

# Data Analytics

## Series editors

Longbing Cao, Advanced Analytics Institute, University of Technology Sydney,  
Broadway, NSW, Australia

Philip S. Yu, University of Illinois, Chicago, IL, USA

**Aims and Goals:**

Building and promoting the field of data science and analytics in terms of publishing work on theoretical foundations, algorithms and models, evaluation and experiments, applications and systems, case studies, and applied analytics in specific domains or on specific issues.

**Specific Topics:**

This series encourages proposals on cutting-edge science, technology and best practices in the following topics (but not limited to):

Data analytics, data science, knowledge discovery, machine learning, big data, statistical and mathematical methods for data and applied analytics,

New scientific findings and progress ranging from data capture, creation, storage, search, sharing, analysis, and visualization,

Integration methods, best practices and typical examples across heterogeneous, interdependent complex resources and modals for real-time decision-making, collaboration, and value creation.

More information about this series at <http://www.springer.com/series/15063>

Longbing Cao

# Data Science Thinking

The Next Scientific, Technological  
and Economic Revolution

 Springer

Longbing Cao   
Advanced Analytics Institute  
University of Technology Sydney  
Sydney, NSW, Australia

ISSN 2520-1859

ISSN 2520-1867 (electronic)

Data Analytics

ISBN 978-3-319-95091-4

ISBN 978-3-319-95092-1 (eBook)

<https://doi.org/10.1007/978-3-319-95092-1>

Library of Congress Control Number: 2018952348

© Springer International Publishing AG, part of Springer Nature 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To my family and beloved ones for their  
generous time and sincere love,  
encouragement, and support which  
essentially form part of the core driver for  
completing this book.*

# Preface

When you migrated to the twenty-first century, did you ever consider what today's world would look like? And what would inspire and drive the development and transformation of almost every aspect of our daily lives, study, work, and entertainment—in fact, every discipline and domain, including government, business, and society in general?

The most relevant answer may be data, and more specifically so-called “big data,” the data economy, the science of data: *data science*, and data scientists. This is without doubt the age of big data, data science, data economy, and data profession.

The past several years have seen tremendous hype about the evolution of cloud computing, big data, data science, and now artificial intelligence. However, it is undoubtedly true that the volume, variety, velocity, and value of data continue to increase every millisecond. It is data and data intelligence that is transforming everything, integrating the past, present, and future. Data is regarded as the new Intel Inside, the new oil, and a strategic asset. Data drives or even determines the future of science, technology, economy, and possibly everything in our world today.

This desirable, fast-evolving, and boundless data world has triggered the debate about *data-intensive scientific discovery*—data science—as a new paradigm, i.e., the so-called “fourth science paradigm,” which unifies experiment, theory, and computation (corresponding to “empirical” or “experimental,” “theoretical,” and “computational” science). At the same time, it raises several fundamental questions: What is data science? How does data science connect to other disciplines? How does data science translate into the profession, education, and economy? How does data science transform existing science, technologies, industry, economy, profession, and education? And how can data science compete in next-generation science, technologies, economy, profession, and education? More specific questions also arise, such as what forms the mindset and skillset of data scientists?

The research, innovation, and practices seeking to address the above and other relevant questions are driving *the fourth revolution* in scientific, technological, and economic development history, namely *data science, technology, and economy*.

These questions motivate the writing of this book from a high-level perspective.

There have been quite a few books on data science, or books that have been labeled in the book market as belonging under the data science umbrella. This book does not address the technical details of any aspect of mathematics and statistics, machine learning, data mining, cloud computing, programming languages, or other topics related to data science. These aspects of data science techniques and applications are covered in another book—*Data Science: Techniques and Applications*—by the same author.

Rather, this book is inspired by the desire to explore answers to the above fundamental questions in the era of data science and data economy. It is intended to paint a comprehensive picture of data science as a new scientific paradigm from the scientific evolution perspective, as data science thinking from the scientific thinking perspective, as a transdisciplinary science from the disciplinary perspective, and as a new profession and economy from the business perspective.

