



Chris Fregly, Antje Barth

## Data Science mit AWS

**End-to-End-Pipelines für  
Continuous Machine Learning  
implementieren**

O'Reilly (Heise-Tochter) 2022

550 Seiten

Taschenbuch 52,90 €

E-Book 47,99 €

Bundle 57,90 €

■ Antje Barth und Chris Fregly zeigen mit viel Liebe zum Detail den AWS-KI- und -ML-Stack. Sie stellen dabei kundenspezifisch gestaltete Machine-Learning-Modelle vor und führen in die benötigten Tools ein. Beispielhaft zeigen sie, wie eine Integration der Jupyter-Notebooks durch das AWS-SDK für Python funktionieren kann, um auf AWS-Ressourcen zuzugreifen, wie dabei der Datenbankclient PyAthena hilft und der AWS Data Wrangler den Datentransfer zwischen pandas und Amazon S3, Athena, Redshift, Glue und EMR vereinfacht. So lassen sich ad hoc komplexe SQL-Abfragen mit mehreren Joins und Teilabfragen über

viele Datenquellen hinweg auf Daten in S3 durchführen. Neben den vorgestellten Tools steht der Lebenszyklus der Modellentwicklung im Vordergrund. Als typisches Beispiel nutzen die Autoren Amazons Bewertungssystem, dessen Datenaufnahme und -analyse, die Auswahl von Features, das Modelltraining, die Abstimmung sowie die Bereitstellung. Es folgt die Automatisierung mittels Pipelines.

Ein Kapitel zu Echtzeit-ML und Streaminganalyse von Datenströmen, um mittels Kinesis-Streaming Anomalien in Echtzeit erkennen zu können, rundet den Teil zu Data Science ab.

Das letzte Kapitel widmet sich der Sicherheit von AWS-Projekten. Ein wichtiger Baustein ist die geteilte Verantwortung, der Zugriff nach dem Least-Privilege-Prinzip. Das meint in diesem Fall, AWS kümmert sich um den Schutz der AWS-Cloud-Infrastruktur und -services, der Kunde hat die Verantwortung für die Sicherheit innerhalb der Cloud. Dabei helfen Dienste wie AWS Identity and Access Management (IAM). Die Autoren bringen dazu neben vielen Quelltexten Ausführungen zur Virtual Private Cloud, zum Schutz von S3-Zugriffen, zu Zugriffseinschränkungen über S3-Access-Points und zum weiten Feld der Verschlüsselung.

Das Buch ist reichlich bebildert und mit viel Quellcode angereichert. Allerdings sei angemerkt, dass es dem Werk zuträglich wäre, an der einen oder anderen Stelle einen aussagefähigen Screenshot zu zeigen. Das würde mehr Klarheit schaffen, wenn Software vorgestellt wird, und ist zielführender, als das GUI in Prosaform zu beschreiben. *Karsten Kisser (js@ix.de)*