

Detlev Frick · Andreas Gadatsch
Jens Kaufmann · Birgit Lankes
Christoph Quix · Andreas Schmidt
Uwe Schmitz *Hrsg.*

Data Science

Konzepte, Erfahrungen, Fallstudien und Praxis

EBOOK INSIDE



Springer Vieweg

Data Science

Detlev Frick · Andreas Gadatsch ·
Jens Kaufmann · Birgit Lankes · Christoph Quix ·
Andreas Schmidt · Uwe Schmitz
(Hrsg.)

Data Science

Konzepte, Erfahrungen, Fallstudien
und Praxis

Hrsg.

Detlev Frick
FB Wirtschaftswissenschaften
Hochschule Niederrhein
Mönchengladbach, Deutschland

Andreas Gadatsch
FB Wirtschaftswissenschaften
Hochschule Bonn-Rhein-Sieg
Sankt Augustin, Deutschland

Jens Kaufmann
FB Wirtschaftswissenschaften
Hochschule Niederrhein
Mönchengladbach, Deutschland

Birgit Lankes
FB Wirtschaftswissenschaften, FH Niederrhein
Mönchengladbach, Deutschland

Christoph Quix
FB Elektrotechnik/Informatik
Hochschule Niederrhein
Krefeld, Deutschland

Andreas Schmidt
FB Wirtschaftswissenschaften
Hochschule Bonn-Rhein-Sieg
Sankt Augustin, Deutschland

Uwe Schmitz
FB Wirtschaft, FH Dortmund
Dortmund, Deutschland

ISBN 978-3-658-33402-4 ISBN 978-3-658-33403-1 (eBook)

<https://doi.org/10.1007/978-3-658-33403-1>

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/Lektorat: 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

Geleitwort: Den Menschen im Fokus – Datenschutz als Erfolgsfaktor für Big Data Technologien

Big Data ist nun wirklich kein neuer Trend mehr. Big Data befasst sich mit der Frage, wie enorme Mengen unterschiedlichster Daten aus unterschiedlichsten Quellen möglichst in Echtzeit so ausgewertet werden können, dass sich hierauf tragfähige Entscheidungen stützen lassen. Es geht hier also immer darum, Bestehendes zu nutzen, um daraus Mehrwerte zu generieren. Wir reden über enorme Chancen. Chancen, das Unsichtbare sichtbar zu machen. Und dies idealerweise zum Wohle aller.

Diese Chancen gehen mit Risiken einher. Ich denke, wir sind gut beraten, die Chancen und Risiken technologischer Entwicklungen differenziert zu betrachten, insbesondere, wenn hierbei personenbezogene Daten genutzt werden.

Personenbezogene Daten sind im höchsten Maße individuelle Informationen. Gerade in der digitalisierten Welt haben personenbezogene Daten eine herausragende Bedeutung. Denn sie fallen hier besonders viel und vielfältig an, sie bilden unser Leben ab. Digitalisierung wird integraler Bestandteil unseres Lebens – und damit auch die digitale Datenvielfalt.

Es ist auch gerade erst diese Individualität, aus der sich der – oft auch wirtschaftliche – Nutzen personenbezogener Daten speist. Diese Individualität ist es aber auch, die diese Daten besonders schützenswert macht. Deshalb ist es bei allen datengetriebenen Geschäftsmodellen mit Personenbezug wichtig, die Interessen des Individuums stets und zuallererst im Blick zu behalten.

Mir ist wichtig, dass der Datenschutz nicht als destruktives Element verstanden wird. Es geht ihm nicht darum, Innovationen einzuschränken oder zu erschweren. Datenschutz sucht vielmehr den Ausgleich. Den Ausgleich zwischen den Interessen einer Datennutzung durch Dritte und des Grundrechts der Souveränität eines jeden Einzelnen. Dies wird schnell vergessen, wenn vom „neuen Öl“ die Rede ist und Datenschutz fälschlicherweise als Bremschuh für Innovationen gebrandmarkt wird.

Es ist vielmehr so: Chancen nutzen heißt auch, den Datenschutz als Erfolgsfaktor zu verstehen. Europäischer Datenschutz ist kein Show-Stopper, sondern kann globaler Game-Changer werden. Denn der europäische Datenschutz bietet zahlreiche

gute Gestaltungsmöglichkeiten für den skizzierten Interessenausgleich, etwa die Anonymisierung oder Pseudonymisierung personenbezogener Daten.

Wir alle tun gut daran, die Privatsphäre zu schützen, auch und gerade um einen Freiraum zur unbeobachteten persönlichen Entfaltung zu belassen. Der Datenschutz hat also stets das Individuum im Fokus und schafft damit gleichzeitig die Voraussetzungen für eine freiheitliche digitale Gesellschaft. Wer Innovationen mit Big Data schaffen will, die dem Menschen dienen, berücksichtigt deshalb naturgemäß die Regeln des Datenschutzes. Proaktiver Datenschutz, also den Schutz der individuellen Daten von Anfang an mitdenken, ist deshalb keine Innovationsbremse, sondern das Mittel um Vertrauen in neue Technologien und ihre Möglichkeiten zu schaffen, das notwendig ist, um sie erfolgreich in den Markt zu bringen. Dieses Vertrauen ist dann tatsächlich das „neue Öl“ für digitale Technologien.

Bonn

Prof. Ulrich Kelber

Bundesbeauftragter für den Datenschutz und die Informationsfreiheit

Vorwort

„Daten sind das neue Öl, aber Informationen sind das echte Gold!“ Das Schürfen dieses Goldes bedarf im digitalen Zeitalter keiner Westernmentalität, sondern neben technischer Lösungen ausgereiftem Fachwissen und digitaler Kompetenzen. Daten müssen effizient verwaltet, im Unternehmen systematisch analysiert und zur Digitalisierung von Geschäftsmodellen erfolgreich implementiert werden. Diese neuen Herausforderungen für Fachkräfte und Experten in Wirtschaft und Wissenschaft sowie fortgeschrittene Studierende mit Interesse an Big Data und Data Science werden in diesem Buch erstmals in den drei Rollen des „Data Strategist“, „Data Architect“ und „Data Analyst“ auf wissenschaftlichem Niveau mit dem erforderlichen Praxisbezug beschrieben.

Die Autoren sind dazu nicht zufällig ausgewählt worden, sondern kommen aus einem gemeinsamen Projekt der wissenschaftlichen Weiterbildung für die digitale Wirtschaft an der Hochschule Niederrhein in Kooperation mit der Hochschule Bonn-Rhein-Sieg und der FH Dortmund. In diesem auf den Bereich Data Science fokussierten Projekt wurde zu jeder der drei genannten Rollen ein sog. Certificate of Advanced Studies (CAS) als wissenschaftliches Weiterbildungsangebot für die Wirtschaft entwickelt, das sich aus einzelnen Zertifikatskursen zusammensetzt.

In moderierten Workshops wurde ausgehend von den für die genannten Tätigkeiten erforderlichen Kompetenzen die Inhalte, Fallbeispiele und Methoden für die Kurse erarbeitet und die einzelnen Curricula passgenau aufeinander abgestimmt. Über die Qualitätssicherung in Form zu genehmigender Prüfungsordnungen und Modulbeschreibungen für jeden Kurs wird die Kompetenz der Autoren durch die Fachgruppen der beteiligten Fachbereiche der Hochschulen weiter ergänzt. Die Kursdurchführung wurde vom Projektteam wissenschaftlich begleitet, intensiv evaluiert und abschließend ausgewertet. Diese Ergebnisse und Erfahrungen sind hier in dem vorliegenden Buch aufbereitet und um anwendungsorientierte Praxisbeispiele zielführend ergänzt. Der Leser erhält somit nicht nur eine wissenschaftliche Beschreibung der Fachkräfteprofile im Markt von Big Data und Data Science aus der Sicht von Lehrenden von drei Hochschulen für angewandte Wissenschaften mit ihrem expliziten Praxisbezug. In diesem Buch sind auch die Erfahrung der ersten Kursdurchführung und das Feedback der

Teilnehmer, die sämtlich aus der Wirtschaft mit einer entsprechenden Berufspraxis ausgewählt wurden, enthalten. Somit ist der Praxisbezug von zwei Seiten garantiert, aus der anwendungsorientierten Wissenschaft und aus der Berufspraxis.

Ein bunter Strauß spannender Themen wartet auf den Leser. Die Implementierung von Big Data-Technologien, die Gestaltung von Informationsarchitekturen und systematische Analyse von Unternehmensdaten bis hin zur Konzeption datenbasierter Geschäftsmodelle, alle Bereiche werden so praxisnah dargestellt, das Fach- und Führungskräfte aller Branchen die erworbenen Kenntnisse und Fähigkeiten direkt in ihrem Aufgabengebiet einsetzen können. Diese Lektüre wird sich lohnen!

Prof. Dr. Thomas Meuser
Leitungsteam des Cyber Management Campus Mönchengladbach
der Hochschule Niederrhein

Grußwort

Im Zuge der digitalen Transformation investieren viele Unternehmen in den Aufbau von Dateninfrastrukturen und Data Science-Teams, die den Weg zum „datengetriebenen“ Unternehmen ebnen sollen. Dies ist die konsequente Weiterentwicklung der klassischen Business Intelligence und scheint ein kleiner Schritt zu sein, schließlich ist man die strukturierte Arbeit mit Daten bereits gewohnt.

