Business Analytics
Methods & Cases for Data-Driven Decisions

Richard Huntsinger

University of California, Berkeley

Preface

Business Analytics: Methods & Cases for Data-Driven Decisions is about leveraging data analysis and analytical modeling methods to better achieve business objectives. Successful business analysts, managers, and executives are increasingly required to leverage newly available data sources – voluminous and varied – to inform their decisions about how to best run their businesses. This book presents a practical data-to-decision process, a rigorous study of 43 data analysis and analytical modeling methods, 14 business cases, and a practicum on implementation skills. It introduces and delves into exploratory data analysis based on statistics and data visualization, descriptive and predictive modeling based on machine learning, and data-driven decision making based on decision science principles and model evaluation. Discussions and examples guide students through how these methods work, expose their individual strengths and weaknesses, and show how to apply them for better business results. The business cases, all based on real industry data, underscore their usefulness.

This book is intended for undergraduate and graduate business students with special interest in data science, data science students with special interest in business, and other students with interest in both. It sits at the sweet spot between a cursory survey of business-oriented data science concepts and an in-depth study of statistical learning theory. Methods are presented by appealing to intuition and their business applications, backed up by an appropriate level of mathematical rigor, making the material accessible to students with a broad range of backgrounds: those looking for "So that's how it works!", those looking for "So that's what it's for!", and those looking for "So that's how to do it!" Students will come away well-positioned and well-differentiated for their future industry, consulting, academic, and government roles as data-savvy business practitioners.

Key Features

Designed for the business-oriented and data-savvy. Methods are explained largely through demonstration, independent of specific technology and always with an eye to how they can be used to inform business decisions. Concepts are developed both bottom-up, building upon simple principles, and top-down, deconstructing complex structures. Details relevant to business decisions, like what assumptions go into the methods and how they perform in various scenarios, are covered rigorously, while other theoretical aspects are covered to an appropriate degree. Copious data tables and data visualizations support the discourse.

Analytical methods with a purpose. In practice, methods are useful only when they are applied to analyses for which they are well-suited. Through copious examples, guidance is provided about how to choose the right methods and data, or combinations of methods and data, to best address a wide variety of business decisions.

Comprehensive collection of analytical methods. A wide array of well-established and leading-edge methods from the exploratory data analysis and machine learning repertoires are represented. The data-to-decision process serves as a context, clarifying how each method fits into the bigger picture. Method presentations are sequenced logically according to the process, but written to also serve later as stand-alone references.

Professional-grade business cases based on real industry data. Several robust business cases complement the method presentations, representing a variety of business functions and industries. All datasets, large and small, are sourced from real industry operations. Only minimal pre-processing is performed on the datasets, so as to provide students practice dealing with the kinds of data sourcing issues they will likely encounter in their future business roles.

Downloadable reference of method implementations in R. A downloadable reference of over 50 examples of method implementations in R code accompanies this book, along with a library of custom R utility functions and links to select open source libraries of other useful R functions, that demonstrate exactly how to realize the methods and apply them in professional analyses.

Complete course package. A rich set of additional online resources accompanies this book, so that instructors can conveniently stand-up new course offerings and enhance existing ones built around the methods and business cases, leveraging classroom-tested lesson plans, homework and exam questions, and other administrative materials. For course offerings that will include coding-based analysis, instructors and students have access to primers, lab work, software tools, guest speaker videos, and examples in interactive Jupyter notebooks running R code.

Experience

The material in this book has been used successfully in several offerings of Advanced Business Analytics, a popular one-semester, upper-division course in the globally top-10 ranked business analytics program at the Haas School of Business, University of California-Berkeley. The course has been attended by students with a variety of backgrounds – some Business and Data Science majors, but also Economics, Mathematics, Statistics, Computer Science, Operations Research, Engineering, Public Policy, and other Letters & Science majors. The material has also been used in offerings of Applied Machine Learning, a one-semester graduate course in the globally top-ranked data science program at the Graduate School of Information, University of California-Berkeley.

Topics

This book is organized into an executive overview and 11 chapters, each comprising lessons on related methods and business cases.

Executive Overview

We start by setting expectations with a concise overview of everything you will learn about business analytics.

Data & Decisions

We set the stage with an introduction to an end-to-end data-to-decision process that will serve as context and show the relationships between the variety of specific methods that we will cover. Also, we describe how to construct decision models using influence diagrams to analyze decision performance sensitivity to business parameters. This material appears early to help keep our subsequent lessons on more technical material expressed in ways relevant to making business decisions.

Introduction • Data-to-Decision Process Model • Decision Models • Sensitivity Analysis

Data Analysis | Data Preparation, Data Exploration, & Data Transformation

We start our discourse in chapters 2 through 4 with methods for data analysis and associated cases demonstrating their business application. These methods enable us to get the right part of the data, expose patterns by examining the data from a variety of perspectives, expose the underlying process that may have generated the data, and change the representation of the data to suit a variety of data analysis approaches.

Data Objects • Selection • Amalgamation • Synthetic Variables • Normalization • Dummy Variables • Descriptive Statistics • Similarity • Cross-Tabulation • Data Visualization • Kernel Density Estimation • Balance • Imputation • Alignment • Principal Component Analysis

Predictive Modeling | Classification, Regression, & Ensemble Assembly

We proceed in chapters 5 through 9 onto methods for predictive modeling. We overview how to build models to predict binary categorical results and evaluate quantitatively how much they could help our business. We look into several specific binary classifier construction methods: k-nearest neighbors based on similarities between various parts of the data, logistic regression based on a mathematical function approximated by the data, decision tree based on probabilities inferred from the data, naive Bayes based on probabilities inferred from the data in another way, support vector machine based on boundaries inferred from the data, and perceptron and neural network inspired by how biological neural networks work. Each of these methods can be enhanced to produce multinomial categorical results. Following this, we overview how to build models to predict numerical results and evaluate quantitatively how much they could help our business, and look into linear regression and regression forms of the classifier construction methods. Regression is treated as a variation on classification, with several concepts introduced in the lessons on classification carrying over to regression. Ultimately, we see how to combine predictive models using bagging, boosting, and stacking ensemble assembly methods. For both classification and regression, we see how to systematically tune predictive models to most help our business.

