

Angewandte Wirtschaftsinformatik

Thomas Barton  
Christian Müller *Hrsg.*

# Data Science anwenden

Einführung, Anwendungen und Projekte

EBOOK INSIDE

 Springer Vieweg

---

# Angewandte Wirtschaftsinformatik

## Reihe herausgegeben von

Thomas Barton, FB Informatik, Hochschule Worms, Worms, Deutschland

Christian Müller, Technische Hochschule Wildau, Wildau, Deutschland

Die Buchreihe „Angewandte Wirtschaftsinformatik“ bereitet das Themengebiet Wirtschaftsinformatik anhand grundlegender Konzepte, praxisnaher Anwendungen und aktueller Themen auf. Dabei wird auf der einen Seite die Perspektive der betrieblichen Anwendungsentwicklung beleuchtet, welche die Erstellung von betriebswirtschaftlicher Software und deren Einsatz in Unternehmen zum Gegenstand hat. Auf der anderen Seite stellt die Perspektive der Organisationsgestaltung sicher, dass die eingesetzte Software auch eine bestmögliche Einbindung in die betriebliche Organisation erfährt. Das Ziel der vorliegenden Reihe besteht darin, angewandte Wirtschaftsinformatik in Form von betrieblichen Szenarien, Best Practices und anwendungsorientierter Forschung aufzubereiten und in kompakter und verständlicher Form darzustellen. Zielgruppe sind sowohl Studierende und Lehrende als auch Praktiker.

**Prof. Dr. Thomas Barton** ist Professor an der Hochschule Worms. Seine Schwerpunkte liegen in den Bereichen Entwicklung betrieblicher Anwendungen, E-Business, Cloud Computing und Data Science.

**Prof. Dr. Christian Müller** ist Professor an der Technischen Hochschule Wildau. Seine Schwerpunkte liegen in den Bereichen Operations Research, Simulation von Geschäftsprozessen und Internet-Technologien.

Weitere Bände in der Reihe <http://www.springer.com/series/13757>

---

Thomas Barton · Christian Müller  
(Hrsg.)

# Data Science anwenden

Einführung, Anwendungen und Projekte

*Hrsg.*

Thomas Barton  
FB Informatik, Hochschule Worms  
Worms, Deutschland

Christian Müller  
FB Wirtschaft, Informatik, Recht (WIR)  
Technische Hochschule Wildau  
Wildau, Deutschland

ISSN 2522-0497                      ISSN 2522-0500 (electronic)  
Angewandte Wirtschaftsinformatik  
ISBN 978-3-658-33812-1              ISBN 978-3-658-33813-8 (eBook)  
<https://doi.org/10.1007/978-3-658-33813-8>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2021

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung der Verlage. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung: Sybille Thelen

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

---

# Inhaltsverzeichnis

## Teil I Einleitung

<b>1 Data Science: Vom Begriff zur Anwendung</b> .....	3
Thomas Barton und Christian Müller	
1.1 Was ist Data Science? .....	4
1.2 Was ist und was macht ein Data Scientist? .....	5
1.3 Einführung in Data Science .....	7
1.4 Systeme, Werkzeuge und Methoden .....	7
1.5 Anwendungen .....	8
Literatur .....	9

## Teil II Einführung in Data Science

<b>2 Visualisierung und Deep Learning in der Data Science</b> .....	13
Jens Kaufmann und Daniel Retkowitz	
2.1 Einleitung .....	14
2.2 Verfahren für die visuelle Aufbereitung von Daten .....	15
2.2.1 Einfache Daten und Texte darstellen .....	15
2.2.2 Komplexe Daten vereinfachen und darstellen .....	18
2.2.2.1 Matrixplot .....	18
2.2.2.2 Hauptkomponentenanalyse und multidimensionale Skalierung .....	20
2.2.2.3 t-SNE .....	22
2.3 Bildinformationen extrahieren .....	22
2.3.1 Visuelle Strukturen mit Deep Learning erkennen .....	24
2.3.2 Architekturen für die Praxis .....	25
2.4 Zusammenführung von Bild und Daten .....	25
2.4.1 Generierung anreichernder Detailinformationen .....	25
2.4.2 Transformation visueller Repräsentationen .....	26
2.4.3 Einsatzmöglichkeiten .....	28

2.5	Zusammenfassung .....	29
	Literatur .....	29
<b>3</b>	<b>Digitale Ethik in datengetriebenen Organisationen und deren Anwendung am Beispiel von KI-Ethik .....</b>	<b>33</b>
	Claudia Lemke, Dagmar Monett und Manuel Mikoleit	
3.1	Einführung .....	34
3.2	Datengetriebene Organisationen .....	35
	3.2.1 Begriff der datengetriebenen Organisation .....	35
	3.2.2 Technologienutzung datengetriebener Organisationen .....	36
	3.2.3 Datengetriebene Unternehmenskultur .....	39
3.3	Digitale Ethik .....	40
	3.3.1 Begriff und Moralthorien .....	40
	3.3.2 Überblick über digital-ethische Grundsätze .....	41
3.4	Digitale Ethik und datengetriebene Organisationen .....	42
	3.4.1 Digital-ethische Grundsätze und Datenwertschöpfung .....	42
	3.4.2 Konsequenzen für die Gestaltung datengetriebener Organisationen .....	43
3.5	Fallbeispiel Deutsche Telekom AG: Operationalisierung einer KI-Ethik .....	45
	3.5.1 Motivation des Konzerns zur Entwicklung einer digitalen Ethik .....	45
	3.5.2 KI-Ethik bei der DTAG .....	45
3.6	Zusammenfassung und Ausblick .....	49
	Literatur .....	50
<b>4</b>	<b>Multiple Perspektiven bei der Implementierung innovativer technologischer Lösungen im Kontext datengesteuerter Entscheidungsfindung .....</b>	<b>53</b>
	Anna-Maria Nitsche, Christian-Andreas Schumann, Christoph Laroque und Olga Matthias	
4.1	Warum die Implementierung innovativer Technologien eine umfassende Herangehensweise notwendig macht .....	54
4.2	Modelle aus der Literatur und ihre Schwachstellen .....	56
4.3	Das Technological and Organisational Coherence Implementation-Modell (TOCI-Modell) .....	59
4.4	Vorteile und Besonderheiten des TOCI-Modells .....	62
4.5	Mögliche nützliche Erweiterungen des TOCI-Modells .....	64
4.6	Ausblick .....	65
	Literatur .....	66

