

Przedmowa	11
Część I. Techniki analityki danych	17
1. I co z tego? Generowanie wartości dzięki danologii	19
Czym jest wartość?	19
„Co?”, czyli zrozumieć biznes	20
„Co z tego?”, czyli istota generowania wartości dzięki danologii	22
„Co teraz?”, czyli bądź przebojowy	23
Pomiar wartości	23
Najważniejsze wnioski	25
Dalsza lektura	25
2. Projektowanie wskaźników	27
Pożądane właściwości wskaźników	27
Mierzalność	27
Możliwość podejmowania działań	28
Trafność	28
Aktualność	28
Dekompozycja wskaźników	29
Lejek analityczny	29
Dekompozycje przepływów i zapasów	30
Dekompozycje typu $P \times Q$	30
Przykład: inny sposób dekompozycji przychodów	31
Przykład: platformy sprzedażowe	31
Najważniejsze wnioski	32
Dalsza lektura	32

3. Dekompozycje wzrostu — zrozumienie przeszkód i sprzyjających czynników	34
Dlaczego dekompozycje wzrostu?	34
Dekompozycja addytywna	34
Przykład	35
Interpretacja i przypadki użycia	36
Dekompozycja multiplikatywna	36
Przykład	37
Interpretacja	37
Dekompozycja zmian wag i wartości	38
Przykład	39
Interpretacja	40
Wyprowadzanie równań matematycznych	40
Dekompozycja addytywna	41
Dekompozycja multiplikatywna	41
Dekompozycja mix-rate	41
Najważniejsze wnioski	42
Dalsza lektura	42
4. Projekty 2x2	43
Argumenty za upraszczaniem	43
Czym jest projekt 2x2?	44
Przykład: testowanie modelu i nowej cechy	45
Przykład: zrozumienie zachowań użytkownika	47
Przykład: udzielanie i akceptacja ofert kredytów	49
Przykład: ustalanie priorytetów w procesie pracy	50
Najważniejsze wnioski	51
Dalsza lektura	51
5. Tworzenie uzasadnienia biznesowego	53
Wybrane zasady tworzenia uzasadnień biznesowych	53
Przykład: proaktywna strategia zatrzymywania klientów	54
Zapobieganie oszustwom	55
Zakup zewnętrznych zbiorów danych	56
Praca nad projektem z obszaru danologii	57
Najważniejsze wnioski	57
Dalsza lektura	58
6. Czym jest wskaźnik przyrostu?	59
Definicja wskaźnika przyrostu	59
Przykład: model klasyfikatora	60
Błędy wynikające z samoselekcji i przeżywalności	60

Inne zastosowania wskaźników przyrostu	62
Najważniejsze wnioski	62
Dalsza lektura	63
7. Narracje	64
Co kryje się w narracji? Opowiadanie historii za pomocą danych	64
Jasna i rzeczowa	65
Wiarygodność	66
Zapadająca w pamięć	67
Możliwość podejmowania działań	68
Tworzenie narracji	68
Nauka jako opowiadanie historii	68
„Co?”, „co z tego?” i „co teraz?”	69
Ostatnia prosta	71
Streszczenia TL;DR	71
Wskazówki dotyczące pisania zapadających w pamięć streszczeń TL;DR	71
Przykład: pisanie streszczenia TL;DR tego rozdziału	72
Skuteczne krótkie prezentacje	74
Prezentowanie narracji	74
Najważniejsze wnioski	75
Dalsza lektura	76
8. Wizualizacje danych — wybór właściwego wykresu do przekazania komunikatu	77
Kilka przydatnych i rzadko używanych wizualizacji danych	77
Wykres słupkowy a wykres liniowy	77
Wykres nachylenia	79
Wykres kaskadowy	79
Funkcje wygładzania dla wykresów punktowych	81
Prezentowanie rozkładów na wykresie	82
Ogólne zalecenia	83
Dobierz odpowiednią wizualizację dla przekazu	83
Mądrze dobieraj kolory	84
Różne wymiary na wykresie	85
Staraj się uzyskać odpowiednio wysoki współczynnik dane/atrament	85
Personalizacja a półautomatyzacja	86
Na samym początku dobierz odpowiedni rozmiar czcionki	86
Interaktywne czy nie?	86
Zachowaj prostotę	87
Zacznij od wyjaśnienia wykresu	87
Najważniejsze wnioski	87
Dalsza lektura	88