As a result, the book covers a very wide spectrum of essential and relevant aspects of data science, spanning the evolution, concepts, thinking and challenges, discipline and foundation of data science to its role in industrialization, profession, and education, and the vast array of opportunities it offers. The book is decomposed into three parts to cover these aspects.

In Part I, we introduce the evolution, concepts and misconceptions, and thinking of data science. This part consists of three chapters. In Chap. 1, the evolution, characteristics, features, trends, and agenda of the data era are reviewed. Chapter 2 discusses the question “What is data science?” from a high-level, multidisciplinary, and process perspective. The hype surrounding big data and data science is evidenced by the many myths and misconceptions that prevail, which are also discussed in this chapter. Data science thinking plays a significant role in the research, innovation, and applications of data science and is discussed in Chap. 3.

Part II introduces the challenges and foundations of doing data science. These important issues are discussed in three chapters. First, the various challenges are explored in Chap. 4. In Chap. 5, the methodologies, disciplinary framework, and research areas in data science are summarized from the disciplinary perspective. Chapter 6 explores the roles and relationships of relevant disciplines and their knowledge base in forming the foundations of data science. Lastly, Chap. 7 summarizes the main research issues, theories, methods, and applications of analytics and learning in the various domains and applications.

The last part, Part III, concerns data science-driven industrialization and opportunities, discussed in four chapters. Data science and its ubiquitous applications drive the data economy, data industry, and data services, which are explored in Chap. 8. Data science, data economy, and data applications propel the development of the data profession, fostering data science roles and maturity models, which are highlighted in Chap. 10. The era of data science has to be built by data scientists and engineers; thus the required qualifications, educational framework, and capability set are discussed in Chap. 11. Lastly, Chap. 12 explores the future of data science.

As illustrated above, this book on data science differs significantly from any book currently on the market by the breadth of its coverage of comprehensive data

science, technology, and economic perspectives. This all-encompassing intention makes compiling a book like this an extremely challenging and risky venture. Basic theories and algorithms in machine learning and data mining are not discussed, nor are most of the related concepts and techniques, as readers can find these in the book *Data Science: Techniques and Applications*, and other more dedicated books, for which a rich set of references and materials is provided.

The book is intended for data managers (e.g., analytics portfolio managers, business analytics managers, chief data analytics officers, chief data scientists, and chief data officers), policy makers, management and decision strategists, research leaders, and educators who are responsible for pursuing new scientific, innovation, and industrial transformation agendas, enterprise strategic planning, or next-generation profession-oriented course development, and others who are involved in data science, technology, and economy from a higher perspective. Research students in data science-related disciplines and courses will find the book useful for conceiving their innovative scientific journey, planning their unique and promising career, and for preparing and competing in the next-generation science, technology, and economy.

Can you imagine how the data world and data era will continue to evolve and how our future science, technologies, economy, and society will be influenced by data in the second half of the twenty-first century? To claim that we are data scientists and “doing data,” we need to grapple with these big, important questions to comprehend and capitalize on the current parameters of data science and to realize the opportunities that will arise in the future. We thus hope this book will contribute to the discussion.

Sydney, NSW, Australia  
July 2018

Longbing Cao



# Acknowledgments

Writing a book like this has been a long journey requiring the commitment of tremendous personal, family, and institutional time, energy, and resources. It has been built on a dozen years of the author's limited, evolving but enthusiastic observations, thinking, experience, research, development, and practice, in addition to a massive amount of knowledge, lessons, and experience acquired from and inspired by colleagues, research and business partners and collaborators. The author would therefore like to thank everyone who has worked, studied, supported, and discussed the relevant research tasks, publications, grants, projects, and enterprise analytics practices with him since he was a data manager of business intelligence solutions and then an academic in the field of data science and analytics.

