The figure shows that churners tend to have higher order intervals than non churners in both definitions. This indicates that customers who order more frequent tend to be more loyal. The median for the static definition is 7 days for non-churners and

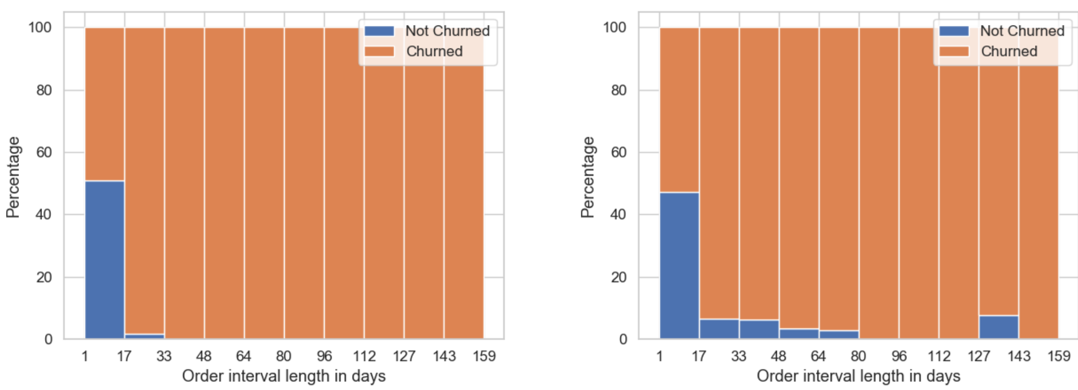
23 for churners. For the dynamic definition the median is 7 days for non-churners and 21 for churners. This indicates that the typical non-churner in both definitions orders weekly and the typical churner orders every three to three and a half weeks. Figure 4.3a also shows that the range of order intervals for non-churners is smaller than for churners and that the static non-churners show a smaller range than the dynamic non-churners. Figure 4.3b shows that the unimodal distribution of order interval is different for non-churners and churners when using the static churn definition. However when using the dynamic definition, the unimodal distribution of order interval for non-churners and churners show similarity. This indicates that this is not a good predictor for dynamic churn.



(a) (b)

Figure 4.3: Distribution of median order interval for non-churners and churners

Figure 4.4a shows that for the static churn definition the customer with an order interval of 1 to 17 days are more loyal and customers with higher order interval all tend to churn within four years. Figure 4.4b shows that for the dynamic churn, there are some customers with a order interval higher than 17 which do not churn within the four years. Customer with an order interval of 60 days or higher would always churn according to the static churn definition. However, comparing figure 4.4a and 4.4b shows that only a small percentage of customers with an order interval of 60 days or higher do churn when using the dynamic churn definition. Based on this exploration can be concluded that the variable *Order interval* shows predictive power for static churn and no predictive power for dynamic churn.

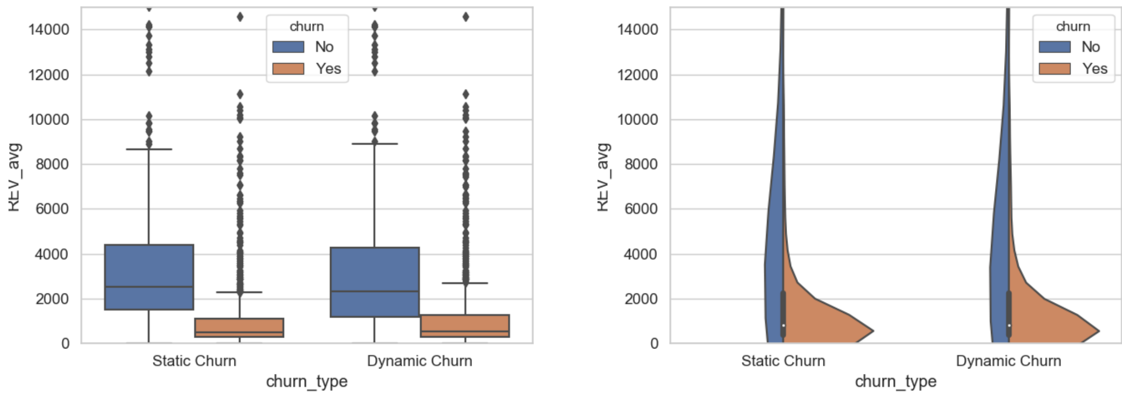


(a) Static Churn (b) Dynamic Churn

Figure 4.4: Percentage non churners and churners for different median order interval

Revenue

Figure 4.5a shows the distribution of the average *Revenue* per order for non-churners and churners using both the static churn definition and the dynamic churn definition. What can be seen is that non-churners tend to have a higher average revenue per order than churners. This indicates that customers who order for a higher value tend to be more loyal than customers who order for a lower value.

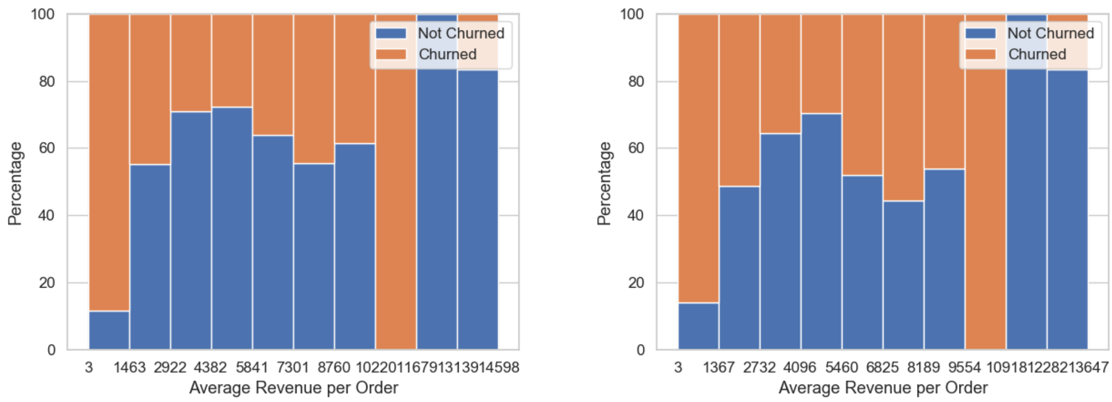


(a) (b)

Figure 4.5: Distribution of average revenue per order for non-churners and churners

The typical non-churner has an average revenue per order of 2528 euro per order for the static definition and 2313 euro for the dynamic definition. The typical churner has an average revenue of 515 euro per order for the static definition and 546 euro for the dynamic definition. Figure 4.5b shows that the churners in both definitions show a more compact unimodal distribution than the non-churners, indicating that revenue per order can be used as predictor for churn.

Figure 4.6a and figure 4.6b show a similar trend. Customers with a very low average revenue per order have high chance of churning. The group from low to high average revenue per order have a medium chance of churning and very high average revenue per order have low chance of churning. *Revenue* is therefore considered a valuable predictor for churn in both definitions.

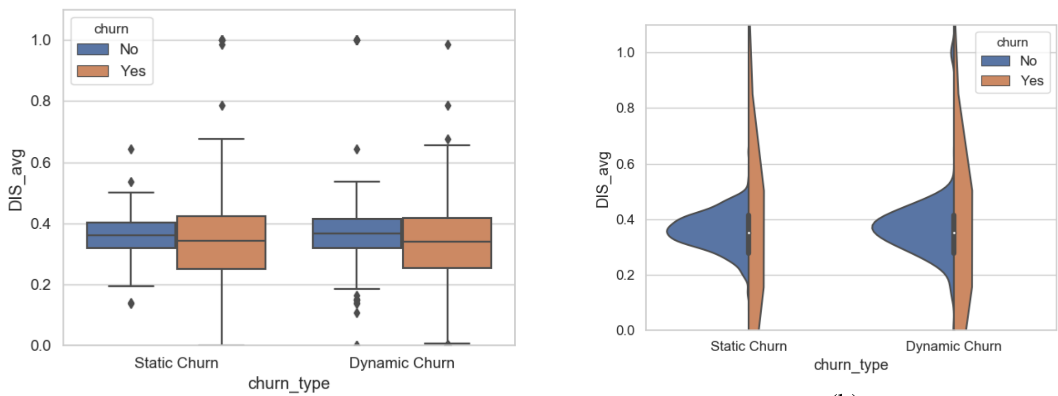


(a) Static Churn (b) Dynamic Churn

Figure 4.6: Percentage non churners and churners for different average sizes of revenue

Discount

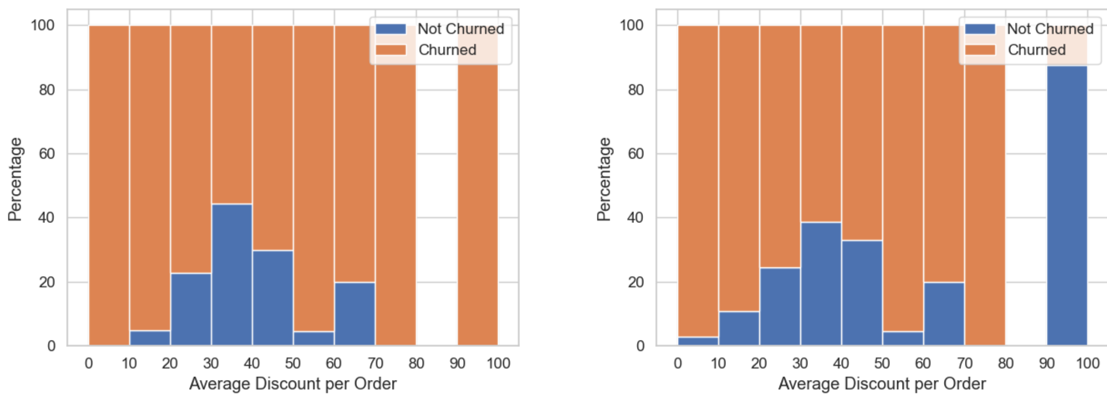
The customers receive different percentages discount per order. Figure 4.7a shows the distributions of the average discount per order for non-churners and churners. The churners and non-churners show similar average percentage discount and the distributions show overlap. This means that there does not seem to be a clear relation between percentage of discount and churn. The range of average discount per order is smaller for non-churners than for non-churners in both definitions. A typical non-churner receives a discount of 36% for the static definition and 37% for the dynamic definition. The typical churner receives 34% discount for both definitions.



(a) (b)

Figure 4.7: Distribution of average discount per order for non-churners and churners

Figure 4.8a and figure 4.8b show that both definition have similar distributions of average discount per order and percentage of non-churners and churners per group. The figures show that a moderate discount between 30 and 40 percent leads to the largest percentage of non-churners. A decrease in effectiveness is visible for lower discount and for higher discount. This indicates that there is a range in which the discount is effective and therefore the discount variable is considered to be a predictor for churn.

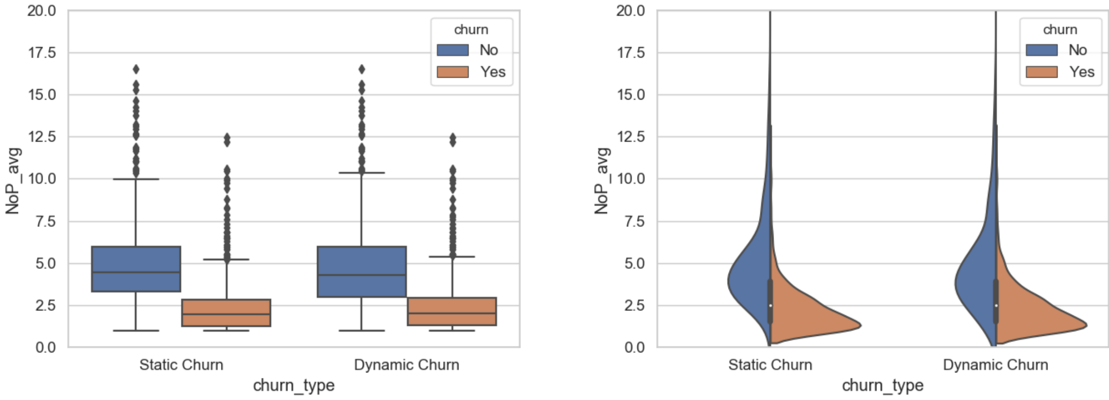


(a) Static Churn (b) Dynamic Churn

Figure 4.8: Percentage non churners and churners for different average sizes discount

Number of Products

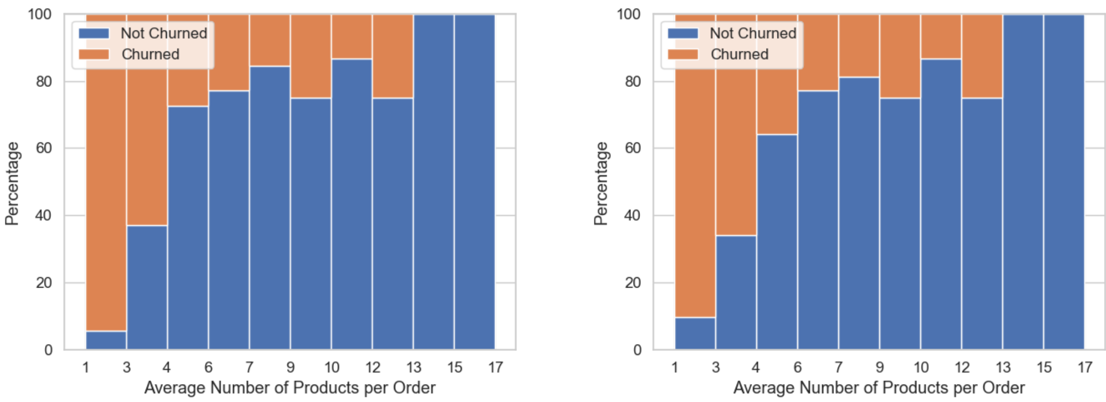
Figure 4.9a shows the average number of products per order for non-churners and churners. It can be seen that non-churners tend to buy more different products than churners. This means that customers that buy more different brands of products tend to be more loyal. A reason for this could be that the customer becomes more dependent on the company for different products. This trend is visible for both the static churn definition as for the dynamic churn definition. A typical non-churner orders products of 4 different brands for both the static definition and the dynamic definition. The typical churner orders products of 2 different brands for both the static as the dynamic definition. Figure 4.9b shows that the variable has an unimodal distribution which differs for non-churners and churners in both definitions. This indicates that number of products can be used as predictor for churn.



(a) (b)

Figure 4.9: Distribution of average number of products per order for non-churners and churners

Figure 4.10a and figure 4.10b show that the higher the average number of products per order the less customers churn. Especially the customers with 1 to 3 products per order on average tend to be less loyal than other customers with a higher average of products per order.

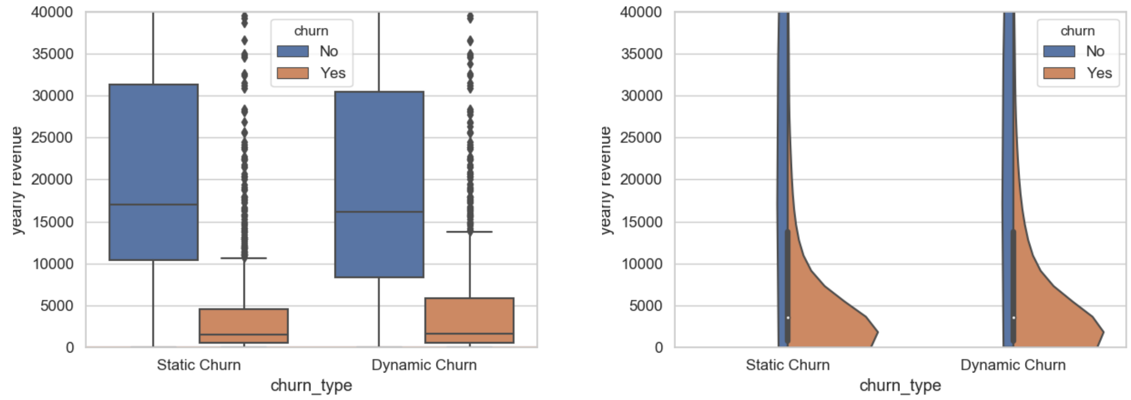


(a) Static Churn (b) Dynamic Churn

Figure 4.10: Percentage non churners and churners for different average number of products

Average yearly revenue

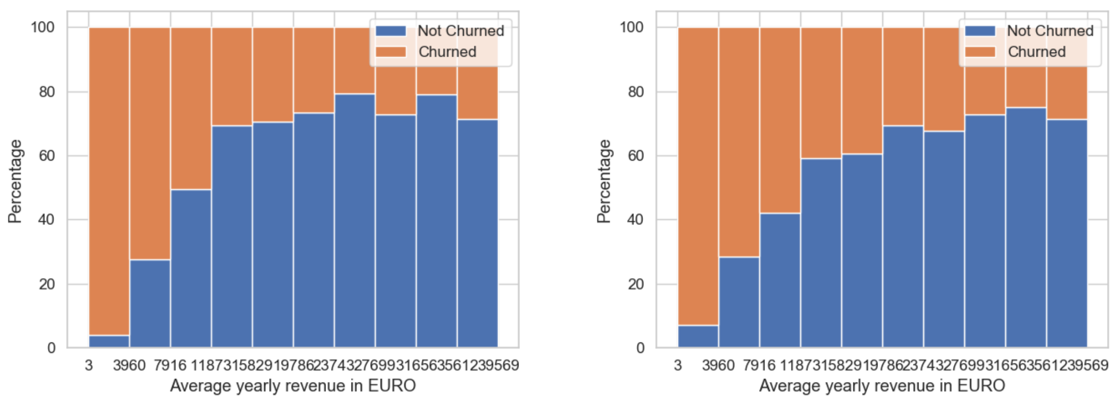
The average yearly revenue is calculated by summing all orders per year and taking the average of the non-zero years for every customer. Figure 4.11a shows that the non-churners tend to have higher yearly revenue than the churners. A reason for this could be that churners usually do not order throughout the year which could lower the yearly average revenue. Another explanation could be that customers with a higher yearly revenue are more loyal to the company because they are more dependent on the products provided by the company. Figure 4.11b shows that the distribution of the average yearly revenue is unimodal in both definitions for the non-churners and the churners.



