

Practical Data Science with Python 3

**Synthesizing Actionable
Insights from Data**

Ervin Varga

Apress®

Table of Contents

About the Authorxi

About the Technical Reviewerxiii

Acknowledgments xv

Introduction xvii

Chapter 1: Introduction to Data Science.....1

 Main Phases of a Data Science Project2

 Brown Cow Model Case Study.....4

 Big Data9

 Big Data Example: MOOC Platforms10

 How to Learn Data Science.....12

 Domain Knowledge Attainment—Example14

 Programming Skills Attainment—Example16

 Overview of the Anaconda Ecosystem18

 Managing Packages and Environments20

 Sharing and Reproducing Environments23

 Summary.....25

 References.....26

Chapter 2: Data Engineering.....29

 E-Commerce Customer Segmentation: Case Study30

 Creating a Project in Spyder33

 Downloading the Dataset34

TABLE OF CONTENTS

Exploring the Dataset	36
Inspecting Results	56
Persisting Results.....	60
Restructuring Code to Cope with Large CSV Files.....	62
Public Data Sources.....	64
Summary.....	69
References	70
Chapter 3: Software Engineering.....	73
Characteristics of a Large-Scale Software System	75
Software Engineering Knowledge Areas	79
Rules, Principles, Conventions, and Standards	81
Context Awareness and Communicative Abilities	85
Reducing Cyclomatic Complexity	89
Cone of Uncertainty and Having Time to Ask.....	91
Fixing a Bug and Knowing How to Ask	93
Handling Legacy Code.....	102
Understanding Bug-Free Code	103
Understanding Faulty Code	105
The Importance of APIs	107
Fervent Flexibility Hurts Your API.....	109
The Socio-* Pieces of Software Production	111
Funny Elevator Case Study	112
Summary.....	118
References	119

Chapter 4: Documenting Your Work.....	121
JupyterLab in Action	125
Experimenting with Code Execution	126
Managing the Kernel	134
Descending Ball Project.....	136
Refactoring the Simulator’s Notebook.....	148
Document Structure.....	150
Wikipedia Edits Project.....	152
Summary.....	156
References.....	157
Chapter 5: Data Processing	159
Augmented Descending Ball Project.....	159
Version 1.1.....	160
Version 1.2.....	172
Version 1.3.....	184
Abstractions vs. Latent Features	200
Compressing the Ratings Matrix	201
Summary.....	205
References.....	207
Chapter 6: Data Visualization	209
Visualizing Temperature Data Case Study.....	210
Showing Stations on a Map.....	211
Plotting Temperatures	213
Closest Pair Case Study	218
Version 1.0.....	223
Version 2.0.....	229
Version 3.0.....	233

TABLE OF CONTENTS

Enquiry of Algorithms Evolution	241
Interactive Information Radiators	241
The Power of Domain-Specific Languages.....	244
Summary.....	252
References.....	253
Chapter 7: Machine Learning	255
Exposition of Core Concepts and Techniques	258
Overfitting.....	271
Underfitting and Feature Interaction	276
Collinearity.....	278
Residuals Plot.....	281
Regularization.....	285
Predicting Financial Movements Case Study	286
Data Retrieval	288
Data Preprocessing	288
Feature Engineering	303
Implementing Streaming Linear Regression	308
Summary.....	314
References.....	316
Chapter 8: Recommender Systems	317
Introduction to Recommender Systems.....	318
Simple Movie Recommender Case Study	322
Introduction to LensKit for Python	329
Summary.....	338
References.....	339

Chapter 9: Data Security.....	341
Checking for Compromise.....	342
Introduction to the GDPR.....	349
Machine Learning and Security	359
Membership Inference Attack	359
Poisoning Attack.....	363
Summary.....	365
References.....	366
Chapter 10: Graph Analysis	369
Usage Matrix As a Graph Problem.....	370
Opposite Quality Attributes	376
Partitioning the Model into a Bipartite Graph	377
Scalable Graph Loading.....	380
Social Networks	385
Summary.....	395
References.....	396
Chapter 11: Complexity and Heuristics	397
From Simple to Complicated.....	401
Counting the Occurrences of a Digit.....	402
Estimating the Edge Betweenness Centrality.....	409
The Count of Divisible Numbers	413
From Disorder to Complex	415
Exploring the KDD Cup 1999 Data	416
Cynefin and Data Science	420
Summary.....	425
References.....	425

TABLE OF CONTENTS

Chapter 12: Deep Learning427

 Intelligent Machines.....428

 Intelligence As Mastery of Symbols431

 Manual Feature Engineering432

 Machine-Based Feature Engineering436

 Summary.....449

 References.....449

Index.....451