<b>5</b>	<b>Keine Angst vor Fehlschlägen – Erkenntnisse aus einer Umfrage zum Scheitern von Data-Science-Projekten</b> . . . . .	69
	Jule Aßmann, Joachim Sauer und Michael Schulz	
5.1	Einleitung . . . . .	70
5.2	Merkmale von und Hypothesen zu Data-Science-Projekten . . . . .	71
5.3	Konzeption und Durchführung der Umfrage . . . . .	73
5.4	Auswertung der Umfrage . . . . .	75
5.5	Fazit und Ausblick . . . . .	79
	Literatur . . . . .	80
<b>Teil III Systeme, Werkzeuge und Methoden</b>		
<b>6</b>	<b>Empfehlungssysteme und der Einsatz maschineller Lernverfahren</b> . . . . .	85
	Andreas Peuker und Thomas Barton	
6.1	Einleitung . . . . .	86
6.2	Kollaborative Empfehlungssysteme . . . . .	88
	6.2.1 Ansätze . . . . .	88
	6.2.2 Methoden . . . . .	89
6.3	Inhaltsbasierte Empfehlungssysteme . . . . .	92
	6.3.1 Ansatz . . . . .	92
	6.3.2 Methoden . . . . .	93
6.4	Weitere Konzepte . . . . .	94
	6.4.1 Demografische Empfehlungssysteme . . . . .	94
	6.4.2 Wissensbasierte Empfehlungssysteme . . . . .	95
	6.4.3 Hybride Empfehlungssysteme . . . . .	95
6.5	Aktuelle Entwicklungen . . . . .	96
6.6	Zusammenfassung . . . . .	97
	Literatur . . . . .	97
<b>7</b>	<b>Vergleich der Machine-Learning-Funktionalitäten von Business-Intelligence- und Analytics-Tools</b> . . . . .	101
	Gabriele Roth-Dietrich, Michael Gröschel und Benedikt Reiner	
7.1	Einleitung . . . . .	103
7.2	Bewertungsrahmen der Business-Intelligence-Tools . . . . .	104
	7.2.1 Auswahl der BI-Tools . . . . .	104
	7.2.2 Personas . . . . .	104
	7.2.2.1 Persona 1: Experte/in . . . . .	104
	7.2.2.2 Persona 2: Laie . . . . .	105
	7.2.3 Vergleichskriterien . . . . .	105
	7.2.4 Testdatensets . . . . .	107
7.3	Vergleich der ML-Methoden . . . . .	108
	7.3.1 SAP Analytics Cloud . . . . .	108

7.3.2	Tableau Online/Tableau Desktop . . . . .	111
7.3.3	Qlik Sense Business/Qlik Sense Desktop . . . . .	112
7.3.4	TIBCO Cloud Spotfire . . . . .	113
7.3.5	RapidMiner . . . . .	117
7.4	Empfehlungen . . . . .	121
	Literatur. . . . .	124
<b>8</b>	<b>Data-Science-Projekte mit dem Vorgehensmodell</b>	
	<b>„DASC-PM“ durchführen: Kompetenzen, Rollen und Abläufe. . . . .</b>	<b>127</b>
	Emal M. Alekozai, Jens Kaufmann, Stephan Kühnel, Uwe Neuhaus und Michael Schulz	
8.1	Einleitung . . . . .	128
8.2	Ablauf eines Projektes mit DASC-PM. . . . .	129
8.2.1	DASC-PM im Überblick . . . . .	129
8.2.2	Projektauftrag. . . . .	131
8.2.3	Datenbereitstellung . . . . .	132
8.2.4	Analyse. . . . .	134
8.2.5	Nutzbarmachung . . . . .	136
8.2.6	Nutzung . . . . .	137
8.3	Phasenübergreifende Schlüsselbereiche. . . . .	138
8.4	Kompetenzorientierte Teamsteuerung mit Rollen . . . . .	140
8.5	Fazit . . . . .	142
	Literatur. . . . .	143
	<b>Teil IV Anwendungen</b>	
<b>9</b>	<b>Integration erneuerbarer Energien – KI-basierte</b>	
	<b>Vorhersageverfahren zur Stromerzeugung durch Photovoltaikanlagen . . . . .</b>	<b>147</b>
	Boris Brandherm, Matthieu Deru, Alassane Ndiaye, Gian-Luca Kiefer, Jörg Baus und Ralf Gampfer	
9.1	Einleitung und Motivation: Integration Erneuerbarer Energien. . . . .	148
9.2	Datenaufbereitung . . . . .	150
9.2.1	Datenerfassung. . . . .	151
9.2.2	Datenerkundung. . . . .	154
9.2.3	Datenbereinigung. . . . .	156
9.2.4	Datentransformation. . . . .	156
9.3	KI-basierte Vorhersageverfahren . . . . .	157
9.3.1	Ansätze auf Basis von künstlichen neuronalen Netzen . . . . .	159
9.3.2	Ansätze auf Basis von Ensemble Machine Learning. . . . .	160
9.4	Fusion der Ergebnisse . . . . .	163
9.5	Anwendungsbeispiele und Ausblick . . . . .	165
	Literatur. . . . .	167

