

Unikalne wprowadzenie do nauki o danych!

W dzisiejszych czasach najcenniejszym dobrem jest informacja. Ogromne ilości danych są przechowywane w przepastnych bazach danych, a kluczem do sukcesu jest ich umiejętna analiza i wyciąganie wniosków. To dynamicznie rozwijająca się dziedzina wiedzy, w której do tej pory brakowało solidnych podręczników, pozwalających na dogłębne poznanie tego obszaru. Na szczęście to się zmieniło!

To unikalna książka, w której badacze z największych firm branży IT dzielą się skutecznymi technikami **analizy danych**. Z kolejnych rozdziałów dowiesz się, czym jest nauka o danych, **model danych** oraz **test A/B**. Ponadto zdobędziesz wiedzę na temat **wnioskowania statystycznego, algorytmów, języka R oraz wizualizacji danych**. Sięgnij po tę książkę, jeżeli chcesz się dowiedzieć, jak wykrywać oszustwa, korzystać z **MapReduce** oraz badać przyczynowość. To obowiązkowa pozycja na półce czytelników zainteresowanych badaniem danych.

Wśród tematów poruszonych w książce odnajdziesz:

- Wnioskowanie statystyczne, eksploracyjną analizę danych i proces (metodologię) nauki o danych
- Algorytmy
- Filtry spamu, **naiwny algorytm Bayesa** i wstępną obróbkę danych
- Regresję logistyczną**
- Modelowanie finansowe**
- Mechanizmy rekomendacji i przyczynowość
- Wizualizowanie danych**
- Sieci społecznościowe i dziennikarstwo danych
- Inżynierię danych, systemy **MapReduce, Pregel i Hadoop**

Wyciągnij wartościowe wnioski z posiadanych informacji!

Przedmowa (9)

Rozdział 1. Wprowadzenie: czym jest nauka o danych? (19)

- Wielkie dane i szum wokół badania danych (19)
- Pokonywanie szumu (21)
- Dlaczego teraz? (22)
- Obecny horyzont (z domieszką historii) (23)
- Profil nauki o danych (27)
- Eksperyment myślowy - metadefinicja (28)
- Kim zatem jest badacz danych? (29)

Rozdział 2. Wnioskowanie statystyczne, eksploracyjna analiza danych i proces badania danych (33)

- Myślenie statystyczne w epoce wielkich danych (33)
- Eksploreacyjna analiza danych (46)
- Proces badania danych (51)
- Eksperyment myślowy: jak zasymulować chaos? (54)
- Studium przypadku: RealDirect (55)

Rozdział 3. Algorytmy (61)

- Algorytmy uczenia maszynowego (62)
- Trzy algorytmy podstawowe (63)
- Zadanie: podstawowe algorytmy uczenia maszynowego (88)
- Podsumowując to wszystko (92)
- Eksperyment myślowy - automatyczny statystyk (93)

Rozdział 4. Filtry spamu, naiwny Bayes i obróbka danych (95)

- Eksperyment myślowy - nauczanie przez przykład (95)
- Naiwna metoda Bayesa (99)
- Zróbmy to z polotem - wygładzanie metodą Laplace'a (103)
- Porównanie naiwnej metody Bayesa z k-NN (104)
- Przykładowy kod w bashu (105)
- Skrobiać po Sieci - interfejsy API i inne narzędzia (106)

Rozdział 5. Regresja logistyczna (111)

- Eksperymenty myślowe (112)
- Klasyfikatory (113)
- Przypadek regresji logistycznej w M6D (115)
- Zadanie z Media6Degrees (124)

Rozdział 6. Znaczniki czasu i modelowanie finansowe (129)

- Kyle Teague i GetBlue (129)
- Znaczniki czasu (131)
- Cathy O'Neil (136)
- Eksperyment myślowy (136)
- Modelowanie finansowe (137)
- Zadanie: GetGlue i zdarzenia opatrzone znacznikami czasu (150)