In der Praxis tauchen jedoch ungewohnte Hindernisse auf: Die angehäuften Daten müssen erst einmal zugänglich gemacht werden oder in auswertungsfähige Formate verwandelt werden. Dies erfordert Personen, die Bild- oder Textdaten in strukturierte Tabellen transformieren, um entscheidungsrelevante Informationen daraus generieren zu können.

Statistische Applikationen werden zudem in relativ neuen Software- und Programmierungsumgebungen entwickelt. Wenn eine solche Anwendung in den täglichen Betrieb überführt werden soll, muss diese in die Konzern-IT eingebettet werden und plötzlich müssen viele Anforderungen aus Governance- oder Compliance-Sicht erfüllt werden: Daran scheitern leider viele gute Data Science-Ideen.

Der Schlüssel zum datengetriebenen Unternehmen ist daher Multidisziplinarität. Neues Wissen und neue Rollen sind notwendig, um Data Science erfolgreich zu machen. Gleichzeitig müssen Prozesse, Organisation und IT-Strukturen überdacht werden, so dass sich der erwähnte „kleine Schritt“ sehr schnell auf die gesamte Enterprise Architektur auswirken kann.

Junge Unternehmen demonstrieren, wie eine analytische Organisation von Anfang an aufgebaut werden kann. Viele etablierte Unternehmen müssen dies oft noch lernen, um den Anschluss nicht zu verlieren. Da nicht genug auf diese Herausforderungen der Digitalisierung hingewiesen werden kann, freue ich mich über dieses Werk, das informiert, sensibilisiert und viele praktische Tipps beinhaltet!

Ulrich Dommer
Partner Consulting, KPMG AG Wirtschaftsprüfungsgesellschaft, Düsseldorf

Grußwort: Data Science – Weiterbildung für die Zukunft

Data Science ist nicht länger nur ein anhaltender Trend, sondern inzwischen auch in der Praxis angekommen. Viele Unternehmen setzen mathematisch-statistische Methoden sowie verschiedene Formen Künstlicher Intelligenz ein, um neue Geschäftsmodelle zu entwickeln, Prozesse zu optimieren und neue Formen der Kundeninteraktion einzuführen.

Das verfügbare Potenzial an Wissen in den Unternehmen reicht Stand heute häufig nicht aus. Es mangelt vielfach an Personal in der IT und in den Fachabteilungen. Daraus resultierend lassen sich zahlreiche Möglichkeiten von Data Science noch nicht ausschöpfen.

Der vorliegende Sammelband „Data Science“ spannt einen weiten Bogen, in dem er sich mit der gleichnamigen Thematik aus verschiedenen Perspektiven und auf mehreren Ebenen auseinandersetzt. Die historische Einführung aus Sicht der Wissenschaft ist für die Einordnung vieler Spezialthemen von Bedeutung. Die Fachbeiträge aus Wissenschaft und Praxis decken vielfältige Themenfelder ab und bieten einen spannenden Einblick in die vielfältigen Möglichkeiten der zukunftsweisenden Data Science.

Das Buch baut auf den wissenschaftlichen Zertifikatskursen zu „Big Data und Data Science“ der Hochschule Niederrhein auf. Die Autorenschaft kommt sowohl aus dem wissenschaftlichen als auch aus dem praktischen Umfeld, was das Werk für eine breite Zielgruppe besonders interessant macht. So kann es gleichermaßen als Einführung wie auch als Vertiefung oder als Nachschlagewerk genutzt werden.

Als Anbieter von IT-Lösungen und Services sind wir auf sehr gut ausgebildete Mitarbeiterinnen und Mitarbeiter angewiesen, ohne die wir unsere Leistungen nicht erbringen könnten. Wir wünschen dem Werk daher eine hohe Verbreitung in der wissenschaftlichen Ausbildung und in der Praxis.

Waldemar Zgrzebski
Geschäftsführer, Bechtle GmbH & Co. KG, IT-Systemhaus Bonn/Köln

Data Science – Entwicklungslinien und Trends

Zusammenfassung

Data Science ist als Begriff genauso viel oder wenig Trend, Hype oder Mode wie viele andere Begriffe der (Wirtschafts-)Informatik zuvor. Sie steht in bester Tradition aller Forschungs- und Anwendungsfelder der letzten Jahrzehnte, die sich mit der Generierung und Bereitstellung entscheidungsrelevanter Informationen befasst haben. Im ständigen Wechselspiel zwischen Technologieorientierung und Geschäftsorientierung folgt die Data Science den aktuellen technischen Möglichkeiten und umfassenden Datenbeständen und nutzt als übergreifendes Konstrukt die Methoden maschinellen Lernens genauso wie die geschäftlich motivierten Analysemethoden von Business Analytics. Schafft sie es, die vielen unterschiedlichen Disziplinen erfolgreich zu einem Zusammenwirken zu bewegen, zeichnet sich eine vielversprechende Zukunft für die Data Science ab.

Will man die Entwicklungslinien von innovativen Verfahren in der Informatik nachvollziehen, muss man in der Regel deutlich weiter in die Vergangenheit gehen als man gemeinhin vermutet. So hat sich bei fast allen Hype-Themen gezeigt, dass sie selten bahnbrechend oder überraschend neu sind. Wie in allen anderen Wissenschaften bauen neue Erkenntnisse auf den Errungenschaften früherer Generationen auf – seien es geglückte oder weniger geglückte Ansätze ehemaligen Ideenreichtums. Für die Informatik ist es symptomatisch, dass neue Ideen sich wegen noch fehlender technologischer Unterstützung nicht durchsetzen können, später unter besseren Rahmenbedingungen aber dann als Innovationen umsetzbar werden. Hinzu kommt ein Effekt, dass geschickte Kombinationen von verfügbaren Algorithmen, Verfahren oder Technologien diese Innovationen erst entstehen lassen. Typisch für die IT-Branche mit ihren stark umworbenen Märkten ist die Proklamation solcher Hype-Themen in kurzen Abständen. Der Wirtschaftsinformatik bleibt dann meist nur die nachträgliche Klärung, ob es sich um Trends oder Moden handelte.

Folgt man den Hype-Zyklen von Gartner zur vermeintlichen Vorhersage von aufkommenden Technologien, so vermittelt sich ein durchaus dynamisches Bild dieser Wellen. Über viele Jahre konnte man einen festen Trend und damit auch Wachstumstreiber feststellen. Dabei war und ist die Bereitstellung entscheidungsrelevanter

Informationen für das Management von Unternehmen und Organisationen offensichtlich ein andauerndes Problemfeld, das buchstäblich Generationen von (Wirtschafts-) Informatikern beschäftigt hat. Gerade hier ist die Abfolge der „neuen“ Ansätze zur Problemlösung auffällig. Die Pendelbewegung zwischen Technologieorientierung und Businessorientierung findet man zum Beispiel u. a. bei den Begriffen „Data Warehousing“ und „Business Intelligence“ um die Jahrtausendwende. Der seit einigen Jahren andauernde Hype um „Data Science“ schließt sich hier lückenlos an – diesmal mit einem Ausschlag des Pendels in Richtung Technologie, insbesondere Daten und Algorithmen.

Eigentlich ist der Begriff schon in den Anfängen der Informatik durch Peter Naur in den 1960er Jahren geprägt worden. Der algorithmische Kern entspringt den bekannten Verfahren der Statistik und der künstlichen Intelligenz (KI), welche zum maschinellen Lernen geführt haben. Wiederum haben die Entwicklung von KI und Expertensystemen in den 1980er Jahren die prinzipielle Machbarkeit gezeigt, sind aber nicht zur Marktreife gekommen. Lediglich in Marktnischen hielten sich derartige wissensbasierte Systeme, ansonsten sorgte das Scheitern für ein signifikantes Abschwellen der KI-Welle. Erst mit dem Aufkommen der ersten Erfolge von „Big Data“ war das Interesse wieder vorhanden. Datafication oder die allumfassende Sammlung und Auswertung von polystrukturierten Daten beliebigen Formats in Echtzeit haben gezeigt, dass sich entscheidungsrelevante Informationen aus diversen Quellen generieren lassen. Die vielfältigen Anwendungsgebiete, welche anfänglich noch als „use cases“ krampfhaft gesucht wurden, überzeugten nicht zuletzt im Bereich „social media analytics“. Die hierzu eingesetzten „Data Scientists“ bei den Internetgiganten Google, Amazon etc. prägten ein neues Berufsbild, welches vertiefte Statistik- und KI-Kenntnisse forderte. Verbunden mit den hybriden Datenhaltungskonzepten aus klassischem Data Warehousing und Hadoop-Clustern sowie parallelen Hochleistungsrechnern formierten sich die neuen digitalen Ökosysteme. Eine frühere Ausrede der Systementwickler, dass die Algorithmen verfügbar wären, aber die Daten nicht, entfiel damit. Somit brachte die zweite KI-Welle unter dem Big-Data-Dach eine Renaissance der Künstlich Neuronalen Netze (KNN). Die Mustererkennung mit Deep Learning, wobei zahlreiche Zwischenschichten in die KNN eingesetzt werden, ist vielversprechend und überzeugend. Wiederum sind die Verfahren nicht neu, aber aufgrund der Datenverfügbarkeit und Rechengeschwindigkeit nun einsetzbar. Der digitale Fußabdruck jeglicher Objekte und Subjekte ist per Datenanalyse aufspürbar, was für zunehmenden Sprengstoff in der öffentlichen Diskussion um ethische Grundfragen sorgt. Somit hat die technologische Lösung unseres Informationsproblems zu einem neuen Problem bei der Informationsnutzung geführt. Hinzu kommen vielschichtige Fragestellungen um den Einsatz von KI in autonomen Systemen, die bisher nur ansatzweise beantwortet sind.

