haufe akademie

Factsheet

# Das ABC der Data Science

Dein Einstieg in die Welt der Datenwissenschaft

Made with AI

## Management Summary

In unserer Welt existiert kaum eine Branche, die nicht von der digitalen Transformation betroffen ist. Seien es Sensoren in Produktionsmaschinen, Kundeninformationen in einer CRM-Software oder Einblicke in die Bedürfnisse der Mitarbeiter:innen: Unternehmen und Kund:innen produzieren jede Menge Daten, die sich bei fachlicher Kompetenz und verständlicher Aufbereitung gewinnbringend einsetzen lassen. Aber was müssen Verantwortliche wissen, um diese wertvollen Erkenntnisse zu bergen?
Steige mit uns ein in die Welt der Data Science und erfahre, wie Du Daten zu Gold machst.

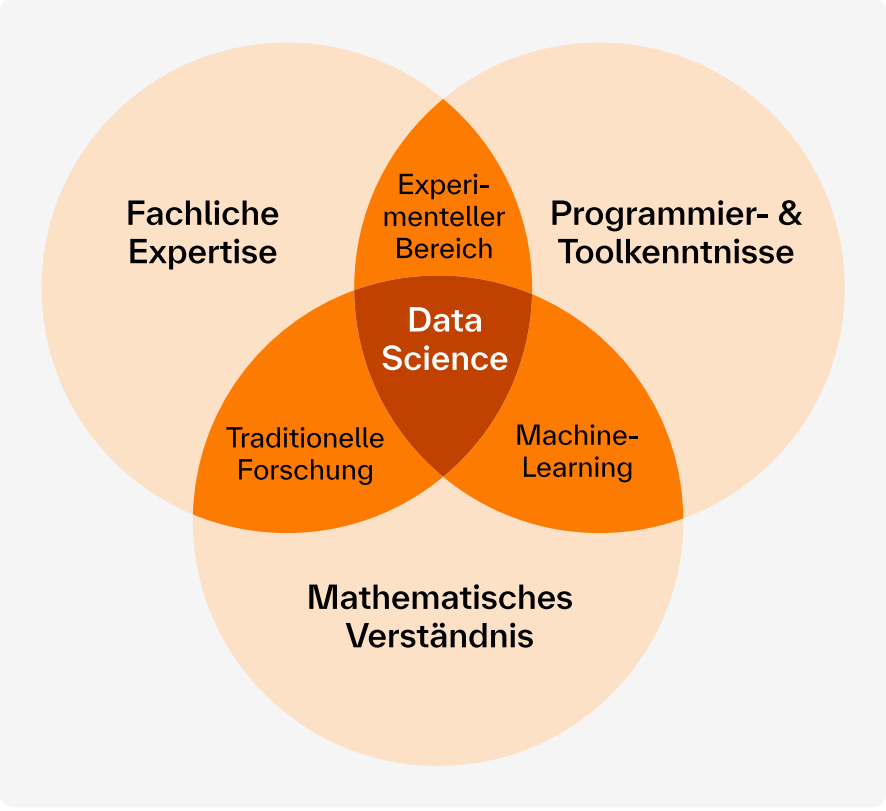
## Inhalt


| | |
|---|---|
| Vom ersten Byte bis zu Big Data: Die Entstehung der Data Science | 3 |
| Warum unsere Welt Data Science braucht | 4 |
| Die Data Science-Bedürfnispyramide | 5 |
| Die wichtigsten Begriffe der Data Science auf einen Blick | 6 |
| Tools und Programmiersprachen in der Data Science | 8 |
| Wie Unternehmen und Organisationen Datenpotenziale nutzen | 10 |
| Die Data Science-Trainings der Haufe Akademie | 11 |

# Vom ersten Byte bis zu Big Data: Die Entstehung der Data Science

**Die Geschichte der wissenschaftlichen Auseinandersetzung mit Daten und der Siegeszug der Computertechnologie sind logischerweise eng miteinander verknüpft.**

