

Spis treści

Przedmowa.....	13
1. Wprowadzenie do danologii na platformie AWS.....	19
Zalety przetwarzania w chmurze	19
Potoki i procesy w danologii	21
Zalecane praktyki z obszaru MLOps	25
Usługi SI Amazona i zautomatyzowane uczenie maszynowe w narzędziu Amazon SageMaker	28
Pobieranie, eksploracja i przygotowywanie danych na platformie AWS	31
Uczenie i dostrajanie modelu za pomocą narzędzia Amazon SageMaker	36
Instalowanie modeli za pomocą usługi Amazon SageMaker i funkcji AWS Lambda	38
Analizy i uczenie maszynowe dla strumieni danych na platformie AWS	39
Infrastruktura platformy AWS i niestandardowy sprzęt	40
Ograniczanie kosztów za pomocą tagów, budżetów i alertów	44
Podsumowanie	44
2. Zastosowania danologii.....	46
Innowacje w każdej branży	46
Spersonalizowane rekomendacje produktów	47
Wykrywanie niestosownych materiałów wideo za pomocą usługi Amazon Rekognition	53
Prognozowanie zapotrzebowania	54
Identyfikowanie fałszywych kont za pomocą usługi Amazon Fraud Detector	58
Używanie usługi Amazon Macie do wykrywania wycieków wrażliwych danych	59
Urządzenia konwersacyjne i asystenci głosowi	61
Analiza tekstu i NLP	61
Wyszukiwanie kognitywne i rozumienie języka naturalnego	66
Inteligentne centra obsługi klienta	67
Przemysłowe usługi SI i konserwacja predykcyjna	68
Automatyzacja domu za pomocą narzędzi AWS IoT i Amazon SageMaker	69

	Pobieranie informacji medycznych z dokumentów służby zdrowia	70
	Samooptymalizująca i inteligentna infrastruktura chmury	71
	Kognitywna i predykcyjna analityka biznesowa	72
	Edukacja następnego pokolenia programistów SI i UM	76
	Zaprogramuj naturalny system operacyjny za pomocą przetwarzania kwantowego	81
	Wzrost wydajności i obniżenie kosztów	85
	Podsumowanie	88
3.	Zautomatyzowane uczenie maszynowe.....	89
	Zautomatyzowane uczenie maszynowe w usłudze SageMaker Autopilot	90
	Śledzenie wyników eksperymentów za pomocą usługi SageMaker Autopilot	91
	Uczenie i instalowanie klasyfikatora tekstu za pomocą usługi SageMaker Autopilot	92
	Zautomatyzowane uczenie maszynowe w usłudze Amazon Comprehend	104
	Podsumowanie	107
4.	Pobieranie danych do chmury.....	108
	Jeziora danych	109
	Kierowanie zapytań do jeziora danych w S3 za pomocą usługi Amazon Athena	115
	Ciągłe pobieranie nowych danych za pomocą narzędzia AWS Glue Crawler	120
	Stosowanie architektury Lake House za pomocą usługi Amazon Redshift Spectrum	121
	Wybór między narzędziami Amazon Athena a Amazon Redshift	128
	Zmniejszanie kosztów i zwiększanie wydajności	129
	Podsumowanie	135
5.	Eksplorowanie zbioru danych.....	137
	Narzędzia do eksplorowania danych w AWS	138
	Wizualizowanie jeziora danych w środowisku SageMaker Studio	138
	Zapytania dotyczące hurtowni danych	151
	Tworzenie paneli kontrolnych za pomocą usługi Amazon QuickSight	159
	Wykrywanie problemów z jakością danych	
	za pomocą narzędzi Amazon SageMaker i Apache Spark	159
	Wykrywanie tendencji w zbiorze danych	167
	Wykrywanie zmian różnego rodzaju za pomocą usługi SageMaker Clarify	174
	Analizowanie danych za pomocą usługi AWS Glue DataBrew	176
	Zmniejszanie kosztów i zwiększanie wydajności	178
	Podsumowanie	180
6.	Przygotowywanie zbioru danych do uczenia modelu.....	181
	Wybieranie i inżynieria cech	181
	Skalowanie inżynierii cech za pomocą zadań SageMaker Processing	194
	Udostępnianie cech za pomocą repozytorium cech z platformy SageMaker	200