(a) (b)

Figure 4.11: Distribution of average yearly revenue for non-churners and churners

Figure 4.12 shows that especially the customers with lower average yearly revenue have a high tendency to churn for both the static churn as the dynamic churn. This tendency to churn decreases when the average yearly revenue increases and becomes stable for the static churn and shows a diminishing increase for dynamic churn. It is therefore considered to be a predictor for churn for both definitions.

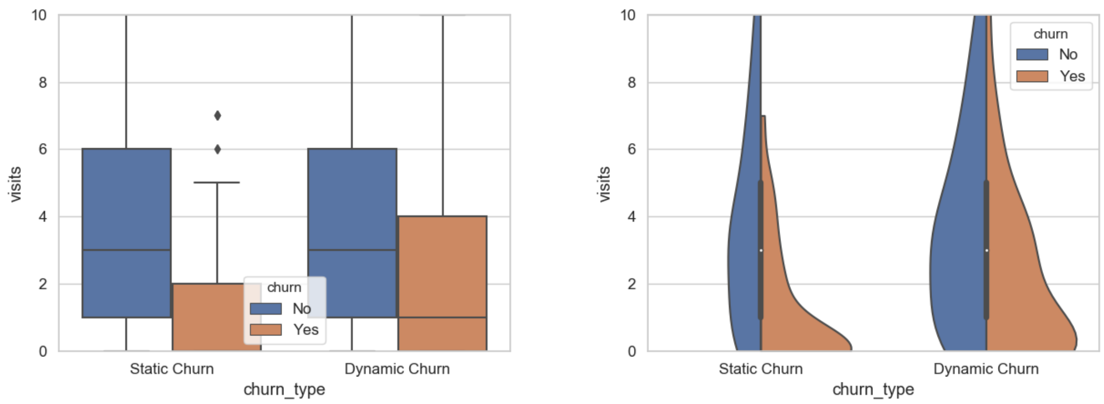


(a) Static Churn (b) Dynamic Churn

Figure 4.12: Percentage non churners and churners for different average yearly revenue

Visits

The salesagents of the company visit the customers throughout the year. Since only the visit data from October 2017 until October 2018 is available, this exploration is based on the customers that are active in this period. Figure 4.13a shows that non-churners are more frequently visited than churners. The difference between non-churners and churners is more present when using the static churn than using dynamic churn. The typical non-churner had three visits, where the typical churner had zero visits for the static definition. For the dynamic definition, the typical non-churner is visited three times and the typical churner is visited one time.



(a) (b)

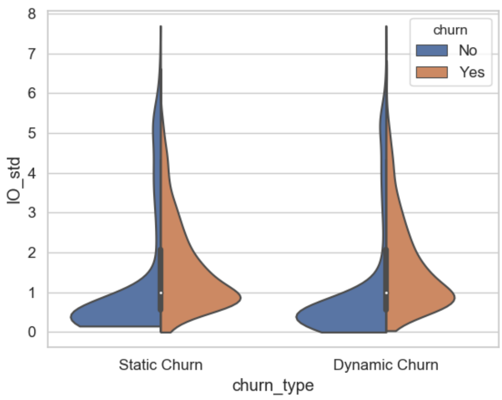
Figure 4.13: Distribution of number of visits for non-churners and churners

Figure 4.13b shows that the distributions of visits for churners and non-churners differs per definition. For the dynamic churn the distributions of non-churners and churners are more similar than the same distributions for static churn. This could indicate that visits are a better predictor when using the static churn definition. Overall the visit data indicates that visits have an effect on churn and are therefore considered a predictor for churn.

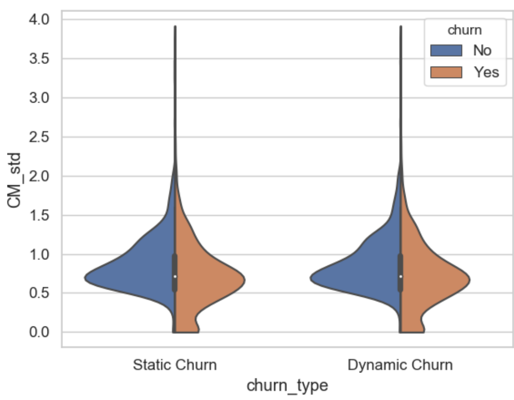
Behavior volatility

The focus in this exploration has been on mean and median values of different variables. However, the deviation of this mean or median can also be informative about the behavior of the customer. It is possible that customers who have a more irregular buying pattern in order interval, revenue and number of products per order have different tendencies to churn. Therefore the predictive power of the volatility of these variables will be explored, where the standardized standard deviation (SSD) of the three variables is studied.

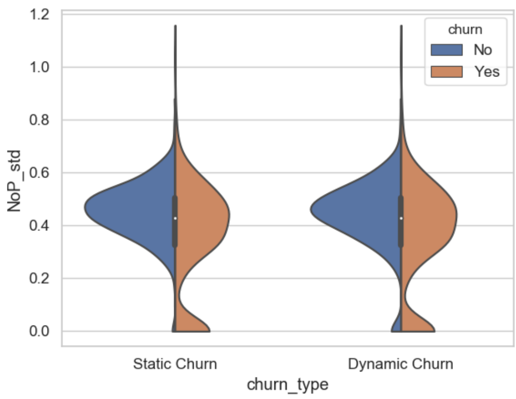
Figure 4.14 shows the distributions of the standardized standard deviation of *Order interval*, *Contribution margin* and *Number of products*. What can be seen from figure 4.14a is that non-churners have a distribution which lies lower than the distribution of the churners in both definitions. This indicates that customers which order pattern is more stable tend to be more loyal. Figure 4.14b and 4.14c show similar distributions for non-churners and churners for the SSD of contribution margin and number of products.



(a) Order Interval SSD



(b) Contribution Margin SSD



(c) Number of Products SSD

Figure 4.14: Standardized Standard Deviation for OI, CM and NoP

This indicates that non-churners and churners both differ in size of order in revenue and number of products in a similar way. Based on this exploration, the order interval SSD seems usable as predictor and both revenue SSD and number of products SSD do not seem usable as predictor.

Conclusion

Table 4.8 is an overview of the different variables and usability based on the data exploration. What can be seen is that most tested variables seem usable predictors for churn. The *Order interval* is usable as predictor for static churn but not for dynamic churn and the standard deviation of revenue and number of products are not usable for both definitions. Since this is a mean analysis, it is possible that variables which seem irrelevant in this analysis can be relevant indicators for churn in a time series. The variables labeled as used in table 4.8 are used for clustering the customers.

It can also be concluded that the studied variables are very similar for churner and non-churners using both definitions. Since the static definitions has more usable predictors than the dynamic churn, from this point the static definition is used.

Table 4.8: Variables and usability as predictor

Variable	Static Churn	Dynamic Churn
Order interval	Used	Not Used
Revenue	Used	Used
Discount	Used	Used
Number of Products	Used	Used
Average yearly revenue	Used	Used
SSD Order interval	Used	Used
SSD Revenue	Not Used	Not Used
SSD Number of Products	Not Used	Not Used

4.2.5 Data Quality

The quality of the data is considered to be good. There are no missing values in the dataset and data from different datasets can be linked by using the unique identifier which is the same over all datasets. There are however three concerns with the data being, high number of churners, visit data of one year and no data prior to 2015.

As can be seen in figure 4.2, there is a high number of churn in the dataset. This is atypical in churn research, where scarcity of churners is usually the problem. However, because the model will be forecasting churn on a periodic level instead of if a customer will ever churn, this should not be problematic. If this disbalance is problematic, a solution could be to oversample the non-churners.

An unforeseen concern with the visit data is that although it seems that the visit data ranges from 27-08-2015 until 29-09-2029, the completed visits only range from October 2017 to October 2018. This means that the usable visit data is only available for one year. This could decrease the power of the optimization model.

The last concern is that there is no data available before 2015. Therefore, it can not be determined when a customer became a customer when this was before 2015. This limits the possibility for calculating the exact length of customerlife. To overcome this concern, an assumption can be made for the start of the customerlife before 2015.

The effect of these concerns and possible solutions will be visible in the modelling phase.

Chapter 5

Churn and CLV Modelling

This chapter describes the predictive modelling of the churn probability and customer lifetime value of all customers. The chapter starts with the data preparation section which describes how the data is prepared for the modelling. This is followed by the churn modelling section and a customer lifetime value modelling section, where both models are parameterized and trained.

5.1 Data Preparation

This section describes the steps that are taken to prepare the dataset for modelling. This consists of data-selection, -cleaning, -construction, -formatting, variable selection and clustering

5.1.1 Data selection

As addressed in the data quality section, the completed visits are limited to the period October 2017 until October 2018. Since this case study aims to create a prescriptive system for visits, the companies that are not active in the period 01-10-2017 to 31-10-2018 are excluded. 861 companies are excluded resulting in a total of 511 companies in the final data set. Figure 5.1 illustrates which companies are included and excluded.

A large number of the excluded companies are companies with a small lifetime. Since the final dataset consist of monthly data per customer, the impact of the exclusion on the final dataset will be limited. The total number of months for all 1371 customers are 22330 months, where the total number of months for the included customers is 17813 months. This indicates that 20 percent of the months are excluded.

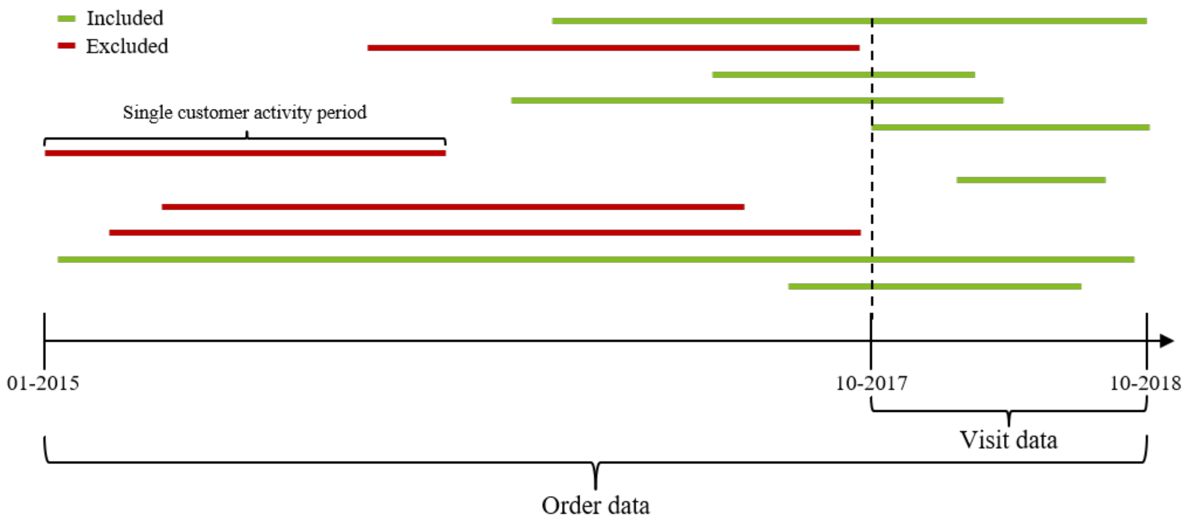


Figure 5.1: Customer inclusion

5.1.2 Data cleaning

The original dataset contains inconsistent data which can be a result of errors.

The first concern is that there are negative values for *Revenue*, *Discount* and *Contribution margin*. These values are considered to be errors based on expert knowledge of the company and the orders containing these values are therefore excluded.

The second concern is that the *Discount* for some orders exceeds the revenue of that order. Similar to the orders containing negative values, these orders are excluded based on expert knowledge.

The third concern is that there are *Visits* with another status than completed. It is unknown if these visits are completed or that they were only scheduled. Therefore these visits are excluded and only the visits with a completed status are included.

The fourth concern is that the dataset contains outlier customers. The outlier customers are identified using the 1.5*IQR rule on the variables from table 4.8 and deleted.

5.1.3 Data construction

The different datasets are merged to construct monthly data per customer. One row contains one month of data for one customer. Table 5.1 shows the included monthly data and the included customer life variables. The monthly variables depend on the orders and activity in the specific month, where the customer life variables depend on the behavior of the customer until that month. All variables are numeric, except for the churned variable which can take either be True or False. This variable is True if the customer churned after the observed month and False if the customer did not churn after that month.

To include information from past months, additional features are constructed based on the original monthly data. The constructed variables are the values of *Total Revenue*, *Total Discount*, *Total Number*

Table 5.1: Variables per row

Included monthly variables	Included customer life variables
CustomerID	*Typical Order interval*
Total Revenue	*Average yearly revenue*
Total Discount	*Average Number of Products*
Total Number of Products	*Average Number of Orders*
Total Number of Orders	*SSD Order interval*
Total Number of Visits	*SSD Revenue*
Churned	*SSD Number of Products*

of Products, *Total Number of Orders* and *Total Number of Visits* for 10 months prior to the observed month. If there were no orders in the months prior to the observed month, this variable will be zero.

Relative change to previous months are included to include deviation in order behavior. Similar to the prior constructed features, are variables included for *Total Revenue*, *Total Discount*, *Total Number of Products* and *Total Number of Orders* for 10 months prior to the observation. The variables are calculated by using formula 5.1.

$$\Delta Rev_{-1} = \frac{Rev_t - Rev_{t-1}}{Rev_{t-1}} \tag{5.1}$$

Table 5.2 shows the constructed variables, these are added to every row in the dataset and results in a total of 104 variables.

Table 5.2: Constructed Variables per row

	Historic monthly variables	Relative change variables
Revenue	$Rev_{-1}, Rev_{-2}, ..., Rev_{-10}$	$\Delta Rev_{-1}, \Delta Rev_{-2}, ..., \Delta Rev_{-10}$
Discount	$Dis_{-1}, Dis_{-2}, ..., Dis_{-10}$	ΔDis_{-1} , $\Delta Dis_{-2}, ..., \Delta Dis_{-10}$
Number of Products	$NoP_{-1}, NoP_{-2}, ..., NoP_{-10}$	ΔNoP_{-1} , $\Delta NoP_{-2}, ..., \Delta NoP_{-10}$
Number of Orders	$NoO_{-1}, NoO_{-2}, ..., NoO_{-10}$	ΔNoO_{-1} , $\Delta NoO_{-2}, ..., \Delta NoO_{-10}$
Number of Visits	$NoV_{-1}, NoV_{-2}, ..., NoV_{-10}$	

For each customer life variables in table 5.2 one binary variable is included, which indicated if the customer has a high or low value for this variable. A value is considered to be low when it is below the median of the variable and high if it is above the median.

5.1.4 Data formatting

For the variable selection Cox time varying variable model is used. The input for this model needs to include a start of the period and an end of the period. The start variable is inclusive and the end variable is exclusive. Therefore, the first month a customer is active is labeled as start 1 and end 2. The next month is labelled start 2 and end 3, which is continued until the last month the customer is active. An example of the data format can be found in appendix B.

5.1.5 Variable selection

The method for variable selection is a variation of the method from Srigopal (2018). As explained in the literature review is the Cox proportional hazard regression not suitable for situation where variables of customers change over time. To overcome this problem, the Cox time varying variables proportional hazard regression will be used instead. This method can account for variables that change over time and are therefor more accurate in predicting the effect of variables. All variables are included in the initial model and the variable with the highest p-value is excluded. Then the model is retrained and the procedure is repeated until all variables are significant.

Table 5.3 shows the significant variables that result from the variable selection. These variables form the base for the churn prediction and indicate the effect of visits on the churn chance. The selected variables consist of a combination of customer life variables which remain the same and monthly data which changes every month. All monthly variables are variables from months prior to the month of interest, making it possible to use the model for predictions.

Table 5.3: Significant variables from Cox time varying variable model

Variable	coef	exp(coef)	se(coef)	z	p	lower .95	upper .95	
$Order\,interval$	0.0036	1.0036	0.0006	5.89	0.00	0.0024	0.0047	***
$NoP\,avg$	-0.7482	0.4732	0.1105	-6.77	0.00	-0.9647	-0.5317	***
$REV\,avg\,high$	-0.8466	0.4289	0.2423	-3.49	0.00	-1.3215	-0.3718	**
NoP_{-3}	0.1086	1.1148	0.0139	7.82	0.00	0.0814	0.1359	***
VIS_{-1}	-0.4541	0.6350	0.2155	-2.10	0.03	-0.8764	-0.0317	.
NoO_{-3}	-0.5340	0.5862	0.1099	-4.85	0.00	-0.7495	-0.3185	***
$REV\,delta_{-8}$	-1.8626	0.1553	0.4477	-4.16	0.00	-2.7402	-0.9851	***
$NoO\,delta_{-9}$	-2.0767	0.1253	0.5299	-3.91	0.00	-3.1154	-1.0381	***

Signif. codes: 0 '***' 0.0001 '**' 0.001 '*' 0.01 '.' 0.05 ' ' 1
Likelihood ratio test = 297.768 on 8 df, p=0.00000

5.1.6 Clustering

As described in the purposed model, the customers are clustered to further specify the effect of visits on the chance of churn for different customers. These clusters are made based on the variables that are described as usable predictors in table 4.8 of the data exploration section. To be able to compare the effect of clustering on the churn prediction, the customers will be clustered based on Gaussian mixture modelling clustering and k-means clustering.

