
Inhaltsübersicht

1	Einleitung	1
	Uwe Haneke · Stephan Trahasch · Michael Zimmer · Carsten Felden	
2	(Advanced) Analytics is the new BI?	15
	Uwe Haneke	
3	Data Science und künstliche Intelligenz – der Schlüssel zum Erfolg?	29
	Marc Beierschoder · Benjamin Diemann · Michael Zimmer	
4	Konzeption und Entwicklung von Data-driven Products/ Datenprodukten	45
	Christoph Tempich	
5	Grundlegende Methoden der Data Science	65
	Stephan Trahasch · Carsten Felden	
6	Deep Learning	101
	Klaus Dorer	
7	Von einer BI-Landschaft zum Data & Analytics-Ökosystem	121
	Michael Zimmer	
8	Self-Service im Data-Science-Umfeld: der emanzipierte Anwender	133
	Uwe Haneke · Michael Zimmer	
9	Data Privacy	147
	Victoria Kayser · Damir Zubovic	
10	Gespräch zur digitalen Ethik	161
	Matthias Haun · Gernot Meier	

Fallstudien

11	Customer Churn mit Keras/TensorFlow und H2O	183
	Shirin Glander	
12	Wirtschaftlichkeitsbetrachtung bei der Auswahl & Entwicklung von Data Science	
	Eine Fallstudie im Online-Lebensmitteleinzelhandel	199
	Nicolas March	
13	Analytics im Onlinehandel	209
	Mikio Braun	
14	Predictive Maintenance	225
	Marco Huber	
15	Scrum in Data-Science-Projekten	245
	Caroline Kleist · Olaf Pier	
16	Der Analytics-Beitrag zu einer Added-Value-Strategie am Beispiel eines Kundenkartenunternehmens	273
	Matthias Meyer	

Anhang

A	Autoren	289
B	Abkürzungen	295
C	Literaturverzeichnis	299
	Index	317

Inhaltsverzeichnis

1	Einleitung	1
	Uwe Haneke · Stephan Trahasch · Michael Zimmer · Carsten Felden	
1.1	Von Business Intelligence zu Data Science	1
1.2	Data Science und angrenzende Gebiete	6
1.3	Vorgehen in Data-Science-Projekten	9
1.4	Struktur des Buches	11
2	(Advanced) Analytics is the new BI?	15
	Uwe Haneke	
2.1	Geschichte wiederholt sich?	15
2.2	Die DIKW-Pyramide erklimmen	21
2.3	Vom Nebeneinander zum Miteinander	24
2.4	Fazit	27
3	Data Science und künstliche Intelligenz – der Schlüssel zum Erfolg?	29
	Marc Beierschoder · Benjamin Diemann · Michael Zimmer	
3.1	Zwischen Euphorie und Pragmatismus	29
3.2	Wann ist Data Science und KI das Mittel der Wahl?	31
3.3	Realistische Erwartungen und klare Herausforderungen	33
3.4	Aus der Praxis	36
3.4.1	Die Automobilbranche als Beispiel	37
3.4.1.1	Machen Sie Ihren Kunden ein Angebot, das sie nicht ausschlagen können	37
3.4.1.2	Spinning the Customer Life Cycle – Schaffen Sie mehr als eine Runde?	38
3.5	Fazit	43

4	Konzeption und Entwicklung von Data-driven Products/ Datenprodukten	45
	Christoph Tempich	
4.1	Einleitung	45
4.2	Datenprodukte	46
4.2.1	Definition	46
4.2.2	Beispiele für Datenprodukte	48
4.2.3	Herausforderungen des Produktmanagements für Datenprodukte	50
4.3	Digitale Produktentwicklung	50
4.3.1	Produktmanagement	50
4.3.2	Agile Entwicklung	51
4.3.3	Lean Startup	51
4.3.4	Data Science	52
4.3.5	Data-centric Business Models	52
4.4	Datenprodukte definieren	53
4.4.1	Ideengenerierung für Datenprodukte entlang der Customer Journey	53
4.4.2	Value Propositions von Datenprodukten	54
4.4.3	Ziele und Messung	55
4.4.4	Die Erwartung an die Güte des Modells bestimmen	56
4.4.5	Mit dem Datenprodukt beginnen	56
4.4.6	Kontinuierliche Verbesserung mit der Datenwertschöpfungskette	57
4.4.7	Skalierung und Alleinstellungsmerkmal	58
4.5	Kritischer Erfolgsfaktor Feedbackschleife	58
4.6	Organisatorische Anforderungen	61
4.7	Technische Anforderungen	63
4.8	Fazit	63
5	Grundlegende Methoden der Data Science	65
	Stephan Trahasch · Carsten Felden	
5.1	Einleitung	65
5.2	Data Understanding und Data Preparation	66
5.2.1	Explorative Datenanalyse	68
5.2.2	Transformation und Normalisierung	70

