

# Indice generale

|  |                                       |          |
|--|---------------------------------------|----------|
| <b>Prefazione</b>                                  | <b>xiii</b>                           |          |
| A chi è rivolto questo libro                       | xiii                                  |          |
| Argomenti trattati                                 | xiii                                  |          |
| Requisiti  | xvi                                   |          |
| File degli esempi                                  | xvi                                   |          |
| Convenzioni utilizzate                             | xvi                                   |          |
| L'autore   | xvii                                  |          |
| Il revisore tecnico                                | xvii                                  |          |
| <b>Parte I</b>                                     | <b>Fondamenti e algoritmi di base</b> | <b>1</b> |
| <b>Capitolo 1</b>                                  | <b>Panoramica sugli algoritmi</b>     | <b>3</b> |
| Che cos'è un algoritmo?                            | 4                                     |          |
| Le fasi di un algoritmo                            | 4                                     |          |
| La logica di un algoritmo                          | 5                                     |          |
| Che cos'è lo pseudocodice                          | 6                                     |          |
| Utilizzo degli snippet                             | 8                                     |          |
| Creazione di un piano di esecuzione                | 8                                     |          |
| Introduzione ai pacchetti Python                   | 9                                     |          |
| Pacchetti Python                                   | 9                                     |          |
| Implementazione di Python tramite Jupyter Notebook | 11                                    |          |
| Tecniche di progettazione degli algoritmi          | 12                                    |          |
| La dimensione dei dati                             | 12                                    |          |
| La dimensione del calcolo                          | 13                                    |          |
| Analisi delle prestazioni                          | 14                                    |          |
| Analisi della complessità in termini di spazio     | 15                                    |          |
| Analisi della complessità in termini di tempo      | 15                                    |          |
| Stima delle prestazioni                            | 16                                    |          |
| Scelta di un algoritmo                             | 17                                    |          |
| Notazione Big O                                    | 17                                    |          |
| Convalida di un algoritmo                          | 21                                    |          |

|                   |  |           |
|-------------------|--|-----------|
|                   | Algoritmi esatti, approssimativi e randomizzati .....                                | 21        |
|                   | Spiegabilità .....   | 22        |
|                   | Riepilogo .....  | 22        |
| <b>Capitolo 2</b> | <b>Strutture di dati utilizzate negli algoritmi .....</b>                            | <b>25</b> |
|                   | Le strutture di dati in Python .....   | 25        |
|                   | Liste .....  | 26        |
|                   | Tuple .....  | 30        |
|                   | Dizionari .....  | 31        |
|                   | Insiemi .....  | 33        |
|                   | Dataframe .....  | 34        |
|                   | Matrici .....  | 37        |
|                   | I tipi di dati astratti in Python .....  | 37        |
|                   | Vettori .....  | 38        |
|                   | Stack .....  | 38        |
|                   | Code .....   | 40        |
|                   | L'idea di base dietro l'uso di stack e code .....                                    | 42        |
|                   | Alberi .....   | 43        |
|                   | Riepilogo .....  | 45        |
| <b>Capitolo 3</b> | <b>Algoritmi di ordinamento e ricerca .....</b>                                      | <b>47</b> |
|                   | Introduzione agli algoritmi di ordinamento .....                                     | 47        |
|                   | Scambiare le variabili in Python .....   | 48        |
|                   | Bubble sort: ordinamento "a bolle" .....   | 48        |
|                   | Insertion sort .....   | 50        |
|                   | Merge sort .....   | 52        |
|                   | Shell sort .....   | 54        |
|                   | Selection sort .....   | 56        |
|                   | Scelta di un algoritmo di ordinamento .....  | 58        |
|                   | Introduzione agli algoritmi di ricerca .....   | 58        |
|                   | Ricerca lineare .....  | 58        |
|                   | Ricerca binaria .....  | 59        |
|                   | Ricerca per interpolazione .....   | 60        |
|                   | Applicazioni pratiche .....  | 61        |
|                   | Riepilogo .....  | 63        |
| <b>Capitolo 4</b> | <b>Progettazione di algoritmi .....</b>  | <b>65</b> |
|                   | Introduzione ai concetti di base della progettazione<br>di un algoritmo .....        | 65        |
|                   | Problema 1: l'algoritmo progettato produrrà il risultato<br>che ci aspettiamo? ..... | 66        |
|                   | Problema 2: è questo il modo ottimale per ottenere questi<br>risultati? .....        | 67        |
|                   | Problema 3: come si comporterà l'algoritmo su dataset più<br>grandi? .....           | 69        |
|                   | Strategie per la progettazione di algoritmi .....                                    | 70        |

