

Contents

Part I Introduction

1 Data Science: From Concept to Application	3
Thomas Barton and Christian Müller	
1.1 What is Data Science?	4
1.2 What is and What Does a Data Scientist?	5
1.3 Introduction to Data Science	6
1.4 Systems, Tools and Methods	7
1.5 Applications	7
References	8

Part II Introduction to Data Science

2 Visualization and Deep Learning in Data Science	13
Jens Kaufmann and Daniel Retkowitz	
2.1 Introduction	14
2.2 Methods for the Visual Preparation of Data	15
2.2.1 Representing Simple Data and Text	15
2.2.2 Simplifying and Representing Complex Data	17
2.2.2.1 Matrixplot	18
2.2.2.2 Principal Component Analysis and Multidimensional Scaling	18
2.2.2.3 t-SNE	21
2.3 Extract Image Information	23
2.3.1 Recognizing Visual Structures with Deep Learning	23
2.3.2 Architectures for Practice	24
2.4 Bringing Together Image and Data	24
2.4.1 Generation of Enriching Detail Information	24
2.4.2 Transformation of Visual Representations	25
2.4.3 Applications	26

2.5	Summary	27
	References.....	28
3	Digital Ethics in Data-Driven Organizations and AI Ethics as Application Example	31
	Claudia Lemke, Dagmar Monett and Manuel Mikoleit	
3.1	Introduction	32
3.2	Data-driven Organizations	33
	3.2.1 The concept of the Data-driven Organization	33
	3.2.2 Technology Use of Data-driven Organizations	34
	3.2.3 Data-driven Corporate Culture.....	36
3.3	Digital Ethics	37
	3.3.1 Terminology and Moral Theories.....	37
	3.3.2 Overview of Digital Ethical Principles	38
3.4	Digital Ethics and Data-driven Organizations	39
	3.4.1 Digital Ethical Principles and Data Value Creation	39
	3.4.2 Consequences for the Design of Data-driven Organizations	40
3.5	Case Study Deutsche Telekom AG: Operationalization of AI Ethics	42
	3.5.1 The Company’s Motivation for Developing Digital Ethics	42
	3.5.2 AI Ethics at the DTAG	42
3.6	Summary and Outlook	46
	References.....	47
4	Multiple Perspectives for the Implementation of Innovative Technological Solutions in the Context of Data-Driven Decision-Making	49
	Anna-Maria Nitsche, Christian-Andreas Schumann, Christoph Laroque and Olga Matthias	
4.1	Why the Implementation of Innovative Technologies Requires a Comprehensive Approach	50
4.2	Models from the Literature and their Weaknesses.....	52
4.3	The Technological and Organizational Coherence Implementation Model (TOCI Model).....	55
4.4	Benefits and Features of the TOCI Model	57
4.5	Possible Useful Extensions to the TOCI Model	59
4.6	Outlook	60
	References.....	61

5 Don't Be Afraid of Failure—Insights from a Survey on the Failure of Data Science Projects 65
Jule Aßmann, Joachim Sauer and Michael Schulz
5.1 Introduction 66
5.2 Characteristics of and Hypotheses about Data Science Projects 67
5.3 Design and Conduct of the Survey 68
5.4 Evaluation of the Survey 71
5.5 Conclusion and Outlook 74
References. 75

Part III Systems, Tools and Methods

6 Recommendation Systems and the Use of Machine Learning Methods . . . 79
Thomas Barton and Andreas Peuker
6.1 Introduction 80
6.2 Collaborative Recommendation Systems 82
6.2.1 Approaches. 82
6.2.2 Methods 83
6.3 Content-based Recommendation Systems. 85
6.3.1 Approach 85
6.3.2 Methods 86
6.4 Other Concepts. 87
6.4.1 Demographic Recommendation Systems 87
6.4.2 Knowledge-based Recommendation Systems 88
6.4.3 Hybrid Recommendation Systems. 88
6.5 Current Developments 89
6.6 Summary 90
References. 90

7 Comparison of Machine Learning Functionalities of Business Intelligence and Analytics Tools. 95
Gabriele Roth-Dietrich, Michael Gröschel and Benedikt Reiner
7.1 Introduction 96
7.2 Evaluation Framework for Business Intelligence Tools. 97
7.2.1 Selection of BI Tools 97
7.2.2 Personas 98
7.2.2.1 Persona 1: Expert. 98
7.2.2.2 Persona 2: Layperson 98
7.2.3 Comparison Criteria. 99
7.2.4 Test Datasets. 101
7.3 Comparison of ML Methods 101
7.3.1 SAP Analytics Cloud 101
7.3.2 Tableau Online/Tableau Desktop 104

