

Studies in Big Data

Volume 50

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series “Studies in Big Data” (SBD) publishes new developments and advances in the various areas of Big Data—quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence incl. neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/11970>

Francesco Corea

An Introduction to Data

Everything You Need to Know About AI,
Big Data and Data Science

 Springer

Francesco Corea
Department of Management
Ca' Foscari University
Venice, Italy

ISSN 2197-6503

ISSN 2197-6511 (electronic)

Studies in Big Data

ISBN 978-3-030-04467-1

ISBN 978-3-030-04468-8 (eBook)

<https://doi.org/10.1007/978-3-030-04468-8>

Library of Congress Control Number: 2018961695

© Springer Nature Switzerland AG 2019

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 you, now and forever

Preface

This book aims to be an introduction to big data, artificial intelligence and data science for anyone who wants to learn more about those domains. It is neither a fully technical book nor a strategic manual, but rather a collection of essays and lessons learned doing this job for a while.

In that sense, this book is not an organic text that should be read from the first page onwards, but rather a collection of articles that can be read at will (or at need). The structure of the chapter is very similar, so I hope the reader won't find difficulties in establishing comparisons or understanding the differences between specific problems AI is being used for. I personally recommend reading the first three-four chapters in a row to have a general overview of the technologies and then jump around depending on what topic interests you the most.

The book also replicates some of the contents already introduced in previous research as well as shows new material created working as a data scientist, as a startup advisor as an investor. It is therefore to some extent both a new book and a 2.0 version of some previous work of mine, but for sure the content is reorganized in a completely new way and gets new meaning when read in a different context.

Artificial intelligence is certainly a hot topic nowadays, and this book wants to be both a guide on the past and a tool to look into the future. I always tried to maintain a balance between explaining concepts, tools and ways in which AI has been used, and *potential* applications or trends for future. I hope the reader may find himself not only grasping how relevant AI, big data and data science are for our progress as a society, but also wondering *what's next*.

The book is structured in such a way that the first few chapters explain the most relevant definitions and business contexts where AI and big data can have an impact. The rest of the book looks instead at specific sectorial applications, issues or more generally subjects that AI is meaningfully changing.

Finally, I am writing this book hoping that it will be valuable for some readers in how they think and use technologies to improve our lives and that it could stimulate conversations or projects that could produce a positive impact in our society.

Venice, Italy

Francesco Corea

Contents

| | | |
|----------|--|----|
| 1 | Introduction to Data | 1 |
| | References | 4 |
| 2 | Big Data Management: How Organizations Create and Implement Data Strategies | 7 |
| | References | 13 |
| 3 | Introduction to Artificial Intelligence | 15 |
| | 3.1 Basic Definitions and Categorization | 15 |
| | 3.2 A Bit of History | 18 |
| | 3.3 Why AI Is Relevant Today | 20 |
| | References | 22 |
| 4 | AI Knowledge Map: How to Classify AI Technologies | 25 |
| | References | 29 |
| 5 | Advancements in the Field | 31 |
| | 5.1 Machine Learning | 31 |
| | 5.2 Neuroscience Advancements | 34 |
| | 5.3 Hardware and Chips | 36 |
| | References | 38 |
| 6 | AI Business Models | 41 |
| | Reference | 46 |
| 7 | Hiring a Data Scientist | 47 |
| | References | 51 |
| 8 | AI and Speech Recognition | 53 |
| | 8.1 Conversation Interfaces | 53 |
| | 8.2 The Challenges Toward Master Bots | 54 |

