Thomas Barton
Christian Müller
Editors

Apply Data Science

Introduction, Applications and Projects

Springer Vieweg

Apply Data Science

Thomas Barton · Christian Müller
Editors

Apply Data Science

Introduction, Applications and Projects

Springer Vieweg

Editors
Thomas Barton
FB Informatik, Hochschule Worms
Worms, Germany

Christian Müller
FB Wirtschaft, Informatik, Recht (WIR)
Technische Hochschule Wildau
Wildau, Germany

ISBN 978-3-658-38797-6 ISBN 978-3-658-38798-3 (eBook)
https://doi.org/10.1007/978-3-658-38798-3

Responsible Editor: Petra Steinmueller
This Springer Vieweg imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH, part of Springer Nature.
The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Contents

Editors and Contributors

About the Editors

Prof. Dr. Thomas Barton studied and obtained his doctorate at TU Kaiserlautern. He then worked for SAP SE for approximately 10 years with a focus on application development, consulting, training and project management. Since 2006 he has been working at the University of Applied Sciences Worms as a professor of computer science with a focus on business informatics. His areas of activity include the development of business applications, e-business and data science. He is the author and editor of numerous publications. In addition, he is active in various committees and working groups. He is also the spokesman for the GI Board of Universities of Applied Sciences.

Prof. Dr. Christian Müller studied mathematics at the Freie Universität Berlin and received his doctorate in 1989 on network flows with side conditions. From 1990 to 1992 he worked for Schering AG and from 1992 to 1994 for the Berliner Verkehrsbetriebe (BVG) in the area of travel and duty plan optimization. In 1994 he received a call to the Technische Hochschule Wildau, Germany. He is a professor in the field of business informatics and dean of the faculty of business, computer science and law. His research focuses are the conception of information systems, mathematical optimization and the simulation of business processes.

Contributors

Dr. Emal M. Alekozai Robert Bosch GmbH, Finance Shared Service Digital, Stuttgart, Germany

Jule Aßmann Hamburg, Germany

Prof. Dr. Thomas Barton Homburg, Germany

Dr. Jörg Baus Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Germany

Dr. Boris Brandherm Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Germany

Dr. Matthieu Deru Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Germany

Prof. Dr. Ralf Gampfer Hochschule Worms, Worms, Germany

Prof. Dr. Rainer Gerten Hochschule Mannheim, Mannheim, Germany

Prof. Dr. Michael Gröschel Hochschule Mannheim, Mannheim, Germany

Joshua Hammesfahr Walldorf, Germany

Prof. Dr. Jens Kaufmann Niederrhein University of Applied Sciences, FB Economics, Mönchengladbach, Germany

Gian-Luca Kiefer Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Germany

Arthur Kokoev Worms, Germany

Prof. Dr. Stefan Kubica Zernsdorf, Germany

Dr. Stephan Kühnel Martin-Luther-Universität Halle-Wittenberg, Lehrstuhl für Wirtschaftsinformatik, insb. Betriebliches Informationsmanagement, Halle (Saale), Germany

Prof. Dr. Christoph Laroque University of Applied Sciences Zwickau, Faculty of Economics, Zwickau, Germany

Prof. Dr. Claudia Lemke Berlin School of Economics and Law (HWR Berlin), Department 2, Berlin, Germany

Prof. Dr. Olga Matthias Leeds Beckett University, Leeds Business School, City Campus, Leeds, UK

Manuel Mikoleit Gückingen, Germany

Prof. Dr. Dagmar Monett Hochschule für Wirtschaft und Recht Berlin (HWR Berlin), Fachbereich 2, Berlin, Germany

Prof. Dr. Christian Müller Berlin, Germany

Dr. Alassane Ndiaye Kognitive Assistenzsysteme, DFKI GmbH, Saarbrücken, Germany

Dipl.-Inform. Uwe Neuhaus NORDAKADEMIE Hochschule der Wirtschaft, Elmshorn, Germany