This book was particularly written in alignment with the author's vision and decades of effort and dedication to the development of data science, culminating in the creation and directorship of the Advanced Analytics Institute (AAi) at the University of Technology Sydney in 2011. This was the first Australian group dedicated to big data analytics, and the author would thus like to thank the university for its strategic leadership in supporting his vision and success in creating and implementing the Institute's Research, Education and Development business model, the strong research culture fostered in his team, the weekly meetings with students and visitors which significantly motivated and helped to clarify important concepts, issues, and questions, and the support of his students, fellows, and visiting scholars. Many of the ideas, perspectives, and early thinking included in this book were initially brought to the author's weekly team meetings for discussion. It has been a very great pleasure to engage in such intensive and critical weekly discussions with young and smart talent. The author indeed appreciates and enjoys these discussions and explorations, and thanks those students, fellows, and visitors who have attended the meetings over the past 10+ years.

In addition, heartfelt thanks are given to my family for their endless support and generous understanding every day and night of the past 4 years spent compiling this book, in addition to their dozens of years of continuous support to the author's research and practice in the field.

The author is grateful to professional editor Ms. Sue Felix who has made significant effort in editing the book.

Last but not least, my sincere thanks to Springer, in particular Ms. Melissa Fearon at Springer US, for their kindness in supporting the publication of this monograph in its *Book Series on Data Analytics*, edited by Longbing Cao and Philip S Yu.

Writing this book has been a very brave decision, and a very challenging and risky journey due to many personal limitations. There are still many aspects that have not been addressed, or addressed adequately, in this edition, and the book may have incorporated debatable aspects, limitations, or errors in the thinking, conceptions, opinions, summarization, and proposed value and opportunities of the data-driven fourth revolution: data science, technology, and economy. The author welcomes comments, discussion, suggestions, or criticism on the content of the book, including being alerted to errors or misunderstandings. Discussion boards and materials from this book are available at [www.datasciences.info](http://www.datasciences.info), a data science portal created and managed by the author and his team for promoting data science research, innovation, profession, education, and commercialization. Direct feedback to the author at [Longbing.Cao@gmail.com](mailto:Longbing.Cao@gmail.com) is also an option for commenting on possible improvements to the book and for the benefit of the data science discipline and communities.

# Contents

## Part I Concepts and Thinking

<b>1</b>	<b>The Data Science Era</b> .....	<b>3</b>
1.1	Introduction .....	3
1.2	Features of the Data Era .....	5
1.2.1	Some Key Terms in Data Science.....	5
1.2.2	Observations of the Data Era Debate .....	5
1.2.3	Iconic Features and Trends of the Data Era .....	7
1.3	The Data Science Journey .....	9
1.3.1	New-Generation Data Products and Economy .....	13
1.4	Data-Empowered Landscape .....	14
1.4.1	Data Power .....	14
1.4.2	Data-Oriented Forces .....	16
1.5	New X-Generations .....	17
1.5.1	X-Complexities .....	18
1.5.2	X-Intelligence .....	18
1.5.3	X-Opportunities.....	19
1.6	The Interest Trends .....	20
1.7	Major Data Strategies by Governments .....	21
1.7.1	Governmental Data Initiatives .....	23
1.7.2	Australian Initiatives .....	23
1.7.3	Chinese Initiatives .....	24
1.7.4	European Initiatives .....	25
1.7.5	United States' Initiatives .....	25
1.7.6	Other Governmental Initiatives .....	26
1.8	The Scientific Agenda for Data Science .....	26
1.8.1	The Scientific Agenda by Governments .....	26
1.8.2	Data Science Research Initiatives.....	27
1.9	Summary.....	28

