

---

## Vorwort zur 2. Auflage

Data Science findet in den unterschiedlichsten Wirtschaftsbereichen eine immer weitere Verbreitung. Verstärkt setzen Unternehmen heute auf die Nutzung von Data Science, um ihre Wettbewerbsfähigkeit zu verbessern. Neue Werkzeuge erlauben es auch Anwendern aus Fachabteilungen, die keine Data-Science-Experten sind, erste datengetriebene Analysen durchzuführen und *Proof of Concepts* zu entwickeln. Dieser Trend wird auch dadurch verstärkt, dass Data-Science-Werkzeuge zunehmend in der Cloud betrieben werden und daher weniger hohe IT-Hürden darstellen. Seit der Veröffentlichung der ersten Auflage im Mai 2019 hat sich aber nicht nur die Verbreitung von Data Science insgesamt geändert. Auch der Grad der organisatorischen Einbindung von Data Science hat sich weiterentwickelt. Data Science ist keine Spielwiese mehr, auf der Data-Science-Teams, abgekoppelt von der sonstigen Organisation, große Datenbestände analysieren, um zu neuen Erkenntnissen über Kunden, Produkte, Wartung, Preisgestaltung etc. zu gelangen, die dann wiederum zum Teil recht mühsam, wenn überhaupt, in den Wertschöpfungsprozess des Unternehmens einfließen. Data Science ist heute schon oft ein direkter Bestandteil der Wertschöpfungskette. Die hier gemeinsam mit den Fachabteilungen erzielten Erkenntnisse fließen direkt in das Produkt, das Produktportfolio, die Produktion ein und macht *Industrialized Data Science* vielfach bereits zur Realität. Die entdeckten Muster, Möglichkeiten und potenziellen Produkte finden nun schneller Eingang in die Wertschöpfung und stärken damit die Position des Unternehmens.

Verstärkt lassen sich nun darüber hinaus nicht nur Fort- und Weiterbildungsangebote finden. In den vergangenen 18 Monaten sind auch an Hochschulen spezialisierte Studiengänge (Bachelor und Master) entwickelt worden, um die Nachfrage nach qualifiziertem Personal zu befriedigen. Ein Ende dieser Entwicklung, in der Data Science zunehmend an Bedeutung gewinnt, ist nach wie vor nicht absehbar. Aus diesem Grund freut es uns, dass wir in der vorliegenden zweiten Auflage des Buches mit erweiterten und neuen Grundlagenkapiteln sowie Fallstudien dies weiter ausbauen konnten.

Das Kapitel »Feature Selection« diskutiert diesen wichtigen Aspekt im Data-Science-Prozess und ergänzt damit den bestehenden Grundlagenteil des Buches.

Vor dem Hintergrund immer weiter zunehmender Datenquellen und -mengen ist Feature Selection mittlerweile zu einer Stellschraube geworden, will man die Güte und die Laufzeiten der Modelle und damit nicht zuletzt die damit verbundenen Kosten im Blick behalten.

Das Kapitel »Deep Learning« wurde um den Aspekt »Deep Reinforcement Learning« erweitert. Hier lässt sich sehen, wie sich methodische Ansätze in der Data Science weiterentwickeln, um den Anforderungen der Praxis gerecht zu werden. Dieser neu integrierte Part zeigt die Verbindung von Deep Learning mit einer Reinforcement-Learning-Strategie an einem konkreten Anwendungsfall.

Die Kapitel »Von einer BI-Landschaft zum Data & Analytics-Ökosystem« und »Self-Service und Governance im Data-Science-Umfeld: der emanzipierte Anwender« wurden für die neue Auflage überarbeitet und aktualisiert. Unter anderem wurde dabei dem Bereich Data Governance mehr Gewicht verliehen. Darüber hinaus lassen sich gerade bei den Self-Service-Angeboten neue Trends und Tendenzen feststellen, die hier vorgestellt werden. Ebenso wurde das Kapitel zu Analytics-Ökosystemen um weitere Aspekte der Industrialisierung und Erfahrungen aus der Praxis beim Aufbau der zugrunde liegenden Architekturen ergänzt.

