

Spis treści

O autorze	7
O recenzentach	8
Przedmowa	9
Rozdział 1. Klasyfikacja na podstawie najbliższego sąsiedztwa	13
Subiektywne odczuwanie temperatury	14
Implementacja algorytmu k najbliższych sąsiadów	16
Mapa Włoch — przykład doboru wartości k	20
Analiza	21
Skalowanie danych — prognozowanie statusu własności	25
Analiza	26
Nieeuklidesowe metryki odległości punktów — klasyfikowanie tekstów	27
Analiza	28
Klasyfikowanie tekstów ciąg dalszy — wielowymiarowy algorytm k-NN	30
Analiza	31
Podsumowanie	32
Problemy	33
Subiektywne odczuwanie temperatury	33
Mapa Włoch — przykład doboru wartości k	33
Status własności	34
Analiza	34
Rozdział 2. Naiwny klasyfikator bayesowski	37
Testy medyczne — podstawowe zastosowanie twierdzenia Bayesa	38
Analiza	38
Podstawowe twierdzenie Bayesa i jego rozszerzenie	39
Twierdzenie Bayesa	39
Rozszerzone twierdzenie Bayesa	40
Zagramy w szachy? — niezależne zdarzenia warunkujące	41
Analiza	42
Implementacja naiwnego klasyfikatora bayesowskiego	43
Zagramy w szachy? — częściowo zależne zdarzenia warunkujące	45
Analiza	45
Chłopak czy dziewczyna? — twierdzenie Bayesa dla ciągłych zmiennych losowych	48
Analiza	48

Podsumowanie	50
Problemy	51
Analiza	53
Rozdział 3. Drzewa decyzyjne	59
Pływamy? — reprezentowanie danych w postaci drzewa decyzyjnego	59
Elementy teorii informacji	61
Entropia informacyjna	61
Zysk informacyjny	63
Pływamy? — obliczanie zysku informacyjnego	63
Algorytm ID3 — konstruowanie drzewa decyzyjnego	65
Pływamy? — budowanie drzewa decyzyjnego	65
Implementacja w języku Python	66
Klasyfikowanie danych za pomocą drzew decyzyjnych	71
Przykład — pływamy czy nie?	72
Przykład — gra w szachy pod chmurką	72
Analiza	72
Na zakupy — przykład niespójnych danych	77
Analiza	77
Podsumowanie	78
Problemy	78
Analiza	80
Rozdział 4. Lasy losowe	83
Ogólne zasady konstruowania lasów losowych	84
Pływamy? — klasyfikacja za pomocą lasu losowego	84
Analiza	84
Konstruowanie lasu losowego	85
Klasyfikowanie cechy na podstawie lasu losowego	89
Implementacja algorytmu konstruowania lasu losowego	90
Przykład — zagramy w szachy?	93
Analiza	93
Konstruowanie lasu losowego	94
Klasyfikacja w drodze głosowania	99
Idziemy na zakupy? — wnioskowanie z niespójnych danych i miara wiarygodności wyniku	100
Analiza	100
Podsumowanie	101
Problemy	102
Analiza	103
Rozdział 5. Klasteryzacja	107
Dochód gospodarstwa domowego — niski czy wysoki?	107
Algorytm k-średnich	108
Początkowy zbiór centroidów	109

Wyznaczanie centroidu klastra	109
Przykład — wykorzystanie algorytmu k-średnich do klasyfikacji dochodów	110
Klasyfikowanie przez klasteryzację — prognozowanie płci nieznanej osoby	111
Analiza	112
Implementacja algorytmu k-średnich	115
Status własności — dobór optymalnej liczby klastrów	118
Analiza	119
Klasyfikowanie dokumentów — semantyczne znaczenie klasteryzacji	125
Analiza	126
Podsumowanie	132
Problemy	132
Analiza	133
Rozdział 6. Analiza regresji	143
Konwersja temperatur — regresja liniowa dla danych doskonałych	144
Rozwiązanie analityczne	144
Metoda najmniejszych kwadratów w regresji liniowej	145
Implementacja analizy regresji liniowej w Pythonie	146
Regresja dla danych pomiarowych — prognozowanie wagi na podstawie wzrostu	149
Analiza	149
Metoda spadku gradientowego i jej implementacja	151
Szczegóły algorytmu	151
Implementacja w Pythonie	152
Przewidywanie czasu przelotu na podstawie odległości	154
Analiza	154
Obliczenia balistyczne — model nieliniowy	156
Analiza	156
Podsumowanie	158
Problemy	158
Analiza	159
Rozdział 7. Analiza szeregów czasowych	163
Zysk w biznesie — analiza trendu	164
Analiza	164
Konkluzja	165
Sprzedaż w sklepie internetowym — analiza sezonowości	166
Analiza	166
Analiza trendu	166
Analiza sezonowości	169
Podsumowanie	175
Problemy	176
Analiza	177

Dodatek A. Podstawy języka Python	181
Przykład	181
Komentarze	182
Typy danych	182
int	182
float	183
Napis	183
Krotka	183
Lista	184
Zbiór	184
Słownik	185
Przepływ sterowania	185
Instrukcje warunkowe	186
Pętla for	187
Pętla while	188
Instrukcje break i continue	189
Funkcje	190
Wejście-wyjście programu	192
Argumenty wywołania programu	192
Operacje na plikach	192
Dodatek B. Statystyka	195
Podstawowe koncepcje	195
Notacja	195
Podstawowe pojęcia	195
Wnioskowanie bayesowskie	197
Rozkład normalny Gaussa	197
Walidacja krzyżowa	197
Testowanie A/B	198
Dodatek C. Słownik pojęć, algorytmów i metod Data Science	199
Skorowidz	203