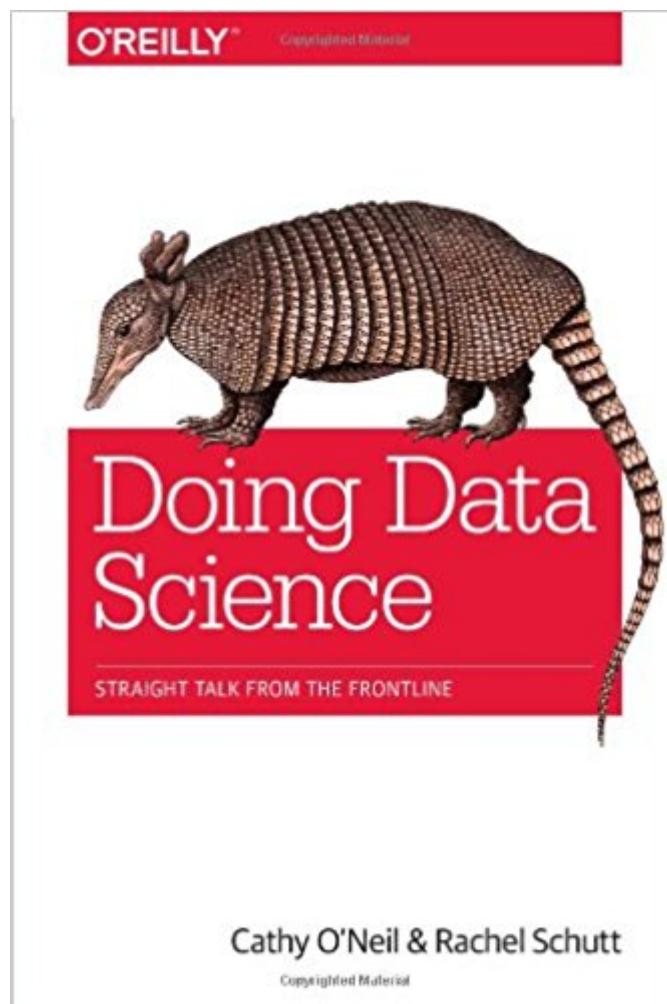


The book was found

Doing Data Science: Straight Talk From The Frontline



Synopsis

Now that people are aware that data can make the difference in an election or a business model, data science as an occupation is gaining ground. But how can you get started working in a wide-ranging, interdisciplinary field that's so clouded in hype? This insightful book, based on Columbia University's Introduction to Data Science class, tells you what you need to know. In many of these chapter-long lectures, data scientists from companies such as Google, Microsoft, and eBay share new algorithms, methods, and models by presenting case studies and the code they use. If you're familiar with linear algebra, probability, and statistics, and have programming experience, this book is an ideal introduction to data science. Topics include:

- Statistical inference
- Exploratory data analysis
- The data science process
- Algorithms
- Spam filters, Naive Bayes, and data wrangling
- Logistic regression
- Financial modeling
- Recommendation engines and causality
- Data visualization
- Social networks and data journalism
- Data engineering, MapReduce, Pregel, and Hadoop

Doing Data Science is collaboration between course instructor Rachel Schutt, Senior VP of Data Science at News Corp, and data science consultant Cathy O'Neil, a senior data scientist at Johnson Research Labs, who attended and blogged about the course.

Book Information

Paperback: 408 pages

Publisher: O'Reilly Media; 1 edition (November 3, 2013)

Language: English

ISBN-10: 1449358659

ISBN-13: 978-1449358655

Product Dimensions: 6 x 0.8 x 9 inches

Shipping Weight: 1.3 pounds (View shipping rates and policies)

Average Customer Review: 4.0 out of 5 stars [See all reviews](#) (49 customer reviews)

Best Sellers Rank: #37,291 in Books (See Top 100 in Books) #1 in Books > Science & Math > Mathematics > Applied > Stochastic Modeling #11 in Books > Textbooks > Computer Science > Algorithms #17 in Books > Computers & Technology > Programming > Algorithms

Customer Reviews

Book review - Doing Data Science by O'Neil and Schutt, O'Reilly Media. More breadth than depth

What is data science? The book Doing Data Science not only explains what data science is but also provides a broad overview of methods and techniques that one must master in order to call one self a data scientist. The book is based on a course about data science given at Columbia

University. However it is not to be considered as a text book about data science but more as a broad introduction to a number of topics in data science. In the spring of 2013 I followed two Coursera courses. One about the statistical programming language R and one on Data Analysis. I had for some time been looking for a book that could be used as a follow-up reading on topics in data science. This was the reason I picked up "Doing Data Science". The book begins with a chapter about what data science is all about is followed by four chapters on topics like statistical inference, explanatory data analysis, various machine learning algorithms, linear and logistic regression, and Naive Bayes. I have a background in both mathematics and statistics and I was able to understand these chapters but the material is covered in such broad terms that I find it hard to believe that a newcomer to this topics will understand or gain much knowledge from reading these chapters. Basic math is presented about the models but without some kind of detailed explanation one cannot develop any deeper intuition for the approach explained. The best parts of the book is definitely chapter 6 to 8 and 10. In here we find interesting discussion about coverage of data science applied to financial modeling, extracting information from data, and social networks.