|   |            |
|---|------------|
| La strategia divide et impera.....  | 70         |
| La strategia di programmazione dinamica.....  | 72         |
| La strategia ad algoritmo greedy.....   | 73         |
| Applicazione pratica: risoluzione del problema<br>del commesso viaggiatore.....           | 73         |
| Usare una strategia a forza bruta.....  | 74         |
| Usare un algoritmo greedy .....   | 77         |
| L'algoritmo PageRank .....  | 78         |
| Definizione del problema.....   | 79         |
| Implementazione dell'algoritmo PageRank.....  | 79         |
| La programmazione lineare .....   | 81         |
| Formulazione di un problema di programmazione lineare .....                               | 81         |
| Applicazione pratica: pianificazione della capacità<br>con la programmazione lineare..... | 82         |
| Riepilogo .....   | 84         |
| <b>Capitolo 5</b>   |            |
| <b>Algoritmi per grafi .....</b>  | <b>85</b>  |
| Rappresentazione dei grafi.....   | 86         |
| Tipi di grafi .....   | 87         |
| Tipi speciali di archi .....  | 90         |
| Reti egocentriche.....  | 90         |
| Analisi di reti sociali.....  | 91         |
| Introduzione all'analisi delle reti.....  | 92         |
| Scoprire il percorso più breve .....  | 92         |
| Creare un intorno.....  | 93         |
| Le metriche di centralità.....  | 94         |
| Calcolo delle metriche di centralità usando Python.....                                   | 96         |
| Attraversamento dei grafi .....   | 98         |
| Ricerca in ampiezza: breadth-first .....  | 98         |
| Ricerca in profondità: depth-first .....  | 100        |
| Caso di studio: analisi delle frodi.....  | 104        |
| Una semplice analisi delle frodi.....   | 106        |
| La metodologia di analisi delle frodi watchtower .....                                    | 107        |
| Riepilogo .....   | 109        |
| <b>Parte II</b>   |            |
| <b>Algoritmi di machine learning.....</b>   | <b>111</b> |
| <b>Capitolo 6</b>   |            |
| <b>Algoritmi di machine learning<br/>senza supervisione .....</b>                         | <b>113</b> |
| Introduzione all'apprendimento senza supervisione .....                                   | 113        |
| L'apprendimento senza supervisione nel ciclo di vita<br>del data mining.....              | 114        |
| Tendenze attuali della ricerca nell'apprendimento<br>senza supervisione.....              | 116        |
| Esempi pratici.....   | 117        |

|  |     |
|--|-----|
| Gli algoritmi di clustering .....                                | 118 |
| Quantificare le somiglianze .....                                | 119 |
| L'algoritmo di clustering k-means .....                          | 121 |
| Clustering gerarchico.....                                       | 125 |
| Valutazione dei cluster .....                                    | 127 |
| Applicazioni del clustering.....                                 | 128 |
| Riduzione della dimensionalità .....                             | 128 |
| PCA (Principal Component Analysis) .....                         | 128 |
| Limitazioni della PCA .....                                      | 131 |
| Mining delle regole associative .....                            | 131 |
| Esempi di utilizzo .....   | 131 |
| Analisi del paniere di mercato .....                             | 131 |
| Regole associative.....  | 133 |
| Valutazione delle regole .....                                   | 134 |
| Algoritmi per l'analisi delle associazioni .....                 | 135 |
| Applicazione pratica: clustering di tweet simili.....            | 140 |
| Modellazione dei temi .....                                      | 141 |
| Clustering .....   | 141 |
| Algoritmi per il rilevamento delle anomalie.....                 | 141 |
| Utilizzo del clustering.....                                     | 142 |
| Utilizzo del rilevamento delle anomalie basato sulla densità.... | 142 |
| Utilizzo di macchine a vettori di supporto .....                 | 142 |
| Riepilogo .....  | 142 |