Anna-Maria Nitsche University of Applied Sciences Zwickau, Faculty of Economics, Zwickau, Germany

Andreas Peuker Harthausen, Germany

Tobias Peuschke-Bischof Wildau, Germany

Benedikt Reiner Hochschule Karlsruhe, Karlsruhe, Germany

Prof. Dr. Daniel Retkowitz Hochschule Niederrhein, FB Wirtschaftswissenschaften, Mönchengladbach, Germany

Prof. Dr. Gabriele Roth-Dietrich Hochschule Mannheim, Mannheim, Germany

Prof. Dr. Joachim Sauer NORDAKADEMIE Elmshorn, Fachbereich Informatik, Elmshorn, Germany

Prof. Dr. Michael Schulz NORDAKADEMIE Hochschule der Wirtschaft, Elmshorn, Germany

Prof. Dr.-Ing. habil. Christian-Andreas Schumann Westsächsische Hochschule Zwickau, Faculty of Economics, Zwickau, Germany

Prof. Dr. Martin Spott HTW Berlin, Berlin, Germany

Part I

Introduction

Data Science: From Concept to Application

1

Thomas Barton and Christian Müller

Contents

Abstract

This article provides an introduction to the topic of applying data science, which appears in the series "Applied Business Informatics". After a discussion of the terms data science and data scientist, the article presents the contributions that appear in the thematic areas. The starting point is an introduction to the field of data science. This is followed by various articles on systems, tools and methods. One focus of this book is the presentation of concrete applications.

T. Barton (✉)
Homburg, Germany
e-mail: barton@hs-worms.de

C. Müller
Berlin, Germanye-mail: christian.mueller@th-wildau.de

T. Barton and C. Müller (eds.), *Apply Data Science*,
https://doi.org/10.1007/978-3-658-38798-3_1

Keywords

Data Science · Data Scientist · Methods · Tools · Systems · Applications

1.1 What is Data Science?

A first visual representation describing the term data science based on the competencies of the people who practice this discipline is shown in Fig. 1.1. According to Drew Conway, data science is shown as the intersection of three sets in a Venn diagram, with each set representing a competence of the people who practice data science [1]. A competence consists of hacking skills, which include abilities like editing text files at the command line level and an understanding of vector operations, as well as algorithmic thinking. Here, a hacker is seen as a person who enjoys having a deeper understanding of the inner workings of a system, a computer, or especially a computer network [2]. Other competencies are mathematical and statistical knowledge. Substantial competencies complete the profile. They should embody the willingness to discover and create knowledge for which motivated questions about the world are needed, from which hypotheses can be derived that can be verified using data.

From a statistician's point of view, data science can be seen as a combination of statistics and computer science [3]. Statistics is one of the most important disciplines that provides tools and methods to gain deeper insights from data; it is the most important discipline to investigate and quantify uncertainty [4].

For the German Academy of Technical Sciences acatech, data science can be divided into four core areas [5]: data engineering, data analytics, data prediction and machine learning. Data engineering includes processes and methods for storing, accessing and traceability of data, while data analytics focuses on data analysis. The prediction of topics and situations based on experience is the subject of data prediction. Machine learning is seen as a cross-sectional area to these three areas.

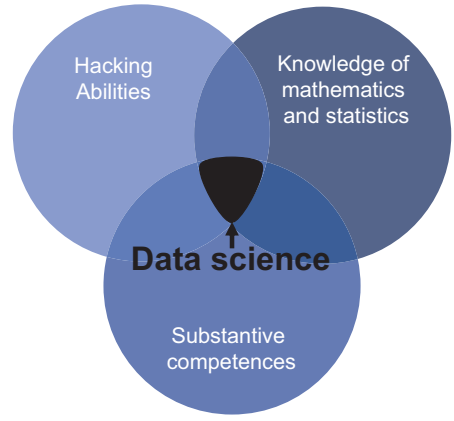

Fig. 1.1 Venn diagram for data science based on [1]

A newer definition for data science, based on an interdisciplinary approach, dates from 2017 [6]:

▶ Data Science = (Statistics + Applied Informatics + Computing + Communication + Sociology + Management | (Data + Environment + Thinking)).