9. Symulacje i bootstrapping	91
Podstawy symulacji	92
Symulacja modelu liniowego i regresji liniowej	94
Czym są wykresy zależności częściowych?	96
Błąd systematyczny z powodu pominięcia zmiennej	100
Symulacja problemu klasyfikacji	103
Modele zmiennych ukrytych	103
Porównanie różnych algorytmów	104
Bootstrapping	106
Najważniejsze wnioski	108
Dalsza lektura	108
10. Regresja liniowa — powrót do podstaw	110
Co kryje się za współczynnikiem?	110
Twierdzenie Frischa-Waugh-Lovella	113
Dlaczego twierdzenie FWL jest ważne?	115
Czynniki zakłócające	116
Dodatkowe zmienne	118
Centralna rola wariancji w uczeniu maszynowym	120
Najważniejsze wnioski	123
Dalsza lektura	124
11. Wyciekanie danych	125
Czym jest wyciekanie danych?	125
Wynik również jest cechą	125
Funkcja wyniku sama też jest cechą	126
Złe zmienne kontrolne	126
Niewłaściwe oznaczenie znacznika czasu	126
Wiele zbiorów danych z nieprecyzyjnymi agregacjami czasowymi	127
Wyciekanie innych informacji	127
Wykrywanie wyciekania danych	128
Całkowita separacja	130
Metoda okien	132
Wybór długości okien	133
Etap treningu odzwierciedla etap oceny punktowej	134
Wdrażanie metody okien	135
Mam wyciek. Co teraz?	136
Najważniejsze wnioski	136
Dalsza lektura	137

12. Stosowanie modeli w środowisku produkcyjnym	138
Co oznacza „gotowość produkcyjna”?	138
Wsadowa ocena punktowa (w trybie offline)	138
Obiekty modeli czasu rzeczywistego	140
Dryf danych i modelu	141
Etapy niezbędne w każdym potoku produkcyjnym	142
Pobieranie i przekształcanie danych	143
Sprawdzanie poprawności danych	144
Etapy treningu i oceny punktowej	145
Sprawdzanie poprawności modelu i ocen punktowych	146
Zapisywanie modelu i ocen punktowych	146
Najważniejsze wnioski	146
Dalsza lektura	147
13. Opowiadanie historii w uczeniu maszynowym	149
Holistyczne spojrzenie na opowiadanie historii w uczeniu maszynowym	149
Opowiadanie historii przed opracowaniem modelu i w trakcie tego procesu	150
Tworzenie hipotez	151
Inżynieria cech	153
Opowiadanie historii po opracowaniu modelu: otwieranie czarnej skrzynki	156
Kompromis między interpretowalnością a skutecznością	156
Regresja liniowa: ustalenie punktu odniesienia	158
Znaczenie cech	160
Mapa cieplna	161
Wykresy zależności częściowych	163
Skumulowane efekty lokalne	165
Najważniejsze wnioski	166
Dalsza lektura	167
14. Od predykcji do decyzji	168
Analiza procesu podejmowania decyzji	168
Proste reguły decyzyjne oparte na inteligentnym wyznaczaniu wartości progowych	170
Precyzja i czułość	171
Przykład: pozyskiwanie list kontaktów	172
Optymalizacja macierzy błędów	174
Najważniejsze wnioski	176
Dalsza lektura	176
15. Zmiany dodatkowe — Święty Graal danologii?	177
Definiowanie zmian dodatkowych	177
Wnioskowanie przyczynowe w celu poprawy predykcji	178
Wnioskowanie przyczynowe jako wyróżnik	178
Usprawnione podejmowanie decyzji	179

Czynniki zakłócające i kolidery	179
Błąd doboru	182
Założenie o braku zmiennych zakłócających	186
Radzenie sobie z błędem doboru — randomizacja	187
Dopasowywanie	188
Uczenie maszynowe i wnioskowanie przyczynowe	191
Kod otwartoźródłowy	191
Podwójne uczenie maszynowe	192
Najważniejsze wnioski	194
Dalsza lektura	195
16. Testy A/B	198
Czym są testy A/B?	198
Kryterium decyzyjne	199
Minimalne wykrywalne efekty	202
Ustalanie mocy statystycznej, poziomu istotności i wartości P	205
Szacowanie wariancji wyniku	205
Symulacje	206
Przykład: współczynniki konwersji	207
Określanie wartości MWE	208
Lista hipotez do zbadania	209
Wskaźnik	209
Hipoteza	210
Uszeregowanie	210
Zarządzanie eksperymentami	210
Najważniejsze wnioski	211
Dalsza lektura	212
17. Modele LLM i praktyka danologii	213
Obecny stan sztucznej inteligencji	213
Czym zajmują się danologowie?	215
Ewolucja opisu stanowiska danologa	217
Studium przypadku: testy A/B	218
Studium przypadku: oczyszczanie danych	219
Studium przypadku: uczenie maszynowe	219
Modele LLM a ta książka	220
Najważniejsze wnioski	221
Dalsza lektura	222
Skorowidz	223