

1. Record Nr.	UNINA9910784656403321
Autore	Refaat Mamdouh
Titolo	Data preparation for data mining using SAS [[electronic resource] / / Mamdouh Refaat
Pubbl/distr/stampa	Amsterdam ; ; Boston, : Morgan Kaufmann Publishers, c2007
ISBN	1-281-00538-X 9786611005382 0-08-049100-6
Descrizione fisica	1 online resource (425 p.)
Collana	The Morgan Kaufmann series in data management systems
Disciplina	005.74 006.3/12 22 006.312
Soggetti	Data mining
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 (p. 373-374) and index.
Nota di contenuto	Front Cover; Data Preparation for Data Mining Using SAS; Copyright Page; Contents; List of Figures; List of Tables; Preface; CHAPTER 1. INTRODUCTION; 1.1 The Data Mining Process; 1.2 Methodologies of Data Mining; 1.3 The Mining View; 1.4 The Scoring View; 1.5 Notes on Data Mining Software; CHAPTER 2. TASKS AND DATA FLOW; 2.1 Data Mining Tasks; 2.2 Data Mining Competencies; 2.3 The Data Flow; 2.4 Types of Variables; 2.5 The Mining View and the Scoring View; 2.6 Steps of Data Preparation; CHAPTER 3. REVIEW OF DATA MINING MODELING TECHNIQUES; 3.1 Introduction; 3.2 Regression Models 3.3 Decision Trees3.4 Neural Networks; 3.5 Cluster Analysis; 3.6 Association Rules; 3.7 Time Series Analysis; 3.8 Support Vector Machines; CHAPTER 4. SAS MACROS: A QUICK START; 4.1 Introduction: Why Macros?; 4.2 The Basics: The Macro and Its Variables; 4.3 Doing Calculations; 4.4 Programming Logic; 4.5 Working with Strings; 4.6 Macros That Call Other Macros; 4.7 Common Macro Patterns and Caveats; 4.8 Where to Go From Here; CHAPTER 5. DATA ACQUISITION AND INTEGRATION; 5.1 Introduction; 5.2 Sources of Data; 5.3 Variable Types; 5.4 Data Rollup; 5.5 Rollup with Sums, Averages, and Counts 5.6 Calculation of the Mode5.7 Data Integration; CHAPTER 6.

INTEGRITY CHECKS; 6.1 Introduction; 6.2 Comparing Datasets; 6.3 Dataset Schema Checks; 6.4 Nominal Variables; 6.5 Continuous Variables; CHAPTER 7. EXPLORATORY DATA ANALYSIS; 7.1 Introduction; 7.2 Common EDA Procedures; 7.3 Univariate Statistics; 7.4 Variable Distribution; 7.5 Detection of Outliers; 7.6 Testing Normality; 7.7 Cross-tabulation; 7.8 Investigating Data Structures; CHAPTER 8. SAMPLING AND PARTITIONING; 8.1 Introduction; 8.2 Contents of Samples; 8.3 Random Sampling; 8.4 Balanced Sampling; 8.5 Minimum Sample Size
8.6 Checking Validity of Sample
CHAPTER 9. DATA TRANSFORMATIONS; 9.1 Raw and Analytical Variables; 9.2 Scope of Data Transformations; 9.3 Creation of New Variables; 9.4 Mapping of Nominal Variables; 9.5 Normalization of Continuous Variables; 9.6 Changing the Variable Distribution; CHAPTER 10. BINNING AND REDUCTION OF CARDINALITY; 10.1 Introduction; 10.2 Cardinality Reduction; 10.3 Binning of Continuous Variables; CHAPTER 11. TREATMENT OF MISSING VALUES; 11.1 Introduction; 11.2 Simple Replacement; 11.3 Imputing Missing Values; 11.4 Imputation Methods and Strategy
11.5 SAS Macros for Multiple Imputation
11.6 Predicting Missing Values; CHAPTER 12. PREDICTIVE POWER AND VARIABLE REDUCTION I; 12.1 Introduction; 12.2 Metrics of Predictive Power; 12.3 Methods of Variable Reduction; 12.4 Variable Reduction: Before or During Modeling; CHAPTER 13. ANALYSIS OF NOMINAL AND ORDINAL VARIABLES; 13.1 Introduction; 13.2 Contingency Tables; 13.3 Notation and Definitions; 13.4 Contingency Tables for Binary Variables; 13.5 Contingency Tables for Multicategory Variables; 13.6 Analysis of Ordinal Variables; 13.7 Implementation Scenarios
CHAPTER 14. ANALYSIS OF CONTINUOUS VARIABLES

Sommario/riassunto

Are you a data mining analyst, who spends up to 80% of your time assuring data quality, then preparing that data for developing and deploying predictive models? And do you find lots of literature on data mining theory and concepts, but when it comes to practical advice on developing good mining views find little "how to?" information? And are you, like most analysts, preparing the data in SAS? This book is intended to fill this gap as your source of practical recipes. It introduces a framework for the process of data preparation for data mining, and presents the detailed implementation o