Data science is based on statistics, applied informatics, computing, communication, management and sociology, with the term sociology standing for social aspects. The subject is the investigation of data and their environment, which includes in particular the associated domains in order to transform data into knowledge and decisions. Here, thinking and methods should be used that make it possible to generate knowledge on the basis of data. The vertical bar | means "under the condition that" and is known from conditional probability.

The assessment of social aspects is based on morality. Morality can be understood as the "total of observable behavior, behavioral norms and attitudes and values related to behavior" [7]. The subject of ethics is morality. As a science, ethics is a discipline of philosophy. Digital ethics is more or less seen as a synonym for information ethics, and "information ethics has the morality of those as its subject who offer and use information and communication technologies (ICT) and new media" [8].

Data science is also the subject of degree programs and continuing education offerings at universities. Proposals for the content design of degree programs and continuing education offerings for data science are available from the GI working group "Data Science / Data Literacy" [9].

1.2 What is and What Does a Data Scientist?

Data Scientist is advertised as the most attractive job of the 21st century [10]. In 2015, a Chief Data Scientist was appointed during the tenure of President Barack Obama [11]. But what does a Data Scientist do, and what qualifications are required? The duties of a Data Scientist can be very diverse. In order to describe the extensive knowledge that Data Scientists need to perform their jobs, a list of qualifications was drawn up and assigned to different subject areas. A representation of the qualifications in the assigned subject areas is shown in Table 1.1 [12].

Eighteen qualifications are assigned to five topic areas. Under the topic area Machine Learning/Big Data, qualifications for structured, unstructured and distributed data are also summarized. The topic area Mathematics/Operation Research also includes qualifications for algorithms, Bayesian statistics and Monte Carlo methods, graphical models and simulations. The topic area Programming/System Administration also includes system administration and programming of back-end and front-end. The topic area Statistics includes statistics as well as qualifications for surveys and marketing as well as for visualization.

Table 1.1 Topics for Data Scientists' qualifications

Business/Product Development	Machine Learning/ Big Data	Mathematics/Operation Research	Programming / System- administration	Statistics and Visualization
- Business - Product Development	- Big Data and distributed data - Machine Learning - Structured data - Unstructured data	- Algorithms - Bayesian statistics and Monte Carlo methods - Graphical models - Mathematics - Optimization - Simulation	- Back-end programming - Front-end programming - System administration	- Statistics - Surveys and marketing - Visualization

Four specializations can be derived for data scientists. Depending on the extent to which the qualifications can be assigned to the five topic areas, the following specializations result [12]:

- Data-Business-Person: A person with a focus on qualifications for business
- Data Creative: A person for whom the qualifications are evenly distributed across the five topic areas
- Data Developer: A person with a focus on qualifications for programming
- Data Researcher: A person with advanced qualifications for statistics

Good data scientists should [13]

- have technical expertise that can be demonstrated, for example, by completing a scientific degree,
- be curious with a desire to discover and go into depth to break down a problem into hypotheses that can be tested,
- practice storytelling by using data to tell a story and communicate it effectively,
- be clever to approach problems creatively and in different ways.

In addition, a team that carries out data science is characterized by intensive collaborative work [14].

1.3 Introduction to Data Science

The introduction to the topic of data science begins in Chap. 2 with a contribution by the authors Jens Kaufmann and Daniel Retkowitz. Their contribution entitled "Visualization and Deep Learning in Data Science" introduces the visual presentation of data and the extraction of information from images.

