

# Bei Data Science durchblicken



## In diesem Kapitel

- ▶ Data Science definieren
- ▶ Data Science mithilfe seiner Schlüsselkomponenten definieren
- ▶ Machbare Lösungen für Ihre eigenen Datenanforderungen bestimmen

Seit geraumer Zeit werden wir alle mit Daten total überschwemmt. Sie stammen von jedem Computer, jedem mobilen Gerät, jeder Kamera und jedem Sensor – und mittlerweile auch von Uhren und anderen tragbaren technischen Geräten. Sie werden bei jeder Aktion in den sozialen Netzwerken, bei jeder von uns gespeicherten Datei, bei jedem geschossenen Foto, jeder eingereichten Suchanfrage erzeugt. Auch wenn wir so etwas Einfaches machen, wie bei Google nach der nächsten Eisdiele zu suchen, entstehen viele Daten.

Obwohl das Eintauchen in Daten nichts Neues ist, haben Sie sicher bemerkt, dass sich das Phänomen beschleunigt. Seen, Pfützen und Flüsse von Daten haben sich in Fluten und wahre Tsunamis aus strukturierten, halbstrukturierten und nichtstrukturierten Daten verwandelt, die von fast jeder Aktivität ausströmen, die in der digitalen oder realen Welt stattfinden. Willkommen in der Welt von *Big Data*!

Sie haben sich vielleicht schon gewundert, »Was ist der Sinn dieser Daten? Warum wertvolle Quellen verwenden, um sie zu erzeugen und zu sammeln?« Obwohl schon vor einem Jahrzehnt niemand in der Lage war, groß Gebrauch von den erzeugten Daten zu machen, haben sich die Zeiten heute definitiv geändert. Als Daten-Ingenieure bezeichnete Spezialisten finden fortwährend innovative und leistungsstarke neue Wege, um unvorstellbar gewaltige Datenmengen zu fangen, zuzuordnen und zu verdichten. Andere Spezialisten, Datenforscher genannt, arbeiten daran, wertvolle und verfolgbare Einblicke aus diesen Daten zu gewinnen.

In ihrer reinsten Form steht Data Science für eine Optimierung von Prozessen und Quellen. Anhand von Data Science gewinnen Sie *Einblicke in Daten* – Einblicke, die Sie verwenden können, um Ihr Unternehmen, Ihre Investitionen, Ihre Gesundheit und auch Ihre Lebensführung und Ihr soziales Leben zu verstehen und zu verbessern. Für jedes vorstellbare Ziel oder Bestreben können Sie mit der Data Science Methoden finden, die Ihnen helfen, den direktesten Weg von dort, wo Sie sind, zu dem Punkt, wo Sie sein wollen, zu kennen und vorherzusagen und jedes Schlagloch auf dem Weg dazwischen vorzusehen.

## Schauen, wer von Data Science Gebrauch machen kann

Die Ausdrücke »Data Science« und »Data Engineering« werden oftmals falsch und inkonsistent verwendet, also lassen Sie mich hier beginnen, indem ich erkläre, dass diese beiden Bereiche in der Tat getrennte und unterschiedliche Fachgebiete sind. *Data Science* ist die Verwendung

von Berechnungsverfahren, um nützliche und verfolgbare Erkenntnisse aus Ausgangsdaten abzuleiten. *Data Engineering* ist dagegen ein Ingenieurbereich, der den Zweck hat, Engpässe bei der Datenverarbeitung und Probleme bei der Behandlung von Daten für Anwendungen zu überwinden, die große Mengen, Arten und Geschwindigkeiten von Daten nutzen. Sowohl bei der Data Science als auch beim Data Engineering ist es üblich, mit den folgenden drei Arten von Daten zu arbeiten:

- ✓ **Strukturierte Daten:** Daten, die in einem herkömmlichen relationalen Datenbanksystem gespeichert und verarbeitet wurden.
- ✓ **Unstrukturierte Daten:** Daten, die gewöhnlich durch menschliche Aktivitäten erzeugt wurden und nicht in eine strukturierte Datenbank passen.
- ✓ **Halbstrukturierte Daten:** Daten, die nicht in eine strukturierte Datenbank passen, die aber trotzdem durch Markierungen strukturiert sind, die nützlich sind, um eine Form von Ordnung und Hierarchie bei ihnen zu erzeugen.

