

# 1 Einleitung

*Uwe Haneke · Stephan Trahasch · Michael Zimmer · Carsten Felden*

## 1.1 Von Business Intelligence zu Data Science

Seit dem Jahr 2015 hat sich die Welt der Business Intelligence (BI) schnell und signifikant verändert. Big Data und die damit zusammenhängenden Entwicklungen im Bereich der Data Science haben auch die Business Intelligence nicht unberührt gelassen. Und so sehen wir aktuell eine Erweiterung der bisherigen BI-Systeme und Architekturen, die die betrieblichen Informationssysteme agiler, schneller, mächtiger und passgenauer machen. Die neue BI-Welt enthält heute eine integrierte analytische Komponente, die weit über das hinausgeht, was man bis vor Kurzem noch kannte.

Dabei ist es nicht so, dass Analytics etwas grundlegend Neues in der Business Intelligence wäre. Allerdings vermochte es Data Science mit ihrem Hintergrund auf der wissenschaftlich, technischen Ebene, einen Innovationsschub auszulösen, dessen Ende noch nicht absehbar zu sein scheint. Die nachfolgenden Ausarbeitungen stellen daher zunächst dar, wie die bisherige BI-Entwicklung beginnend in den 1960er-Jahren bis heute verlief. Dabei wird ein besonderes Augenmerk auf die Business Analytics gelegt, die sich im Grunde genommen als das Pendant der Data Science in der Business Intelligence interpretieren lässt. Stubbs sieht dabei Business Analytics wie folgt [Stubbs 2013]:

»The cornerstone of business analytics is pure analytics. Although it is a very broad definition, analytics can be considered any data-driven process that provides insight. It may report on historical information or it may provide predictions about future events; the end goal of analytics is to add value through insight and turn data into information.«

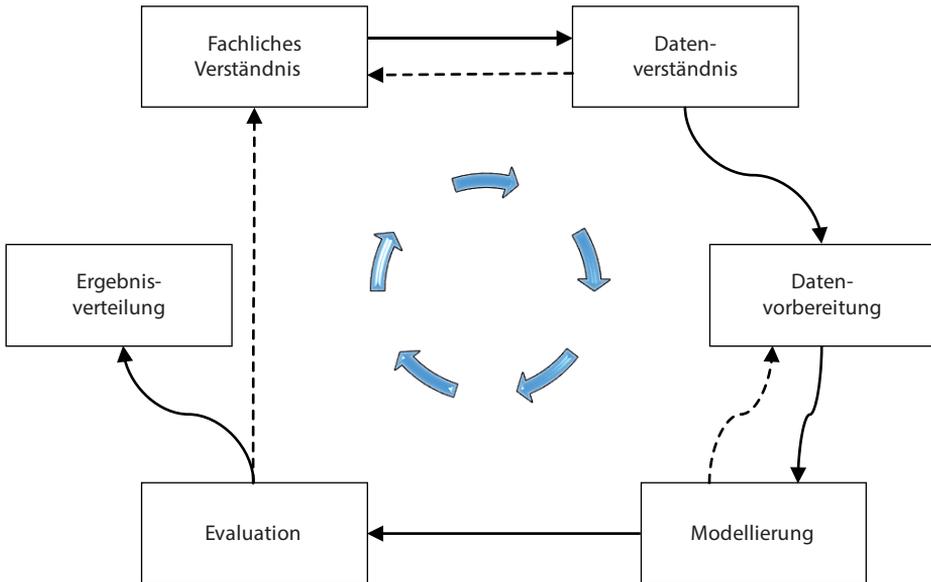
Stubbs Definition und unser Verständnis von Data Science, das wir in diesem Buch zugrunde legen wollen, überlappen sich damit größtenteils. Im Folgenden wird im Buch der Begriff Business Analytics zwar zugunsten von Data Science (vgl. Abschnitt 1.2) aufgegeben, der für die datenanalytischen Methoden und Vorgehensweisen stehen soll. Zum besseren Verständnis und um nicht zuletzt die Ähn-

lichkeiten im Vorgehen zu veranschaulichen, erfolgt aber zunächst eine Herleitung des Begriffs Business Analytics.

Was aber ist das Ziel der Business Analytics und inwieweit wird sich die Rolle von Business Analytics durch Methoden und Technologien aus dem Bereich Big Data und Data Science verändern? Haben die Unternehmen mit Business Analytics nicht auch Data Mining betrieben? Diese Fragen lassen sich beliebig erweitern. Leider stehen den Fragen nur wenige präzise Antworten gegenüber. Wenn man versucht, sich diesem Thema von einer fachlichen Seite zu nähern, stellt man schnell fest, dass die Datenorientierung im betriebswirtschaftlichen Handeln zugenommen hat. Diese Zunahme entsteht auch durch die wachsende Integration unterschiedlicher unternehmensinterner und -externer Systeme. Basierend auf entstehenden Datensammlungen werden im Unternehmen schon von jeher Entscheidungen getroffen. Aktuell ist jedoch eine deutliche Zunahme der Datenorientierung bei Entscheidungen auf allen Unternehmensebenen zu verzeichnen. Dabei gerät nun auch zunehmend die technische und methodische Unterstützung bei der Entscheidungsfindung in die Diskussion – und im BI-Umfeld finden wir diese Diskussion unter der Überschrift *Business Analytics*.