Die Entwicklungslinien von Data Science hängen direkt mit der Entstehung der Statistik und der künstlichen Intelligenz in Form von maschinellem Lernen zusammen. Gepaart mit Datenhaltungskonzepten und hochperformanten Computern, welche

wiederum ihre eigene Entwicklungsgeschichte haben, ist damit eine komplexe Werkzeugbank entstanden, die den handelnden Akteuren viel abverlangt.

Nicht zuletzt seit den Publikationen von Fayyad ab 1996, der den Begriff „Knowledge Discovery in Databases“ (KDD) prägte, ist allgemein bekannt, dass die Generierung von Wissen aus Daten einen Prozess darstellt. Dieser führt stufenweise von der inhaltlichen Fragestellung über die Datenvorverarbeitung, die eigentliche algorithmische Mustererkennung (Data Mining) und Interpretation (Erklärungsmodell) bis zur Modellimplementierung. Im Rahmen des Modelleinsatzes werden Anpassungsnotwendigkeiten entstehen, die zu einem erneuten Durchlauf des Prozesses führt. Dieser Betrachtung folgen viele Prozessmodelle wie der etablierte Industriestandard CRISP-DM („cross-industry standard process for data mining“) und auch das neu entwickelte DASC-PM („Data Science Process Model“).

Aufgrund der komplexen Aufgabenstellung entlang der Prozessphasen stellen sich vielfältige Anforderungen an die Datenanalysten. Die Bemühung um curriculare Erweiterungen einschlägiger Studiengänge zum Abschluss „Data Scientist“ ist daher allerorten an Hochschulen erkennbar. Man darf sich aber nicht täuschen, denn die algorithmische Befähigung alleine reicht nicht. Die lange Geschichte der Ausbildung zu Business-Intelligence-Experten hat gezeigt, dass ein Gleichklang von Technologie, Business und Organisation gefordert ist, um im Alltagsinsatz der Projekte gewappnet zu sein. Die Führung und die Integration von agilen Teams mit spezifischen Fachkenntnissen sind unverzichtbar, denn ein Einzelkämpfer steht den komplexen Aufgaben machtlos gegenüber.

Nicht nur die Kernaktivitäten der Datenanalyse sind herausfordernd, auch die Ausgestaltung der Digitalisierungsstrategie und die Klärung architektonischen Fragen zur Datenhaltung müssen behandelt werden. Nachfolgende Beiträge im Sammelwerk werden hierzu Antworten liefern.

Die starke Akzentuierung auf KI und Machine Learning (ML) deuten auf eine enge Bindung des Forschungsfeldes an die Kerninformatik hin. Dennoch hat das Thema „Business Analytics (BA)“ aus der Wirtschaftsinformatik dieses Teilgebiet immer eingeschlossen. Dabei kann BA als Sammlung unterschiedlicher Methoden und Technologien verstanden werden, welche dazu dienen, Erkenntnisse aus verfügbaren Daten für unternehmerische Entscheidungen zur Steuerung der Geschäftsprozesse zu gewinnen. BA grenzt sich von Business Intelligence (BI) dadurch ab, dass verstärkt auf die datengetriebene Analyse zur Planung und Prognoserechnung gesetzt wird. Damit steht die Zukunftsorientierung im Vordergrund. Unter dem Begriff Advanced Analytics werden in diesem Zusammenhang gerade die Methoden des maschinellen Lernens und der Statistik erfasst, welche die Ableitung von Vorhersagemodellen mit Kausalzusammenhängen ermöglichen, die deutlich über die Fähigkeiten von explorativen und vergangenheitsorientierten Datenanalysen der Business Intelligence hinausgehen. Also sind Data Science und Advanced Analytics prinzipiell wesensgleich. Der Unterschied entsteht dort, wo die Anwendungsdomäne den ökonomischen Bereich verlässt.

Im Fokus von Data Science steht die Entwicklung von einsetzbaren Entscheidungsmodellen, die wahlweise als interaktive Entscheidungsunterstützung oder als autonome „Entscheidungsmaschine“ genutzt werden können. Im Rahmen der Datenanalyse erhofft man sich als Resultat plausible und interpretierbare Muster, welche als Regelwerke die Entscheidungsmodelle bilden. Zur Aufdeckung der Ursache-Wirkungs-Beziehungen kommen vielfältige Verfahren des Data Mining zum Einsatz. Diese können grob in überwachte und unüberwachte Lernverfahren gegliedert werden. In die erste Gruppe fallen Vorhersagen auf der Basis von klassischen Regressionsverfahren, Klassifikationen mit Entscheidungsbäumen oder Künstlich Neuronale Netzen sowie Zeitreihenanalysen. Den Verfahren ist gemeinsam, dass sie aus Datenbeständen die bekannte Abhängigkeit der zugrunde liegenden Variablen erlernen und als Prognosemodell zur Verfügung stellen. In die zweite Gruppe der unüberwachten Lernverfahren gehören Assoziationsanalysen sowie Clustering. Diesen Algorithmen stehen keine Lernmuster zur Verfügung, stattdessen ermitteln sie eigenständig Hypothesen aus dem Datenmaterial. Entscheidungsmodelle entstehen erst aus der Kombination von derartigen Datenanalysen und der mathematischen Optimierung. Die Verfahren des Operations Research (Simulation, lineare und nichtlineare Optimierung, stochastische Optimierung etc.) können auf den Kausalanalysen aufsetzen und den Erklärungsmodellen eine Zielfunktion hinzufügen. Hierdurch können optimale Handlungsalternativen ermittelt werden, so dass ein Übergang von Deskription über Prädiktion zur Präskription stattfindet.

Insgesamt finden sich vielfältige und verzweigte Wurzeln der Entwicklungsgeschichte von Data Science, die zumindest schlaglichtartig angeklungen sind. Zurzeit zeichnet sich das Bild einer prospektiven Zukunft des Teilbereichs der Informatik ab. Voraussetzung bleibt aber ein positives Zusammenwirken der unterschiedlichen Disziplinen und Akteure, um dem gemeinsamen Ziel der Automatisierung und Digitalisierung näher zu kommen.

Univ.-Prof. Dr. Peter Chamoni

Inhaltsverzeichnis

**Teil I Data Strategist Digitalisierung von Geschäftsmodellen –
Big Data Technologien erfolgreich implementieren**

1	Big Data	3
	Uwe Schmitz	
1.1	Grundlagen	3
1.2	Architektur und Bausteine	6
1.3	Datengetriebene Geschäftsmodelle	15
1.4	Exemplarische Einsatzmöglichkeiten	17
	Literatur	24
2	Data Literacy als ein essenzieller Skill für das 21. Jahrhundert	27
	Andreas Schmidt, Thomas Neifer und Benedikt Haag	
2.1	Notwendigkeit von Data Literacy	28
2.2	Data Literacy als Begriff	30
2.3	Data Literacy Skills im Detail	32
2.4	Konzepte zur Implementation von Data Literacy in Lehre und Praxis	34
2.5	Fazit	38
	Literatur	39
3	Management von Big Data Projekten	41
	Andreas Gadatsch und Dirk Schreiber	
3.1	Konzeptioneller Rahmen des Informationsmanagements	41
3.1.1	Überblick	42
3.1.2	Aufgabenorientiertes Ebenenmodell	42
3.1.3	Integriertes Informationsmanagement	44
3.1.4	Einordnung von Big Data.	45
3.2	Digitalisierung von Geschäftsmodellen mit Big Data	46
3.2.1	IT-Governance und Digitalisierung	46
3.2.2	Von der IT-Strategie zur Business Digitalstrategie	49