Während in den 1970er-Jahren Intel mit der Serienproduktion von Mikroprozessoren begann und Apple seinen ersten Heimcomputer auf den Markt brachte, prägten Wissenschaftler:innen und Professor:innen auf der ganzen Welt zeitgleich den Begriff der Data Science (zu Deutsch Datenwissenschaft). Was damals als synonymer Ausdruck für die Informatik stand, behauptete sich innerhalb der folgenden Jahre als eigene Disziplin. Spätestens seit den 2000er Jahren versteht sich die Data Science als eigenes Forschungsgebiet, das vor allem bei Unternehmen und Organisationen auf immer mehr Zuspruch trifft. Die Erkenntnisse aus großen Datenmengen zu nutzen, um Prozesse zu optimieren, Ressourcen zu schonen und Services zu verbessern: Dank großer Speicherkapazitäten und rasanter Rechenleistungen sind die Versprechungen der Data Science für viele Betriebe inzwischen nur noch einen Klick entfernt. Doch für eine erfolgreiche Data Science benötigt es mehr als nur ein Verständnis von reinen Daten, Statistik und Algorithmen. Unabdingbar ist vor allem die Expertise von Mitarbeiter:innen, in deren Fachgebiet Data Science eingesetzt werden soll. Denn nur diese Fachexpert:innen sind in der Lage, Ergebnisse von Data Science-Projekten sinnvoll zu interpretieren und gewinnbringend einzusetzen.




Die moderne Data Science ist eine Mischung aus fachlicher Expertise, Computer- und Programmierwissen sowie mathematischen und statistischen Inhalten.

# Warum unsere Welt Data Science braucht

**Tera- und Petabyte waren gestern. Untersuchen Sie heute den weltweiten industriellen Datenverkehr, ist der „Zettabyte“ die passende Datengröße.**

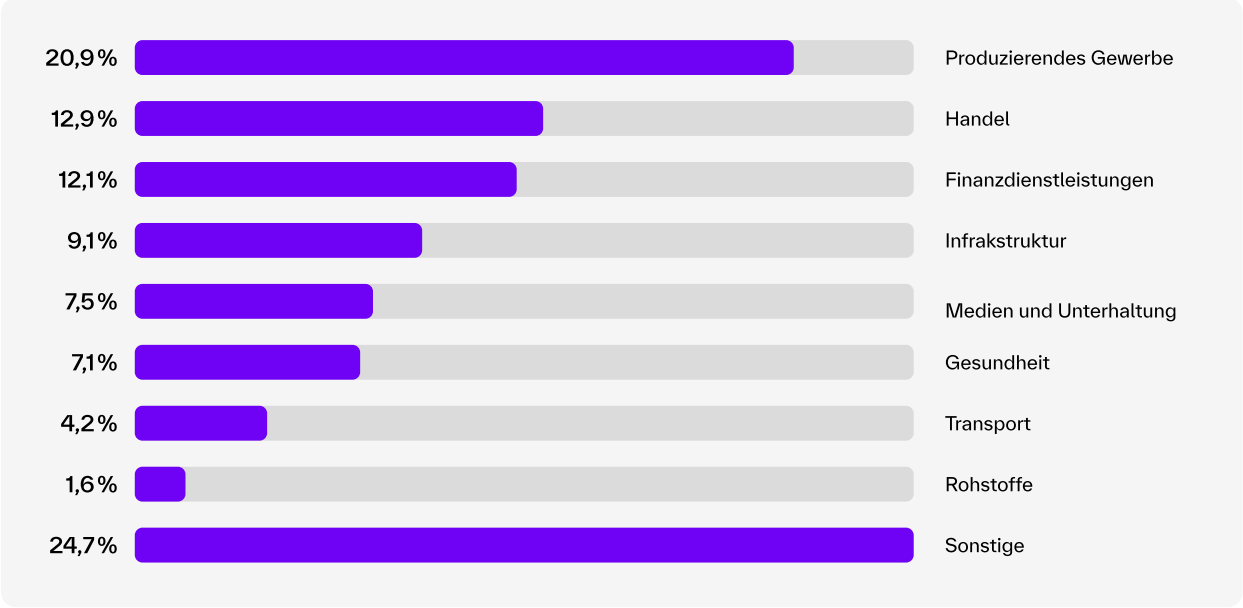
Während in den 1970er-Jahren Intel mit der Serienproduktion von Mikroprozessoren begann und Apple seinen ersten Heimcomputer auf den Markt brachte, prägten Wissenschaftler:innen und Professor:innen auf der ganzen Welt zeitgleich den Begriff der Data Science (zu Deutsch Datenwissenschaft). Was damals als synonymer Ausdruck für die Informatik stand, behauptete sich innerhalb der folgenden Jahre als eigene Disziplin.