Viele Leute glauben, dass nur große Organisationen, die über ausreichende finanzielle Mittel verfügen, Methoden der Data Science anwenden, um ihr Unternehmen zu optimieren und zu vervollkommen, aber das ist nicht der Fall. Die Vermehrung der Daten hat eine Nachfrage nach Einblicken erzeugt, und diese Nachfrage ist in viele Aspekte unserer modernen Kultur eingebettet. Daten und die Notwendigkeit von Einblicken in diese Daten sind allgegenwärtig. Da Organisationen jeder Größe zu erkennen beginnen, dass sie in einer daten-gesteuerten, auf Wettbewerb ausgerichteten Umgebung mit einer Friss-oder-Stirb-Mentalität eingetaucht sind, entwickelt sich Daten-Fachwissen zum Herzstück und Rüstzeug in nahezu allen Geschäftsbereichen.

Was aber bedeutet das für einen normalen Menschen? Es heißt zunächst, dass sich unsere Kultur verändert hat und Sie sich auf dem Laufenden halten müssen. Das bedeutet nicht, dass Sie zurück in die Schule gehen und einen Abschluss in Statistik, Computerwissenschaften oder Data Science machen müssen. In dieser Hinsicht unterscheidet sich die Datenrevolution nicht so sehr von jeder anderen Veränderung, die die Industrie in der Vergangenheit getroffen hat. Tatsache ist: Um weiter aktuell zu bleiben, müssen Sie nur die Zeit und die Anstrengung aufbringen, um die Fähigkeiten zu erwerben, die Sie auf dem Laufenden halten. Wenn es darauf hinausläuft, zu lernen, wie Data Science anzuwenden ist, können Sie einige Kurse belegen, sich selbst mit im Netz zur Verfügung stehenden Quellen etwas beibringen, Bücher wie dieses lesen und Veranstaltungen besuchen, bei denen Sie lernen, was Sie wissen müssen, um in Bestform zu bleiben.

Wer kann Data Science verwenden? Sie können es. Ihr Betrieb kann es. Ihr Arbeitgeber kann es. Jeder, der ein wenig Verständnis und Übung hat, kann beginnen, Daten-zentrierte Sichtweisen zu verwenden, um sein Leben, seine Karriere und das Wohlergehen seines Unternehmens zu verbessern. Data Science stellt eine Änderung der Art und Weise dar, wie Sie an die Welt herangehen. Menschen sind es gewohnt, zu arbeiten und auf Ergebnisse zu hoffen, aber Dateneinblicke bieten die Vision, die die Menschen benötigen, um Veränderungen starten und

Dinge gut ablaufen zu lassen. Sie können eine Daten-zentrierte Sichtweise verwenden, um die folgenden Änderungen zu bewirken:

- ✓ Optimierung von betrieblichen Systemen und Kapitalrenditen (englisch *Return on Investment*, ROI) für jede messbare Aktivität.
- ✓ Verbesserung der Effektivität von Verkäufen und Marketing-Initiativen, ob sie nun Teil einer betrieblichen Marketing-Kampagne oder lediglich ein persönlicher Aufwand sind, um bessere Beschäftigungsmöglichkeiten für sich selbst zu sichern.
- ✓ Der Konkurrenz bei den allerneuesten Entwicklungen in jedem Bereich eine Nasenlänge voraus zu sein.
- ✓ Gemeinschaften sicherer zu erhalten.
- ✓ Die Welt für die zu einem besseren Platz zu machen, die weniger wohlhabend sind.