Ethical considerations are playing an increasingly important role in the digital transformation of companies. “Digital Ethics in Data-Driven Organizations and Their Application to the Example of AI Ethics” is the title of the contribution by Claudia Lemke, Dagmar Monett and Manuel Mikoleit. Digital ethics and its application in intelligent systems of a company is the subject of their contribution in Chap. 3.

Digital transformation leads to the implementation of technological solutions to support decision making. Anna-Maria Nitsche, Christian-Andreas Schumann, Christoph Laroque and Olga Matthias present a process framework for the implementation and evaluation of these technologies. Their contribution is entitled “Multiple Perspectives on the Implementation of Innovative Technological Solutions in the Context of Data-Driven Decision Making” and includes Chap. 4.

Investigations into the failure of data science projects are the subject of the contribution by Jule Aßmann, Joachim Sauer and Michael Schulz. Their contribution, which is assigned to Chap. 5, is entitled “Don’t be afraid of failures—findings from a survey on the failure of data science projects”.

1.4 Systems, Tools and Methods

The first contribution in this section comes from Thomas Barton and Andreas Peuker. In their contribution in Chap. 6 entitled “Recommendation systems and the use of machine learning methods”, the authors present the basics and the use of recommendation systems taking into account recent developments.

Gabriele Roth-Dietrich, Michael Gröschel and Benedikt Reiner compare BI systems with regard to their ability to provide machine learning functionality for applications. Chap. 7 consists of their contribution “Comparison of the machine learning functionality of business intelligence and analytics tools”.

The contribution of the author team Stephan Kühnel, Uwe Neuhaus, Jens Kaufmann, Michael Schulz and Emal M. Alekozai in Chap. 8 is about the use of a data science process model. The title of the article is “Using the Data Science Process Model Version 1.1. (DASC-PM v1.1) for executing Data Science projects: procedures, competencies and roles”.

1.5 Applications

The presented applications deal with the topics of renewable energy and optimization of energy management, scientific literature evaluation, customer satisfaction in the automotive industry and driver assistance system development.

The energy transition with the exit from nuclear energy and its challenges for the electricity system in Germany are the starting point for the contribution by Boris Brandherm, Matthieu Deru, Alassane Ndiaye, Gian-Luca Kiefer, Jörg Baus and Ralf Gampfer. The integration of renewable energies using the example of photovoltaic systems and the

use of artificial intelligence is the subject of their contribution “Integration of renewable energies—AI-based prediction methods for electricity generation by photovoltaic systems” in Chap. 9.

Sustainability and a more efficient use of heat energy are the starting point for the contribution by Gabriele Roth-Dietrich and Rainer Gerten. Their contribution “Machine learning for energy management optimization” describes the optimization of an air conditioning system using Data Science based on a concrete project. Chap. 10 consists of this contribution.

Chapter 11 shows how the analysis of publications can be supported by text mining as part of a scientific literature review. To do this, authors Thomas Barton and Arthur Kokoev present a concrete application example. Their contribution is entitled “Text Mining in a Scientific Literature Review: Extraction of Keywords to Describe Content”.

The use of data to determine relevant relationships is the subject of the contribution by authors Joshua Hammesfahr and Martin Spott. As part of their contribution “Identification of Relevant Relationships in Data Using Machine Learning”, they establish a connection between, on the one hand, configurations of products or an infrastructure and, on the other hand, errors in order to reduce errors. Chap. 12 presents this contribution.

The management and analysis of vehicle data in a digital learning factory is the subject of the contribution by Tobias Peuschke-Bischof and Stefan Kubica. Their contribution in Chap. 13 entitled “Framework for the Management and Analysis of Vehicle Data for Model-based Driver Assistance System Development in Teaching and Research” illustrates mechanisms in the development and project management.