## Capitolo 7 **Algoritmi classici di machine learning con supervisione.....143**

|   |     |
|---|-----|
| Il machine learning con supervisione .....                      | 144 |
| Formulazione del machine learning con supervisione.....         | 144 |
| Le condizioni abilitanti .....                                  | 146 |
| Differenziazione fra classificatori e regressori .....          | 147 |
| Gli algoritmi di classificazione .....                          | 147 |
| La sfida fra i classificatori.....                              | 148 |
| Valutazione dei classificatori.....                             | 151 |
| Le fasi dei classificatori.....                                 | 155 |
| Algoritmo di classificazione ad albero decisionale .....        | 156 |
| I metodi a ensemble.....  | 159 |
| Regressione logistica.....                                      | 163 |
| L'algoritmo SVM.....  | 165 |
| L'algoritmo naive Bayes .....                                   | 167 |
| Fra gli algoritmi di classificazione, il vincitore è.....       | 170 |
| Gli algoritmi di regressione .....                              | 170 |
| La sfida fra i regressori .....                                 | 170 |
| Regressione lineare.....  | 172 |
| L'algoritmo ad albero di regressione .....                      | 176 |
| L'algoritmo di regressione con amplificazione del gradiente.... | 177 |
| Fra gli algoritmi di regressione, il vincitore è.....           | 178 |

|                    |   |            |
|--------------------|---|------------|
|                    | Esempio pratico: previsioni del tempo .....   | 178        |
|                    | Riepilogo .....   | 180        |
| <b>Capitolo 8</b>  | <b>Algoritmi a rete neurale .....</b>   | <b>181</b> |
|                    | Le reti neurali .....   | 182        |
|                    | Evoluzione delle reti neurali.....  | 183        |
|                    | Addestramento di una rete neurale .....   | 185        |
|                    | Anatomia di una rete neurale .....  | 185        |
|                    | Definizione della discesa del gradiente .....   | 186        |
|                    | Funzioni di attivazione.....  | 188        |
|                    | Strumenti e framework .....   | 193        |
|                    | Keras .....   | 193        |
|                    | TensorFlow.....   | 196        |
|                    | I vari tipi di reti neurali.....  | 198        |
|                    | Reti neurali convoluzionali.....  | 198        |
|                    | Trasferimento dell'apprendimento.....   | 200        |
|                    | Caso di studio: utilizzo del deep learning per il rilevamento<br>delle frodi .....        | 201        |
|                    | Metodologia.....  | 201        |
|                    | Riepilogo .....   | 204        |
| <b>Capitolo 9</b>  | <b>Algoritmi per l'elaborazione del linguaggio<br/>naturale .....</b>                     | <b>205</b> |
|                    | Introduzione all'elaborazione del linguaggio naturale .....                               | 205        |
|                    | La terminologia dell'elaborazione del linguaggio naturale.....                            | 206        |
|                    | NLTK .....  | 208        |
|                    | Elaborazione del linguaggio naturale BoW-based .....                                      | 208        |
|                    | Introduzione al word embedding .....  | 211        |
|                    | L'intorno di una parola .....   | 211        |
|                    | Proprietà dei word embedding.....   | 211        |
|                    | Utilizzo delle reti neurali ricorrenti per l'elaborazione<br>del linguaggio naturale..... | 212        |
|                    | Utilizzo dell'elaborazione del linguaggio naturale per l'analisi<br>del sentiment .....   | 212        |
|                    | Caso di studio: analisi del sentiment nelle recensioni di film.....                       | 215        |
|                    | Riepilogo .....   | 217        |
| <b>Capitolo 10</b> | <b>Motori di raccomandazione .....</b>  | <b>219</b> |
|                    | Introduzione ai sistemi di raccomandazione.....   | 219        |
|                    | Tipi di motori di raccomandazione .....   | 220        |
|                    | Motori di raccomandazione basati sui contenuti.....                                       | 220        |
|                    | Motori di raccomandazione a filtri collaborativi.....                                     | 222        |
|                    | Motori di raccomandazione ibridi .....  | 224        |
|                    | I limiti dei sistemi di raccomandazione.....  | 226        |

|  |     |
|--|-----|
| Il problema dell'avviamento a freddo.....                        | 226 |
| Il problema dei metadati .....                                   | 226 |
| Il problema della scarsità di dati.....                          | 226 |
| Il bias dovuto all'influenza sociale .....                       | 227 |
| Dati limitati .....  | 227 |
| Campi di applicazione.....                                       | 227 |
| Esempio pratico: creazione di un motore di raccomandazione ..... | 227 |
| Riepilogo .....  | 229 |