<b>10</b>	<b>Machine Learning für die Energiemanagementoptimierung</b> . . . . .	171
	Gabriele Roth-Dietrich und Rainer Gerten	
10.1	Digital Twin für eine Klimatisierungsanlage mit passiver und aktiver Wärmerückgewinnung . . . . .	172
10.2	Konzeption und Architektur . . . . .	176
10.3	Analyse und Evaluation der Datenbearbeitungsschritte. . . . .	177
10.3.1	Schritt 1: Daten erheben. . . . .	177
10.3.2	Schritt 2: Daten bereinigen . . . . .	178
10.3.3	Schritt 3: Daten klassifizieren . . . . .	179
10.3.4	Schritt 4: Daten filtern . . . . .	182
10.3.5	Schritt 5: Vorhersage berechnen. . . . .	183
10.4	Proof-of-Concept . . . . .	186
10.4.1	Methoden- und Technologien-Stack . . . . .	186
10.4.2	Visualisierung der Ergebnisse . . . . .	187
10.5	Fazit . . . . .	188
10.6	Ausblick . . . . .	189
10.6.1	Weitere Analyseansätze . . . . .	189
10.6.2	Anwendungsmöglichkeiten . . . . .	189
	Literatur. . . . .	190
<b>11</b>	<b>Text Mining bei einer wissenschaftlichen Literaturlauswertung: Extraktion von Schlüsselwörtern zur Beschreibung von Inhalten.</b> . . . . .	193
	Thomas Barton und Arthur Kokoev	
11.1	Einführung . . . . .	194
11.2	Explainable Artificial Intelligence . . . . .	194
11.3	Extraktion von Schlüsselwörtern . . . . .	195
11.4	Extraktion von Schlüsselwörtern für eine Literaturlauswertung zu „Explainable AI“ . . . . .	195
11.5	Fazit . . . . .	197
	Literatur. . . . .	198
<b>12</b>	<b>Identifikation relevanter Zusammenhänge in Daten mit maschinellem Lernen.</b> . . . . .	201
	Joshua Hammesfahr und Martin Spott	
12.1	Einleitung . . . . .	202
12.2	Fachliche Problemstellung. . . . .	203
12.3	Ansätze zur Reduzierung von Regelmengen . . . . .	205
12.3.1	Association Rule Discovery. . . . .	206
12.3.2	Subgroup Discovery. . . . .	208
12.4	Gütebestimmung von reduzierten Regelmengen . . . . .	210
12.5	Kombinationssystematik . . . . .	211
12.6	Ergebnisse . . . . .	213
12.7	Zusammenfassung . . . . .	216
	Literatur. . . . .	216

<b>13 Framework für das Management und die Analyse von Fahrzeugdaten für die modellbasierte Fahrerassistenzsystementwicklung in Lehre und Forschung</b> . . . . .	219
Tobias Peuschke-Bischof und Stefan Kubica	
13.1 Motivation . . . . .	220
13.2 Wildauer Maschinen Werke an der TH Wildau . . . . .	221
13.3 Vorstellung Fahrzeugflotte . . . . .	222
13.3.1 Trikes . . . . .	222
13.3.2 Trucks . . . . .	224
13.4 Vorstellung Infrastruktur . . . . .	225
13.4.1 ROS . . . . .	225
13.4.2 Node-RED . . . . .	225
13.4.3 MQTT-Bridge . . . . .	226
13.4.4 ROS Car2X . . . . .	226
13.4.5 Lichtsignalanlagen . . . . .	226
13.4.6 VDI . . . . .	227
13.5 Entwicklungsframework . . . . .	228
13.5.1 Umsetzung Kommunikation Fahrzeuge . . . . .	228
13.5.2 Modellbasierte Entwicklung und Codegenerierung auf Fahrzeuge . . . . .	229
13.5.3 Agiles Projektmanagement, Wissensmanagement und Sourcecode-Management . . . . .	233
13.6 Szenarien-basierte Lehre und Forschung . . . . .	235
13.6.1 ROS Car2X als Datenaggregation und Funktionsverhalten über Fahrzeuge hinweg . . . . .	235
13.6.2 NodeRED zur Datenanalyse . . . . .	238
13.6.3 Interdisziplinäres Szenario am Beispiel Materialwirtschaft . . . . .	239
13.7 Zusammenfassung und Ausblick . . . . .	239
Literatur . . . . .	241
<b>Stichwortverzeichnis</b> . . . . .	243

---

# Herausgeber- und Autorenverzeichnis

---

## Über die Herausgeber

**Prof. Dr. Thomas Barton** studierte und promovierte an der TU Kaiserslautern. Anschließend war er ca. 10 Jahre bei der SAP SE tätig mit Schwerpunkt Anwendungsentwicklung, auch Beratung, Schulung und Projektleitung. Seit 2006 arbeitet er an der Hochschule Worms als Professor für Informatik mit dem Schwerpunkt Wirtschaftsinformatik. Seine Tätigkeitsschwerpunkte liegen in den Bereichen Entwicklung betrieblicher Anwendungen, E-Business, Cloud Computing und Data Science. Er ist Autor und Herausgeber zahlreicher Publikationen. Darüber hinaus ist er in verschiedenen Gremien und Fachausschüssen aktiv. So engagiert er sich auch als Sprecher des GI-Beirates für Hochschulen für Angewandte Wissenschaften.

**Prof. Dr. Christian Müller** studierte Mathematik an der Freien Universität Berlin und promovierte 1989 über Netzwerkflüsse mit Nebenbedingungen. Von 1990 bis 1992 arbeitete er bei der Schering AG und von 1992 bis 1994 bei den Berliner Verkehrsbetrieben (BVG) im Bereich Fahr- und Dienstplan Optimierung. 1994 erhielt er einen Ruf an die Technische Hochschule Wildau, Deutschland. Er ist Professor im Studiengang Wirtschaftsinformatik und Dekan des Fachbereichs Wirtschaft, Informatik und Recht. Seine Forschungsschwerpunkte sind die Konzeption von Informationssystemen, Mathematische Optimierung und die Simulation von Geschäftsprozessen.

---

## Autorenverzeichnis

**Dr. Emal M. Alekozai** Robert Bosch GmbH, Finance Shared Service Digital, Stuttgart, Deutschland

**Jule Aßmann** Hamburg, Deutschland

**Prof. Dr. Thomas Barton** Homburg, Deutschland

**Dr. Jörg Baus** Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Deutschland

**Dr. Boris Brandherm** Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Deutschland

**Dr. Matthieu Deru** Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Deutschland

**Prof. Dr. Ralf Gampfer** Hochschule Worms, Worms, Deutschland

**Prof. Dr. Rainer Gerten** Hochschule Mannheim, Mannheim, Deutschland

**Prof. Dr. Michael Gröschel** Hochschule Mannheim, Mannheim, Deutschland

**Joshua Hammesfahr** Walldorf, Deutschland

**Prof. Dr. Jens Kaufmann** Hochschule Niederrhein, FB Wirtschaftswissenschaften, Mönchengladbach, Deutschland

**Gian-Luca Kiefer** Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Deutschland

**Arthur Kokoev** Worms, Deutschland

**Prof. Dr. Stefan Kubica** Zernsdorf, Deutschland

**Dr. Stephan Kühnel** Martin-Luther-Universität Halle-Wittenberg, Lehrstuhl für Wirtschaftsinformatik, insb. Betriebliches Informationsmanagement, Halle (Saale), Deutschland