References

1. Conway D (2010) The data science venn diagram. http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram. Accessed: 21. Nov. 2022
2. Malkin G, Parker TL (1993) Request for comments 1392. Internet users’ glossary. https://tools.ietf.org/html/rfc1392. Accessed: 21. November 2022
3. Kauermann G (2019) Data Science – Einige Gedanken aus Sicht eines Statistikers. Informatik Spektrum 42(6):387–393
4. Weihs C, Ickstadt K (2018) Data science: the impact of statistics. Int J Data Sci Analytics 6:189–194
5. Gausemeier J, Guggemos M, Kreimeyer A (Hrsg) (2017) Auswahl, Beschreibung, Bewertung und Messung der Schlüsselkompetenzen für das Technologiefeld Data Science. acatec. https://www.acatech.de/wp-content/uploads/2019/02/acatech_NKM_Data_Science_WEB-2.pdf. Accessed: 21. November 2022
6. Cao L (2017) Data science: a comprehensive overview. ACM Comput Surv 50(3): Article 43
7. Werner MH (2021) Einführung in die Ethik. Metzler. https://doi.org/10.1007/978-3-476-05293-3
8. Bendel O (2019) 400 Keywords Informationsethik, 2nd edn. Springer Gabler, Wiesbaden

9. GI (2019) Data Science: Lern- und Ausbildungsinhalte (Arbeitspapier). https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Arbeitspapier_Data-Science_2019-12_01.pdf. Accessed: 21. November 2022
10. Davenport TH, Patil DJ (2012) Data scientist: the sexiest job of the 21st century. Harv Bus Rev 90(5):70–76
11. The White House (2015) The white house names Dr. DJ Patil as the first U.S. chief data scientist. https://obamawhitehouse.archives.gov/blog/2015/02/18/white-house-names-dr-dj-patil-first-us-chief-data-scientist. Accessed: 21. November 2022
12. Harris HD Murphy SP, Vaisman M (2013) Analyzing the analyzers. O'Reilly Media, Sevastopol
13. Patil DJ (2011) Building data science teams. O'Reilly Media, Sebastopol
14. Zhang AX, Muller M, Wang D (2020) How do data science workers collaborate? Roles, workflows, and tools. Proc ACM on Hum-Comput Interact 4(CSCW1):1–23

Thomas Barton studied and obtained his doctorate at the TU Kaiserlautern. He then worked for SAP SE for approximately 10 years with a focus on application development, also consulting, training and project management. Since 2006 he has been working at the Hochschule Worms as a professor for computer science with a focus on business informatics. His areas of activity include the development of business applications, e-business and data science. He is the author and editor of numerous publications. In addition, he is active in various committees and working groups. He is also engaged as a speaker of the GI Board for Universities of Applied Sciences.

Christian Müller studied mathematics at the Freie Universität Berlin and obtained his doctorate in 1989 on network flows with side conditions. From 1990 to 1992 he worked at Schering AG and from 1992 to 1994 at the Berliner Verkehrsbetriebe (BVG) in the area of timetable and service planning optimization. In 1994 he was appointed to the Technische Hochschule Wildau, Germany. He is a professor in the field of business informatics and dean of the Faculty of Economics, Computer Science and Law. His research focus is on the conception of information systems, mathematical optimization and the simulation of business processes.

Part II

Introduction to Data Science

Visualization and Deep Learning in Data Science

2

Jens Kaufmann and Daniel Retkowitz

Contents

Abstract

People quickly and efficiently take in visually processed information. The processing of data of any kind is therefore an important and heavily researched area in data science and all its surrounding fields. The aim is to simplify complex information as

J. Kaufmann (✉) · D. Retkowitz
Hochschule Niederrhein, FB Wirtschaftswissenschaften, Mönchengladbach, Germany
e-mail: jens.kaufmann@hs-niederrhein.de
D. Retkowitz e-mail: daniel.retkowitz@hs-niederrhein.de

T. Barton and C. Müller (eds.), *Apply Data Science*,
https://doi.org/10.1007/978-3-658-38798-3_2