5.3	Überwachte Lernverfahren	71
5.3.1	Datenaufteilung	71
5.3.2	Bias-Variance-Tradeoff	74
5.3.3	Klassifikationsverfahren	75
5.4	Unüberwachte Lernverfahren und Clustering	79
5.5	Reinforcement Learning	85
5.5.1	Aspekte des Reinforcement Learning	86
5.5.2	Bestandteile eines Reinforcement-Learning-Systems	89
5.6	Evaluation	91
5.6.1	Ausgewählte Qualitätsmaße im Kontext von Klassifikationsaufgabenstellungen	92
5.6.2	Ausgewählte Qualitätsmaße im Kontext von Clusterungen ..	98
5.7	Weitere Ansätze	100
5.7.1	Deep Learning	100
5.7.2	Cognitive Computing	100
5.8	Fazit	100
6	Deep Learning	101
	Klaus Dorer	
6.1	Grundlagen neuronaler Netzwerke	103
6.1.1	Menschliches Gehirn	103
6.1.2	Modell eines Neurons	104
6.1.3	Perzeptron	105
6.1.4	Backpropagation-Netzwerke	107
6.2	Deep Convolutional Neural Networks	109
6.2.1	Convolution-Schicht	110
6.2.2	Pooling-Schicht	112
6.2.3	Fully-Connected-Schicht	113
6.3	Anwendung von Deep Learning	113
6.3.1	RoboCup	113
6.3.2	AudiCup	115
6.3.3	Deep-Learning-Frameworks	116
6.3.4	Standarddatensätze	118
6.3.5	Standardmodelle	118
6.3.6	Weitere Anwendungen	119
6.4	Fazit	120

7	Von einer BI-Landschaft zum Data & Analytics-Ökosystem	121
	Michael Zimmer	
7.1	Einleitung	121
7.2	Komponenten analytischer Ökosysteme	122
7.3	Vom Reporting zur industrialisierten Data Science	125
7.4	Data Science und Agilität	129
7.5	Entwicklungs-, Test- und Produktionsumgebungen für Data Science	129
7.6	Fazit	132
8	Self-Service im Data-Science-Umfeld: der emanzipierte Anwender	133
	Uwe Haneke · Michael Zimmer	
8.1	Einleitung	133
8.2	Self-Service-Angebote für Data & Analytics	135
8.3	Self-Service-Datenaufbereitung und Data Science	137
8.4	Self-Service-Datenaufbereitung vs. ETL	140
8.5	Data & Analytics Governance	142
8.6	Bimodale Data & Analytics: Segen oder Fluch?	144
8.7	Fazit	146
9	Data Privacy	147
	Victoria Kayser · Damir Zubovic	
9.1	Die Rolle von Data Privacy für Analytics und Big Data	147
9.2	Rechtliche und technische Ausgestaltung von Data Privacy	149
9.2.1	Rechtliche Bestimmungen zu Data Privacy	149
9.2.2	Technische und methodische Ansätze zur Schaffung von Data Privacy	150
9.3	Data Privacy im Kontext des Analytics Lifecycle	152
9.3.1	Ideen generieren	153
9.3.2	Prototypen entwickeln	154
9.3.3	Implementieren der Lösung	155
9.4	Diskussion und Fazit	157
10	Gespräch zur digitalen Ethik	161
	Matthias Haun · Gernot Meier	

