

Teil I Grundlagen von Data Science

1	Geschichte von Data Science	3
	Göran Kauermann	
	Literatur.	6
2	Data Science in der Ausbildung	7
	Göran Kauermann	
3	Quantifizierung von Unsicherheit	13
	Göran Kauermann	
	Literatur.	17
4	Maschinelles Lernen	19
	Damian Borth, Eyke Hüllermeier und Göran Kauermann	
4.1	Einleitung und Prinzipien des ML	19
4.2	Regression	23
4.3	Bayes Modelle und MCMC	26
4.4	Entscheidungsbäume und Regellernen	28
4.5	Neuronale Netze.	35
4.6	Automatisiertes Maschinelles Lernen	44
	Literatur.	47
5	Strukturierte und unstrukturierte Daten	51
	Wolfgang Reuter, Anna Korotkova und Linh Nguyen	
5.1	Wie Computer sehen (lernten)	53
5.2	Natural Language Processing	58
	Literatur.	68
6	Visualisierung und Exploration	69
	Kathrin Guckes	
6.1	Vorteile visualisierungsgetriebener Ansätze	72
6.2	Visuelle Variablen	77
6.3	Visualisierung in der Praxis	86
	Literatur.	88

7	Datenethik zwischen gesellschaftlichem Anspruch und betrieblicher Praxis	93
	Ulrich Hemel	
7.1	Datenethik – Wirkungsgeschichte und Geltungsbereich	93
7.2	Digitale Datenethik in Institutionen und digitale Unternehmensethik	96
7.3	Berufsspezifische Datenethik und digitale Professionsethik	97
7.4	Digitale Datenethik von Individuen und grundsätzliche Ethikmodelle.	98
7.5	Politische Datenethik	98
7.6	Ethische Herausforderungen durch KI und digitale Bereichsethiken.	99
7.7	Digitale Risikoethik und die ethische Toolbox	101
7.8	Das Ziel: digitale Verantwortung	102
	Literatur.	103

Teil II Data Science in der Praxis

8	Data Science in der Praxis	109
	Andreas Gillhuber	
8.1	Organisation	110
8.2	Prozesse und Use Cases	111
8.3	Rollen & Verantwortungen	112
8.4	Governance	114
8.5	IT-Systemlandschaft.	115
8.6	Zusammenfassung	116
	Literatur.	116
9	Data Science Use Cases	117
	Ludwig Brummer und Andreas Gillhuber	
9.1	Vorgehensmodell	117
9.2	Auswahl von Use Cases.	119
9.3	Priorisierung von Use Cases	120
9.4	Detaillierung der Use Cases.	122
9.5	Erfolgskennzahlen/Metriken/KPIs.	124
9.6	Ausblick.	126
10	Skalierbare und wertschaffende KI-Implementierung in der Luftfahrt	129
	Susan Wegner	
10.1	Analytics Factory – „Menschen und Organisation“	130
10.2	Analytics Factory – Daten, Technologie und Tools	133
10.3	AI as a Service	137
	Literatur.	138

11 Künstliche Intelligenz: Anwendungen und Werkzeuge in der Automobilindustrie	141
Andre Luckow, Joscha Eirich und Kai Demtröder	
11.1 Einleitung	141
11.2 Hintergrund: Künstliche Intelligenz und Maschinelles Lernen	142
11.3 Anwendungen in der automobilen Wertschöpfungskette	147
11.4 Diskussion	151
11.5 Zusammenfassung	152
Literatur	153
12 Architekturen und Technologien für den Data Lake	157
Carsten Dittmar und Peter Schulz	
12.1 Historie der dispositiven Datenplattformen	157
12.2 Das Data-Lake-Konzept	158
12.3 Reifegrad von Data-Lakes	160
12.4 Architekturprinzipien eines Data-Lakes	161
12.5 Datenarchitektur eines Data-Lakes	164
12.6 Zusammenfassung	165
Literatur	165
13 Big Data bringt Einsparungen in Milliardenhöhe	167
Thomas Froese	
13.1 Einführung	167
13.2 Was ist der konkrete Nutzen von Daten in der Produktion?	168
13.3 Einsparungen in Milliardenhöhe	169
13.4 Datenqualität als Schlüsselgröße	169
13.5 „What-If-Optimizer“	170
13.6 Auswirkungen der Datenqualität auf die Effizienz des Maschinellen Lernens	170
Teil III Beispiele aus der Praxis	
14 (Rück-)Versicherung	175
Wolfgang Hauner	
14.1 Daten- und KI-gestütztes Schadenmanagement in der Versicherung	175
15 Die AI Factory der ERGO	187
Felix Wenzel und Alexander Haag	
15.1 Daten und AI in Erstversicherungsunternehmen	188
15.2 Daten als Treibstoff der digitalen Transformation	189
15.3 AI-Implementierung als „Selbstläufer“?	189
15.4 Nutzenerwartung bei ERGO	190
15.5 Zielsetzung und Scope der AI Factory	191
15.6 Machine Learning Operations als übergeordnetes Paradigma	192