Unter Business Analytics wird die kontinuierliche Erforschung und Untersuchung von vergangenheitsorientierten Geschäftsdaten verstanden, um darin Erkenntnisse sowohl über die abgelaufene als auch die kommende Geschäftstätigkeit zu erlangen, die wiederum in die einzelnen zu planenden Geschäftsaktivitäten einfließen [Felden 2012]. Die Kontinuität entsteht durch die regelmäßige Ausführung von Analysetätigkeiten, die sich entsprechend in einer Ablauforganisation implementieren lassen. Iterativ sind derartige Aktivitäten, weil im Analyseprozess häufig eher neue Fragen als abschließende Antworten entstehen, die letztlich zu untersuchen sind. So kann die bisherige Geschäftstätigkeit nachvollzogen werden, um Verbesserungen bei neuen Handlungen zu ermöglichen.

Letztlich ist Business Analytics ein Prozess, der aus den in der folgenden Abbildung gezeigten Schritten besteht und eng an das in Abschnitt 1.3 vorgestellte CRISP-DM angelehnt ist.



**Abb. 1-1** Schritte der Business Analytics

Das fachliche Verständnis bestimmt die Auswahl der Daten, wobei dabei oftmals Rückfragen bzw. Nachbesserungen erforderlich sind, sodass fachliches Verständnis und Datenverständnis interdependent sind. Die vorliegenden Daten werden entsprechend aufbereitet in ein Modell überführt. Dabei bringt es die Modellbildung mit sich, dass die Aufbereitung neuerlich durchzuführen ist, da beispielsweise ein anderer Algorithmus als ursprünglich geplant genutzt wird. Die erzeugten Modelle sind zu evaluieren und deren Ergebnisse zur Nutzung an die jeweiligen Anwender weiterzuleiten. Die Erkenntnisse aus deren Nutzung fließen wieder als fachliches Verständnis in einen neuen Durchlauf ein.

Bereits seit Ende der 1990er-Jahre ist der KDD-Prozess (KDD = Knowledge Discovery in Databases) mit seinen Schritten der Datenauswahl, Vorverarbeitung, Transformation, Data Mining und der Ergebnisinterpretation theoretische Grundlage marktgängiger Software. Letztlich basiert auch die Business Analytics auf diesen Vorgehensschritten und erweitert diesen KDD-Prozess um eine Quellenbewirtschaftung zu Beginn und fachliche Handlung im Sinne einer zu treffenden Entscheidung und deren Durchsetzung am Ende des Prozesses. Im Weiteren wird die KDD um den Evaluationsschritt ergänzt, er dient dem Vergleich der erzeugten Modellvarianten anhand eines sogenannten Gütemaßes.

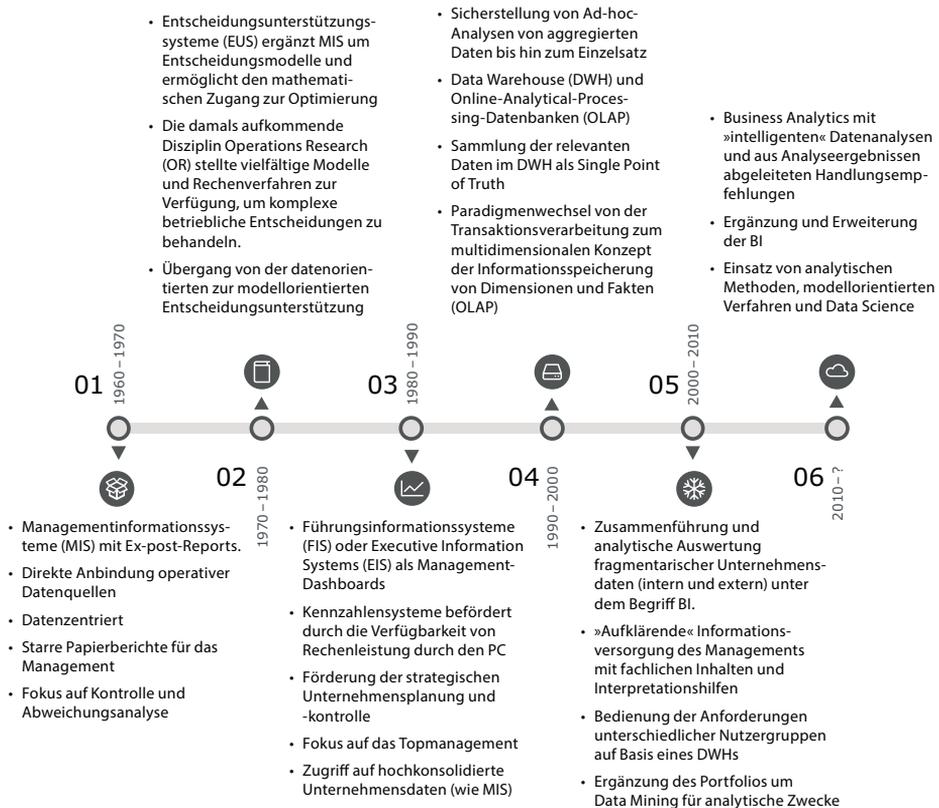
Somit liegt nun ein Prozess vor, der eine Langfristigkeit und damit eine strategische Komponente inhärent in sich birgt, da die Ergebnisse Entscheidungsgrundlage für das unternehmerische Handeln darstellen. Fachliche Analyseanforderungen und technische Komponenten zur zielgruppen- und aufgabenadäquaten Unterstützung sind in diesem Prozess gemeinschaftlich zu betrachten, um im

