



METADATA

Title: Data science through the R language

Other Titles: -

Language: Greek

ISBN: 978-960-603-394-0

Subject: MATHEMATICS AND COMPUTER SCIENCE,
ENGINEERING AND TECHNOLOGY

Keywords: Data Mining / Statistics / Data Science / Big Data
Analytics / Machine Learning

Bibliographic Reference: Verykios, V., Kagklis, V., & Stavropoulos, E. (2015). Data science through the R language [Undergraduate textbook]. Kallipos, Open Academic Editions. <http://dx.doi.org/10.57713/kallipos-734>

Abstract

The increasing accumulation of large amounts of data creates new opportunities in the following areas: science, economy, education, research, etc. In the past, either the storage of such data was not feasible or its analysis was far beyond the computational capabilities of modern computer systems. Today, with the technological convergence in the above mentioned areas, there is a need to train scientists who will be able to cope with modern needs, providing solutions to serious issues through intelligent data analysis. The purpose of this book is to document the basic principles underlying this new data science, also known as Data Mining. The book starts with a description of the current state of the art in data mining, the domains on which it is based and the new challenges. It then covers issues of multidimensional data and report generation, as well as issues of new database

formats, (NoSQL systems). Data types and data quality are an integral part of the book, which together with processing and similarity measures form the basis of a data mining project. Summary statistics and visualization are the first step in data exploration. The main part of the book deals with the detailed description and analysis of alternative algorithms for performing the basic functions of Data Mining, which are Categorization, Correlation Analysis and Cluster Analysis. More advanced issues concerning the estimation of the generated models, their comparison and faster execution are also covered in the last part of the book. All topics are discussed in depth through programs in the R language. A large number of solved exercises help to understand and consolidate the concepts, while solvable problems help the student to develop a comprehensive view on the subject.

