

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/57982> holds various files of this Leiden University dissertation.

Author: Radosavljevik, D.

Title: Applying data mining in telecommunications

Issue Date: 2017-12-11

Applying Data Mining in Telecommunications

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op maandag 11 December 2017
klokke 10:00 uur
door
Dejan Radosavljevik
geboren te Skopje, Macedonië
in 1975

Promotor: Prof. dr. J.N. Kok Universiteit Leiden
Co-promotor: Dr. P.W.H. van der Putten Universiteit Leiden

Doctorate committee

Prof. dr. A. Plaat	Universiteit Leiden
Prof. dr. T. Bäck	Universiteit Leiden
Prof. dr. H.J. van den Herik (Secr.)	Universiteit Leiden
Prof. dr. C. Soares	Porto University
Dr. M. Baratchi	Universiteit Leiden
Prof. dr. M. Pechenizkiy	Technische Universiteit Eindhoven

This thesis is based upon work supported by T-Mobile Netherlands B.V.

Printing: Ridderprint BV | www.ridderprint.nl

ISBN: 978-94-6299-814-8

© 2017 Dejan Radosavljevik, The Hague, The Netherlands. All rights reserved.

Contents

1	Introduction	7
1.1	Thesis Structure	11
2	The Impact of Experimental Setup in Prepaid Churn Prediction for Mobile Telecommunications: What to Predict, for Whom and Does the Customer Experience Matter?	15
2.1	Introduction	16
2.2	(Prepaid) Churn Modeling	18
2.3	Customer Experience Management	20
2.3.1	Measuring the Customer Experience	20
2.4	Research Setup	21
2.4.1	End to End Data Mining Process	23
2.4.2	Definition of Prepaid Churn and Initial Sample	26
2.4.3	Experiment A: Addition of CEM Parameters	27
2.4.4	Experiment B: Variations in Population Sample	28
2.4.5	Experiment C: Change in Outcome Definition	28
2.5	Results	29
2.6	Discussion and Future Research	33
2.7	Conclusion	36
3	Combining Customer Attribute and Social Network Mining for Prepaid Mobile Churn Prediction	37
3.1	Introduction	38
3.2	Related Work	39
3.2.1	The Call Graph	41
3.2.2	Extended Tabular Churn Models	42
3.2.3	Extended Social Propagation Models	43
3.3	Experimental Setup	45
3.3.1	Operational Definition of Churn	45
3.3.2	Data Set	46
3.3.3	Churn Predictive Models	48

3.4	Results	49
3.5	Conclusions and Future work	51
4	Preventing Churn in Telecommunications: The Forgotten Network	53
4.1	Introduction	54
4.2	Telecom Churn in Literature	55
4.3	Data Set and Methodology	56
4.3.1	Data Set	57
4.3.2	Methodology	57
4.4	Results, Application and Discussion	59
4.5	Limitations and Future Work	63
4.6	Conclusions	64
5	Large Scale Predictive Modeling for Micro-Simulation of 3G Air Interface Load	65
5.1	Introduction	66
5.2	Defining the Load Parameters	68
5.2.1	Output Parameters	68
5.2.2	Input Parameters	70
5.3	Approximating the Load	72
5.4	Building the Load Formulas	74
5.4.1	Tools	74
5.4.2	Process Description	76
5.4.3	Results and Discussion	77
5.4.4	Forecasting the Load	80
5.4.5	Applications- Simulation Scenarios	81
5.5	Limitations and Future Work	83
5.6	Conclusions	84
6	Service Revenue Forecasting in Telecommunications: A Data Science Approach	87
6.1	Introduction	88
6.2	Related Work	89
6.3	Overview of Our Approach	90
6.4	Data Collection and Understanding	91
6.4.1	Data Set Description	93
6.5	Clustering, Modeling and Deployment	95
6.5.1	Clustering the Data	95
6.5.2	Generating the Service Revenue Formulas per Cluster	95
6.5.3	Generating the Monthly State of the Clusters- Inflow, Changes in Contract and Outflow of Customers	97
6.5.4	Generating the Values of Input Parameters- Scaling Factors	98

<i>CONTENTS</i>	5
6.5.5 Putting it All Together	99
6.6 Results	100
6.6.1 Results of Service Revenues Modeling	100
6.6.2 Results of the Forecasting Process for a Full Year	102
6.7 Discussion and Lessons Learned	102
6.8 Limitations and Future work	105
6.9 Conclusion	105
7 Summary and Conclusion	107
Bibliography	114
Samenvatting	125
Curriculum Vitae	129