3.2.3	Management von Big Data.	53
3.2.4	Vorgehensmodelle zur Einführung von Big Data	54
3.2.5	Messung des Reifegrades von Organisationen.	57
3.2.6	Auswirkungen von Big Data auf die Organisation	59
	Literatur.	60
4	Digital Leadership.	63
	Wilhelm Mülder	
4.1	Führung im Digitalzeitalter	63
4.2	New Work.	64
4.2.1	Mobile Arbeitsplätze	65
4.2.2	Flexible Arbeitszeiten.	65
4.2.3	Veränderte Arbeitsinhalte.	66
4.2.4	Neue Arbeitsorganisation.	67
4.3	New Workforce	68
4.3.1	Beschäftigungseffekte der Digitalisierung.	68
4.3.2	Rekrutierung von Generation Z.	68
4.4	Digital Leader.	71
4.4.1	Persönlichkeitsmerkmale	71
4.4.2	Führungskompetenzen	72
4.4.3	Virtuelle Führung.	72
4.5	Konzepte und Methoden für Digital Leadership	73
4.5.1	SCRUM	73
4.5.2	Design Thinking.	76
4.5.3	Servant Leadership.	77
4.5.4	VOPA+Modell	77
4.6	Fazit	79
	Literatur.	80
 Teil II Data Architect: Informationsarchitekturen gestalten – Daten effizient verwalten		
5	Data Engineering	85
	Christoph Quix	
5.1	Aufgaben des Data Engineering.	86
5.2	Architekturen zum Daten-Management.	87
5.3	Datenmodellierung und Metadaten-Management	91
5.4	Datenaufbereitung und Datenintegration.	93
5.5	Datenbank-Management-Systeme: SQL, NoSQL und Big Data.	99
5.6	Fazit	102
	Literatur.	103

6	Data Governance	105
	Detlev Frick	
6.1	Einführung	105
6.1.1	Begriffliche Einordnung	105
6.1.2	Datenstrategie	107
6.2	Data Governance Framework	109
6.2.1	Strategie	109
6.2.2	Aufbauorganisation	111
6.2.3	Richtlinien, Prozesse und Standards	112
6.2.4	Messen und Beobachten	113
6.2.5	Technologie	114
6.2.6	Kommunikation	116
6.3	Data Quality Management (DQM)	117
6.4	Fazit	118
	Literatur	118
7	Einsatz von In-Memory Technologien	121
	Uwe Schmitz	
7.1	Einleitung	121
7.2	Definition und Abgrenzung In-Memory Technologien	123
7.3	Anforderungen an den Einsatz einer In-Memory-Technologie	127
7.4	Bewertung	129
7.5	Fazit	131
	Literatur	131
8	Big-Data-Technologien	133
	Christoph Quix	
8.1	Einleitung	133
8.2	Skalierbarkeit und Fehlertoleranz	134
8.3	Volume – Management von großen Datenmengen	137
8.4	Velocity – Kontinuierliche Verarbeitung von Datenströmen	142
8.5	Variety – Unterstützung für die Zusammenführung von heterogenen Daten	145
8.6	Fazit	148
	Literatur	148
9	Information Data Models: Das Fundament einer guten Information Strategy	149
	Christian Rupert Maierhofer	
9.1	Drei Thesen aus Sicht eines Praktikers	150
9.2	It's all about the information	152

9.3	Das Heute und seine Hürden	152
9.4	Wie es dazu gekommen ist.	153
9.5	Die Enterprise Architektur	154
9.6	Drei Formen der Informations-Architektur und deren Auswirkungen	155
9.6.1	Das Gestern und leider noch das Heute. Der anwendungs- zentrierte Ansatz (The Application Centric Approach)	155
9.6.2	Das Heute und die Morgendämmerung, der datengesteuerte Ansatz (The Data Driven Approach)	156
9.6.3	Das überfällige Übermorgen, die datenzentrische Architektur (The Data Centric Architecture)	159
	Literatur.	162

Teil III Data Analyst: Auswerten, Präsentieren, Entscheiden – Systematische Datenanalyse im Unternehmen

10	Reporting multidimensionaler Daten und Kennzahlen.	167
	Detlev Frick und Birgit Lankes	
10.1	Betriebswirtschaftliche Motivation	167
10.1.1	Kennzahlen und ihre Anwendung	168
10.1.2	Auswahl von Kennzahlen.	169
10.2	Daten und Business Intelligence	170
10.2.1	Datenmodellierung.	171
10.2.2	Datensicherung.	172
10.2.3	Harmonisierung	173
10.2.4	Daten-/Informationsqualität.	173
10.2.5	Datenbereitstellung	174
10.3	Reporting/Berichtswesen	174
10.3.1	Berichtsgrundformen	176
10.3.2	Anforderungen an Berichte	177
	Literatur.	177
11	Fundamentale Analyse- und Visualisierungstechniken.	179
	Jens Kaufmann	
11.1	Einleitung und Begriffswelt	179
11.2	Lineare Regression.	182
11.2.1	Basisidee und Begrifflichkeiten	182
11.2.2	Beispiel und Ergebnisinterpretation.	183
11.2.3	Prüfen der Voraussetzungen und Variablentransformation	185
11.3	Einfache Klassifikationsverfahren	186
11.3.1	k-Nearest-Neighbors	186
11.3.2	Naive Bayes	187

11.3.3	Entscheidungsbäume	188
11.4	Clustering-Verfahren	189
11.4.1	Hierarchische Verfahren	189
11.4.2	Partitionierende Verfahren	191
11.5	Assoziationsanalyse	191
11.6	Ergänzende Überlegungen, Software und Tools	192
	Literatur	193
12	Fortgeschrittene Verfahren zur Analyse und Datenexploration, Advanced Analytics und Text Mining	195
	Jens Kaufmann	
12.1	Einleitung	195
12.2	Datenexploration und -darstellung	196
12.3	Principal Component Analysis	197
12.4	Random Forests	200
12.5	Logistische Regression	200
12.6	Entscheidungsbewertung	201
12.7	Zeitreihenanalyse	202
12.8	Text Mining	205
12.9	Weitere Analysemöglichkeiten	207
	Literatur	208
13	Datenbasierte Algorithmen zur Unterstützung von Entscheidungen mittels künstlicher neuronaler Netze	209
	Daniel Retkowitz	
13.1	Datenbasierte Algorithmen und maschinelles Lernen	209
13.1.1	Maschinelles Lernen	210
13.1.2	Lernverfahren	211
13.2	Künstliche neuronale Netze	212
13.2.1	Netzarchitekturen	212
13.2.2	Grenzen künstlicher neuronaler Netze	213
13.3	Beispielhafte Anwendungsfelder	214
13.4	Entwicklungsprozess	216
13.5	Entwicklungsplattformen und Werkzeuge	217
13.5.1	TensorFlow und PyTorch	218
13.5.2	Ausführungsmodi	219
13.5.3	Deployment und Betrieb	220
13.6	Fazit und Ausblick	222
	Literatur	223
14	Künstliche Neuronale Netze – Aufbau, Funktion und Nutzen	225
	Anja Tetzner, Tom Kühne, Peter Gluchowski und Melanie Pfoh	
14.1	Einleitung	226

14.2	Aufbau	227
14.2.1	Künstliches Neuron	227
14.2.2	Künstliche neuronale Netze	229
14.3	Lernen künstlicher neuronaler Netze	233
14.3.1	Überwachtes Lernen – Lernen mittels Backpropagation	234
14.3.2	Unüberwachtes Lernen – Lernen mittels Wettbewerbslernen	235
14.4	Nutzenpotenziale und Herausforderungen	236
14.5	Fazit	238
	Literatur	238
15	Bayesian Thinking in Machine Learning	241
	Thomas Neifer, Andreas Schmidt, Dennis Lawo, Lukas Böhm und Özge Tetik	
15.1	Bayesian Thinking	242
15.2	Bayes in Machine Learning	245
15.2.1	Bayes in Regressionsverfahren	245
15.2.2	Bayes in Klassifikationsverfahren	249
15.3	Naive Bayes Classifier	251
15.3.1	Grundlagen	251
15.3.2	Methodik	252
15.4	Fazit	254
	Literatur	254
 Teil IV Anwendungsorientierte Data Science		
16	Text Mining: Durchführung einer Sentiment Analysis mit SAP HANA	259
	Patrick Bachmaier	
16.1	Einleitung	259
16.2	Grundlagen	260
16.3	Umsetzung	261
16.3.1	Vorgehensmodell	261
16.3.2	Implementierung	263
16.4	Fazit	273
	Literatur	274
17	Weiterbildung in Data Science	277
	Christoph Quix	
17.1	Kompetenz-Rahmenwerke für Data Science	278
17.2	Studiengänge zu Data Science	280

17.3	Berufliche Weiterbildung zu Data Science.	283
17.3.1	Zertifikatsprogramm der Fraunhofer Gesellschaft zu Data Science	284
17.3.2	Zertifikatsstudien der Hochschule Niederrhein	285
17.3.3	Zertifikatslehrgang zum Data Scientist der Bitkom Akademie	287
17.4	Fazit	288
	Literatur.	289
18	Plattformökonomie für Data Plattformen	291
	Valeria Knoll und Alexa Scheffler	
18.1	Motivation	291
18.2	Begriffshaushalt	292
18.2.1	Plattformen und Plattformökonomie	292
18.2.2	Data Plattform	294
18.3	Design-Prinzipien für Data Plattformen.	296
18.3.1	Netzwerkeffekte durch gemeinsam genutzte Datenobjekte.	296
18.3.2	Strategien für die Aktivierung von Plattformteilnehmern	297
18.3.3	Einfacher Zugang durch Self-Service	298
18.3.4	Effektives Matching durch Metadaten.	299
18.4	Monetarisierung	299
18.5	Zusammenfassung und Fazit	300
	Literatur.	302
19	Akzeptanz und Nutzung von maschinellem Lernen und Analytics im Rechnungswesen und Controlling	305
	Markus Eßwein, Domenica Martorana, Martina Reinersmann und Peter Chamoni	
19.1	Eine Herausforderung für die Finanzfunktion	306
19.2	Nutzerakzeptanzforschung zu maschinellem Lernen.	307
19.3	Befragung von Führungskräften	308
19.3.1	Strukturgleichungsmodell	308
19.3.2	Umfrage	308
19.4	Aktuelle Nutzung und Treiber	310
19.4.1	Ergebnisse der Befragung	310
19.4.2	Treibermodell zur Nutzung und Akzeptanz	315
19.5	Handlungsempfehlungen und Ausblick	317
	Literatur.	318
20	Durch Daten zu neuen Geschäftsmodellen und Prozessoptimierungen – im Kontext von Car-Sharing	321
	Eva Schoetzau	
20.1	Kurze Einführung.	321

