

Einleitung: Ein Handbuch zum Management von Data Science	13
---	-----------

Teil I: Data-Science-Grundlagen	27
--	-----------

1 Eine Einführung in Data Science aus Projektsicht	29
Verlauf eines Data-Science-Projekts (Prozessmodell)	30
Von einfachen Analysen zur Automatisierung (Analytics Continuum) ..	32
Welche Kompetenzen brauchen wir in einem Data-Science-Projekt? ...	34
2 Wie wir über Daten sprechen	37
Strukturierte Daten	37
Semistrukturierte Daten	38
Unstrukturierte Daten	40
Skalenniveaus und besondere Datenformate	40
Verschiedene Aspekte der Qualität von Daten	42
Big Data und Smart Data	43
3 Datenbeschaffung und -aufbereitung	45
Datenquellen und Datenerhebung	45
Datenzugriff ist nicht nur eine technische Angelegenheit	46
Integration und Aufbereitung verschiedener Datenquellen	47
Trainings- und Testdaten für das Training von Machine-Learning- Algorithmen	48
Feature Engineering	48
4 Deskriptive Analysen	51
Univariate Basisstatistiken und Kennzahlen	51
Bivariate Darstellungen und Korrelationen	53
Visualisierung von Daten	55
Explorative Datenanalyse (EDA)	58

5	Modellbildung in der klassischen Statistik	61
	Grundgesamtheiten und Stichproben	61
	Die Regressionsanalyse als Beispiel für ein erklärendes Modell	63
	Wie funktioniert eine Regressionsanalyse aus mathematischer Sicht?	64
	Die Flexibilität der Regressionsanalyse	65
	Spezielle Anwendungsfälle: Zeitreihenanalyse und Vorhersagen	67
6	Vorhersagen im Machine Learning	71
	Supervised Learning	73
	Regressionsanalyse	73
	Entscheidungsbäume	74
	K-Nearest-Neighbors	74
	Datenqualität und verwandte Herausforderungen	75
	Unsupervised Learning	77
	Dimensionsreduktion	77
	Clusteranalyse	77
	Deep Learning, Reinforcement Learning und neuronale Netze	78
	Predictive, Prescriptive, Automation	80
7	Aufbereitung der Ergebnisse für die weitere Verwendung	81
	Dokumentation, Wiederverwendung und Replizierbarkeit	81
	Reporting	83
	Statischer Report	83
	Dashboards	84
	Storytelling und visuelle Kommunikation mit Daten	85
	Mehrwert von Daten im Unternehmen	86
	Impact, Evaluation und Feedback	86
8	Aspekte einer Basisinfrastruktur	89
	Datenformate und Datenbanken	89
	Plain Text	90
	Binary Files	91
	SQL-Datenbanken	91
	NoSQL	91
	Datenverarbeitung und Analyse	91
	Collaboration und Arbeit in der Cloud	92
9	Hands-on: Beispielprojekt	95
	Studiendesign	95
	Datenbeschaffung und -aufbereitung	96
	Analyse der Daten	98
	Dokumentation und Reporting	99
	Handlungsempfehlung (Impact)	100

Teil II: Data-Science-Management	101
10 Fallstricke für Data-Science-Projekte	105
Fallstricke in Technologie und Infrastruktur	105
Data Engineering wird unterschätzt	106
Datensilos	106
Fallstricke in der Modellierung	107
Zu komplexe Modelle	107
Fluch der Dimensionalität	108
Ausreißer	109
Fallstricke im Management	110
Law of Instrument	110
Zu viel, zu früh	111
Unklare Ziele	111
Ein Projekt ist keine produktive Anwendung	112
Fehlende Skills und Data-Science-Kultur	112
11 Grundlagen des Projektmanagements	115
Klassisches Anforderungsmanagement	117
Agiles Management und Lean Mindset	120
Mehrwert und Kundenzentrierung	121
Kollaboration	121
Iteratives und inkrementelles Vorgehen	122
Kontinuierliche Verbesserung	122
Dezentralität und Selbstorganisation	123
PoC und MVP	123
Agiles Mindset	124
Erkenntnisse aus der agilen Praxis	124
Agiles Anforderungsmanagement	125
Zeit- und Ressourcenmanagement	127
Finanzielle Ressourcen	128
Zeitliche Ressourcen	129
Infrastrukturelle Ressourcen	132
Daten	133
Kontextualisierung und Kommunikation	134
Team-Bubble	135
12 Data-Science-Teams	137
Funktionen von Teams	137
Teamstrukturen	140
Team of Teams und New Work	143
Verortung von Data-Science-Teams	145

