

# *spis treści*

---

przedmowa xv

podziękowania xvii

o tej książce xix

o autorze xxii

o okładce xxiii

## **CZĘŚĆ 1 PRZYGOTOWANIE DO SYSTEMÓW REKOMENDACJI ..... 1**

### **1 Co to jest system rekomendacji? 3**

#### 1.1. Rekomendacje z życia wzięte 4

Systemy rekomendacji pasują do Internetu 5 ■ Długi ogon 6 ■ System rekomendacji Netfliksa 6 ■ Definicja systemu rekomendacji 13

#### 1.2. Taksonomia systemów rekomendacji 15

Domena 16 ■ Cel 16 ■ Kontekst 17 ■ Poziom personalizacji 17 ■ Czyje opinie 18 ■ Prywatność i wiarygodność 19 ■ Interfejs 20 ■ Algorytmy 22

#### 1.3. Uczenie maszynowe i nagroda Netfliksa 24

#### 1.4. Witryna MovieGEEKs 25

Projekt i specyfikacja 26 ■ Architektura 27

#### 1.5. Budowanie systemu rekomendacji 29

## 2 Zachowanie użytkowników i jak zbierać o tym dane 31

### 2.1. Jak (według mnie) Netflix zbiera dowody podczas przeglądania 32

Dowody, które zbiera Netflix 34

### 2.2. Znajdowanie przydatnych zachowań użytkowników 36

Przechwytywanie wrażeń odwiedzających 37 ■ Czego można się dowiedzieć od osoby rozglądającej się po sklepie 37 ■ Akt zakupu 42 ■ Konsumpcja produktów 43 ■ Oceny odwiedzających 44 ■ Poznawanie swoich klientów (starym) sposobem Netfiksa 47

### 2.3. Identyfikowanie użytkowników 47

### 2.4. Uzyskiwanie danych odwiedzających z innych źródeł 48

### 2.5. Moduł zbierający dane 49

Budowanie plików projektu 50 ■ Model danych 50 ■ Snitch: moduł zbierający dowody po stronie klienta 51 ■ Integrowanie modułu z MovieGEEKs 53

### 2.6. Czym są użytkownicy w systemie i jak ich modelować 55

## 3 Monitorowanie systemu 59

### 3.1. Dlaczego dodanie pulpitu nawigacyjnego jest dobrym pomysłem 60

Odpowiadanie na pytanie „jak nam idzie?” 60

### 3.2. Obliczanie statystyk 62

Analityka internetowa 62 ■ Podstawowe statystyki 63 ■ Konwersje 64 ■ Analizowanie ścieżki do konwersji 67 ■ Ścieżka konwersji 69

### 3.3. Persony 71

### 3.4. Pulpit nawigacyjny MovieGEEKs 74

Automatyczne generowanie danych w dzienniku 74 ■ Specyfikacja i projektowanie pulpitu nawigacyjnego 75 ■ Analityczny pulpit nawigacyjny – szkielet projektu 75 ■ Architektura 76

## 4 Oceny i jak je obliczać 80

### 4.1. Preferencje użytkownik-przedmiot 81

Definicja ratingu 81 ■ Macierz użytkownik-przedmiot 82

### 4.2. Jawne lub niejawne oceny 84

Jak wykorzystujemy zaufane źródła do rekomendacji 85

### 4.3. Ponowne odwiedzanie ocen jawnych 86

### 4.4. Czym są oceny niejawne? 86

Rekomendowanie osób 88 ■ Uwagi dotyczące obliczania ocen 88

- 4.5. Obliczanie ocen niejawnych 91
  - Spojrzenie na dane behawioralne 92 ■ To można uznać za problem uczenia maszynowego 96
- 4.6. Jak zaimplementować oceny niejawne 97
  - Dodawanie aspektu czasu 101
- 4.7. Mniej popularne elementy mają większą wartość 103

## 5 Niespersonalizowane rekomendacje 107

- 5.1. Co to są niespersonalizowane rekomendacje? 108
  - Co to jest reklama? 108 ■ Co robi rekomendacja? 110
- 5.2. Jak tworzyć rekomendacje, gdy nie masz danych 110
  - Top 10: lista najpopularniejszych przedmiotów 112
- 5.3. Implementacja listy najpopularniejszych pozycji i podstawy dla komponentu systemu rekomendacji 114
  - Komponent systemu rekomendacji 114 ■ Kod MovieGEEKs z GitHub 115 ■ System rekomendacji 115 ■ Dodawanie listy najpopularniejszych pozycji do MovieGEEKs 116 ■ Uatrakcyjnienie wyglądu treści 117
- 5.4. Rekomendacje oparte na wartościach inicjujących 119
  - Często kupowane przedmioty podobne do oglądanych 120 ■ Reguły asocjacyjne 121 ■ Implementowanie reguł asocjacyjnych 126 ■ Zapisywanie reguł asocjacyjnych w bazie danych 131 ■ Uruchamianie kalkulatora reguł asocjacyjnych 132 ■ Używanie różnych zdarzeń do tworzenia reguł asocjacyjnych 134