20.2	Durch Daten zu neuen Ideen und Optimierungen	322
20.3	Umdenken im Unternehmen	325
20.4	Durch ständige Überwachung zur stetigen Anpassung	328
20.5	Mit ‚Lessons Learned‘ zur Optimierung von Geschäftsmodellen und -prozessen	331
20.6	Fazit	335
	Literatur.	335
21	Einsatz von Logit- und Probit-Modellen in der Finanzindustrie	337
	Uwe Rudolf Fingerlos und Alexander Pastwa	
21.1	Einleitung	337
21.2	Logit- und Probit-Modelle	338
21.3	Datengrundlage	340
21.4	Modellierung	343
21.5	Überprüfung der Modellannahmen	347
21.6	Vorstellung der Ergebnisse.	348
21.7	Vergleichende Beurteilung.	351
	Literatur.	354
	Stichwortverzeichnis.	357

Herausgeber- und Autorenverzeichnis

Über die Herausgeber



Prof. Dr. Detlev Frick (Jahrgang 1956). Studium der Wirtschaftswissenschaften mit Schwerpunkt Wirtschaftsinformatik bei Prof. Dr. Jörg Biethahn an der Universität Gesamthochschule Duisburg mit Abschluss als Diplom-Ökonom absolviert. Anschließend wissenschaftlicher Mitarbeiter an der Ruhr-Universität Bochum am Lehrstuhl von Prof. Dr. Roland Gabriel und Promotion zum Dr. rer. oec. an der Gerhard-Mercator-Universität Duisburg (Gutachter: Prof. Dr. Roland Gabriel und Prof. Dr. Bernd Rolfes).

Tätigkeit als festangestellter und freiberuflicher SAP-Berater. Ab 1995 Projektleiter in der Softwareentwicklung (Individualsoftware). Beteiligung an Softwareprojekten in der Größenordnung von 10 bis 140 Mitarbeitern. Von 1999 bis 2001 verantwortlich für den Bereich Methoden und Standards der SAP-Systeme im zentralen Informationsmanagement des Konzerns Deutsche Telekom AG. Von 2001 bis 2004 Kompetenzmanager und Projektleiter der T-Systems Nova in der BU Essen und dort verantwortlich für den Themenbereich SAP. Durchführung von zahlreichen SAP-Projekten. Engagement beim Aufbau des Qualitätsmanagementsystems.

Zum SS 2004 Berufung als Professor für Betriebswirtschaftslehre, insb. Wirtschaftsinformatik an die HS Niederrhein.

Die anwendungsbezogene Lehre und Forschung umfasst die Fachgebiete Standardanwendungssoftware (insb. SAP), Projektmanagement, Business Intelligence und Data Science.

Zahlreiche Beratungsprojekte, Vorträge, Seminare, Workshops und Publikationen zu den vorgenannten Fachgebieten.



Prof. Dr. Andreas Gadatsch ist Inhaber der Professur für Betriebswirtschaftslehre, insbesondere Wirtschaftsinformatik, Leiter des Masterstudiengangs Innovations- und Informationsmanagement sowie Wissenschaftlicher Leiter des Data Innovation Labs im Institut für Management der Hochschule Bonn-Rhein-Sieg.

Er ist Gründungsmitglied des Big Data Innovation Centers der Hochschulen Bonn-Rhein-Sieg, Niederrhein und der FH Dortmund. Die aktuellen Projekte beschäftigen sich mit den Auswirkungen von Big Data auf das Informations- und Geschäftsprozessmanagement.

Er ist Autor von über 340 Fachpublikationen zur Wirtschaftsinformatik, davon 28 Bücher und Herausgeberbände, z. T. in mehreren Auflagen.



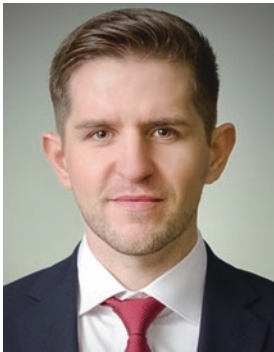
Prof. Dr. Jens Kaufmann ist Inhaber der Professur für Wirtschaftsinformatik, insb. Data Science an der Hochschule Niederrhein. Zuvor war er mehrere Jahre in der Beratung bei Horváth & Partners sowie im Bereich des Global CIO bei der ERGO Group AG in Düsseldorf tätig. Er dozierte als Gastprofessor an der University of North Carolina in Charlotte, NC, USA, und beschäftigt sich in Lehre und Forschung schwerpunktmäßig mit der Anwendung von Data Science und ihrem Transfer in die betriebliche Praxis.



Birgit Lankes ist seit 2013 Lehrkraft für besondere Aufgaben am Fachbereich Wirtschaftswissenschaften der Hochschule Niederrhein. Nach einem Fachhochschulstudium der BWL hat sie zunächst in der IT der Hochschule gearbeitet und hier erste Erfahrungen mit SAP gesammelt. Später hat sie die Lehrenden in der Forschung unterstützt und sich stetig im Bereich SAP weitergebildet. Mit einem internationalen Team hat sie gemeinsam mit Prof. Dr. Frick ein Curriculum zum SAP Solution Manager entwickelt, dass weltweit von Lehrenden eingesetzt wird. Seit 2013 hält verschiedene Veranstaltungen im SAP- und BI-Kontext.



Prof. Dr. Christoph Quix ist seit 2019 Professor für Wirtschaftsinformatik und Data Science im Fachbereich Elektrotechnik und Informatik an der Hochschule Niederrhein. Am Fraunhofer-Institut für Angewandte Informationstechnik FIT in St. Augustin leitet er im Forschungsbereich Life Science Informatics die Abteilung High-Content Analysis. Vorher hatte er eine Vertretungsprofessur für Data Science an der RWTH Aachen inne, an der er auch seine Habilitation (2013) und Promotion (2003) abgeschlossen hat. Seine Forschungsschwerpunkte sind Datenintegration, Management von großen, heterogenen Datenmengen und Metadaten-Management. Er hat mehr als 100 Publikationen in internationalen, wissenschaftlichen Zeitschriften und Konferenzen.



Andreas Schmidt ist wissenschaftlicher Mitarbeiter an der Hochschule Bonn-Rhein-Sieg und dort neben Lehr- und Forschungstätigkeiten im Bereich Data Literacy und dem Themenfeld Future Skills insbesondere mit dem Aufbau des Data Innovation Labs im Institut für Management betraut. Er hat an der Hochschule Bonn-Rhein-Sieg Innovations- und Informationsmanagement studiert und seither verschiedene berufliche Stationen im Bereich der Trend- und Zukunftsforschung durchlaufen – zuletzt dabei bei KPMG als Berater im Bereich Trend Analytics im Zuge der Entwicklung und Anwendung innovativer, datenbasierter Trendanalyse-Methoden zur Beantwortung aktueller gesellschaftlicher, wirtschaftlicher und technologischer Fragestellungen.



Prof. Dr. Uwe Schmitz studierte bis 1998 Betriebswirtschaftslehre an der Hochschule Niederrhein. Direkt im Anschluss hatte er bis zum Jahr 2009 verschiedene Positionen und Funktionen bei der SAP AG inne und war verantwortlich für diverse internationale Großprojekte bei namhaften DAX-Unternehmen. Berufsbegleitend erfolgte seine externe Promotion an der TU Chemnitz im Jahr 2005. Seit 2009 ist Dr. Uwe Schmitz Professor für Wirtschaftsinformatik an der Fachhochschule Dortmund, wo er zurzeit Vorsitzender der Fachgruppe Wirtschaftsinformatik und Leiter der Wirtschaftsinformatikstudiengänge (Bachelor und Master) ist. Er ist auch Vorsitzender der Big Data Innovation Centers der Hochschulen Bonn-Rhein-Sieg, Niedersachsen und der FH Dortmund, sowie Autor diverser wissenschaftlicher Artikel und Buchbeiträge.

Autorenverzeichnis



Patrick Bachmaier Jahrgang 1993, absolvierte das Bachelor- sowie Masterstudium in Wirtschaftsinformatik an der Hochschule Niederrhein in Mönchengladbach. Parallel zum Studium konnte Herr Bachmaier Praxiserfahrung im Bereich Data Warehouse, Business Intelligence und Data Science sammeln. Er arbeitet seit mehr als zwei Jahren als IT-Architekt im Bereich Data Analytics. Parallel zu dieser Tätigkeit besuchte er, neben weiteren Fortbildungen im Bereich Data Analytics, den Zertifikatsstudiengang CAS Data Analyst an der Hochschule Niederrhein, welchen er 2020 erfolgreich abschloss. Herr Bachmaier lebt mit seiner Frau in Moers am Niederrhein.