Fallstudien

11	Customer Churn mit Keras/TensorFlow und H2O	183
	Shirin Glander	
11.1	Was ist Customer Churn?	183
11.1.1	Wie kann Predictive Analytics bei dem Problem helfen? ..	184
11.1.2	Wie können wir Customer Churn vorhersagen?	185
11.2	Fallstudie	185
11.2.1	Der Beispieldatensatz	186
11.2.2	Vorverarbeitung der Daten	189
11.2.3	Neuronale Netze mit Keras und TensorFlow	190
11.2.4	Stacked Ensembles mit H2O	192
11.3	Bewertung der Customer-Churn-Modelle	193
11.3.1	Kosten-Nutzen-Kalkulation	194
11.3.2	Erklärbarkeit von Customer-Churn-Modellen	196
11.4	Zusammenfassung und Fazit	198
12	Wirtschaftlichkeitsbetrachtung bei der Auswahl & Entwicklung von Data Science Eine Fallstudie im Online-Lebensmitteleinzelhandel	199
	Nicolas March	
12.1	Herausforderungen in der Praxis	199
12.1.1	Data-Science-Anwendungen im Online-LEH	199
12.1.2	Auswahl und Umsetzung wirtschaftlicher Anwendungsfälle	200
12.2	Fallstudie: Kaufempfehlungssysteme im Online-Lebensmitteleinzelhandel	204
12.2.1	Vorabanalysen zur Platzierung von Empfehlungen	205
12.2.2	Prototypische Entwicklung eines Empfehlungsalgorithmus	206
12.2.3	MVP und testgetriebene Entwicklung der Recommendation Engine	207
12.3	Fazit	208

13	Analytics im Onlinehandel	209
	Mikio Braun	
13.1	Einleitung	209
13.2	Maschinelles Lernen: von der Uni zu Unternehmen	211
13.3	Wie arbeiten Data Scientists und Programmierer zusammen?	213
13.4	Architekturmuster, um maschinelle Lernmethoden produktiv zu nehmen	218
13.4.1	Architekturmuster des maschinellen Lernens	218
13.4.2	Architekturmuster, um Modelle auszuliefern	219
13.4.3	Datenvorverarbeitung und Feature-Extraktion	220
13.4.4	Automation und Monitoring	222
13.4.5	Integrationsmuster für maschinelles Lernen	222
13.5	Was kann man sonst auf Firmenebene tun, um Data Science zu unterstützen?	223
13.6	Fazit	224
14	Predictive Maintenance	225
	Marco Huber	
14.1	Einleitung	225
14.2	Was ist Instandhaltung?	227
14.2.1	Folgen mangelhafter Instandhaltung	228
14.2.2	Wettbewerbsfähige Produktion	229
14.3	Instandhaltungsstrategien	230
14.3.1	Reaktive Instandhaltung	231
14.3.2	Vorbeugende Instandhaltung	231
14.3.3	Vorausschauende Instandhaltung (Predictive Maintenance)	232
14.4	Prozessphasen der vorausschauenden Instandhaltung	233
14.4.1	Datenerfassung und -übertragung	234
14.4.2	Datenanalyse und Vorhersage	235
14.4.2.1	Unüberwachte Verfahren	236
14.4.2.2	Überwachte Verfahren	238
14.4.3	Planung und Ausführung	239

14.5	Fallbeispiele	240
14.5.1	Heidelberger Druckmaschinen	240
14.5.2	Verschleißmessung bei einem Werkzeugmaschinenhersteller	242
14.5.3	Vorausschauende Instandhaltung in der IT	243
14.6	Fazit	244
15	Scrum in Data-Science-Projekten	245
	Caroline Kleist · Olaf Pier	
15.1	Einleitung	245
15.2	Kurzüberblick Scrum	246
15.3	Data-Science-Projekte in der Praxis	248
15.4	Der Einsatz von Scrum in Data-Science-Projekten	250
15.4.1	Eigene Adaption	251
15.4.2	Realisierte Vorteile	254
15.4.3	Herausforderungen	261
15.5	Empfehlungen	266
15.6	Fazit	271
16	Der Analytics-Beitrag zu einer Added-Value-Strategie am Beispiel eines Kundenkartenunternehmens	273
	Matthias Meyer	
16.1	Geschäftsmodell eines Multipartnerprogramms	273
16.2	Kundenbindung und Kundenbindungsinstrumente	273
16.3	Funktionen und Services eines Multipartnerprogrammbetreibers ..	276
16.3.1	Funktionen	276
16.3.2	Services und Vorteile aus Nutzer- und aus Partnerperspektive	277
16.4	Konkrete Herausforderungen des betrachteten Multipartnerprogrammbetreibers	278
16.5	Added-Value-Strategie	279
16.5.1	Hintergrund und Zielsetzung	279
16.5.2	Ausgangspunkt Datenbasis	280

16.6	Pilotierung ausgewählter Analytics-Ansätze	281
16.6.1	Analytische Ansatzpunkte	281
16.6.2	Pilotierung	282
16.7	Fazit	286

Anhang

A	Autoren	289
B	Abkürzungen	295
C	Literaturverzeichnis	299
	Index	317