

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	Feature Selection	101
	Bianca Huber	
6.1	Weniger ist mehr	101
6.2	Einführung in die Feature Selection	102
6.2.1	Definition	103
6.2.2	Abgrenzung	104
6.3	Ansätze der Feature Selection	105
6.3.1	Der Filter-Ansatz	107
6.3.2	Der Wrapper-Ansatz	109
6.3.3	Der Embedded-Ansatz	111
6.3.4	Vergleich der drei Ansätze	112
6.4	Feature Selection in der Praxis	113
6.4.1	Empfehlungen	113
6.4.2	Anwendungsbeispiel	114
6.5	Fazit	117

7	Deep Learning	119
	Klaus Dorer	
7.1	Grundlagen neuronaler Netzwerke	121
7.1.1	Menschliches Gehirn	121
7.1.2	Modell eines Neurons	122
7.1.3	Perzeptron	123
7.1.4	Backpropagation-Netzwerke	125
7.2	Deep Convolutional Neural Networks	127
7.2.1	Convolution-Schicht	128
7.2.2	Pooling-Schicht	130
7.2.3	Fully-Connected-Schicht	131
7.3	Deep Reinforcement Learning	131
7.4	Anwendung von Deep Learning	132
7.4.1	Sweaty	133
7.4.2	AudiCup	134
7.4.3	DRL im RoboCup	136
7.4.4	Deep-Learning-Frameworks	137
7.4.5	Standarddatensätze	139
7.4.6	Standardmodelle	139
7.4.7	Weitere Anwendungen	140
7.5	Fazit	141
8	Von einer BI-Landschaft zum Data & Analytics-Ökosystem	143
	Michael Zimmer · Benjamin Diemann · Andreas Holzhammer	
8.1	Einleitung	143
8.2	Komponenten analytischer Ökosysteme	144
8.3	Vom Reporting zur industrialisierten Data Science	147
8.4	Data Science und Agilität	151
8.5	Entwicklungs-, Test- und Produktionsumgebungen für Data Science	151
8.6	Vom Spielplatz für Innovation zur Serienfertigung	154
8.7	Anwendungsbeispiel	156
8.8	Fazit	159

9	Self-Service und Governance im Data-Science-Umfeld: der emanzipierte Anwender	161
	Uwe Haneke · Michael Zimmer	
9.1	Einleitung	161
9.2	Self-Service-Angebote für Data & Analytics	163
9.3	Data Governance und Self-Service	165
9.4	Self-Service-Datenaufbereitung und Data Science	167
9.5	Self-Service-Datenaufbereitung vs. ETL	170
9.6	Bimodale Data & Analytics: Segen oder Fluch?	172
9.7	Entwicklungen im Self-Service-Bereich	174
	9.7.1 AutoML als Data-Scientist-Ersatz?	174
	9.7.2 Augmented Analytics	175
9.8	Fazit	176
10	Data Privacy	177
	Victoria Kayser · Damir Zubovic	
10.1	Die Rolle von Data Privacy für Analytics und Big Data	177
10.2	Rechtliche und technische Ausgestaltung von Data Privacy	179
	10.2.1 Rechtliche Bestimmungen zu Data Privacy	179
	10.2.2 Technische und methodische Ansätze zur Schaffung von Data Privacy	180
10.3	Data Privacy im Kontext des Analytics Lifecycle	182
	10.3.1 Ideen generieren	183
	10.3.2 Prototypen entwickeln	184
	10.3.3 Implementieren der Lösung	185
10.4	Diskussion und Fazit	187
11	Gespräch zur digitalen Ethik	191
	Matthias Haun · Gernot Meier	

Fallstudien	211
12 Customer Churn mit Keras/TensorFlow und H2O	213
Shirin Glander	
12.1 Was ist Customer Churn?	213
12.1.1 Wie kann Predictive Analytics bei dem Problem helfen? . . .	214
12.1.2 Wie können wir Customer Churn vorhersagen?	215
12.2 Fallstudie	215
12.2.1 Der Beispieldatensatz	216
12.2.2 Vorverarbeitung der Daten	219
12.2.3 Neuronale Netze mit Keras und TensorFlow	220
12.2.4 Stacked Ensembles mit H2O	222
12.3 Bewertung der Customer-Churn-Modelle	223
12.3.1 Kosten-Nutzen-Kalkulation	224
12.3.2 Erklärbarkeit von Customer-Churn-Modellen	226
12.4 Zusammenfassung und Fazit	228
13 Wirtschaftlichkeitsbetrachtung bei der Auswahl & Entwicklung von Data Science Eine Fallstudie im Online-Lebensmitteleinzelhandel	229
Nicolas March	
13.1 Herausforderungen in der Praxis	229
13.1.1 Data-Science-Anwendungen im Online-LEH	229
13.1.2 Auswahl und Umsetzung wirtschaftlicher Anwendungsfälle	230
13.2 Fallstudie: Kaufempfehlungssysteme im Online-Lebensmitteleinzelhandel	234
13.2.1 Vorabanalysen zur Platzierung von Empfehlungen	235
13.2.2 Prototypische Entwicklung eines Empfehlungsalgorithmus	236
13.2.3 MVP und testgetriebene Entwicklung der Recommendation Engine	237
13.3 Fazit	238

