

Wilfried Grossmann • Stefanie Rinderle-Ma

Fundamentals of Business Intelligence

 Springer

Contents

1	Introduction	1
1.1	Definition of Business Intelligence	1
1.2	Putting Business Intelligence into Context	4
1.2.1	Business Intelligence Scenarios	4
1.2.2	Perspectives in Business Intelligence	6
1.2.3	Business Intelligence Views on Business Processes	8
1.2.4	Goals of Business Intelligence	11
1.2.5	Summary: Putting Business Intelligence in Context	13
1.3	Business Intelligence: Tasks and Analysis Formats	14
1.3.1	Data Task	14
1.3.2	Business and Data Understanding Task	15
1.3.3	Modeling Task	17
1.3.4	Analysis Task	19
1.3.5	Evaluation and Reporting Task	20
1.3.6	Analysis Formats	20
1.3.7	Summary: Tasks and Analysis Formats	24
1.4	Use Cases	24
1.4.1	Application in Patient Treatment	25
1.4.2	Application in Higher Education	28
1.4.3	Application in Logistics	29
1.4.4	Application in Customer Relationship Management	30
1.5	Structure and Outline of the Book	31
1.6	Recommended Reading (Selection)	32
	References	32
2	Modeling in Business Intelligence	35
2.1	Models and Modeling in Business Intelligence	35
2.1.1	The Representation Function of Models	36
2.1.2	Model Presentation	39
2.1.3	Model Building	41

2.1.4	Model Assessment and Quality of Models.....	44
2.1.5	Models and Patterns.....	45
2.1.6	Summary: Models and Modeling in Business Intelligence	46
2.2	Logical and Algebraic Structures	46
2.2.1	Logical Structures	46
2.2.2	Modeling Using Logical Structures	48
2.2.3	Summary: Logical Structures.....	51
2.3	Graph Structures.....	51
2.3.1	Model Structure	51
2.3.2	Modeling with Graph Structures	54
2.3.3	Summary: Graph Structures	57
2.4	Analytical Structures	58
2.4.1	Calculus.....	58
2.4.2	Probabilistic Structures	61
2.4.3	Statistical Structures	67
2.4.4	Modeling Methods Using Analytical Structures	70
2.4.5	Summary: Analytical Structures.....	73
2.5	Models and Data.....	74
2.5.1	Data Generation	74
2.5.2	The Role of Time.....	76
2.5.3	Data Quality	78
2.5.4	Summary: Models and Data	82
2.6	Conclusion and Lessons Learned	82
2.7	Recommended Reading (Selection)	83
	References	83
3	Data Provisioning	87
3.1	Introduction and Goals	87
3.2	Data Collection and Description.....	88
3.3	Data Extraction	90
3.3.1	Extraction-Transformation-Load (ETL) Process	90
3.3.2	Big Data	93
3.3.3	Summary on Data Extraction.....	98
3.4	From Transactional Data Towards Analytical Data.....	98
3.4.1	Table Formats and Online Analytical Processing (OLAP) ...	100
3.4.2	Log Formats	104
3.4.3	Summary: From Transactional Towards Analytical Data	108
3.5	Schema and Data Integration	108
3.5.1	Schema Integration.....	108
3.5.2	Data Integration and Data Quality.....	112
3.5.3	Linked Data and Data Mashups	113
3.5.4	Summary: Schema and Data Integration.....	114
3.6	Conclusion and Lessons Learned	115

3.7	Recommended Reading	115
	References	115
4	Data Description and Visualization	119
4.1	Introduction	119
4.2	Description and Visualization of Business Processes	120
4.2.1	Process Modeling and Layout	121
4.2.2	The BPM Tools' Perspective	122
4.2.3	Process Runtime Visualization	123
4.2.4	Visualization of Further Aspects	123
4.2.5	Challenges in Visualizing Process-Related Information	126
4.2.6	Summary: Description and Visualization of Business Processes	127
4.3	Description and Visualization of Data in the Customer Perspective	127
4.3.1	Principles for Description and Visualization of Collections of Process Instances	127
4.3.2	Interactive and Dynamic Visualization	131
4.3.3	Summary: Visualization of Process Instances	133
4.4	Basic Visualization Techniques	133
4.4.1	Description and Visualization of Qualitative Information	134
4.4.2	Description and Visualization of Quantitative Variables	137
4.4.3	Description and Visualization of Relationships	140
4.4.4	Description and Visualization of Temporal Data	143
4.4.5	Interactive and Dynamic Visualization	145
4.4.6	Summary: Basic Visualization Techniques	146
4.5	Reporting	147
4.5.1	Description and Visualization of Metadata	147
4.5.2	High-Level Reporting	149
4.5.3	Infographics	151
4.5.4	Summary: Reporting	152
4.6	Recommended Reading	153
	References	153
5	Data Mining for Cross-Sectional Data	155
5.1	Introduction to Supervised Learning	155
5.2	Regression Models	159
5.2.1	Model Formulation and Terminology	159
5.2.2	Linear Regression	161
5.2.3	Neural Networks	166
5.2.4	Kernel Estimates	169
5.2.5	Smoothing Splines	171
5.2.6	Summary: Regression Models	172

