THE DECISION MAKER'S HANDBOOK TO DATA SCIENCE

A GUIDE FOR NON-TECHNICAL EXECUTIVES, MANAGERS, AND FOUNDERS

SECOND EDITION

Stylianos Kampakis

Apress®

The Decision Maker's Handbook to Data Science: A Guide for Non-Technical Executives, Managers, and Founders

Stylianos Kampakis
London, UK

ISBN-13 (pbk): 978-1-4842-5493-6 ISBN-13 (electronic): 978-1-4842-5494-3
https://doi.org/10.1007/978-1-4842-5494-3

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Shiva Ramachandran
Development Editor: Rita Fernando
Coordinating Editor: Rita Fernando

Cover designed by eStudioCalamar

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail rights@apress.com, or visit http://www.apress.com/rights-permissions.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/978-1-4842-5493-6. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

Contents

About the Author

Dr. Stylianos (Stelios) Kampakis is a data scientist who is living and working in London, UK. He holds a PhD in Computer Science from the University College London as well as an MSc in Informatics from the University of Edinburgh. He also holds degrees in Statistics, Cognitive Psychology, Economics, and Intelligent Systems. He is a member of the Royal Statistical Society and an honorary research fellow in the UCL Centre for Blockchain Technologies.[1] He has many years of academic and industrial experience in all fields of data science like statistical modeling, machine learning, classic AI, optimization, and more.

Throughout his career, Stylianos has been involved in a wide range of projects: from using deep learning to analyze data from mobile sensors and radar devices, to recommender systems, to natural language processing for social media data, to predicting sports outcomes. He has also done work in the areas of econometrics, Bayesian modeling, forecasting, and research design. He also has many years of experience in consulting for startups and scale-ups, having successfully worked with companies of all stages, some of which have raised millions of dollars in funding. He is still providing services in data science and blockchain as a partner in Electi Consulting.

In the academic domain, he is one of the foremost experts in the area of sports analytics, having done his PhD in the use of machine learning for predicting football injuries. He has also published papers in the areas of neural networks, computational neuroscience, and cognitive science. Finally, he is also involved in blockchain research and more specifically in the areas of tokenomics, supply chains, and securitization of assets.

Stylianos is also very active in the area of data science education. He is the founder of The Tesseract Academy,[2] a company whose mission is to help decision makers understand deep technical topics such as machine learning and blockchain. He is also teaching "Social Media Analytics" and "Quantitative

[1] http://blockchain.cs.ucl.ac.uk/
[2] http://tesseract.academy

Methods and Statistics with R" in the Cyprus International Institute of Management[3] and runs his own data science school in London called Datalyst.[4]

Finally, he often writes about data science, machine learning, blockchain, and other topics at his personal blog: The Data Scientist.[5]

In his spare time, Stylianos enjoys (among other things) composing music, traveling, playing sports (especially basketball and training in martial arts), and meditating.

[3] www.ciim.ac.cy/
[4] www.dataly.st/
[5] http://thedatascientist.com/

Introduction

What is *data science*? What is *artificial intelligence*? What is the difference between *artificial intelligence and machine learning*? What is the best algorithm to use for X? How many people should I hire for my data science team? Do I need a recommender system? Is a *deep neural network* a good idea for this use case?

Having dedicated my career to understanding data, and modeling uncertainty, these questions (and many similar ones) have popped up very often in conversations I am having with CEOs, startup founders, and product managers. There are always three common elements:

1. The people involved have a non-technical background.
2. Their business collects data, or is in a position to collect data.
3. They want to use data science but they don't know where to start.

Data science (and all the fields it encompasses such as AI and machine learning) can transform our world on every level: business, political, and individual. In more than one way, this is already happening. Online retailers know what you are going to like, through the use of recommendation engines. Your photographs get automatically tagged through the use of computer vision. Autonomous vehicles can drive us around with no driver in the seat.

However, this is still only a fraction of the things that are possible with data science. The benefits of this powerful technology will never be reaped, unless the entrepreneurs and the decision makers fully understand how to use it.

Data science is unique in the space of technology in two ways. First in contrast to software development, it is intangible. You can't see a flashy front end, but only the results of a model. Secondly, it is *science*, which means that, in contrast to engineering, it is difficult to set out a perfectly laid plan in advance. Uncertainty is an integral part of data science, and this can make estimations and decisions more difficult. These two factors make the understanding of data science more challenging.

The cloud of buzzwords currently dominating technology does not really help at all with that. I have seen buzzwords such as "big data," "analytics," "prediction," "forecasting," and many more used without any real understanding of the context surrounding them. This is partly due to aggressive sales tactics that end up confusing rather than illuminating. The result is that many entrepreneurs end up feeling more insecure, since they can't understand what they need and how much is a fair price to pay. This is why I realized that it is upon us, the data scientists, to take up the role of educators.

This book is meant to be the ultimate short handbook for a decision maker who wants to use data science but is not sure where to start. All case studies outlined are always described with the decision maker in mind. The problem in business is not how to choose the right model from a scientific viewpoint but how to deliver *value*. The data scientist has to make decisions based on trade-offs such as the cost of development (which can include time and hiring), the interplay with business decisions, and the cost of data. The book explains how the decision maker can better understand these dilemmas and help the data scientist make the most beneficial choices for the business.

I hope that after reading this book, the world of data science will no longer be a dangerous landscape dominated by buzzwords and incomprehensible algorithms, but rather a place of wonder, a place where the future lies. I do hope that you will enjoy reading it as much as I enjoyed writing it.