

Mit diesem Buch lernen KI- und Machine-Learning-Praktikerinnen und -Praktiker, wie sie erfolgreich Data-Science-Projekte auf Amazon Web Services (AWS) entwickeln und deployen können. Der KI- und Machine-Learning-Stack von Amazon gibt Ihnen die Möglichkeit, durch die Vereinigung von Data Science, Data Engineering und Anwendungsentwicklung Ihre Fähigkeiten zu steigern. Dieser Leitfaden zeigt Ihnen, wie Sie Pipelines in der Cloud erstellen und ausführen und anschließend die Ergebnisse innerhalb von Minuten statt Tagen in Anwendungen integrieren können. Chris Fregly und Antje Barth machen zudem im gesamten Buch deutlich, wie Sie Ihre Kosten senken und die Leistung verbessern können.

- Wenden Sie Amazons KI- und ML-Stack auf reale Anwendungsfälle aus den Bereichen der natürlichen Sprachverarbeitung (*Natural Language Processing*, NLP), der Bildverarbeitung (*Computer Vision*), der Erkennung von Betrug (*Fraud Detection*) sowie im Rahmen des Einsatzes intelligenter Kommunikationsgeräte (*Conversational Devices*) und mehr an.
- Greifen Sie auf automatisierte ML-Algorithmen (AutoML) zurück, um bestimmte Anwendungsfälle mit Amazon SageMaker Autopilot umzusetzen.
- Erhalten Sie einen tiefen Einblick in den gesamten Lebenszyklus der Modellentwicklung für einen BERT-basierten Anwendungsfall der natürlichen Sprachverarbeitung (NLP), einschließlich, neben vielem anderen, Datenaufnahme (engl. *Data Ingestion*) und -analyse.
- Bündeln Sie alles in eine wiederverwendbare MLOps-Pipeline (*ML Operations*).
- Erkunden Sie die Möglichkeiten des Einsatzes von ML in Echtzeit, Anomalieerkennung und Streaming-Analysen auf Basis von Echtzeitdatenströmen mit Amazon Kinesis und *Amazon Managed Streaming for Apache Kafka* (Amazon MSK).
- Lernen Sie bewährte Sicherheitspraktiken für Data-Science-Projekte und -Workflows kennen, einschließlich *AWS Identity and Access Management* (IAM), Authentifizierung, Autorisierung, darunter Datenaufnahme und -analyse, Modelltraining und Deployment.

# Die Kapitel im Überblick

Kapitel 1 bietet einen allgemeinen Überblick über den sehr umfang- und facettenreichen KI- und ML-Stack von Amazon, der ein enorm leistungsfähiges und vielfältiges Angebot an Diensten, Open-Source-Bibliotheken und Infrastrukturen bietet, die für Data-Science-Projekte jeder Komplexität und Größe genutzt werden können.

Kapitel 2 beschreibt, wie Amazons KI- und ML-Stack in realen Anwendungen aus den Bereichen Empfehlungssysteme, Computer Vision, Betrugserkennung, Verstehen natürlicher Sprache (*Natural Language Understanding*, NLU), Conversational Devices, Cognitive Search, Kundenbetreuung, vorausschauende Wartung (*Predictive Maintenance*) in der Industrie, Hausautomatisierung, Internet der Dinge (*Internet of Things*, IoT), aus dem Gesundheitswesen und auch dem Bereich Quantencomputing eingesetzt werden kann.

Kapitel 3 zeigt, wie Sie mit SageMaker Autopilot AutoML nutzen und einige dieser Anwendungsfälle implementieren können.

In den Kapiteln 4 bis 9 wird der komplette Lebenszyklus der Modellentwicklung (*Model Development Life Cycle*, MDLC) für einen BERT-basierten NLP-Anwendungsfall ausführlich vorgestellt. Die Vorstellung schließt die Datenaufnahme und -analyse, die Auswahl von Features (engl. *Feature Selection*) und das Feature Engineering, das Modelltraining sowie die Modellabstimmung und -bereitstellung mit Amazon SageMaker, Amazon Athena, Amazon Redshift, Amazon EMR, TensorFlow, PyTorch und serverloses Apache Spark mit ein.

In Kapitel 10 wird gezeigt, wie sich alle zuvor gezeigten Teilschritte mithilfe von MLOps auf Basis von SageMaker Pipelines, Kubeflow Pipelines, Apache Airflow, MLflow oder TFX in wiederverwendbaren Pipelines zusammenführen lassen.