5.3	Classification Models	173
5.3.1	Model Formulation and Terminology	173
5.3.2	Classification Based on Probabilistic Structures	177
5.3.3	Methods Using Trees	182
5.3.4	K-Nearest-Neighbor Classification	185
5.3.5	Support Vector Machines	186
5.3.6	Combination Methods	190
5.3.7	Application of Classification Methods	191
5.3.8	Summary: Classification Models	192
5.4	Unsupervised Learning	193
5.4.1	Introduction and Terminology	193
5.4.2	Hierarchical Clustering	195
5.4.3	Partitioning Methods	199
5.4.4	Model-Based Clustering	201
5.4.5	Summary: Unsupervised Learning	203
5.5	Conclusion and Lessons Learned	204
5.6	Recommended Reading	204
	References	205
6	Data Mining for Temporal Data	207
6.1	Terminology and Approaches Towards Temporal Data Mining	207
6.2	Classification and Clustering of Time Sequences	212
6.2.1	Segmentation and Classification Using Time Warping	214
6.2.2	Segmentation and Classification Using Response Features	217
6.2.3	Summary: Classification and Clustering of Time Sequences	220
6.3	Time-to-Event Analysis	220
6.4	Analysis of Markov Chains	224
6.4.1	Structural Analysis of Markov Chains	226
6.4.2	Cluster Analysis for Markov Chains	230
6.4.3	Generalization of the Basic Model	231
6.4.4	Summary: Analysis of Markov Chains	233
6.5	Association Analysis	233
6.6	Sequence Mining	237
6.7	Episode Mining	240
6.8	Conclusion and Lessons Learned	242
6.9	Recommended Reading	243
	References	244
7	Process Analysis	245
7.1	Introduction and Terminology	245
7.2	Business Process Analysis and Simulation	247
7.2.1	Static Analysis	248
7.2.2	Dynamic Analysis and Simulation	248

7.2.3	Optimization	251
7.2.4	Summary: Process Analysis and Simulation	252
7.3	Process Performance Management and Warehousing	252
7.3.1	Performance Management	252
7.3.2	Process Warehousing	253
7.3.3	Summary: Process Performance Management and Warehousing	255
7.4	Process Mining	255
7.4.1	Process Discovery	256
7.4.2	Change Mining	263
7.4.3	Conformance Checking	266
7.4.4	Summary: Process Mining	267
7.5	Business Process Compliance	268
7.5.1	Compliance Along the Process Life Cycle	268
7.5.2	Summary: Compliance Checking	270
7.6	Evaluation and Assessment	270
7.6.1	Process Mining	270
7.6.2	Compliance Checking	271
7.7	Conclusion and Lessons Learned	271
7.8	Recommended Reading	272
	References	272
8	Analysis of Multiple Business Perspectives	275
8.1	Introduction and Terminology	275
8.2	Social Network Analysis and Organizational Mining	277
8.2.1	Social Network Analysis	277
8.2.2	Organizational Aspect in Business Processes	282
8.2.3	Organizational Mining Techniques for Business Processes	284
8.2.4	Summary: Social Network Analysis and Organizational Mining	290
8.3	Decision Point Analysis	290
8.4	Text Mining	294
8.4.1	Introduction and Terminology	294
8.4.2	Data Preparation and Modeling	296
8.4.3	Descriptive Analysis for the Document Term Matrix	301
8.4.4	Analysis Techniques for a Corpus	303
8.4.5	Further Aspects of Text Mining	307
8.4.6	Summary: Text Mining	313
8.5	Conclusion and Lessons Learned	313
8.6	Recommended Reading	315
	References	315
9	Summary	319

A Survey on Business Intelligence Tools	329
A.1 Data Modeling and ETL Support	329
A.2 Big Data	330
A.3 Visualization, Visual Mining, and Reporting	334
A.4 Data Mining	337
A.5 Process Mining	338
A.6 Text Mining	339
References	340
Index	343