14	Analytics im Onlinehandel	239
	Mikio Braun	
14.1	Einleitung	239
14.2	Maschinelles Lernen: von der Uni zu Unternehmen	241
14.3	Wie arbeiten Data Scientists und Programmierer zusammen?	243
14.4	Architekturmuster, um maschinelle Lernmethoden produktiv zu nehmen	248
14.4.1	Architekturmuster des maschinellen Lernens	248
14.4.2	Architekturmuster, um Modelle auszuliefern	249
14.4.3	Datenvorverarbeitung und Feature-Extraktion	250
14.4.4	Automation und Monitoring	252
14.4.5	Integrationsmuster für maschinelles Lernen	252
14.5	Was kann man sonst auf Firmenebene tun, um Data Science zu unterstützen?	253
14.6	Fazit	254
15	Predictive Maintenance	255
	Marco Huber	
15.1	Einleitung	255
15.2	Was ist Instandhaltung?	257
15.2.1	Folgen mangelhafter Instandhaltung	258
15.2.2	Wettbewerbsfähige Produktion	259
15.3	Instandhaltungsstrategien	260
15.3.1	Reaktive Instandhaltung	261
15.3.2	Vorbeugende Instandhaltung	261
15.3.3	Vorausschauende Instandhaltung (Predictive Maintenance)	262
15.4	Prozessphasen der vorausschauenden Instandhaltung	263
15.4.1	Datenerfassung und -übertragung	264
15.4.2	Datenanalyse und Vorhersage	265
15.4.2.1	Unüberwachte Verfahren	266
15.4.2.2	Überwachte Verfahren	268
15.4.3	Planung und Ausführung	269

15.5	Fallbeispiele	270
15.5.1	Heidelberger Druckmaschinen	270
15.5.2	Verschleißmessung bei einem Werkzeugmaschinenhersteller	272
15.5.3	Vorausschauende Instandhaltung in der IT	273
15.6	Fazit	274
16	Scrum in Data-Science-Projekten	275
	Caroline Kleist · Olaf Pier	
16.1	Einleitung	275
16.2	Kurzüberblick Scrum	276
16.3	Data-Science-Projekte in der Praxis	278
16.4	Der Einsatz von Scrum in Data-Science-Projekten	280
16.4.1	Eigene Adaption	281
16.4.2	Realisierte Vorteile	284
16.4.3	Herausforderungen	291
16.5	Empfehlungen	296
16.6	Fazit	301
17	Der Analytics-Beitrag zu einer Added-Value-Strategie am Beispiel eines Kundenkartenunternehmens	303
	Matthias Meyer	
17.1	Geschäftsmodell eines Multipartnerprogramms	303
17.2	Kundenbindung und Kundenbindungsinstrumente	303
17.3	Funktionen und Services eines Multipartnerprogrammbetreibers ...	306
17.3.1	Funktionen	306
17.3.2	Services und Vorteile aus Nutzer- und aus Partnerperspektive	307
17.4	Konkrete Herausforderungen des betrachteten Multipartnerprogrammbetreibers	308
17.5	Added-Value-Strategie	309
17.5.1	Hintergrund und Zielsetzung	309
17.5.2	Ausgangspunkt Datenbasis	310
17.6	Pilotierung ausgewählter Analytics-Ansätze	311
17.6.1	Analytische Ansatzpunkte	311
17.6.2	Pilotierung	312
17.7	Fazit	316

18	Künstliche Intelligenz bei der Zurich Versicherung – Anwendungen und Beispiele	317
	Michael Zimmer · Jörg Narr · Ariane Horbach · Markus Hatterscheid	
18.1	Herausforderungen innerhalb der Versicherungsbranche	317
18.2	KI bei der Zurich Versicherung	319
18.3	Anwendungsfälle	320
18.3.1	Analyse von Leistungsinformationen mithilfe von MedEye	320
18.3.2	Bildererkennung im Antragsprozess der Motorfahrzeugversicherung in der Schweiz	323
18.3.3	Betrugserkennung im Kfz-Bereich	325
18.3.4	Verbesserung der Kundeninteraktion und des Kundenmanagements mit den Swiss Platform for Analytical and Cognitive Enterprise (SPACE) Services . . .	326
18.4	Fazit	329
Anhang		331
A	Autoren	333
B	Abkürzungen	341
C	Literaturverzeichnis	345
	Index	367