

Uwe Haneke · Stephan Trahasch · Michael Zimmer · Carsten Felden (Hrsg.)

# Data Science

Grundlagen, Architekturen und Anwendungen

Edition TDWI



**dpunkt.verlag**

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
	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
<b>2</b>	<b>(Advanced) Analytics is the new BI?</b>	<b>15</b>
	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
<b>3</b>	<b>Data Science und künstliche Intelligenz – der Schlüssel zum Erfolg?</b>	<b>29</b>
	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

<b>4</b>	<b>Konzeption und Entwicklung von Data-driven Products/ Datenprodukten</b>	<b>45</b>
	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
<b>5</b>	<b>Grundlegende Methoden der Data Science</b>	<b>65</b>
	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
<b>6</b>	<b>Deep Learning</b> . . . . .	<b>101</b>
	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

<b>7</b>	<b>Von einer BI-Landschaft zum Data &amp; Analytics-Ökosystem</b>	<b>121</b>
	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
<b>8</b>	<b>Self-Service im Data-Science-Umfeld: der emanzipierte Anwender</b>	<b>133</b>
	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
<b>9</b>	<b>Data Privacy</b>	<b>147</b>
	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
<b>10</b>	<b>Gespräch zur digitalen Ethik</b>	<b>161</b>
	Matthias Haun · Gernot Meier	

## Fallstudien

<b>11</b>	<b>Customer Churn mit Keras/TensorFlow und H2O</b>	<b>183</b>
	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
<b>12</b>	<b>Wirtschaftlichkeitsbetrachtung bei der Auswahl &amp; Entwicklung von Data Science Eine Fallstudie im Online-Lebensmitteleinzelhandel</b>	<b>199</b>
	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

<b>13</b>	<b>Analytics im Onlinehandel</b>	<b>209</b>
	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
<b>14</b>	<b>Predictive Maintenance</b>	<b>225</b>
	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
<b>15</b>	<b>Scrum in Data-Science-Projekten</b>	<b>245</b>
	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
<b>16</b>	<b>Der Analytics-Beitrag zu einer Added-Value-Strategie am Beispiel eines Kundenkartenunternehmens</b>	<b>273</b>
	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

<b>A</b>	<b>Autoren</b>	<b>289</b>
<b>B</b>	<b>Abkürzungen</b>	<b>295</b>
<b>C</b>	<b>Literaturverzeichnis</b>	<b>299</b>
	<b>Index</b>	<b>317</b>