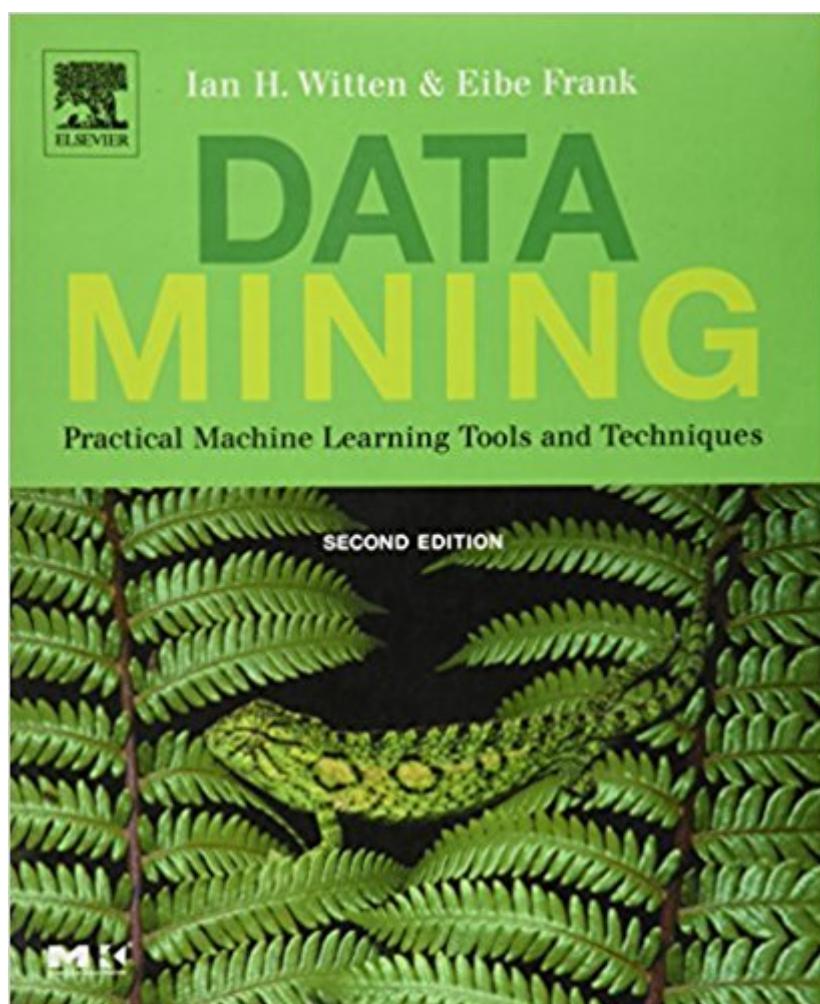


The book was found

# **Data Mining: Practical Machine Learning Tools And Techniques, Second Edition (Morgan Kaufmann Series In Data Management Systems)**



## **Synopsis**

Data Mining, Second Edition, describes data mining techniques and shows how they work. The book is a major revision of the first edition that appeared in 1999. While the basic core remains the same, it has been updated to reflect the changes that have taken place over five years, and now has nearly double the references. The highlights of this new edition include thirty new technique sections; an enhanced Weka machine learning workbench, which now features an interactive interface; comprehensive information on neural networks; a new section on Bayesian networks; and much more. This text is designed for information systems practitioners, programmers, consultants, developers, information technology managers, specification writers as well as professors and students of graduate-level data mining and machine learning courses. Algorithmic methods at the heart of successful data mining— including tried and true techniques as well as leading edge methods—Performance improvement techniques that work by transforming the input or output

## **Book Information**

Series: The Morgan Kaufmann Series in Data Management Systems

Paperback: 560 pages

Publisher: Morgan Kaufmann; 2 edition (June 22, 2005)

Language: English

ISBN-10: 0120884070

ISBN-13: 978-0120884070

Product Dimensions: 7.5 x 1.1 x 9.2 inches

Shipping Weight: 2.4 pounds (View shipping rates and policies)

Average Customer Review: 3.7 out of 5 stars 38 customer reviews

Best Sellers Rank: #448,277 in Books (See Top 100 in Books) #143 in Books > Textbooks > Computer Science > Artificial Intelligence #169 in Books > Computers & Technology > Networking & Cloud Computing > Network Administration > Storage & Retrieval #278 in Books > Computers & Technology > Databases & Big Data > Data Mining

## **Customer Reviews**

“This book presents this new discipline in a very accessible form: both as a text to train the next generation of practitioners and researchers, and to inform lifelong learners like myself. Witten and Frank have a passion for simple and elegant solutions. They approach each topic with this mindset, grounding all concepts in concrete examples, and urging the reader to consider the simple techniques first, and then progress to the more sophisticated ones if the simple ones prove

inadequate. If you have data that you want to analyze and understand, this book and the associated Weka toolkit are an excellent way to start. • From the foreword by Jim Gray, Microsoft Research  
It covers cutting-edge, data mining technology that forward-looking organizations use to successfully tackle problems that are complex, highly dimensional, chaotic, non-stationary (changing over time), or plagued by. The writing style is well-rounded and engaging without subjectivity, hyperbole, or ambiguity. I consider this book a classic already! • Dr. Tilman Bruckhaus, StickyMinds.com

Highly anticipated second edition of the highly-acclaimed reference on data mining and machine learning.

