

Spis treści

Wstęp	13
I Wszystko, co chciałeś wiedzieć o arkuszu kalkulacyjnym, ale bałeś się o to zapytać	21
Przykładowe proste dane	22
Szybkie statystyki opisowe	23
Tabele Excela	24
Filtrowanie i sortowanie	26
Formatowanie tabeli	27
Odwołania strukturalne	28
Dodawanie kolumn do tabeli	31
Formuły przeznaczone do wyszukiwania	32
WYSZUKAJ.POZIOMO	32
INDEKS/PODAJ.POZYCJĘ	34
X.WYSZUKAJ	36
Tabele przestawne	38
Korzystanie z formuł tablicowych	40
Rozwiązywanie problemów za pomocą narzędzia Solver	41
2 Ustaw i zapomnij! Wprowadzenie do Power Query	49
Czym jest Power Query?	49
Przykładowe dane	50
Rozpoczęcie pracy z Power Query	51
Filtrowanie wierszy	54
Usuwanie kolumn	55
Znajdź i zastąp	56
Zamknij i załaduj... do tabeli	58

3	Naiwny klasyfikator bayesowski i niezwykła lekkość bycia idiotą	61
	Najszybsze na świecie wprowadzenie do rachunku prawdopodobieństwa	61
	Obliczanie prawdopodobieństwa warunkowego	62
	Prawdopodobieństwo części wspólnej, reguła łańcuchowa i niezależność	62
	A co, jeżeli sytuacje są zależne od siebie?	63
	Twierdzenie Bayesa	64
	Oddzielanie sygnału od szumu	65
	Tworzenie modelu sztucznej inteligencji za pomocą twierdzenia Bayesa	66
	Zwykle zakłada się, że wysokopoziomowe prawdopodobieństwa klas są sobie równe	68
	Kilka innych drobnostek	68
	Czas rozpocząć zabawę z Excelem	70
	Porządkowanie danych za pomocą Power Query	70
	Dzielenie na znakach spacji: każde słowo musi dostać to, co mu się należy	73
	Zliczanie leksemów i obliczanie prawdopodobieństw	78
	Zbudowaliśmy model. Skorzystajmy z niego!	80
4	Analiza skupień. Część I — zastosowanie algorytmu centroidów do segmentowania bazy klientów	87
	Zabawy taneczne na obozie letnim	88
	Prawdziwy problem: implementacja algorytmu centroidów w e-mail marketingu	93
	Początkowy zbiór danych	93
	Określanie tego, co chcemy mierzyć	94
	Zacznij od czterech grup	97
	Odległość euklidesowa — pomiar odległości w linii prostej	98
	Określanie położenia środków klastrów	102
	Analiza uzyskanych wyników	104
	Ustalanie najlepszej oferty dla danego klastra	106
	Sylwetka podziału — dobry sposób na określenie optymalnej liczby klastrów	110
	A może potrzebujesz pięciu klastrów?	117
	Dzielenie klientów na pięć klastrów za pomocą narzędzia Solver	118
	Ustalanie najlepszych ofert dla wszystkich pięciu klastrów	119
	Określanie sylwetki podziału na pięć klastrów	122

Podział na grupy za pomocą algorytmu K-mediodów i asymetryczny pomiar odległości	123
Podział na grupy za pomocą metody K-mediodów	123
Stosowanie lepszego sposobu pomiaru odległości	123
Implementacja za pomocą Excela	126
Najlepsze oferty przy podziale na pięć klastrow za pomocą median	128
5 Analiza skupień. Część II — grafy i analiza sieci	133
Czym jest graf sieci?	134
Wizualizacja prostego grafu	134
Wyjście poza dodatek GiGraph i listy sąsiedztwa	138
Tworzenie grafu na podstawie danych sprzedaży wina	140
Tworzenie macierzy podobieństwa kosinusowego	142
Generowanie grafu r-sąsiedztwa	144
Wprowadzenie do Gephi	146
Tworzenie statycznej macierzy sąsiedztwa	146
Macierz r-sąsiedztwa w Gephi	147
Stopień rozgałęzienia	151
Edycja danych grafu	152
Jaka jest wartość krawędzi? Nagradzanie i karanie krawędzi — modularność grafu	154
Czym jest punkt, a czym kara?	155
Tworzenie arkusza punktacji	159
Czas dokonać podziału na grupy	161
Podział 1.	161
Podział 2. — kontratak	167
Podział 3. — zemsta	169
Grupy — kodowanie i analiza	170
Tam i z powrotem — czas na Gephi	174
6 Regresja jako przodek nadzorowanego uczenia maszynowego i sztucznej inteligencji	181
Przewidywanie ciąży klientów na podstawie regresji liniowej	183
Zbiór cech	184
Tworzenie treningowego zbioru danych	186