To facilitate clustering, the variables are first normalized. This is necessary because both algorithms use distance as determining factor for which cluster a customer belongs to. Without normalization, the variables with high values and therefore higher absolute deviation would become more important than variables with low values and low absolute deviation. Normalizing eliminates this problem and makes every variable equally important. The data is normalized using equation 5.2, where all features are normalized independently.

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{5.2}$$

After the clusters are made, the data is transformed back to the original values and will be validated by company experts.

K-means clustering

The first clustering method used is k-means clustering. The free parameter is the amount of clusters k, which needs to be set. This cannot be done by choosing the k which results in the lowest distance from the centroids to the observations because this method will lead to $k = n$ where the distance is zero. k is determined by using the elbow criterion. In this method the optimal trade-off between explained variance and number of clusters is used as number of clusters. The sum of squared error is calculated for a range of k, after which the point is located where adding an extra cluster does not outweigh the decrease in error. This point is visible in a graph as the point with the largest bend.

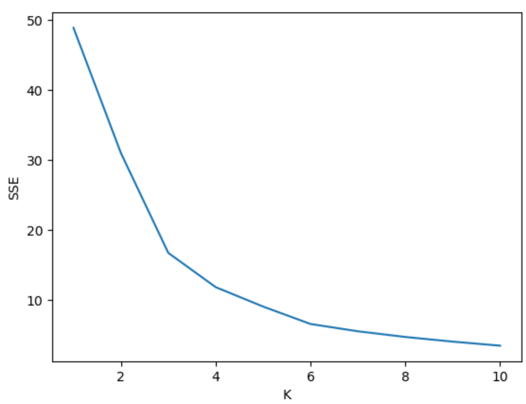
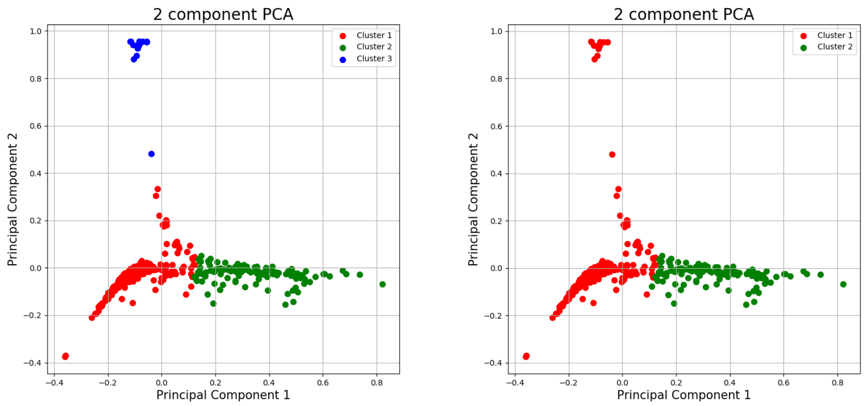


Figure 5.2: Sum of squared errors for k clusters

Figure 5.2 shows the sum of squared errors for different number of clusters. The elbow point is located at $k = 3$, which based on the elbow criterion would be the optimal number of clusters.

Using $k = 3$ leads to the clustering in figure 5.3b, where the observations are projected on the two principal components. As can be seen, there is a small number of observations in cluster 3. A small number of observations combined with high class imbalance is problematic for modelling. The reason for this is that this can result in clusters without churn observations, which would make it not possible to create a churnmodel for this cluster. Therefore $k = 2$ is chosen as optimal k. This clustering is visible in figure 5.3b.



(a) $k = 3$ (b) $k = 2$

Figure 5.3: K-means clustering with (a) $k = 3$ and (b) $k = 2$

Validation of clusters

Figure 5.4 shows how the cluster center of cluster 1 deviates from the average of the non clustered dataset. What can be seen is that cluster 1 consists of customers with a higher than average order interval and a far lower than average order interval deviation. The average contribution margin, revenue, number of products and yearly revenue are close to the original average. This cluster could be labeled as the infrequent but very consistent customer.

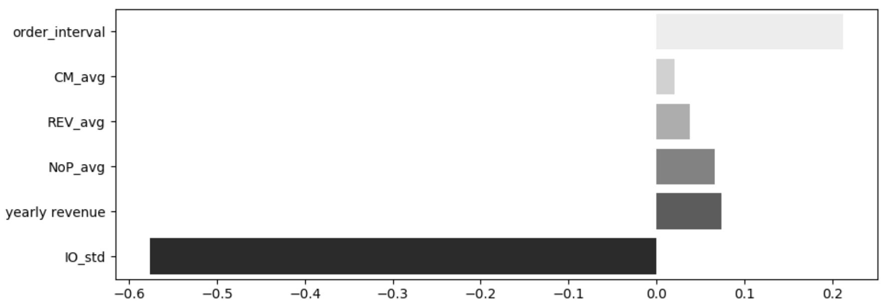


Figure 5.4: Cluster 1 center using KM

Figure 5.5 shows the same figure for cluster 2. The figure shows the opposite deviation as cluster 1 and represents the frequent but very inconsistent customers.

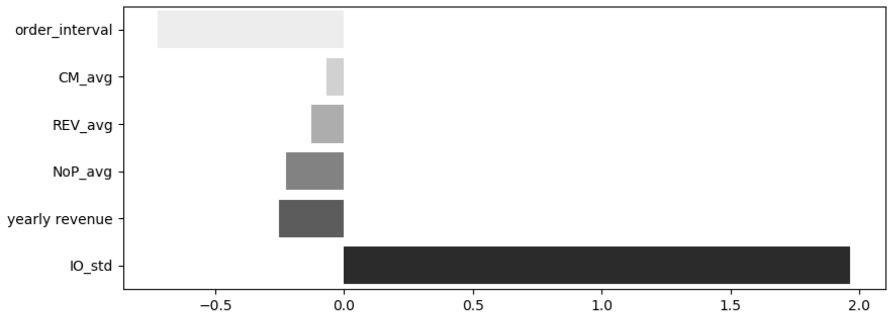


Figure 5.5: Cluster 2 center using KM

The domain experts recognize customergroups with consistent and inconsistent order behavior, but indicate that customers with inconsistent behavior usually also buy less products and have low revenue per order.

Gaussian mixture modelling clustering

The second used clustering method is Gaussian mixture modelling clustering. The free variables in Gaussian mixture modelling are the number of components n and the covariance type. The number of components n is similar to k in k-means clustering and depicts the number of clusters. The covariance type determines the shape, length and direction of the distribution. This can be spherical, tied, diagonal or full, for more information see Appendix C. The Bayesian information criterion, BIC, can be used to compare the different number of components and covariance type. The lowest BIC depicts the best number of components and covariance type.

Figure 5.6 shows the big scores for different number of n and different covariance type. Because the algorithm for GMM optimization is prone to local optima, the clustering is repeated 1000 times and the average values are calculated. This ensures that the found values are valid. Figure 5.6 shows that clustering with $n = 4$ and full covariance leads to the smallest BIC value and can therefore be seen as the optimum setting for GMM.

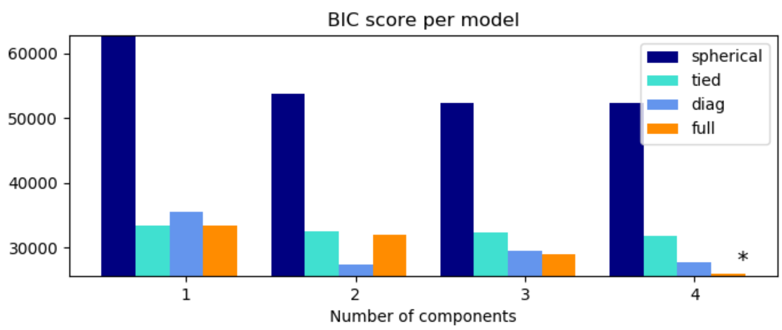


Figure 5.6: BIC for different GMM settings repeated 1000 times

Similar to the k-means clustering, choosing $n = 4$ leads to clusters with a small number of observations.

Therefore the second best settings for the GMM is used, which is $n = 2$ and covariance type is diagonal. This leads to the probabilities as shown in figure 5.7.

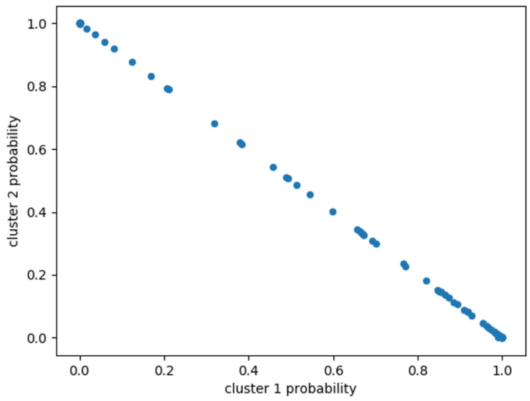


Figure 5.7: Probabilities for cluster 1 or cluster 2 using GMM

Validation of clusters

Figure 5.8 shows how the cluster center of cluster 1 deviates from the average of the non clustered dataset. Cluster 1 consists of low yearly revenue customers with a high order interval, high standard deviation of their order interval and low contribution margin, revenue and number of products per month. These customers can be labelled as small infrequent and inconsistent customers.

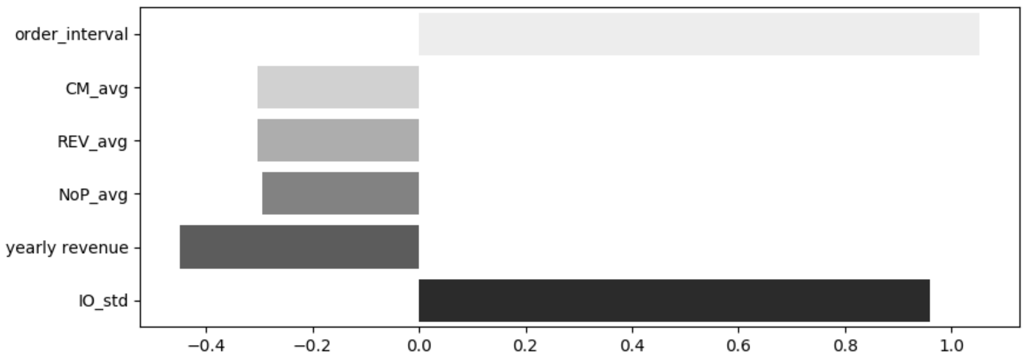


Figure 5.8: Cluster 1 center using GMM

Figure 5.9 describes the cluster center of cluster 2 compared to the average variable value of the non clustered dataset. This cluster consists of large customers with a low order interval, low standard deviation of order interval and high average contribution margin, revenue and number of products per month. This cluster can be labeled as large frequent and consistent customers.

These clusters correspond with the opinion of company experts, which indicate that large frequent and consistent customers behave differently than small infrequent and inconsistent customers.

The k-means with $k = 2$ and GMM clusters with $n = 2$ will be used in the modelling section.

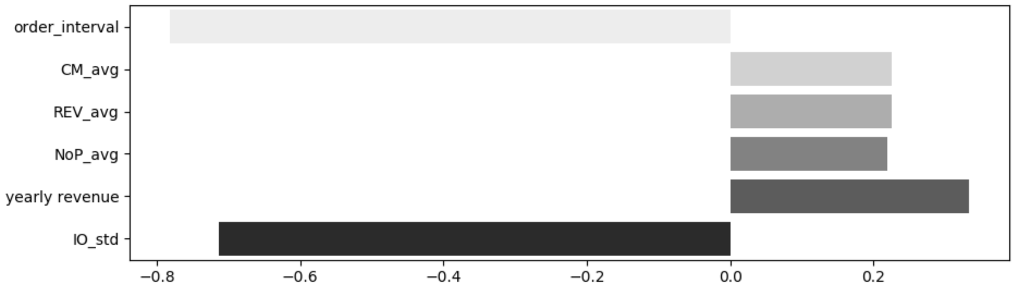


Figure 5.9: Cluster 2 center using GMM

5.2 Churn Modelling

This section describes the modelling phase and uses the inputs from the previous phases to create different churn models. The proposed method in chapter 3 uses the churn model as starting point on which all other variables are calculated. The methods that will be used are decision tree and logistic regression. The churn prediction model can be seen as a classification model which classifies for belonging to the churn class or the non-churn class, where churn is 1 and non-churn is 0. The classification models will be used to generate the probability that an observation belongs to the churn class. This will be used as the churn probability of a customer. The two models will be tested with different parameters using 10-fold cross validation, to find the optimal parameter values per model and are then compared. There are three situations for which churn models are trained, being a non-clustered dataset, a k-means clustered dataset with $k = 2$ and a GMM clustered dataset with $n = 2$. First the decision trees will be modelled after which the logistic regression will be modelled.

5.2.1 Churn modelling with Decision Tree

Parameters

The first churn model which is tested is the decision tree. The decision tree parameters which are tuned are the split criterion and the maximum tree depth. The split criterion determines how a split is generated and can be the Gini impurity measure or the entropy information gain. These measures are calculated using equation 5.3.

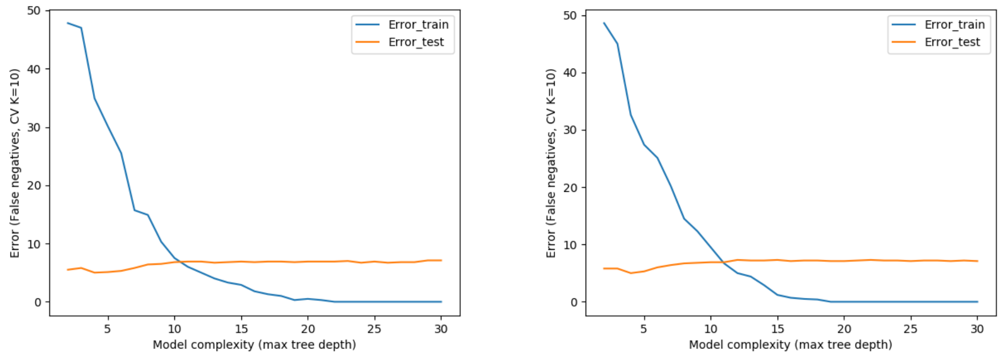
$$Gini(E) = 1 - \sum_{j=1}^{c} p_j^2, \quad Entropy(E) = \sum_{j=1}^{c} p_j \log p_j \tag{5.3}$$

The three maximum depth determines the maximum amount of times a set of observations is split before reaching the end not. This prevents the three from growing until every node consists only one observation, with a high chance of overfitting.

Model training

These decision trees are trained for maximum depths ranging from 2 to 30 with both the Gini measure and the entropy measure. This results in 3*2*29 = 174 decision trees that need to be trained. These trees are evaluated based on the false negatives, where a lower number of false negative is better. The reason for this is that it is preferred to have a model which can correctly classify the churn observations.

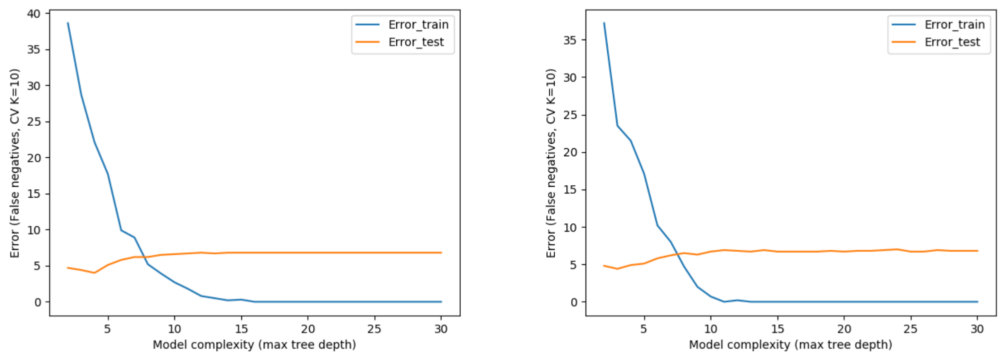
The non-clustered situation is modelled first. Figure 5.10a shows the false negative for the different maximum depths and using entropy for splitting and figure 5.10b shows the false positives when using the Gini criterion for splitting. What can be seen is that for both splitting criteria a maximum depth of 4 results in the lowest false negatives.



(a) Entropy, $Depth_{optimal} = 4$ (b) Gini, $Depth_{optimal} = 4$

Figure 5.10: False negative for different maximum depths without clustering

The k-means clustering situation with $k = 2$ is modelled secondly. Figure 5.11a shows the false negative for the different maximum depths and using entropy for splitting and figure 5.11b shows the false positives when using the Gini criterion for splitting. What can be seen is that for the entropy criterion, a maximum depth of 4 performs best and a maximum depth of 3 performs best for the Gini criterion.



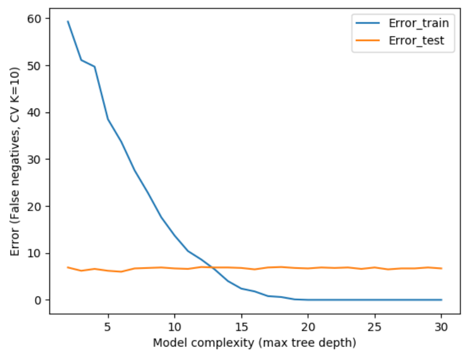
(a) Entropy, $Depth_{optimal} = 4$ (b) Gini, $Depth_{optimal} = 3$

Figure 5.11: False negative for different maximum depths using k-means clustering

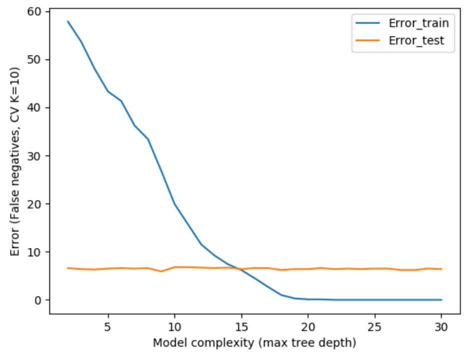
The third situation that is modelled is where Gaussian mixture modelling clustering with $n = 2$ is used. To be able to train the models, the observations are first assigned to the class with the highest probability.