## ***Die Teile des Data-Science-Puzzles betrachten***

Um Data Science im eigentlichen Sinne des Wortes anzuwenden, benötigt man die analytischen Kenntnisse der Mathematik und Statistik, Programmierfertigkeiten und Fachkompetenz. Ohne Fachkompetenz könnten Sie sich auch Mathematiker oder Statistiker nennen. Genauso wie Software-Programmierer ohne Fachkompetenz und analytisches Fachwissen besser als Software-Ingenieur oder -Entwickler denn als Datenwissenschaftler betrachtet werden.

Da die Nachfrage nach einer Daten-zentrierten Sichtweise exponentiell wächst, ist jeder Bereich gezwungen, Data Science anzuwenden. Von daher haben sich verschiedene Richtungen von Data Science herauskristallisiert. Hier folgen nur ein paar Namen, unter denen Experten jeder Disziplin Data Science verwenden: Ad-Tech-Datenwissenschaftler, Banking Digital Analyst, Klinischer Datenwissenschaftler, Datenwissenschaftler der Geo-Ingenieurwissenschaften, Datenwissenschaftler der Geografie, Datenwissenschaftler im Personalbereich des Einzelhandels, Klinischer Informationsanalytiker im Bereich Pharmakometrie. Da es oftmals so scheint, dass man nicht die Übersicht darüber behalten kann, wer ein Datenwissenschaftler ohne Scorecard (= Ergebniskarte) ist, nehme ich mir in den folgenden Abschnitten die Zeit, um die Schlüsselkomponenten hier darzulegen, die ein Teil jeder Anwendung der Data Science sind.

## ***Daten sammeln, abfragen und bearbeiten***

Dateningenieure haben die Aufgabe, große Mengen von strukturierten, unstrukturierten und halbstrukturierten *Big Data* zu erfassen und zu sortieren; Daten, die die Verarbeitungskapazität herkömmlicher Datenbanken übersteigen, da sie zu umfangreich sind, sich zu schnell bewegen oder nicht die strukturellen Anforderungen herkömmlicher Datenbanken erfüllen. Um es noch einmal zu betonen, die Aufgaben von Dateningenieuren unterscheiden sich von der in der

Data Science durchgeführten Arbeit, die ihren Schwerpunkt eher in der Analyse, Vorhersage und Sichtbarmachung hat. Trotz dieser Unterscheidung verrichtet ein Datenwissenschaftler, wenn er die zur Analyse notwendigen Daten sammelt, befragt und gebraucht, eine Arbeit, die große Ähnlichkeit mit der eines Dateningenieurs hat.

Obwohl man auch aus einer einzigen Datenquelle nützliche Einsichten erhalten kann, liefert die Kombination verschiedener relevanter Quellen oftmals die kontextbezogenen Informationen, die für bessere datenkundige Entscheidungen erforderlich sind. Ein Datenwissenschaftler kann mehrere Datensätze abarbeiten, die in einer Datenbank oder auch in mehreren verschiedenen Datenspeichern gespeichert sind. (Mehr über die Arbeit mit zusammengesetzten Datensätzen erfahren Sie in Kapitel 3.) In anderen Fällen werden Originaldaten über ein von Software und Dateningenieuren aufgebautes Cloud-Netzwerk gespeichert und verarbeitet.

Unabhängig davon, wie die Daten zusammengesetzt und wo sie gespeichert sind: Wenn Sie Data Science betreiben, müssen Sie die Daten nahezu immer abfragen. Mit anderen Worten, Sie schreiben Befehle, um die relevanten Datensätze aus dem Speichersystem herauszuziehen. Meistens werden Sie eine Datenbankabfragesprache wie die Structured Query Language (SQL) verwenden. (Kapitel 16 enthält alle Informationen über SQL; wenn Sie das Kurzwort erschreckt, springen Sie einfach jetzt zu diesem Kapitel.) Ob Sie nun eine Anwendung benutzen oder eine kundenspezifische Analyse mithilfe einer Programmiersprache wie R oder Python verwenden, Sie können aus einer Reihe von allgemein akzeptierten Dateiformaten wählen. Zu diesen Formaten gehören:

- ✓ **Das Dateiformat CSV (englisch *Comma-Separated Values*):** Nahezu jede Plattform und webbasierte Anwendung akzeptiert dieses Dateiformat, genauso wie allgemein gebräuchliche Skriptsprachen wie Python und R.
- ✓ **Skripts:** Die meisten Datenwissenschaftler wissen, wie man Programmiersprachen wie Python und R zur Analyse und Veranschaulichung von Daten verwendet. Diese Skriptdateien haben jeweils die Endung `.py` bzw. `.r`.
- ✓ **Anwendungsdateien:** Excel ist zur kurzen und schmerzlosen Stichprobenuntersuchung von kleinen bis mittleren Datensätzen nützlich. Excel-Dateien haben die Endung `.xls` oder `.xlsx`. Anwendungen zur Untersuchung von Geodaten wie ArcGIS und QGIS speichern mit firmeneigenen File-Formaten (mit der Endung `.mxd` bei ArcGIS und `.qgs` bei QGIS).
- ✓ **Dateien in der Web-Programmierung:** Wenn Sie kundenspezifische Visualisierungen in Webbrowsern erstellen, dann arbeiten Sie vielleicht mit D3.js; dies steht für datengesteuerte Dokumente, eine JavaScript-Bibliothek für die Visualisierung von Daten. Wenn Sie mit D3.js arbeiten, werden Ihre Dateien mit der Endung `.html` gespeichert.

## ***Von der Mathematik und Statistik Gebrauch machen***

Data Science verlässt sich stark auf die Mathematik- und Statistikkenntnisse einer Fachkraft, da man diese Fähigkeiten benötigt, um die Daten und ihre Signifikanz zu verstehen. Diese Fähigkeiten sind darüber hinaus in der Data Science nützlich, da man mit damit Prognosen und Entscheidungsmodelle erstellen sowie Hypothesen testen kann.

Bevor eine ausführliche Diskussion der mathematischen und statistischen Methoden beginnen kann, ist es notwendig, hier kurz Pause zu machen und den Unterschied zwischen den Gebieten der Mathematik und der Statistik zu erklären. Mathematiker verwenden deterministische numerische Methoden und eine schlussfolgernde Argumentation, um zu einer quantitativen Beschreibung der Welt zu gelangen. Dagegen ist die Statistik eine Art von Wissenschaft, die von der Mathematik abgeleitet ist, deren Schwerpunkt aber die Verwendung einer stochastischen Methode ist, das heißt einer Methode auf der Grundlage von Wahrscheinlichkeiten. Somit verwendet sie eine induktive Argumentation, um zu einer quantitativen Beschreibung der Welt zu gelangen.

### ***Mathematische Modellierungen für Aufgaben im Bereich Data Science verwenden***

Datenwissenschaftler verwenden mathematische Methoden, um Entscheidungsmodelle zu erstellen sowie Näherungen und Vorhersagen bezüglich der Zukunft zu machen. In Kapitel 7 werden einige komplexe angewandte mathematische Herangehensweisen vorgestellt, die für die Arbeit im Bereich der Data Science nützlich sind.



Dieses Buch geht davon aus, dass Sie recht zuverlässige Grundkenntnisse in Mathematik besitzen. Es wäre vorteilhaft, wenn Sie Kurse in Analysis oder auch Linearer Algebra belegt hatten. Ich habe mich jedoch sehr bemüht, mich auf die Leser einzustellen. Mir ist schon klar, dass Sie vielleicht nur über begrenzte mathematische Kenntnisse verfügen (Fortgeschrittene Algebra oder Analysis im Geschäftsleben), daher habe ich versucht, fortgeschrittene mathematische Konzepte in einer einfachen Sprache so darzustellen, dass sie für jeden leicht zu verstehen sind.

### ***Mit statistischen Methoden Einblicke erhalten***

In der Data Science sind statistische Methoden nützlich, um ein besseres Verständnis für die Signifikanz der Daten zu bekommen sowie zur Überprüfung von Hypothesen, zum Simulieren von Szenarien und zum Erstellen von Prognosen über zukünftige Ereignisse. Fortgeschrittene statistische Fähigkeiten sind auch unter quantitativen Analytikern, Ingenieuren und Naturwissenschaftlern eher selten. Wenn Sie es im Bereich der Data Science zu etwas bringen wollen, nehmen Sie sich die Zeit, um sich bezüglich einiger statistischer Methoden wie der linearen Regression, der normalen quadratischen Regression, Monte-Carlo-Simulationen und Zeitreihenanalysen auf den neuesten Stand zu bringen. All diese Methoden werden in Kapitel 4 im Zusammenhang mit Wahrscheinlichkeiten und Statistik erläutert.