Lukas Böhm ist Wissenschaftlicher Mitarbeiter und Doktorand an der Universität Siegen. Dort forscht er zu nachhaltiger Mobilität. Weiterhin ist er Wissenschaftlicher Mitarbeiter an der Hochschule Bonn-Rhein-Sieg, wo er im Kompetenzzentrum Usability als KI-Trainer Workshops zu Data Science und Text Mining durchführt. Nach einem Bachelorstudium in Wirtschaftsinformatik an Europäischen Fachhochschule Brühl hat er Information Systems an der Universität zu Köln studiert.



Prof. Dr. Peter Chamoni war seit 1995 Inhaber des Lehrstuhls für Wirtschaftsinformatik, insbesondere Business Intelligence an der Mercator School of Management der Universität Duisburg-Essen. Nach dem Studium der Mathematik und Betriebswirtschaft promovierte er an der Ruhr-Universität Bochum in Operations Research und habilitierte sich dort zum Thema „Entscheidungsunterstützungssysteme und Datenbanken“. Seitdem erschienen von ihm zahlreiche Publikationen zum Thema „Data Warehouse und Business Intelligence“. Auf einschlägigen nationalen und internationalen Tagungen ist er Organisator, Autor und Fachgutachter. Neben der Wissenschaft und der Lehre im Masterstudiengang „Business Analytics“ nimmt die Arbeit in Praxisprojekten für ihn einen hohen Stellenwert ein. Er war

Mitgründer und Vorsitzender des Aufsichtsrats der cundus AG sowie Präsident des TDWI Germany e. V.

Seit dem Wintersemester 2019/2020 ist er im Ruhestand. Als Honorarprofessor an der TU Bergakademie Freiberg in Sachsen nimmt er weiterhin Lehraufgaben in der Wirtschaftsinformatik wahr.



Ulrich Dommer, Jahrgang 1974, Dipl.-Kfm. und MBA, beschäftigt sich seit über 20 Jahren mit SAP-Technologie, Datenarchitekturen, Business Intelligence und Predictive Analytics.

Als Unternehmensberater erarbeitet er gemeinsam mit seinen Kunden effiziente Lösungen für das Management und die Aufbereitung von Daten zu Steuerungszwecken. Herr Dommer begleitet insbesondere Transformationsprogramme im SAP-Umfeld aus Sicht der Unternehmensarchitektur und dem Aufbau von nachhaltigen Informations- und Steuerungssystemen.

Seine beruflichen Stationen umfassten seit seiner Ausbildung bei den Duisburger Grillo-Werken verschiedene Unternehmensberatungen. Ab 2007 führte er die Geschäfte der auf SAP BI-Lösungen spezialisierten CONOGY GmbH, die 2018 in die KPMG AG Wirtschaftsprüfungsgesellschaft integriert wurde. Seitdem verantwortet er Aktivitäten zur digitalen Transformation als Partner im Beratungsbereich der KPMG.

Herr Dommer ist Autor zum Thema Predictive Analytics mit SAP und Beirat im Big Data Innovation Center, über das er mit den Herausgebern dieses Werks verbunden ist.



Dr. Markus Eßwein ist interner Auditor für Finance & Accounting bei Henkel AG & Co. KGaA. Nach seinem Studium des Wirtschaftsingenieurwesens mit technischer Fachrichtung Elektrotechnik an der Technischen Universität Darmstadt arbeitete er als Strategieberater für den Bereich CFO & Enterprise Value bei Accenture Strategy. 2019 promovierte er zur Digitalisierung des Rechnungswesens und Controlling am Lehrstuhl für Wirtschaftsinformatik, insbesondere Business Intelligence an der Mercator School of Management der Universität Duisburg-Essen. Seit 2013 ist er Mitglied und Mitorganisator des Kompetenzzentrums Unternehmenssteuerungssysteme/Arbeitskreises Digital Finance der Schmalenbach-Gesellschaft für Betriebswirtschaft e. V.



Aktuell arbeitet **Dr. Uwe Rudolf Fingerlos** als Risikomanager mit Fokus auf der Performance-Messung und Governance interner Kreditrisikomodelle bei einer spanischen Großbank. Zuvor war er ebenfalls bei einer spanischen Großbank als Teamleiter im Bereich Forschung und Entwicklung für die Kreditrisikomodellierung mit Schwerpunkt auf internen Modellen, IFRS9-Modellen sowie Modellen zur Quantifizierung operationeller Risiken tätig. Überdies verfügt Dr. Fingerlos über Berufserfahrung als Manager und Data Scientist im Geschäftsfeld Risk Advisory (Service Line Regulatory Risk) bei der Deloitte GmbH Wirtschaftsprüfungsgesellschaft (Frankfurt am Main, Deutschland) und der Niederösterreichischen Gebietskrankenkasse (St. Pölten, Österreich) mit Schwerpunkten auf der statistisch-ökonomischen Datenanalyse sowie dem Datenmanagement. Nach seinem Studium promovierte Dr. Fingerlos im Jahr 2014 im Fachbereich Volkswirtschaftslehre an der Wirtschaftsuniversität Wien (Österreich).



Prof. Dr. Peter Gluchowski leitet den Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung und Anwendungssysteme, an der Technischen Universität in Chemnitz und konzentriert sich dort mit seinen Forschungsaktivitäten auf das Themengebiet Business Intelligence & Analytics. Er beschäftigt sich seit mehr als 25 Jahren mit Fragestellungen, die den praktischen Aufbau dispositiver bzw. analytischer Systeme zur Entscheidungsunterstützung betreffen. Seine Erfahrungen aus unterschiedlichsten Praxisprojekten sind in zahlreichen Veröffentlichungen zu diesem Themenkreis dokumentiert.



Benedikt Haag studiert Betriebswirtschaftslehre an der Hochschule Bonn-Rhein-Sieg. Er arbeitet als Studentische Hilfskraft im Bereich Wirtschaftsinformatik. Dort unterstützt er beim Aufbau eines Data Innovation Labs sowie in der Erstellung eines Data Literacy Curriculums.



Prof. Ulrich Kelber ist seit dem 7. Januar 2019 der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit. Er ist verheiratet und hat fünf Kinder. Der Dipl.-Informatiker arbeitete nach dem Studium zunächst am Forschungszentrum Informationstechnik, danach als Wissensmanagement-Berater in einem Software-Unternehmen.

Im September 2000 rückte er in den Bundestag nach und vertrat als direkt gewählter Abgeordneter seine Heimatstadt Bonn bis zum Januar 2019.

Von 2005 bis 2013 war er stellvertretender Vorsitzender der SPD-Bundestagsfraktion und koordinierte die Politikbereiche Verbraucherschutz, Ernährung, Landwirtschaft, Umwelt, Naturschutz, Reaktorsicherheit sowie Nachhaltigkeit. Vom Dezember 2013 bis April 2018 war er Parlamentarischer Staatssekretär im Bundesministerium der Justiz und für Verbraucherschutz mit dem Schwerpunkt Verbraucher- und Datenschutz.



Valeria Knoll ist seit 2019 als Data Consultant bei der AXA Konzern AG in Köln tätig. Sie unterstützt die Bereiche Vertrieb und Finance bei strategischen und operativen Fragestellungen rund um die Daten. Davor war sie bei der Allianz Technology AG als Projektmanagerin tätig, nach fast 4 Jahren als Beraterin für Finance bei BearingPoint. Valeria erwarb ihr Bachelor an der Kiev-Mohyla Akademie in der Ukraine und Master an der Otto-von-Guericke Universität Magdeburg, beides im Bereich Finance. Zusammen mit ihrem Mann erzieht sie eine Tochter.



Tom Kühne ist als wissenschaftlicher Mitarbeiter und Nachwuchsforscher an der Professur für Wirtschaftsinformatik, insb. Systementwicklung und Anwendungssysteme, der Technischen Universität Chemnitz tätig. Neben den Themenbereichen Informationssicherheit und Datenbanken liegen seine Forschungsschwerpunkte in der Anwendung und Nutzung von Verfahren des maschinellen Lernens und der Künstlichen Intelligenz. Insbesondere Künstliche Neuronale Netzwerke stehen dabei im Fokus seiner Forschung.



Dennis Lawo ist Wissenschaftlicher Mitarbeiter und Doktorand an der Universität Siegen. Dort forscht er zu nachhaltigem Lebensmittelkonsum. Weiterhin ist er Wissenschaftlicher Mitarbeiter an der Hochschule Bonn-Rhein-Sieg, wo er im Kompetenzzentrum Usability als KI-Trainer Workshops zu Data Science und Text Mining durchführt. Er hat Information Systems an der Universität zu Köln studiert.



Christian Rupert Maierhofer ist 46 Jahre alt und hat einen Abschluss als Industriekaufmann (IHK) sowie als Betriebswirt (VWA). Er ist seit 2017 General Director A/V Software Solutions 360° bei der Bechtle GmbH & Co. KG, IT-Systemhaus Bonn/Köln, nachdem er zuvor dort andere Managementpositionen innehatte. Davor leitete er mehrere Jahre sein eigenes Unternehmen (CRM Design).

