

<b>Vorwort</b> .....	<b>11</b>
<b>1 Data Science mit AWS – eine Einführung</b> .....	<b>19</b>
Vorzüge des Cloud Computing .....	19
Data-Science-Pipelines und -Workflows .....	22
Best Practices für MLOps .....	26
Amazons KI-Services und AutoML mit Amazon SageMaker .....	30
Datenaufnahme, -exploration und -aufbereitung in AWS .....	33
Modelle mit Amazon SageMaker trainieren und feintunen .....	39
Modelle mit Amazon SageMaker und AWS Lambda Functions deployen .....	42
Streaming-Analysen und Machine Learning mit AWS .....	43
AWS-Infrastruktur und individuell zusammengestellte Hardware .....	45
Kosten mit Tags, Budgets und Alerts einsparen .....	49
Zusammenfassung .....	49
<b>2 Anwendungsbeispiele aus dem Bereich Data Science</b> .....	<b>51</b>
Innovationen in allen Branchen .....	51
Personalisierte Produktempfehlungen .....	52
Unangemessene Videos mit Amazon Rekognition erkennen .....	58
Bedarfsprognose .....	60
Betrügerische Benutzerkonten mit Amazon Fraud Detector identifizieren .....	64
Datenschutzlücken mit Amazon Macie erkennen .....	66
Conversational Devices und Sprachassistenten .....	67
Textanalyse und Natural Language Processing .....	68
Cognitive Search und Natural Language Understanding .....	73
Intelligente Kundenbetreuungszentren .....	75
Industrielle KI-Services und vorausschauende Wartung (Predictive Maintenance) .....	75

Heimautomatisierung mit AWS IoT und Amazon SageMaker. . . . .	76
Medizinische Informationen aus Gesundheitsdokumenten auslesen. . . . .	78
Selbstoptimierende und intelligente Cloud-Infrastruktur. . . . .	79
Kognitive und prädiktive Business Intelligence (BI) . . . . .	80
Die nächste Generation von KI- und ML-Entwicklern ausbilden. . . . .	84
Mithilfe von Quantencomputern das Betriebssystem der Natur programmieren . . . . .	89
Kosten einsparen und die Leistung verbessern . . . . .	94
Zusammenfassung . . . . .	97
<b>3 Automatisiertes Machine Learning . . . . .</b>	<b>99</b>
Automatisiertes Machine Learning mit SageMaker Autopilot . . . . .	100
Experimente mit SageMaker Autopilot tracken . . . . .	102
Einen Textklassifikator mit SageMaker Autopilot trainieren und deployen . . . . .	103
Automatisiertes Machine Learning mit Amazon Comprehend . . . . .	116
Zusammenfassung . . . . .	120
<b>4 Datenaufnahme in die Cloud . . . . .</b>	<b>123</b>
Data Lakes . . . . .	124
Amazon-S3-basierte Data Lakes mit Amazon Athena abfragen . . . . .	131
Mit dem AWS Glue Crawler kontinuierlich neue Daten aufnehmen. . . . .	137
Mit Amazon Redshift Spectrum ein Lake House aufbauen . . . . .	138
Zwischen Amazon Athena und Amazon Redshift wählen . . . . .	146
Kosten einsparen und die Leistung verbessern . . . . .	147
Zusammenfassung . . . . .	155
<b>5 Exploration des Datensatzes . . . . .</b>	<b>157</b>
Tools für die explorative Datenanalyse in AWS . . . . .	158
Mit SageMaker Studio Daten aus dem Data Lake visualisieren . . . . .	159
Abfragen auf unserem Data Warehouse durchführen . . . . .	172
Dashboards mit Amazon QuickSight erstellen. . . . .	180
Probleme im Hinblick auf die Datenqualität mithilfe von Amazon SageMaker und Apache Spark erkennen . . . . .	181
Bias in unserem Datensatz erkennen . . . . .	188
Verschiedene Arten von Drift mit SageMaker Clarify erkennen. . . . .	196
Unsere Daten mit AWS Glue DataBrew analysieren . . . . .	198
Kosten einsparen und die Leistung verbessern . . . . .	200
Zusammenfassung . . . . .	203

<b>6</b>	<b>Vorbereitung des Datensatzes für das Modelltraining</b>	<b>205</b>
	Feature Selection und Feature Engineering	205
	Das Feature Engineering mithilfe von SageMaker Processing Jobs skalieren	220
	Features über den SageMaker Feature Store gemeinsam nutzen	227
	Daten mit SageMaker Data Wrangler einlesen und transformieren	231
	Artefakt- und Experiment-Lineage mit Amazon SageMaker tracken	233
	Daten mit AWS Glue DataBrew aufnehmen und transformieren	238
	Zusammenfassung	240
<b>7</b>	<b>Das erste Modell trainieren</b>	<b>241</b>
	Die Infrastruktur von SageMaker verstehen	241
	Ein vortrainiertes BERT-Modell mit SageMaker JumpStart deployen	246
	Modelle in SageMaker entwickeln	248
	Ein kurzer Überblick über die historische Entwicklung des Natural Language Processing	251
	Die Transformer-Architektur von BERT	253
	BERT von Grund auf trainieren	256
	Feintuning eines vortrainierten BERT-Modells	258
	Das Trainingskript erstellen	261
	Das Trainingskript aus einem SageMaker-Notebook ausführen	268
	Modelle evaluieren	275
	Debugging und Profiling des Modelltrainings mit SageMaker Debugger	281
	Modellvorhersagen interpretieren und erklären	286
	Bias in Modellen erkennen und Vorhersagen erklären	291
	Weitere Möglichkeiten im Rahmen des Trainings von BERT	296
	Kosten einsparen und die Leistung verbessern	305
	Zusammenfassung	312
<b>8</b>	<b>Modelle in großem Maßstab trainieren und optimieren</b>	<b>313</b>
	Automatisch nach den besten Hyperparametern von Modellen suchen	313
	Einen Warmstart für zusätzliche SageMaker-HPT-Jobs verwenden	321
	Das Training mit SageMaker Distributed Training verteilen und skalieren	325
	Kosten einsparen und die Leistung verbessern	332
	Zusammenfassung	336

