

Spis treści

O autorze	11
O konsultacjach merytorycznych	11
Podziękowania	12
Wstęp	13
<i>Co ja tutaj robię?</i>	13
<i>Praktyczna definicja analizy danych</i>	14
<i>Chwila, chwila... A co z tym data?</i>	15
<i>Kim jestem?</i>	16
<i>Kim jeste?</i>	16
<i>Nie wszystko byłoby pracować tylko w arkuszu kalkulacyjnym</i>	17
<i>Ala arkusze kalkulacyjne są takie staremodne!</i>	18
<i>Konieczność programu Excel lub pakietu LibreOffice</i>	18
<i>Konwencje typograficzne przyjęte w tej książce</i>	19
<i>Zaczynamy</i>	20
1. Wskazówki, co chciałbyś wiedzieć o arkuszu kalkulacyjnym, ale bałeś się o to zapytać	21
<i>Przykładowe proste dane</i>	22
<i>Szybkie przeglądanie arkusza i filarów Excel</i>	23
<i>Szybkie kopiowanie danych i format</i>	24
<i>Formatowanie komórek</i>	26
<i>Wklejanie wartości specjalnych</i>	27
<i>Wstawianie wykresów</i>	28
<i>Mowa Znajdź i mowa Zmień</i>	29

Formuły przekształceń do wyznakiwania i wyzyskiwania wartości	30
Stosowanie formuły $\text{N}(\text{X})\text{Z}(\text{U},\text{K},\text{A})\text{P}(\text{K},\text{O},\text{O},\text{H},\text{H},\text{O})$ do łączenia danych	32
Filtrowanie i sortowanie	33
Stosowanie tabel przestawnych	36
Korzystanie z formuł tabliczowych	39
Rozwiązywanie problemów za pomocą narzędzia Solver	40
Opisulstwo — chciałbym, abyśmy go nie potrzebowali, ale...	46
Podsumowanie	47
2. Analiza skupień. Część I — zastosowanie algorytmu centroidów do segmentowania bazy klientów	49
Dławiący temat z dwiema nazwami, a skupię skupię się po kolei	51
Przedziwy problem: implementacja algorytmu centroidów w e-mail marketingu	56
Handel w internecie	56
Początkowy zbiór danych	57
Określenie tego, co chcemy mierzyć	57
Zacznij od czterech grup	61
Odległość rachunkowa — pomiar odległości w linii prostej	61
Odległość dla wszystkich	64
Określenie położenia środków klastrów	66
Analiza uzyskanych wyników	68
Ustalenie najlepszej oferty dla danego klienta	69
Słowniki podziału — dobry sposób na określenie optymalnej liczby klastrów	74
A może potrzebujesz pięciu klastrów?	80
Dzielenie klientów na pięć klastrów za pomocą narzędzia Solver	81
Ustalenie najlepszych ofert dla wszystkich pięciu klastrów	82
Określenie słowniki podziału na pięć klastrów	85
Podział na grupy za pomocą algorytmu k-medoidów i asymetryczny pomiar odległości	87
Podział na grupy za pomocą metody k-medoidów	87
Stosowanie innego sposobu pomiaru odległości	87
Implementacja za pomocą Excela	90
Najlepsze oferty przy podziale na pięć klastrów za pomocą Excela	92
Podsumowanie	93
3. Mały klasyfikator bayesowski i nieszyfrowana lekkość języka ludzkiego	97
Jakiś mały, ale swój produkt Microsoft, to wykonał za pomocą informacji zawartej	97
Najpierw na świecie sprawdzanie do rachunku prawdopodobieństwa	100
Obliczenie prawdopodobieństwa warunkowego	100
Prawdopodobieństwo części wspólnej: reguła Bayesa i niezależność	101
A co, jeżeli sytuacja się zmienia od czasu?	102
Tworzenie Bayesa	102

Tworzenie modelu sztucznej inteligencji za pomocą twierdzenia Bayesa	183
Zwykle zakłada się, że występowanie pewnego zdarzenia klasę są sobie równe	185
Kilka innych doświadczeń	186
Czas rozpoznaj i zabawy z Raczem	187
Utworzenie nieliniowych modeli interpretacyjnych	188
Dzielenie na próbki spoży	189
Eksperymenty i obliczanie prawdopodobieństwa	192
Zbudowanie modelu. Skracający i ciągły	194
Podsumowanie	199
4. Modelowanie optymalizacyjne — „Zwierzę wyćnięte” tak nie zamieszka się sam	123
Dlaczego analiza danych powinna wiedzieć, czym jest optymalizacja?	124
Zaczynamy od prostego komputera	125
Przekształcenie problemu w formie matematycznej	126
Rozwiązanie problemu poprzez przetwarzanie postawki	128
Metoda simplex — dlaczego się walczy ciągle	129
Praca w Excelu	130
Na koniec tego rozdziału będzie się potwie	140
Solówka (zwierzę) celia pomiarowego	
prosto z drzewa... z przystankiem na modelowanie	141
Tworzenie skryptów z modelu nieliniowego	142
Zaczynamy od specyfikacji celów	142
Składowe produkty wyćnięte	144
Wypracowanie danych do Excela	145
Określenie problemu w Excelu Solver	148
Obliczanie standardów	150
Utworzenie rachunkowego problemu — minimalizacja maksymalnych odchyleń	154
Warunki i ograniczenia „zwierzę M”	156
Modelowanie zmiennych — skracający do 100% mocy Racz	160
Modelowanie ryzyka	166
Dane pochodzące z rozkładu normalnego	168
Podsumowanie	176
5. Analiza skupień. Część II — grafy i analiza sieci	179
Czym jest graf sieci?	180
Wizualizacja prostego grafu	181
Krótka wprowadzenie do Gephi	184
Instalacja Gephi i przygotowanie pliku	184
Budowa grafu	185
Stopień rozgałęzienia	188
Elegancki wybór	189
Edycja danych grafu	192

