

Spis treści

Przedmowa 13

Zanim zaczniesz 33

1. Wprowadzenie do komputerów i języka Python 37

1.1. Wstęp 38

1.2. Podstawy technologii obiektowych 39

1.3. Język Python 41

1.4. Biblioteki 44

1.5. IPython i notatniki Jupytera - pierwsze spotkanie 47

1.6. Chmury i internet rzeczy 55

1.7. Big Data - ile to jest "Big"? 58

1.8. Analiza przypadku - mobilna aplikacja Big Data 66

1.9. Wprowadzenie do Data Science: gdy informatyka spotyka się z Data Science 68

1.10. Podsumowanie 71

2. Wprowadzenie do programowania w języku Python 73

2.1. Wstęp 74

2.2. Zmienne i instrukcje przypisania 74

2.3. Obliczenia arytmetyczne 75

2.4. Funkcja "print" i łańcuchy znaków 81

2.5. Łańcuchy z potrójnymi ogranicznikami 83

2.6. Wprowadzanie danych z klawiatury 84

2.7. Podejmowanie decyzji: instrukcja "if" i operatory porównania 86

2.8. Obiekty i typowanie dynamiczne 92

2.9. Wprowadzenie do Data Science - podstawowe statystyki opisowe 93

2.10. Podsumowanie 96

3. Instrukcje przepływu sterowania 97

- 3.1. Wstęp 98
- 3.2. Słowa kluczowe języka Python 99
- 3.3. Instrukcja "if" 99
- 3.4. Instrukcje "if ... else" i "if ... elif ... else" 101
- 3.5. Instrukcja "while" 104
- 3.6. Instrukcja "for" 104
- 3.7. Rozszerzone przypisania 107
- 3.8. Iterowanie po ciągach. Formatowane łańcuchy 107
- 3.9. Nadzorowane iterowanie 109
- 3.10. Wbudowana funkcja "range" - nieco dokładniej 112
- 3.11. Obliczenia finansowe - typ "Decimal" 112
- 3.12. Instrukcje "continue" i "break" 116
- 3.13. Operatory boolowskie 117
- 3.14. Wprowadzenie do Data Science: miary tendencji centralnej 120
- 3.15. Podsumowanie 122
- 4. Funkcje 123
 - 4.1. Wstęp 124
 - 4.2. Funkcje definiowane w programie 124
 - 4.3. Funkcje z wieloma parametrami 127
 - 4.4. Generowanie liczb pseudolosowych 129
 - 4.5. Analiza przypadku - gra losowa 132
 - 4.6. Standardowa biblioteka Pythona 135
 - 4.7. Funkcje modułu "math" 136
 - 4.8. Wspomagane uzupełnianie kodu 138
 - 4.9. Domyślne wartości parametrów 140
 - 4.10. Argumenty kluczowe 141

- 4.11. Zmienne listy parametrów 142
- 4.12. Metody - funkcje należące do obiektów 144
- 4.13. Zasięg definicji 145
- 4.14. O importowaniu nieco dokładniej 147
- 4.15. Przekazywanie argumentów - nieco szczegółów 149
- 4.16. Rekurencja 152
- 4.17. Funkcyjny styl programowania 156
- 4.18. Wprowadzenie do Data Science: miary rozproszenia 158
- 4.19. Podsumowanie 160
- 5. Ciągi: listy i krotki 161
 - 5.1. Wstęp 162
 - 5.2. Listy 162
 - 5.3. Krotki 167
 - 5.4. Rozpakowywanie ciągów 170
 - 5.5. Wyodrębnianie podciągów 173
 - 5.6. Instrukcja "del" 176
 - 5.7. Listy jako argumenty wywołań funkcji 177
 - 5.8. Sortowanie list 179
 - 5.9. Multiplikacja ciągu 180
 - 5.10. Przeszukiwanie ciągów 180
 - 5.11. Inne metody listy 183
 - 5.12. Symulowanie stosu za pomocą listy 186
 - 5.13. Odwzorowywanie list 187
 - 5.14. Wyrażenia generatorowe 189
 - 5.15. Natywne filtrowanie, mapowanie i redukcja 189
 - 5.16. Inne funkcje do przetwarzania ciągów 192
 - 5.17. Listy dwuwymiarowe 193

