

Contents

Introduction: Data Science in Tourism	xxxii
Industry Insights from Data Scientists: Q&A Session	xlvii
Part I Theoretical Fundaments	
AI and Big Data in Tourism	3
Luisa Mich	
Epistemological Challenges	17
Roman Egger and Joanne Yu	
Data Science and Interdisciplinarity	35
Roman Egger and Joanne Yu	
Data Science and Ethical Issues	51
Roman Egger, Larissa Neuburger, and Michelle Mattuzzi	
Web Scraping	67
Roman Egger, Markus Kroner, and Andreas Stöckl	
Part II Machine Learning	
Machine Learning in Tourism: A Brief Overview	85
Roman Egger	
Feature Engineering	109
Pablo Duboue	
Clustering	129
Matthias Fuchs and Wolfram Höpken	
Dimensionality Reduction	151
Nikolay Oskolkov	

Classification	169
Ulrich Bodenhofer and Andreas Stöckl	
Regression	209
Andreas Stöckl and Ulrich Bodenhofer	
Hyperparameter Tuning	231
Pier Paolo Ippolito	
Model Evaluation	253
Ajda Pretnar Žagar and Janez Demšar	
Interpretability of Machine Learning Models	275
Urszula Czerwinska	
Part III Natural Language Processing	
Natural Language Processing (NLP): An Introduction	307
Roman Egger and Enes Gokce	
Text Representations and Word Embeddings	335
Roman Egger	
Sentiment Analysis	363
Andrei P. Kirilenko, Luyu Wang, and Svetlana O. Stepchenkova	
Topic Modelling	375
Roman Egger	
Entity Matching: Matching Entities Between Multiple Data Sources	
Ivan Bilan	405
Knowledge Graphs	423
Mayank Kejriwal	
Part IV Additional Methods	
Network Analysis	453
Rodolfo Baggio	
Time Series Analysis	467
Irem Onder and Wenqi Wei	
Agent-Based Modelling	481
Jillian Student	
Geographic Information System (GIS)	513
Andrei P. Kirilenko	

Visual Data Analysis	527
Johanna Schmidt	
Software and Tools	547
Roman Egger	
Glossary	589
Index	599