1.	Record Nr.	UNINA9910817415203321
	Autore	Larose Daniel T.
	Titolo	Discovering knowledge in data : an introduction to data mining / / Daniel T. Larose, Chantal D. Larose
	Pubbl/distr/stampa	Hoboken, New Jersey : , : IEEE, , 2014 ©2014
	ISBN	1-118-87357-2 1-118-87405-6 1-118-87358-0
	Edizione	[2nd ed.]
	Descrizione fisica	1 online resource (336 p.)
	Collana	Wiley Series on Methods and Applications in Data Mining
	Classificazione	COM021040COM021030
	Disciplina	006.3/12
	Soggetti	Data mining
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Note generali	Includes index.
	Nota di bibliografia	Includes bibliographical references at the end of each chapters and index.
	Nota di contenuto	DISCOVERING KNOWLEDGE IN DATA; Contents; Preface; 1 An Introduction to Data Mining; 1.1 What is Data Mining?; 1.2 Wanted: Data Miners; 1.3 The Need for Human Direction of Data Mining; 1.4 The Cross-Industry Standard Practice for Data Mining; 1.4.1 Crisp-DM: The Six Phases; 1.5 Fallacies of Data Mining; 1.6 What Tasks Can Data Mining Accomplish?; 1.6.1 Description; 1.6.2 Estimation; 1.6.3 Prediction; 1.6.4 Classification; 1.6.5 Clustering; 1.6.6 Association; References; Exercises; 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; References; Exercises; Hands-On Analysis; 3 Exploratory Data

| Sommario/riassunto | "This is a new edition of a highly praised, successful reference on data mining, now more important than ever due to the growth of the field and wide range of applications. This edition features new chapters on multivariate statistical analysis, covering analysis of variance and chi-square procedures; cost-benefit analyses; and time-series data analysis. There is also extensive coverage of the R statistical programming language. Graduate and advanced undergraduate students of computer science and statistics, managers/CEOs/CFOs, marketing executives, market researchers and analysts, sales analysts, and medical professionals will want this comprehensive reference"-- |