

AI

Podejście pragmatyczne

Wprowadzenie do uczenia
maszynowego opartego na chmurze

Noah Gift

Przekład
Marek Włodarz

APN Promise
Warszawa 2019

Spis treści

Wstęp ix

Podziękowania xiii

O autorze xv

Źródła ilustracji xvi

Część I: Wprowadzenie do pragmatycznej AI 1

1 Pragmatyczne podejście do AI 3

Funkcjonalne wprowadzenie do Pythona 4

Instrukcje proceduralne 5

Drukowanie 5

Tworzenie i używanie zmiennych 5

Wiele instrukcji proceduralnych 6

Obliczenia arytmetyczne 6

Łączenie fraz (tekstów) 6

Złożone instrukcje 6

Ciągi i formatowanie ciągów 7

Liczby i operacje arytmetyczne 10

Struktury danych 11

Słowniki 11

Listy 13

Funkcje 13

Pisanie funkcji 14

Prosta funkcja 14

Dokumentowanie funkcji 14

Argumenty funkcji: pozycyjne lub nazwane 14

Używanie struktur sterujących w Pythonie 22

Pętla for 22

Pętle while 23

if/elif/else 24

Wyrażenie generatora 24

Wyrażenia listowe 25

Zagadnienia pośrednie 26

Tworzenie biblioteki w Pythonie 26

Importowanie biblioteki 26

Instalowanie innych bibliotek przy użyciu pip install 27

Klasy 28

Finalne spostrzeżenia 29

2 Narzędzia AI i ML 31

Ekosystem analiz danych w Pythonie: IPython, Pandas, NumPy, Jupyter Notebook, Sklearn 32

R, RStudio, Shiny i ggplot 33

Arkusze kalkulacyjne: Excel oraz Google Sheets 33

Projektowanie rozwiązania chmurowego AI przy użyciu Amazon Web Services 34

DevOps w AWS 34

 Ciągłe dostarczanie 35

 Tworzenie środowiska projektowania oprogramowania dla AWS 35

 Integracja z Jupyter Notebook 42

 Integrowanie narzędzi wiersza polecenia 44

 Integrowanie narzędzi AWS CodePipeline 48

Konfigurowanie Docker na potrzeby analiz danych 53

Inne serwery kompilacji: Jenkins, CircleCI oraz Travis 54

Podsumowanie 54

3 Spartański cykl życia AI 55

Pragmatyczne sprzężenie zwrotne 56

AWS SageMaker 59

Sprzężenie zwrotne AWS Glue 61

AWS Batch 65

Sprzężenia zwrotne oparte na Docker 66

Podsumowanie 68

Część II: AI w chmurze 69

4 Chmurowe projektowanie AI w Google Cloud Platform 71

Przegląd GCP 72

Colaboratory 73

Datalab 75

 Rozszerzanie Datalab przy użyciu Docker i Google Container Registry 75

 Uruchamianie wydajnych maszyn przy użyciu Datalab 76

BigQuery 78

 Przenoszenie danych do BigQuery z wiersza polecenia 79

Usługi AI w chmurze Google 81

 Klasyfikowanie (wielu) ras mojego psa przy użyciu Cloud Vision API 82

Cloud TPU i TensorFlow 84

 Uruchamianie MNIST w Cloud TPU 85

Podsumowanie 87

5 Projektowanie chmurowej AI przy użyciu Amazon Web Services 89

Budowanie rozwiązań rzeczywistości rozszerzonej (AR) i wirtualnej (VR) w AWS 91

 Rozpoznawanie obrazów: potoki AR/VR z EFS i Flask 92

Potok inżynierii danych z wykorzystaniem EFS, Flask i Pandas	94
Podsumowanie	109

Część III: Tworzenie praktycznych aplikacji AI 111

6 Przewidywanie wpływów mediów społecznościowych w NBA 113

Formułowanie problemu	114
Gromadzenie danych	114
Gromadzenie danych z trudnych źródeł	135
Gromadzenie informacji o wyświetleniach stron Wikipedii dla sportowców	135
Gromadzenie informacji o zaangażowaniu sportowców na Twitterze	140
Przeglądanie danych sportowców NBA	143
Nienadzorowane uczenie maszynowe dotyczące zawodników NBA	147
Tworzenie wykresu kafelkowego w R	148
Zebranie wszystkiego razem: drużyny, zawodnicy, siła i dodatki	150
Kolejne pragmatyczne kroki	152
Podsumowanie	152

7 Tworzenie inteligentnego slackbota w AWS 153

Tworzenie bota	153
Konwertowanie biblioteki w narzędzie wiersza polecenia	154
Przenoszenie bota na następny poziom przy użyciu funkcji krokowych AWS	156
Konfigurowanie poświadczeń IAM	157
Korzystanie z chalice	157
Budowanie funkcji krokowej	166
Podsumowanie	169

8 Wyszukiwanie pomysłów na zarządzanie projektami w organizacji GitHub 171

Przegląd problemów związanych z zarządzaniem projektami oprogramowania	171
Pytania do rozważenia	172
Tworzenie wstępnego szkieletu projektu Data Science	173
Zbieranie i transformowanie danych	175
Komunikowanie się z całą organizacją GitHub	177
Tworzenie statystyk specyficznych dla dziedziny	178
Wiązanie projektu Data Science ze środowiskiem wiersza polecenia	180
Wykorzystanie notatnika Jupyter do eksplorowania organizacji	182
Projekt Pallets	182
Przeglądanie metadanych pliku w projekcie CPython	184
Wyszukiwanie usuniętych plików w projekcie CPython	188
Wdrażanie projektu w Python Package Index	191
Podsumowanie	193

- 9 Dynamiczna optymalizacja instancji EC2 w AWS 195**
 - Uruchamianie zadań w AWS 195
 - Instancje Spot 195
 - Podsumowanie 213
- 10 Nieruchomości 215**
 - Eksplorowanie rynku nieruchomości w Stanach Zjednoczonych 215
 - Interaktywna wizualizacja danych w Pythonie 217
 - Grupowanie według rozmiarów i cen 219
 - Podsumowanie 226
- 11 Produkcyjna AI dla treści generowanych przez użytkowników 227**
 - Netflix Prize nie została zaimplementowana w produkcji 228
 - Kluczowe koncepcje systemów rekomendacji 229
 - Korzystanie z platformy Surprise w Pythonie 230
 - Rozwiązania chmurowe dla systemów rekomendacji 232
 - Problemy świata rzeczywistego w mechanizmach rekomendacji 233
 - Praktyczne problemy systemów rekomendacji: integracja z produkcyjnymi API 234
 - Chmurowe NLP i analiza opinii 238
 - NLP w Azure 238
 - NLP w GCP 241
 - Eksplorowanie API jednostek 241
 - Produkcyjny bezserwerowy potok AI dla NLP w AWS 244
 - Podsumowanie 250
- 12 Akceleratory AI 251**
- 13 Wybór liczby grup 253**
 - Indeks 255*