

Spis treści

O redaktorach technicznych	ix
Podziękowania	xi
Przedmowa	xxi
Wprowadzenie	xxv

CZĘŚĆ I

Myśl jak spec od danych

ROZDZIAŁ 1

Na czym polega problem? 3

Pytania, które powinien zadawać spec od danych	4
<i>Dlaczego problem jest ważny?</i>	4
<i>Na kogo wpływa ten problem?</i>	6
<i>Co, jeśli nie mamy właściwych danych?</i>	7
<i>Kiedy projekt się zakończy?</i>	7
<i>Co, jeśli nie spodobają nam się rezultaty?</i>	7
Dlaczego projekty związane z danymi kończą się niepowodzeniem?	8
<i>Wrażenia klientów</i>	8
<i>Omówienie</i>	10
Praca nad problemami, które mają znaczenie	11
Podsumowanie rozdziału	11

ROZDZIAŁ 2

Czym są dane? 13

Dane a informacje	13
<i>Przykładowy zbiór danych</i>	14
Typy danych	15
Jak gromadzi się dane i jaką mają strukturę?	16
<i>Dane obserwacyjne i eksperymentalne</i>	16
<i>Dane ustrukturyzowane i nieustrukturyzowane</i>	17
Podstawowe statystyki zbiorcze	18
Podsumowanie rozdziału	19

ROZDZIAŁ 3

Przygotowanie do myślenia statystycznego	21
Zadawaj pytania	22
Wszystko jest zmienne	23
<i>Scenariusz: wrażenia klientów (kontynuacja)</i>	24
<i>Studium przypadku: zachorowalność na raka nerki</i>	26
Prawdopodobieństwo i statystyka	28
<i>Prawdopodobieństwo a intuicja</i>	29
<i>Odkrywanie informacji za pomocą statystyki</i>	31
Podsumowanie rozdziału	33

CZĘŚĆ II**Mów jak spec od danych****ROZDZIAŁ 4**

Polemizuj z danymi	37
Co byś zrobił(a)?	38
<i>Katastrofa spowodowana brakiem danych</i>	39
Jaka jest historia pochodzenia danych?	43
<i>Kto zebrał dane?</i>	44
<i>Jak zebrano dane?</i>	44
Czy dane są reprezentatywne?	45
<i>Czy poprawnie dobrano próbę?</i>	46
<i>Co zrobiono z wartościami odstającymi?</i>	46
Jakich danych nie widzę?	47
<i>Jak rozwiązano problem brakujących wartości?</i>	47
<i>Czy dane mogą zmierzyć to, co ma być mierzone?</i>	48
Polemizuj z danymi każdej wielkości	48
Podsumowanie rozdziału	49

ROZDZIAŁ 5

Eksploruj dane	51
Eksploracyjna analiza danych i Ty	52
Przyjmij nastawienie eksploracyjne	52
<i>Pytania naprowadzające</i>	53
<i>Scenariusz</i>	53
Czy dane mogą odpowiedzieć na pytanie?	54
<i>Określ oczekiwania i użyj zdrowego rozsądku</i>	54
<i>Czy wartości mają intuicyjny sens?</i>	54
<i>Uważaj! Wartości odstające i brakujące</i>	58

Czy odkryliście jakieś związki?	58
<i>Korelacja</i>	59
<i>Uważaj! Błędne interpretowanie korelacji</i>	60
<i>Uważaj! Korelacja nie implikuje przyczynowości</i>	62
Czy znaleźliście w danych nowe możliwości?	63
Podsumowanie rozdziału	63

ROZDZIAŁ 6

Badaj prawdopodobieństwa	65
Zgadnij odpowiedź	66
Reguły gry	67
<i>Notacja</i>	67
<i>Prawdopodobieństwo warunkowe i zdarzenia niezależne</i>	69
<i>Prawdopodobieństwo wielu zdarzeń</i>	69
Dwie rzeczy, które zdarzają się razem	69
Jedno albo drugie	71
Ćwiczenie myślowe z zakresu prawdopodobieństwa	72
<i>Następne kroki</i>	73
Uważaj z zakładaniem niezależności	74
<i>Nie popełniaj błędu hazardzysty</i>	75
Wszystkie prawdopodobieństwa są warunkowe	75
<i>Nie przedstawiaj zależności</i>	76
<i>Twierdzenie Bayesa</i>	77
Upewnij się, że prawdopodobieństwa mają znaczenie	80
<i>Kalibracja</i>	80
<i>Rzadkie zdarzenia mogą się zdarzać i się zdarzają</i>	80
Podsumowanie rozdziału	81

ROZDZIAŁ 7

Kwestionuj statystyki	83
Krótkie lekcje o wnioskowaniu	83
<i>Zostaw sobie trochę przestrzeni</i>	84
<i>Więcej danych, więcej dowodów</i>	84
<i>Kwestionuj status quo</i>	85
<i>Dowody na twierdzenie przeciwne</i>	86
<i>Równoważenie błędów decyzyjnych</i>	88
Proces wnioskowania statystycznego	89
Pytania, które pomogą Ci kwestionować statystyki	90
<i>Jaki jest kontekst tych statystyk?</i>	90
<i>Jaki jest rozmiar próby?</i>	91
<i>Co testujecie?</i>	91

<i>Jaka jest hipoteza zerowa?</i>	92
<i>Zakładanie równoważności</i>	93
<i>Jaki jest poziom istotności?</i>	93
<i>Ile przeprowadzacie testów?</i>	94
<i>Czy mogę zobaczyć przedziały ufności?</i>	94
<i>Czy jest to praktycznie istotne?</i>	96
<i>Czy zakładacie przyczynowość?</i>	96
Podsumowanie rozdziału	97