## **Parte III Argomenti avanzati .....231**

### **Capitolo 11 Algoritmi per i dati.....233**

|   |     |
|---|-----|
| Introduzione agli algoritmi per i dati.....                             | 233 |
| Classificazione dei dati .....  | 234 |
| Gli algoritmi di archiviazione dei dati.....                            | 234 |
| Le strategie di archiviazione dei dati.....                             | 235 |
| Gli algoritmi per i dati in streaming.....                              | 237 |
| Applicazioni operanti su dati in streaming .....                        | 237 |
| Gli algoritmi di compressione dei dati.....                             | 238 |
| Algoritmi di compressione senza perdita di dati.....                    | 238 |
| Esempio pratico: analisi del sentiment dei tweet<br>in tempo reale..... | 239 |
| Riepilogo .....   | 242 |

### **Capitolo 12 Crittografia.....245**

|  |     |
|--|-----|
| Introduzione alla crittografia .....   | 245 |
| L'importanza dell'anello debole.....   | 246 |
| Terminologia di base.....  | 246 |
| I requisiti di sicurezza.....  | 247 |
| Identificare le entità.....  | 247 |
| Definire gli obiettivi di sicurezza .....  | 247 |
| Valutare la sensibilità dei dati.....  | 248 |
| Progettazione di cifrature .....   | 249 |
| Cifratura a sostituzione.....  | 249 |
| Cifratura di Cesare.....   | 249 |
| Rotazione 13 (ROT13).....  | 250 |
| Criptoanalisi dei cifrari a sostituzione.....  | 251 |
| La cifratura a trasposizione .....   | 251 |
| Tipi di tecniche crittografiche.....   | 251 |
| La crittografia hash .....   | 252 |
| La crittografia simmetrica .....   | 254 |
| La crittografia asimmetrica.....   | 256 |
| Esempio: problemi di sicurezza nella distribuzione<br>di un modello di machine learning..... | 260 |
| Attacchi Man-in-the-Middle .....   | 260 |

|   |            |
|---|------------|
| Attacchi masquerading.....  | 262        |
| Crittografia di dati e modelli.....   | 262        |
| Riepilogo .....   | 265        |
| <b>Capitolo 13 Algoritmi per dati su larga scala .....</b>                  | <b>267</b> |
| Introduzione agli algoritmi per dati su larga scala.....                    | 267        |
| Che cos'è un algoritmo su larga scala ben progettato .....                  | 268        |
| Terminologia.....   | 268        |
| Progettazione di algoritmi paralleli .....                                  | 269        |
| Legge di Amdahl.....  | 269        |
| Condurre analisi di esecuzione parallela .....                              | 270        |
| Granularità delle attività.....   | 271        |
| Bilanciamento del carico.....   | 272        |
| Localizzazione dei dati.....  | 272        |
| Elaborazione concorrente in Python .....                                    | 272        |
| Strategia di elaborazione multi-risorsa .....                               | 273        |
| Che cos'è CUDA .....  | 273        |
| Cluster computing.....  | 277        |
| Strategia ibrida .....  | 278        |
| Riepilogo .....   | 279        |
| <b>Capitolo 14 Considerazioni pratiche .....</b>                            | <b>281</b> |
| Introduzione.....   | 281        |
| La triste storia di un bot di intelligenza artificiale<br>per Twitter ..... | 282        |
| La spiegabilità di un algoritmo .....                                       | 283        |
| Algoritmi di machine learning e spiegabilità.....                           | 283        |
| Algoritmi ed etica.....   | 287        |
| Problemi con gli algoritmi ad apprendimento .....                           | 287        |
| Le implicazioni etiche.....   | 287        |
| Ridurre i pregiudizi nei modelli.....                                       | 288        |
| Affrontare i problemi NP-difficili.....                                     | 289        |
| Semplificare il problema.....   | 290        |
| Adattare una soluzione nota a un problema simile .....                      | 290        |
| Usare un metodo probabilistico .....  | 290        |
| Quando usare gli algoritmi.....   | 291        |
| Un esempio pratico: eventi cigno nero.....                                  | 291        |
| Riepilogo .....   | 293        |
| <b>Indice analitico.....</b>  | <b>295</b> |