Chapter 1

Introduction

1.1 Motivation

With larger data sets, higher computing capacities, and better algorithms, opportunities to base business decisions on data grow. To tackle problems in operations management, both mathematical optimization and machine learning algorithms are valuable tools, which have unique strengths. Mathematical optimization can provide decision support by determining optimal decisions for a large number of variables under explicitly defined constraints. Machine learning can provide decision support by learning complex relationships to make predictions.

In this thesis, we apply both optimization and machine learning as well as combinations of both techniques to enable data-driven decision-making in two application areas, customer churn prevention and crew scheduling. In both areas, we consider human behavior and preferences that play a role and are difficult to specify mathematically as patterns or rules.

Customer behavior that leads to churn is a major concern for companies. To avoid churn, taking proactive measures to retain customers who might leave is crucial. Hence, it is valuable for companies to have accurate churn prediction models and churn prevention processes. Available data to build a model typically consists of time series data such as customer purchase histories or customer relationship management activities. Churn processes can be subject to trends or seasonality and the drivers of churn as well as the environment potentially change over time. To address these challenges, we develop an approach to use multiple time slices of data for training a machine learning model, allowing more accurate churn predictions than with a single time slice. For evaluating the business impact of our model in practice, we show how to integrate it into operations and evaluate how well churn can be prevented by basing targeting decisions on our model's predictions.

Another application area that faces complex decisions is crew scheduling. There has been extensive research on using optimization algorithms to create train driver duties which cover train trips while minimizing costs under constraints such as tariff regulations. Apart from cost minimization, other goals, such as driver satisfaction, become more and more important, partly due to labor shortage. Preferences can already be included in many existing models by adding penalty cost terms to the objective functions. However, this approach requires to set a combination of penalty parameters, which planners generally find difficult. We suggest learning their preferences with a machine learning model and integrating it into the optimization model to increase both planner and driver satisfaction.

In this thesis, we build and apply machine learning models to predict human behavior and preferences for two application areas. We develop approaches to improve predictive performance of machine learning models, suggest new ways to combine machine learning and mathematical optimization, and derive concepts for integrating predictions into operations processes. Thereby, we enable and enhance data-driven decision-making.

1.2 Outline and contribution

This thesis is structured into four main chapters (Chapters 2 – 5), followed by a conclusion (Chapter 6) that gives a summary on the results and an outlook on future research.

Figure 1.1 illustrates how chapters are organized by application area (customer churn and crew scheduling) and analytics focus. In each application area, one chapter focuses on building a machine learning prediction model and one chapter focuses on integrating the developed model into decision-making processes. Each chapter is written as a stand-alone paper and consists of an abstract, an introduction, a literature review, a methodological part, a results section, and a conclusion.

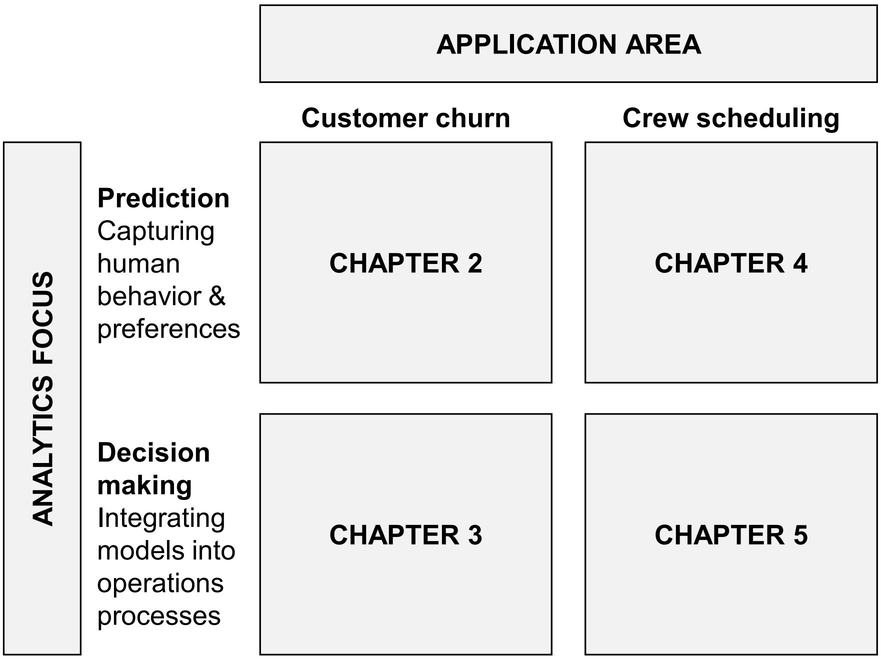

Figure 1.1: Outline of the thesis.

In **Chapter 2**, we present a multi-slicing approach to train a customer churn prediction model. We use data on business-to-business (B2B) customers of a convenience food wholesaler to train and test our approach. We refer to a data extract needed for computing features and labels as a time slice. Our developed multi-slicing approach entails building the training set by stacking feature and label matrices from different time slices. We compare the results of different versions of our multi-slicing approach with a single-slicing approach and show that training on data from multiple time slices improves churn prediction performance. Our analysis reveals that two effects contribute to the performance improvement: An increase in the sample size reduces the risk of overfitting and training on samples from different time slices makes a model more generalizable. We evaluate the impact of varying the number of time slices included in the training set and find that adding additional time slices increases predictive performance but that the marginal benefit of adding time slices is decreasing. Multi-slicing improves out-of-period predictive performance compared to single-slicing in our setting, which makes it a promising approach also for other churn prediction settings, particularly those subject to time-varying conditions. A similar version of Chapter 2 was published in the European Journal of Operational Research (Gattermann-Itschert and Thonemann, 2021).

