# Book Review

## Data Science for Business – What You Need To Know About Data Mining and Data-Analytic Thinking by Foster Provost & Tom Fawcett, published by O'Reilly Media, Inc., 1006 Gravenstein Highway North, Sebastopol, CA 95472, Pages 386, Hardcover, $49.99

Girish Pai B.
*Assistant Professor, Marketing, MSNIM, Mangalore.*
*Email: girish.pai@msnim.edu.in*

---

In the beginning of my journey into learning Data Science, I needed a better understanding of its fundamental principles without it being too technical by including mathematical jargons and coding. I extensively researched many of the blog sites sharing articles about Data Science and came across the mention of this book several times. I found this extremely useful book available on Amazon. This book popularly known as the Provost & Fawcett's "Data Science for Business" is a non-technical guide to the fundamental principles of Data Science. It walks the reader through the "data analytical thinking" necessary for extracting useful knowledge and business value from the data collected which eventually helps with effective problem solving. This book discusses the fundamental concepts and principles that form the basic analysis of data-centered business problems and serve as the foundation for many well-known algorithms of data mining.

This book is very useful for individuals aspiring to be Data Scientists including those in senior management positions looking to implement Data science in their organization. The authors of this book who are known experts in the field of Data Science have simplified the complexity of the topics making them accessible to individuals at all levels of expertise. One of the global industry leaders, Mr. Craig Vaughan, the ex-Global Vice President at SAP has mentioned this book as "A must-read resource for anyone who is serious about embracing the opportunity of Big Data".

This book explains several key algorithms like decision trees, support vector machines, k-nearest neighbors, logistic regression and Term Frequency Inverse Document (TF-ID) without using coding terminologies making it a good read for individuals who are beginners in the world of data science. This book contains 14 chapters with chapter 1 and 2 introducing us to data analytical thinking and data science solutions to business problems. Chapter 3 to 5 explain the concepts of predictive modelling ranging from correlation to supervised segmentation and fitting a model to the data while providing explanation of the over fitting of the model and its avoidance. Chapters 6 to 8 walk us through different algorithms based on similarity, neighbors and clusters, discussing decision analytic thinking and deciding on a good model and visualizing the model performance. Chapters 9 to 12 explain the probabilistic models based on Naïve Baye's rules and applying it in data science, data mining and different

techniques employed in it. It explains analytical engineering, tools and techniques available and other data science tasks and techniques. Chapters 13 and 14 walk us through data science and business strategy explaining the fundamental concepts of data science and its applications in solving business problems, thus changing the way we think about the solutions to them.

The readers of this book are expected to have basic knowledge in the areas of statistics including descriptive and inferential statistics, simple, multiple and logistic regression, mathematics including algebraic concepts like matrices and calculus – both integral and differential. The concepts in this book are very well explained giving a lot of real life business examples.

Overall, this book is highly recommended for data science students and can be used as a text book for subjects like Fundamentals of Data Science. This book also holds relevance to senior management leaders who want to build and lead a team of data scientists and implement data science in solving complex business problems.