**Prof. Dr. Christoph Laroque** Westsächsische Hochschule Zwickau, Fakultät Wirtschaftswissenschaften, Zwickau, Deutschland

**Prof. Dr. Claudia Lemke** Hochschule für Wirtschaft und Recht Berlin (HWR Berlin), Fachbereich 2, Berlin, Deutschland

**Prof. Dr. Olga Matthias** Leeds Beckett University, Leeds Business School, City Campus, Leeds, UK

**Manuel Mikoleit** Gückingen, Deutschland

**Prof. Dr. Dagmar Monett** Hochschule für Wirtschaft und Recht Berlin (HWR Berlin), Fachbereich 2, Berlin, Deutschland

**Prof. Dr. Christian Müller** Berlin, Deutschland

**Dr. Alassane Ndiaye** Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Deutschland

**Dipl.-Inform. Uwe Neuhaus** NORDAKADEMIE Hochschule der Wirtschaft, Elmshorn, Deutschland

**Anna-Maria Nitsche** Westsächsische Hochschule Zwickau, Fakultät Wirtschaftswissenschaften, Zwickau, Deutschland

**Andreas Peuker** Harthausen, Deutschland

**Tobias Peuschke-Bischof** Wildau, Deutschland

**Benedikt Reiner** Hochschule Karlsruhe, Karlsruhe, Deutschland

**Prof. Dr. Daniel Retkowitz** Hochschule Niederrhein, FB Wirtschaftswissenschaften, Mönchengladbach, Deutschland

**Prof. Dr. Gabriele Roth-Dietrich** Hochschule Mannheim, Mannheim, Deutschland

**Prof. Dr. Joachim Sauer** NORDAKADEMIE Elmshorn, Fachbereich Informatik, Elmshorn, Deutschland

**Prof. Dr. Michael Schulz** NORDAKADEMIE Hochschule der Wirtschaft, Elmshorn, Deutschland

**Prof. Dr.-Ing. habil. Christian-Andreas Schumann** Westsächsische Hochschule Zwickau, Fakultät Wirtschaftswissenschaften, Zwickau, Deutschland

**Prof. Dr. Martin Spott** HTW Berlin, Berlin, Deutschland

---

**Teil I**  
**Einleitung**



# Data Science: Vom Begriff zur Anwendung

1

Thomas Barton und Christian Müller

## Inhaltsverzeichnis

1.1 Was ist Data Science? .....	4
1.2 Was ist und was macht ein Data Scientist? .....	5
1.3 Einführung in Data Science .....	7
1.4 Systeme, Werkzeuge und Methoden .....	7
1.5 Anwendungen .....	8
Literatur .....	9

### Zusammenfassung

Dieser Beitrag bietet eine Einführung in den Themenband „Data Science anwenden“, der in der Reihe „Angewandte Wirtschaftsinformatik“ erscheint. Nach einer Einordnung der Begriffe Data Science und Data Scientist erfolgt die Vorstellung der Beiträge, die in den thematischen Schwerpunkten erscheinen. Ausgangspunkt ist eine Einführung in das Themengebiet Data Science. Es folgen verschiedene Beiträge zu Systemen, Werkzeugen und Methoden. Ein Fokus dieses Buches liegt in der Vorstellung konkreter Anwendungen.

T. Barton (✉)  
Homburg, Deutschland  
E-Mail: [barton@hs-worms.de](mailto:barton@hs-worms.de)

C. Müller  
Berlin, Deutschland  
E-Mail: [christian.mueller@th-wildau.de](mailto:christian.mueller@th-wildau.de)

---

**Schlüsselwörter**Data Science · Data Scientist · Methoden · Werkzeuge · Systeme · Anwendungen

---

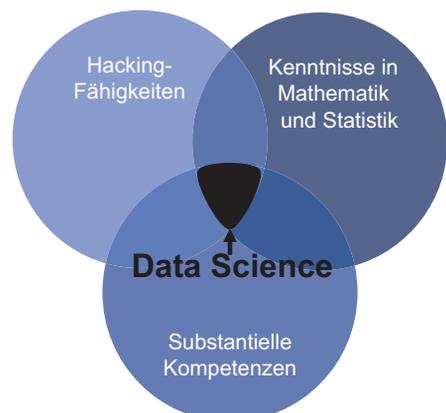
## 1.1 Was ist Data Science?

Eine erste visuelle Darstellung, die den Begriff Data Science auf Basis von Kompetenzen beschreibt, über welche die Menschen verfügen, die diese Disziplin betreiben, ist in Abb. 1.1 dargestellt. Drew Conway zufolge wird Data Science als Schnittmenge dreier Mengen als Venn-Diagramm gezeigt, wobei jede Menge eine Kompetenz der Personen darstellt, die Data Science betreibt [1]. Eine Kompetenz besteht aus Hacking-Fähigkeiten, die Fertigkeiten wie beispielsweise das Bearbeiten von Textdateien auf Ebene der Kommandozeile und ein Verständnis für Vektoroperationen umfassen sowie algorithmisches Denken einschließen. Hierbei wird ein Hacker als eine Person gesehen, die Vergnügen dabei empfindet, ein tieferes Verständnis für das Innere eines Systems, eines Computers oder insbesondere eines Computernetzwerkes zu besitzen [2]. Weitere Kompetenzen sind mathematische und statistische Kenntnisse. Substantielle Kompetenzen vervollständigen das Profil. Sie sollen die Bereitschaft verkörpern, etwas entdecken und Wissen schaffen zu wollen, wofür motivierte Fragen über die Welt benötigt werden, woraus Hypothesen abgeleitet werden können, die mithilfe von Daten überprüft werden.

