Contents

1 Data Processing Procedures in Strategic Planning **9**

2 On Error Probability and Measurement Reliability **15**
2.1 Comparing solutions by distance 15
2.2 Measuring reliability 18
2.3 Some more notions and concepts 22

3 Fuzzy approach to the evaluation of measured values **23**
3.1 Uncertainty in the measurement 23
3.2 Averaging of measured values 24
3.3 Expressing the level of consensus among the measurements 26
3.4 Optimization of the consensus 27

4 Implementation of Algorithm for Missing Values Imputation in Categorical Data Using Association Rules **29**
4.1 Missing values imputation methods 29
4.2 Algorithm for Missing Values Imputation in Categorical Data Using Association Rules 32
4.2.1 Generating Association Rules 32
4.2.2 Obtaining Complete Data Set for Generating Association Rules 33

4.2.3 Algorithm for missing values imputation in categorical data using association rules 33
4.2.4 Testing procedure of missing values imputation algorithm 35
4.3 Implementation of the Algorithm for Missing Values Imputation in Categorical Data using Association Rules 35
4.3.1 Implementation of the Algorithm for Missing Values Imputation in Categorical Data Using Association Rules in MS Excel and MS Visual Basic for Applications. 35
4.3.2 Possibilities of Weka software for missing values imputation in categorical data using association rules 38
4.4 Summary 40

5 Elimination of inaccuracies in the data 41
5.1 Standard ČSN 73 4055 41
5.2 Data used in the calculation 42
5.2.1 Payment zones 43
5.3 Description of the formula 43
5.4 Verification of Model accuracy 44
5.4.1 Example building 45
5.5 Improvement of the input data 46
5.5.1 Location based improvement 46
5.5.2 Time based improvement 47

6 Knowledge Management 49
6.1 Can improved concept of metadata management help to enhance systems as BIM? . . . 49
6.1.1 Curent knowledge management issues 50
6.1.2 Knowledge management 51
6.1.3 Building Information Modeling 52
6.1.4 Knowledge & Metadata Management Systems 53

6.1.5 How to introduce and implement Knowledge management system within a company 55
6.1.6 Concept and role of Metadata 57
6.1.7 Metadata management 58
6.2 Conclusion 62

7 Coupling of Data Mining with Ontology: A Happy Marriage 63
7.1 Introduction 63
7.2 What is an ontology 64
7.2.1 Basic elements of the ontology 65
7.2.2 Types of ontologies 65
7.2.3 Deep and shallow ontologies 68
7.2.4 Reasoning 68
7.3 Use of ontologies in data mining 69
7.3.1 Possible roles of ontology in data mining . . . 70
7.3.2 Ontology learning 72
7.3.3 Application of ontologies in data mining (and vice versa) 72
7.4 Conclusion 76

8 Data Mining With Neural Networks 77
8.1 Introduction 77
8.2 Multilayer Feedforward Network 78
8.2.1 Artificial Neuron 78
8.2.2 Neural Network Learning (Training) 80
8.2.3 Backpropagation Algorithm 82
8.3 Stock Price Prediction and Missing Stock Price Imputation by Neural Networks 83
8.3.1 Predicting Stock Price 84
8.3.2 Network Architecture and Settings 84
8.3.3 Network Inputs 84
8.3.4 Training Data Set and Testing Data Set . . . 85
8.3.5 Why Is the Choice of Inputs so Important . . 85

8.3.6 How to Choose Proper Inputs 86

9 Designing data warehouses for long-term experiments 89
9.1 Terminology 90
9.1.1 OLAP 90
9.1.2 DWH 90
9.2 OLTP vs. DWH 91
9.3 DWH Architectures 92
9.4 Logical Versus Physical Design in Data Warehouses 93
9.4.1 Logical Design 94
9.5 Physical Design 95
9.5.1 Index 98

10 RapidMiner - description of the methods 101
10.1 Outliers 101
10.2 Missing values 103
10.3 TIMODAZ data cleaning procedures in the program RapidMiner 104
10.3.1 MAT - Moving Average, Thresholding 104
10.3.2 TMAT – Thresholding, Moving Average, Thresholding 105
10.3.3 Thresholding 107
10.3.4 Matmata (Moving Average Moving Average Thresholding) 107
10.3.5 Aggregation of data calculate the average hourly and daily average 107
10.3.6 Evaluation procedures 108

List of Figures

4.1 Missing Values Imputation in Categorical Data version 2.00 [27] 36
4.2 Association rules generated using Weka software . . 40

5.1 Example situation for distance approximation 47
5.2 Price indexes of One-dwelling buildings according to CC [34] . 2005 average = 100. Czech Statistical Office. 15.11.2012. Data for graph was retrieved 25. 12. 2012 from `http://www.czso.cz/csu/redakce.nsf/i/ipc_cr` 48

7.1 Ontology types. Adapted from [44] 66
7.2 Steps of the knowledge discovery in databases (KDD). Adapted from [46]. 70
7.3 General framework of data mining with ontologies. Adapted from [47]. 71

8.1 Multilayer feedforward network 79
8.2 Artificial neuron 80
8.3 Sigmoidal function 81
8.4 Comparing network's prediction with real price . . . 88

9.1 One layer solution 92

9.2 Two layer solution 92
9.3 Three layer solution 93
9.4 Logical Versus Physical Design in Data Warehouses 93
9.5 Star Schemas 94
9.6 Snowflake schema 95
9.7 1 layer DWH 96
9.8 2 layer DWH 97
9.9 3 layer DWH 97
9.10 Relationship between an index and a table 98

10.1 Graph Δ indicating outliers 105
10.2 Graph Δ indicating outliers 106

List of Tables

5.1 Sample of Table of Price Indexes 42
5.2 Sample of Table of Price Indexes 43
5.3 Example of the table of population indexes 47

8.1 Unsupervised learning - Training patterns 81
8.2 Supervised learning - Training patterns 81
8.3 Similar inputs in CEZ stock price prediction 86
8.4 Input weights in hidden neurons 86
8.5 Input weights in hidden neurons 87
8.6 Sum of absolute values of network errors with different inputs for two time intervals - Sum of Errors $= \sum |realprice - predictedprice|$ 88

10.1 Assessment process in terms of missing values 108
10.2 Evaluation methods in terms of purity data 108