Kapitel 11 gibt einen Einblick in den Themenkomplex rund um Echtzeit-ML, Anomalieerkennung und Streaming-Analysen für Echtzeitdatenströme unter Einsatz von Amazon Kinesis und Apache Kafka.

Kapitel 12 stellt eine Vielzahl von bewährten Sicherheitspraktiken für Data-Science-Projekte und -Workflows vor, darunter IAM, Authentifizierung, Autorisierung, Netzwerkisolierung, Verschlüsselung von Daten im Ruhezustand (*Data Encryption at Rest*), Post-Quanten-Netzwerkverschlüsselung bei der Übertragung von Daten, Governance und Auditierbarkeit.

Im gesamten Buch finden Sie zahlreiche Tipps dazu, wie Sie die Kosten senken und die Ergebnisse bzw. die Leistungsfähigkeit von Data-Science-Projekten auf AWS verbessern können.

## An wen sich dieses Buch richtet

Dieses Buch richtet sich an alle, die auf der Grundlage von Datenanalysen wichtige Geschäftsentscheidungen treffen. Der Leitfaden hilft Data Analysts, Data Scientists, Data Engineers, Machine Learning Engineers, Research Scientists, Anwendungsentwicklerinnen und -entwicklern sowie DevOps Engineers dabei, ihre Kenntnisse des modernen Data-Science-Stacks auszuweiten und ihre Fähigkeiten im Hinblick auf die Entwicklung in der Cloud zu verbessern.

Amazons KI- und Machine-Learning-Stack vereint die Disziplinen Data Science, Data Engineering und Anwendungsentwicklung und hilft Benutzerinnen und Benutzern, ihre Fähigkeiten über ihre gegenwärtigen Tätigkeiten hinaus zu erweitern. Wir zeigen, wie man Pipelines in der Cloud erstellt und ausführt und dann die Ergebnisse innerhalb von Minuten – nicht in Tagen – in Anwendungen integriert.

Um den größtmöglichen Nutzen aus diesem Buch zu ziehen, sollten Sie idealerweise über die folgenden Kenntnisse verfügen:

- ein grundlegendes Verständnis von Cloud Computing,
- grundlegende Programmierkenntnisse in Python, R, Java/Scala oder SQL sowie
- Grundkenntnisse im Umgang mit Data-Science-Tools wie Jupyter Notebook, Pandas, NumPy oder scikit-learn.

## Weitere Ressourcen

Es gibt eine Reihe großartiger Autorinnen und Autoren sowie Quellen, die uns im Hinblick auf dieses Buch inspiriert haben:

- Aurélien Géron's Buch *Praxiseinstieg Machine Learning mit Scikit-Learn, Keras und TensorFlow* (<https://oreilly.de/produkt/praxiseinstieg-machine-learning-mit-scikit-learn-keras-und-tensorflow/>) (O'Reilly, aktuell in 2. Auflage) ist ein hervorragender praktischer Leitfaden für den Aufbau intelligenter ML-Systeme mit gängigen Tools wie Python, scikit-learn und TensorFlow.
- *Deep Learning for Coders with fastai and PyTorch* (<https://www.oreilly.com/library/view/deep-learning-for/9781492045519/>) (O'Reilly) von Jeremy Howard und Sylvain Gugger bietet eine exzellente Einführung in die Erstellung von Deep-Learning-Anwendungen mit PyTorch – und zwar ohne dass ein Dokortitel vonnöten wäre, um dem Buch folgen zu können.
- Das Buch *Building Machine Learning Pipelines* (<https://www.oreilly.com/library/view/building-machine-learning/9781492053187/>) (O'Reilly) von Hannes Hapke und Catherine Nelson ist ein ausgezeichnetes und einfach zu lesendes Nachschlagewerk zum Aufbau von AutoML-Pipelines mit TensorFlow und TFX.
- Das Buch *Programming Quantum Computers* (<https://www.oreilly.com/library/view/programming-quantum-computers/9781492039679/>) (O'Reilly) von Eric R. Johnston, Nic Harrigan und Mercedes Gimeno-Segovia ist eine hervorra-

gende Einführung in Quantencomputer mit leicht verständlichen Beispielen, die den Nutzen von Quanten aufzeigen.

- Micha Gorelick und Ian Ozsvald haben ein Buch für Fortgeschrittene namens *High Performance Python* (<https://www.oreilly.com/library/view/high-performance-python/9781492055013/>) (O'Reilly) verfasst, das viele wertvolle Tipps und Tricks zum Profilen und Optimieren von Python-Code enthält, insbesondere im Hinblick auf eine hochleistungsfähige Datenverarbeitung, das Feature Engineering und das Modelltraining.