Aus Sicht eines Statistikers ist Data Science als eine Kombination aus Statistik und Informatik anzusehen [3]. Statistik ist eine der wichtigsten Disziplinen, die Werkzeuge und Methoden liefert, um tiefere Einsichten aus Daten zu gewinnen; sie ist die wichtigste Disziplin, um Unsicherheit zu untersuchen und zu quantifizieren [4].

Für die Deutsche Akademie der Technikwissenschaften acatech lässt sich Data Science in vier Kernbereiche einteilen [5]: Data Engineering, Data Analytics, Data Prediction und maschinelles Lernen. Data Engineering umfasst Prozesse und Methoden

**Abb. 1.1** Venn-Diagramm für Data Science in Anlehnung an [1]



für die Speicherung, für den Zugriff und für die Rückverfolgbarkeit von Daten, während bei Data Analytics die Datenanalyse im Mittelpunkt steht. Die Vorhersage von Themen und Situationen auf Basis von Erfahrungswissen ist Gegenstand von Data Prediction. Maschinelles Lernen wird als Querschnittsbereich zu diesen drei Bereichen gesehen.

Eine neuere Definition für Data Science, die auf einem interdisziplinären Ansatz basiert, stammt aus dem Jahr 2017 [6]:

► Data Science = (Statistik + angewandte Informatik + Computing + Kommunikation + Soziologie + Management | (Daten + Umgebung + Denkweise)).

Data Science stützt sich auf Statistik, angewandte Informatik, Computing, Kommunikation, Management und Soziologie, wobei der Begriff Soziologie für soziale Aspekte stehen soll. Gegenstand ist die Untersuchung von Daten und ihrer Umgebung, wozu insbesondere die zugehörigen Domänen zählen, um Daten in Erkenntnisse und Entscheidungen zu transformieren. Hierbei sollen Denkweisen und Methodiken zugrunde liegen, die es ermöglichen, Erkenntnisse und Wissen auf der Basis von Daten zu generieren. Der senkrechte Strich | bedeutet „unter der Voraussetzung, dass“ und ist aus der bedingten Wahrscheinlichkeit bekannt.

Die Beurteilung von sozialen Aspekten basiert auf der Moral. Moral kann als die „Gesamtheit feststellbarer Verhaltensweisen, Verhaltensnormen und verhaltensbezogener Einstellungen und Werturteile“ verstanden werden [7]. Gegenstand der Ethik ist die Moral. Als Wissenschaft ist Ethik eine Disziplin der Philosophie. Digitale Ethik wird mehr oder weniger als Synonym für Informationsethik angesehen, und „Informationsethik hat die Moral derjenigen zum Gegenstand, die Informations- und Kommunikationstechnologien (IKT) und neue Medien anbieten und nutzen“ [8].

Data Science ist auch Gegenstand von Studiengängen und Weiterbildungsangeboten an Hochschulen. Für die inhaltliche Gestaltung von Studiengängen und Weiterbildungsangeboten für Data Science liegen Vorschläge des GI-Arbeitskreises „Data Science/Data Literacy“ vor [9].

---

## 1.2 Was ist und was macht ein Data Scientist?

Data Scientist wird als attraktivster Job des 21. Jahrhunderts angepriesen [10]. Im Jahr 2015 wurde in der Amtszeit von Präsident Barack Obama ein Chief Data Scientist ernannt [11]. Aber was macht ein Data Scientist, und welche Qualifikation wird dazu benötigt? Die Tätigkeiten eines Data Scientist können sehr vielfältig sein. Um die umfangreichen Kenntnisse zu beschreiben, die Data Scientists für die Durchführung ihres Jobs benötigen, wurde eine Liste von Qualifikationen aufgestellt und verschiedenen Themenbereichen zugeordnet. Eine Darstellung der Qualifikationen in den zugeordneten Themenbereichen ist in Tab. 1.1 gezeigt [12].

**Tab. 1.1** Themenbereiche für Qualifikationen von Data Scientists

Business / Produktent- wicklung	Machine Learning/Big Data	Mathematik/ Operation Research	Programmierung /System- administration	Statistik und Visualisierung
– Business – Produktent- wicklung	– Big Data und verteilte Daten – Machine Learning – Strukturierte Daten – Unstrukturierte Daten	– Algorithmen – Bayes'sche Statistik und Monte-Carlo- Methoden – Grafische Modelle – Mathematik – Optimierung – Simulation	– Back-End- Programmierung – Front-End- Programmierung – System- administration	– Statistik – Umfragen und Marketing – Visualisierung

Hierbei sind achtzehn Qualifikationen fünf Themenbereichen zugeordnet. Unter dem Themenbereich Machine Learning / Big Data sind auch Qualifikationen für strukturierte, unstrukturierte und verteilte Daten zusammengefasst. Der Themenbereich Mathematik / Operation Research beinhaltet auch Qualifikationen für Algorithmen, Bayes'sche Statistik und Monte-Carlo-Methoden, grafische Modelle sowie Simulation. Der Themenbereich Programmierung / Systemadministration schließt neben der Systemadministration die Programmierung von Back-End und Front-End ein. Der Themenbereich Statistik umfasst neben Statistik auch Qualifikationen für Umfragen und Marketing sowie für Visualisierung.

Für Data Scientists lassen sich vier Spezialisierungen ableiten. In Abhängigkeit davon, in welchem Umfang sich die Qualifikationen den fünf Themenbereichen zuordnen lassen, ergeben sich folgende Spezialisierungen [12]:

- Data-Business-Person: Eine Person mit Fokus auf Qualifikationen für Business
- Data Creative: Eine Person, bei der die Qualifikationen ungefähr gleichmäßig auf die fünf Themenbereiche verteilt ist
- Data Developer: Eine Person mit Fokus auf Qualifikationen für Programmierung
- Data Researcher: Eine Person mit vertieften Qualifikationen für Statistik

Gute Data Scientists sollen [13]

- über technische Expertise verfügen, die beispielsweise über den Abschluss eines naturwissenschaftlichen Studiums nachgewiesen werden kann,
- neugierig sein mit einem Verlangen, zu entdecken und in die Tiefe zu gehen, um ein Problem in Hypothesen aufzuschlüsseln, die getestet werden können,

- Storytelling betreiben, indem sie Daten dazu verwenden, um eine Geschichte zu erzählen und diese effektiv zu kommunizieren,
- Cleverness besitzen, um Probleme kreativ und auf unterschiedliche Weise anzugehen.