5.18. Wprowadzenie do Data Science: symulacje i ich statyczna wizualizacja	195
5.19. Podsumowanie	205
6. Słowniki i zbiory	207
6.1. Wstęp	208
6.2. Słowniki	208
6.3. Zbiory	218
6.4. Wprowadzenie do Data Science: dynamiczna wizualizacja symulacji	224
6.5. Podsumowanie	232
7. Biblioteka NumPy i tablice ndarray	233
7.1. Wstęp	234
7.2. Tworzenie tablic na podstawie istniejących danych	235
7.3. Atrybuty tablic	235
7.4. Wypełnianie tablicy zadaną wartością	237
7.5. Tworzenie tablicy na podstawie zakresu danych	238
7.6. %timeit - porównanie efektywności tablic i list	240
7.7. Inne "magiczne" polecenia IPython	241
7.8. Operatory tablicowe	242
7.9. Metody obliczeniowe biblioteki "NumPy"	244
7.10. Funkcje uniwersalne biblioteki "NumPy"	245
7.11. Indeksowanie i wyodrębnianie	247
7.12. Widoki tablic jako płytkie kopie	249
7.13. Głębokie kopiowanie	250
7.14. Restrukturyzacja i transponowanie tablic	251
7.15. Wprowadzenie do Data Science: szeregi i ramki danych biblioteki Pandas	255
7.16. Podsumowanie	269
8. Łańcuchy znaków	271
8.1. Wstęp	272

- 8.2. Formatowanie łańcuchów 272
- 8.3. Konkatenowanie i zwielokrotnianie łańcuchów 278
- 8.4. Usuwanie białych znaków otaczających łańcuch 278
- 8.5. Zmiana wielkości liter w łańcuchu 279
- 8.6. Operatory porównywania łańcuchów 279
- 8.7. Wyszukiwanie podłańcuchów 280
- 8.8. Zastępowanie podłańcuchów 283
- 8.9. Dzielenie i składanie łańcuchów 283
- 8.10. Testowanie specyficznych właściwości łańcucha i jego znaków 286
- 8.11. Surowe łańcuchy 287
- 8.12. Podstawy wyrażeń regularnych 288
- 8.13. Wprowadzenie do Data Science: wyrażenia regularne i preparacja danych w bibliotece Pandas 298
- 8.14. Podsumowanie 303
- 9. Pliki i wyjątki 305
 - 9.1. Wstęp 306
 - 9.2. Pliki 307
 - 9.3. Przetwarzanie plików tekstowych 308
 - 9.4. Aktualizowanie plików tekstowych 311
 - 9.5. Serializacja obiektów w formacie JSON 313
 - 9.6. Niebezpieczny moduł "pickle" 315
 - 9.7. Dodatkowe uwagi o plikach 316
 - 9.8. Obsługa wyjątków 318
 - 9.9. Klauzula "finally" 323
 - 9.10. Jawne generowanie wyjątków 325
 - 9.11. Odwijanie stosu i ślad wykonania 326
 - 9.12. Wprowadzenie do Data Science: przetwarzanie plików CSV 328
 - 9.13. Podsumowanie 335

- 10. Programowanie zorientowane obiektowo 337
 - 10.1. Wstęp 338
 - 10.2. Przykład: klasa "Account" 340
 - 10.3. Kontrolowanie dostępu do atrybutów 344
 - 10.4. Właściwości organizują dostęp do atrybutów. Przykład: klasa "Time" 345
 - 10.5. Symulowanie "prywatności" atrybutów 353
 - 10.6. Analiza przypadku: symulacja tasowania i rozdawania kart 354
 - 10.7. Dziedziczenie: klasy bazowe i podklasy 364
 - 10.8. Hierarchia dziedziczenia a polimorfizm 366
 - 10.9. "Kacze typowanie" a polimorfizm 374
 - 10.10. Przeciążanie operatorów 375
 - 10.11. Klasy wyjątków - hierarchia i definiowanie podklas 379
 - 10.12. Nazwane krotki 380
 - 10.13. Nowość wersji 3.7: klasy danych 381
 - 10.14. Testy jednostkowe przy użyciu łańcuchów dokumentacyjnych i modułu "doctest" 388
 - 10.15. Przestrzenie nazw i widoczność identyfikatorów 391
 - 10.16. Wprowadzenie do Data Science: szeregi czasowe i prosta regresja liniowa 394
 - 10.17. Podsumowanie 402
- 11. Przetwarzanie języka naturalnego (NLP) 405
 - 11.1. Wstęp 406
 - 11.2. Klasa "TextBlob" 407
 - 11.3. Wizualizacja statystyki słów 422
 - 11.4. Ocena czytelności tekstu - biblioteka "Textatistic" 428
 - 11.5. Rozpoznawanie nazwanych encji - biblioteka "spaCy" 430
 - 11.6. Podobieństwo dokumentów 431
 - 11.7. Inne biblioteki i narzędzia NLP 432
 - 11.8. Zastosowanie NLP w uczeniu maszynowym i głębokim uczeniu 433

