

Algorithms for data science

Steele, Brian, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20510037&lokasi=lokal>

Abstrak

This textbook on practical data analytics unites fundamental principles, algorithms, and data. Algorithms are the keystone of data analytics and the focal point of this textbook. Clear and intuitive explanations of the mathematical and statistical foundations make the algorithms transparent. But practical data analytics requires more than just the foundations. Problems and data are enormously variable and only the most elementary of algorithms can be used without modification. Programming fluency and experience with real and challenging data is indispensable and so the reader is immersed in Python and R and real data analysis. By the end of the book, the reader will have gained the ability to adapt algorithms to new problems and carry out innovative analyses.

This book has three parts:

- (a) **Data Reduction:** Begins with the concepts of data reduction, data maps, and information extraction. The second chapter introduces associative statistics, the mathematical foundation of scalable algorithms and distributed computing. Practical aspects of distributed computing is the subject of the Hadoop and MapReduce chapter.
- (b) **Extracting Information from Data:** Linear regression and data visualization are the principal topics of Part II. The authors dedicate a chapter to the critical domain of Healthcare Analytics for an extended example of practical data analytics. The algorithms and analytics will be of much interest to practitioners interested in utilizing the large and unwieldy data sets of the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System.
- (c) **Predictive Analytics** Two foundational and widely used algorithms, k-nearest neighbors and naive Bayes, are developed in detail. A chapter is dedicated to forecasting. The last chapter focuses on streaming data and uses publicly accessible data streams originating from the Twitter API and the NASDAQ stock market in the tutorials.

This book is intended for a one- or two-semester course in data analytics for upper-division undergraduate and graduate students in mathematics, statistics, and computer science. The prerequisites are kept low, and students with one or two courses in probability or statistics, an exposure to vectors and matrices, and a programming course will have no difficulty. The core material of every chapter is accessible to all with these prerequisites. The chapters often expand at the close with innovations of interest to practitioners of data science. Each chapter includes exercises of varying levels of difficulty. The text is eminently suitable for self-study and an exceptional resource for practitioners.