Die gute Neuigkeit ist, dass Sie nicht alles wissen müssen. Es ist nicht so, dass Sie sofort losgehen müssen, um einen Master-Abschluss in Statistik zu machen, damit Sie Data Science betreiben können. Sie sollten nur ein paar grundlegende Konzepte und Methoden aus der Statistik kennen, um Probleme zu lösen. (Ich erläutere einige dieser Konzepte und Methoden in Kapitel 4.)

## ***Programmierung: Teil des Spiels***

Programmieren ist unvermeidlich, wenn Sie auf dem Gebiet der Data Science arbeiten. Sie müssen in der Lage sein, Codes zu schreiben, damit Sie Ihren Computer anleiten können, wie er Ihre Daten handhaben, analysieren und visualisieren soll. Programmiersprachen wie Python und R sind zum Schreiben von Skripten zur Handhabung, Untersuchung und Visualisierung von Daten wichtig; SQL ist bei der Bearbeitung von Daten nützlich. Die JavaScript-Bibliothek D3.js ist eine neue großartige Möglichkeit, um richtig coole kundenspezifische, interaktive webbasierte Datenvisualisierungen zu erstellen.

Obwohl Programmieren eine Voraussetzung für Data Science ist, muss es wirklich nicht das große unheimliche Thema sein, das manche Leute daraus machen. Ihr Programm kann so raffiniert und komplex sein, wie Sie es haben möchten, aber Sie können auch eine eher einfache Methode verwenden. Obwohl diese Fähigkeiten für den Erfolg von großer Bedeutung sind, können Sie ziemlich mühelos genug über das Programmieren lernen, um auf hohem Niveau Data Science zu betreiben. In den Kapiteln 10, 14, 15 und 16 helfe ich Ihnen, Ihre Geschwindigkeit bei der Verwendung von D3.js zur webbasierten Visualisierung von Daten, beim Programmieren in Python und R und beim Befragen mit SQL zu erhöhen.

## ***Data Science in Ihrem Fachgebiet anwenden***

Es gibt bei Statistikern ein gewisses Maß an Eigensinn, wenn es um den Stellenwert von Data Science geht. Viele Statistiker haben schon aufgeschrien: »Data Science ist nichts Neues! Es ist nur ein anderer Name für das, was wir schon immer getan haben.« Obwohl ich für diese Einstellung Verständnis habe, bin ich gezwungen, mich auf die Seite der Datenwissenschaftler zu stellen, die deutlich erklären, dass Data Science eigenständig ist und sich definitiv von den statistischen Ansätzen unterscheidet, die sie einschließt.

Meine Meinung zur einzigartigen Natur von Data Science basiert zu einem gewissen Maß auf der Tatsache, dass Datenwissenschaftler oftmals Programmiersprachen verwenden, die in der traditionellen Statistik nicht gebraucht werden; zudem setzen sie Methoden ein, die sich vom Bereich der Mathematik ableiten. Doch der Hauptunterschied zwischen Statistik und Data Science ist die Notwendigkeit von Kenntnissen der Thematik.

Da Statistiker gewöhnlich ein begrenztes Maß an Fachkenntnissen außerhalb der Statistik haben, sind sie fast immer gezwungen, sich mit Fachleuten zu beraten, um genau zu prüfen, was ihre Ergebnisse bedeuten, und die Richtung festzulegen, in der sie fortfahren. Datenwissenschaftler müssen dagegen ein großes Fachwissen in dem Bereich besitzen, in dem sie arbeiten. Datenwissenschaftler erschaffen tiefe Einblicke und verwenden dann ihre fachspezifischen Kenntnisse, um genau zu verstehen, was diese Einsichten in Bezug auf den Bereich bedeuten, in dem sie arbeiten. Die folgende Liste enthält einige Wege, die aufzeigen, auf welche Weise Fachleute Data Science verwenden, um die Leistungsfähigkeit in den jeweiligen Industriezweigen zu verbessern:

- ✓ Ingenieure verwenden maschinelles Lernen, um die Energieeffizienz bei der Entwicklung von modernen Gebäuden zu erhöhen.