Classification Methodology • Classifier Evaluation • k-Nearest Neighbors • Logistic Regression • Decision Tree • Naive Bayes • Support Vector Machine • Neural Network • Multinomial Classification • Regression Methodology • Regressor Evaluation • Linear Regression • Regression Versions • Bagging • Boosting • Stacking

Descriptive Modeling | Cluster Analysis

In chapter 10, we introduce methods for cluster analysis that can enable us to discover how customers and other entities are similar to and dissimilar from each other. We start with an overview of cluster analysis and then proceed to look deeper into three methods to cluster the data, each using its own unique approach.

Cluster Analysis Methodology • Cluster Model Evaluation • k-Means • Hierarchical Agglomeration • Gaussian Mixture

Special Data Types

The lessons on modeling methods have so far assumed the availability of data in cross-sectional form. In chapter 11, we extend the applicability of these methods and introduce new methods to deal with text data, time series data, and network data represented in other forms.

Text Data • Time Series Data • Network Data • PageRank • Collaborative Filtering

Business Cases

Several robust business cases complement the method presentations, representing a variety of business functions and industries.

High-Tech Stocks: Data preparation with selection, amalgamation, and synthetic variables to discover relationships among companies' financial performance in the high-tech industry.

Fundraising Strategies: Data exploration with descriptive statistics, synthetic variables, descriptive statistics, cross-tabulation, and data visualization to discover effective fundraising strategies from a historical US presidential campaign.

Iowa Liquor Sales: Data exploration with kernel density estimation to discover historical sales patterns that could influence future product management strategies in the beverage industry.

Loan Portfolio: Data transformation with kernel density estimation and principal component analysis to discover relationships predictive of loan defaults in the banking industry.

Loan Portfolio Revisited: Predictive modeling with principal component analysis and k-nearest neighbors to inform loan purchase decisions in the banking industry.

Telecom Customer Churn: Predictive modeling with balance, naive Bayes, cross-validation, and sensitivity analysis to inform a new customer churn reduction strategy in the telecommunications industry.

Truck Fleet Maintenance: Predictive modeling (IIoT industrial internet of things) with balance, principal component analysis, support vector machine, classifier cutoff tuning, and custom test data to inform a new adaptive truck fleet maintenance schedule in the logistics industry.

Facial Recognition: Predictive modeling with principal component analysis and multinomial support vector machine to recognize people in photographs in the gaming industry.

Credit Card Fraud Detection: Comparison of predictive modeling methods to inform a credit card fraud reduction strategy in the financial services industry.

Call Center Scheduling: Predictive modeling with linear regression to inform a partnership agreement between a large retail pet supply client and a business process outsourcing provider.

Fortune 500 Diversity: Cluster analysis to expose race and gender diversity in Fortune 500 companies.

Music Market Segmentation: Cluster analysis with principal component analysis and sensitivity analysis to segment a market for new bands in the music industry.

Deceptive Hotel Reviews: Predictive modeling on text data to detect fraudulent customer surveys and inform a customer service improvement strategy in the hospitality industry.

Targeted Marketing: Network analysis with descriptive statistics for network data to simulate market adoption and inform a marketing strategy in the enterprise software industry.

Online Resources

A rich set of online resources accompanies this book, so that instructors can conveniently stand-up new course offerings and enhance existing ones built around the methods and business cases presented in text.

For Instructors & Students

These resources are available to instructors and students.

Primers: Primers introducing key concepts and coding skills around using Jupyter interactive notebooks, the R language for data analysis, special functions for text and table presentation,

functions for data object type manipulation, ggplot2 functions for data visualization, and rgl functions for interactive three-dimensional data visualization.

Method Implementations: Example implementations of all methods in R, which collectively serve as a comprehensive reference guide and can be leveraged for lab work and project work.

Lab Assignments: Sample lab assignments in R, which can be incorporated into a course offering. Each lab provides students with practice implementing and exploring a specific method.

Project Assignment: A sample term project assignment in R, which can be incorporated into a course offering. The project objective is to recommend a portfolio of equity investments informed by original predictive models based on real financial fundamentals data from thousands of public companies. The project includes sections focusing on data exploration, data transformation, classification and model tuning, regression and model tuning, and model deployment, which can be assigned piecemeal over a term. The project provides students practice implementing the full data-to-decision process. Students have fun pitting their original predictive models against each other, culminating in a friendly contest for the best-performing portfolio.

Function Library: A library of custom R functions to facilitate lessons, labs, and the project.

Datasets: Source data for the business cases and project in csv format.

More Business Cases: Additional business cases, created and available on an ongoing basis, which can be incorporated into a course offering.

Guest Speakers: Video recordings of the author interviewing business practitioners on business analytics applied in real industry settings, created and available on an ongoing basis, which can be incorporated into a course offering.

For Instructors

More resources are available only to instructors.

Syllabi: Sample syllabi for terms of various lengths.

Lesson Plans: Sample lesson plans for terms of various lengths, detailing pace of topic coverage.

Exam Questions & Solutions: Sample exam questions and their solutions.

Lab Solutions: Solutions for the sample lab assignments.

Project Solutions & Reporting Tool: Solutions for the sample term project and a tool to report on students' submissions. Students have fun viewing the report together to see how their recommended portfolios stand up to each other.