I chose this book after looking at a number of options. I was not disappointed. The text is clearly written for individuals with a bachelor-level education in computer science. The author prefers pseudocode and text explanations of algorithms to equations, and when he does use equations they use clear, commonly understandable notation rather than the terse greek alphabet soup preferred by many of the more mathematically oriented authors. It should be pointed out that about 10% of the text of this book is devoted simply as a user manual for an open source Java package called Weka. When I first realized this I almost flipped; I really didn't want a book that was devoted to gaining a surface understanding of a particular implementation of a set of algorithms. After reading through, I can tell you it is not. All the algorithms are explained well enough that you could implement them and work out simple examples on paper. I should note also that Weka, as well as a lot of the algorithms in this book, don't parallelize well (or obviously). This is an excellent point to get your feet wet and do some exploratory analysis, but if you're past that point and want to learn about crunching big (TB+) data you should look elsewhere. One area that the text does not cover (and, for many software engineers this is not a fault) is some of the mathematics behind some of the algorithms the author proposes. For instance, in the author's description of linear regression using SGD he glosses over the math of actually calculating the gradient by saying "there's a matrix inversion involved and it's available in prepackaged software." I'm not saying this is bad, because if you're a good software engineer the first thing you'll do is look for an existing implementation that you can alter to fit your needs, so he's right. It just may not be what mathematicians or more theory-oriented computer scientists expect.

This book is very easy to read and understand. Unlike Hastie's Statistical Learning book, it is not

geared towards those with an expert level knowledge of statistics, and instead takes time to explain functions and formulas for the person with a decent but not extraordinary understanding of statistical/math concepts. For example, their description of a Gaussian was the clearest I've seen. On the other hand, if you're math/statistics background is considerable, you may find this book somewhat simplistic or tedious. The book has a good coverage of techniques and algorithms, although I was somewhat disappointed that they do not mention Influence Diagrams, considering the amount of coverage of both decision trees and Bayesian techniques. Their discussion of Combining Multiple Models, however, is well done, and is not covered to this extent in most books I've seen. I also like how they broke out the discussion of input and output (knowledge representation) into their own chapters.

**Addendum 10/30:** After reading a good hunk of this book I still agree with most of what I said earlier, but I do think the authors could have gone into graphical models a lot more. At the end of the discussion on Bayesian networks, Markov networks and other graphical models are mentioned very briefly and the author says they are very big in ML right now, but he doesn't say why they didn't describe them further. It might have something to do with the organization of the book. Graphical models almost need a chapter of their own but the book's chapters discuss all techniques in one chapter but with varying levels of detail.

I bought this book in the hopes that it would help me better explore the data from the Netflix Prize contest, which it did. I had been reading numerous Wikipedia articles, scientific papers, etc. on line and felt it would be useful to have a more general tome on the subject. This book covers many of the common, overarching themes i.e. clustering, neural networks, linear regression, etc. to varying degree. I only wish the examples involved slightly more complex data sets and more pseudo code was provided. I suppose since the book is very closely tied to WEKA, one could always dig through the source code of that application; but I feel that the authors could have provided a bit more of the strictly algorithm relevant code in the book.

This book was pretty good. I have to admit that for the first hundred or so pages, I was feeling very impatient. All of that information could have been conveyed in about 25 pages, and been much easier to read. But there are some very good examples in here, and it is worth reading. If you are looking for something more technical, try "Pattern Recognition and Machine Learning", by Christopher M. Bishop or "The Elements of Statistical Learning" by Hastie, Tibshirani, and Friedman.

Highly recommend this book for a practical introduction to the theory and applications of Machine Learning. Great book if you are looking to ACTUALLY implement some machine learning systems, prefer to learn via diagrams, a "how-stuff-works"-style explanation, and skip much of the equations and heavy math that fills similar books. Obviously, this book is a perfect companion to the Weka machine toolbox, which is quickly becoming a standard, invaluable research toolbox for many.

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

Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems) Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (Morgan Kaufmann Series in Data Management Systems) Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) Big Data For Business: Your Comprehensive Guide to Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Discovering Knowledge in Data: An Introduction to Data Mining (Wiley Series on Methods and Applications in Data Mining) Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton Series in Modern Observational Astronomy) Learning Processing, Second Edition: A Beginner's Guide to Programming Images, Animation, and Interaction (The Morgan Kaufmann Series in Computer Graphics) Computer Networks, Fifth Edition: A Systems Approach (The Morgan Kaufmann Series in Networking) VLSI Test Principles and Architectures: Design for Testability (The Morgan Kaufmann Series in Systems on Silicon) Computer Networks: A Systems Approach (The Morgan Kaufmann Series in Networking) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) See MIPS Run, Second Edition (The Morgan Kaufmann Series in Computer Architecture and Design) Computer Organization and Design MIPS Edition, Fifth Edition: The Hardware/Software Interface (The Morgan Kaufmann Series in Computer Architecture and Design) Computer Organization and Design, Fourth Edition: The Hardware/Software Interface (The Morgan Kaufmann Series in Computer Architecture and Design) Machine Learning: An Algorithmic Perspective, Second Edition (Chapman & Hall/Crc Machine Learning & Pattern Recognition) Foundations of Analog and Digital Electronic Circuits (The Morgan Kaufmann Series in Computer Architecture and Design) Self-Checking and Fault-Tolerant Digital Design (The Morgan Kaufmann Series in Computer Architecture and Design) Logical Effort:

Designing Fast CMOS Circuits (The Morgan Kaufmann Series in Computer Architecture and Design)

[Contact Us](#)

[DMCA](#)

[Privacy](#)

[FAQ & Help](#)