7.3.3	Qlik Sense Business/Qlik Sense Desktop	105
7.3.4	TIBCO Cloud Spotfire	106
7.3.5	RapidMiner	110
7.4	Recommendations	113
	References.	116
8	Using the Data Science Process Model Version 1.1. (DASC-PM v1.1) for Executing Data Science Projects : Procedures, Competencies, and Roles	119
	Stephan Kuehnel, Uwe Neuhaus, Jens Kaufmann, Michael Schulz and Emal M. Alekozai	
8.1	Introduction	120
8.2	The Project Flow when Using DASC-PM v1.1	121
8.2.1	DASC-PM v1.1 at a Glance	121
8.2.2	Project Order	121
8.2.3	Data Provision	124
8.2.4	Analysis	125
8.2.5	Deployment	127
8.2.6	Application.	128
8.3	Overarching Key Areas	129
8.4	Competence-driven Team Management Using Roles	130
8.5	Concluding Remarks	133
	References.	133
 Part IV Applications		
9	Integration of Renewable Energies—AI-Based Prediction Methods for Electricity Generation from Photovoltaic Systems	137
	Boris Brandherm, Matthieu Deru, Alassane Ndiaye, Gian-Luca Kiefer, Jörg Baus and Ralf Gampfer	
9.1	Introduction and Motivation: Integration of Renewable Energies.	138
9.2	Data Preparation.	140
9.2.1	Data Collection.	141
9.2.2	Data Exploration	143
9.2.3	Data Cleansing	144
9.2.4	Data Transformation.	144
9.3	AI-based Prediction Methods	146
9.3.1	Approaches Based on Artificial Neural Networks	148
9.3.2	Approaches Based on Ensemble Machine Learning	149
9.4	Fusion of Results	152
9.5	Application Examples and Outlook.	154
	References.	156

10	Machine Learning for Energy Management Optimization	159
	Gabriele Roth-Dietrich and Rainer Gerten	
10.1	Digital Twin for an Air Conditioning System with Passive and Active Heat Recovery	160
10.2	Conception and Architecture	162
10.3	Analysis and Evaluation of the Data Processing Steps	165
10.3.1	Step 1: Data Collection	165
10.3.2	Step 2: Data Cleansing	165
10.3.3	Step 3: Classify Data	167
10.3.4	Step 4: Filter Data	170
10.3.5	Step 5: Calculate Prediction	172
10.4	Proof-of-Concept	173
10.4.1	Methods and Technologies Stack	173
10.4.2	Visualization of results	174
10.5	Conclusion	176
10.6	Outlook	176
10.6.1	Further Analysis Approaches	176
10.6.2	Applications	177
	References	177
11	Text Mining in Scientific Literature Evaluation: Extraction of Keywords for Describing Content	181
	Thomas Barton and Arthur Kokoev	
11.1	Introduction	182
11.2	Explainable Artificial Intelligence	182
11.3	Extraction of Keywords	182
11.4	Extraction of keywords for a literature review on “Explainable AI”	183
11.5	Conclusion	185
	References	186
12	Identification of Relevant Relationships in Data Using Machine Learning	189
	Joshua Hammesfahr and Martin Spott	
12.1	Introduction	190
12.2	Expertise Problem	191
12.3	Approaches to Reducing the Number of Rules	193
12.3.1	Association Rule Discovery	193
12.3.2	Subgroup Discovery	195
12.4	Determining the Quality of Reduced Rule Sets	197
12.5	Combination Schema	198
12.6	Results	200
12.7	Summary	203
	References	204

13 Framework for the Management and Analysis of Vehicle Data for Model-Based Driver Assistance System Development in Teaching and Research 207

Tobias Peuschke-Bischof and Stefan Kubica

13.1 Motivation 208

13.2 Wildauer Maschinen Werke at TH Wildau 209

13.3 Presentation of the Vehicle Fleet 210

 13.3.1 Trikes 210

 13.3.2 Trucks 212

13.4 Introduction to Infrastructure 212

 13.4.1 ROS 212

 13.4.2 Node-RED 213

 13.4.3 MQTT-Bridge 213

 13.4.4 ROS Car2X 213

 13.4.5 Traffic Light Systems 214

 13.4.6 VDI 214

13.5 Development Framework 215

 13.5.1 Implementation of Vehicle Communication 215

 13.5.2 Model-based Development and Code Generation for Vehicles 216

 13.5.3 Agile Project Management, Knowledge Management and Source Code Management 220

13.6 Scenario-based Teaching and Research 222

 13.6.1 ROS Car2X as Data Aggregation and Function Behavior Across Vehicles 222

 13.6.2 NodeRED for Data Analysis 223

 13.6.3 Interdisciplinary Scenario using the Example of Material Management 225

13.7 Summary and Outlook 226

References 228

Index 231