In **Chapter 3**, we focus on using a churn prediction model for proactive customer retention. By evaluating the feature importances of our churn prediction model for non-contractual B2B customer relationships of a convenience wholesaler, we present new insights on drivers of churn in such a setting. We find that features regarding recency, frequency and monetary value are important and features derived from CRM activity logs also hold valuable information. We illustrate how the results of the machine learning classifier can be integrated into the customer retention process and used for decision-making. In a field experiment, we compare targeting customers with the highest predicted churn probabilities to randomly targeting customers. The results from our field experiment provide evidence on the predictive quality of our model and show that basing targeting decisions on the machine learning predictions reduces the churn rate by more than 10% compared to random targeting. We also demonstrate a positive impact in terms of revenues.

In **Chapter 4**, we focus on another application area, crew scheduling, and develop a machine learning model to learn planners' preferences regarding the output of a crew scheduling optimization tool used by a railway freight carrier. We gather planner feedback on train driver duties with a thumbs up or thumbs down rating option and learn patterns in their feedback by training a random forest model. Using methods from interpretable machine learning, we detect nonlinear relationships between duty characteristics and duty acceptance. This supports our outlined approach for integrating the machine learning predictions into the crew schedule optimization algorithm by replacing the existing construct of linear penalty terms in the objective function. Our trained model predicts the probability of acceptance for unseen duties with high quality and lays the foundation for a beneficial combination of machine learning and optimization in the crew scheduling context.

In **Chapter 5**, we evaluate our approach to include planners' preferences into the crew scheduling optimization model on 20 real-world data sets. Our approach increases the planner acceptance probability of duties compared to the results obtained with the original penalty parameter settings. In the original approach, a complex setting of penalty parameters was necessary and led to multiple planning iterations. Our integrated approach simplifies and improves the decision-making process. We determine the effects on planner acceptance probabilities, actual costs, and duty characteristics and find that a similar cost level can be maintained while improving the mean planner acceptance probability and increasing the share of duties with favorable characteristics compared to the original benchmark schedules.

This chapter is joint work with Laura Maria Poreschack. The study involved several tasks: Conceptual design, literature review, mathematical formulation, implementation of combining prediction and optimization model, experiment runs, and result analysis. I started and conceptually designed the research project in 2018 and Laura Maria Poreschack joined in October 2020 when she started her position at the chair of Supply Chain Management and Management Science. The literature review was mainly done by Laura Maria Poreschack. Implementation-wise, the development of the prediction model and the transformation into a web service was done by me, and the adjustments to the optimization model were done by Laura Maria Poreschack. The mathematical formulation, running the experiments and interpreting the results were joint tasks and both contributed the same share in terms of effort.

In summary, our research contributes to the field of customer churn prediction and prevention (Chapters 2 and 3) and crew scheduling optimization (Chapters 4 and 5). Methodologically, we contribute to research on how to build machine learning models that learn patterns in human behavior (in terms of customer churn) and human preferences (in terms of planner acceptance of crew schedule duties). We develop novel ways of data-driven decision-making by integrating machine learning predictions into operations and optimization processes.

Chapter 2

How training on multiple time slices improves performance in churn prediction

2.1 Abstract

Customer churn prediction models using machine learning classification have been developed predominantly by training and testing on one time slice of data. We train models on multiple time slices of data and refer to this approach as multi-slicing. Our results show that given the same time frame of data, multi-slicing significantly improves churn prediction performance compared to training on the entire data set as one time slice. We demonstrate that besides an increased training set size, the improvement is driven by training on samples from different time slices. For data from a convenience wholesaler, we show that multi-slicing addresses the rarity of churn samples and the risk of overfitting to the distinctive situation in a single training time slice. Multi-slicing makes a model more generalizable, which is particularly relevant whenever conditions change or fluctuate over time. We also discuss how to choose the number of time slices.

2.2 Introduction

In industries with high customer acquisition costs, customer retention is a crucial success factor (Reichheld, 1996; Zeithaml et al., 1996). For targeting customers with retention activities, companies have to identify customers who might churn. Machine learning classification models have been used for churn prediction tasks in different industries. Well known examples of such industries are telecommunications (Verbeke et al., 2012; Coussement et al., 2017), banking (Larivière and Van den Poel, 2005; Kumar and Ravi, 2008), and TV/newspaper subscription (Burez and Van den Poel, 2007; Coussement and Van den Poel, 2009; Ballings and Van den Poel, 2012). By training on a large sample of customers, models learn to identify customers who are likely to stay with the company and those who are likely to churn.

Most churn prediction studies consider a single time window of data and train and test their models on a single time slice. By doing so, they potentially miss extracting relevant information from the data and cannot properly capture that conditions and drivers of churn change over time. Changing environments make it difficult to train a model that upholds its performance in the future (Risselada et al., 2010).

In this study, we structure data into time slices and use multiple time slices for