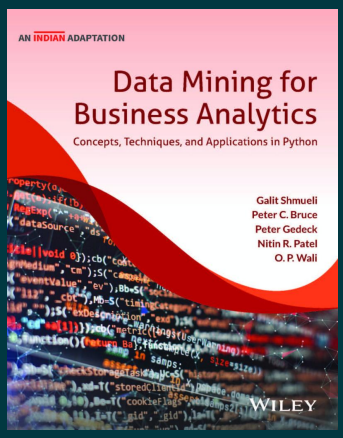
WILEY




# Data Mining for Business Analytics: Concepts, Techniques and Applications in Python (An Indian Adaptation)

By Galit Shmueli, Peter C. Bruce, Peter Gedeck, Nitin R. Patel, O.P. Wali

Paperback 9789357461672 ₹1,099.00

## Description

Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python presents an applied approach to data mining concepts and methods, using python software for illustration. Readers will learn how to implement a variety of popular data mining algorithms in Python to tackle business problems and opportunities. It is a comprehensive resource for students pursuing courses in data mining, business analytics, and related courses within the domain of AI.

Click to Download Datasets

## About the Author

**Galit Shmueli, Peter C. Bruce, Peter Gedeck, Nitin R. Patel, O.P. Wali**

Galit Shmueli, PhD

## Table of Contents

Part I Preliminaries

Chapter 1 Introduction

1.1 What Is Business Analytics?

1.2 What Is Data Mining?

1.3 Data Mining and Related Terms

1.4 Big Data

1.5 Data Science

1.6 Why Are There So Many Different Methods?

1.7 Terminology and Notation

1.8 Road Maps to This Book

Chapter 2 Overview of the Data Mining Process

2.1 Introduction

2.2 Core Ideas in Data Mining

2.3 The Steps in Data Mining

2.4 Preliminary Steps

2.5 Predictive Power and Overfitting

2.6 Building a Predictive Model

2.7 Using Python for Data Mining on a Local Machine
2.8 Automating Data Mining Solutions
2.9 Ethical Practice in Data Mining
Part II Data Exploration And Dimension Reduction
Chapter 3 Data Visualization
3.1 Introduction
3.2 Data Examples
3.3 Basic Charts: Bar Charts, Line Graphs, and Scatter Plots
3.4 Multidimensional Visualization
3.5 Specialized Visualizations
3.6 Summary: Major Visualizations and Operations, by Data Mining Goal
Chapter 4 Dimension Reduction
4.1 Introduction
4.2 Curse of Dimensionality
4.3 Practical Considerations
4.4 Data Summaries
4.5 Correlation Analysis
4.6 Reducing the Number of Categories in Categorical Variables
4.7 Converting a Categorical Variable to a Numerical Variable
4.8 Principal Components Analysis
4.9 Dimension Reduction Using Regression Models
4.10 Dimension Reduction Using Classification and Regression Trees
Part III Performance Evaluation
Chapter 5 Evaluating Predictive Performance
5.1 Introduction
5.2 Evaluating Predictive Performance
5.3 Judging Classifier Performance
5.4 Judging Ranking Performance
5.5 Oversampling
Part IV Prediction And Classification Methods
Chapter 6 Multiple Linear Regression
6.1 Introduction
6.2 Explanatory vs. Predictive Modeling
6.3 Estimating the Regression Equation and Prediction
6.4 Variable Selection in Linear Regression
Chapter 7 k-Nearest Neighbors (k-NN)
7.1 The k-NN Classifier (Categorical Outcome)
7.2 k-NN for a Numerical Outcome