CZĘŚĆ III

Przybornik specjalisty data science

ROZDZIAŁ 8

W poszukiwaniu ukrytych grup	101
Uczenie nienadzorowane	102
Redukcja wymiarowości	102
<i>Tworzenie cech złożonych</i>	103
Analiza składowych głównych	105
<i>Składowe główne zdolności sportowych</i>	105
<i>Podsumowanie PCA</i>	108
<i>Potencjalne pułapki</i>	109
Klasteryzacja	109
Klasteryzacja metodą k-średnich	111
<i>Klasteryzacja sklepów detalicznych</i>	111
<i>Potencjalne pułapki</i>	113
Podsumowanie rozdziału	114

ROZDZIAŁ 9

Model regresji	117
Uczenie nadzorowane	117
Jak działa regresja liniowa?	119
<i>Regresja metodą najmniejszych kwadratów:</i>	
<i>nie tylko pomysłowa nazwa</i>	120
Regresja liniowa: co Ci daje?	123
<i>Rozszerzanie modelu na wiele cech</i>	124
Regresja liniowa: jakie powoduje nieporozumienia?	125
<i>Pominięte zmienne</i>	126
<i>Współliniowość</i>	126
<i>Przeciek danych</i>	127
<i>Błędy ekstrapolacji</i>	128
<i>Relacje nie zawsze są liniowe</i>	128

Wyjaśniasz czy przewidujesz?	128
Skuteczność regresji	130
Inne modele regresji	131
Podsumowanie rozdziału	131

ROZDZIAŁ 10

Model klasyfikacji 133

Wprowadzenie do klasyfikacji	133
<i>Czego się nauczysz?</i>	134
<i>Przykładowy problem klasyfikacji</i>	135
Regresja logistyczna	135
<i>Regresja logistyczna — i co z tego?</i>	138
Drzewa decyzyjne	139
Metody zespołowe	142
<i>Lasy losowe</i>	143
<i>Drzewa wzmacniane gradientowo</i>	143
<i>Interpretowalność modeli zespołowych</i>	145
Strzeż się pułapek	145
<i>Złe podejście do problemu</i>	146
<i>Przeciek danych</i>	146
<i>Brak podziału danych</i>	146
<i>Wybór odpowiedniego progu decyzyjnego</i>	147
Błędne rozumienie dokładności	147
<i>Macierze błędów</i>	148
Podsumowanie rozdziału	150

ROZDZIAŁ 11

Analiza tekstu 151

Oczekiwania wobec analizy tekstu	151
Jak tekst staje się liczbami	153
<i>Wielki worek słów</i>	153
<i>N-gramy</i>	157
<i>Osadzenia słów</i>	158
Modelowanie tematyczne	160
Klasyfikacja tekstu	162
<i>Naiwny klasyfikator bayesowski</i>	164
<i>Analiza odczuć</i>	166
Kwestie praktyczne podczas pracy z tekstem	167
<i>Giganci technologiczni mają przewagę</i>	168
Podsumowanie rozdziału	169

ROZDZIAŁ 12

Uczenie głębokie	171
Sieci neuronowe	172
<i>Pod jakimi względami sieci neuronowe przypominają ludzki mózg?</i>	172
<i>Prosta sieć neuronowa</i>	173
<i>Jak uczy się sieć neuronowa?</i>	174
<i>Nieco bardziej złożona sieć neuronowa</i>	175
Zastosowania uczenia głębokiego	178
<i>Korzyści z uczenia głębokiego</i>	179
<i>Jak komputery „widzą” obrazy?</i>	180
<i>Konwolucyjne sieci neuronowe</i>	181
<i>Uczenie głębokie w języku i sekwencjach</i>	183
Uczenie głębokie w praktyce	184
<i>Czy masz dane?</i>	184
<i>Czy Twoje dane są ustrukturyzowane?</i>	185
<i>Jak będzie wyglądać sieć?</i>	186
Sztuczna inteligencja i Ty	187
<i>Giganci technologiczni mają przewagę</i>	188
<i>Etyka w uczeniu głębokim</i>	188
Podsumowanie rozdziału	189

CZĘŚĆ IV**Droga do sukcesu****ROZDZIAŁ 13**

Strzeż się pułapek	193
Tendencyjność i dziwne zjawiska w danych	194
<i>Błąd przeżywalności</i>	194
<i>Regresja do średniej</i>	195
<i>Paradoks Simpsona</i>	195
<i>Błąd konfirmacji</i>	197
<i>Błąd utopionych kosztów</i>	197
<i>Dyskryminacja algorytmiczna</i>	197
<i>Nieskategoryzowane przejawy tendencyjności</i>	198
Wielka lista pułapek	199
<i>Pułapki związane ze statystyką i uczeniem maszynowym</i>	199
<i>Pułapki związane z projektem</i>	200
Podsumowanie rozdziału	202

ROZDZIAŁ 14

Znaj ludzi i osobowości	203
Siedem scenariuszy fiaska komunikacyjnego	204
<i>Post mortem</i>	204
<i>Wieczorynka</i>	205
<i>Głuchy telefon</i>	206
<i>W gąszczu szczegółów</i>	206
<i>Konfrontacja z rzeczywistością</i>	207
<i>Wrogie przejęcie</i>	207
<i>Egocentryk</i>	207
Osobowości w świecie danych	208
<i>Entuzjasta</i>	208
<i>Cynik</i>	209
<i>Spec od danych</i>	209
Podsumowanie rozdziału	209

ROZDZIAŁ 15

Co dalej?	211
------------------	------------