Rahmen der Informationslogistik, also die Daten zur richtigen Zeit dem richtigen Empfänger in der richtigen Qualität zur Verfügung zu stellen [Dinter & Winter 2008], eine sinnhafte Vollautomation zu erzeugen. Das informationslogistische Verständnis der Business Intelligence, also des Prozesses, Daten zu sammeln, aufzubereiten und zur Entscheidungsfindung zur Verfügung zu stellen [Chamoni & Gluchowski 2006], mündet in der praktischen Umsetzung eher in eine Standardorientierung mit konsistenten Kennzahlen (Metriken) und Analysen. Sie ist Dashboard-basiert mit vordefinierten Berichtsstrukturen zur Beantwortung vorab definierter Fragestellungen, sodass ein indirekter Zugriff auf die multidimensionalen Strukturen, Berichte und aggregierte Daten stattfindet, was jedoch auch zu einem Exception Reporting, also dem Triggern von automatisierten Informationsbereitstellungen bei Schwellenwertüberschreitungen [Felden & Buder 2012, S. 17 ff.], weitergedacht werden kann. Business Analytics ergänzt das Business-Intelligence-Verständnis um weitere Analysen von z. B. Geschäftsaktivitäten und richtet dabei den Fokus auf die Unterstützung von interaktiven und erforschenden Analysen durch Endanwender. Das Ziel ist die Sammlung neuer Erkenntnisse und damit eine Verständniskerngewinnung über vergangene Aktivitäten zur Entdeckung unbekannter Muster/Strukturen in den Datenbeständen. Dabei basiert Business Analytics auf Detaildaten, um einzelne Aktivitäten entsprechend betrachten und analysieren zu können.

Daten bzw. bereits daraus generierte Informationen zu besitzen, ist in den Unternehmen nicht mehr ein Wert an sich, vielmehr besteht der Wert darin, die Möglichkeit und Fähigkeit zu haben, Informationen aus unübersichtlichen Mengen von Daten und deren heterogenen Strukturen zu identifizieren und Entscheidungsträgern als Grundlage für unternehmerische Entscheidungen zur Verfügung zu stellen. Mit dem Fokus auf eine Datenauswertung ist dabei zwangsläufig das Thema der Business Analytics zunehmend in den Mittelpunkt gerückt. Auf dieser Basis verbindet Business Analytics moderne Verfahren der Auswertung von großen Datenvorräten, vor allem Data Mining, und maschinelles Lernen auf Grundlage der künstlichen Intelligenz und statistischer Methoden. Mittlerweile kombiniert Business Analytics einzelne Komponenten wie Kennzahlenkonzepte, Active/Realtime Warehousing, Data und Text Mining, User-Interface-Konzepte oder Systemintegration. Hierin liegt der eigentliche Nutzen; die Zusammenführung einzelner Komponenten bringt es mit sich, dass der Entscheider heute viel schneller auf Veränderungen in seinem Unternehmen oder der Unternehmensumwelt reagieren kann. Der strategische Mehrwert von Business Analytics wird damit deutlich. Entwicklungen der letzten Jahre haben das Image und den Agitationsrahmen von Business Analytics erweitert: Stichworte wie Systemintegration, Geschäftsprozessorientierung oder Benutzeroberflächendesign werden mit Business Analytics in Verbindung gebracht [Olson & Delen 2008, S. 151 ff.].

Sowohl Business Intelligence (BI) wie auch Business Analytics (BA) sind Begriffe, die am Ende einer langen Entwicklungsgeschichte der Managementun-

terstützungssysteme (MUS) stehen (siehe Abb. 1–2). Chronologisch wird die Genese der MUS in unterschiedliche Phasen eingeteilt, die jeweils vor dem Hintergrund der verfügbaren IT-Ressourcen zu sehen sind. Allen Phasen gemeinsam ist, dass nach Werkzeugen für eine adäquate Informationsversorgung für das Management gesucht wird. Vorrangig steht dabei die Unterstützung des Managements in der Entscheidungssituation an. Die folgende zeitliche Zuordnung ist nicht trennscharf, da sich die jeweiligen Konzepte überlagern und teilweise latent existieren. Es wird lediglich die dominante Begriffsprägung einer Epoche zugewiesen. Insgesamt stellt der Komplex MUS als Sammelbegriff aller Strömungen ein Kontinuum dar.