7.3 Advantages and Shortcomings of k-NN Algorithms
Chapter 8 The Naive Bayes Classifier
8.1 Introduction
8.2 Applying the Full (Exact) Bayesian Classifier
8.3 Advantages and Shortcomings of the Naive Bayes Classifier
Chapter 9 Classification and Regression Trees
9.1 Introduction
9.2 Classification Trees
9.3 Evaluating the Performance of a Classification Tree
9.4 Avoiding Overfitting
9.5 Classification Rules from Trees
9.6 Classification Trees for More Than Two Classes
9.7 Regression Trees
9.8 Improving Prediction: Random Forests and Boosted Trees
9.9 Advantages and Weaknesses of a Tree
Chapter 10 Logistic Regression
10.1 Introduction
10.2 The Logistic Regression Model
10.3 Example: Rating Indian Quick Food Services
10.4 Evaluating Classification Performance
10.5 Logistic Regression for Multi-Class Classification
10.6 Example of Complete Analysis: Predicting Delayed Flights
Chapter 11 Neural Nets
11.1 Introduction
11.2 Concept and Structure of a Neural Network
11.3 Fitting a Network to Data
11.4 Required User Input
11.5 Exploring the Relationship Between Predictors and Outcome
11.6 Deep Learning
11.7 Advantages and Weaknesses of Neural Networks
Chapter 12 Discriminant Analysis
12.1 Introduction
12.2 Distance of a Record from a Class
12.3 Fisher’s Linear Classification Functions
12.4 Classification Performance of Discriminant Analysis
12.5 Prior Probabilities
12.6 Unequal Misclassification Costs
12.7 Classifying More Than Two Classes

12.8 Advantages and Weaknesses
Chapter 13 Combining Methods: Ensembles and Uplift Modeling
13.1 Ensembles
13.2 Uplift (Persuasion) Modeling
13.3 Summary
Part V Mining Relationships Among Records
Chapter 14 Association Rules and Collaborative Filtering
14.1 Association Rules
14.2 Collaborative Filtering
14.3 Summary
Chapter 15 Cluster Analysis
15.1 Introduction
15.2 Measuring Distance Between Two Records
15.3 Measuring Distance Between Two Clusters
15.4 Hierarchical (Agglomerative) Clustering
15.5 Non-hierarchical Clustering: The k-Means Algorithm
Part VI Forecasting Time Series
Chapter 16 Handling Time Series
16.1 Introduction
16.2 Descriptive vs. Predictive Modeling
16.3 Popular Forecasting Methods in Business
16.4 Time Series Components
16.5 Data-Partitioning and Performance Evaluation
Chapter 17 Regression-Based Forecasting
17.1 A Model with Trend
17.2 A Model with Seasonality
17.3 A Model with Trend and Seasonality
17.4 Autocorrelation and ARIMA Models
Chapter 18 Smoothing Methods
18.1 Introduction
18.2 Moving Average
18.3 Simple Exponential Smoothing
18.4 Advanced Exponential Smoothing
Part VII Data Analytics
Chapter 19 Social Network Analytics
19.1 Introduction
19.2 Directed vs. Undirected Networks
19.3 Visualizing and Analyzing Networks

19.4 Social Data Metrics and Taxonomy
19.5 Using Network Metrics in Prediction and Classification
19.6 Collecting Social Network Data with Python
19.7 Advantages and Disadvantages
Chapter 20 Text Mining
20.1 Introduction
20.2 The Tabular Representation of Text: Term-Document Matrix and “Bag-of-Words”
20.3 Bag-of-Words vs. Meaning Extraction at Document Level
20.4 Preprocessing the Text
20.5 Implementing Data Mining Methods
20.6 Example: Online Discussions on Autos and Electronics
20.7 Summary
Part VIII Cases
Chapter 21 Cases
21.1 Charles Book Club
21.2 German Credit
21.3 Tayko Software Cataloger
21.4 Political Persuasion
21.5 Taxi Cancellations
21.6 Segmenting Consumers of Bath Soap
21.7 Direct-Mail Fundraising
21.8 Predicting Tourist Travel Packages
21.9 Time Series Case: Forecasting Public Transportation Demand
21.10 Predicting Attrition
21.11 Attitude Toward Therapy Suggestions in Covid-19 Times
References
Data Files Used in the Book
Python Utilities Functions
Index

**To purchase this product, please visit:**
https://www.wileyindia.com/data-mining-for-business-analytics-concepts-techniques-and-applications-in-python-an-indian-adaptation.html

Scan to buy