Spätestens seit den 2000er Jahren versteht sich die Data Science als eigenes Forschungsgebiet, das vor allem bei Unternehmen und Organisationen auf immer mehr Zuspruch trifft. Die Erkenntnisse aus großen Datenmengen zu nutzen, um Prozesse zu optimieren, Ressourcen zu schonen und Services zu verbessern:

Dank großer Speicherkapazitäten und rasanter Rechenleistungen sind die Versprechungen der Data Science für viele Betriebe inzwischen nur noch einen Klick entfernt.

Doch für eine erfolgreiche Data Science benötigt es mehr als nur ein Verständnis von reinen Daten, Statistik und Algorithmen. Unabdingbar ist vor allem die Expertise von Mitarbeiter:innen, in deren Fachgebiet Data Science eingesetzt werden soll. Denn nur diese Fachexpert:innen sind in der Lage, Ergebnisse von Data Science-Projekten sinnvoll zu interpretieren und gewinnbringend einzusetzen.

## Welche Branchen produzieren weltweit die meisten Daten?



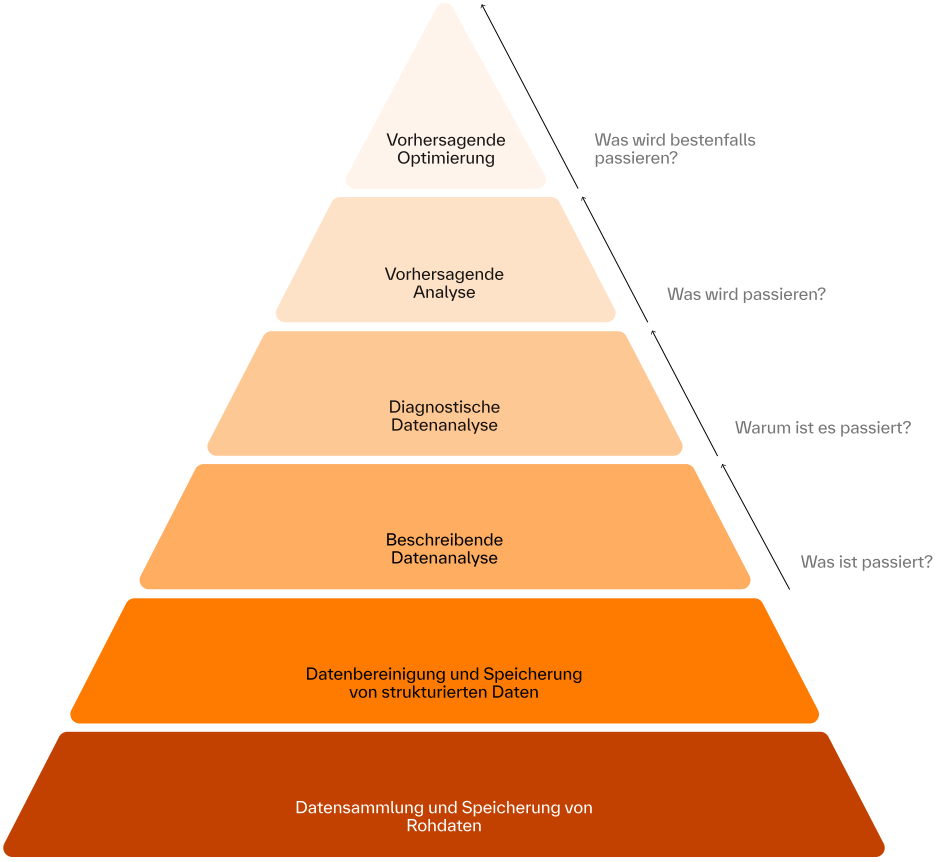

Ein Überblick über die Verteilung der weltweiten Datenmenge in einzelnen Branchen im Jahr 2018 prozentual verteilt.
Quelle: International Data Corporation, 2019 →

# Die Data Science-Bedürfnispyramide

**„Big Data“ zu analysieren und zu verstehen kann eine wunderbare und vor allem wertvolle Entscheidungshilfe sein – aber nur, wenn Verantwortliche die Informationen überblicken und einordnen können.**

Um den Rohstoff Daten gewinnbringend einzusetzen, braucht es Expert:innen, die sich mit dem gesamten Prozess der Datenverarbeitung auskennen. Für eine bessere Übersicht entwickelten Datenwissenschaftler:innen deshalb eine Bedürfnispyramide, mit welcher sich die Data Science in ihre einzelnen Bestandteile zerlegen lässt.