**Abb. 1–2** Die Phasen von MIS (Phase 01) zu Business Analytics (Phase 06)

Der Begriffswandel in Business Analytics verspricht einen intensiveren Einsatz von »intelligenten« Datenanalysen, verbunden mit direkten Handlungsempfehlungen, die aus den Analyseergebnissen abgeleitet werden. Dabei wird BI nicht diskreditiert, sondern eher in den Kontext der performanten Informationslieferung und aktiven Analyse gesetzt. Hingegen verspricht Business Analytics eine Aufklärung mittels Algorithmen über bestmögliche zukünftige Handlungen. Womit

bekannte Prognoseverfahren und Optimierungsrechnung (siehe Phase 2) erneut in den Fokus rücken. Die neue Qualität von Business Analytics wird in der sinnvollen Kombination von Methoden der Datenanalyse und Modellen liegen, die vor allem dem Umfeld der Data Science zuzurechnen sind. Die Konvergenz von datenorientierten und modellorientierten Verfahren scheint daher naheliegend und bringt tatsächlich neue Aspekte in die Betrachtung von MUS auf dem Zeitstrahl. Vergleichbar der Phase 2 treten Algorithmen in den Vordergrund, die automatisierte Entscheidungsprozesse ermöglichen, die auf großen polystrukturierten Datenbeständen (Big Data) in Realzeit Empfehlungen für bestmögliche Entscheidungen geben oder selbst entscheiden.

## 1.2 Data Science und angrenzende Gebiete

In der aktuellen Diskussion rund um die neuen Entwicklungen im Bereich der Informations- und Entscheidungssysteme kann man eine polyphone Stimmenvielfalt feststellen, die so manchen Betrachter verwirrt und manchmal sogar ratlos zurücklässt. Dabei stehen gerade die Abgrenzungen der Begriffe künstliche Intelligenz (aka KI, AI oder Artificial Intelligence), Data Science und Machine Learning im Fokus.

Historisch betrachtet wurde zunächst der Begriff künstliche Intelligenz geschaffen. Im Sommer 1956 fand am Dartmouth College in den Vereinigten Staaten eine von John McCarthy organisierte Konferenz zum Thema »Artificial Intelligence« statt. Im Laufe der nächsten Jahre wurden verschiedene Konzepte im Bereich der KI-Forschung verfolgt und zum Teil heftige Dispute über die Ausrichtung der KI und die zu verwendenden Werkzeuge ausgetragen.<sup>1</sup> Nachdem verschiedene Forschungsansätze auf konzeptionelle, zunächst unüberwindlich erscheinende Probleme gestoßen waren, folgte der sogenannte »AI-Winter« in den 1980er-Jahren. Neue Forschungsansätze (z.B. mehrschichtige neuronale Netze, der Backpropagation-Algorithmus oder rekurrente neuronale Netze), stark verbesserte Technologien in Form von Rechenleistung sowie das aufkommende Big-Data-Phänomen mit der damit einhergehenden Flut an zur Verfügung stehenden Daten führten nicht nur zu einem Revival der KI, sondern dazu, dass KI heute als die wichtigste und möglicherweise entscheidende Kompetenz für die wirtschaftliche Entwicklung eines Landes gesehen wird.<sup>2</sup>

- 
1. Erwähnt sei an dieser Stelle nur der bekannte Disput zwischen Marvin Minsky, der an der oben erwähnten Konferenz am Dartmouth College teilgenommen hatte und der über 50 Jahre hinweg das Forschungsgebiet künstliche Intelligenz am MIT vorangebracht hat, und Frank Rosenblatt, der das Konzept des Perzeptrons eingeführt hat.
  2. Man betrachte hier nur die massiven Investitionen Chinas in die KI-Forschung oder auch die KI-Initiative der Deutschen Bundesregierung ([https://www.bmbf.de/files/180718%20Eckpunkte\\_KI-Strategie%20final%20Layout.pdf](https://www.bmbf.de/files/180718%20Eckpunkte_KI-Strategie%20final%20Layout.pdf)).