Rozdział 7. Wydobywanie znaczeń z danych (153)

- William Cukierski (153)
- Model Kaggle (156)
- Eksperyment myślowy: jakie są etyczne następstwa Robo-Gradera? (159)
- Wybór cech (161)
- David Huffaker: hybrydowe podejście Google do badań społecznych (176)

Rozdział 8. Doradzarki ? budowanie na styku z użytkownikiem produktu danych na miarę (181)

- Doradzarka z prawdziwego zdarzenia (182)
- Eksperyment myślowy ? bąbelki filtrowania (192)
- Zadanie: zbuduj własną doradzkę (192)

Rozdział 9. Wizualizacja danych i wykrywanie oszustw (195)

- Historia wizualizacji danych (195)
- Czym jest nauka o danych? Nowym rozdziałem? (197)
- Przykładowe projekty wizualizacji danych (199)
- Marka projekty wizualizacji danych (202)
- Nauka o danych i ryzyko (209)
- Wizualizacja danych w Square (219)
- Eksperyment myślowy Iana (220)
- Wizualizacja danych dla takich jak my (220)

Rozdział 10. Sieci społeczne i dziennikarstwo danych (223)

- Analiza sieci społecznych w Morningside Analytics (223)
- Analiza sieci społecznych (225)
- Terminologia z obszaru sieci społecznych (226)
- Eksperyment myślowy (228)
- Metody analityczne w Morningside (229)

Szersze tło statystyczne analizy sieci społecznych (232)
Dziennikarstwo danych (236)

Rozdział 11. Przyczynowość (239)

Korelacja nie implikuje przyczynowości (240)
Starania witryny OK Cupid (242)
Złoty standard - losowe próby kliniczne (243)
Testy A/B (245)
Z braku czegoś lepszego: badania obserwacyjne (247)
Trzy zalecenia (252)

Rozdział 12. Epidemiologia (253)

Wykształcenie i kariera zawodowa Madigana (253)
Eksperyment myślowy (254)
Współczesna statystyka akademicka (254)
Literatura medyczna i badania obserwacyjne (255)
Stratyfikacja nie rozwiązuje problemu czynników zaburzających (256)
Czy jest lepsze wyjście? (258)
Eksperyment badawczy (partnerstwo w wynikach obserwacji medycznych) (259)
Finalny eksperyment myślowy (263)

Rozdział 13. Wnioski z konkursów danych: wycieki danych i ocenianie modelu (265)

Profil Claudii jako badaczki danych (265)
Zawody w wydobywaniu danych (267)
Jak być dobrym modelarzem (268)
Wyciek danych (268)
Jak unikać wycieków (273)
Ocenianie modeli (273)
Wybór algorytmu (278)
Przykład końcowy (278)
Przemyślenia na pożegnanie (279)

Rozdział 14. Inżynieria danych - MapReduce, Pregel i Hadoop (281)

O Davidzie Crawshaw (282)
Eksperyment myślowy (282)
MapReduce (283)
Problem częstości słów (284)
Inne przykłady użycia systemu MapReduce (288)
Pregel (289)
O Joshu Willsie (289)
Eksperyment myślowy (290)
Gdy się jest badaczem danych (290)
Interludium ekonomiczne - Hadoop (291)
Wracając do Josha - tok pracy (292)
Jak zatem zacząć z Hadoopem? (293)

Rozdział 15. Głos studentów (295)

Proces myślowy (295)
Już nie naiwny (296)
Pomocne dłonie (298)
Twoje koszty mogą być różne (299)

Tunele spinające (301)

Z naszych prac (301)

Rozdział 16. Następna generacja badaczy danych, arogancja i etyka (303)

Co zostało zrobione? (303)

Czym jest (spytajmy raz jeszcze!) nauka o danych? (303)

Jacy są badacze danych następnej generacji? (306)

Jak być etycznym badaczem danych (308)

Rada dotycząca kariery (313)

Skorowidz (315)