Zusätzlich zum Buch haben wir Ihnen eine Webseite (*Data Science on AWS*, <https://datascienceonaws.com>) bereitgestellt, die Workshops für Fortgeschrittene, monatliche Webinare, Meet-ups, Videos und Folien zu den Inhalten dieses Buchs bietet.

Außerdem teilen wir regelmäßig relevante Blogbeiträge, Konferenzvorträge, Folien, Termine für Meet-ups und Workshops auf Twitter oder LinkedIn:

- Folgen Sie uns auf Twitter: <https://twitter.com/cfregly> und <https://twitter.com/anbarth>
- Auf LinkedIn finden Sie uns ebenfalls: <https://www.linkedin.com/in/cfregly> und <https://www.linkedin.com/in/antje-barth>

## In diesem Buch verwendete Konventionen

Die folgenden typografischen Konventionen werden in diesem Buch eingesetzt:

### *Kursiv*

Kennzeichnet neue Begriffe, URLs, E-Mail-Adressen, Dateinamen und Dateierendungen.

### Konstante Zeichenbreite

Wird für Programmlistings und Programmelemente in Textabschnitten wie Variablen- oder Funktionsnamen, Datenbanken, Datentypen, Umgebungsvariablen, Anweisungen und Schlüsselwörter verwendet.

### **Konstante Zeichenbreite, fett**

Kennzeichnet Befehle oder anderen Text, der vom Benutzer wörtlich eingegeben werden sollte.



Dieses Symbol steht für einen Tipp oder eine Empfehlung.



Dieses Symbol steht für einen allgemeinen Hinweis.

## Verwenden von Codebeispielen

Ergänzendes Material (Codebeispiele, Übungen usw.) steht in dem Repository *oreilly\_book* unter <https://github.com/data-science-on-aws> zum Download bereit. Einige der in diesem Buch gezeigten Codebeispiele sind gekürzt, um eine bestimmte Implementierung hervorzuheben. Das Repository enthält zusätzliche Notebooks, die nicht in diesem Buch behandelt werden, aber für den Leser nützlich sind. Die Notebooks sind jeweils nach den einzelnen Kapiteln des Buchs gegliedert und sollten leicht nachvollziehbar sein. Sie können die Zellen in den einzelnen Notebooks auf bequeme Weise ausführen, indem Sie die Umschalt- und die Eingabetaste gleichzeitig drücken.

Im Rahmen der Ersteinrichtung folgen Sie bitte den Anweisungen bzw. Screenshots in der *README.md*-Datei (unter *Instructions*) sowie den durchnummerierten Jupyter-Notebooks im Ordner *01\_introduction*, um

- zunächst ein AWS-Konto zu erstellen (falls noch nicht vorhanden),
- sich in Ihrer AWS Management Console anzumelden,
- SageMaker Studio einzurichten,
- zum Dashboard des IAM-Services zu navigieren,
- als Rolle *SageMakerExecutionRole* auszuwählen,
- die IAM-Richtlinie (*IAM Policy*) *AdministratorAccess* zuzuordnen, um die Zugriffsberechtigungen festzulegen,
- das GitHub-Repository zu klonen und
- die in diesem Buch verwendeten Bibliotheken zu installieren.

Beachten Sie, dass die hier verwendeten Berechtigungen nur Demonstrationszwecken dienen. Folgen Sie ansonsten bei der Vergabe von Zugriffsrechten in Ihrer Umgebung bitte stets dem Least-Privilege-Prinzip.

Dieses Buch dient dazu, Ihnen beim Erledigen Ihrer Arbeit zu helfen. Im Allgemeinen dürfen Sie die Codebeispiele aus diesem Buch in Ihren eigenen Programmen und der dazugehörigen Dokumentation verwenden. Sie müssen uns dazu nicht um Erlaubnis bitten, solange Sie nicht einen beträchtlichen Teil des Codes reproduzieren. Beispielsweise benötigen Sie keine Erlaubnis, um ein Programm zu schreiben, in dem mehrere Codefragmente aus diesem Buch vorkommen. Wollen Sie dagegen ein Produkt mit Beispielen aus Büchern von O'Reilly verkaufen oder verteilen, benötigen Sie eine Erlaubnis. Eine Frage zu beantworten, indem Sie aus diesem Buch zitieren und ein Codebeispiel wiedergeben, benötigt keine Erlaubnis. Eine beträchtliche Menge Beispielcode aus diesem Buch in die Dokumentation Ihres Produkts aufzunehmen, bedarf hingegen unserer ausdrücklichen Zustimmung.