Je weiter oben sich ein Unternehmen oder eine Organisation auf der Data Science-Bedürfnispyramide einordnet, desto fortgeschrittener ist dort das Know-how über und die Nutzung von Data Science. Daher gilt es zunächst, sich als Unternehmen, Organisation oder Abteilung auf der Bedürfnispyramide einzuordnen: Wie werden Daten aktuell bei uns gespeichert, verarbeitet und analysiert, kurz: nutzbar gemacht? Welche Potenziale ergeben sich für uns, wenn wir durch mehr Data Science-Expertise und mehr Data Science-Projekte die Pyramide erklimmen?




Von der Sammlung über die Verarbeitung bis zur Nutzung: Die Bedürfnispyramide der Data Science verdeutlicht die komplexen Zusammenhänge.
Angelehnt an „The AI Hierarchy of Needs“ von Monica Rogati, 2017.

# Die wichtigsten Begriffe der Data Science auf einen Blick

**Wer sich zum ersten Mal mit Data Science beschäftigt, ist wahrscheinlich überwältigt von der immensen Anzahl an verschiedenen Ausdrücken. Mit der Verwirrung ist jetzt Schluss! Hier finden Sie eine Übersicht wichtiger Fachbegriffe der Data Science:**

## Künstliche Intelligenz

Dieser Begriff besitzt keine allgemeingültige Definition. Allgemein ist die künstliche Intelligenz ein Teilgebiet der Informatik, das sich mit der Automatisierung von intelligentem Verhalten befasst. Zu beachten ist hierbei, dass eine künstliche Intelligenz nur ein scheinbar intelligentes Verhalten zeigen kann. Künstliche Intelligenz wird heutzutage sehr weit gefasst und kann vieles umfassen: von klassischen, regelbasierten Algorithmen bis hin zu selbstlernenden künstlichen neuronalen Netzen.

## Machine Learning

Machine Learning (zu Deutsch Maschinelles Lernen) steht für den Prozess von Algorithmen, anhand von Beispielen Muster zu erkennen und Regeln abzuleiten. Die zwei Grundarten des maschinellen Lernens sind das überwachte Lernen und das unüberwachte Lernen. Der Unterschied besteht – grob gesagt – darin, dass beim überwachten Lernen nach Mustern gesucht wird, die ein bereits bekanntes Endergebnis erklären können. Beim unüberwachten Lernen gibt es ein solches bekanntes Endergebnis allerdings noch nicht.

## Algorithmus

Im Kern ist ein Algorithmus eine eindeutige Beschreibung, wie man mit einer bestimmten Folge von Anweisungen ein Problem in einer begrenzten Zeit löst. Im Bereich der Data Science werden mit Algorithmen kleine Programme beschrieben, die mit Eingabedaten „gefüttert“ werden und darauf basierend ein eindeutiges und konstantes Ergebnis liefern.

## Künstliches Neuronales Netz

Als Künstliches Neuronales Netz bezeichnet man in der Informatik eine Verknüpfung mathematischer Modelle, die dem Aufbau biologischer Gehirne nachempfunden sind. Auch sie bestehen aus (künstlichen) Neuronen, die einzelne Aufgaben übernehmen und miteinander verbunden sind.

## Deep Learning

Eng verknüpft mit den künstlichen neuronalen Netzen ist das Deep Learning (zu Deutsch tiefes Lernen), das noch einen Schritt weitergeht und besonders viele künstliche Neuronen in Schichten aneinanderreiht. So entsteht ein neuronales Netzwerk, das sich selbstständig weiterentwickelt und immer intelligentere Entscheidungen treffen kann.

## Predictive Analytics

Ob in der Wettervorhersage oder in der Medizin: Die Predictive Analytics (zu Deutsch vorausschauende Analyse) ermöglicht dank mathematischer Formeln, historische Daten auszuwerten und damit vorherzusagen, wie wahrscheinlich ein bestimmtes Ereignis in Zukunft eintreffen wird. Dabei ist natürlich zu beachten, dass die Zukunft immer nur auf Grundlage der Vergangenheit vorhergesagt werden kann. Ändern sich Umweltparameter, müssen auch Vorhersagemodelle angepasst werden.