- ✓ Klinische Datenwissenschaftler arbeiten an der Personalisierung von Behandlungsplänen und verwenden die medizinische Informatik, um zukünftige Probleme bei Risikopatienten vorherzusagen und auszuschließen.
- ✓ Datenwissenschaftler im Vertrieb verwenden logistische Regressionen, um Kundenabwanderungen (die Abwanderung von Kunden für Ihr Produkt oder Ihre Dienstleistung zu einem Konkurrenten) vorherzusagen und zu vermeiden. Mehr zur Vermeidung der Kundenabwanderung finden Sie in den Kapiteln 3 und 20.
- ✓ Datenwissenschaftler im Journalismus durchforsten Webseiten nach neuen Daten, um die neuesten Eilmeldungen und Neuigkeiten zu entdecken und zu berichten (in Kapitel 18 erfahren Sie mehr darüber).
- ✓ Datenwissenschaftler in der Kriminalprävention verwenden ortsbezogene vorhersagende Modelle, um kriminelle Aktivitäten vorherzusagen, ihnen zuvorzukommen und sie zu verhindern (in Kapitel 21 folgen Einzelheiten über die Verwendung der Data Science zur Beschreibung und Vorhersage krimineller Aktivitäten).
- ✓ Weltverbesserer verwenden maschinelles Lernen, um entscheidende Informationen über Unglücksgebiete einzuordnen und zu melden, damit Entscheidungen über die Unterstützung in humanitären Angelegenheiten zeitnah getroffen werden können, was Sie in Kapitel 19 nachlesen können.

## ***Einblicke kommunizieren***

Eine weitere Fähigkeit ist für den Erfolg von Datenwissenschaftlern ebenfalls sehr wichtig (aber vielleicht nicht sofort einleuchtend): Als Datenwissenschaftler muss man über gute Fähigkeiten in der mündlichen und schriftlichen Kommunikation verfügen. Wenn ein Datenwissenschaftler nicht kommunizieren kann, werden all seine Fachkenntnisse und Einsichten in die Welt Ihrer Einrichtung nichts nützen.

Datenwissenschaftler müssen in der Lage sein, Dateneinblicke in einer Weise zu erklären, in der Mitarbeiter sie verstehen können. Nicht nur das, sie müssen in der Lage sein, eindeutige und aussagekräftige Visualisierungen und schriftliche Berichte anzufertigen. Meistens müssen die Menschen etwas für sich selbst betrachten, um es zu verstehen. Datenwissenschaftler müssen bei den Mitteln und Methoden ihrer Kommunikation kreativ und pragmatisch sein. (In den Kapiteln 9 und 18 werden die Themen der Visualisierung von Daten und des datengesteuerten Geschichtenerzählens ausführlich erläutert.)

## ***Die Grundlagen schaffen***

Organisationen und ihre Leiter schlagen sich trotzdem mit der Frage herum, wie man Big Data und Data Science am besten verwendet. Die meisten von ihnen wissen, dass moderne Analytik ihrer Organisation einen gewaltigen Vorsprung vor der Konkurrenz verschaffen kann, doch nur wenige haben eine Idee von den verfügbaren Möglichkeiten oder dem genauen



Nutzen, den Data Science bieten kann. In den folgenden Abschnitten werden die wichtigsten Vorgehensweisen und Vorteile vorgestellt, die durch die Einführung von Data Science bewirkt werden können.