- 2 What Is Data Science** ..... 29
  - 2.1 Introduction ..... 29
  - 2.2 Datafication and Data Quantification ..... 29
  - 2.3 Data, Information, Knowledge, Intelligence and Wisdom ..... 30
  - 2.4 Data DNA ..... 32
    - 2.4.1 What Is Data DNA ..... 32
    - 2.4.2 Data DNA Functionalities ..... 33
  - 2.5 Data Science Views ..... 34
    - 2.5.1 The Data Science View in Statistics ..... 34
    - 2.5.2 A Multidisciplinary Data Science View ..... 35
    - 2.5.3 The Data-Centric View ..... 35
  - 2.6 Definitions of Data Science ..... 36
    - 2.6.1 High-Level Data Science Definition ..... 36
    - 2.6.2 Trans-Disciplinary Data Science Definition ..... 37
    - 2.6.3 Process-Based Data Science Definition ..... 38
  - 2.7 Open Model, Open Data and Open Science ..... 43
    - 2.7.1 Open Model ..... 44
    - 2.7.2 Open Data ..... 45
    - 2.7.3 Open Science ..... 46
  - 2.8 Data Products ..... 48
  - 2.9 Myths and Misconceptions ..... 48
    - 2.9.1 Possible Negative Effects in Conducting Data Science ... 49
    - 2.9.2 Conceptual Misconceptions ..... 50
    - 2.9.3 Data Volume Misconceptions ..... 52
    - 2.9.4 Data Infrastructure Misconceptions ..... 53
    - 2.9.5 Analytics Misconceptions ..... 53
    - 2.9.6 Misconceptions About Capabilities and Roles ..... 55
    - 2.9.7 Other Matters ..... 56
  - 2.10 Summary ..... 58
- 3 Data Science Thinking** ..... 59
  - 3.1 Introduction ..... 59
  - 3.2 Thinking in Science ..... 60
    - 3.2.1 Scientific vs. Unscientific Thinking ..... 60
    - 3.2.2 Creative Thinking vs. Logical Thinking ..... 62
  - 3.3 Data Science Structure ..... 66
  - 3.4 Data Science as a Complex System ..... 68
    - 3.4.1 A Systematic View of Data Science Problems ..... 68
    - 3.4.2 Complexities in Data Science Systems ..... 71
    - 3.4.3 The Framework for Data Science Thinking ..... 72
    - 3.4.4 Data Science Thought ..... 73
    - 3.4.5 Data Science Custody ..... 74
    - 3.4.6 Data Science Feed ..... 74
    - 3.4.7 Mechanism Design for Data Science ..... 75
    - 3.4.8 Data Science Deliverables ..... 76
    - 3.4.9 Data Science Assurance ..... 76

- 3.5 Critical Thinking in Data Science ..... 77
  - 3.5.1 Critical Thinking Perspectives ..... 77
  - 3.5.2 We Do Not Know What We Do Not Know ..... 77
  - 3.5.3 Data-Driven Scientific Discovery ..... 80
  - 3.5.4 Data-Driven and Other Paradigms ..... 83
  - 3.5.5 Essential Questions to Ask in Data Science ..... 88
- 3.6 Summary ..... 89

**Part II Challenges and Foundations**

- 4 Data Science Challenges ..... 93**
  - 4.1 Introduction ..... 93
  - 4.2 X-Complexities in Data Science ..... 94
    - 4.2.1 Data Complexity ..... 94
    - 4.2.2 Behavior Complexity ..... 95
    - 4.2.3 Domain Complexity ..... 95
    - 4.2.4 Social Complexity ..... 96
    - 4.2.5 Environment Complexity ..... 96
    - 4.2.6 Human-Machine-Cooperation Complexity ..... 97
    - 4.2.7 Learning Complexity ..... 97
    - 4.2.8 Deliverable Complexity ..... 98
  - 4.3 X-Intelligence in Data Science ..... 99
    - 4.3.1 Data Intelligence ..... 99
    - 4.3.2 Behavior Intelligence ..... 100
    - 4.3.3 Domain Intelligence ..... 100
    - 4.3.4 Human Intelligence ..... 100
    - 4.3.5 Network Intelligence ..... 101
    - 4.3.6 Organization Intelligence ..... 101
    - 4.3.7 Social Intelligence ..... 102
    - 4.3.8 Environment Intelligence ..... 103
  - 4.4 Known-to-Unknown Data-Capability-Knowledge Cognitive Path ..... 103
    - 4.4.1 The Data Science Cognitive Path ..... 103
    - 4.4.2 Four Knowledge Spaces in Data Science ..... 104
    - 4.4.3 Data Science Known-to-Unknown Evolution ..... 105
    - 4.4.4 Opportunities for Significant Original Invention ..... 105
  - 4.5 Non-IIDness in Data Science Problems ..... 106
    - 4.5.1 IIDness vs. Non-IIDness ..... 106
    - 4.5.2 Non-IID Challenges ..... 108
  - 4.6 Human-Like Machine Intelligence Revolution ..... 109
    - 4.6.1 Next-Generation Artificial Intelligence: Human-Like Machine Intelligence ..... 110
    - 4.6.2 Data Science-Enabled Human-Like Machine Intelligence ..... 111