Tworzenie grafu na podstawie danych sprzedaży wina	193
Tworzenie macierzy podobieństwa binarnowego	193
Generowanie grafu z sąsiedztwa	197
<i>Jaka jest wartość krawędzi? Krawędźami i krawędziowymi — modułami grafu</i>	202
<i>Czy jest punkt, a czy linia?</i>	202
Tworzenie arkusza punktacji	206
Czas dokonać podziału na grupy	208
Podział 1.	209
Podział 2. — krawędzi	214
Podział 3. — centra	219
Grupy — kodowanie i analiza	219
Tam i z powrotem — czas na Gplot	220
Podsumowanie	223
6. Regresja jako przedmiot nadzorowanego uczenia maszynowego i sztucznej inteligencji	227
<i>Cel: linia w ciępi?</i>	227
<i>Nie-oczekuj sukcesu</i>	228
Przewidywanie ciągły klientów na podstawie regresji liniowej	229
Zbiór cech	230
Tworzenie treningowego zbioru danych	232
Tworzenie testowych filarych	232
Polowanie się regresji liniowej	233
Parametry regresji liniowej: współczynnik determinacji, test F i t test	244
Przewidywanie ciągły na nowym zbiorze danych i sprawdzanie jakości modelu	251
Przewidywanie ciągły klientów za pomocą regresji logistycznej	260
Najpierw model: określ funkcję wiążącą	263
Tworzenie funkcji logistycznej i pomocna optymalizacja	266
Praca nad przewidywaniem regresji logistycznej	270
Wybór modelu — porównywanie skuteczności regresji liniowej i regresji logistycznej	272
Dobry kłótnia	274
Podsumowanie	275
7. Modele zespołowe — dużo nie najlepszego giczy	277
Kompozycja z danych z rezultatu 6.	278
Agregacja — linia, linia, powód	280
Planek decyzyjny to niekiedy ładnie stworzenie głębokiego modelu	280
To może nie wygląda się tak głęboko	281
Wgryz maczy	283
Czas rozpocząć proces trenowania	284
Cena działania modelu zespołowego	293

Wymaganie — jeżeli wyliczeń nieustępliwość jest wymagana	258
Jeżeli wymagany jest model i uwzględnia go jeszcze raz	258
Zmiana modelu — jeżeli trzeba ma więcej niż jeden	259
Wyliczenia modelu w ramach reguł decyzyjnych	267
Podsumowanie	271
II. Prognozowanie — oddechaj spokojnie, i tak nie wygrasz	313
Masa na rynku sprzedaży młodych	314
Szeregi czasowe	315
Zaczaj od prostego wykładniczego wykładniczego	317
Przygotowanie arkusza pracy prostego wykładniczego wykładniczego	318
Być może dane zawierają trend	323
Podwójne wykładnicze wykładnicze (metoda Holt)	327
Metoda Holt w arkusza kalkulacyjnym	329
To wszystko? Analiza autokorelacji	335
Wielokrotne wykładnicze wykładnicze — model Holt-Wintnera	342
Określenie początkowych wartości poziomu, trendu i sezonowości	343
Tworzenie prognozy	349
Czas na optymalizację	354
Pierwsza część, że to jest koniec. Przewidywanie	356
Interwały prognozy	356
Tworzenie wykresu wariancji wariancji wartości	360
Podsumowanie	362
3. Wykrywanie obserwacji odstających — to, że jakiś element jest inny od pozostałych, nie oznacza, że jest nieistotny	365
Element odstający to też obserwacja	366
Zaczynając od pracy Flakamala	367
Metoda Tukeya	368
Implementacja metody Tukeya w arkusza kalkulacyjnym	368
Ograniczenia tej prostej techniki	371
Nie bójmy się, ale słuchajmy uważnie	372
Przygotowanie danych do utworzenia wykresu	373
Tworzenie grafu	376
Określenie i naciśnięcie spacji	378
Pierwsza metoda wykrywania elementów odstających grafu	
— skorysuj te dane wchodzącego	379
Druga metoda wykrywania elementów odstających grafu	
— złączanie elementów na pomocą i odległości	383
Trzecia metoda wykrywania elementów odstających grafu	
— lokalny wskaźnik stopnia odległości obserwacji	385
Podsumowanie	391