## ***Mögliche Vorgehensweisen klären***

Wenn Sie Data Science bei einer Organisation oder auch in einer Abteilung einführen wollen, stehen drei Möglichkeiten zur Verfügung: Sie können ein Team aus Datenwissenschaftlern vor Ort einsetzen, die Arbeit auslagern und einem externen Team anvertrauen oder auf Cloud-Dienste zurückgreifen, die die Datenanalyse von Fachleuten für Mitarbeiter machen lassen, die vielleicht nur mäßige Datenkenntnisse haben.

## ***Ein eigenes betriebsinternes Team bilden***

Hier folgen drei Möglichkeiten zur Bildung eines betriebseigenen Teams aus Datenwissenschaftlern:

- ✓ **Ausbildung eigener Angestellter.** Die ist eine kostengünstige Möglichkeit. Wenn Sie Ihre Firma mit der Leistung von Data Science und Datenanalyse ausrüsten wollen, kann eine Ausbildung vorhandener Mitarbeiter datenkundige, hochspezialisierte Experten für Ihr firmeneigenes Team hervorbringen.
- ✓ **Ausbildung eigener Angestellter und Einstellung einiger Experten.** Eine weitere gute Möglichkeit besteht darin, vorhandene Mitarbeiter so auszubilden, dass sie datenwissenschaftliche Aufgaben auf hohem Niveau übernehmen können, und darüber hinaus einige neue Mitarbeiter zur Lösung fortgeschrittener Datenprobleme und für strategische Überlegungen einzustellen.
- ✓ **Experten einstellen.** Manche Firmen versuchen ihre Anforderungen durch das Einstellen von erfahrenen Datenwissenschaftlern oder Hochschulabgängern mit einem Abschluss in Data Science zu erfüllen. Das Problem bei dieser Herangehensweise besteht darin, dass es nicht genug von diesen Leuten gibt, um vorwärtszukommen, und wenn Sie jemanden finden, der einsteigen möchte, dann werden er oder sie sehr hohe Gehaltsforderungen stellen. Denken Sie daran: Zusätzlich zur Mathematik, Statistik und Programmierung müssen Datenwissenschaftler einen hohen Grad von Fachwissen in dem Gebiet haben, in dem sie tätig sind. Daher ist es außerordentlich schwer, solche Leute zu finden. Bis Universitäten Datenkenntnisse zum festen Bestandteil jeder Ausbildung machen, wird es nahezu unmöglich sein, hochspezialisierte und fähige Datenwissenschaftler zu finden, die die beruflichen Anforderungen erfüllen.

## ***Die Arbeit an private Beratungsinstitute für Data Science übertragen***

Viele Firmen bevorzugen es, die erforderlichen Arbeiten bezüglich Data Science und Datenanalyse an außenstehende Experten zu übergeben. Dazu gibt es zwei gebräuchliche Wege: Auslagern zur Entwicklung einer großangelegten Strategie, die Ihrem gesamten Unternehmen dient, oder Auslagern für einzelne datenwissenschaftliche Lösungen für bestimmte Probleme, die innerhalb Ihres Unternehmens entstehen oder entstanden sind.



### **Auslagerung zur Entwicklung einer großangelegten datenwissenschaftlichen Strategie**

Wenn Sie in Ihrem Unternehmen ein auf der Data Science beruhendes System einführen wollen, können Sie einen privaten Fachberater engagieren, der Ihnen bei der Entwicklung einer umfassenden Strategie hilft. Diese Art von Leistung wird Sie voraussichtlich etwas kosten, aber Sie erhalten im Gegenzug häufig wertvolle Einsichten. Ein Strategie wird die verfügbaren Möglichkeiten zur Erfüllung Ihrer Anforderungen sowie deren Vor- und Nachteile kennen. Mit einer Strategie an der Hand und einem Experten, der Ihnen für Beratung zur Verfügung steht, können Sie die Aufgabe zur Bildung eines internen Teams viel besser steuern.

### **Auslagerung der Lösung einzelner Probleme**

Wenn Sie nicht zu dem eher komplexen Prozess der Entwicklung und Umsetzung einer umfassenden Strategie bereit sind, haben Sie die Möglichkeit, einen privaten datenwissenschaftlichen Fachberater mit kleineren Arbeiten zu beauftragen. Diese punktuelle Methode könnte die Vorteile von Data Science bieten, ohne dass eine Umorganisation der Struktur und Finanzierung Ihres gesamten Unternehmens nötig ist.

