

Fundamentals of Clinical Data Science

Pieter Kubben · Michel Dumontier
Andre Dekker
Editors

Fundamentals of Clinical Data Science

 Springer Open

Editors

Pieter Kubben
Department of Neurosurgery
Maastricht University
Maastricht, Limburg
The Netherlands

Michel Dumontier
Institute of Data Science
Maastricht University
Maastricht, Limburg
The Netherlands

Andre Dekker
Maastricht Clinic
Maastricht, Limburg
The Netherlands



Corrected Publication 2019. This book is an open access publication.
ISBN 978-3-319-99712-4 ISBN 978-3-319-99713-1 (eBook)
<https://doi.org/10.1007/978-3-319-99713-1>

Library of Congress Control Number: 2018963226

© The Editor(s) (if applicable) and The Author(s) 2019, corrected publication 2019

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

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.

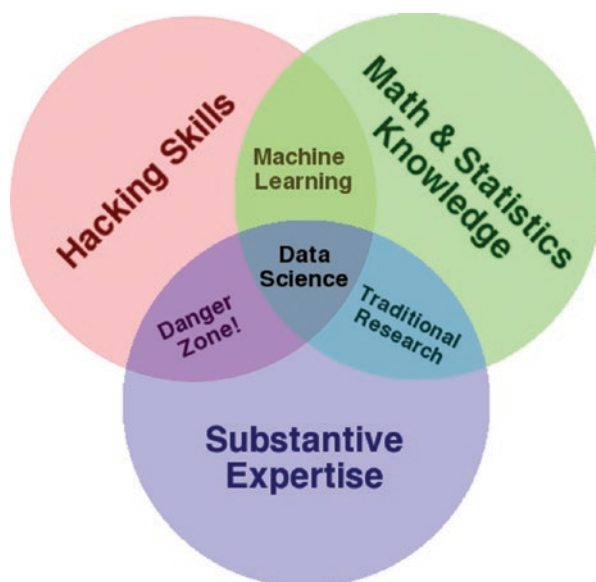
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained 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

Introduction “Fundamentals of Clinical Data Science”

In the era of eHealth and personalized medicine, “big data” and “machine learning” are increasingly becoming part of the medical world. Algorithms are capable of supporting diagnostic and therapeutic processes and offer added value for both health-care professionals and patients. The field of big data, machine learning, deep learning, and algorithm development and validation is often referred to as “data science,” and “data scientist” was mentioned in *Harvard Business Review* as “the sexiest job of the 21st century” (<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>). A commonly used visual representation of the field is Drew Conway’s Venn diagram (Fig. 1), which describes data science as a mix of content expertise, methodological knowledge, and IT skills.

Fig. 1 Data science Venn diagram by Drew Conway. (Reproduced with permission)



Unfortunately, most healthcare professionals still consider the field of clinical data science as highly technical and something “for the IT whizzkids.” That leaves many interesting and valuable opportunities unexplored and could even contribute to serious flaws in developed algorithms. Chen and Asch described machine learning’s “peak of inflated expectations” and suggest that “we can soften a subsequent crash into a ‘trough of disillusionment’ by fostering a stronger appreciation of the technology’s capabilities and limitations” (Chen and Asch 2017). They conclude that “combining machine-learning software with the best human clinician ‘hardware’ will permit delivery of care that outperforms what either can do alone.” We could not agree more.

This book is for you, the healthcare professional and “best human clinician hardware” who would like to embrace the field of clinical data science but who is still looking for a resource that explains the topic in nonengineering terminology. This book’s promise is “*no math, no code.*” It contains three sections that help you understand the transformation of data to model and to applications. It should be sufficient to give you a decent grasp on the topic for understanding and a solid foundation if you are to continue with active mastery of the field by taking programming courses online or in a classroom setting. Either way, we want you to get aboard.

Our thanks go to the NFU Citrienfonds who made it financially possible to publish this e-book as open access. Citrienfonds of the NFU and ZonMw helps to develop sustainable solutions in Dutch healthcare to all authors for their valuable time and contributions, to Studio Piranha for the website, and to Springer for their help in the publishing process.



Pieter Kubben, Michel Dumontier, and André Dekker
www.clinicaldatasciencebook.com

Reference

Chen JH, Asch SM. Machine learning and prediction in medicine – beyond the peak of inflated expectations. *N Engl J Med.* 2017;376(26):2507–9. <https://doi.org/10.1056/NEJMp1702071>.

Contents

Part I Data Collection

1	Data Sources	3
	Pieter Kubben	
2	Data at Scale	11
	Alberto Traverso, Frank J. W. M. Dankers, Leonard Wee, and Sander M. J. van Kuijk	
3	Standards in Healthcare Data	19
	Stefan Schulz, Robert Stegwee, and Catherine Chronaki	
4	Research Data Stewardship for Healthcare Professionals	37
	Paula Jansen, Linda van den Berg, Petra van Overveld, and Jan-Willem Boiten	
5	The EU’s General Data Protection Regulation (GDPR) in a Research Context	55
	Christopher F. Mondschein and Cosimo Monda	

Part II From Data to Model

6	Preparing Data for Predictive Modelling	75
	Sander M. J. van Kuijk, Frank J. W. M. Dankers, Alberto Traverso, and Leonard Wee	
7	Extracting Features from Time Series	85
	Christian Herff and Dean J. Krusienski	
8	Prediction Modeling Methodology	101
	Frank J. W. M. Dankers, Alberto Traverso, Leonard Wee, and Sander M. J. van Kuijk	

9 Diving Deeper into Models 121
 Alberto Traverso, Frank J. W. M. Dankers, Biche Osong,
 Leonard Wee, and Sander M. J. van Kuijk

**10 Reporting Standards and Critical Appraisal
 of Prediction Models** 135
 Leonard Wee, Sander M. J. van Kuijk, Frank J. W. M. Dankers,
 Alberto Traverso, Mattea Welch, and Andre Dekker

Part III From Model to Application

11 Clinical Decision Support Systems 153
 A. T. M. Wasylewicz and A. M. J. W. Scheepers-Hoeks

12 Mobile Apps 171
 Pieter Kubben

**13 Optimizing Care Processes with Operational
 Excellence & Process Mining** 181
 Henri J. Boersma, Tiffany I. Leung, Rob Vanwersch, Elske Heeren,
 and G. G. van Merode

14 Value-Based Health Care Supported by Data Science 193
 Tiffany I. Leung and G. G. van Merode

Correction to: Prediction Modeling Methodology C1

Index 213