4.7	Data Quality .....	113
4.7.1	Data Quality Issues .....	113
4.7.2	Data Quality Metrics .....	115
4.7.3	Data Quality Assurance and Control .....	116
4.7.4	Data Quality Analytics .....	118
4.7.5	Data Quality Checklist .....	119
4.8	Data Social and Ethical Issues .....	121
4.8.1	Data Social Issues .....	121
4.8.2	Data Science Ethics.....	123
4.8.3	Data Ethics Assurance.....	124
4.9	The Extreme Data Challenge .....	125
4.10	Summary.....	127
<b>5</b>	<b>Data Science Discipline .....</b>	<b>129</b>
5.1	Introduction.....	129
5.2	Data-Capability Disciplinary Gaps .....	129
5.3	Methodologies for Complex Data Science Problems.....	131
5.3.1	From Reductionism and Holism to Systematism .....	132
5.3.2	Synthesizing X-Intelligence.....	135
5.3.3	Qualitative-to-Quantitative Metasynthesis .....	136
5.4	Data Science Disciplinary Framework .....	138
5.4.1	Interdisciplinary Fusion for Data Science .....	138
5.4.2	Data Science Research Map.....	140
5.4.3	Systematic Research Approaches.....	143
5.4.4	Data A-Z for Data Science .....	144
5.5	Some Essential Data Science Research Areas .....	145
5.5.1	Developing Data Science Thinking .....	146
5.5.2	Understanding Data Characteristics and Complexities ...	148
5.5.3	Discovering Deep Behavior Insight .....	150
5.5.4	Fusing Data Science with Social and Management Science .....	153
5.5.5	Developing Analytics Repositories and Autonomous Data Systems .....	156
5.6	Summary.....	160
<b>6</b>	<b>Data Science Foundations .....</b>	<b>161</b>
6.1	Introduction.....	161
6.2	Cognitive Science and Brain Science for Data Science .....	163
6.3	Statistics and Data Science .....	164
6.3.1	Statistics for Data Science.....	165
6.3.2	Data Science for Statistics.....	166
6.4	Information Science Meets Data Science .....	167
6.4.1	Analysis and Processing.....	168
6.4.2	Informatics for Data Science.....	169
6.4.3	General Information Technologies.....	170

- 6.5 Intelligence Science and Data Science ..... 171
  - 6.5.1 Pattern Recognition, Mining, Analytics and Learning .... 172
  - 6.5.2 Nature-Inspired Computational Intelligence ..... 173
  - 6.5.3 Data Science: Beyond Information and Intelligence Science ..... 173
- 6.6 Computing Meets Data Science ..... 175
  - 6.6.1 Computing for Data Science ..... 175
  - 6.6.2 Data Science for Computing ..... 177
- 6.7 Social Science Meets Data Science ..... 179
  - 6.7.1 Social Science for Data Science ..... 180
  - 6.7.2 Data Science for Social Science ..... 183
  - 6.7.3 Social Data Science..... 188
- 6.8 Management Meets Data Science ..... 190
  - 6.8.1 Management for Data Science ..... 191
  - 6.8.2 Data Science for Management ..... 194
  - 6.8.3 Management Analytics and Data Science ..... 196
- 6.9 Communication Studies Meets Data Science ..... 197
- 6.10 Other Fundamentals and Electives ..... 199
  - 6.10.1 Broad Business, Management and Social Areas ..... 200
  - 6.10.2 Domain and Expert Knowledge..... 200
  - 6.10.3 Invention, Innovation and Practice..... 201
- 6.11 Summary..... 202
- 7 Data Science Techniques ..... 203**
  - 7.1 Introduction..... 203
  - 7.2 The Problem of Analytics and Learning ..... 204
  - 7.3 The Conceptual Map of Data Science Techniques ..... 204
    - 7.3.1 Foundations of Data Science ..... 205
    - 7.3.2 Classic Analytics and Learning Techniques..... 208
    - 7.3.3 Advanced Analytics and Learning Techniques..... 210
    - 7.3.4 Assisting Techniques ..... 214
  - 7.4 Data-to-Insight-to-Decision Analytics and Learning ..... 219
    - 7.4.1 Past Data Analytics and Learning ..... 220
    - 7.4.2 Present Data Analytics and Learning ..... 220
    - 7.4.3 Future Data Analytics and Learning ..... 221
    - 7.4.4 Actionable Decision Discovery and Delivery ..... 221
  - 7.5 Descriptive-to-Predictive-to-Prescriptive Analytics ..... 222
    - 7.5.1 Stage 1: Descriptive Analytics and Business Reporting .. 223
    - 7.5.2 Stage 2: Predictive Analytics/Learning and Business Analytics ..... 224
    - 7.5.3 Stage 3: Prescriptive Analytics and Decision Making .... 225
    - 7.5.4 Focus Shifting Between Analytics/Learning Stages ..... 226
    - 7.5.5 Synergizing Descriptive, Predictive and Prescriptive Analytics ..... 228