### ***Auf Cloud-Dienste zurückgreifen***

Manche Leute haben die Explosion von Big Data und Data Science seit Langem kommen sehen. Obwohl es für die meisten neu ist, haben Fachkräfte und Unternehmen, die im Bilde waren, rasend schnell gearbeitet, um sich vorzubereiten. Einige Unternehmen haben große Anstrengungen und Kosten aufgewandt, um datenwissenschaftliche Lösungen zu entwickeln, die für alle zugänglich sind. Cloud-Dienste wie Watson Analytics von IBM (<http://www.ibm.com/analytics/watson-analytics/>) bieten Nutzern automatische Datenleistungen ohne Codierung, die von der Bereinigung von Daten über statistische Modellierungen bis zur Analyse und Visualisierung von Daten reichen. Obwohl Sie trotzdem noch die statistische, mathematische und substanzielle Relevanz der Daten-zentrierten Sicht verstehen müssen, können Anwendungen wie Watson Analytics einige leistungsstarke Ergebnisse liefern, ohne von den Nutzern zu verlangen, dass sie Programme oder Skripts schreiben können.



Wenn Sie sich entscheiden, auf Cloud-Dienste zurückzugreifen, um Ihrem Unternehmen zu helfen, die datenwissenschaftlichen Ziele zu erreichen, denken Sie daran, dass Sie ausgebildete und fähige Mitarbeiter im eigenen Haus benötigen, um die quantitativen Ergebnisse dieser Dienste zum Laufen zu bringen und zu interpretieren. Der Cloud-Dienst ersetzt nicht die Notwendigkeit ausgebildeter Mitarbeiter und datenwissenschaftlichen Fachwissens im eigenen Unternehmen. Er wird lediglich Ihr Unternehmen so erweitern, dass es seine Ziele schneller erreicht.

### ***Die offensichtlichen Gewinne ermitteln***

Ich hoffe, im Verlauf dieses Buches Ihnen die Leitungsfähigkeit der Data Science zeigen zu können sowie, wie Sie mit ihrer Hilfe Ihre persönlichen und beruflichen Ziele schneller erreichen. Unabhängig davon, in welchem Feld Sie arbeiten, der Erwerb von datenwissenschaftlichen Kenntnissen kann Sie in eine marktfähigere Fachkraft verwandeln. Die folgende

Liste enthält nur einen kleinen Teil des Nutzens, den die Data Science sowie die Analyse von Daten für wesentliche Industriebereiche bietet:

- ✓ **Nutzen für Konzerne, kleine und mittlere Unternehmen (englisch *small and medium-sized enterprises*, SMEs) und den elektronischen Geschäftsverkehr (E-Commerce):** Optimierung von Produktionskosten, Maximierung des Umsatzes, Verbesserung der Kapitalrenditen (englisch *Return on Investment*, ROI), Optimierung der Produktivität der Mitarbeiter, Vermeidung von Kundenabwanderungen, Erhöhung der Dauer der Kundenbeziehungen, Voraussagen über den Bestand und den Verkauf, Optimierung in der Preisermittlung, Betrugsaufdeckung und logistische Verbesserungen.
- ✓ **Nutzen für öffentliche Einrichtungen:** Optimierung der Geschäftsvorgänge und der Produktivität der Mitarbeiter, Verbesserung der Entscheidungshilfen im Management, Vorhersagen bezüglich der Finanzen, Rückverfolgung und Optimierung von Aufgaben, Betrugsaufdeckung.
- ✓ **Nutzen für die akademische Welt:** Verbesserung der Aufteilung von Ressourcen, Verbesserung des Abschneidens von Studierenden, Verringerung von Studienabbrüchen, Optimierung der Geschäftsvorgänge, bessere Vorhersagen bezüglich der Finanzen und Anwerbung von Geldern.