Darüber hinaus zeichnet sich ein Team, das Data Science durchführt, durch eine intensive kollaborative Zusammenarbeit aus [14].

---

### 1.3 Einführung in Data Science

Die Einführung in das Thema Data Science startet in Kap. 2 mit einem Beitrag der Autoren Jens Kaufmann und Daniel Retkowitz. Ihr Beitrag mit dem Titel „Visualisierung und Deep Learning in der Data Science: Von Daten zu Bildern und von Bildern zu Daten“ führt in die visuelle Aufbereitung von Daten und in die Extraktion von Informationen aus Bildern ein.

Ethische Betrachtungen spielen in der digitalen Transformation von Unternehmen eine immer größere Rolle. „Digitale Ethik in datengetriebenen Organisationen und deren Anwendung am Beispiel von KI-Ethik“ heißt der Beitrag von Claudia Lemke, Dagmar Monett und Manuel Mikoleit. Digitale Ethik und ihre Anwendung in intelligenten Systemen eines Unternehmens ist Gegenstand ihres Beitrages in Kap. 3.

Die digitale Transformation führt zu der Implementierung technologischer Lösungen zur Unterstützung der Entscheidungsfindung. Anna-Maria Nitsche, Christian-Andreas Schumann, Christoph Laroque und Olga Matthias stellen ein Prozessrahmenwerk vor für die Umsetzung und Bewertung dieser Technologien. Ihr Beitrag heißt „Multiple Perspektiven bei der Implementierung innovativer technologischer Lösungen im Kontext datengesteuerter Entscheidungsfindung“ und umfasst Kap. 4.

Untersuchungen zum Scheitern von Data-Science-Projekten sind Gegenstand des Beitrages von Jule Aßmann, Joachim Sauer und Michael Schulz. Ihr Beitrag, der Kap. 5 zugeordnet ist, hat den Titel „Keine Angst vor Fehlschlägen – Erkenntnisse aus einer Umfrage zum Scheitern von Data-Science-Projekten“.

---

### 1.4 Systeme, Werkzeuge und Methoden

Der erste Beitrag in dieser Rubrik stammt von Andreas Peuker und Thomas Barton. In ihrem Beitrag in Kap. 6 mit dem Titel „Empfehlungssysteme und der Einsatz maschineller Lernverfahren“ stellen die Autoren die Grundlagen und den Einsatz von Empfehlungssystemen unter Berücksichtigung aktueller Entwicklungen vor.

Gabriele Roth-Dietrich, Michael Gröschel und Benedikt Reiner vergleichen BI-Systeme bezüglich ihrer Eignung, Funktionalität aus dem Bereich Machine Learning

für Fachanwendungen bereitzustellen. Kap. 7 besteht aus ihrem Beitrag „Vergleich der Machine-Learning-Funktionalitäten von Business Intelligence- und Analytics-Tools“.

Der Beitrag des Autorenteam Emal M. Alekozai, Jens Kaufmann, Stephan Kühnel, Uwe Neuhaus, Michael Schulz in Kap. 8 hat den Einsatz eines Vorgehensmodells für Data Science zum Gegenstand. Der Titel des Beitrages lautet „Data-Science-Projekte mit dem Vorgehensmodell ‚DASC-PM‘ durchführen: Kompetenzen, Rollen und Abläufe“.

---

## 1.5 Anwendungen

Die vorgestellten Anwendungen befassen sich mit den Themen erneuerbare Energien und Optimierung des Energiemanagements, wissenschaftliche Literaturobwertung, Kundenzufriedenheit in der Automobilindustrie und Fahrerassistenzsystementwicklung.

Die Energiewende mit dem Ausstieg aus der Kernenergie und ihre Herausforderungen bezüglich des Stromsystems in Deutschland sind Ausgangspunkt für den Beitrag von Boris Brandherm, Matthieu Deru, Alassane Ndiaye, Gian-Luca Kiefer, Jörg Baus und Ralf Gampfer. Die Integration erneuerbarer Energien am Beispiel von Photovoltaikanlagen und der Einsatz von künstlicher Intelligenz ist Gegenstand ihres Beitrages „Integration erneuerbarer Energien – KI-basierte Vorhersageverfahren zur Stromerzeugung durch Photovoltaikanlagen“ in Kap. 9.

Nachhaltigkeit und ein effizienterer Umgang mit Wärmeenergie ist Ausgangspunkt für den Beitrag von Gabriele Roth-Dietrich und Rainer Gerten. Ihr Beitrag „Machine Learning für die Energiemanagementoptimierung“ beschreibt die Optimierung einer Klimatisierungsanlage mithilfe von Data Science anhand eines konkreten Projektes. Kap. 10 besteht aus diesem Beitrag.

Kap. 11 zeigt, wie die Analyse von Publikationen im Rahmen einer wissenschaftlichen Literaturobwertung mithilfe von Text Mining unterstützt werden kann. Dazu stellen die Autoren Thomas Barton und Arthur Kokoev ein konkretes Anwendungsbeispiel vor. Ihr Beitrag heißt „Text Mining bei einer wissenschaftlichen Literaturobwertung: Extraktion von Schlüsselwörtern zur Beschreibung von Inhalten“.

Die Verwendung von Daten zur Ermittlung von relevanten Zusammenhängen ist Gegenstand des Beitrages der Autoren Joshua Hammesfahr und Martin Spott. Im Rahmen ihres Beitrages „Identifikation relevanter Zusammenhänge in Daten mit maschinellem Lernen“ stellen sie einen Zusammenhang zwischen einerseits Konfigurationen von Produkten bzw. einer Infrastruktur und andererseits Fehlern her, um Fehler zu reduzieren. Kap. 12 stellt diesen Beitrag vor.

Das Management und die Analyse von Fahrzeugdaten in einer digitalen Lernfabrik ist Gegenstand des Beitrages von Tobias Peuschke-Bischof und Stefan Kubica. Ihr Beitrag in Kap. 13 mit dem Titel „Framework für das Management und die Analyse von Fahrzeugdaten für die modellbasierte Fahrerassistenzsystementwicklung in Lehre und Forschung“ veranschaulicht Mechanismen bei der Entwicklung und beim Projektmanagement.