| | | |
|-----------|---|-----------|
| 8.3 | How Is the Market Distributed? | 55 |
| 8.4 | Final Food for Thoughts | 56 |
| | References | 56 |
| 9 | AI and Insurance | 57 |
| 9.1 | A Bit of Background | 57 |
| 9.2 | So How Can AI Help the Insurance Industry? | 58 |
| 9.3 | Who Are the Sector Innovators? | 59 |
| 9.4 | Concluding Thoughts | 61 |
| 10 | AI and Financial Services | 63 |
| 10.1 | Financial Innovation: Lots of Talk, Little Action? | 63 |
| 10.2 | Innovation Transfer: The Biopharma Industry | 64 |
| 10.3 | Introducing AI, Your Personal Financial Disruptor | 65 |
| 10.4 | Segmentation of AI in Fintech | 66 |
| 10.5 | Conclusions | 68 |
| | References | 68 |
| 11 | AI and Blockchain | 69 |
| 11.1 | Non-technical Introduction to Blockchain | 69 |
| 11.2 | A Digression on Initial Coin Offerings (ICOs) | 70 |
| 11.3 | How AI Can Change Blockchain | 71 |
| 11.4 | How Blockchain Can Change AI | 72 |
| 11.5 | Decentralized Intelligent Companies | 73 |
| 11.6 | Conclusion | 75 |
| | References | 75 |
| 12 | New Roles in AI | 77 |
| 12.1 | Hiring New Figures to Lead the Data Revolution | 77 |
| 12.2 | The Chief Data Officer (CDO) | 77 |
| 12.3 | The Chief Artificial Intelligence Officer (CAIO) | 79 |
| 12.4 | The Chief Robotics Officer (CRO) | 80 |
| 13 | AI and Ethics | 83 |
| 13.1 | How to Design Machines with Ethically-Significant Behaviors | 83 |
| 13.2 | Data and Biases | 83 |
| 13.3 | Accountability and Trust | 85 |
| 13.4 | AI Usage and the Control Problem | 88 |
| 13.5 | AI Safety and Catastrophic Risks | 89 |
| 13.6 | Research Groups on AI Ethics and Safety | 89 |
| 13.7 | Conclusion | 90 |
| | References | 91 |
| 14 | AI and Intellectual Property | 93 |
| 14.1 | Why Startups Patent Inventions (and Why Is Different for AI) | 93 |

- 14.2 The Advantages of Patenting Your Product 94
- 14.3 Reasons Behind not Looking for Patent Protection 96
- 14.4 The Patent Landscape 98
- 14.5 Conclusions 99
- References 99
- 15 AI and Venture Capital 101**
 - 15.1 The Rationale 101
 - 15.2 Previous Studies 102
 - 15.2.1 Personal and Team Characteristics 102
 - 15.2.2 Financial Considerations 104
 - 15.2.3 Business Features 104
 - 15.2.4 Industry Knowledge 105
 - 15.2.5 An Outsider Study: Hobos and Highfliers 105
 - 15.3 Who Is Using AI in the Private Investment Field 107
 - 15.4 Conclusions 108
 - References 108
- 16 A Guide to AI Accelerators and Incubators 111**
 - 16.1 Definitions 111
 - 16.2 Are They Worth Their Value? 112
 - 16.2.1 Entrepreneur Perspective: To Join or not to Join 112
 - 16.2.2 Investor Perspective: Should I Stay or
Should I Go 113
 - 16.2.3 Accelerators Assessment Metrics: Is the Program
Any Good? 114
 - 16.3 A Comparison Between Accelerators 115
 - 16.4 Final Thoughts 115
 - References 118
- Appendix A: Nomenclature for Managers 119**
- Appendix B: Data Science Maturity Test 123**
- Appendix C: Data Scientist Extended Skills List
(Examples in Parentheses) 127**
- Appendix D: Data Scientist Personality Questionnaire 129**

List of Figures

| | | |
|-----------|--|----|
| Fig. 2.1 | Big data lean deployment approach | 8 |
| Fig. 2.2 | Big data maturity map | 11 |
| Fig. 2.3 | Maturity stage transitions | 12 |
| Fig. 2.4 | Data analytics organizational models | 13 |
| Fig. 3.1 | Artificial intelligence trend for the period 2012–2016 | 19 |
| Fig. 4.1 | AI knowledge map. | 26 |
| Fig. 6.1 | Artificial intelligence classification matrix | 45 |
| Fig. 7.1 | Data scientist core skills set. | 48 |
| Fig. 7.2 | Data science value chain | 50 |
| Fig. 8.1 | Bots classification matrix | 55 |
| Fig. 10.1 | Innovation transfer: the biopharma industry | 64 |

List of Tables

| | | |
|------------|---|-----|
| Table 2.1 | Data stage of development structure | 9 |
| Table 7.1 | Data scientists’ personality assessment and classification. | 50 |
| Table 15.1 | Taxonomy of signals to predict probability of startup success. | 106 |
| Table 16.1 | AI accelerators and incubators. | 116 |
| Table B.1 | Data science maturity test classification. | 125 |
| Table D.1 | Data scientist personality classification | 131 |