- 11.9. Zbiory danych dla NLP 434
- 11.10. Podsumowanie 434
- 12. Eksploracja masowych danych - Twitter 437
 - 12.1. Wstęp 438
 - 12.2. Ogólnie o API Twittera 440
 - 12.3. Pierwszy krok - konto deweloperskie 441
 - 12.4. Drugi krok - aplikacja i poświadczenia 441
 - 12.5. Tweety jako obiekty 443
 - 12.6. Biblioteka "Tweepy" 447
 - 12.7. Uwierzytelnianie za pomocą biblioteki "Tweepy" 447
 - 12.8. Informacja o koncie Twittera 449
 - 12.9. Kursory biblioteki "Tweepy" 451
 - 12.10. Przeszukiwanie niedawnych tweetów 454
 - 12.11. Odkrywanie trendów - Twitter Trends API 458
 - 12.12. Preparacja tweetów przed analizą 462
 - 12.13. Strumieniowanie tweetów - Twitter Streaming API 464
 - 12.14. Tweety i analiza sentymentu - klasa "SentimentListener" 469
 - 12.15. Mapy i geokodowanie 474
 - 12.16. Przechowywanie tweetów 483
 - 12.17. Twitter a szeregi czasowe 483
 - 12.18. Podsumowanie 484
- 13. Przetwarzanie poznawcze - IBM Watson 485
 - 13.1. Wstęp 486
 - 13.2. Konto IBM Cloud i konsola usług 487
 - 13.3. Usługi platformy Watson 488
 - 13.4. Przydatne narzędzia platformy Watson 492
 - 13.5. Watson Developer Cloud Python SDK 493

- 13.6. Analiza przypadku: dwujęzyczna aplikacja konwersacyjna 494
- 13.7. Zasoby powiązane z IBM Watson 509
- 13.8. Podsumowanie 511
- 14. Uczenie maszynowe: klasyfikacja, regresja i klasteryzacja 513
 - 14.1. Wstęp 514
 - 14.2. Analiza przypadku: algorytm k najbliższych sąsiadów - klasyfikacja w zbiorze "Digits". Część pierwsza 520
 - 14.3. Analiza przypadku: algorytm k najbliższych sąsiadów - klasyfikacja w zbiorze "Digits". Część druga 531
 - 14.4. Analiza przypadku: prosta regresja liniowa na szeregu czasowym 539
 - 14.5. Analiza przypadku: wielokrotna regresja liniowa na zbiorze "California Housing" 545
 - 14.6. Analiza przypadku: uczenie nienadzorowane. Część pierwsza - redukcja wymiarowości 558
 - 14.7. Analiza przypadku: uczenie nienadzorowane. Część druga - klasteryzacja za pomocą algorytmu k średnich 562
 - 14.8. Podsumowanie 576
- 15. Głębokie uczenie 579
 - 15.1. Wstęp 580
 - 15.2. Pliki danych wbudowane w bibliotekę "Keras" 583
 - 15.3. Alternatywne środowiska Anacondy 584
 - 15.4. Sieci neuronowe 586
 - 15.5. Tensory 588
 - 15.6. Konwolucyjne sieci neuronowe i widzenie komputerowe: wieloklasyfikacja w zbiorze "MNIST" 590
 - 15.7. TensorBoard - wizualizacja trenowania sieci 610
 - 15.8. ConvNetJS: wizualizacja trenowania w oknie przeglądarki WWW 613
 - 15.9. Rekurencyjne sieci neuronowe i ciągi danych: analiza sentymentu w zbiorze "IMDb" 614
 - 15.10. Dostrajanie modeli głębokiego uczenia 622
 - 15.11. Modele wstępnie wytrenowane 623
 - 15.12. Podsumowanie 624
- 16. Big Data: Hadoop, Spark, NoSQL i IoT 627
 - 16.1. Wstęp 628

16.2. Relacyjne bazy danych i język SQL 632

16.3. Bazy danych NoSQL i NewSQL dla Big Data - krótki rekonesans 644

16.4. Analiza przypadku: dokumenty JSON w bazie MongoDB 648

16.5. Hadoop 660

16.6. Spark 672

16.7. Strumieniowanie Sparka: zliczanie hashtagów przy użyciu "pyspark-notebook" 683

16.8. Internet rzeczy (IoT) i dashboardy 693

16.9. Podsumowanie 707