Eine weitere Fallstudie baut unseren Praxisteil aus. Im Kapitel »Künstliche Intelligenz bei der Zurich Versicherung« wird die Nutzung von KI und Data Science in der Versicherungsbranche beispielhaft präsentiert. Im Vordergrund steht hier, unterschiedliche KI-Anwendungsfälle vorzustellen und aufzuzeigen, dass auch Versicherungen künstliche Intelligenz sehr erfolgreich im operativen Geschäft einsetzen.

Wir möchten uns an dieser Stelle auch beim dpunkt.verlag (hier insbesondere bei Christa Preisendanz) und beim TDWI für das uns entgegengebrachte Vertrauen und natürlich für die Unterstützung bedanken! Ohne diesen Einsatz wäre auch die zweite Auflage nicht möglich gewesen.

Auch bei Ihnen, liebe Leserinnen und Leser, möchten wir uns bedanken. Einerseits natürlich dafür, dass Sie unser Buch aus dem mittlerweile sehr großen Portfolio der Fachliteratur zum Thema Data Science ausgewählt haben. Andererseits aber auch für die Rückmeldungen, die wir erhalten haben. Diese haben uns auf unserem Weg bestärkt, den wir mit diesem Buch eingeschlagen haben. Konstruktive Hinweise und Vorschläge haben wir versucht, so weit wie möglich in der neuen Auflage zu berücksichtigen.

Wir wünschen Ihnen viel Spaß bei der Lektüre dieser zweiten Auflage und hoffen, dass Sie das Buch gut auf Ihrer Reise in und durch die Welt der Data Science begleitet. »Nichts entwickelt die Intelligenz wie das Reisen«, bemerkte einst Émile Zola. Was gibt es Schöneres, als Interessantes und Nützliches zu verbinden!

*Uwe Haneke, Stephan Trahasch, Michael Zimmer, Carsten Felden*  
Karlsruhe, Offenburg, Köln, Freiberg im Dezember 2020

---

## Vorwort

Data Science, Machine Learning und auch künstliche Intelligenz sind in aller Munde und sorgen gerade im Rahmen der Digitalisierung für viel Gesprächsstoff. Von den angesprochenen Begriffen sind nicht nur die Produktionsprozesse, die bearbeiteten Geschäftsfelder, das Produktportfolio und die damit verbundene Wertschöpfung der Unternehmen betroffen. Auch für die bereits bestehende analytische Landschaft innerhalb eines Unternehmens ergeben sich neue Chancen, die sowohl in der zunehmenden Datenbereitstellung und Datennutzung als auch in den Möglichkeiten der neuen Technologien selbst zu verorten sind. Gerade die oftmals über viele Jahre hinweg aufgebauten analyseorientierten Systeme lassen sich durch Data Science und Big Data ergänzen und qualitativ verbessern, was dem wachsenden Analyseverständnis in Unternehmen Rechnung trägt. Dazu hat sich neben der Business Intelligence (BI) und den typischen BI-Analysen mit der Data Science eine weitere Analysewelt entwickelt, die sich mit intensiven Datenauswertungen und insbesondere auch mit prädiktiven Analysen auseinandersetzt. Während in der BI überwiegend historische Daten und teilweise auch Echtzeitdaten ausgewertet und entsprechende Kennzahlen ermittelt werden, ermöglicht Data Science mit prädiktiven Methoden unter anderem auch eine Vorhersage von Kennzahlen. Unternehmen, die Business Intelligence und Data Science verwenden, werden somit in die Lage versetzt, umfangreiche Analysen durchzuführen, die einen Blick in die Vergangenheit, auf den aktuellen Zustand und mit Data Science auch in die »nahe« Zukunft ermöglichen. Bei BI und Data Science handelt es sich nicht um unabhängige Systemwelten – die Schnittstelle zwischen Data Science und Business Intelligence ist durchaus bidirektional zu sehen. Ebenso wie die konsolidierte Datenbasis im etablierten Data Warehouse neben Roh- und weiteren Detaildaten eine wichtige Quelle für tieferegehende Analysen ist, so können auch die Ergebnisse der Data Science wieder zurück in die BI fließen.