Wczytywanie i przekształcanie danych w usłudze SageMaker Data Wrangler	204
Śledzenie historii artefaktów i eksperymentów na platformie Amazon SageMaker	205
Wczytywanie i przekształcanie danych za pomocą usługi AWS Glue DataBrew	209
Podsumowanie	211
7. Uczenie pierwszego modelu.....	213
Infrastruktura platformy SageMaker	213
Instalowanie wyuczonego modelu BERT za pomocą usługi SageMaker JumpStart	217
Tworzenie modelu w platformie SageMaker	219
Krótka historia przetwarzania języka naturalnego	221
Architektura Transformer w algorytmie BERT	223
Uczenie modelu BERT od podstaw	225
Dostrajanie wstępnie wyuczonego modelu BERT	227
Tworzenie skryptu uczenia	230
Uruchamianie skryptu uczenia w usłudze SageMaker Notebook	236
Ocena modeli	242
Debugowanie i profilowanie procesu uczenia modelu w usłudze SageMaker Debugger	247
Interpretowanie i wyjaśnianie predykcji modelu	251
Wykrywanie tendencyjności modelu i wyjaśnianie predykcji	257
Dodatkowe metody uczenia algorytmu BERT	261
Zmniejszanie kosztów i zwiększanie wydajności	269
Podsumowanie	274
8. Uczenie i optymalizowanie modeli na dużą skalę.....	276
Automatyczne znajdowanie optymalnych hiperparametrów dla modelu	276
Stosowanie ciepłego startu dla dodatkowych zadań dostrajania hiperparametrów na platformie SageMaker	283
Skalowanie poziome uczenia rozproszonego na platformie SageMaker	287
Zmniejszanie kosztów i zwiększanie wydajności	294
Podsumowanie	297
9. Instalowanie modeli w środowisku produkcyjnym.....	299
Predykcje w czasie rzeczywistym czy w trybie wsadowym?	299
Generowanie predykcji w czasie rzeczywistym za pomocą punktów końcowych platformy SageMaker	300
Automatyczne skalowanie punktów końcowych platformy SageMaker za pomocą usługi Amazon CloudWatch	308
Strategie instalowania nowych i zaktualizowanych modeli	312
Testowanie i porównywanie nowych modeli	316
Monitorowanie pracy modelu i wykrywanie zmian	327
Monitorowanie jakości danych w punktach końcowych platformy SageMaker	330

Monitorowanie jakości modelu w zainstalowanych punktach końcowych platformy SageMaker	335
Monitorowanie zmian tendencyjności w zainstalowanych punktach końcowych platformy SageMaker	339
Monitorowanie zmian wkładu cech w zainstalowanych punktach końcowych platformy SageMaker	342
Wsadowe generowanie predykcji za pomocą usługi przekształcania wsadowego na platformie SageMaker	345
Funkcje AWS Lambda i usługa Amazon API Gateway	350
Optymalizowanie modeli i zarządzanie nimi na obrzeżach sieci	350
Instalowanie modelu opartego na platformie PyTorch za pomocą narzędzia TorchServe	351
Generowanie predykcji przez algorytm BERT oparty na platformie TensorFlow na platformie AWS Deep Java Library	353
Zmniejszanie kosztów i zwiększanie wydajności	355
Podsumowanie	360
10. Potoki i MLOps.....	361
MLOps	361
Potoki programowe	362
Potoki uczenia maszynowego	363
Koordynowanie potoku za pomocą usługi SageMaker Pipelines	367
Automatyzacja w usłudze SageMaker Pipelines	378
Inne sposoby tworzenia potoków	382
Procesy z udziałem człowieka	391
Zmniejszanie kosztów i zwiększanie wydajności	396
Podsumowanie	397
11. Analizy i uczenie maszynowe dla danych przesyłanych strumieniowo.....	398
Uczenie w trybach online i offline	399
Aplikacje strumieniowe	399
Zapytania oparte na oknach dotyczące strumieniowanych danych	400
Analiza i uczenie maszynowe na podstawie strumieni na platformie AWS	403
Klasyfikowanie recenzji produktów w czasie rzeczywistym za pomocą narzędzi Amazon Kinesis, AWS Lambda i Amazon SageMaker	405
Implementowanie pobierania strumieniowanych danych za pomocą usługi Kinesis Data Firehose	406
Podsumowywanie recenzji produktów w czasie rzeczywistym na podstawie analizy strumienia	410
Konfigurowanie usługi Amazon Kinesis Data Analytics	411
Aplikacje w usłudze Kinesis Data Analytics	419

Klasyfikowanie recenzji produktów	
za pomocą narzędzi Apache Kafka, AWS Lambda i Amazon SageMaker	425
Zmniejszanie kosztów i zwiększanie wydajności	426
Podsumowanie	428
12. Bezpieczna danologia na platformie AWS.....	430
Model podziału odpowiedzialności między platformę AWS i klientów	430
Korzystanie z usługi IAM na platformie AWS	431
Izolacja środowisk obliczeniowych i sieciowych	439
Zabezpieczanie dostępu do danych w S3	442
Szyfrowanie danych w spoczynku	449
Szyfrowanie danych w tranzycie	453
Zabezpieczanie instancji z notatnikami platformy SageMaker	455
Zabezpieczanie środowiska SageMaker Studio	456
Zabezpieczanie zadań i modeli platformy SageMaker	459
Zabezpieczanie usługi AWS Lake Formation	462
Zabezpieczanie danych uwierzytelniających do bazy	
za pomocą AWS Secrets Manager	463
Nadzór	463
Audytowalność	466
Zmniejszanie kosztów i zwiększanie wydajności	468
Podsumowanie	470