The decision trees are trained using these clusters, where after the churn or non-churn classification is based on the weighted sum of the outcome from the different trees. For more details see chapter 3.

Figure 5.12a and figure 5.12b show the results of the decision trees for the entropy and Gini criterion using GMM clustering and different maximum widths. What can be seen is that the optimal depth for for the decision tree with entropy as splitting criteria and GMM clustering is 6 and the optimal depth when using the Gini splitting criterion is 9.



(a) Entropy, $Depth_{optimal} = 6$



(b) Gini, $Depth_{optimal} = 9$

Figure 5.12: False negative for different maximum depths using GMM clustering

The decision trees with the optimal maximum depth will be used.

5.2.2 Churn modelling with Logistic Regression

The logistic regression does not need parameter tuning but is sensitive to class imbalance. Therefore the logistic regression will be trained using different oversampling ratios. The churn observations will be oversampled using Synthetic Minority Oversampling Technique, SMOTE. This method takes a desired ratio between the majority class and minority class as input and generates extra samples of the minority class by inserting new samples based on existing samples with a random alteration until the desired ratio is reached. The used ratios are: no oversampling for LR, oversampling until 25 percent of the dataset is a churn observation for $LR_{0.25}$ and oversampling until the dataset is balanced for $LR_{0.5}$.

The logistic regressions are therefore trained for the three different situations, where the classification for churn and non-churn are based on the highest probability. The results of the regression are shown in table 5.4.

Table 5.4: Coefficients of logistic regressions with churn=1 as dependent variable

Cluster method	No cluster			K-means		
Model	LR	$LR_{0.25}$	$LR_{0.5}$	LR	$LR_{0.25}$	$LR_{0.5}$
$Customer\,lifetime$	0.0147	0.0509	0.0652	0.0292/0.1298	0.0919/0.1439	0.1319/0.1368
	(0.010)	***(0.002)	***(0.002)	.(0.013)/***(0.024)	***(0.003)/***(0.010)	***(0.003)/***(0.007)
$Order\,interval$	0.0102	0.0934	0.1628	0.0111/0.032	0.0584/0.1120	0.1158/0.1754
	***(0.003)	***(0.004)	***(0.004)	***(0.003)/(0.025)	***(0.003)/***(0.011)	***(0.003)/***(0.010)
$NoP\,avg$	-1.1876	-0.59302	-0.4756	-1.3418/-1.0963	-0.9613/-0.5994	-0.8427/-0.4396
	***(0.084)	***(0.022)	***(0.015)	***(0.125)/***(0.170)	***(0.035)/***(0.065)	***(0.025)/***(0.044)
$REV\,avg\,high$	-0.8017	-0.9090	-1.0634	-0.7986/-1.599	-1.6374/-1.6791	-2.3783/-1.4509
	**(0.241)	***(0.070)	***(0.053)	.(0.342)/***(0.405)	***(0.092)/***(0.201)	***(0.081)/***(0.150)
NoP_{-3}	0.1522	0.1054	0.0863	0.1661/0.1943	0.1568/0.1176	0.1482/0.0716
	***(0.010)	***(0.004)	***(0.003)	***(0.014)/***(0.030)	***(0.005)/***(0.014)	***(0.004)/***(0.010)
VIS_{-1}	-0.8576	-0.87450	-0.7494	-1.098/-0.5142	-0.8603/-0.6330	-0.4386/-0.7432
	***(0.205)	***(0.059)	***(0.044)	**(0.321)/(0.335)	***(0.087)/***(0.143)	***(0.066)/***(0.113)
NoO_{-3}	-1.0584	-0.8758	-0.7826	-1.3217/-1.2692	-1.0753/-1.0646	-1.0541/-0.8686
	***(0.108)	***(0.026)	***(0.018)	***(0.172)/***(0.212)	***(0.036)/***(0.086)	***(0.028)/***(0.055)
$REV\,delta_{-8}$	-1.1680	-1.6798	-1.9460	–0.985/-1.3695	-1.8389/-1.8466	-1.7544/-2.5101
	***(0.306)	***(0.070)	***(0.046)	*(0.382)/*(0.436)	***(0.090)/***(0.218)	***(0.056)/***(0.0182)
$NoO\,delta_{-9}$	-1.1789	-1.8490	-1.7104	-1.1686/-1.4209	-2.2260/-1.2666	-2.3947/-1.2371
	*(0.382)	***(0.090)	***(0.058)	.(0.502)/*(0.577)	***(0.114)/***(0.259)	***(0.082)/***(0.178)

Signif. codes: 0 '***' 0.0001 '**' 0.001 '*' 0.01 '.' 0.05 ' ' 1

Table 5.5: Coefficients of logistic regressions continued with churn=1 as dependent variable

Cluster method		GMM	
Model	LR	$LR_{0.25}$	$LR_{0.5}$
$Customer\ lifetime$	0.2765/0.0317	0.4299/0.0648	0.5420/0.0761
	(0.098)/(0.011)	***(0.018)/***(0.005)	***(0.018)/***(0.003)
$Order\ interval$	-2.4294/0.0088	-2.0386/0.0416	-2.4181/0.0730
	(0.613)/(0.002)	***(0.108)/***(0.003)	***(0.110)/***(0.003)
$NoP\ avg$	-0.5195/-1.0685	-1.6391/-05854	-2.1898/-0.3711
	(0.347)/***(0.089)	***(0.085)/***(0.038)	***(0.089)/***(0.027)
$REV\ avg\ high$	-0.0635/-0.5401	0.5557/-0.6105	1.1895/-0.5613
	(1.164)/.(0.261)	*(0.206)/***(0.118)	***(0.205)/***(0.089)
NoP_{-3}	0.0876/0.2149	0.2388/0.1664	0.3074/0.1143
	.(0.043)/***(0.018)	***(0.012)/***(0.008)	***(0.012)/***(0.006)
VIS_{-1}	-0.4735/-0.6606	-0.3325/-0.9017	-0.0874/-0.9639
	(0.634)/*(0.239)	.(0.156)/***(0.113)	(0.158)/***(0.085)
NoO_{-3}	-0.4525/-1.3137	-1.2514/-1.2006	-1.5276/-0.9770
	(0.373)/***(0.137)	***(0.080)/***(0.057)	***(0.085)/***(0.037)
$REV\ delta_{-8}$	-4.3933/-1.0911	-5.8000/-1.3531	-6.9363/-1.4521
	.(2.00)/**(0.314)	***(0.274)/***(0.125)	***(0.275)/***(0.081)
$NoO\ delta_{-9}$	-3.0979/-0.9232	-6.2987/-1.4053	-8.4199/-1.3405
	(1.682)/.(0.382)	***(0.321)/***(0.154)	***(0.339)/***(0.101)

Signif. codes: 0 '***' 0.0001 '**' 0.001 '*' 0.01 '.' 0.05 ' ' 1

5.3 Customer lifetime value Modelling

This section describes the customer lifetime value model, which is the second model of the system. The customer lifetime value is calculated using a combination of the KaplanMeijer estimate and Bayesian updating as described in chapter 3. The KaplanMeier estimate is calculated using all customers. The reason for this is that the data selection as described in section 5.1 creates a bias towards customers with a longer customer life. This is not a problem for the churn modelling since that is calculated for a specific month, but it would bias the KaplanMeier estimate in that it would estimate a longer customer life duration than the real customer life duration. The survival curve is shown in figure 5.13. This shows that there is high chance that customer will have a short customer life and 50 percent of the customer remain customer for more than 7 months. What can be seen in figure 5.13 is that the KaplanMeijer estimate shows high resemblance with the percentage of customers that is still alive at time t. This indicates validity of the KM-estimate.

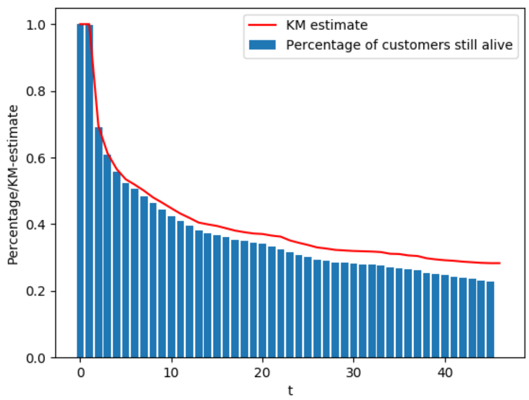


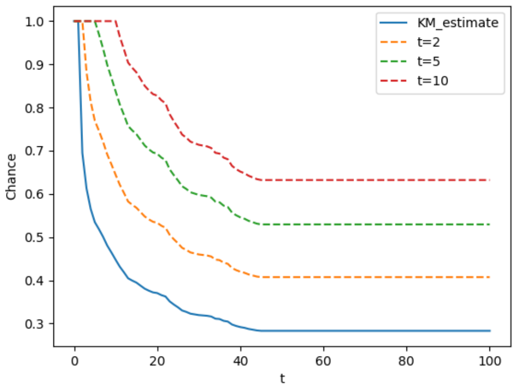
Figure 5.13: KaplanMeier Estimate using all customers

There is no data after 46 months, therefore the assumption is made that the the survival chance after 46 months remains constant.

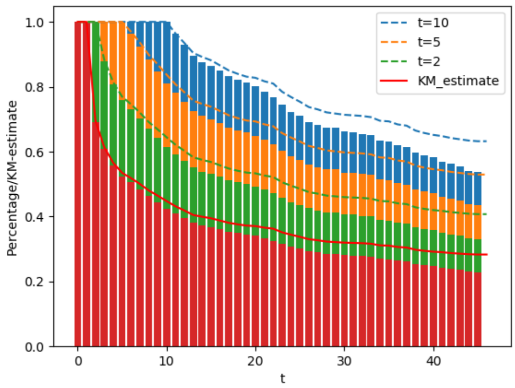
The probability $p_{i,t}$ at t is updated using the method as described in equation 3.8. The resulting curves are shown in figure 5.14. This results in a KaplanMeijer estimate p for a specific customer i at a specific period t. The remaining customer lifetime value is than calculated by using equation 5.4, where the discounted expected future contribution margin is calculated.

$$CLV_i = CM_i \cdot \sum_{t=CCL}^{T} \frac{p_{i,t}}{(1+\frac{r}{12})^{t-1}} \tag{5.4}$$

Where CLV is customer lifetime value, CM_i is the average monthly contribution margin of customer i, CCL is current customer lifetime, T is the forecast range, $p_{i,t}$ is the KaplanMeijer estimate for customer i at period t and r is the discount rate.



(a)



(b)

Figure 5.14: Chance of reaching a period conditioned for t (a) without and (b) with percentage of customers still alive

Chapter 6

System Evaluation

In this chapter, the models will be evaluated. The churn models without clustering, with k-means clustering and with GMM clustering will first be evaluated independently. After this section, the best model of the three different situations will be evaluated. This chapter will end with an evaluation of the prescriptive model as explained in chapter 3.

6.1 Churn Model evaluation

This section will start with the evaluation criteria on which the different churn models are evaluated. Then the churn models with the same clustering method are compared, starting with the churn models without clustering. This is followed by the churn models with k-means clustering and the GMM clustering, after which the best model is chosen. This section ends with an interpretation of the models.

6.1.1 Evaluation criteria

The churn model will be evaluated using the area under the receiver operator curve (AUC), precision, recall and the F_1 score. These are all classification measures and measure the performance of the model. The measures are based on the confusion matrix as shown in figure 6.1, where the predicted values are compared with the actual values. The positive class in this study is the minority churn class and the negatives are the non-churn observations.

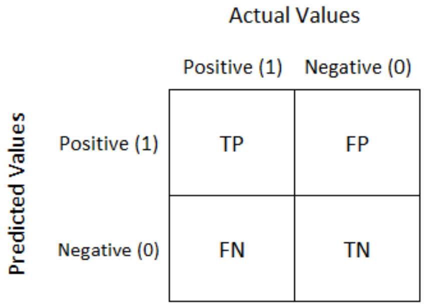


Figure 6.1: Confusion matrix

The precision as defined in equation 6.1 describes the number of correctly classified churners in the observation which are classified as churners. This measure indicates the probability that an observation which is classified as churn by the model is an actual churn observation. The precision can take a value between zero and one, where higher is better.

$$Precision = \frac{TP}{TP+FP} \tag{6.1}$$

Recall or true positive rate as defined in equation 6.2 measures how much of the churners are captured by the model. This is a measure which shows what portion of the actual churn observations is labelled as such by the models. The recall can take a value between zero and one, where higher is better.

$$TPR/Recall = \frac{TP}{TP+FN} \tag{6.2}$$

Recall and precision usually show opposite behavior and are therefore a trade-off. A model which classifies all observations as churn will have captured all actual churn observation and have a recall of one, but a very low precision. Vice versa will a model that classifies one actual churn observation as churn observation have a precision of one and a low recall. The F_1 score, as shown in equation 6.3, overcomes this problem by calculating a harmonic mean of precision and recall. The harmonic mean is a fair representation of the trade off between precision and recall that punishes for values which are far apart. The F_1 score can take a value between zero and one where higher is better.

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{6.3}$$

The area under the receiver operator curve is the area under the curve formed by different true positive rate (TPR) and false positive rate (FPR) combinations. The TPR is the same metric as recall and the FPR is calculated as equation 6.4. This metric measures how much of the non churn observations are classified as churn observations. The FPR can take a value between zero and one where lower is better.

$$FPR = \frac{FP}{FP+TN} \tag{6.4}$$

Figure 6.2 shows a receiver operator curve and the area under the curve (AUC). The dotted line shows the curve resulting from random guessing and the green line the curve resulting from a model. An optimal model would have an AUC equal to one, which would indicate that the model is perfect in classifying the positive and negatives. The AUC can take a value between zero and one where higher is better and higher than 0.5 indicates that the model is better than random guessing.

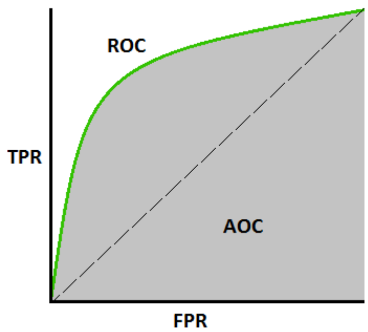


Figure 6.2: Receiver operator curve

6.1.2 Churn model without clustering

Table 6.1 shows the results from the churn models without clustering. The $Logistic\,Regression_{0.50}$ shows the best result for the AUC, which means that this model is the best in classifying both classes. However, this model is outperformed by the other four models based on the F_1 score. The $Decision\,Tree_{Gini}$ shows the best performance based on the F_1 score. Its churn classifications are more accurate than the churn classification of the $Logistic\,Regression_{0.50}$, but it is also able to correctly classify less churn observations.

Table 6.1: Evaluation metrics of the churn models without clustering

Model	AUC	Precision	Recall	F_1
$Logistic\,Regression$	0.902	0.750	0.153	0.254
$Logistic\,Regression_{0.25}$	0.925	0.074	0.673	0.133
$Logistic\,Regression_{0.50}$	0.926	0.035	0.836	0.067
$Decision\,Tree_{Entropy}$	0.887	0.365	0.500	0.422
$Decision\,Tree_{Gini}$	0.844	0.370	0.541	0.434

Figure 6.3 shows the ROC curves of the different models. What can be seen is that the decision trees perform slightly better for the low thresholds where the logistic regression performs better for the higher thresholds. This indicates that the logistic regression is better in identifying the positive observations. However, the low F_1-score of these models indicate that the precision decreases to a low point. Both $Logistic\,Regression_{0.50}$ and $Decision\,Tree_{Gini}$ are seen as good models and will be used in further evaluation.

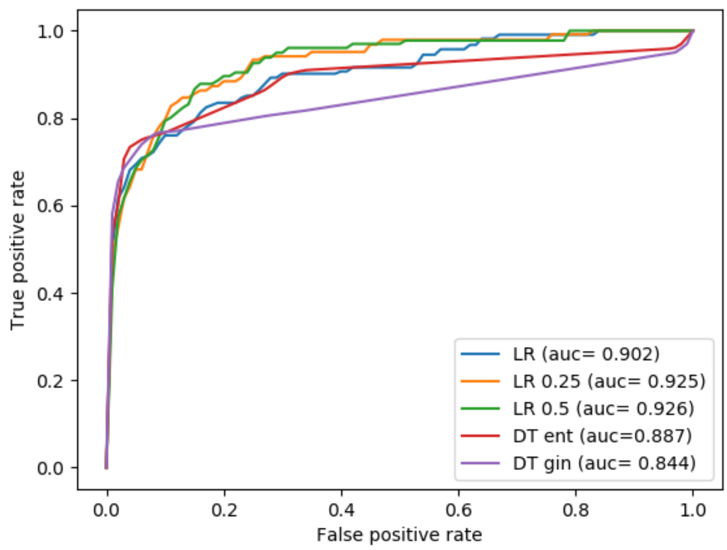


Figure 6.3: Receiver operator curve for models without clustering

6.1.3 Churn model with k-means clustering

Table 6.2 shows the evaluation metrics for the churn models using k-means clustering with $k = 2$. What can be seen is that the $Decision\,Tree_{Entropy}$ has the highest AUC and F_1 score. This indicates that this model has the highest accuracy out of the tested models and that it is best in identifying churners while not labeling too many observations as churn observations.