15.7	Adaption von ML-Ops für ERGO	192
15.8	Implementierung in der AI Factory	194
15.9	IT-Architektur	197
15.10	AI-Governance- und Compliance-Aspekte	199
15.11	Fazit	200
	Literatur.	201
16	Daten und Rückversicherung: Eine ewige Liebe	203
	Monica Epple und Maya Bundt	
16.1	Swiss Re: Einer der größten Rückversicherer weltweit.	203
16.2	Rückversicherung: Was ist das eigentlich?	203
16.3	Im Innern der Rückversicherung	206
16.4	Daten und Analytics in der Rückversicherung.	208
16.5	Zusammenfassung und Ausblick	214
17	Automotive & Manufacturing	215
	Joscha Eirich, Andre Luckow und Kai Demtröder	
17.1	Visual Analytics zur Optimierung von Prüfprozessen in der Serienfertigung elektrischer Motoren	215
	Literatur.	225
18	Computer Vision based Deep Learning zur Reifenwandererkennung.	227
	Holger Großkreutz und Dubravko Dolic	
18.1	Vom OCR zur AI-getriebenen Computer Vision	228
18.2	Knowledge-Aufbau – Deep Learning als Technologieprojekt.	230
18.3	Konsolidierung und Aufbau eines Frameworks zur DL-gestützten Bilderkennung	232
18.4	Aktueller Status und Ausblick	233
19	Autonomes Fahren und das Projekt KI-Wissen.	235
	Luca Bruder und Dina Krayzler	
19.1	Autonomes Fahren: Aktuelle Entwicklung und Herausforderungen	235
19.2	Das Projekt KI-Wissen: Idee und Ziele	237
19.3	Explainable AI-Methoden in KI-Wissen	238
19.4	Fazit	244
	Literatur.	244
20	Energiewirtschaft	247
	Philipp Ditzel, Michael Mommert und Wolfgang Korosec	
20.1	Optimierung der Photovoltaik-Prognose mittels Deep Learning	247

21	Transport & Logistik	255
	Christian Spannbauer und Sebastian Weber	
21.1	Alternative Daten zur kommerziellen Steuerung der Lufthansa Airlines	255
21.2	Zusammenfassung und Ausblick	265
	Literatur	265
22	Die Entwicklung eines Kartellscreening-Tools bei der Deutsche Bahn AG	267
	Hannes Beth und Livia Kaiser	
22.1	Einleitung	267
22.2	Set-Up des Kartellscreening-Projekts	270
22.3	Entwicklung des daten- und algorithmusbasierten Kartellscreening-Tools	273
22.4	Ausblick	280
	Literatur	281
23	Datenstrategie „Mehrwert durch Daten und KI“ professionalisiert die Digitalisierung der Bahn	283
	Oliver Wittmaier, Anna Einhorn, Ursula Besbak und Nicholas Drude	
23.1	Einführung	284
23.2	Fünf Handlungsfelder der DB-Datenstrategie	284
23.3	Schlagkräftige Data-Governance zur Sicherstellung der notwendigen Datenqualität	285
23.4	Wesentliche Governance-Rollen für die Geschäftsfeld- Bereiche	286
23.5	Zentraler Datenkatalog für mehr Datentransparenz und einfache Datenverfügbarkeit bei der DB.	288
23.6	Erfolgreicher Start im Bahnalltag	288
23.7	Operating Model des Datenkatalogs	289
23.8	Einjährige Pilotierungsphase zur Erprobung des Datenkatalogs	290
23.9	Erfolgsfaktoren bei der Einführung	290
23.10	User Enablement und Content-Generierung im Fokus	291
23.11	Fazit und Ausblick	291
24	Wie Künstliche Intelligenz helfen kann, die Pünktlichkeit der Bahn zu verbessern	293
	Hannah Richta	
24.1	Einleitung	293
24.2	Dispositionsentscheidungen	294
24.3	Ansatzpunkt und Ansätze für Künstliche Intelligenz.	296

24.4	Die automatische Dispositionsassistentz auf Grundlage des Produktionsmodells Betrieb	300
24.5	Ausblick	303
	Literatur.	303
25	Gesundheitswesen, Chemie/Pharma	305
	Wolfgang Reuter	
25.1	Wie falsche ärztliche Diagnosen den Einsatz Künstlicher Intelligenz in der Medizin behindern	305
	Literatur.	318
26	Retail & Consumer Products, Media	321
	Karin Immenroth und Marc Egger	
26.1	Let's Talk Data: SmartSearch als Beispiel für innovative End-to-End-Datenproduktentwicklung bei RTL	321
	Literatur.	332
27	Automatische Klassifikation von Points-of-Interest für passgenaue Empfehlungen von Reise- und Freizeitaktivitäten in der ADAC Trips App.	333
	Astrid Rohmfeld, Christina Sievers, Florian Leser und Nico Becker	
27.1	Die Freizeitmotivationen	334
27.2	Automatische Klassifizierung von POIs durch Maschinelles Lernen	335
27.3	Matching POIs und Nutzerprofil	341
27.4	Ausblick	343
	Literatur.	343
28	Modellierung der Nachfrageunsicherheit zur gezielten Steuerung auf Ziel-Servicelevel.	345
	Matthias Ulrich und Robin Senge	
28.1	Problemstellung	348
28.2	Modellierung	349
28.3	Technologie Stack	352
28.4	Ergebnisse und Ausblick	353
	Literatur.	354