## Literatur

1. Conway D (2010) The data science venn diagram. <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>. Zugegriffen: 4. Febr. 2021
2. Malkin G, Parker TL (1993) Request for comments 1392. Internet users' glossary. <https://tools.ietf.org/html/rfc1392>. Zugegriffen: 4. Febr. 2021
3. Kauermann G (2019) Data Science – Einige Gedanken aus Sicht eines Statistikers. Informatik Spektrum 42(6):387–393
4. Weihs C, Ickstadt K (2018) Data science: the impact of statistics. Int J Data Sci Analytics 6:189–194
5. Gausemeier J, Guggemos M, Kreimeyer A (Hrsg) (2017) Auswahl, Beschreibung, Bewertung und Messung der Schlüsselkompetenzen für das Technologiefeld Data Science. acatec. [https://www.acatech.de/wp-content/uploads/2019/02/acatech\\_NKM\\_Data\\_Science\\_WEB-2.pdf](https://www.acatech.de/wp-content/uploads/2019/02/acatech_NKM_Data_Science_WEB-2.pdf). Zugegriffen: 4. Febr. 2021
6. Cao L (2017) Data science: a comprehensive overview. ACM Comput Surv 50(3): Article 43
7. Werner MH (2021) Einführung in die Ethik. Metzler Verlag. <https://doi.org/10.1007/978-3-476-05293-3>
8. Bendel O (2019) 400 Keywords Informationsethik, 2. Aufl. Springer Gabler Wiesbaden
9. GI (2019) Data Science: Lern- und Ausbildungsinhalte (Arbeitspapier). [https://gi.de/fileadmin/GI/Allgemein/PDF/GI\\_Arbeitspapier\\_Data-Science\\_2019-12\\_01.pdf](https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Arbeitspapier_Data-Science_2019-12_01.pdf). Zugegriffen: 4. Febr. 2021
10. Davenport TH, Patil DJ (2012) Data scientist: the sexiest job of the 21st century. Harv Bus Rev 90(5):70–76
11. The White House (2015) The White House Names Dr. DJ Patil as the First U.S. Chief Data Scientist. <https://obamawhitehouse.archives.gov/blog/2015/02/18/white-house-names-dr-dj-patil-first-us-chief-data-scientist>. Zugegriffen: 23. Jan. 2021
12. Harris HD, Murphy SP, Vaisman M (2013) Analyzing the analyzers. O'Reilly Media, Sebastopol
13. Patil DJ (2011) Building data science teams. O'Reilly Media, Sebastopol
14. Zhang AX, Muller M, Wang D (2020) How do data science workers collaborate? Roles, workflows, and tools. Proc ACM on Hum-Comput Interact 4(CSCW1):1–23

**Thomas Barton** studierte und promovierte an der TU Kaiserslautern. Anschließend war er ca. 10 Jahre bei der SAP SE tätig mit Schwerpunkt Anwendungsentwicklung, auch Beratung, Schulung und Projektleitung. Seit 2006 arbeitet er an der Hochschule Worms als Professor für Informatik mit dem Schwerpunkt Wirtschaftsinformatik. Seine Tätigkeitsschwerpunkte liegen in den Bereichen Entwicklung betrieblicher Anwendungen, E-Business, Cloud Computing und Data Science. Er ist Autor und Herausgeber zahlreicher Publikationen. Darüber hinaus ist er in verschiedenen Gremien und Fachausschüssen aktiv. So engagiert er sich auch als Sprecher des GI-Beirates für Hochschulen für Angewandte Wissenschaften.

**Christian Müller** studierte Mathematik an der Freien Universität Berlin und promovierte 1989 über Netzwerkflüsse mit Nebenbedingungen. Von 1990 bis 1992 arbeitete er bei der Schering AG und von 1992 bis 1994 bei den Berliner Verkehrsbetrieben (BVG) im Bereich Fahr- und Dienstplan Optimierung. 1994 erhielt er einen Ruf an die Technische Hochschule Wildau, Deutschland. Er ist Professor im Studiengang Wirtschaftsinformatik und Dekan des Fachbereichs Wirtschaft, Informatik und Recht. Seine Forschungsschwerpunkte sind die Konzeption von Informationssystemen, Mathematische Optimierung und die Simulation von Geschäftsprozessen.

---

**Teil II**

**Einführung in Data Science**



# Visualisierung und Deep Learning in der Data Science

# 2

Von Daten zu Bildern und von Bildern zu Daten

Jens Kaufmann und Daniel Retkowitz

## Inhaltsverzeichnis

2.1	Einleitung	14
2.2	Verfahren für die visuelle Aufbereitung von Daten	15
2.2.1	Einfache Daten und Texte darstellen	15
2.2.2	Komplexe Daten vereinfachen und darstellen	18
2.3	Bildinformationen extrahieren	22
2.3.1	Visuelle Strukturen mit Deep Learning erkennen	24
2.3.2	Architekturen für die Praxis	25
2.4	Zusammenführung von Bild und Daten	25
2.4.1	Generierung anreichernder Detailinformationen	25
2.4.2	Transformation visueller Repräsentationen	26
2.4.3	Einsatzmöglichkeiten	28
2.5	Zusammenfassung	29
	Literatur	29

## Zusammenfassung

Menschen nehmen visuell aufbereitete Informationen schnell und effizient wahr. Die Aufbereitung von Daten jeder Art ist daher ein wichtiges und stark beforschtes Gebiet in der Data Science und all ihren umliegenden Feldern. Dabei gilt es, komplexe

J. Kaufmann (✉) · D. Retkowitz  
Hochschule Niederrhein, FB Wirtschaftswissenschaften,  
Mönchengladbach, Deutschland  
E-Mail: [jens.kaufmann@hs-niederrhein.de](mailto:jens.kaufmann@hs-niederrhein.de)

D. Retkowitz  
E-Mail: [daniel.retkowitz@hs-niederrhein.de](mailto:daniel.retkowitz@hs-niederrhein.de)

Informationen so weit zu vereinfachen, dass ohne signifikanten Sinnverlust die Kerninformationen möglichst einfach und anschaulich transportiert werden können. Im Gegenzug besteht eine steigende Notwendigkeit, Bilder und Bildinformationen automatisch zu verarbeiten, sei es für die Gesichtserkennung als biometrisches Merkmal, für persönliche Assistenten oder für die Auswertung der Kamerabilder beim fahrerlosen Auto. Der Beitrag stellt dar, welche Möglichkeiten jeweils existieren und wie Algorithmen, z. B. im Deep Learning, sich selbst oder gegenseitig trainieren und verbessern.