Table 6.2: Evaluation metrics of the churn models with k-means clustering

Model	AUC	Precision	Recall	F_1
Logistic Regression	0.876	0.727	0.163	0.254
$Logistic\ Regression_{0.25}$	0.880	0.101	0.776	0.133
$Logistic\ Regression_{0.50}$	0.886	0.062	0.850	0.116
$Decision\ Tree_{Entropy}$	0.890	0.299	0.592	0.397
$Decision\ Tree_{Gini}$	0.840	0.245	0.490	0.327

The $Decision\,Tree_{Entropy}$ is the best tested model for churn prediction when using k-means clustering. However, it does not have a higher AUC than the $Logistic\,Regression_{0.50}$ without clustering and it does not have a higher F_1 score than the $Decision\,Tree_{Gini}$ without clustering. Therefore, this model is not seen as an improvement.

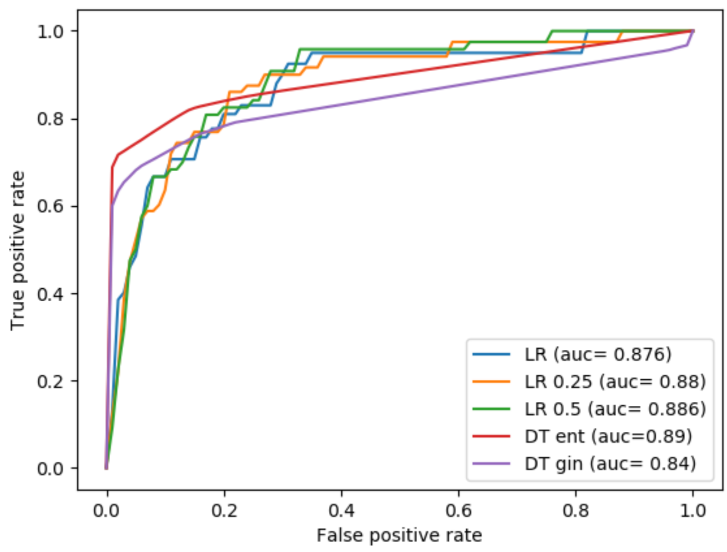


Figure 6.4: Receiver operator curve for models with k-means clustering

6.1.4 Churn model with Gaussian mixture modelling clustering

Table 6.3 shows the evaluations metrics for the churn models using Gaussian mixture modelling clustering with $n = 2$. The $Decision\,Tree_{Gini}$ performs best for both the AUC score and the F_1 score.

Table 6.3: Evaluation metrics of the churn models with GMM clustering

Model	AUC	Precision	Recall	F_1
Logistic Regression	0.955	0.792	0.194	0.311
$Logistic\ Regression_{0.25}$	0.953	0.081	0.806	0.147
$Logistic\ Regression_{0.50}$	0.944	0.064	0.806	0.119
$Decision\ Tree_{Entropy}$	0.942	0.329	0.500	0.397
$Decision\ Tree_{Gini}$	0.957	0.378	0.571	0.455

Figure 6.5 shows that the ROC curves of the models are very similar. The receiver operator curve of the $Decision\,Tree_{Entropy}$ is slightly higher for the low false positive rates, but is crossed by all other curves when the false positive rate increases. The $Decision\,Tree_{Gini}$ curve is close to the optimal curve for all false positive rates. This indicates that the $Decision\,Tree_{Gini}$ is the best model to use.

The AUC of the models with GMM clustering are all higher than the AUC of the models with k-means clustering and without clustering. This is an indication that these are better models. The AUC and F_1 score of the $Decision\,Tree_{Gini}$ are the highest when compared to all other models. For this reason, the $Decision\,Tree_{Gini}$ with GMM clustering is considered to be the best model and will therefore be used as churn model in the prescriptive model. The individual churn models are created using equation 3.2, where the weights from equation 3.1 are the cluster weights from the GMM clustering.

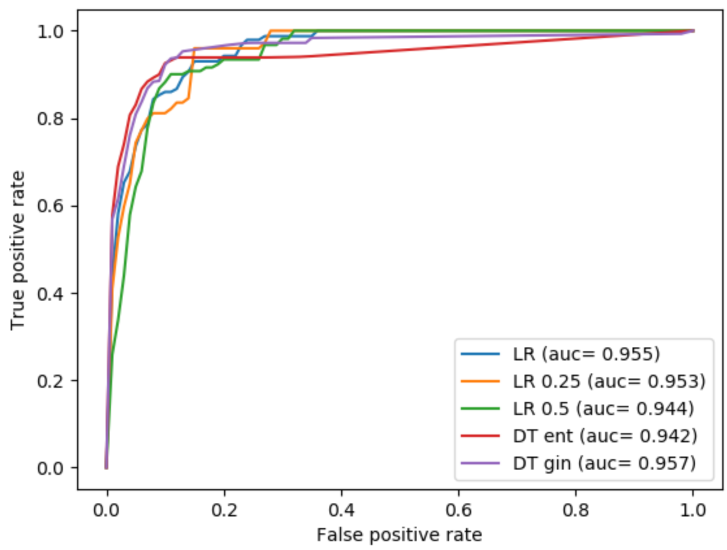


Figure 6.5: Receiver operator curve for models with GMM clustering

6.1.5 Model interpretation

Decision tree

The decision trees for cluster 1 and 2 using the GMM clustering and Gini splitting criterion can be found in appendix D. For cluster 1 the most important splitting variable is the order interval which is splitted at an order interval of 14.5. This means that customers that order at most bi-weekly behave similarly. This corresponds with the data exploration, where was found that customers with an order interval up to 17 days are more loyal than customers with a higher order interval. The second most important factor is the lifetime of the customer for both the customers with an order interval below 14.5 days and the customer with a higher order interval than 14.5 days. The long term customers and short term customers are split. This corresponds with the KaplanMeijer estimate as shown in figure 5.13. 50 percent of the customers churned after 7 months, which makes customers that did not churn until this month more likely to remain customer. After these splits the decision tree uses different variables to continue splitting.

For cluster 2, The first splitting criteria is the difference in revenue with $t-8$. This is a long term difference and indicates that customers with a revenue drop compared to 8 months before behave differently. This can have two reasons. Firstly the difference in revenue compared to $t-8$ is 0 when a customers lifetime is below 8, therefore this is also a split on lifetime. The second reason is that customers who show a large decrease in revenue are also more likely to churn. The second split is on lifetime higher than 41.5 months. This splits the very long term customers from the other customers. As shown in figure 5.13 does the probability customers staying around 40 months stabilize. This indicates that the customers are more likely to stay after this period and therefore show other behavior.

Logistic regression

The estimates of the logistic regression model as found in table 5.4 and 5.5 shows a positive relation between $customer\,lifetime$ and churn. The reason for this could be that customers always churn at their last month and therefore a positive relation is found between higher customer life and churn. The estimates also show a positive relation with $order\,interval$ and NoP_{-3}, which indicates that a higher order interval increases the probability of churning. This corresponds with findings in section 4.2, where churners where associated with higher order intervals. The NoP_{-3} however does not correspond with earlier findings. A negative effect was expected, where a higher number of products would decrease the churn probability. The reason for this could be that the average number of products, $NoP\,avg$, is also included in the model and that NoP_{-3} reduces the effect of $NoP\,avg$. Al other variables show the expected relation between the variable and churn probability. Higher average revenue, more visits, higher number of order, an increase in revenue compared to $t-8$ and an increase in the number of orders compared to $t-9$ are all associated with a decrease in churn probability.

6.2 The prescriptive system

This section describes the prescriptive system with the inputs from the previous section. It starts with the churn model and is followed with the action effectiveness, customer lifetime value and the optimization.

6.2.1 Churn model

The prescriptive system uses the $DecisionTree_{Gini}$ with GMM clustering as the input for the optimization model. The variables used for the churn model allow for a prediction of the churn probability at the end of $t+2$ and a visit to be planned in $t+1$. Figure 3.2 is therefore changed into figure 6.6.

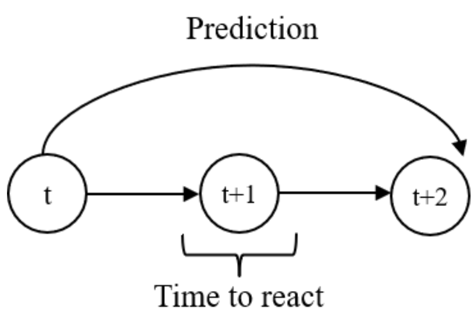


Figure 6.6: Graphical representation of final prediction method

6.2.2 Action effectiveness

The action effectiveness of the visits are calculated using equation 3.4. Because a decision tree is used as underlying model, the effect of a visit when calculated with equation 3.4, can lead to no effect, a positive effect or a negative effect. The reason for this is that it is possible that the visit variable is not used as a splitting variable for some routes of the decision tree. This means that the alteration of the visit variable in $churn\ probability_{i,t+2,a+1}$ does not lead to a different route and that the observation ends in the same end node as when using $churn\ probability_{i,t+2,a}$. The positive or negative effect of the visit is the result of the visit variable leading to a different path with a different end node. The difference in chance between the end note is seen as the effect of the visit variable.

6.2.3 Customer lifetime value

The last input for the optimization model is the customer lifetime value. As described in the modelling section does the KaplanMeijer estimate closely resemble the true probability of customers staying until a specific period. To further assess the effect of the CLV on the prescriptive system, the CLV will be calculated using a minimum, medium and maximum scenario. Where the minimum scenario holds the survival chance of $t = 46$ constant until $t = 75$, the medium scenario holds this survival chance constant until $t = 100$ as shown in figure 5.14. The maximum scenario holds the survival chance of $t = 46$ constant until $t = 125$.

6.2.4 Optimization

Integer programming is used for the optimization using equation 3.10, in which multiple restraints can be used. Examples are the maximum number of visits per month, a minimum period between visiting the same customer and different costs of a visit. The optimization problem can be solved by calculating the return of a visit for all customers and sort these based on descending value. The maximum number of visits are than selected by choosing the visits with the highest return.

The output of the prescriptive system is a list of customers to visit in the coming month.

6.3 Evaluation of the prescriptive system

As described in section 4.1.1 are the success criteria for the system that it increases the the customer lifetime equity and that the outcome is actionable. Therefore, the prescriptive system is evaluated for its monetary results and its practical use.

6.3.1 Monetary Evaluation

The monetary effect of the prescriptive system is evaluated by calculating the customer lifetime equity when using the optimal visits of the system. This is compared with a situation without visits and with the actual visits of the company. The results are shown in table 6.4.

The customer lifetime equity in October 2018 was €74,663,533 if no visits where held. The 85 times the salesagents of the company visited a customer led to an increase of CLE by €437,694 to €75,124,699.

Table 6.4: Evaluation of the prescriptive system

Situation	CLE	Number of Visits	Return of visits
Without visits	€74,663,533	0	-
Using actual visits	€75,124,699	85	€437,694
Using prescriptive system	€77,045,032	19	€2,381,498

If however the prescriptive system was used, the return of the visits would be €2,381,498 making the customer equity €77,045,032. This shows that it is beneficial to use the prescriptive system. It both increases the customer lifetime equity while decreasing the number of visits and therefore increases the efficiency of the visits.

Table 6.5: Customers to visit sorted based on priority using prescriptive system

	Customer to visit this month:
1	549
2	668
3	30564
4	30203
5	30439
6	29895
7	33558
8	30215
9	463565
10	30194
11	471065
12	30458
13	56
14	30028
15	351021
16	351435
17	30226
18	514168
19	687417

The prescriptive system creates an output in the form of a list of customers to visit, sorted on priority. The customer with the highest return of visit is ranked as the first on the list. An example output is shown in table 6.5. In this list all visits with a positive effect are displayed, resulting in 19 visit. This can be generated at the start of every month to determine what customers to visit.

6.3.2 Practical evaluation

Domain experts and end users concluded that the output of the prescriptive system is directly usable in practice. The model can be implemented in the CRM system where the visits are automatically planned and assigned to salesagents. In practice this model can be used to support the visit planning but at this stage can not be considered a replacement due to its accuracy. If accuracy can be improved, complete replacement could be possible in the future.

New customers can be included in a new planning by determining their cluster weights and using the existing models. However, every new and churning customer will affect the performance of the models, because they reduce the relevance of the used trainset. Therefore the clusters, CLV and churn model should be retrained when the evaluation criteria of the models are below a threshold set by the company.

6.4 Robustness check

The prescriptive system depends on the churn model and the customer lifetime value model. It is therefore important to test the robustness of the system in order to evaluate how reliable the outcomes are and how dependent the outcomes are on the accuracy of both models. The system will first be tested for robustness of the churn model and secondly by the customer lifetime value model.

6.4.1 Robustness of the churn model

The robustness of the churn model is tested by comparing the prescriptive system as modelled, with a prescriptive system with an overfit churn model. The used method is an alteration of the lift analysis and shows how the model compares to an optimal model. The overfitted churn model is similar to the used churn model but has a maximum depth of 1000. Due to the overfitting, the model is not usable as predictive model. It can however serve as an optimal model assuming that this model is able to correctly classify the churners and non-churners.

Figure 6.7 shows the effect of a maximum number of visits on the customer lifetime equity. What can be seen is that the prescriptive model follows a similar trend as the optimal model. This indicates that the a similar effect of this visits is found by the prescriptive system and the optimal model. The optimal model shows a higher CLE than the prescriptive system, illustrating that the effect of the visits on the CLE is slightly higher than the prescriptive system predicts. The similar results indicate that the churn model is close to optimal and that the prescriptive system prescribes the correct visits.

The actual visit line in figure 6.7 shows the actual visits as made by the company, where the assumptions is made that the visits with the highest return on visit are made first. This shows that the prescriptive system is better in identifying effective visits than the current procedure.

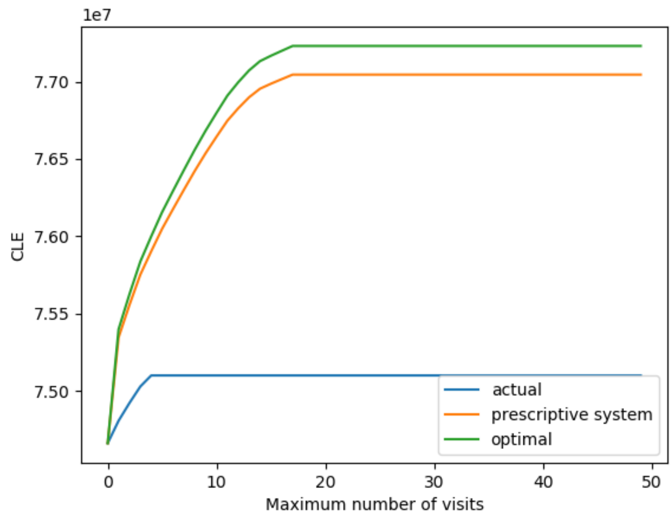


Figure 6.7: CLE of number of visits

6.4.2 Robustness for different customer lifetime values

As described in chapter 3 can the T in equation 3.9 be changed in order to change the CLV for all customers. This will be used to test the robustness of the system. A robust system would work for different CLV calculations and lead to similar resource allocation.

Table 6.6 shows the customers to visit using $T = 75$ as minimum scenario, $T = 100$ as medium scenario and $T = 125$ as maximum scenario. What can be seen is that the three CLV calculations in combination with the prescriptive system results in the same actions with the same priorities. This indicates that the prescriptive system is robust for different CLV calculations. What does change is the return on visits. The reason for this is that the CLV calculation affects the valuation of all customer and a lower T leads to a lower future value. An implication for this is that the accuracy of the CLV calculation is not important for the prescriptive system. The CLV calculation can become important when the costs of the visits are included in the model and only the visits leading to a return higher than the costs are considered.

6.4.3 Conclusion

The conclusion of this section is that the prescriptive model performs close to the optimal curve. This shows that although the prescriptive system underestimates the effect of the visits it still follows the same curve. Also can be concluded that the system is robust for different customer lifetime equity calculations. Different calculations lead to the same visits in the same order. This indicates that the accuracy of the CLV calculation is not important when no costs are included in the model.

Table 6.6: Priority of Customers to visit using different CLV calculations

Customers	Customer to visit this month:		
	Minimum (T=75)	Medium (T=100)	Maximum (T=125)
549	1	1	1
668	2	2	2
30564	3	3	3
30203	4	4	4
30439	5	5	5
29895	6	6	6
33558	7	7	7
30215	8	8	8
463565	9	9	9
30194	10	10	10
471065	11	11	11
30458	12	12	12
56	13	13	13
30028	14	14	14
351021	15	15	15
351435	16	16	16
30226	17	17	17
514168	18	18	18
687417	19	19	19
Return of visits:	€1,337,010	€2,381,498	€3,425,986

Chapter 7

Conclusion

This chapter will conclude the study and start with a summary of research and its outcomes. This will be followed by how this study can be applied in a business context and limitations. This chapter will end with possible future work.

7.1 Summary

This research presents a prescriptive analytics study focused on preventing churn in a non-subscriptual context. A system was created that clusters customers based on their customer characteristics after which churn models are created to predict the churn probability of all customers at the end of $t+2$. The number of visits in $t+1$ are included in the model and serve as a future action that can affect the churn probability of a customer. This churn model is combined with a customer lifetime value model that determines the future value of a specific customer to estimate the expected value loss when the customer churns. The system generates a list of customers to visit based on the increase in customer lifetime equity resulting from the specific visit.