## Data Warehouse

Mit einem Data Warehouse (zu Deutsch Datenwarenlager) bezeichnet man eine zentrale Datenbank, die Daten aus mehreren unterschiedlichen Quellen zur einfacheren Analyse an einem einzigen Speicherort zusammenführt.

# Tools und Programmiersprachen in der Data Science

**Was wäre ein:e Friseur:in ohne Schere oder ein:e Konditor:in ohne ein Nudelholz? Ohne ihre Werkzeuge hätten sie keine Chance, die vorliegenden Ressourcen so zu verarbeiten, dass andere Menschen davon einen Nutzen haben. So ähnlich ist das auch bei Data Scientists, die für die Organisation, Verarbeitung und Visualisierung von Daten eigene Programmiersprachen und Tools benötigen.**

Doch es gibt in der Data Science nicht die eine Programmiersprache oder Software-Lösung, mit der sich alle Aufgaben vollumfänglich lösen lassen. Manche Funktionen sind essenziell und in jeder Sprache hinterlegt, manche Tools können nur als Schnittstelle im Web-Browser genutzt werden, und wieder andere Programme lassen sich nur als lokale Software installieren. Innerhalb der vergangenen Jahre haben sich Datenwissenschaftler:innen auf eine Handvoll an Lösungen fokussiert. Eine Auswahl davon finden Sie auf der folgenden Seite.

## Python

Python ist die wohl etablierteste Programmiersprache in der Data Science und vor allem aufgrund des klaren Aufbaus besonders für noch unerfahrene Datenwissenschaftler:innen empfehlenswert. Python ist als Open Source Software kostenlos erhältlich. Die Sprache eignet sich für eine Vielzahl von Anwendungen und auch große Tech-Firmen wie Google oder Instagram nutzen Python.

## KNIME

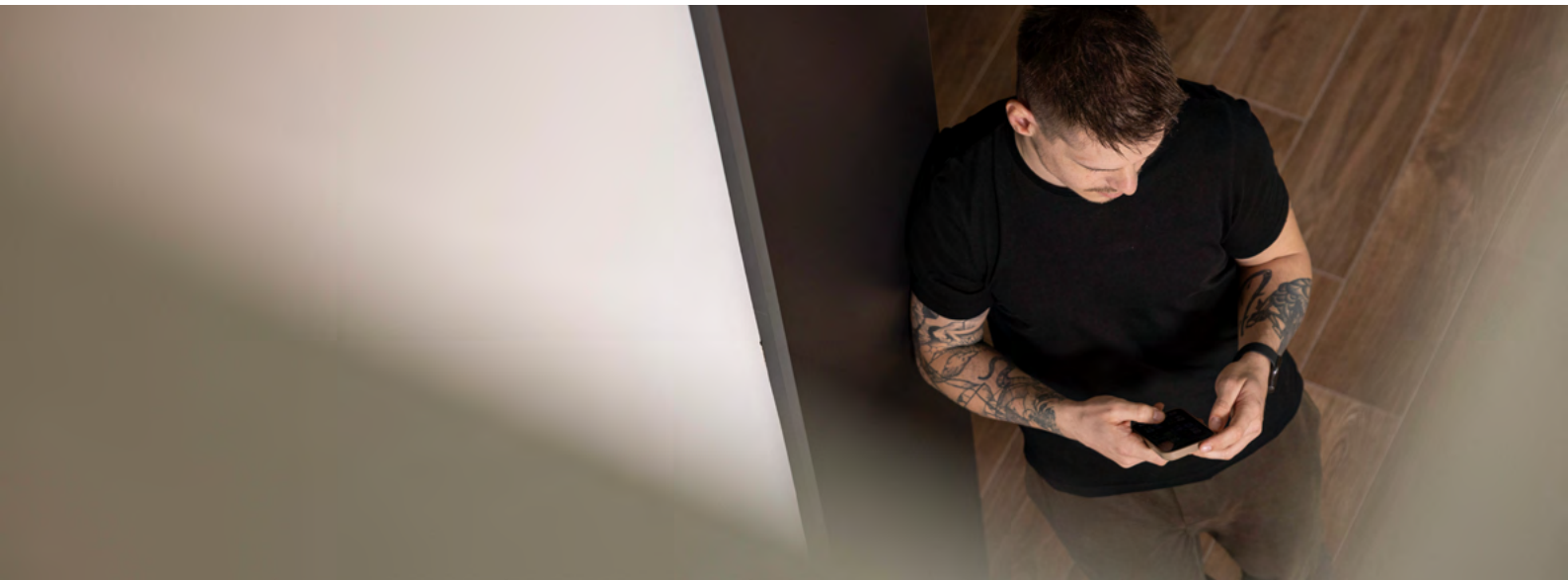
KNIME ist eine aus Deutschland stammende kostenlose Software, mit der Anwender:innen auch ohne Programmierkenntnisse komplexe Datenanalysen und Machine Learning-Prozesse umsetzen können. Die einzelnen Arbeitsschritte innerhalb einer Datenanalyse lassen sich grafisch als „Workflows“ aneinanderreihen und modular zusammensetzen. KNIME basiert auf der Programmiersprache Java und ist deshalb mit vielen anderen Analysesoftwares kompatibel.

