

Contents

List of Figures	V
List of Tables	VII
List of Abbreviations	IX
1 Introduction	1
1.1 Motivation	1
1.2 Outline and contribution	2
2 How multi-slicing improves performance in churn prediction	7
2.1 Abstract	8
2.2 Introduction	8
2.3 Literature review	10
2.3.1 Churn prediction with classification techniques	10
2.3.2 Model testing approaches	11
2.3.3 Model training with data from multiple periods	12
2.4 Methodology	13
2.4.1 Time slices	13
2.4.2 Multi-slicing	14
2.4.3 Cross-validation with time slices	15
2.5 Empirical study	17
2.5.1 Dataset	17
2.5.2 Churn definition	18
2.5.3 Features	19
2.6 Experimental setting	20
2.6.1 Feature selection	21
2.6.2 Evaluation metrics	21
2.6.3 Model training	22
2.6.4 Cross-validation and statistical testing	24
2.6.5 Single- and multi-slicing versions	24

2.7	Results	25
2.7.1	Classifier selection	25
2.7.2	Performance enhancement with multi-slicing	27
2.7.3	Effects of sample size and including samples from multiple time slices	28
2.7.4	Impact of number of time slices	29
2.8	Conclusion	29
2.A	Features	32
2.B	Details on feature selection	33
2.C	Feature importances	34
3	Proactive retention management based on churn prediction	37
3.1	Abstract	38
3.2	Introduction	38
3.3	Case study	40
3.3.1	Customer churn	41
3.3.2	Customer retention	42
3.4	Churn prediction model	43
3.4.1	Model training and evaluation	43
3.4.2	Feature engineering	45
3.4.2.1	Recency, frequency and monetary value	45
3.4.2.2	Features specific for B2B relationships	46
3.4.3	Performance measures	49
3.4.3.1	Threshold-dependent measures	49
3.4.3.2	Threshold-independent measures	49
3.4.3.3	Domain-specific measures	50
3.4.4	Prediction results	50
3.4.5	Feature importances	51
3.5	Experimental setting	53
3.6	Experimental design	55
3.7	Results and discussion	56
3.7.1	Churn prediction	57
3.7.2	Churn prevention	59
3.7.3	Revenue change	59
3.8	Conclusions	60
3.8.1	Theoretical implications	61
3.8.2	Managerial implications	62
3.8.3	Limitations	63

3.A	Details on model training	65
4	Learning planners' preferences in crew scheduling optimization	67
4.1	Abstract	68
4.2	Introduction	68
4.3	Literature review	70
4.4	Problem setting	71
4.5	Integration approach	72
4.6	Methodology	73
4.6.1	Data collection	73
4.6.2	Data preparation	74
4.6.3	Model building	76
4.6.4	Model evaluation	77
4.7	Results	78
4.7.1	Predictive performance	78
4.7.2	Model interpretation	79
4.7.2.1	Feature importances	79
4.7.2.2	Partial dependence plots	80
4.7.2.3	SHAP values	82
4.8	Conclusion	84
5	Including planners' preferences in crew scheduling optimization	87
5.1	Abstract	88
5.2	Introduction	88
5.3	Literature review	90
5.3.1	Crew scheduling	90
5.3.2	Preference learning	91
5.3.3	Machine learning and combinatorial optimization	91
5.4	Problem setting	93
5.5	Solution approach	93
5.6	Prediction model	94
5.7	Empirical study	94
5.7.1	Data set	95
5.7.2	Benchmark and evaluation method	96
5.8	Results	96
5.8.1	Planner acceptance probability	97
5.8.2	Cost objective	98
5.8.3	Duty characteristics	99

5.9	Conclusion	101
6	Conclusion	105
6.1	Summary of key results	105
6.2	Critical review and future research	107
	Bibliography	110