A case study showed that by visiting the customers selected by the proposed system, the customer lifetime equity can be increased by 3.2 percent compared to the customer lifetime equity without visits. The actual visits as done by the company increased the customer lifetime equity by 0.6 percent when compared to the customer lifetime equity without visits. This indicates that value can be generated from using the prescriptive system. The result of the system was achieved with 78 percent less visits when compared to the actual number of visits, which indicates that the system is able to identify effective visits. The study shows that prescriptive resource allocation can be beneficial in a non-subscriptual context for preventing churn and increases efficiency.

7.2 Applications

The system can directly be used as a support system that identifies what customers to visit in the coming month. All current customers are included and can be excluded after a customer has churned. New customers can be assigned to the cluster most applicable, use the models and be included in the visit selection for the next month.

Figure 7.1 shows a possible setup of the system in the visits selection process. The prescriptive system serves as a support system which identifies beneficial visits and is combined with expert knowledge from sales agents. The prescriptive system selects customers and the sales agents select customers to visit which results a number of selected visits. The customers which are selected by both the prescriptive system and the sales agents can have a high priority and the customers selected by only the sales agent or only by the prescriptive selection can have a medium priority.

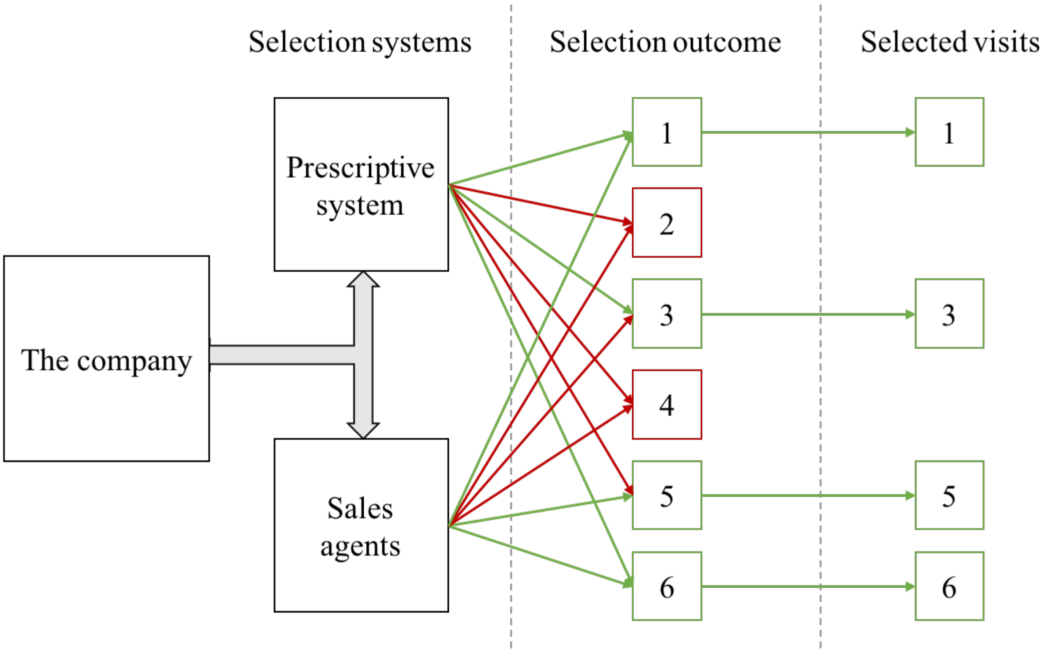


Figure 7.1: Possible practical setup of prescriptive system as decision support system

Since the prescriptive system is a system with static clusters and pretrained models, it is advisable to monitor the performance of the churn model for the newly predicted month. The performance of the churn model will change due to customers churning and the acquisition of new customer, because this will reduce the relevance of the used trainset. The churn model can be revaluated using the evaluation criteria of section 6.1.1. A lower limit for the evaluation criteria needs to be set by the company, which when passed would require retraining of the models.

7.3 Limitations

The first limitation of this study is that the proposed model is tested with limited data with a high imbalance in churn and non-churn observations and a partial visit data set. Although the imbalance was corrected for in the modelling phase, this affects the performance of the model. It is possible that the real effect of visits is higher than predicted in the model because some visits are not logged. The result can be that some visits are not considered to be effective visits by the model when they would be effective visits in reality. The visits that are considered effective visits in this study would still be considered effective but possible more effective visits could be identified. This could be overcome by having a complete visit dataset.

A second limitation is that it is possible that important variables are not studied. The available data and variables are based on expert knowledge, but it is possible that there are unknown relations between other variables and churn probability. The result of this can be that more accurate models are possible if more relevant variables are included.

A third limitation is that the clustering methods use normalization and are based on one distance measure. It is possible that the distance between the different points is no longer valid after normalization which could affect the clusters. It is possible that different distance measures would lead to other clusters that are more valid.

A fourth limitation is that it is not known what would be the historical effect of not visiting a specific customer. The model is trained based on the historical data where an effect is assigned to being visited. It is however not known what would have happened if that customer was not visited in that period. The result of this could be that the effect of a visit is underestimated.

A fifth limitation is that the probability in the decision trees can be biased due to the small number of churn observations. The churn probability is calculated using the proportion of churn observations in the end node. This can however lead to biased results when a small number of observations are in the end node. The result of this could be that the churn probabilities are more extreme. A possible solution for this is to use $Churn\,probability = \frac{n+1}{n+m+2}$ instead (Provost & Fawcett & Provost, 2013), where n are the churn observations in the endnode and m are the non-churn observations in the endnode.

The sixth limitation is that the model selection is based on the area under the receiver curve and F_1 score. The result of using these selection criteria is that the model is selected which has the best trade-off between precision and recall. It is however possible that a selection based on precision or recall solely could be better in specific business situations. In a situation where the costs of missing a churn observations are higher than classify non-churners as churners, it could be beneficial to visit more customers and thereby reducing the probability of missing a churn observation. This would mean that recall should be a criteria. It could however also be possible that non-churn customers are alerted by a visit and activated to rethink their ordering pattern which could result in a customer churning after a visit. In this case it is important to have a high precision. The effect of these visits to non-churners is not studied and therefore the F_1 score and AUC are used. If more information is known, another selection criteria could be selected.

A seventh limitation is that the future cost are not included in the customer lifetime value calculation. The CLV is calculated using the expected future cashflows and therefore represents the profit generated by the customer based on the products. There are however also other costs associated with customers like future actions. This result of this is that the customer lifetime valuation is overestimated and therefore also the monetary effect of the visits is overestimated. By including the future costs a more realistic customer lifetime value can be calculated.

The last limitation is that the chosen churn model, although being the best tested model, does not show a high accuracy when taking the precision and recall into account. It is able to correctly classify 57 percent of all churn observations and 38 percent of the observations classified as churn is an actual churn observation. The result of this is that the system is build on a churn model that is able to identify a portion of the churn observations and some are missed. It is therefore not directly usable as full replacement for manual visit planning but can serve as a support system. Improving the model could make this possible in the future.

7.4 Future Work

Since the research in prescriptive analytics is still in an immature phase, there are possibilities for future work in this field. More research can be done that focuses on using a combination of forecasting and optimization on business processes and show the benefits of this type of analytics. This can improve the awareness of the benefits of prescriptive systems at companies and can help to shift the focus from descriptive and predictive analytics to prescriptive analytics.

Future work can also focus on extensions of the system. The first extension could be to include more different actions. As described is the used method flexible and allows for the use of different retention actions. Example actions are calling the customer, sending emails, send specific marketing content and product demonstrations. By including the data of these actions the prescriptive system can allocate the resources over more actions which could be beneficial. The objective function as shown in equation 7.1 can be used for this extension.

$$max(CLE) = max\sum_{i=1}^{I} CLV_i\,(1-(Churn\,probability_{i,t+1} - \sum_{j=1}^{J} Action\,Effect_{i,j} \cdot a_{i,j})) \tag{7.1}$$

Where j is a specific action, J are the total number of different actions and $a_{i,j}$ is action j for customer i.

The second extension is to include action costs. The used model includes a constraint about the maximum number of visits that are held every month, which needs to be set by the company. When the costs of the actions are also included, this maximum number of visit constraint could be excluded and the system could decide which actions are profitable and which are not. To be able to do this, future work needs to focus on how to measure the cost of an action for different levels of staff. Assuming that low level staff are associated with lower costs than high level staff, it could be possible that some actions lead to profit when executed by low level staff and not when executed by high level staff. Including these costs for actions when executed by different staff levels could improve the decision on what actions to undertake and by what kind of staff. The objective function as shown in equation 7.2 can be used for this extension.

$$max(CLE) = max\sum_{i=1}^{I} CLV_i\,(1-(Churn\,probability_{i,t+1} - Action\,Effect_i \cdot a_i)) - C_a \cdot a_i \tag{7.2}$$

Where C_a is the cost of the action.

The third extension could be to include a response function of the combination of action, customer and salesagent. This response function describes the effect of a specific action on a specific customer when executed by a specific salesagent. By including this function, a more detailed resource allocation system can be created which optimizes for the optimal actions at customer and also dictates which salesagent needs to execute this action. The objective function as shown in equation 7.3 can be used for this extension.

$$max(CLE) = max \sum_{i=1}^{I} CLV_i \left(1 - (Churn\,probability_{i,t+1} - \sum_{s=1}^{S} Action\,Effect_{i,s} \cdot a_{i,s})\right) \tag{7.3}$$

Where s is a specific sales agent, S the total number of different sales agents and $a_{i,s}$ an action performed by a specific salesagent at customer i.

Lastly, future work could also focus on using other definitions of churn. An example is to use partial revenue churn, where customers churned when the revenue decreases by a certain percentage. This could be especially interesting in a retail context without a directly measurable churn moment.

Appendix A

Review protocol prescriptive analytics

2. Review Protocol

This chapter will provide the review protocol which specifies the method used to search, select and extract literature for this literature review. The protocol will follow the guidelines of Kitchenham et al. (2007) and start with the research questions followed by the search strategy, the extraction strategy and the result of the applied review protocol.

Research Questions:

The aim of this literature review is to answer the following questions:

1. For what purpose is prescriptive analytics currently used in business processes?
2. What methods for applying prescriptive analytics are currently used?
3. What prescriptive analytics method performs best in what context?

Search Strategy

Search Terms

To answer the research questions, the following keywords were extracted from the first question: "prescriptive analytics" and "business processes". From the second and third question the keywords, "method" and "performance" were extracted. Since, "method" and "performance" are very general keywords, the decision was made to exclude these from the search terms. Instead, the found literature will be consulted for these keywords. Because prescriptive analytics is usually a combination of predictive analytics and optimization (Soltanpoor & Sellis, 2016), the keyword "prescriptive analytics" is complemented with "predictive analytics" and "optimization". No synonyms were found in the Oxford English Dictionary or in Thesaurus. Table 1 shows an overview of the keywords and search terms.

Table 1: Keywords and Search terms

Index	Keyword	Synonyms, hyponyms and variants	Source
1	Prescriptive Analytics		n/a
2	Predictive Analytics	Supervised Learning, Unsupervised Learning, Neural Networks, Nearest-Neighbour, Decision Tree, Fuzzy Modelling, Fuzzy Clustering, Crisp Clustering, K-means Clustering, Elastic Net, Neural Net, Bayes Classifier	Hastie et al. (2017)
3	Optimization	Linear Programming, Stochastic Programming, Mixed Integer Programming, Dynamic Programming, Swarm Optimization, Ant Colony Optimization Algorithms	Hiller and Lieberman (2015)
4	Business Processes	Business Process, Decision Support	n/a

Both the keywords predictive analytics and optimization led to a large number of articles without application. Therefore, the decision was made to include specific methods for predictive analytics and optimization. The included predictive analytics and optimization techniques are obtained from literature from Hastie, Tibshirani and Friedman (2017), Hiller and Lieberman (2015) and Hutchison (2016).

The keyword “Prescriptive Analytics” is used to focus the search on the main subject of interest. Since the terms “Predictive Analytics” and “Optimization” are included to broaden the search, these three keywords, including their synonyms, hyponyms and variants, will have to be present in the search results. Therefore, these keywords will be separated by an “OR” operator, were “Predictive Analytics” and “Optimization” need to be present simultaneously. To focus on applications in business processes, the keyword “Business Processes” needs to be present as well.

This leads to the following query were all synonyms, hyponyms and variants are used:

- Search for: “Business Processes” AND (“Prescriptive Analytics” OR (“Predictive Analytics” AND “Optimization”))

Sources

The previously mentioned search terms will be used to form a reproducible systematic literature review. To get full coverage of the relevant articles, a selection was made based on the number of documents, relevant field coverage and possibility of using advanced search queries. The following sources were selected:

1. **Scopus:** Scopus is a database with 21,950 journals, more than 150 thousand books and more than 8 million conference papers in all fields of study. It claims to be the largest abstract and citation database of peer reviewed literature and is created by Elsevier. Scopus incorporated tools that allow the user to use advanced search and refine techniques, making it a fast and easy to use database.
2. **Web of Science:** Web of Science is a combined citation database of several citation databases with citations of more than 20 thousand Journals, over 80 thousand books and over 180 thousand Conference Papers in all scientific disciplines. It is a subscription based database that is created by Clarivate Analytics.
3. **Springer Journals:** Springer Journals is a subscription based database that aims to provide researchers with access to millions of scientific documents. It covers more than 3,4 thousand Journals, more than 240 thousand Books and over 1 million Conference Papers of all field studied by academics. Springer Journals is owned by Springer and its main focus is on the own published books, providing full text versions.
4. **IEEE Explorer:** The IEEE Explorer digital library provides access to all IEEE journals and technical content. It is a subscription library with over 195 journals, more than 6 thousand Technical Standards and 1,800 Conference Papers. The focus of the IEE content is on technical subjects ranging from bioengineering to computer science. It is owned by the Institute of Electrical and Electronics Engineers.
5. **ACM Digital Library:** The ACM Digital Library is a library of all content published by ACM, full text publications from selected publishers and the ACM guide to Computing Literature. This guide is a bibliographic database focused solely on computer science. The ACM Digital Library contains 1,4 thousand journals, over 170 thousand books and over 25 thousand technical reports. It is owned by the Association for Computing Machinery and full text articles are available for members.

The research of Cavacini (2015) indicates that the databases Scopus and Web of Science have high computer science article coverage combined with high quality indexing. Furthermore, Springer Journals was added to include more books in the search. Lastly IEEE explore and ACM Digital Library were added because of their specific relevance for Computer Science (Brereton et al., 2006). Using these digital sources will give a good coverage of the online available literature on the topic.

Selection Protocol

The list resulting from the queries needs to be narrowed down to a shortlist only containing the most relevant articles. For this a selection protocol is applied which consists of three steps.

The first step is to use the queries specified in the search terms section to compose a combined longlist of the unique articles found at the five sources.

The second step is to use the selection criteria to create a middle list. These selection criteria will function as a filter where only the articles that are in line with the criteria are kept and the other articles are put on the backup list.

Lastly, the middle short list will be manually refined by labelling every article to the research questions, based on the title and abstract. The article can be labelled to one or more research questions if applicable, or not be labelled if the article is not applicable to any of the research questions. The literature without a label will be left out the final short list to create a list of the most applicable literature. The manual refinement will be done by the author of this literature review and done in two stages. In the first refinement stage all the relevant articles are labelled and put in the final short list. In the second round all the articles without a label are rechecked and if found relevant for one of the research questions, labelled and added to the final short list. This way articles are only rejected if they were not found relevant two times. The manually rejected literature will be put on a rejection list and kept to allow for use when deemed useful.

Selection criteria

The selection criteria to be used are based on relevance, quality and usability:

To increase the relevance of the literature:

1. **All search terms have to be in the title, abstract or keywords**
2. **The subject area needs to be "Computer Science" or "Business"**

To increase the quality and accurateness of the literature:

3. **The literature needs to be a journal article, a book or a conference paper**
4. **The literature needs to be published after 2009**

To ensure usability of the literature:

5. **The literature needs to be in English**

Middle list

Table 2 shows the search queries and number of results per database. A total of 92 results were found of which 18 were duplicates and three were not accessible. The used query for the Springer Journals database resulted in over 400 results, therefore the decision was made to keep these results on a separate list which will be used as theoretical background information when needed. The 71 unique results form the middle list and will be manually refined as described in step three of the selection protocol.

