

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

Curriculum Vitae

Dejan Radosavljevik was born in 1975 in Skopje, Macedonia. After graduating from a BSc program in Computer Science at the University of Ss. Cyril and Methodius in Skopje in 2001, he has worked as a software developer for several Macedonian companies. In 2009 he completed a Master's degree in ICT in Business with cum laude distinction at Leiden University with a thesis on Prepaid Churn Modeling Using Customer Experience Management Key Performance Indicators. Since then he has worked in multiple positions related to artificial intelligence, data mining and data science at T-Mobile Netherlands B.V., in parallel to working on this PhD research at Leiden University. He currently holds the position of Lead Data Scientist within T-Mobile Netherlands.

List of Figures

| | | |
|-----|--|----|
| 1.1 | CRISP-DM Process Model for Data Mining | 9 |
| 2.1 | Customer Experience Framework for Mobile Telecommunications . . . | 22 |
| 2.2 | Coefficient of Concordance | 24 |
| 2.3 | Coefficient of Concordance of predictors grouped in group 1 for experiment A | 30 |
| 2.4 | Gain chart of models for experiment A, B and C (training set) | 32 |
| 3.1 | Telecom call graph. | 41 |
| 3.2 | Initial energy of the simple and extended propagation technique. . . . | 43 |
| 3.3 | Spreading activation in a weighted graph. | 44 |
| 3.4 | Call Graph Details. | 46 |
| 3.5 | Implementation scenarios. | 48 |
| 3.6 | Gain and Lift chart of all models. | 50 |
| 4.1 | Gain Charts of Models Used | 60 |
| 5.1 | Actual Load vs. Linear approximation | 73 |
| 5.2 | Communication Graph of the Tools used | 75 |
| 6.1 | Overview of the Service Revenue Forecasting Process | 90 |
| 6.2 | The ETL Process in KNIME using RJDDBC | 92 |
| 6.3 | Modeling Workflow in KNIME | 96 |

List of Tables

| | | |
|-----|---|----|
| 1.1 | Mapping of the Focus of the Thesis Chapters to the Stages of the CRISP-DM process | 11 |
| 2.1 | Sample size, churn rate and CoCs in experiments A, B1a, B1b and C . . | 29 |
| 2.2 | Grouping of variables of Model A_Incl_CEM | 31 |
| 3.1 | Social network features used in the extended tabular churn models. . . | 42 |
| 3.2 | Coefficient of Concordance of the scoring and propagation models. . . | 51 |
| 4.1 | List of contractual, demographic and CDR based features | 57 |
| 4.2 | List of network quality features per category | 58 |
| 4.3 | Model Performance | 59 |
| 4.4 | Univariate performance of predictors (CoC) | 61 |
| 5.1 | List of Input Parameters | 71 |
| 5.2 | Regression Modeling Results for Downlink Load (DL) for Country Operator 1 | 78 |
| 5.3 | Regression Modeling Results for Uplink Load (UL) for Country Operator 1 | 78 |
| 5.4 | Regression Modeling Results for Downlink Load (DL) for Country Operator 2 | 78 |
| 5.5 | Regression Modeling Results for Uplink Load (UL) for Country Operator 2 | 78 |
| 5.6 | Regression Modeling Results for Downlink Load (DL) for Country Operator 3 | 79 |
| 5.7 | Regression Modeling Results for Uplink Load (UL) for Country Operator 3 | 79 |
| 5.8 | Regression Modeling Results for Downlink Load (DL) for Country Operator 4 | 79 |
| 5.9 | Regression Modeling Results for Uplink Load (UL) for Country Operator 4 | 79 |

| | | |
|-----|---|-----|
| 6.1 | Inputs used for creating service revenue models | 94 |
| 6.2 | Algorithm performance | 100 |
| 6.3 | Modeling Service Revenue Components | 101 |