

Spis treści

Przedmowa 9

CZĘŚĆ I. OKREŚLENIE WŁAŚCIWEGO PODEJŚCIA ML

1. Od pomysłu do podejścia ML 19

Określenie, co jest możliwe 20

Modele 21

Dane 27

Zarys edytora ML 30

Spróbujmy wszystko zrobić za pomocą ML 30

Najprostsze podejście: wykonanie algorytmu 32

Etap pośredni: uczenie się na podstawie doświadczenia 33

Jak wybierać projekty ML i określać ich priorytety - Monica Rogati 34

Podsumowanie 36

2. Opracowanie planu 37

Mierzenie postępów 37

Wydajność biznesowa 38

Wydajność modelu 39

Aktualność i przesunięcie rozkładu danych 41

Szybkość 43

Szacowanie zakresu i wyzwań 44

Wykorzystanie doświadczenia w danej dziedzinie 44

Podążanie za wielkimi 45

Planowanie edytora ML 48

Początkowy plan edytora 48

Zawsze należy zaczynać od prostego modelu 49

Sposób na systematyczne postępy - prosty początek 50

Prosty początkowy proces 50

Proces dla edytora ML 51

Podsumowanie 53

CZĘŚĆ II. BUDOWANIE PROCESU

3. Zbudowanie pierwszego całościowego procesu 57

Najprostsza architektura 57

Prototyp edytora ML 58

Analizowanie i oczyszczanie danych 59

Tokenizacja testu 60

Generowanie cech 60

Testowanie procesu 62

Wrażenia użytkowników 62

Modelowanie wyników 62

Ocena prototypu edytora ML 63

Model 64

Wrażenia użytkowników 65

Podsumowanie 65

4. Pozyskiwanie początkowego zbioru danych 67

Iterowanie zbioru danych 67

Badanie danych 68

Badanie pierwszego zbioru danych 68

Bądź skuteczny, zacznij od czegoś małego 69

Informacje i produkty 69

Ocena jakości danych 70

Etykiety i wyszukiwanie trendów 75

Statystyki podsumowujące 76

Efektywne badanie i znakowanie danych 78

Wykonanie algorytmu 92

Trendy danych 93

Niech dane informują o cechach i modelach 94

Budowanie cech na podstawie wzorców 94

Cechy edytora ML 97

Jak wyszukiwać, znakować i wykorzystywać dane - Robert Munro 97

Podsumowanie 99

CZĘŚĆ III. ITEROWANIE MODELI

5. Trening i ocena modelu 103

Najprostszy, odpowiedni model 103

Proste modele 103

Od wzorców do modeli 105

Podział zbioru danych 107

Podział danych dla edytora ML 112

Ocena wydajności 113

Ocena modelu: nie tylko dokładność 116

Porównywanie danych i prognoz 116

Tablica pomyłek 117

Krzywa ROC 117

Krzywa kalibracyjna 119

Redukcja wymiarowości w analizie błędów 121

Metoda top-k 121

Inne modele 125

Ocena ważności cech 125

Ocena bezpośrednio z klasyfikatora 126

Analiza czarnej skrzynki 127

Podsumowanie 129

6. Diagnozowanie problemów 131

Dobre praktyki programowania 131

Dobre praktyki w ML 132

Diagnozowanie połączeń: wizualizacja i testy 133

Na początek jeden przykład 133

Testowanie kodu ML 139

Diagnozowanie treningu 143

Trudność zadania 144

Problemy optymalizacyjne 146

Diagnozowanie uogólnienia modelu 147

Wyciek danych 148

Nadmierne dopasowanie 148

Analiza zadania 151

Podsumowanie 152

7. Przygotowywanie zaleceń przy użyciu klasyfikatora 153

Wyodrębnianie zaleceń z modeli 154

Co można osiągnąć bez modelu? 154

Wyodrębnianie globalnych ważności cech 155

Wykorzystanie ocen modelu 156

Wyodrębnianie lokalnych ważności cech 156

Porównanie modeli 158

Wersja 1: karta raportu 159

Wersja 2: lepszy, ale mniej czytelny model 159

Wersja 3: zrozumiałe zalecenia 161

Formułowanie zaleceń edycyjnych 162

Podsumowanie 165

CZĘŚĆ IV. WDRAŻANIE I MONITOROWANIE MODELI

8. Wdrażanie modeli 169

Dane 169

Własność danych 170

Zniekształcenie danych 170

Zniekształcenia systemowe 172

Modele 172

Sprzężenie zwrotne 173

Inkluzyjna wydajność modelu 174

Kontekst 175

Ataki 175

Nadużycia i podwójne zastosowanie 176

Eksperymenty wysyłkowe - Chris Harland 177

Podsumowanie 179

9. Opcje wdrażania modeli 181

Wdrożenie po stronie serwera 181

Aplikacja strumieniowa, czyli interfejs API 181

Prognozowanie wsadowe 184

Wdrożenie po stronie klienta 185

Model w urządzeniu 186

Model w przeglądarce 187

Uczenie federacyjne: podejście hybrydowe 188

Podsumowanie 189

10. Zabezpieczanie modelu 191

Ochrona przed awariami 191

Sprawdzanie danych wejściowych i wyjściowych 191

Scenariusze awaryjne modelu 194

Inżynieria wydajności 198

Obsługa wielu użytkowników 198

Zarządzanie cyklem życia modelu i danych 201

Przetwarzanie danych i skierowany graf acykliczny 203

Opinie użytkowników 204

Wspieranie badaczy danych we wdrażaniu modeli - Chris Moody 206

Podsumowanie 208

11. Monitorowanie i aktualizowanie modeli 209

Monitorowanie oszczędza kłopotów 209

Informowanie o konieczności odświeżenia modelu 209

Wykrywanie nadużyć 210

Co monitorować? 210

Wskaźniki wydajności 211

Wskaźniki biznesowe 213

CI/CD w dziedzinie ML 214

Testy A/B i eksperymenty 215

Inne podejścia 217

Podsumowanie 219