Die aus der Zunahme des datengetriebenen Handelns in Unternehmen kombinierte Nutzung von Business Intelligence auf der einen und Data Science auf der anderen Seite bietet Potenziale, die den Unternehmen sowohl neue Entwicklungschancen, effizientere Entscheidungs- und Steuerungssysteme als auch die Verbesserung ihrer Wettbewerbsfähigkeit ermöglichen. Für Unternehmen, die heute bereits

über ein entwickeltes und etabliertes analytisches System, wie eben BI, verfügen, verbindet sich dies mit unterschiedlichen Fragestellungen: Wie sind Data Science und ihre Möglichkeiten fruchtbringend einzusetzen? Was gibt es zu beachten? Welche Konsequenzen hätte ein solcher Schritt? Ist die Nutzung von Data Science überhaupt sinnvoll für das Unternehmen?

Das vorliegende Buch richtet sich in erster Linie an Leserinnen und Leser aus Studium und Praxis, die bereits Erfahrung mit analytischen Systemen haben. Business-Intelligence-Manager gehören ebenso zu der Zielgruppe des Buches wie auch Daten- und Informationsverantwortliche im Unternehmen sowie Projektleiter aus dem BI- und Analytics-Bereich. Das Buch möchte die Anwenderinnen und Anwender in ihrem bestehenden Umfeld abholen, ihnen Schritt für Schritt die Welt der Data Science nahebringen und Möglichkeiten für die Nutzung von Data Science in ihrem Unternehmen aufzeigen. Dies soll über die Einteilung in einen Grundlagen- und einen praxisnahen Anwendungsteil geleistet werden. Dabei werden zunächst die verschiedenen Facetten der Data Science von der Herkunft über die Möglichkeiten der Nutzung und die dazu notwendigen Architekturen bis hin zu ethischen Aspekten und der Frage der Data Privacy diskutiert. Im Anschluss arbeiten Fallstudien aus der Praxis verschiedene Aspekte heraus, die es bei der Implementierung und beim Einsatz von Data Science im Zusammenspiel mit bestehenden BI-Systemen im Unternehmen zu beachten gilt. Die einzelnen Kapitel können dabei auch in einer anderen als der hier vorliegenden Reihenfolge oder auch selektiv gelesen werden.

Wir als Herausgeber sind sehr froh, dass uns der TDWI und der dpunkt.verlag bei diesem Projekt unterstützt und immer wieder auf unserem Weg bestärkt haben. In erster Linie gilt unser Dank natürlich den beteiligten Autorinnen und Autoren, ohne die uns die Umsetzung des Buches nicht gelungen wäre. Wir freuen uns, dass wir bemerkenswerte Spezialisten für die unterschiedlichen Aspekte von Data Science und ihrer Nutzung im Unternehmen gewinnen konnten, was zu einem ganzheitlichen und runden Bild geführt hat, in dem die Praxis nicht zu kurz kommt.

Beim dpunkt.verlag möchten wir uns speziell bei Christa Preisendanz bedanken, die uns mit ihrer Erfahrung stets auf die Fertigstellung des Buches fokussiert gehalten hat.

Wir hoffen, dass Ihnen, liebe Leserinnen und Leser, die Lektüre des vorliegenden Buches gefällt und Sie viel Nutzen daraus ziehen können, wenn Sie sich für den Einsatz von Data Science interessieren. Da sich der Prozess der Verknüpfung von Data Science und der bestehenden analytischen Landschaft erst am Anfang befindet, wird es spannend werden, zu beobachten, wie schnell und in welchem Umfang die Entwicklung in den Unternehmen voranschreiten wird und welche Transformationen sich daraus ergeben werden.

*Uwe Haneke, Stephan Trahasch, Michael Zimmer, Carsten Felden*  
Karlsruhe, Offenburg, Stuttgart, Freiberg im Januar 2019