... helps the medicine go down, as Mary Poppins used to say. An IT-focused publisher, O'Reilly has twice before used the "book as collection of chapters by different contributors" formula in its foray into the attractive "data" niche, with such titles as "Beautiful data" and "Bad data". "Doing data science" - by the way, I prefer Hastie and Tibshirani's "statistical learning" to the fuzzy and grandiose "data science" - follows the same approach, but, with its subject matter being closer to the academe, the company enlisted two young PhDs to steer the collaborative effort. Rachel Schutt took the lead as author and editor, and, assisted by Cathy O'Neil, produced an engaging, informal - you don't often see "science" in the title and "huge-ass" in the text - yet sufficiently technical to be hands-on, sequence-of-vignettes-styled book. Imagine a mash-up of a magazine article and a textbook. Neither part may be best-in-class, but their combination makes for a "unique selling proposition". Well, maybe not a textbook. Most textbooks are carefully written and carefully checked. In contrast, when I see "Doing data science" introduce the ROC curve in three places, one of which translates the "O" as "operator", I can guess that this is a copy-paste of papers by three contributors. When Dr. O'Neil casually redefines an English word ("causal") to avoid rewriting a couple of sentences, or pronounces, on page 159, that "priors reduce degrees of freedom" - this is painfully meaningless, and neither term is defined, only name-checked - I suspect that she knows better, but just did not feel like spending more time on her half-chapter. Neither author speaks of their own projects - if this is the "frontline", then it's other soldiers' "trenches" that we are visiting.

This is a beautiful, thoughtful survey with excellent references. I am an academic data scientist with nearly 20 years experience and I wanted a book to offer my students who are starting in the field. This is it. The "difficulty" with data science is in the breadth of skills that are needed. Because data scientists need training in art, communication, statistics, and programming nobody is prepared to handle all the tasks and the neophyte (and expert) will need to fill in around their weaknesses. This book does a brilliant job of working around that issue. The writing is superb for a beginning to intermediate reader and the graphics and aside boxes are engaging. More importantly, the references are plentiful and spot on. In the areas I know well the authors suggest the things I recommend and where I am weak the recommendations have proven interesting. While this is a broad survey, there is some depth here. There are formulas throughout but the book does not get bogged down in proofs and derivations. There are programs written in R code scattered throughout. The code is nicely commented but there is not a deep dive into how it works. So, the reader who knows some R will learn a few new tricks but it does not interrupt the flow of the book. A reader who types the R code will run into problems. Clearly the authors/editors did not attempt to run the code after the type setter mangled it. For example, on page 39 there is a line which begins with a + and that character needed to be on the previous line. In other places, (like page 49) functions are invoked (count) but the authors have not included the commands to make the functions available (in this case library(plyr)).

[Download to continue reading...](#)

Doing Data Science: Straight Talk from the Frontline Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data Science, ... (Hacking Freedom and Data Driven Book 2) Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault Conversation: The Gentle Art Of Hearing & Being Heard - HowTo "Small Talk", How To Connect, How To Talk To Anyone (Conversation skills, Conversation starters, Small talk, Communication) Big Data For Beginners: Understanding SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More! The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences The Dead Straight Guide to The Beatles (Dead Straight Guides) Straight Man Seduced by the Gay Priest (Taken by the Gay Priest 1)(MM, MMM, Menage, Straight-to-Gay, First Time Erotica) The Battle for Wau: New Guinea's Frontline 1942-1943 (Australian Army History Series) Syria Speaks: Art and Culture from the Frontline Teacher Created Materials - TIME For Kids Informational Text: Straight Talk: Drugs and Alcohol - Grade 4 - Guided Reading Level R (Time for

Kids Nonfiction Readers: Level 4.5) Teacher Created Materials - TIME For Kids Informational Text: Straight Talk: Drugs and Alcohol - Hardcover - Grade 4 - Guided Reading Level R (Time for Kids Nonfiction Readers) Divorce - Remarriage and the Innocent Spouse: Counseling for Betrayed Believers (Straight Talk Bible Study) (Volume 1) Dirty Talk : Secrets For Women and men, Straight, Gay and Bi, Spice Up Your Sex Life and Have Mindblowing Sex: (Sexuality, Intimacy, Sexting, Confidence, Relationship) (Great Sex Book Series 1) Life, Liberty, and the Pursuit of Healthiness: Dr. Dean's Straight-Talk Answers to Hundreds of Your Most Pressing Health Questions Psychopharmacology: Straight Talk on Mental Health Medications, Third Edition Straight Talk about Psychiatric Medications for Kids, Fourth Edition The Disappearance of the Universe: Straight Talk about Illusions, Past Lives, Religion, Sex, Politics, and the Miracles of Forgiveness Discovering Knowledge in Data: An Introduction to Data Mining (Wiley Series on Methods and Applications in Data Mining) Big Data, MapReduce, Hadoop, and Spark with Python: Master Big Data Analytics and Data Wrangling with MapReduce Fundamentals using Hadoop, Spark, and Python

[Contact Us](#)

[DMCA](#)

[Privacy](#)

[FAQ & Help](#)