Das Gebiet künstliche Intelligenz ist extrem facettenreich und stark interdisziplinär geprägt. Hier liegt auch der Grund, warum eine Definition von KI so schwer ist. Nach Winston lässt sich formulieren:

»Künstliche Intelligenz ist die Untersuchung von Berechnungsverfahren, die es ermöglichen, wahrzunehmen, zu schlussfolgern und zu handeln.«<sup>3</sup>

Damit versucht die KI-Forschung die menschlichen Wahrnehmungs- und Verstandesleistungen zu operationalisieren. Folgt man Görz, Schmid und Wachsmuth [Görz et al. 2013], kann man vereinfacht feststellen, dass es das Ziel der KI ist, Computerprogramme für Problembereiche zu entwickeln, die bislang nur von Menschen lösbar sind. Für sie ist KI als Teil der Informatik eine Ingenieurwissenschaft und als Teil der Kognitionswissenschaft auch Erkenntniswissenschaft. Entsprechend lassen sich zwei Ausprägungen unterscheiden: die starke KI und die schwache KI. Während die starke KI das Ziel hat, menschliche Problemlösungskreativität, Selbstbewusstsein und Emotionen abzubilden, fokussiert die schwache KI auf die Lösung konkreter Anwendungsprobleme durch Simulation von Intelligenz durch Methoden der Informatik, der Statistik und der Mathematik.

Hinsichtlich dieses hohen Maßes an Interdisziplinarität gibt es eine große Überlappung zur Data Science. Der Ursprung dieses noch recht jungen Zweigs wird zeitlich unterschiedlich verortet. Gehen Kelleher und Tierney [Kelleher & Tierney 2018] und andere häufig von Jeff Wus [Wus 1997] gehaltener Vorlesung »Statistics = Data Science?« aus, so führt Cao den Namen auf die Nennung des Begriffs im Vorwort eines 1974 publizierten Buches zu Berechnungsmethoden zurück, in dem es heißt, Data Science sei »the science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and sciences« [Cao 2017, S. 3]. Noch weiter zurück geht Donoho, der erste Ansätze bereits Mitte der 1950er-Jahre sieht [Donoho 2015, S. 1]. Bei Donoho findet sich auch die folgende Definition für Data Science:

»This coupling of scientific discovery and practice involves the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary applications.«

---

3. Winston nach [Görz et al. 2013]

Neben der Interdisziplinarität der Data Science rückt Donoho damit auch die Verknüpfung von wissenschaftlicher Entdeckung und Praxis in den Vordergrund. Die Data Science Association sieht ihre Wissenschaft wie folgt:

»Data Science« means the scientific study of the creation, validation and transformation of data to create meaning. [...] Data science uses scientific principles to get meaning from data and uses machine learning and algorithms to manage and extract actionable, valuable intelligence from large data sets.«<sup>4</sup>

Entsprechend ist der Data Scientist »[...] a professional who uses scientific methods to liberate and create meaning from raw data [...] The data scientist has a solid foundation in machine learning, algorithms, modeling, statistics, analytics, math and strong business acumen [...]«.

Damit wird deutlich, dass Machine Learning oder maschinelles Lernen eine der Methoden ist, die neben zahlreichen anderen in der Data Science zum Einsatz kommt. Maschinelles Lernen ist nach Wrobel, Joachims und Mroczk:

»[...] ein Forschungsgebiet, das sich mit der computergestützten Modellierung und Realisierung von Lernphänomenen beschäftigt« [Wrobel et al. 2013, S. 406].

Bei den eingesetzten Lernverfahren unterscheidet man das überwachte Lernen (supervised learning), das unüberwachte Lernen (unsupervised learning) sowie das Verstärkungslernen (reinforcement learning). Vielfach kommen hier neuronale Netze zum Einsatz, doch werden je nach Kontext und Fragestellung auch andere Verfahren genutzt. Die Autoren sehen Machine Learning, Data Mining und die »Knowledge Discovery in Databases« (KDD) als Teilgebiete der KI, die in den vergangenen Jahren zunehmend Eingang in praktische Anwendungen in Industrie und Wirtschaft gefunden haben. Die klassische Definition von KDD stammt von Fayyad, Piatetsky-Shapiro und Smyth:

»Knowledge Discovery in Databases describes the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data« [Fayyad et al. 1996].

Data Mining ist dabei als der Teilschritt dieses Prozesses zu sehen, der sich mit der Analyse beschäftigt. Im kommerziellen Bereich verschwimmt die Unterscheidung zwischen KDD und Data Mining jedoch häufig.

---

4. <http://www.datascienceassn.org/about-data-science>

Die Entwicklungen rund um Data Science fußen nicht zuletzt auf der enormen Menge an Daten, die Wissenschaftlern, Regierungen und natürlich auch den Unternehmen heute zur Verfügung stehen. Unter dem Schlagwort Big Data wird diese Entwicklung zusammengefasst. Big Data umfasst Methoden und Technologien für die hochskalierbare Integration, Speicherung und Analyse polystrukturierter Daten. Dabei bezieht man sich häufig auf die sogenannten 3Vs (Volume, Velocity und Variety), die zum Teil durch weitere Vs, wie etwa für Value, ergänzt werden (vgl. [Cai & Zhu 2015, S. 2]). Skalierbarkeit bezieht sich insbesondere auf die in der Regel hohen Datenvolumina (Data Volume), das schnelle Anfallen der Daten und die dafür notwendige hohe Datenverarbeitungs- und analysegeschwindigkeit (Data Velocity) sowie eine breite Quellen- und Datenvielfalt (Data Variety) (vgl. [Dittmar 2016, S. 56 f.]).