Rollen und deren Aufgaben in Data-Science-Teams	147
Rollenverständnis nach methodischer Tiefe	147
Rollenverständnis nach Ausbildung und Interessen	148
Rollenverständnis nach Aufgaben	149
Rollen von Data Scientists	150
Data Scientists	151
Data Engineers	151
Fachexpertinnen und -experten	151
Software Engineers und DevOps Engineers	152
Machine Learning Engineers und MLOps Architects	152
Model-Risk-Managerinnen und -Managern	153
Softwarearchitektinnen und -architekten	153
Analystinnen und Analysten	154
Herausforderungen und Konflikte in Teams	155
Digitales Arbeiten und Remote Work	155
Zusammenarbeit und Kommunikation	156
13 Data-Science-Managerinnen und -Manager	159
Aufgaben und Fähigkeiten	161
Modernes Leadership	164
Servant Leadership	164
Agile Leadership	165
Shared Leadership	167
Impact durch Leadership	168
Coaching und Mentoring von Data Scientists	171
14 Hands-on: Empfohlenes Toolkit für das Data-Science-Management	175
Scrum	175
Kanban	177
Scrum oder Kanban nutzen?	178
Team Health Checks	179
AI Project Canvas	181
Checkliste Anforderungsmanagement	182
Problemfelder benennen	182
Herausforderungen ermitteln	183
Mehrwert beschreiben	183

Teil III: Infrastruktur und Architektur	185
15 Automatisierung und Operationalisierung im kybernetischen Regelkreis	187
Das wissenschaftliche Vorgehen: Wissen iterativ weiterentwickeln und vertiefen	188
Proof-of-Concept-Projekte und Design Thinking	188
Operationalisierung und Evaluation von Zielen in laufenden Projekten	189
Der kybernetische Regelkreis	190
Cross Industry Standard Process for Data Mining (CRISP-DM)	192
16 Grundlagen der IT-Infrastruktur	193
Bausteine einer Softwareanwendung	193
Hardware: eigene Rechner vs. Cloud	196
Container und Microservices	199
Platform-as-a-Service (PaaS) und Serverless	200
Software- und Data-Science-as-a-Service (SaaS/DSaaS)	201
17 Data-Science-Architekturen	203
Data Lake	204
Data Warehouse (DWH)	205
Weitere Optionen wie das Analytics Lab	207
Interaktive Visualisierung, EDA und Business Intelligence	208
Data Mesh	209
18 DevOps und MLOps: Entwicklung und Betrieb	211
Versionierung und Versionskontrolle	211
Continuous Integration and Delivery	213
Microservices und Application Programming Interfaces (APIs)	215
Testing und Monitoring	217
Betrieb von Machine-Learning-Modellen (DevOps und MLOps)	219
19 Hands-on: Modellierung von Software und Infrastruktur	221
Bestandsaufnahme im Event-Storming	221
Weiterentwicklung in der Business Process Model and Notation (BPMN)	223
Modellierung einer technischen Infrastruktur	224
Modellierung einer (relationalen) Datenbank	225
Regelkonformität	226

Teil IV: Data Science Governance und Data-driven Culture	227
20 Digitale Transformation der Unternehmen	231
Strategischer Einsatz von Daten	232
Wettbewerbsvorteile durch Data Science	236
As-a-Service-Modelle	239
21 Implementierung im Unternehmen	241
Schritt 1: Ideenfindung	241
Wie findet man geeignete Anwendungsfälle?	241
Schritt 2: Proof-of-Concept	242
Schritt 3: Technische Implementierung	243
Schritt 4: Implementierung auf Bereichsebene	243
Schritt 5: Skalierung auf Unternehmensebene	244
Schritt 6: Verstetigung	245
Change Management	245
Datenmanagement	249
IT-Management	253
22 Sicherheit und Datenschutz	255
Safety	256
Security	257
Governance, Compliance und rechtliche Aspekte	261
Ethische Aspekte und Corporate Responsibility	263
Digitalpolitik	266
23 Digitale Kompetenzen und Data-Science-Kultur	269
New Work	270
Flexibilisierung der Arbeitsorganisation	273
Diversität und Kreativität	274
Netzwerkorganisationen und Leadership	274
Achtsamkeit und Gesundheit	274
Recruiting	275
Upskilling und Reskilling	278
Entrepreneurship, Intrapreneurship und Innovation	280
Literacy, Enablement und Citizen Data Science	282
Grundpfeiler einer kreativen Umgebung	284
24 Hands-on: Toolkit für Strategie und Governance	287
Business Model Canvas	287
AI Canvas	288
Datenstrategie-Designkit	290

25 Schlüsselfaktoren für erfolgreiches Data-Science-Management 293

 Data Scientists als Individuen 293

 Wirtschaftlichkeit 294

 Governance 294

 Kultur 294

 Infrastruktur 295

 Projekte und Teams 296

 Wirtschaftlichkeit 296

 Governance 296

 Kultur 297

 Infrastruktur 298

 Unternehmen und Strategie 298

 Wirtschaftlichkeit 298

 Governance 299

 Kultur 299

 Infrastruktur 300

Index 301