- 7.6 X-Analytics ..... 230
  - 7.6.1 X-Analytics Spectrum ..... 230
  - 7.6.2 X-Analytics Working Mechanism ..... 231
- 7.7 Summary ..... 232

**Part III Industrialization and Opportunities**

- 8 Data Economy and Industrialization ..... 237**
  - 8.1 Introduction ..... 237
  - 8.2 Data Economy ..... 237
    - 8.2.1 What Is Data Economy ..... 238
    - 8.2.2 Data Economy Example: Smart Taxis and Shared e-Bikes ..... 241
    - 8.2.3 New Data Economic Model ..... 243
    - 8.2.4 Distinguishing Characteristics of Data Economy ..... 246
    - 8.2.5 Intelligent Economy and Intelligent Datathings ..... 247
    - 8.2.6 Translating Real Economy ..... 249
  - 8.3 Data Industry ..... 251
    - 8.3.1 Categories of Data Industries ..... 251
    - 8.3.2 New Data Industries ..... 252
    - 8.3.3 Transforming Traditional Industries ..... 254
  - 8.4 Data Services ..... 257
    - 8.4.1 Data Service Models ..... 257
    - 8.4.2 Data Analytical Services ..... 259
  - 8.5 Summary ..... 262
- 9 Data Science Applications ..... 263**
  - 9.1 Introduction ..... 263
  - 9.2 Some General Application Guidance ..... 264
    - 9.2.1 Data Science Application Scenarios ..... 264
    - 9.2.2 General Data Science Processes ..... 264
    - 9.2.3 General vs. Domain-Specific Algorithms and Vendor-Dependent vs. Independent Solutions ..... 265
    - 9.2.4 The Waterfall Model vs. the Agile Model for Data Science Project Management ..... 266
    - 9.2.5 Success Factors for Data Science Projects ..... 268
  - 9.3 Advertising ..... 269
  - 9.4 Aerospace and Astronomy ..... 270
  - 9.5 Arts, Creative Design and Humanities ..... 270
  - 9.6 Bioinformatics ..... 271
  - 9.7 Consulting Services ..... 271
  - 9.8 Ecology and Environment ..... 272
  - 9.9 E-Commerce and Retail ..... 273
  - 9.10 Education ..... 274
  - 9.11 Engineering ..... 274
  - 9.12 Finance and Economy ..... 275