## Tableau

Das Programm Tableau findet in der Data Science vor allem in der Datenanalyse und Visualisierung seinen Einsatz. Die Software lässt sich auch online als Plattform-Lösung nutzen. Per Drag and Drop können Nutzer Datensätze miteinander verbinden und in eigenen Dashboards darstellen.

## Excel

Auch mithilfe des Tabellenprogramms Excel lassen sich große Datensätze zusammenfassen. Und damit sind nicht nur Tortendiagramme und bedingte Zellenformatierungen gemeint, sondern auch Analysen komplexer Datenmengen mit zahlreichen Querverweisen sowie die Visualisierung von statistischen Verteilungen und Korrelationen.



# Wie Unternehmen & Organisationen Datenpotenziale nutzen

**Umsatzentwicklungen vorhersagen, Einsparpotenziale erkennen oder Personalplanungen zu Beginn eines neuen Geschäftsjahres aufstellen: Für Unternehmen sind Daten seit jeher eine wichtige Basis, um langfristige Entscheidungen zu treffen.**

In Zeiten von „Big Data“ verfügen Verantwortliche jedoch über weitaus mehr Informationen, die sich schnell abrufen lassen und aus unterschiedlichen Unternehmensbereichen stammen. Vier einfache Beispiele zeigen, wie sich Data Science gewinnbringend bzw. kostensparend auch bei Ihnen einsetzen lässt.

### Praxisbeispiel: Marketing
**Kundenbedürfnisse frühzeitig erkennen**

Jeder Mausklick im Online-Shop eines Klamottenhändlers führt zu Daten, die Auskunft über die Wünsche der Kund:innen geben. Diese Datenmenge lässt sich nutzen, indem man das Nutzerverhalten in einzelne Kundengruppen anhand ihrer Einkaufsgewohnheiten, der Verweildauer oder sogar sozio-demografischer Daten zusammenfasst. Diese Informationen können anschließend ins Marketing einfließen, um Kundenbedürfnisse in der Kommunikation besser zu berücksichtigen. Personalisierte Inhalte oder eine Vorhersage zukünftiger Käufe sind so nur wenige Klicks entfernt.

### Praxisbeispiel: Logistik
**Fahrtwege der eigenen Flotte optimieren**

Ein städtischer Müllbetrieb möchte die Fahrtwege seiner Müllwagen verbessern, damit die Fahrer:innen schneller ihre Ladefläche befüllen und wieder zurück in die Deponie fahren können. Er digitalisiert deshalb die Fahrtprotokolle und sammelt über einen längeren Zeitraum die Routen seiner Fahrer:innen. Die Daten geben dem Geschäftsführer eine gute Übersicht, welche Straßen ein besonders hohes Müllaufkommen haben und welche Kreuzungen stark befahren sind. So kann er die Routen seiner Müllwägen optimieren und in weniger Zeit mehr Abfälle einsammeln.

### Praxisbeispiel: Personalmanagement
**Bewerber:innen besser beurteilen**

Arbeitnehmer:innen durchlaufen in ihrer Karriere einige Phasen: Sie bewerben sich, feiern ihre Einstellung, bilden sich fort, werden befördert und scheiden irgendwann wieder aus dem Unternehmen aus. In all diesen Schritten produzieren sie Daten, die tief verborgen in der Personalakte versteckt sind. Nutzen HR-Spezialist:innen diese Informationen, erhalten sie mithilfe von Data Science viele nützliche Einschätzungen, die sie für zukünftige Bewerber:innen nutzen können – wie etwa die durchschnittlichen Investitionskosten, die Kosten pro Einstellung oder die Wahrscheinlichkeit einer Kündigung.