<b>9</b>	<b>Deployment von Modellen in die Produktion</b> .....	<b>339</b>
	Zwischen Vorhersagen in Echtzeit oder Batch-Vorhersagen wählen ...	339
	Echtzeitvorhersagen mit SageMaker Endpoints .....	341
	SageMaker Endpoints automatisch mit Amazon CloudWatch skalieren .....	349
	Strategien für das Deployment neuer oder aktualisierter Modelle .....	354
	Neue Modelle testen und vergleichen. ....	358
	Monitoring der Modellleistung und Drift erkennen .....	371
	Die Qualität der ein- und ausgehenden Daten der im Einsatz befindlichen SageMaker Endpoints überwachen .....	374
	Monitoring der Modellqualität von im Einsatz befindlichen SageMaker Endpoints .....	380
	Monitoring der Bias-Drift von im Einsatz befindlichen SageMaker Endpoints .....	385
	Monitoring der Drift der Feature Attribution von im Einsatz befindlichen SageMaker Endpoints .....	388
	Batch-Vorhersagen mit SageMaker Batch Transform durchführen .....	391
	AWS Lambda Functions und Amazon API Gateway .....	397
	Modelle auf Edge-Geräten optimieren und verwalten .....	398
	PyTorch-Modelle mit TorchServe deployen .....	398
	Inferenz für TensorFlow-basierte BERT-Modelle mit der AWS-Deep-Java-Bibliothek .....	400
	Kosten einsparen und die Leistung verbessern .....	402
	Zusammenfassung .....	408
<b>10</b>	<b>Pipelines und MLOps</b> .....	<b>409</b>
	Machine Learning Operations (MLOps) .....	409
	Software-Pipelines .....	411
	Machine-Learning-Pipelines .....	411
	Pipelines mit SageMaker Pipelines orchestrieren .....	416
	Pipelines mit SageMaker Pipelines automatisieren .....	427
	Weitere Optionen für Pipelines .....	432
	Human-in-the-Loop-Workflows .....	442
	Kosten einsparen und die Leistung verbessern .....	448
	Zusammenfassung .....	449
<b>11</b>	<b>Streaming-Analysen und Machine Learning</b> .....	<b>451</b>
	Unterschiede zwischen Online Learning und Offline Learning .....	452
	Streaming-Anwendungen .....	453
	Windowed Queries für Streaming-Daten .....	454
	Streaming-Analysen und Machine Learning mit AWS .....	458

Produktrezensionen in Echtzeit mit Amazon Kinesis, AWS Lambda und Amazon SageMaker klassifizieren . . . . .	459
Streaming-Daten mit Amazon Kinesis Data Firehose aufnehmen. . . . .	460
Zusammenfassende Metriken für Produktrezensionen mithilfe von Streaming-Analysen in Echtzeit ermitteln. . . . .	465
Amazon Kinesis Data Analytics einrichten . . . . .	466
Amazon-Kinesis-Data-Analytics-Anwendungen . . . . .	475
Produktrezensionen mit Apache Kafka, AWS Lambda und Amazon SageMaker klassifizieren . . . . .	482
Kosten einsparen und die Leistung verbessern . . . . .	483
Zusammenfassung. . . . .	485
<b>12 Sicherheit von Data-Science-Projekten auf AWS . . . . .</b>	<b>487</b>
Modell der geteilten Verantwortung zwischen AWS und seinen Kunden . . . . .	487
AWS Identity and Access Management (IAM) anwenden . . . . .	489
Rechen- und Netzwerkumgebungen isolieren. . . . .	497
Zugriff auf Daten von Amazon S3 schützen . . . . .	500
Verschlüsselung im Ruhezustand (Encryption at Rest). . . . .	509
Verschlüsselung bei der Übertragung von Daten (Encryption in Transit) . . . . .	513
SageMaker-Notebook-Instanzen schützen . . . . .	515
SageMaker Studio schützen. . . . .	516
SageMaker-Jobs und Modelle schützen. . . . .	519
Daten mit AWS Lake Formation schützen . . . . .	523
Datenbankzugangsdaten mit AWS Secrets Manager schützen . . . . .	523
Governance . . . . .	524
Auditierbarkeit. . . . .	527
Kosten einsparen und die Leistung verbessern . . . . .	529
Zusammenfassung. . . . .	531
<b>Index . . . . .</b>	<b>533</b>