Die Gründung der Abteilung A/V Software Solutions 360° hatte für ihn eine „Matrix“-ähnliche Erfahrung. Die blaue oder die rote Pille? Er hat sich damals für die rote Pille entschieden und sagt heute, dass er es zwar nicht bereut hat, aber die IT hinter den Kulissen schon als „etwas sinnfremd“ bewertet.

Seitdem hat er sich mit Enterprise Architekturen sowie Sicherheits- und Skalierungsmodellen beschäftigt und musste feststellen, dass der „wahre Wert“ erstens in der Softwareentwicklung liegt und zweitens, dass das schlagfertigste Konstrukt innerhalb der Informationsverarbeitung die „Community“ ist. Er hat gelernt, dass die IT ein Spiegel der Gesellschaft ist und die Hürden der Digitalisierung wir Menschen selbst sind. „It’s all about the information“ sagte Sir Ben Kingsley im Film „Senakers“ – Die lautlosen im Jahre 1992 – besser und treffender könnte man es 2020 auch nicht formulieren.



Dr. Domenica Martorana ist Scientist in der Produktentwicklung bei QIAGEN GmbH. Nach ihrem Studium der molekularen und zellulären Biologie an der Philipps-Universität Marburg promovierte sie zur Stressantwort in Pilz- und Humanzellen im Fachbereich Mikrobiologie und Genetik an der Georg-August Universität Göttingen sowie am Cancer Research Institute des Oslo University Hospitals in Norwegen. Während ihres Studiums erlangte sie in internationalen Wettbewerben in zwei aufeinanderfolgenden Jahren eine Goldmedaille. Im Rahmen ihrer Dissertation beschäftigte sie sich intensiv mit der automatisierten Auswertung großer Datensätze sowie der statistischen Analyse von Daten aus kleinen Stichproben.



Prof. Dr. Thomas Meuser hat seit 1998 eine Stiftungsprofessur für Informatik an der Hochschule Niederrhein in Krefeld inne. Zurzeit ist er dort Dekan des Fachbereichs Elektrotechnik und Informatik.

Prof. Dr. Meuser studierte und promovierte an der RWTH Aachen anschließend folgte der Wechsel zu den Philips Forschungslaboratorien in Aachen. Seine Lehr- und Forschungsinteressen liegen im Bereich der Kommunikationssysteme und Cyber-Sicherheit. Weitere Aktivitäten beziehen sich auf die digitale Lehre. So leitet er seit 2005 das Cisco Networking Academy Program in Deutschland und ist darüber hinaus Mitglied verschiedener internationaler Design- und Forschungsgremien in diesem weltweiten Lehrprogramm. Von 2017 bis 2019 führte er an der Hochschule Niederrhein als akademischer Leiter das Projekt ‚Weiterbildung für die digitale Wirtschaft‘.

Seine praktischen Erfahrungen in der Wirtschaft basieren auf einer mehrjährigen Beratertätigkeit, der Tätigkeit als Fachleiter für Informations-Sicherheitsmanagementsysteme und der langjährigen Beschäftigung als CIO eines mittelständischen Elektronikunternehmens. Seit 2019 ist Prof. Meuser Mitglied des Fraunhofer-Instituts FKIE in Bonn und aktuell arbeitet er im Gründungsteam des Cyber Management Campus MG der Hochschule Niederrhein mit.



Prof. Dr. Wilhelm Mülder studierte Wirtschaftswissenschaften an der Universität Essen und promovierte im Bereich Wirtschaftsinformatik zum Thema „Implementierung von computergestützten Personalinformationssystemen“.

Nach Tätigkeit als Software-Entwickler und Berater bei zwei Software-Unternehmen ist er als Professor für Wirtschaftsinformatik an der Hochschule Niederrhein, Fachbereich Wirtschaftswissenschaften in Mönchengladbach tätig. Die wichtigsten Lehr- und Forschungsschwerpunkte sind E-Business, M-Business, Digitalisierung der Wirtschaft und Internet der Dinge. Zu diesen Themen verfasste er mehrere Fachbücher und zahlreiche Fachartikel.

Er leitet das Forschungsinstitut GEMIT (Geschäftsprozessmanagement und IT). Ferner ist er Sprecher der Fachgruppe „Informationssysteme in der Personalwirtschaft“ innerhalb der Gesellschaft für Informatik e. V., Bonn, sowie Mitherausgeber der Fachzeitschrift „HR-Performance“ im Datakontext-Fachverlag, Frechen.



Thomas Neifer ist Wissenschaftlicher Mitarbeiter und Doktorand an der Universität Siegen. Er hat Innovations- und Informationsmanagement an der Hochschule Bonn-Rhein-Sieg studiert und promoviert im Bereich der Verbraucherinformatik. Er forscht zu Empfehlungs- und Reputationsmechanismen im Kontext von nachhaltigem Lebensmittelkonsum und Mobilität. Darüber hinaus wirkt er als Wissenschaftlicher Mitarbeiter an der Hochschule Bonn-Rhein-Sieg bei der Entwicklung einer offenen Datenintegrationsplattform mit. Als Dozent lehrt er dort Data Analytics, Statistik und Volkswirtschaftslehre.



Dr. Alexander Pastwa ist seit 2015 Senior Manager im Bereich Financial Industry Risk & Regulatory bei der Deloitte GmbH Wirtschaftsprüfungsgesellschaft. Er unterstützt nationale und internationale Kunden bei der fachlichen Konzipierung, technischen Umsetzung und Einführung von Standard- und Ad-hoc-Reporting-Anwendungen zur Erfüllung regulatorischer Anforderungen (z. B. im Kontext der MaRisk, BCBS #239); vorwiegend im Finanzsektor. In seinem Bereich und in den Projekten verantwortet Dr. Pastwa die Themenfelder Business Intelligence, Datenqualitätsmanagement und Data Governance. Zuvor war Dr. Pastwa als Senior Manager und stellvertretender Bereichsleiter bei der

SKS Unternehmensberatung in den Themenbereichen Risikomanagement und BI-basiertes Reporting tätig. Nach seinem Studium der Wirtschaftswissenschaft promovierte er im Jahr 2009 am Lehrstuhl für Wirtschaftsinformatik der Ruhr-Universität Bochum.



Dr. Melanie Pfoh arbeitet als wissenschaftliche Mitarbeiterin an der Professur Wirtschaftsinformatik, insb. Systementwicklung und Anwendungssysteme, und hat bereits mehrjährige Erfahrung im Bereich Data Mining sowie in der Konzeption und der Implementierung analytischer Informationssysteme. Ihren forschungsbezogenen Schwerpunkt setzte sie im Verlauf verschiedener Forschungsprojekte und ihrer Promotion im Bereich der Entscheidungsunterstützung im Kontext privater und öffentlicher Energiesysteme, welche in verschiedenen Veröffentlichungen dokumentiert sind.



Dr. Martina Reinersmann ist Lehrbeauftragte im Bereich Data Science an der FOM Hochschule für Oekonomie & Management und an der Hochschule Niederrhein in Mönchengladbach. Nach ihrer Promotion am Lehrstuhl für Wirtschaftsinformatik der Ruhr-Universität Bochum und der Universität Duisburg-Gesamthochschule arbeitete sie als freiberufliche Dozentin in berufsbegleitenden Weiterbildungsmaßnahmen der Industrie- und Handelskammer NRW (IHK) und der VWA – Verwaltungs- und Wirtschaftsakademie. Von 1997 bis 2007 hatte sie die Leitung der Geschäftseinheit „Analytical Applications“ bei der Alldata Systems GmbH und ab 2002 bei der ScaleOn GmbH & Co. KG (später: Bayer Business Solutions GmbH) inne. In den darauffolgenden Jahren war sie sowohl als selbstständige SAP BI Senior Consultant als auch als wissenschaftliche Mitarbeiterin an der Fakultät für Betriebswirtschaftslehre am Lehrstuhl für Wirtschaftsinformatik, insbesondere Business Intelligence, der Universität Duisburg-Essen tätig.



Prof. Dr. rer. nat. Daniel Retkowitz ist seit 2017 Professor für Wirtschaftsinformatik, insbesondere Software Engineering am Fachbereich Wirtschaftswissenschaften der Hochschule Niederrhein. Er studierte Informatik an der RWTH Aachen und der Chalmers University of Technology in Göteborg, Schweden. In seiner Dissertation beschäftigte er sich mit der Softwareentwicklung für Smart Homes. Anschließend war er mehrere Jahre bei einem IT-Dienstleister in der Versicherungsbranche als Projektleiter in der Systementwicklung sowie als IT-Architekt und Senior Software Developer tätig und dozierte als Lehrbeauftragter an der FH Aachen. In Forschung und Lehre beschäftigt er sich schwerpunktmäßig mit den Bereichen Software Engineering und Machine Learning.