### Praxisbeispiel: Controlling
**Zahlungsläufe intelligent gestalten**

Die Geschäftsführung eines Unternehmens möchte den optimalen Zeitpunkt herausfinden, zu dem offene Verbindlichkeiten beglichen werden sollen. Im Fokus steht die Frage, ob Rechnungen zum vereinbarten Zahlungsdatum bezahlt werden sollen oder ob im Fall einer Skontogewährung eine vorzeitige Bezahlung lukrativer ist, um mögliche Rabatteffekte zu nutzen. Mithilfe von Data Science kann nicht nur das optimale Zahlungsdatum ermittelt werden, es kann auch das Ausführungdatum der Zahlungsläufe optimiert werden und sogar die Zahlungsbedingungen mit Lieferanten anhand einer KI analysiert und im Anschluss nachverhandelt werden.

# Unsere Trainings zum Thema Data Science

**Qualifiziere dich weiter und bringe Datenexpertise in dein Unternehmen! Ob du auf Augenhöhe mit Data Scientists kommunizieren willst oder ob du selbst in die Welt der Data Science eintauchen und programmieren lernen möchtest: wir bieten dir die passende Qualifizierung.**

**Master Class mit Zertifikat | Präsenz + Online oder nur Online**

## Geprüfter Data Analyst

Daten analysieren, Muster erkennen, Vorhersagen treffen

**Smart Data Science – Einstieg in die professionelle Datenanalyse**

Vom Basiswissen bis zum ersten professionellen Datenprojekt – mit allen Grundlagen rund um Big Data, Machine Learning und Datenanalyse mit dem CRISP-DM-Prozess.

**Data Analytics und Data Science praktisch umsetzen mit KNIME**

Lerne, wie du die ersten Schritte in der Datenanalyse meisterst, und setze komplette Datenprozesse mit Hilfe des KNIME-Tools um.

Jetzt informieren →

## Datenkompetenz – vom Einstieg bis zum Zertifikat

### Basic

**Data Awareness & AI Literacy: Die Online-Training-Bibliothek**

Die Jahreslizenz für fünf umfassende Online-Trainings zu den Themen Data Literacy, Data-Driven Management, Data Storytelling, AI Literacy und AI Driven Management

Online 34206 →

### Basic

**SQL und relationale Datenbanken**

Einfache und komplexe Abfragetechniken mit der wichtigsten Datenbanksprache – das Online-Training mit Abschlusszertifikat

Online 34205 →

### Basic

**Datenanalyse und Datenvisualisierung mit Excel und Power BI**

Daten kompetent einsetzen – der Praxiskurs für die Datenanalyse mit Bordmitteln

Online 34159 →

### Basic+

**Data Science in der Praxis**

Datenanalysen und maschinelles Lernen mit Python

Präsenz 34069 →
oder Online 36359 →

### Basic+

**Data Analyst mit Python**

Datenanalyse, Datenvisualisierung und Predictive Analytics mit Python – das Online-Training mit Abschlusszertifikat

Online 30353 →

### Basic+

**Python Basics**

Grundlagen der Automatisierung und erste Programmierpraxis für Ein- und Umsteiger – mit vielen interaktiven Übungsaufgaben

Online 30815 →

### Advanced

**Business Intelligence Analyst**

Professionelle Datenanalysen und Datenreports mit SQL und Power BI – das Online-Training mit Abschlusszertifikat

Online 34046 →

### Advanced

**Der Python-Programmierer**

Qualifizierung für die Softwareentwicklung mit Python – das Online-Training mit Abschlusszertifikat