Table 2: Search Query Results of search on 10 -03-2018

	Scopus	Web of Science	Springer Journals	IEEE Xplore Digital Library	ACM Digital Library
Search Query	TITLE-ABS-KEY (("Business Process" OR "Business Processes" OR "Decision Support") AND ((“Prescriptive Analytics”) OR (("Supervised Learning" OR "Unsupervised Learning" OR "Neural Networks" OR "Nearest-Neighbor" OR "Decision Tree" OR "Fuzzy Modelling" OR "Fuzzy Clustering" OR "Crisp Clustering" OR "K-means Clustering" OR "Elastic Net" OR "Neural Net" OR "Bayes Classifier") AND ("Linear Programming" OR "Stochastic Programming" OR "Mixed Integer Programming" OR "Swarm Optimization" OR "Ant Colony Optimization Algorithms"))))	TS = (("Business Process" OR "Business Processes" OR "Decision Support") AND ((“Prescriptive Analytics”) OR (("Supervised Learning" OR "Unsupervised Learning" OR "Neural Networks" OR "Nearest-Neighbor" OR "Decision Tree" OR "Fuzzy Modelling" OR "Fuzzy Clustering" OR "Crisp Clustering" OR "K-means Clustering" OR "Elastic Net" OR "Neural Net" OR "Bayes Classifier") AND ("Linear Programming" OR "Stochastic Programming" OR "Mixed Integer Programming" OR "Swarm Optimization" OR "Ant Colony Optimization Algorithms"))))	(("Business Process" OR "Business Processes" OR "Decision Support") AND ((“Prescriptive Analytics”) OR (("Supervised Learning" OR "Unsupervised Learning" OR "Neural Networks" OR "Nearest-Neighbor" OR "Decision Tree" OR "Fuzzy Modelling" OR "Fuzzy Clustering" OR "Crisp Clustering" OR "K-means Clustering" OR "Elastic Net" OR "Neural Net" OR "Bayes Classifier") AND ("Linear Programming" OR "Stochastic Programming" OR "Mixed Integer Programming" OR "Swarm Optimization" OR "Ant Colony Optimization Algorithms"))))	(("Business Process" OR "Business Processes" OR "Decision Support") AND ((“Prescriptive Analytics”) OR (("Supervised Learning" OR "Unsupervised Learning" OR "Neural Networks" OR "Nearest-Neighbor" OR "Decision Tree" OR "Fuzzy Modelling" OR "Fuzzy Clustering" OR "Crisp Clustering" OR "K-means Clustering" OR "Elastic Net" OR "Neural Net" OR "Bayes Classifier") AND ("Linear Programming" OR "Stochastic Programming" OR "Mixed Integer Programming" OR "Swarm Optimization" OR "Ant Colony Optimization Algorithms"))))	(("Business Process" OR "Business Processes" OR "Decision Support") AND ((“Prescriptive Analytics”) OR (("Supervised Learning" OR "Unsupervised Learning" OR "Neural Networks" OR "Nearest-Neighbor" OR "Decision Tree" OR "Fuzzy Modelling" OR "Fuzzy Clustering" OR "Crisp Clustering" OR "K-means Clustering" OR "Elastic Net" OR "Neural Net" OR "Bayes Classifier") AND ("Linear Programming" OR "Stochastic Programming" OR "Mixed Integer Programming" OR "Swarm Optimization" OR "Ant Colony Optimization Algorithms"))))
Publication Years	2009-2018	2009-2018	2009-2018	2009-2018	2009-2018
Search Space	Title, Abstract, Key	Title, Topic	Metadata	Metadata	Title, Abstract, Key
Literature Type	Journal Articles, Conference Proceeding and Books	Journal Articles, Review, Conference Proceeding and Books	Journal Articles, Conference Proceeding and Books	Journal Articles, Conference Proceeding and Books	Journal Articles, Conference Proceeding and Books
Subject Area	Computer Science	Computer Science, Operations Research, Business	Computer Science	Computer Science	Computer Science
Language	English	English	English	English	English
Results	12	20	n/a	0	60
Unique Results	71				

Short list
For the third step all 71 results are scanned to assess the relevance of the articles. In this, the criteria are:

1. The result describes application of prescriptive analytics to business processes
2. The result describes a method of prescriptive analytics
3. The result compares at least two different methods of prescriptive analytics

Since criteria 1 and 3 cannot be met without criteria 2 being met as well, there are no results that only meet criteria 1 and 3. The manual refinement led to the following result:

Table 3: Number of papers meeting criteria for the short list

Criteria met	Non	1	2	3	1+2	2+3	1+2+3
Number of results	28	1	1	2	16	12	11

The 11 result meeting all three criteria are put on the short list to be examined further and used as a starting point of the literature review. Figure 2 shows the conversion from middle list to short list sorted by publication year. The number of relevant results for the short list suggest that after a small peak in 2010, there has been a more interest in the topic since 2013. The reason for this could be that more and more companies deem prescriptive analytics to be a new way to become competitive or the increased awareness that data analysis and prescriptive analytics can be used to make better use of resources. As seen by the low number of relevant articles it can be said that the field of prescriptive analytics is in a premature phase and more research is needed.

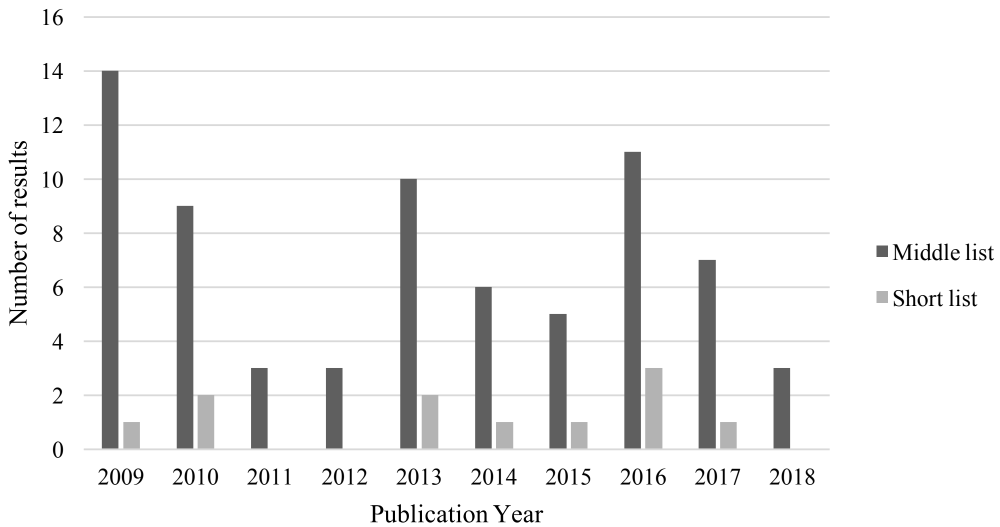


Figure 2: Number of results on middle and short list sorted by publication year

All results on the shortlist has at least one citation and are books, conference papers or articles from journals recognized by the ADBC Master Journal list of 2016 as quality journals or by the 2018 journal quality list of Harzing. This indicates that the articles used for this literature review are of quality.

Extraction Strategy
For the extraction of the literature, the form in table 4 is used. All extraction will be done by the author. The literature shown in table 4 will function as the starting point of the systematic literature review, which will be presented in chapter 4 to 6.

Table 4: Short list

Paper	Author	Year	Method	Application
Advertisement Clicking Prediction by Using Multiple Criteria Mathematical Programming	Lee, Shi, Wang, Lee & Kim	2016	Support vector machine, Support vector regression, Logistic regression, Gaussian process, neural network, regularized regression + Linear programming	Behavioural targeting for online advertisements
Prescriptive Control of Business Processes	Krumeich, Werth & Loos	2016	variety	Process optimization
Early Predictions of Movie Success: The Who, What, and When of Profitability	Lash & Zhao	2016	Regression + Linear programming	Investment decisions in the entertainment industry
Empirical decision model learning	Lombardi, Milano & Bartolini	2017	Neural network, decision trees + Mixed integer non-linear programming, constraint programming, Local search, satisfiability modulo theories	'Hot' and 'cold' job scheduling in the cloud computing industry
Electricity Markets Portfolio Optimization using a Particle Swarm approach	Guedes, Pinto, Vale, Sousa & Sousa	2013	Artificial neural network + Particle Swarm Optimization	Electricity portfolio management for investment bankers
On-line economic optimization of energy systems using weather forecast informatoin	Zavala, Constantinescu, Krause & Anitescu	2009	Gaussian Process + Stochastic Optimization	Energy control
Application of a Hybrid of Genetic Algorithm and Particle Swarm Optimization Algorithm for Order Clustering	Kuo & Lin	2010	Neural network + particle swarm optimization	Reduce production time
An Online Fuzzy Decision Support System for Resource Management in Cloud Environments	Ramezani, Lu, & Hussain	2013	Fuzzy neural networks + multi objective programming	Resource allocation for cloud computing
Supply Chain Analytics	Souza	2014	Regression + Mixed-integer non-linear programming	Supply chain optimization
An Information System for Sales Team Assignments Utilizing Predictive and Prescriptive Analytics	Von Bischhoffshausen, Paatsch, Reuter, Satzger & Fromm	2015	Neural network + Integer Programming	Assigning sales reps to customer accounts
A Decision Support System Using Soft Computing for Modern International Container Transportation Services	Liu, Zhou, Guo, Wang, Pang & Zhai	2010	Linear regression, Exponential smoothing, neural network, genetic algorithm and sequence alignment methods + Linear programming	Stowage planning and shipping line optimization

Appendix B

Data Format

Table B.1: Example data format

Customer	Start	End	Variable 1	Variable 2	...	Variable n	Churn
110	1	2	0	2	...	5	False
110	2	3	3	1	...	8	False
110	3	4	1	4	...	2	True
305	1	2	0	8	...	10	False

Appendix C

Gaussian mixture modelling covariance

Quoted from :whuber (https://stats.stackexchange.com/users/919/whuber), Different covariance types for Gaussian Mixture Models, URL (version: 2018-02-03): https://stats.stackexchange.com/q/326678

A Gaussian distribution is completely determined by its covariance matrix and its mean (a location in space). The covariance matrix of a Gaussian distribution determines the directions and lengths of the axes of its density contours, all of which are ellipsoids.

These four types of mixture models can be illustrated in full generality using the two-dimensional case. In each of these contour plots of the mixture density, two components are located at (0,0) and (4,5) with weights 3/5 and 2/5 respectively. The different weights will cause the sets of contours to look slightly different even when the covariance matrices are the same, but the overall shapes of individual contours will still be similar for identical matrices.

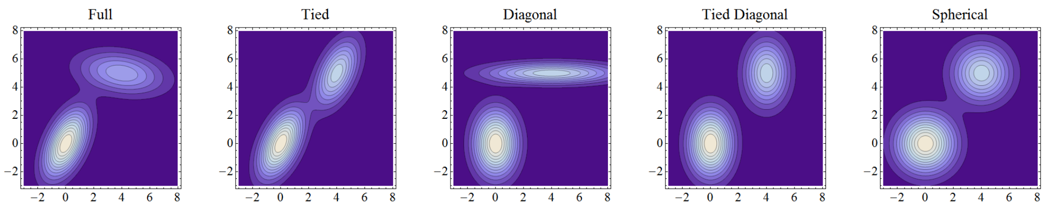


Figure C.1: Covariance types

- **Full** means the components may independently adopt any position and shape.
- **Tied** means they have the same shape, but the shape may be anything.
- **Diagonal** means the contour axes are oriented along the coordinate axes, but otherwise the eccentricities may vary between components.
- **Tied Diagonal** is a ”tied” situation where the contour axes are oriented along the coordinate axes.
- **Spherical** is a ”diagonal” situation with circular contours (spherical in higher dimensions, hence the name).