Dr. Alexa Scheffler verantwortet seit 2020 den spartenübergreifenden Bereich Customer Analytics & Insights bei der AXA Konzern AG in Köln. Davor war sie bei AXA als Leiterin des Data Management Offices tätig, wo sie u. a. die Datenstrategie für den AXA Konzern entwickelte und am Aufbau einer Data Plattform mitwirkte. Vor ihrer Tätigkeit bei AXA war sie bei Capgemini als Beraterin für Enterprise Architektur. Ihr Ausbildungshintergrund ist ein Studium der Informatik an der Universität Augsburg und eine akademische Promotion in Wirtschaftsinformatik, bei der sie sich mit der ökonomischen Bewertung von In-Memory Datenbanken beschäftigte.



Eva Schoetzau hat einen Abschluss als B.A. in Betriebswirtschaft von der HS Niederrhein. Zunächst Ausbildung zur Fremdsprachenkorrespondentin in Englisch. Nach dreijähriger Berufstätigkeit u. a. bei einer internationaltätigen Wirtschaftsprüfungsgesellschaft folgte das Studium der Betriebswirtschaft. Praktische Erfahrungen während des Studiums erfolgten unter anderem in den Bereichen des Produktmanagements, Public Relations und des Online Marketings in der FMCG- und Medienbranche. Nach dem Studium mit dem Schwerpunkt Marketing mehrjährige, fundierte Erfahrungen in den Bereichen Kommunikation, (digitalem) Projektmanagement und der Beratung gesammelt. Weiterbildungen im Bereich des agilen Projektmanagements u. a. als Product Owner und Scrum Master konnten bereits

erfolgreich in der Praxis angewendet werden. Derzeit berufsbegleitendes Masterstudium des Business Development Managements an der Europäischen Fernhochschule Hamburg sowie eine Weiterbildung zur psychologischen Beraterin und Coach.



Prof. Dr. Dirk Schreiber hat seit 2000 eine Professur für Betriebswirtschaftslehre, insb. Informationsmanagement am Fachbereichs Wirtschaftswissenschaften der Hochschule Bonn-Rhein-Sieg inne. Im Rahmen der akademischen Selbstverwaltung ist er seit mehr als 10 Jahren in der Fachbereichsleitung engagiert. Prof. Dr. Schreiber studierte und promovierte an der Universität Siegen. Danach wechselte er zum Sparkasseninformatik-Zentrum in Bonn. Seine praktischen Erfahrungen in der Wirtschaft basieren darüber hinaus auf einem mehrjährigen Engagement als CIO eines mittelständischen Unternehmens der metallverarbeitenden Industrie.

Sein zentrales Lehr- und Forschungsfeld ist die Internet-Ökonomie, zu der er bereits 2010 ein Lehrbuch veröffentlicht hat, das mittlerweile in mehreren Auflagen erschienen ist.

Seit 2016 ist Prof. Schreiber Gründungsdirektor des Instituts für Management der Hochschule Bonn-Rhein-Sieg. Aktuell arbeitet er an der Gründung des Instituts für Verbraucherinformatik an der Hochschule Bonn-Rhein-Sieg mit.



Özge Tetik studiert Betriebswirtschaftslehre an der Hochschule Bonn-Rhein-Sieg und ist als Studentische Hilfskraft und Tutorin in der Wirtschaftsinformatik tätig. Darüber hinaus unterstützt Sie als Studentische Hilfskraft das Mittelstand 4.0-Kompetenzzentrum Usability.



Dr. Anja Tetzner ist als wissenschaftliche Mitarbeiterin an der Professur Wirtschaftsinformatik, insb. Systementwicklung und Anwendungssysteme, an der Technischen Universität Chemnitz tätig. Der Fokus ihrer Tätigkeit liegt auf den Themenbereichen künstliche neuronale Netze und Business Intelligence, insbesondere dem Teilgebiet des analytischen Kundenbeziehungsmanagements. Im Rahmen ihrer Promotion legte sie den Schwerpunkt ihrer Forschungsarbeit auf den Einsatz künstlicher neuronaler Netze zur Identifikation von Ironie, welche die strukturierte Untersuchung der Eignung verschiedener künstlicher neuronaler Netze zur automatisierten Erkennung von Ironie in vordergründig informeller textueller Kommunikation forcierte.



Waldemar Zgrzebski ist seit 2005 Geschäftsführer des Bechtle IT-Systemhauses Bonn.

Bechtle ist ein zukunftsstarker IT-Infrastrukturdienstleister und Digitalisierungspartner für Industrie, Mittelstand und Verwaltung. Der Standort Bonn gehört zur börsennotierten Bechtle AG mit Hauptsitz in Neckarsulm. Mit 75 IT-Systemhäusern in der DACH-Region und IT-Handelsgesellschaften in 14 Ländern Europas ist Bechtle das größte IT-Systemhaus Deutschlands, starker Partner für zukunftsfähige IT-Architekturen und europaweit führend im IT-E-Commerce.

Waldemar Zgrzebski ist verheiratet, hat vier Kinder und lebt in Bonn.

Abbildungsverzeichnis

Abb. 1.1	Klassifizierung von Big Data-Technologien	7
Abb. 1.2	Allgemeine Kosten für das erste Jahr pro TB im Vergleich Entnommen aus: Bitkom, Big-Data-Technologien – Wissen für Entscheider, S. 36	8
Abb. 1.3	Data Mining – Instrumente	10
Abb. 1.4	Eine Donut-Cloud für „Big Data“	11
Abb. 1.5	Ein Flare-Chart für Professor-Student-Beziehungen	11
Abb. 1.6	Ein Beispiel für Dashboards auf Smartphones und mobilen Endgeräten	12
Abb. 1.7	Social Media Prisma	19
Abb. 1.8	Anwendungsziele bei Nutzung von Big Data für Marketing und Vertrieb	20
Abb. 1.9	Anwendungsziele bei Nutzung von Big Data für Distribution und Logistik	21
Abb. 1.10	Anwendungsziele bei Nutzung von Big Data für Finanz- und Risikocontrolling	23
Abb. 1.11	Anwendungsziele bei Nutzung von Big Data für Produktion, Service und Support	24
Abb. 2.1	Häufigkeit der Erwähnung von „Data Literacy“ in der Forschungsdatenbank ScienceDirect im Zeitverlauf	31
Abb. 2.2	Data Literacy als Schnittstellenbegriff	32
Abb. 2.3	Data Literacy Kompetenzrahmen	35
Abb. 2.4	Framework zur Integration von Data Literacy Initiativen in Unternehmen	38
Abb. 3.1	Aufgabenorientiertes Ebenenmodell nach Krcmar (2015)	43
Abb. 3.2	Modell des Integrierten Informationsmanagements nach Zarnekov 2005	44

Abb. 3.3	Ebenen des Integrierten Informationsmanagements nach Zarnekov 2005	46
Abb. 3.4	Auswirkungen der Digitalisierung auf die menschliche Arbeit nach Kornwachs (2018).	47
Abb. 3.5	Reifegradmodell zur Digitalisierung nach Krafft (2018)	48
Abb. 3.6	Rollen im Informationsmanagement im Wandel	50
Abb. 3.7	Klassisches versus Agiles Informationsmanagement.	51
Abb. 3.8	Ableitung der IT-Strategie aus der Unternehmensstrategie	51
Abb. 3.9	Entwicklung einer Digitalstrategie	52
Abb. 3.10	Bitkom-Portfolio für Big Data (2013)	54
Abb. 3.11	Data Science Pipeline.	55
Abb. 3.12	CSC Vorgehensmodell für Big Data.	56
Abb. 3.13	BITKOM Vorgehensmodell	57
Abb. 3.14	DASC-PM v1.0 – Ein Vorgehensmodell für Data-Science-Projekte.	58
Abb. 3.15	Vorgehensmodell für Big Data – Modell Austria	59
Abb. 4.1	Scrum-Methode	74
Abb. 4.2	Design Thinking.	76
Abb. 4.3	Servant Leadership.	78
Abb. 4.4	Digital Leadership Tools im VOPA + Modell	78
Abb. 5.1	Schematische Darstellungen einer Data-Warehouse-(links) und einer Data-Lake-Architektur (rechts).	89
Abb. 5.2	Schritte während der Datenaufbereitung und -integration	94
Abb. 6.1	Data Governance als Teil der Data Economy.	108
Abb. 6.2	Data Governance-Spannungsfeld	108
Abb. 6.3	Treiber für Data Governance	109
Abb. 6.4	Data Governance Framework.	110
Abb. 6.5	Einordnung DGO und Data Governance	111
Abb. 6.6	Prozess zur Nutzung eines Data Catalogs	115
Abb. 6.7	Fachliche Taxonomie	116
Abb. 6.8	Data Asset Catalog	116
Abb. 6.9	Data-Profiling-Analyse als iterativer Prozess.	118
Abb. 7.1	Technische Architektur der RDBMS	124
Abb. 7.2	Beispiele für Logs und Savepoints	125
Abb. 7.3	Spalten- und Zeilenorientierte Datenbanken	126
Abb. 8.1	Beispiel für Sharding und Replikation mit Replikationsfaktor 3.	136
Abb. 8.2	Namenode und Datanodes in HDFS und Ablauf einer Leseoperation	138
Abb. 8.3	Schematischer Ablauf eines Map-Reduce-Jobs	140
Abb. 8.4	Architektur eines Kafka-Systems.	144
Abb. 9.1	Mindset.	151
Abb. 9.2	Application Centric	156