

Spis treści

Wstęp	7
Rozdział 1. Rozumienie Data Science (<i>Grażyna Trzpiot</i>)	9
1.1. Co oznacza Data Science?	9
1.1.1. Co wyróżnia Data Science?	9
1.2. Implikacje Data Science	11
1.3. Procedury Data Science	11
1.4. Składowe Data Science	16
1.4.1. Typy danych	16
1.4.2. Klasy modeli analitycznych	18
1.4.3. Modele uczące się	19
1.4.4. Modele wykonawcze	20
1.5. Wybór modelu analitycznego	22
1.5.1. Dekompozycja problemu	22
1.5.2. Ograniczenia implementacji	23
1.5.3. Wybór metody analitycznej	24
Rozdział 2. Obserwacje odstające w (dużych) zbiorach danych (<i>Justyna Majewska</i>)	31
2.1. Przygotowanie danych do analiz	31
2.2. Obserwacje odstające	34
2.2.1. Czym jest obserwacja odstająca?	34
2.2.2. Problemy i wyzwania występowania obserwacji odstających w dużych zbiorach danych	36
2.3. Różne aspekty problemu identyfikacji obserwacji odstających	37
2.4. Przegląd metod identyfikacji obserwacji odstających w zbiorach danych	39
2.4.1. Metody klasyfikacji	39
2.4.2. Metody najbliższego sąsiada	42
2.4.3. Metody grupowania	43
2.4.4. Metody statystyczne	44
2.4.5. Metody spektralne	45
2.4.6. Identyfikacja anomalii kontekstowych	46
2.5. Obszary zastosowań metod identyfikacji obserwacji odstających	47
Rozdział 3. Modele regresji (<i>Grażyna Trzpiot</i>)	51
3.1. Klasyczne modelowanie zmiennych	51
3.1.1. Standardowy model regresji liniowej	51
3.1.2. Klasyczny model regresji wielowymiarowej	52

3.1.3. Model logitowy dla binarnej zmiennej objaśnianej	53
3.1.4. Mieszany liniowy model regresji	54
3.1.5. Nieparametryczny model regresji	55
3.1.6. Standardowy addytywny model regresji	55
3.1.7. Addytywny model regresji Poissona	56
3.1.8. Przestrzenny model regresji	56
3.1.9. Modelowanie a modele regresji	56
3.2. Rozszerzenia liniowego modelu regresji	59
3.3. GLM - uogólnione modele liniowe	60
3.3.1. Binarne modele regresji	60
3.3.2. Interpretacja modelu logitowego	61
3.3.3. Log-liniowy model Poissona	62
3.3.4. Uogólniony liniowy model	62
3.4. Regulowana regresja	64
3.4.1. Regresja grzbietowa (<i>Ridge Regression</i>)	65
3.4.2. Regresja LASSO (<i>Least Absolute Shrinkage and Selection Operator</i>)	65
Rozdział 4. Redukcja wymiaru (<i>Dominik Krężolek</i>)	67
4.1. Klasyczna analiza czynnikowa	68
4.2. Metoda głównych składowych	77
4.3. Metody grupowania	79
4.3.1. Analiza skupień	79
4.3.2. Algorytm k-średnich	84
4.3.3. Analiza najbliższego sąsiedztwa	86
Rozdział 5. Szeregi czasowe (<i>Alicja Ganczarek-Gamrot</i>)	89
5.1. Prognozuj analitycznie	92
5.1.1. Modele trendu i sezonowości	92
5.1.2. Modele autoregresyjne	93
5.1.3. Modele ekonometryczne	95
5.2. Prognozuj mechanicznie	96
5.2.1. Algorytmy wygładzania	96
5.2.2. Wskaźniki sezonowości	98
5.3. Prognozuj przez analogie	99
Literatura	101
Informacja o autorach	113