## 6 Użytkownik (i treści), który przyszedł z zimna 136

- 6.1. Co to jest zimny start? 137
  - Zimne produkty 138 ■ Zimny gość 139 ■ Szare owce 140 ■ Spójrzmy na przykłady z życia wzięte 140 ■ Co możesz zrobić z zimnym startem? 142
- 6.2. Śledzenie odwiedzających 142
  - Uparcie anonimowi użytkownicy 142
- 6.3. Rozwiązywanie problemów z zimnym startem za pomocą algorytmów 143
  - Używanie reguł asocjacyjnych do tworzenia rekomendacji dla zimnych użytkowników 143 ■ Korzystanie ze znajomości domeny i zasad biznesowych 145 ■ Korzystanie z segmentów 146 ■ Wykorzystywanie kategorii do obejścia problemu szarych owiec i jak wprowadzać zimny produkt 147
- 6.4. Ci, którzy nie pytają, nie będą wiedzieć 149
  - Kiedy odwiedzający nie jest już nowy 150

## 6.5. Wykorzystanie reguł asocjacyjnych do szybkiego przedstawiania rekomendacji 151

Znajdowanie zebranych elementów 152 ■ Wydobywanie reguł asocjacyjnych i porządkowanie ich według ufności 152 ■ Wyświetlanie rekomendacji 153 ■ Ocena implementacji 156

## CZĘŚĆ 2 ALGORYTMY SYSTEMÓW REKOMENDACJI ..... 157

### 7 Znajdowanie podobieństw wśród użytkowników i wśród treści 159

#### 7.1. Dlaczego podobieństwo? 161

Czym jest funkcja podobieństwa? 161

#### 7.2. Podstawowe funkcje podobieństwa 162

Indeks Jaccarda 163 ■ Pomiar odległości z normami Lp 165 ■ Podobieństwo kosinusowe 168 ■ Znajdowanie podobieństwa za pomocą współczynnika korelacji liniowej Pearsona 171 ■ Testowanie podobieństwa Pearsona 172 ■ Korelacja liniowa jest podobna do kosinusowej 174

#### 7.3. Klastrowanie metodą k-średnich 175

Algorytmy klastrowania metodą k-średnich 175 ■ Przekładanie algorytmu klastrowania metodą k-średnich na Pythona 178

#### 7.4. Implementacja podobieństw 183

Implementacja podobieństwa w witrynie MovieGEEKs 185 ■ Implementacja klastrowania w witrynie MovieGEEKs 188

### 8 Wspólna filtracja w sąsiedztwie 193

#### 8.1. Wspólna filtracja: lekcja historii 195

Kiedy informacja stała się wspólnie przefiltrowana 195 ■ Wzajemna pomoc 196 ■ Macierz ocen 198 ■ Potok wspólnej filtracji 199 ■ Czy należy używać wspólnej filtracji użytkownik-użytkownik czy element-element? 199 ■ Wymagania dotyczące danych 200

#### 8.2. Obliczanie rekomendacji 201

#### 8.3. Obliczanie podobieństw 202

#### 8.4. Algorytm Amazona do wstępnego obliczania podobieństwa elementów 202

#### 8.5. Sposoby wybierania sąsiedztwa 207

#### 8.6. Znajdowanie właściwego sąsiedztwa 209

#### 8.7. Sposoby obliczania przewidywanych ocen 209

#### 8.8. Przewidywanie z filtrowaniem opartym na elementach 211

Obliczanie przewidywań elementów 212

- 8.9. Problemy z zimnym startem 213
- 8.10. Kilka słów o pojęciach w uczeniu maszynowym 213
- 8.11. Wspólna filtracja na stronie MovieGEEKs 215
  - Filtrowanie oparte na elementach 216
- 8.12. Jaka jest różnica między rekomendacjami z reguł asocjacyjnych a rekomendacjami opartymi na wspólnej filtracji 223
- 8.13. Dźwignie do majstrowania przy wspólnej filtracji 223
- 8.14. Plusy i minusy wspólnej filtracji 225

## 9 Ewaluacja i testowanie systemu rekomendacji 227

- 9.1. Biznes chce wzrostów, sprzedaży krzyżowej, sprzedaży droższych towarów i konwersji 228
- 9.2. Dlaczego ewaluacja jest ważna? 229
- 9.3. Jak interpretować zachowanie użytkowników 230
- 9.4. Co należy mierzyć 230
  - Rozumienie mojego gustu: minimalizowanie błędów przewidywania 232 ■ Różnorodność 232 ■ Pokrycie 233 ■ Szczęśliwy zbieg okoliczności 236
- 9.5. Przed wdrożeniem systemu rekomendacji... 236
  - Zweryfikuj algorytm 236 ■ Testowanie regresyjne 237
- 9.6. Rodzaje ewaluacji 238
- 9.7. Ewaluacja offline 239
  - Co zrobić, kiedy algorytm nie przedstawia żadnych rekomendacji 240
- 9.8. Eksperymenty offline 240
  - Przygotowywanie danych do eksperymentów 246
- 9.9. Implementacja eksperymentu na witrynie MovieGEEKs 253
  - Lista rzeczy do zrobienia 253
- 9.10. Ewaluacja zbioru testowego 257
  - Rozpoczynanie od predyktora bazowego 258 ■ Znajdowanie odpowiednich parametrów 260
- 9.11. Ewaluacja online 262
  - Eksperyment kontrolowany 262 ■ Testy A/B 262
- 9.12. Testowanie ciągle z eksploatacją/eksploracją 264
  - Pętle informacji zwrotnych 265

