1. Record Nr. UNINA9910795970103321

Autore	Larose Daniel T.
Titolo	Data mining and predictive analytics / / Daniel T. Larose, Chantal D. Larose
Pubbl/distr/stampa	Hoboken, New Jersey : , : John Wiley & Sons, , 2015 2015
ISBN	1-118-86867-6 1-118-86870-6
Edizione	[Second edition.]
Descrizione fisica	1 online resource (824 pages) : illustrations
Collana	Wiley Series on Methods and Applications in Data Mining
Disciplina	006.3/12
Soggetti	Data mining Prediction theory
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Series; Title Page; Copyright; Table of Contents; Dedication; Preface; What is Data Mining? What is Predictive Analytics?; Why is this Book Needed?; Who Will Benefit from this Book?; Danger! Data Mining is Easy to do Badly; "White-Box" Approach; Algorithm Walk-Throughs; Exciting New Topics; The R Zone; Appendix: Data Summarization and Visualization; The Case Study: Bringing it all Together; How the Book is Structured; The Software; Weka: The Open-Source Alternative; The Companion Web Site: www.dataminingconsultant.com; Data Mining and Predictive Analytics as a Textbook; Acknowledgments. Daniel's AcknowledgmentsChantal's Acknowledgments; Part I: Data Preparation; Chapter 1: An Introduction to Data Mining and Predictive Analytics; 1.1 What is Data Mining? What Is Predictive Analytics?; 1.2 Wanted: Data Miners; 1.3 The Need For Human Direction of Data Mining; 1.4 The Cross-Industry Standard Process for Data Mining: CRISP-DM; 1.5 Fallacies of Data Mining; 1.6 What Tasks can Data Mining Accomplish; The R Zone; R References; Exercises; Chapter 2: Data Preprocessing; 2.1 Why do We Need to Preprocess the Data?; 2.2 Data Cleaning; 2.3 Handling Missing Data. 2.4 Identifying Misclassifications2.5 Graphical Methods for Identifying Outliers; 2.6 Measures of Center and Spread; 2.7 Data Transformation; 2.8 Min-Max Normalization; 2.9 Z-Score Standardization; 2.10 Decimal Scaling; 2.11 Transformations to Achieve

Normality; 2.12 Numerical Methods for Identifying Outliers; 2.13 Flag Variables; 2.14 Transforming Categorical Variables into Numerical Variables; 2.15 Binning Numerical Variables; 2.16 Reclassifying Categorical Variables; 2.17 Adding an Index Field; 2.18 Removing Variables that are not Useful; 2.19 Variables that Should Probably not be Removed. 2.20 Removal of Duplicate Records2.21 A Word About ID Fields; The R Zone; R Reference; Exercises; Chapter 3: Exploratory Data Analysis; 3.1 Hypothesis Testing Versus Exploratory Data Analysis; 3.2 Getting to Know The Data Set; 3.3 Exploring Categorical Variables; 3.4 Exploring Numeric Variables; 3.5 Exploring Multivariate Relationships; 3.6 Selecting Interesting Subsets of the Data for Further Investigation; 3.7 Using EDA to Uncover Anomalous Fields; 3.8 Binning Based on Predictive Value; 3.9 Deriving New Variables: Flag Variables; 3.10 Deriving New Variables: Numerical Variables. 3.11 Using EDA to Investigate Correlated Predictor Variables3.12 Summary of Our EDA; The R Zone; R References; Exercises; Chapter 4: Dimension-Reduction Methods; 4.1 Need for Dimension-Reduction in Data Mining; 4.2 Principal Components Analysis; 4.3 Applying PCA to the Houses Data Set; 4.4 How Many Components Should We Extract?; 4.5 Profiling the Principal Components; 4.6 Communalities; 4.7 Validation of the Principal Components; 4.8 Factor Analysis; 4.9 Applying Factor Analysis to the Adult Data Set; 4.10 Factor Rotation; 4.11 User-Defined Composites.

Sommario/riassunto

"This updated second edition serves as an introduction to data mining methods and models, including association rules, clustering, neural networks, logistic regression, and multivariate analysis. The authors apply a unified 'white box' approach to data mining methods and models. This approach is designed to walk readers through the operations and nuances of the various methods, using small data sets, so readers can gain an insight into the inner workings of the method under review. Chapters provide readers with hands-on analysis problems, representing an opportunity for readers to apply their newly-acquired data mining expertise to solving real problems using large, real-world data sets."-- Portion of summary from book.