

1. Record Nr.	UNINA9910818902803321
Autore	Ahlemeyer-Stubbe Andrea
Titolo	A practical guide to data mining for business and industry // Andrea Ahlemeyer-Stubbe, Shirley Coleman
Pubbl/distr/stampa	Chichester, England : , : Wiley, , 2014 ©2014
ISBN	1-118-76337-8 1-118-76370-X 1-118-76372-6
Edizione	[1st edition]
Descrizione fisica	1 online resource (325 p.)
Disciplina	006.3/12
Soggetti	Data mining Marketing - Data processing Management - Mathematical models
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	A Practical Guide to Data Mining for Business and Industry; Copyright; Contents; Glossary of terms; Part I Data Mining Concept; 1 Introduction; 1.1 Aims of the Book; 1.2 Data Mining Context; 1.2.1 Domain Knowledge; 1.2.2 Words to Remember; 1.2.3 Associated Concepts; 1.3 Global Appeal; 1.4 Example Datasets Used in This Book; 1.5 Recipe Structure; 1.6 Further Reading and Resources; 2 Data mining definition; 2.1 Types of Data Mining Questions; 2.1.1 Population and Sample; 2.1.2 Data Preparation; 2.1.3 Supervised and Unsupervised Methods; 2.1.4 Knowledge-Discovery Techniques 2.2 Data Mining Process 2.3 Business Task: Clarification of the Business Question behind the Problem; 2.4 Data: Provision and Processing of the Required Data; 2.4.1 Fixing the Analysis Period; 2.4.2 Basic Unit of Interest; 2.4.3 Target Variables; 2.4.4 Input Variables/Explanatory Variables; 2.5 Modelling: Analysis of the Data; 2.6 Evaluation and Validation during the Analysis Stage; 2.7 Application of Data Mining Results and Learning from the Experience; Part II Data Mining Practicalities; 3 All about data; 3.1 Some Basics; 3.1.1 Data, Information, Knowledge and Wisdom

3.1.2 Sources and Quality of Data; 3.1.3 Measurement Level and Types of Data; 3.1.4 Measures of Magnitude and Dispersion; 3.1.5 Data Distributions; 3.2 Data Partition: Random Samples for Training, Testing and Validation; 3.3 Types of Business Information Systems; 3.3.1 Operational Systems Supporting Business Processes; 3.3.2 Analysis-Based Information Systems; 3.3.3 Importance of Information; 3.4 Data Warehouses; 3.4.1 Topic Orientation; 3.4.2 Logical Integration and Homogenisation; 3.4.3 Reference Period; 3.4.4 Low Volatility; 3.4.5 Using the Data Warehouse
3.5 Three Components of a Data Warehouse: DBMS, DB and DBCS; 3.5.1 Database Management System (DBMS); 3.5.2 Database (DB); 3.5.3 Database Communication Systems (DBCS); 3.6 Data Marts; 3.6.1 Regularly Filled Data Marts; 3.6.2 Comparison between Data Marts and Data Warehouses; 3.7 A Typical Example from the Online Marketing Area; 3.8 Unique Data Marts; 3.8.1 Permanent Data Marts; 3.8.2 Data Marts Resulting from Complex Analysis; 3.9 Data Mart: Do's and Don'ts; 3.9.1 Do's and Don'ts for Processes; 3.9.2 Do's and Don'ts for Handling; 3.9.3 Do's and Don'ts for Coding/Programming; 4 Data Preparation
4.1 Necessity of Data Preparation; 4.2 From Small and Long to Short and Wide; 4.3 Transformation of Variables; 4.4 Missing Data and Imputation Strategies; 4.5 Outliers; 4.6 Dealing with the Vagaries of Data; 4.6.1 Distributions; 4.6.2 Tests for Normality; 4.6.3 Data with Totally Different Scales; 4.7 Adjusting the Data Distributions; 4.7.1 Standardisation and Normalisation; 4.7.2 Ranking; 4.7.3 Box-Cox Transformation; 4.8 Binning; 4.8.1 Bucket Method; 4.8.2 Analytical Binning for Nominal Variables; 4.8.3 Quantiles; 4.8.4 Binning in Practice; 4.9 Timing Considerations; 4.10 Operational Issues
5 Analytics

Sommario/riassunto

Data mining is well on its way to becoming a recognized discipline in the overlapping areas of IT, statistics, machine learning, and AI. Practical Data Mining for Business presents a user-friendly approach to data mining methods, covering the typical uses to which it is applied. The methodology is complemented by case studies to create a versatile reference book, allowing readers to look for specific methods as well as for specific applications. The book is formatted to allow statisticians, computer scientists, and economists to cross-reference from a particular application or method