Tworzenie zmiennych fikcyjnych	188
Pobawmy się regresją liniową	191
Parametry regresji liniowej: współczynnik determinacji, test F i test t	198
Przewidywanie ciąży na nowym zbiorze danych i sprawdzanie jakości modelu	208
Przewidywanie ciąży klientów za pomocą regresji logistycznej	218
Najpierw musisz określić funkcję wiążącą	218
Tworzenie funkcji logistycznej i ponowna optymalizacja	219
Praca nad prawdziwą regresją logistyczną	222
7 Modele zespołowe — dużo nie najlepszej pizzy	229
Korzystanie z danych z rozdziału 6.	230
Agregacja — losuj, trenuj, powtórz	231
Pieniek decyzyjny to kolejne określenie słabego klasyfikatora	231
To wcale nie wydaje się takie słabe!	231
Więcej mocy!	234
Czas rozpocząć proces trenowania	235
Ocena działania modelu zespolonego	248
Wzmacnianie — jeżeli uzyskałeś niesatysfakcjonujące wyniki, to wzmocnij swój model i uruchom go jeszcze raz	252
Trenowanie modelu — każda cecha ma swoje pięć minut	253
Wydałość modelu wzmacnianych reguł decyzyjnych	261
8 Prognozowanie — oddychaj spokojnie, i tak nie wygrasz	265
Hossa na rynku sprzedaży mieczy	266
Szeregi czasowe	267
Zaczynaj od prostego wygładzania wykładniczego	268
Przygotowanie arkusza prognozy prostego wygładzania wykładniczego	271
Być może dane zawierają trend	280
Podwójne wygładzanie wykładnicze (metoda Holta)	282
Metoda Holta w arkuszu kalkulacyjnym	283
To wszystko? Analiza autokorelacji	290

Wielokrotne wygładzanie wykładnicze — model Holta-Wintersa	298
Określanie początkowych wartości poziomu, trendu i sezonowości	300
Tworzenie prognozy	307
Czas na optymalizację	313
Interwały prognozy	315
Tworzenie wykresu warstwowego wachlarza wartości	321
Arkusze prognozy w Excelu	323
9 Modelowanie optymalizacyjne — świeżo wyciśnięty sok	
nie zamiesza się sam	327
Ale czy to w ogóle jest analiza danych?	328
Zacznijmy od prostego kompromisu	329
Przedstawienie problemu w formie wielokomórki	330
Rozwiązywanie problemu poprzez przesuwanie poziomicy	331
Metoda simpleks — kręcenie się wokół rogów	333
Praca w Excelu	335
Szklanka świeżego soku pomarańczowego prosto z drzewa...	
z przystankiem na modelowanie	340
Zacznijmy od specyfikacji soków	342
Stałość produktu wyjściowego	343
Wprowadzanie danych do Excela	344
Określanie problemu w dodatku Solver	347
Obniżanie standardów	349
Usuwanie cuchnącego problemu	
— minimalizacja maksymalnych odchyień	352
Warunki i ograniczenie „wielkiego M”	354
Mnożenie zmiennych — skorzystajmy ze 110% mocy Excela	359
Modelowanie ryzyka	366
Dane pochodzące z rozkładu normalnego	366
10 Wykrywanie obserwacji odstających	375
Element odstający to też człowiek	376
Fascynująca sprawa Hadlumów	376
Metoda Tukeya	378
Implementacja metody Tukeya w arkuszu kalkulacyjnym	379
Ograniczenia tej prostej techniki	381

Nie tragiczny, ale słaby we wszystkim	383
Przygotowywanie danych do utworzenia wykresu	384
Tworzenie grafu	386
Określanie k najbliższych sąsiadów	388
Pierwsza metoda wykrywania elementów odstających grafu	
— skorzystaj ze stopnia wchodzącego	389
Druga metoda wykrywania elementów odstających grafu	
— zgłębianie niuansów za pomocą k-odległości	393
Trzecia metoda wykrywania elementów odstających grafu	
— lokalny miernik stopnia oddalenia obserwacji	395
I I Przejście z arkusza kalkulacyjnego do języka R	401
Przygotowanie środowiska i początek pracy w języku R	402
Szybkie szkolenie z pisanie skryptów w języku R	403
Działania matematyczne na wektorach i faktory	405
Najlepszy typ danych — dataframe	408
Pomoc dla języka R	409
Wyjście poza podstawowe możliwości R	410
Prawdziwa analiza danych	412
Wczytywanie danych do R	412
Sferyczny algorytm k-średnich wywołany	
za pomocą zaledwie kilku linii kodu	413
Budowanie modeli sztucznej inteligencji	
na podstawie danych zakupów (wykrywanie ciąży)	418
Prognozowanie w R	426
Wykrywanie elementów odstających	430
12 Wnioski	435
Gdzie ja jestem? Co się stało?	435
Zanim odłożysz tę książkę	435
Poznaj problem	436
Potrzebujemy więcej tłumaczy	436
Uważaj na trójgłowe monstrum: narzędzia,	
wydajność i perfekcjonizm	437
Nie jesteś najważniejszą osobą w firmie	440
Bądź kreatywny	440