### 1.3 Vorgehen in Data-Science-Projekten

Bei Data-Science-Projekten hat sich ein iteratives, agiles Vorgehen bewährt, das sich in der Regel an dem Vorgehensmodell Cross-Industry Standard Process for Data Mining, kurz CRISP-DM, orientiert (siehe Abb. 1–3).

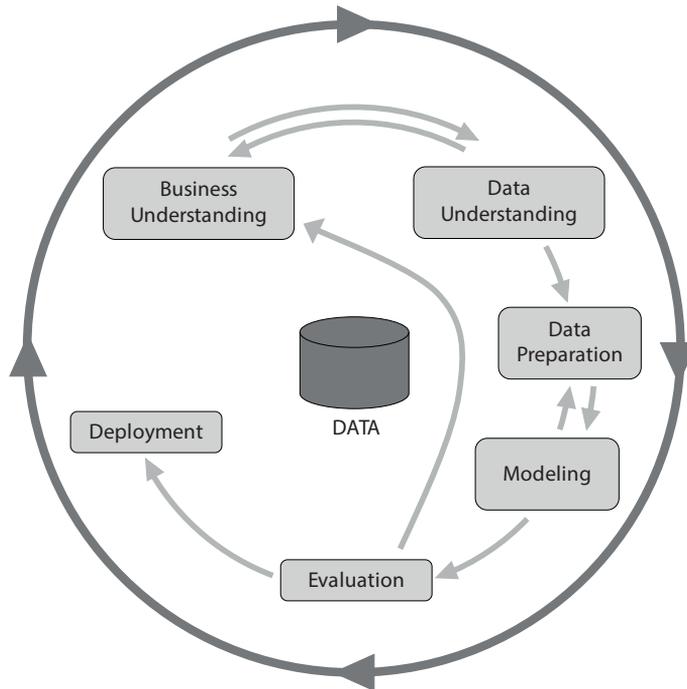
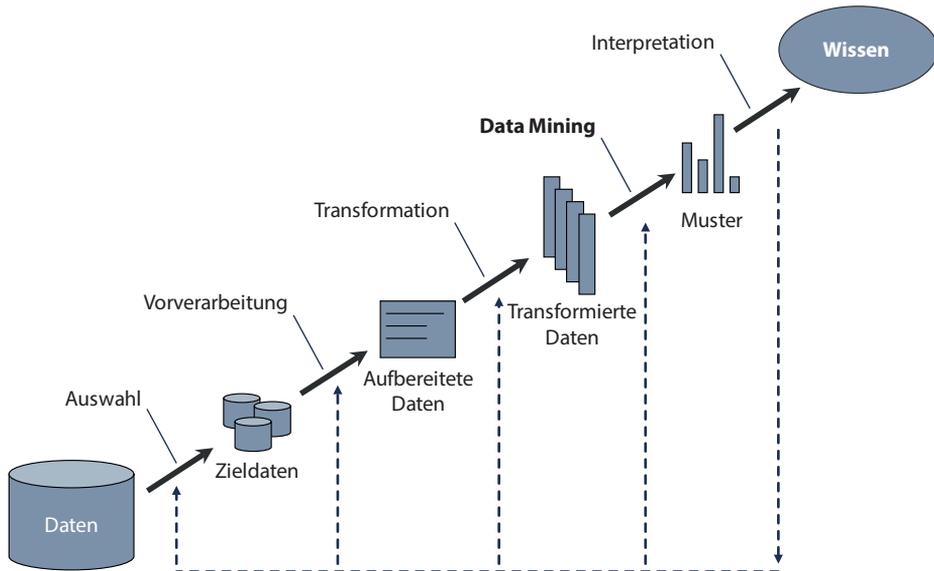


Abb. 1-3 CRISP-DM mit sechs Phasen

CRISP-DM besteht aus sechs Phasen, die als zyklischer Prozess zu verstehen sind. Das Business Understanding (fachliches Verständnis) umfasst die Bestimmung der Geschäftsziele, die Beurteilung der aktuellen Situation sowie die konkreten fachlichen Ziele des Data-Science-Projekts und – verbunden damit – die Planung der weiteren Aktivitäten. Im Data Understanding (Verständnis der Daten) werden die Daten und Datenquellen identifiziert, die zur Beantwortung der analytischen Fragestellung notwendig sind. Dieser Schritt enthält auch eine erste Datenerfassung, Datenbeschreibung und die Überprüfung der Datenqualität. Sind die Datenquellen identifiziert und die Daten zusammengestellt, erfolgt eine explorative Datenanalyse, um erste erkennbare Muster zu sichten. Neben der visuellen Analyse und den deskriptiven statistischen Verfahren können auch BI-typische Datenaufbereitungen und -navigationen hilfreich sein, um erste Erkenntnisse über den vorliegenden Datenbestand zu gewinnen. Grundsätzlich folgen solche Analysen einem Prozess, um einen zielorientierten und nachvollziehbaren Ablauf der jeweiligen Datenanalyse zu ermöglichen. Bereits die Business Intelligence liefert hier einen allgemeinen Ablauf, der mit der Datenextraktion, der Transformation und dem Laden in das Data Warehouse beginnt und im weiteren Vorgehen vorab definierte Auswertungen mit einem entsprechenden Analysewerkzeug ermöglicht.