- 9.13 Gaming Industry..... 276
- 9.14 Government..... 277
- 9.15 Healthcare and Clinics ..... 277
- 9.16 Living, Sports, Entertainment, and Relevant Services..... 278
- 9.17 Management, Operations and Planning ..... 279
- 9.18 Manufacturing ..... 279
- 9.19 Marketing and Sales..... 280
- 9.20 Medicine ..... 281
- 9.21 Physical and Virtual Society, Community, Networks,  
Markets and Crowds ..... 282
- 9.22 Publishing and Media ..... 284
- 9.23 Recommendation Services..... 285
- 9.24 Science..... 286
- 9.25 Security and Safety..... 287
- 9.26 Social Sciences and Social Problems..... 288
- 9.27 Sustainability ..... 288
- 9.28 Telecommunications and Mobile Services ..... 289
- 9.29 Tourism and Travel..... 290
- 9.30 Transportation ..... 291
- 9.31 Summary..... 292
- 10 Data Profession ..... 293**
  - 10.1 Introduction..... 293
  - 10.2 Data Profession Formation..... 294
    - 10.2.1 Disciplinary Significance Indicator ..... 294
    - 10.2.2 Significant Data Science Research..... 294
    - 10.2.3 Global Data Scientific Communities ..... 295
    - 10.2.4 Significant Data Professional Development..... 297
    - 10.2.5 Significant Socio-Economic Development ..... 298
  - 10.3 Data Science Roles ..... 298
    - 10.3.1 Data Science Team ..... 299
    - 10.3.2 Data Science Positions ..... 300
  - 10.4 Core Data Science Knowledge and Skills..... 301
    - 10.4.1 Data Science Knowledge and Capability Set..... 301
    - 10.4.2 Data Science Communication Skills..... 304
  - 10.5 Data Science Maturity ..... 307
    - 10.5.1 Data Science Maturity Model ..... 308
    - 10.5.2 Data Maturity ..... 309
    - 10.5.3 Capability Maturity..... 311
    - 10.5.4 Organizational Maturity ..... 312
  - 10.6 Data Scientists ..... 313
    - 10.6.1 Who Are Data Scientists ..... 313
    - 10.6.2 Chief Data Scientists ..... 314
    - 10.6.3 What Data Scientists Do ..... 315
    - 10.6.4 Qualifications of Data Scientists ..... 318

- 10.6.5 Data Scientists vs. BI Professionals ..... 319
- 10.6.6 Data Scientist Job Survey ..... 320
- 10.7 Data Engineers ..... 320
  - 10.7.1 Who Are Data Engineers ..... 321
  - 10.7.2 What Data Engineers Do ..... 323
- 10.8 Tools for Data Professionals ..... 325
- 10.9 Summary ..... 326
- 11 Data Science Education ..... 329**
  - 11.1 Introduction ..... 329
  - 11.2 Data Science Course Review ..... 330
    - 11.2.1 Overview of Existing Courses ..... 330
    - 11.2.2 Disciplines Offering Courses ..... 331
    - 11.2.3 Course Body of Knowledge ..... 332
    - 11.2.4 Course-Offering Organizations ..... 332
    - 11.2.5 Course-Offering Channels ..... 333
    - 11.2.6 Online Courses ..... 333
    - 11.2.7 Gap Analysis of Existing Courses ..... 334
  - 11.3 Data Science Education Framework ..... 337
    - 11.3.1 Data Science Course Structure ..... 337
    - 11.3.2 Bachelor in Data Science ..... 339
    - 11.3.3 Master in Data Science ..... 343
    - 11.3.4 PhD in Data Science ..... 346
  - 11.4 Summary ..... 347
- 12 Prospects and Opportunities in Data Science ..... 349**
  - 12.1 Introduction ..... 349
  - 12.2 The Fourth Revolution: Data+Intelligence Science, Technology and Economy ..... 350
    - 12.2.1 Data Science, Technology and Economy: An Emerging Area ..... 350
    - 12.2.2 The Fourth Scientific, Technological and Economic Revolution ..... 352
  - 12.3 Data Science of Sciences ..... 355
  - 12.4 Data Brain ..... 356
  - 12.5 Machine Intelligence and Thinking ..... 358
  - 12.6 Advancing Data Science and Technology and Economy ..... 359
  - 12.7 Advancing Data Education and Profession ..... 361
  - 12.8 Summary ..... 362
- References ..... 363**
- Index ..... 381**