---

**Schlüsselwörter**

Visualisierung · Deep Learning · Data Science · Generative Adversarial Networks · Dimensionsreduzierung · Bilddaten

---

## 2.1 Einleitung

Das menschliche Gehirn besitzt eine Kernkompetenz in der Mustererkennung. Seine Leistungsfähigkeit ist so hoch, dass die modernen Verfahren zur Bild- und Objekterkennung auf einer Nachbildung dieses Organs basieren, den künstlichen neuronalen Netzen (KNN). Tief verschachtelt, mit diversen Anpassungen, Ergänzungen und Spezialisierungen liefern Anwendungen im Bereich Deep Learning inzwischen in kürzester Zeit Ergebnisse, die mit denen des menschlichen Gehirns konkurrieren und sie z. B. bei der Auswertung von medizinisch genutzten Fotografien übertreffen. Gleichzeitig beschäftigt sich die Data Science in der Gegenrichtung mit der Frage, wie komplexe Datenstrukturen, vieldimensionale Datenmengen und umfangreiche Analyseergebnisse möglichst ansprechend, verständlich und interpretierbar für die Anwender der Systeme gestaltet werden können.

Dieser Beitrag zeigt in Abschn. 2.2, welche Verfahren existieren, um komplexe Daten visuell aufzubereiten, und stellt dar, wie dimensionsreduzierende Verfahren z. B. Ähnlichkeiten in Daten und statistische Verteilungen nutzen, um Informationsverlust zu vermeiden. Es wird beschrieben, welche Algorithmusparameter die Anwender für welchen Anwendungszweck gestalten können und wie die Ergebnisse praktisch zu interpretieren sind.

Im Gegenzug zur Datendarstellung kann Deep Learning aus visuell vorliegenden Informationen Strukturen auslesen und Objekte erkennen. Auch hier haben Anwender umfangreiche Konfigurationsmöglichkeiten, um mit vertretbarem (Rechen-)Aufwand Antworten auf betriebliche Fragestellungen zu erhalten. Die entsprechenden Betrachtungen und Hinweise zur praktischen Anwendung finden sich in Abschn. 2.3.

In der Kombination von Bildgenerierung und Bilderkennung befinden sich auch sogenannte Generative Adversarial Networks (GAN). Sie sind ein Beispiel dafür, wie KNN eingesetzt werden können, um möglichst perfekte Nachahmungen existierender

Grafiken zu erzeugen oder mit „Deep Fakes“ vertraute Menschen nachzubilden. Abschn. 2.4 stellt vor, welche Auswirkungen dies auf gute Datenvisualisierungen, aber auch auf Sicherheitsaspekte in Unternehmen haben kann.

---

## 2.2 Verfahren für die visuelle Aufbereitung von Daten

Daten können – auch ohne visuelle Aufbereitung – diverse Formen annehmen. Zahlenreihen, Textdaten, Finanztabellen, Bilddaten etc. sind nur einige Beispiele. Dieser Abschnitt befasst sich vorwiegend mit der grafischen Aufbereitung von Daten, die originär in numerischer Form vorliegen. Ergänzend wird die Visualisierung von Text behandelt. Dabei wird grundsätzlich nur eine Schwarz-Weiß-Darstellung gezeigt. Während moderne Visualisierungen auch durch den Einsatz von Farben Informationen transportieren und vor allem ästhetisch ansprechender sind, basieren alle Darstellungsformen zunächst auf Basisprinzipien und Strukturen, die in der Regel farbumabhängig sind und auch mit Grautönen problemlos realisiert werden können. Vorteilhaft ist hier, dass eine nachträgliche Farbgebung immer möglich, aber eben nicht erforderlich ist. So kann z. B. sichergestellt werden, dass die Darstellungen auch bei Schwarz-Weiß-Ausdrucken und unabhängig vom Format gut lesbar bleiben. Zudem wird das Problem umgangen, dass nicht immer bekannt ist, ob die Grafikempfänger eine Farbfehlsichtigkeit aufweisen, z. B. eine Rot-Grün-Schwäche. Insbesondere bei den dargestellten Verfahren für komplexe Daten liegt der Fokus dieses Beitrags daher auf der Methodik und deren Verständnis, nicht aber auf Beispielen für möglichst farbintensive Grafiken.

### 2.2.1 Einfache Daten und Texte darstellen

Einfach strukturierte Daten liegen im (betrieblichen) Alltag in der Regel umfangreich vor. Beispiele sind Verkaufszahlen, Produkteigenschaften, Produktionsdurchlaufzeiten, Benzinverbrauch oder Infektionszahlen in der Bevölkerung eines Landes. Für die Darstellung dieser Daten bieten sich diverse Standarddiagramme an, die in der deskriptiven Statistik genutzt werden. Da Data Science zu einem guten Teil die Anwendung von Statistik ist, sollte jeder Data Scientist auch mit den entsprechenden Diagrammtypen vertraut sein. Abb. 2.1 zeigt als Auswahl sechs unterschiedliche Diagramme eines Datensatzes (ohne dessen nähere Beschreibung).

In der oberen Reihe findet sich zunächst ein *Säulendiagramm* (a). Für mehrere Elemente (dargestellt auf der Abszisse, der x-Achse) wird hier die Häufigkeit des Auftretens auf der Ordinate, der y-Achse, angegeben. Stellen die einzelnen Säulen nicht einzelne Elemente (beispielsweise einzelne Menschen mit einer bestimmten Körpergröße), sondern Klassen dar (155–165 cm, 166–175 cm ...), handelt es sich um ein *Histogramm*. Das *Punktendiagramm* (b) ist eine Variante, bei der die einzelnen Einträge deutlicher dargestellt werden. Zuletzt stellt das *Balkendiagramm* (c) die um 90 Grad