Im Rahmen der Data Preparation (Datenvorbereitung) sind die Daten so aufzubereiten, dass diese im nächsten Schritt für das Training der Modelle verwendet werden können. Modeling (Modellierung) benennt die Parametrisierung und das eigentliche Lernen eines Modells mithilfe von Data-Mining-Algorithmen zur Lösung der Aufgabenstellung. Diese können Regressionsanalyse, Assoziationsanalyse, Klassifikations- oder Clusteranalysen sein. Die Evaluierung erfolgt einerseits bezogen auf die Ergebnisqualität des gelernten Modells und andererseits gegen das Ziel der fachlichen Aufgabenstellung sowie der betriebswirtschaftlichen Bewertung. Die Gewinnung des Geschäftsverständnisses ist ein iteratives Prozedere, in dem die Ergebnisse durch unterschiedliche Algorithmen und Visualisierungen ausgewertet werden, um ein tieferes Verständnis über die erzielten Ergebnisse zu erhalten. Das abschließende Deployment ist die Übertragung der Ergebnisse in die organisationalen Operationen, seien es Vorhersagen zu Marketingaktivitäten oder zu Wartungszyklen der Maschinen in der Fertigung. Zu einem Deployment gehört allerdings auch, dass diese Modelle auf Veränderungen der Betriebsbedingungen zu überwachen sind, da sich Bedingungslagen und Strukturen ändern können, sodass die Gültigkeit von Ergebnissen nicht mehr vorliegt und ein neues Verfahren zu initiieren ist.

Neben CRISP-DM gibt es alternative Ansätze wie beispielsweise der KDD-Prozess nach Fayyad oder SEMMA. Der fayyadsche Ansatz kennzeichnet sich durch die expliziten Phasen Datenauswahl, Datentransformation, Data Mining und die darauffolgende Interpretation (vgl. Abb. 1–4). Implizit wird dabei auch davon ausgegangen, dass Schritte iterativ ausgeführt werden.



**Abb. 1-4** Überblick über den KDD-Prozess (nach [Fayyad et al. 1996])

SEMMA, ein früher herstellernaher Ansatz, geht auch phasenorientiert vor, wobei hier von Datenauswahl (Sampling), Datenverständnis (Explore), Modifikation, Algorithmanwendung (Model) und Ergebnisevaluation (Assess) gesprochen wird.

Die Vorgehensweise ist in fast jedem Data-Science-Projekt iterativ und die Phasen werden mehrmals durchlaufen. Dies bedingt, dass die Nachvollziehbarkeit der einzelnen Schritte wie Datenauswahl, Transformationen etc. und auch das Training in den verschiedenen Phasen ein wesentlicher Punkt ist, der von Projektbeginn an berücksichtigt werden muss. Nur wenn die Nachvollziehbarkeit der Analyse sichergestellt ist, sind eine fundierte Bewertung der Ergebnisse und die Reproduktion der Analyse in der Produktivumgebung und damit das Deployment möglich.

## 1.4 Struktur des Buches

Das vorliegende Werk ist in einen Grundlagenteil und einem Praxisteil mit Fallstudien gegliedert. Im Grundlagenteil werden verschiedene Aspekte von Data Science erläutert und im zweiten Teil des Buches werden die Grundlagen anhand von konkreten Fallstudien aus Data-Science-Projekten mit deren spezifischen praktischen Problemstellungen und Lösungsansätzen dargestellt. Die Projektberichte nehmen Bezug auf die Grundlagen des ersten Teils, sind in sich jedoch geschlossen und können in einer frei wählbaren Reihenfolge gelesen werden.

In Kapitel 2 diskutiert Uwe Haneke, ob Analytics wirklich das neue BI ist und welche Erkenntnisse die Unternehmen daraus ziehen können. Er geht der Frage nach, warum sich Data Science gerade jetzt so rasant verbreitet und in den Unternehmen Fuß fasst. Im Anschluss wird erläutert, warum dieser Entwicklung eine so große Bedeutung zukommt und wie sich eine mögliche Fusion der alten BI-Welt mit der neuen, erweiterten Analytics-Welt in den Informationssystemen der Unternehmen darstellen könnte.

In Kapitel 3 zeigen die Autoren Marc Beierschoder, Benjamin Diemann und Michael Zimmer anhand eines konkreten Beispiels, unter welchen Rahmenbedingungen der Einsatz von Data Science im Allgemeinen und KI im Speziellen zum Erfolg in einem Unternehmen führen kann.