Wir freuen uns über Zitate, verlangen diese aber nicht. Ein Zitat enthält Titel, Autor, Verlag und ISBN, zum Beispiel: »*Data Science mit AWS* von Chris Fregly und Antje Barth (O'Reilly). Copyright 2022 dpunkt.verlag, ISBN 978-3-96009-184-4.«

Wenn Sie glauben, dass Ihr Einsatz von Codebeispielen über die übliche Nutzung hinausgeht oder außerhalb der oben vorgestellten Nutzungsbedingungen liegt, kontaktieren Sie uns bitte unter [komentar@oreilly.de](mailto:komentar@oreilly.de).

## Danksagungen

Wir möchten uns bei Gary O'Brien, Development Editor bei O'Reilly, bedanken, der uns bei der Erstellung des Buchs geholfen hat und, was noch wichtiger ist, uns jedes Mal ein Lächeln ins Gesicht gezaubert hat, wenn wir uns unterhalten haben. Danke, Gary, dass wir den Quellcode und die Low-Level-Hardwarespezifikationen in Kapitel 1 aufnehmen durften! Wir möchten uns auch bei Jessica Haberman, Senior Acquisition Editor, bedanken, die uns wichtige Ratschläge zu allen Aspekten vom ersten Buchvorschlag bis zur endgültigen Seitenzahl gegeben hat. Mit ihrer Hilfe konnten wir nach sieben Jahren der Einreichung von Buchvorschlägen die Messlatte so hoch legen, dass der Vorschlag angenommen wurde! Ein besonderer Dank geht an Mike Loukides und Nicole Taché von O'Reilly für ihre wohlüberlegten Empfehlungen zu Beginn des Schreibprozesses, zur Gliederung der Kapitel, zu den Einleitungen und zu den Zusammenfassungen.

Wir möchten uns ganz herzlich bei den Fachgutachtern bedanken, die unermüdlich jede Seite dieses Buchs begutachtet – und erneut begutachtet – haben. Diese Reviewer sind hier in alphabetischer Reihenfolge nach Vornamen aufgeführt: Ali Arsanjani, Andy Petrella, Brent Rabowsky, Dean Wampler, Francesco Mosconi, Hannah Marlowe, Hannes Hapke, Josh Patterson, Josh Wills, Liam Morrison, Noah Gift, Ramine Tinati, Robert Monarch, Roy Ben-Alta, Rustem Feyzkhanov, Sean Owen, Shelbee Eigenbrode, Sireesha Muppala, Stefan Natu, Ted Dunning und Tim O'Brien. Ihr umfassendes technisches Fachwissen und ihr fundiertes Feedback waren nicht nur für dieses Buch von unschätzbarem Wert, sondern auch für die Art und Weise, wie wir in Zukunft technisches Material präsentieren werden. Sie haben dazu beigetragen, dass dieses Buch nicht nur gut, sondern großartig geworden ist, und es hat uns wirklich Spaß gemacht, mit allen diesen Menschen an diesem Projekt gearbeitet zu haben.

## Chris

Ich möchte dieses Buch meinem verstorbenen Vater, Thomas Fregly, widmen. Dad, du hast meinen ersten Apple-Computer mit nach Hause gebracht, als ich acht Jahre alt war, und damit mein Leben für immer verändert. Du hast mir im Alter von zehn Jahren geholfen, dein Buch über Infinitesimalrechnung zu verinnerlichen, und mein starkes Interesse an der Mathematik weiter gefestigt. Du hast mir beigebracht, fleißig zu lesen, kurz und bündig zu schreiben, effektiv zu sprechen, schnell zu tippen und frühzeitig Fragen zu stellen. Dadurch, dass ich dir bei der Reparatur eines Bootsmotors zusah, als du auf dem Michigansee gestrandet warst, wurde ich immer wieder dazu inspiriert, tiefer einzutauchen und die Hardware zu verstehen, die meine Software zum Laufen bringt. Bei einem Rundgang durch dein