**10 Filtrowanie oparte na treści 267**

## 10.1. Przykład opisowy 268

## 10.2. Filtrowanie oparte na treści 271

## 10.3. Analizator zawartości 272

Wychwytywanie cech dla profilu przedmiotu 272 ■ Dane katagoryczne z małymi liczbami 275 ■ Przekształcanie roku na porównywalne cechy 275

## 10.4. Wydobywanie metadanych z opisów 276

Przygotowywanie opisów 276

## 10.5. Znajdowanie ważnych słów za pomocą TF-IDF 280

## 10.6. Modelowanie tematów przy użyciu LDA 282

Jakie gałki możesz pokręcić, aby dostosować LDA? 289

## 10.7. Znajdowanie podobnych treści 292

## 10.8. Tworzenie profilu użytkownika 293

Tworzenie profilu użytkownika za pomocą LDA 293 ■ Tworzenie profilu użytkownika za pomocą TF-IDF 293

## 10.9. Rekomendacje oparte na treści na witrynie MovieGEEKs 295

Pobieranie danych 295 ■ Uczenie modelu 298 ■ Tworzenie profili elementów 299 ■ Tworzenie profili użytkowników 299 ■ Pokazywanie rekomendacji 301

## 10.10. Ocena systemu rekomendacji opartego na treści 302

## 10.11. Zalety i wady filtrowania opartego na treści 304

**11 Znajdowanie ukrytych gatunków za pomocą faktoryzacji macierzy 306**

## 11.1. Czasami dobrze jest zredukować ilość danych 307

## 11.2. Przykład tego, co chcesz rozwiązać 309

## 11.3. Powiew algebry liniowej 312

Macierz 312 ■ Co to jest faktoryzacja? 315

## 11.4. Tworzenie faktoryzacji za pomocą SVD 316

Dodawanie nowego użytkownika metodą folding-in 322 ■ Jak tworzyć rekomendacje za pomocą SVD 324 ■ Predyktory bazowe 325 ■ Dynamika czasowa 327

## 11.5. Budowanie faktoryzacji za pomocą Funk SVD 328

Średnia kwadratowa błędu 328 ■ Metoda gradientu prostego 329 ■ Stochastyczne zejście wzdłuż gradientu 332 ■ Wreszcie zabieramy się za faktoryzację 333 ■ Dodawanie odchyleń 334 ■ Jak zacząć i kiedy przestać 335

11.6. Budowanie rekomendacji przy użyciu Funk SVD	340
11.7. Implementacja Funk SVD na witrynie MovieGEEKs	342
Co zrobić z odstającymi wartościami	348
Aktualizowanie modelu	350
Szybka implementacja	350
11.8. Dane jawne kontra niejawne	350
11.9. Ewaluacja	350
11.10. Dźwignie do majstrowania dla Funk SVD	353
<b>12 Branie tego co najlepsze ze wszystkich algorytmów: implementacja hybrydowych silników rekomendacji</b>	<b>355</b>
12.1. Zagmatwany świat hybryd	357
12.2. Monolityczny	357
Mieszanie cech opartych na treści z danymi behawioralnymi, aby usprawnić system rekomendacji wspólnej filtracji	358
12.3. Mieszany hybrydowy system rekomendacji	359
12.4. Algorytmy złożone	360
Przełączany system rekomendacji	362
Ważony system rekomendacji	363
Regresja liniowa	364
12.5. Liniowe łączenie modeli wężone cechami (FWLS)	365
Meta-cechy: wagi jako funkcje	365
Algorytm	367
12.6. Implementacja	375
<b>13 Rangowanie i nauka rangowania</b>	<b>385</b>
13.1. Nauka rangowania na przykładzie Foursquare	386
13.2. Rangowanie ze zmianą kolejności	391
13.3. Czym jest nauka rangowania ze zmianą kolejności?	392
Trzy typy algorytmów LTR	392
13.4. Bayesowskie rangowanie spersonalizowane (BPR)	394
Rangowanie z BPR	396
Sztuczki matematyczne (magia zaawansowana)	398
Algorytm BPR	401
BPR z faktoryzacją macierzy	402
13.5. Implementacja BPR	402
Tworzenie rekomendacji	408
13.6. Ewaluacja	410
13.7. Dźwignie do majstrowania przy BPR	413

**14 Przyszłość systemów rekomendacji 415**

14.1. Ta książka w kilku zdaniach 416

14.2. Tematy do dalszej nauki 420

Dalsza lektura 420 ■ Algorytmy 421 ■ Kontekst 421 ■ Interakcje człowiek-komputer 421 ■ Wybór dobrej architektury 422

14.3. Jaka jest przyszłość systemów rekomendacji? 423

14.4. Końcowe przemyślenia 427

**indeks 429**