

Data Science and Machine Learning with R

by Reema Thareja, McGraw Hill Education (India), 2021. Pages: 472, Rs. 650/-

Today, data science & machine learning has become an integral part of both undergraduate as well as postgraduate education, especially in the domains of engineering and information technology & management. In this regard, the textbook by Reema Thareja is a good book that introduces the concepts of data science and machine learning through the open-source statistical computing software R to people having no prior knowledge of the subjects. The book is crisp, well-structured and is lucidly written.

Although not explicitly mentioned, the book can be divided into 4 parts:

Part 1: Chapters 1-3 help students familiarize with the concepts of data science and machine learning algorithms. Machine learning algorithms are further divided into 2 sections: Section 1 (Chapter 2) includes algorithms such as Linear Regression, Logistic Regression, K-Nearest Neighbor, Decision Trees and Random Forests. Section 2 (Chapter 3) deals with algorithms such as K-Means, Naive Bayes, Neural Network and Support Vector Machine.

Part 2: Chapters 4-10 gently introduces the basics of R programming. R is currently the best, industry approved, free and powerful software available for data science. Sections on R programming include fundamentals of data structures, decision controls & loops, and data manipulation in R. Readers also get the exposure to specialized data management R packages (also known as libraries) such as ‘dplyr’ & ‘tidyr’ and data visualization package ‘ggplot2’.

Part 3: Chapter 11 deals with several aspects of Social Media Mining such as Text Mining, Twitter & Facebook Mining and Web Scraping in R.

Part 4: Chapters 12-13 combine the concepts of data science learned in Part 1 and the fundamentals of R programming acquired in Part 2 to implement machine learning algorithms with real datasets for prediction & classification in R.

Some key features of the book include lots of solved examples and concept illustrations, large number of exercises with programming tips at the end of each chapter and seven case studies that provide the experience of real-life scenarios. Overall, the book is best suited for beginners in the field of data science.

Dr. Amarnath Mitra

Associate Professor,
Information Technology (Big Data Analytics),
FORE School of Management,
New Delhi.