Online 34203 →

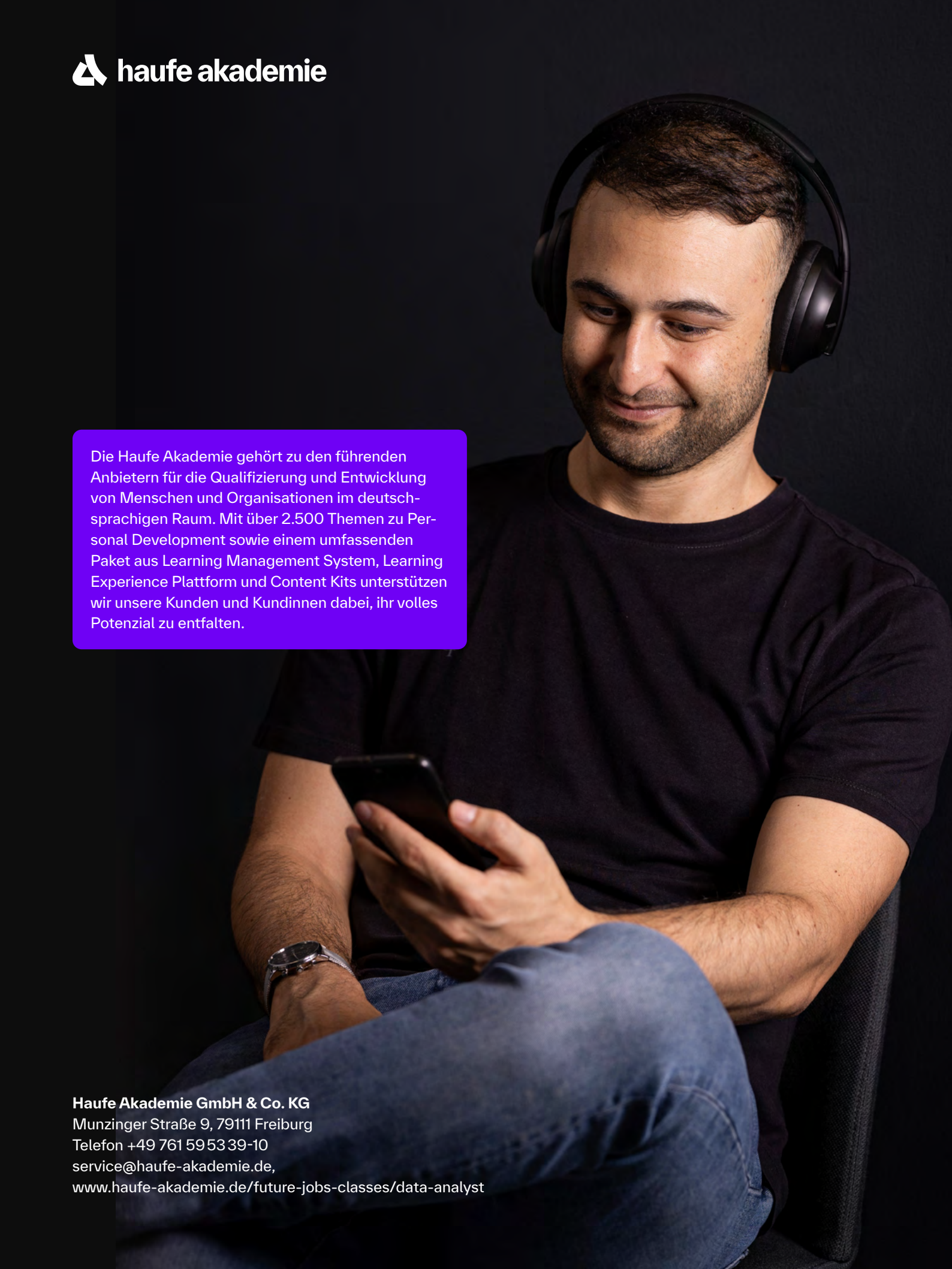
### Advanced

**Data Scientist mit Python**

Data-Pipelines mit Machine-Learning-Algorithmen und Python – das Online-Training mit Abschlusszertifikat

Online 30354 →

haufe akademie



Die Haufe Akademie gehört zu den führenden Anbietern für die Qualifizierung und Entwicklung von Menschen und Organisationen im deutschsprachigen Raum. Mit über 2.500 Themen zu Personal Development sowie einem umfassenden Paket aus Learning Management System, Learning Experience Plattform und Content Kits unterstützen wir unsere Kunden und Kundinnen dabei, ihr volles Potenzial zu entfalten.

**Haufe Akademie GmbH & Co. KG**
Munzinger Straße 9, 79111 Freiburg
Telefon +49 761 59 53 39-10
service@haufe-akademie.de,
www.haufe-akademie.de/future-jobs-classes/data-analyst