Anschließend stellt Christoph Tempich in Kapitel 4 vor, wie die Konzeption und die Entwicklung von Data-driven Products erfolgen kann und auf welche Punkte dabei geachtet werden muss. Unter anderem werden die Aspekte Ideenfindung, Value Proposition Design und Zielgrößen näher untersucht und die Messung der Qualität eines Datenprodukts mithilfe einer Feedbackschleife vorgeschlagen.

In Kapitel 5 stellen Stephan Trahasch und Carsten Felden im Überblick grundlegende Methoden der Data Science vor, die in den Phasen Data Understanding, Data Preparation, Modeling und Evaluation Verwendung finden.

Angesichts weiter zunehmender zur Verfügung stehender Daten kommt der Feature Selection eine immer größere Bedeutung zu. Diesem wichtigen Aspekt wird in Kapitel 6 von Bianca Huber Rechnung getragen.

Klaus Dorer führt in Kapitel 7 in die Grundlagen neuronaler Netzwerke ein und erläutert anhand von Deep Convolutional Neural Networks für die Objekterkennung in Bildern, wie Deep Learning funktioniert. Neben einigen praktischen Anwendungen gibt das Kapitel auch einen Überblick über die zahlreichen verfügbaren Frameworks und Standarddatensätze für Deep Learning.

Nur mit geeigneten Datenarchitekturen als Grundlage können Unternehmen zukünftig Data Science und Artificial-Intelligence-basierte Anwendungsfälle abbilden. Wie solch eine Datenarchitektur aussehen kann, erläutern Michael Zimmer, Benjamin Diemann und Andreas Holzhammer in Kapitel 8.

Self-Service und Befähigung der Anwender sind in der BI ein aktuelles Thema. In Kapitel 9 stellen Uwe Haneke und Michael Zimmer vor, warum gerade Self-Service-Szenarien in Data Science wichtig sind, um im Unternehmen die analytische Sichtweise zu verankern. Daneben stellen die Autoren ein Konzept für eine differenzierte Data & Analytics Governance vor, da das Thema Governance im Data-Science-Umfeld immer mehr an Bedeutung gewinnt.

In Kapitel 10 diskutieren Victoria Kayser und Damir Zubovic die Rolle von Data Privacy für Analytics und Big Data. Neben der rechtlichen und technischen Ausgestaltung von Data Privacy im Unternehmen diskutieren die Autoren auch, wie die Unternehmen mit der Herausforderung umgehen, Analytics und KI in ihre Prozesse zu integrieren.

Anschließend führen Matthias Haun und Pfarrer Gernot Meier in Kapitel 11 ein Gespräch zur digitalen Ethik, geben einen Einblick in die Vielgestaltigkeit der Diskussion und zeigen auf, welche Fragestellungen auf uns zukommen.

Mit Kapitel 12 beginnt der Praxisteil des Buches. In der ersten Fallstudie stellt Shirin Glander dar, wie mit Methoden der Data Science Vorhersagen zum Churn-Verhalten von Kunden getroffen werden können. Zur prädiktiven Analyse wird ein neuronales Netz mit Keras und TensorFlow trainiert und dies mit einem Stacked-Ensemble-Modell auf Basis von H2O verglichen.

In Kapitel 13 gibt Nicolas March einen Einblick in die Erfahrungen mit Data Science und in die Wirtschaftlichkeitsbetrachtungen bei der Auswahl und Entwicklung von Data-Science-Anwendungen im Online-Lebensmittelhandel.

Mikio Braun stellt in Kapitel 14 vor, wie Zalando die Grundlagen für Analytics, BI und Data Science zum unternehmensweiten Einsatz geschaffen hat und welche Herausforderungen das Unternehmen zu meistern hatte.

Predictive Maintenance hat für die industrielle Produktion eine großes Potenzial. Marco Huber erläutert in Kapitel 15 die verschiedenen Strategien der Instandhaltung und wie die Nutzung von unterschiedlichen Daten, die während der Produktion anfallen, für die vorausschauende Instandhaltung unter Einsatz von Verfahren der Statistik und des maschinellen Lernens erfolgen kann.

Caroline Kleist und Olaf Pier beschreiben in Kapitel 16, wie Scrum in Data-Science-Projekten bei der Volkswagen Financial Services AG erfolgreich eingesetzt wird und mit welchen Herausforderungen sie konfrontiert wurden, und geben Empfehlungen zum Einsatz von Scrum für Data-Science-Teams.

In Kapitel 17 zeigt Matthias Meyer, wie durch die Konzeption und Pilotierung zusätzlicher datenbasierter Serviceangebote für einen Betreiber eines Kundenkartenprogramms ein Mehrwert geschaffen werden konnte.

Abschließend beschäftigt sich Kapitel 18 mit dem Einsatz von KI und Data Science in der Versicherungsbranche. Am Beispiel der Zurich Versicherung zeigen die Autoren anschaulich, vor welchen Herausforderungen das Unternehmen stand und wie KI im Wertschöpfungsprozess heute in verschiedenen Anwendungsfällen in der Versicherung zum Einsatz kommt.