Appendix D

Decision trees

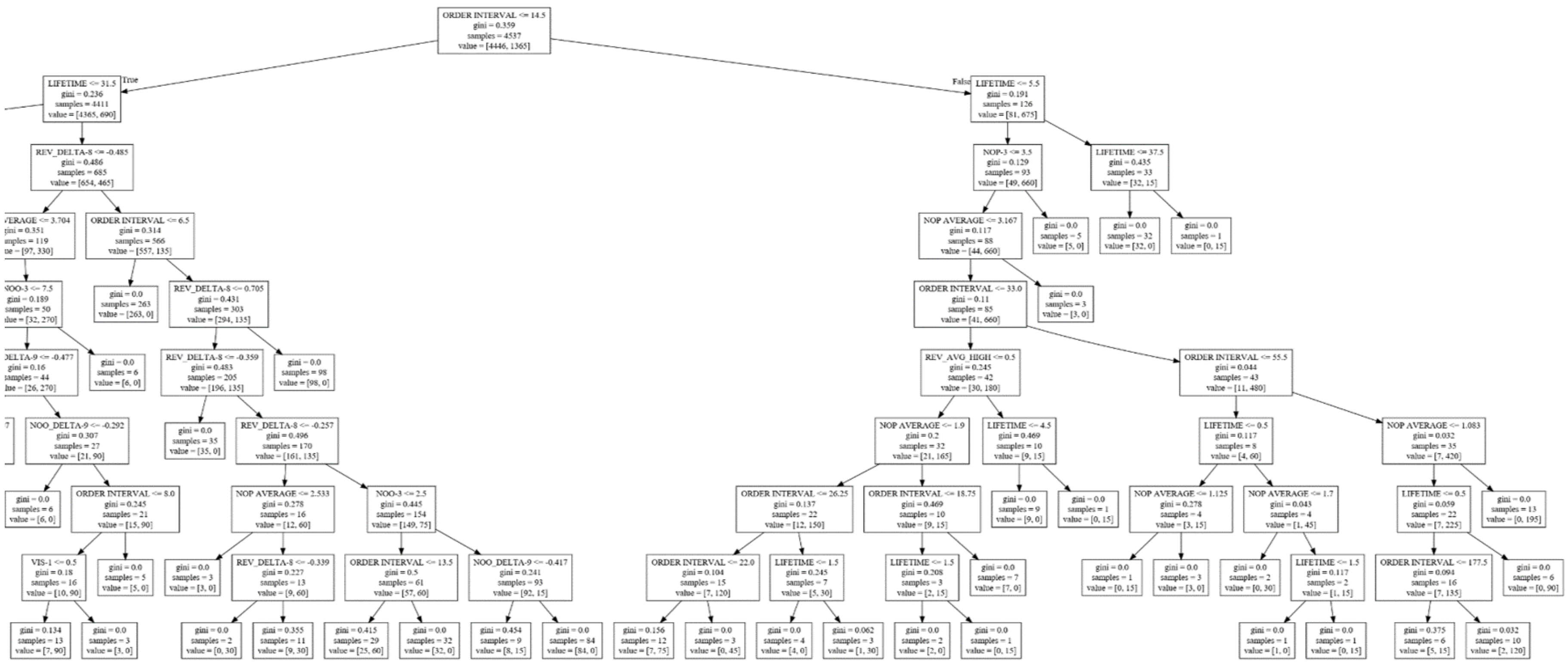


Figure D.1: Decision tree for cluster 1 -Part 1

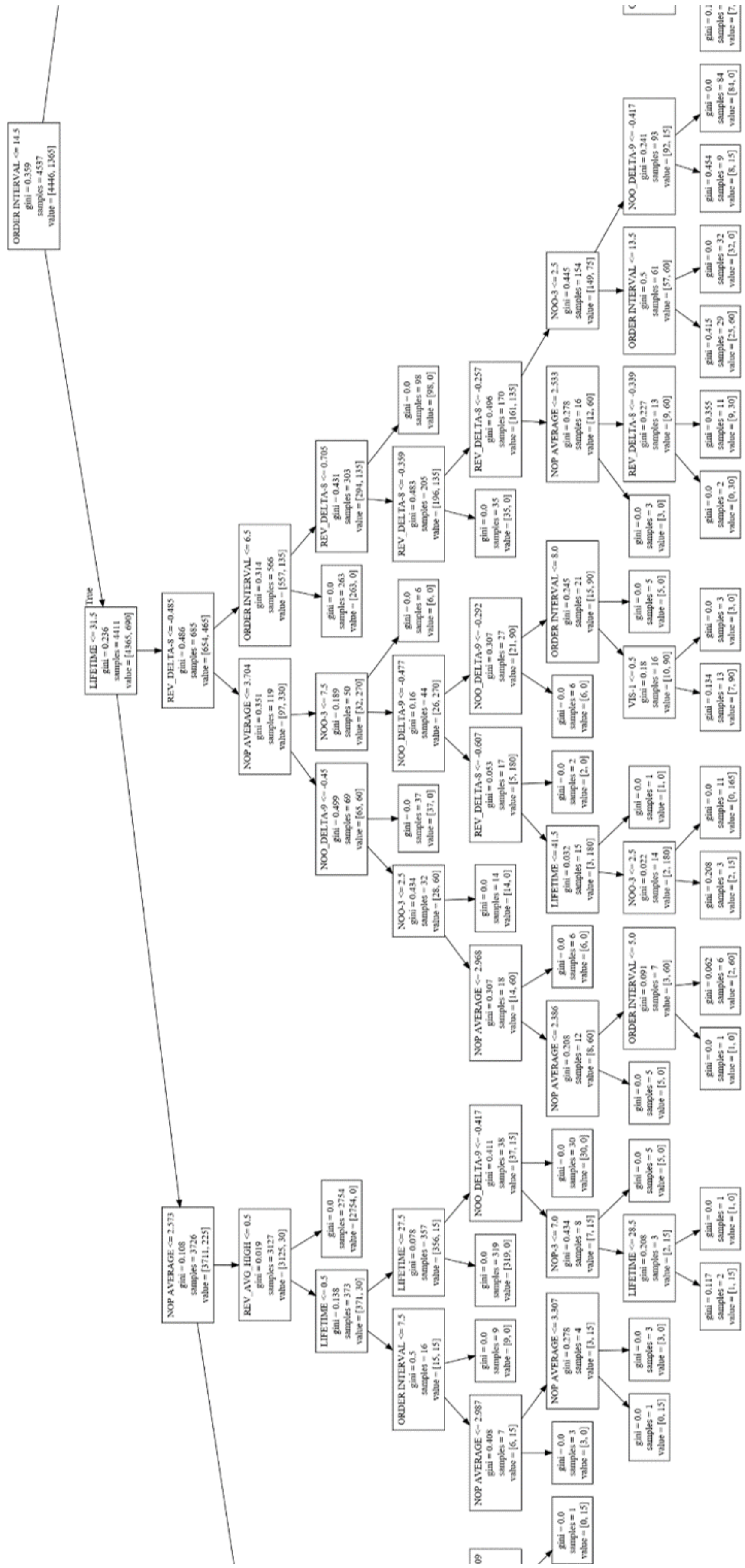


Figure D.2: Decision tree for cluster 1 -Part 2

Figure D.3: Decision tree for cluster 1 -Part 3

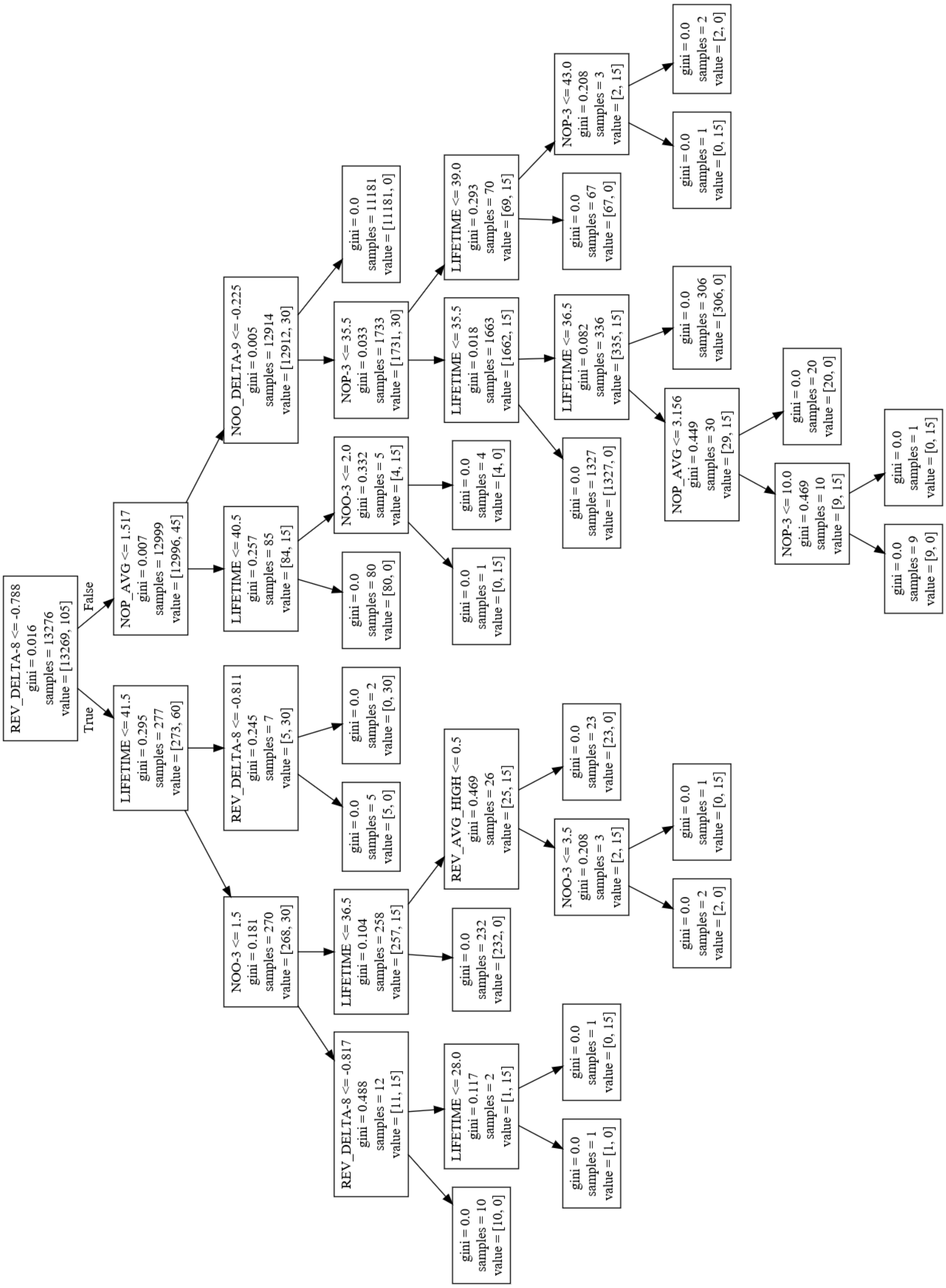


Figure D.4: Decision tree for cluster 2

References

Ahmadi, K., Taherdoost, H., & Fakhravar, S. (2011). A new model for evaluating customer lifetime value in high risk markets. In *International conference on social science and humanity.* Retrieved from https://www.researchgate.net/profile/Kaveh$_A hmadi2/publication/228435700_{A_n}ew_m odel_f or_e valuating$ $new - model - for - evaluating - Customer - Lifetime - Value - in - High - Risk - Markets.pdf$

Baier, M., Carballo, J. E., Chang, A. J., Lu, Y., Mojsilovic, A., Richard, M. J., ... Varshney, K. R. (2012). Sales-force performance analytics and optimization. *IBM Journal of Research and Development*. doi: 10.1147/JRD.2012.2216314

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, *12*(1), 17–30. doi: 10.1002/(SICI)1520-6653(199824)12:1¡17::AID-DIR3¿3.0.CO;2-K

Blattberg, R. C., & Deighton, J. (1996). Manage Marketing by the Customer Euqity. *Harvard business review*, *74*(4), 136. doi: 10.1142/9789814287067_0013

Bolton, R. N. (1998). A Dynamic Model of the Duration of the Customer ' s Relationship with a Continuous Service Provider : The Role of Satisfaction All use subject to http://about.jstor.org/terms A Dynamic Model of the Duration of the Customer ' s Relationship with a Continuo. , *17*(1), 45–65.

Buckinx, W., & Van Den Poel, D. (2005). *Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting.* doi: 10.1016/j.ejor.2003.12.010

Caetano, N., Cortez, P., & Laureano, R. M. (2015). Using data mining for prediction of hospital length of stay: An application of the CRISP-DM methodology. In *Lecture notes in business information processing.* doi: 10.1007/978-3-319-22348-3_9

Chen, S. H. (2016). The gamma CUSUM chart method for online customer churn prediction. *Electronic Commerce Research and Applications*. doi: 10.1016/j.elerap.2016.04.003

Cheng, C.-J., Chiu, S., Cheng, C.-B., & Wu, J.-Y. (2012, 6). Customer lifetime value prediction by a Markov chain based data mining model: Application to an auto repair and maintenance company in Taiwan. *Scientia Iranica*, *19*(3), 849–855. Retrieved from https://www.sciencedirect.com/science/article/pii/S1026309812000107 doi: 10.1016/J.SCIENT.2011.11.045

Colombo, R., & Jiang, W. (1999, 1). A stochastic RFM model. *Journal of Interactive Marketing*, *13*(3), 2–12. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1094996899702367 doi: 10.1002/(SICI)1520-6653(199922)13:3¡2::AID-DIR1¿3.0.CO;2-H

Coussement, K., & De Bock, K. W. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*. doi: 10.1016/j.jbusres.2012.12.008

Coussement, K., & Poel, D. V. d. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2008.07.021

Crone, S. F., Lessmann, S., & Stahlbock, R. (2006). The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*. doi: 10.1016/j.ejor.2005.07.023

Esmaeiligookeh, M., Tarokh, M. J., & Tarokh, M. J. (2013). *"Technical Note" Customer Lifetime Value Models: A literature Survey* (Tech. Rep.). Retrieved from `http://ijiepr.iust.ac.ir/article-1-509-en.pdf`

Evans, J. R., & Lindner, C. H. (2012). Business Analytics: The Next Frontier for Decision Sciences. *Decision Line*, *43*(2), 4–6. Retrieved from `http://www.cbpp.uaa.alaska.edu/afef/business`$_a$*nalytics.htm* doi: 10.1007/978-1-4614-6080-0

Fader, P. S., Hardie, B., & Lee, K. (2004). Counting Your Customers the Easy Way: An Alternative to the Pareto. Retrieved from

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005, 5). Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science*, *24*(2), 275–284. Retrieved from `http://pubsonline.informs.org/doi/10.1287/mksc.1040.0098` doi: 10.1287/mksc.1040.0098

Fü, J., Matzler, K., & Faullant, R. (2006). Asymmetric Effects in Customer Satisfaction. *Annals of Tourism Research*.

Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*. doi: 10.1016/j.ejor.2008.06.027

Groeger, L., & Buttle, F. (2014).). The aggregated CLV of a firms entire customer base equals the firms customer equity. C. (Clv).

Guedes, N., Pinto, T., Vale, Z., Sousa, T. M., & Sousa, T. (2013). Electricity Markets Portfolio Optimization Using a Particle Swarm Approach. *2013 24th International Workshop on Database and Expert Systems Applications*, *2011*, 199–203. Retrieved from `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6621371` doi: 10.1109/DEXA.2013.49

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

Haenlein, M., Kaplan, A. M., & Beeser, A. J. (2007, 6). A Model to Determine Customer Lifetime Value in a Retail Banking Context. *European Management Journal*, *25*(3), 221–234. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0263237307000163` doi: 10.1016/J.EMJ.2007.01.004

Hendrikse, K. (2017). *Predicting customer churn in the healthcare insurance market a case study* (Unpublished doctoral dissertation).

Huang, Kechadi, T. M., Buckley, B., Kiernan, G., Keogh, E., & Rashid, T. (2010). A new feature set with new window techniques for customer churn prediction in land-line telecommunications. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2009.10.025

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2011.08.024

Huigevoort, C. (2015). *Customer churn prediction for an insurance company* (Unpublished doctoral dissertation).

Hung, S.-Y., Yen, D. C., & Wang, H.-y. (2006). Applying data mining to telecom churn. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2005.09.080

Hwang, H., Jung, T., & Suh, E. (2004, 2). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, *26*(2), 181–188. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0957417403001337` doi: 10.1016/S0957-4174(03)00133-7

Kawas, B., Squillante, M. S., Subramanian, D., & Varshney, K. R. (2013). Prescriptive analytics for allocating sales teams to opportunities. In *Proceedings - ieee 13th international conference on data mining workshops, icdmw 2013.* doi: 10.1109/ICDMW.2013.156

Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing Journal*. doi: 10.1016/j.asoc.2014.08.041

Khodabandehlou, S., & Zivari Rahman, M. (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*. doi: 10.1108/JSIT-10-2016-0061

Kim, S. Y., Jung, T. S., Suh, E. H., & Hwang, H. S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2005.09.004

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*. doi: 10.1093/oxfordjournals.pan.a004868

Krumeich, J., Werth, D., & Loos, P. (2016). Prescriptive Control of Business Processes. *Business & Information Systems Engineering*, *58*(4), 261–280. Retrieved from `http://link.springer.com/10.1007/s12599-015-0412-2` doi: 10.1007/s12599-015-0412-2

Kumar, A., Moorthy, J., Aeron, H., & Janakiraman, M. (2010). Application of data mining techniques for customer lifetime value parameters: a review Application of data mining techniques for customer lifetime value parameters: a reviewxxxx) 'Application of data mining techniques for customer lifetime value parameters: a review'. *Article in International Journal of Business Information Systems*, *1*(Y), 0–000. Retrieved from `https://www.researchgate.net/publication/220412657` doi: 10.1504/IJBIS.2010.035744

Kuo, R. J., & Lin, L. M. (2010). Application of a hybrid of genetic algorithm and particle swarm optimization algorithm for order clustering. *Decision Support Systems*, *49*(4), 451–462. Retrieved from `http://dx.doi.org/10.1016/j.dss.2010.05.006` doi: 10.1016/j.dss.2010.05.006

Lash, M. T., & Zhao, K. (2016). Early Predictions of Movie Success: The Who, What, and When of Profitability. *Journal of Management Information Systems*, *33*(3), 874–903. doi:

10.1080/07421222.2016.1243969

Lee, E., Kim, B., Kang, S., Kang, B., Jang, Y., & Kim, H. K. (2018). Profit Optimizing Churn Prediction for Long-term Loyal Customer in Online games. *IEEE Transactions on Games*, *PP*(c), 1–1. Retrieved from `https://ieeexplore.ieee.org/document/8485736/` doi: 10.1109/TG.2018.2871215

Lee, J., Shi, Y., Wang, F., Lee, H., & Kim, H. K. (2016). Advertisement clicking prediction by using multiple criteria mathematical programming. *World Wide Web*, *19*(4), 707–724. doi: 10.1007/s11280-015-0353-1

Lehman, D., & Gupta, S. (2003). Customers as assets. *Journal of interactive marketing*, *17*(1), 9–24.

Liu, D., & Shih, Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information and Management*. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0378720604000394`

Liu, Y., Zhou, C., Guo, D., Wang, K., Pang, W., & Zhai, Y. (2010). A decision support system using soft computing for modern international container transportation services. *Applied Soft Computing Journal*, *10*(4), 1087–1095. doi: 10.1016/j.asoc.2009.06.015

Lombardi, M., Milano, M., & Bartolini, A. (2017). Empirical decision model learning. *Artificial Intelligence*, *244*, 343–367. Retrieved from `http://dx.doi.org/10.1016/j.artint.2016.01.005` doi: 10.1016/j.artint.2016.01.005

Lotfi, F. H., Hatami-Marbini, A., Agrell, J., Aghayi, N., & Gholami, K. (2012). Author's personal copy Allocating fixed resources and setting targets using a common-weights DEA approach q. Retrieved from `http://www.elsevier.com/copyright` doi: 10.1016/j.cie.2012.12.006

Lu, J. (2003). Modeling Customer Lifetime Value Using Survival Analysis An Application in the Telecommunications Industry. *Data Mining Techniques*, *28*, 120–28. Retrieved from `https://pdfs.semanticscholar.org/b497/6a2525d7be2e0f3f51d0fe896e0e56255aad.pdf`

Lu, N., Lin, H., Lu, J., & Zhang, G. (2014). A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*. doi: 10.1109/TII.2012.2224355

Moro, S., & Laureano, R. M. S. (2011). Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology. *European Simulation and Modelling Conference*.

Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education*, *40*(3), 355–371. doi: 10.1023/A:1018755201899

Ozenne, B., Sørensen, A. L., Scheike, T., & Torp-pedersen, C. (2017). riskRegression : Predicting the Risk of an Event using Cox Regression Models. *The R Journal*, *9*(2), 440–460. doi: 10.1113/jphysiol.1992.sp019284

Peppers, D., & Rogers, M. (1995). A new marketing paradigm: Share of customer, not market share. *Managing Service Quality: An International Journal*, *5*(3), 48–51. doi: 10.1108/09604529510796412

Perlich, C., Provost, F., & Simonoff, J. S. (2004). Tree induction vs. Logistic regression: A learning-curve analysis. *Journal of Machine Learning Research*. doi: 10.1162/153244304322972694

Pfeifer, P. E., & Carraway, R. L. (2000). Modeling customer relationships as Markov chains. *Journal of Interactive Marketing*. doi: 10.1002/(SICI)1520-6653(200021)14:2¡43::AID-DIR4¿3.0.CO;2-H

Pladdet, F. (2018). *The use of Prescriptive Analyitcs in Business Processes* (Unpublished doctoral dissertation).

Provost & Fawcett, & Provost, &. F. T., F. (2013). *Data Science for Business. Book.* doi: 10.1593/neo.04781

Ramezani, F., Lu, J., & Hussain, F. (2013). An online fuzzy Decision Support System for Resource Management in cloud environments. *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, 754–759. Retrieved from `http://ieeexplore.ieee.org/document/6608495/` doi: 10.1109/IFSA-NAFIPS.2013.6608495

Rubinstein, P. (1985). A Marketing Approach. *Australian Alcohol/Drug Review*, *4*(2), 267–270. doi: 10.1080/09595238580000441

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2003). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, *68*(1), 109–127. doi: 10.1509/jmkg.68.1.109.24030

Sen, B., & Ucar, E. (2012). Evaluating the achievements of computer engineering department of distance education students with data mining methods. *Procedia Technology*. doi: 10.1016/j.protcy.2012.02.053

Shih, Y. Y., & Liu, D. R. (2008). Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2007.07.055

Soltanpoor, R., & Sellis, T. (2016). Prescriptive analytics for big data. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics).* doi: 10.1007/978-3-319-46922-5_19

Souza, G. C. (2014). Supply chain analytics. *Business Horizons*, *57*(5), 595–605. Retrieved from `http://dx.doi.org/10.1016/j.bushor.2014.06.004` doi: 10.1016/j.bushor.2014.06.004

Srigopal, V. (2018). *Predicting customer churn risks and optimizing retention investments using reliability and maintenance engineering* (Unpublished doctoral dissertation).

Vanderbei, R. J. (2014). *Linear Programming: Foundations and Extensions* (Tech. Rep.). Retrieved from `http://staff.ulsu.ru/semushin/`$_index/_pilocus/_gist/docs/mycourseware/16 - optimeth = ised/2 - reading/lp - foundations - extentions.pdf$

Von Bischhoffshausen, J. K., Paatsch, M., Reuter, M., Satzger, G., & Fromm, H. (2015). An Information System for Sales Team Assignments Utilizing Predictive and Prescriptive Analytics. *Proceedings - 17th IEEE Conference on Business Informatics, CBI 2015*, *1*, 68–76. doi: 10.1109/CBI.2015.38

Yeh, I.-C., Yang, K.-J., & Ting, T.-M. (2009, 4). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, *36*(3), 5866–5871. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0957417408004508` doi: 10.1016/J.ESWA.2008.07.018

Yu, X., Guo, S., Guo, J., & Huang, X. (2011). An extended support vector machine forecasting framework for customer churn in e-commerce. Retrieved from `https://ac.els-cdn.com/S0957417410006779/1-s2.0-S0957417410006779-main.pdf?`$_tid = 34e79546 - 21f5 - 4d6e - b381 - d36b81a45d03acdnat = 1542119415_af6b0685237ecc8b1d9877f6be5c4b9f$ doi: 10.1016/j.eswa.2010.07.049

Zavala, V. M., Constantinescu, E. M., Krause, T., & Anitescu, M. (2009). On-line economic optimization of energy systems using weather forecast information. *Journal of Process Control*. doi: 10.1016/j.jprocont.2009.07.004

Zhao, Y., Li, B., Li, X., Liu, W., & Ren, S. (2005). Customer Churn Prediction Using Improved One-Class Support Vector Machine. , 300–306. Retrieved from http://link.springer.com/10.1007/11527503$_3$6 doi: 10.1007/11527503_36