Büro bei der *Chicago Sun-Times* lernte ich, dass jeder eine interessante Geschichte zu erzählen hat, auch der Rezeptionist, der Geschäftsführer und das Wartungspersonal. Du hast alle gleichermaßen begrüßt, dich nach ihren Kindern erkundigt, ihnen zugehört und sie mit einer eigenen lustigen Geschichte zum Lachen gebracht. Als ich als Kind an deiner Hand über deinen Universitätscampus lief, lernte ich, dass es in Ordnung ist, den Bürgersteig zu verlassen und mir meinen eigenen Weg durch das Gras zu bahnen. Du sagtest: »Keine Sorge, Chris, sie werden diesen Weg irgendwann pflastern, denn es ist eindeutig der kürzeste Weg vom Ingenieurbauwerk zur Cafeteria.« Du hattest recht, Dad. Viele Jahre später gingen wir diesen neu gepflasterten Weg, um uns in der Cafeteria dein Lieblingsgetränk, Pepsi Light, zu holen. Von dir habe ich gelernt, meinen eigenen Weg im Leben zu gehen und nicht immer der Masse zu folgen. Auch wenn du Windows 95 nicht mehr erlebt hast, hast du, ehrlich gesagt, nicht viel verpasst. Und ja, Mac OS ist schließlich zu Linux gewechselt. Auch da hattest du recht.

Ich möchte auch meiner Co-Autorin, Antje Barth, dafür danken, dass sie viele Nächte und Wochenenden investiert hat, um das Schreiben dieses Buchs zu einem fantastischen Erlebnis zu machen. Trotz des Zeitunterschieds von acht bis neun Stunden zwischen San Francisco und Düsseldorf hast du dir für virtuelle Whiteboard-Sitzungen, kurzfristige Quellcodeverbesserungen und Diskussionen im Hinblick auf die Verwendung des Oxford-Kommata immer Zeit genommen. Durch diese Erfahrung sind wir noch bessere Freunde geworden, und ohne dich hätte ich ein so gehaltvolles und hochwertiges Buch nicht erstellen können. Ich freue mich auf die Zusammenarbeit mit dir im Rahmen von zahlreichen zukünftigen Projekten!


## Antje

Ich möchte Ted Dunning und Ellen Friedman dafür danken, dass sie mir stets als großartige Mentoren zur Seite stehen und mich immer wieder dazu ermutigen, neue Herausforderungen in Angriff zu nehmen. Ted, wenn wir uns unterhalten, hast du immer weise Worte parat, die mir helfen, die Dinge aus einer anderen Perspektive zu sehen – sei es bei der Vorbereitung auf einen Demowettbewerb oder bei Gesprächen darüber, wie wir unseren Lesern helfen können, das Beste aus diesem Buch herauszuholen. Ellen, ich erinnere mich noch gut daran, wie du mir geholfen hast, überzeugende Vorschläge für Konferenzvorträge zu erstellen, als ich damit anfang, Vorträge für O'Reilly-Strata- und AI-Konferenzen einzureichen. Und bis zum heutigen Tag mache ich mir, wenn es darum geht, einprägsame Titel zu finden, besonders viele Gedanken. Leider kam O'Reilly meinen Vorschlag, als Titel dieses Buchs *Alexa, bitte trainiere mein Modell* zu nehmen, nicht nach.

Ihr beide geht mit gutem Beispiel voran, wenn ihr sagt: »Unterstützt den Traum eines Mädchens, das zu erreichen, was sie erreichen kann.« Aus demselben Grund möchte ich dieses Buch allen Frauen und Mädchen widmen, die von einer Karriere in der IT-Branche träumen oder diese anstreben. Solange ihr an euch selbst glaubt, steht euch nichts im Wege, eure Träume in diesem Berufsfeld zu verwirklichen.

Es gab noch so viele weitere Personen, die mich auf meinem beruflichen Weg unterstützt und ermutigt haben. Ich danke euch allen.

Ich möchte auch Chris dafür danken, dass er ein unterhaltsamer und kompetenter Mitverfasser war. Von Anfang an hast du immer auf die höchsten Standards gepocht, mich dazu gebracht, die Themen zu vertiefen, und mich ermutigt, neugierig zu sein und viele Fragen zu stellen. Du hast mir geholfen, meinen Code zu vereinfachen, meine Gedanken klar darzustellen und endlich das heiß diskutierte Oxford-Komma zu akzeptieren!

Diese Leseprobe haben Sie beim  
 [edv-buchversand.de](https://www.edv-buchversand.de) heruntergeladen.  
Das Buch können Sie online in unserem  
Shop bestellen.

[Hier zum Shop](#)