

1. Record Nr.	UNINA9910790502403321
Autore	McCormick Keith
Titolo	IBM SPSS modeler cookbook / / Keith McCormick [and four others]
Pubbl/distr/stampa	Birmingham : , : Packt Publishing, , [2013] ©2013
ISBN	1-84968-547-9
Edizione	[1st edition]
Descrizione fisica	1 online resource (382 p.)
Disciplina	005.478
Soggetti	Data mining Database management
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	<p>""Cover""; ""Copyright""; ""Credits""; ""Foreword""; ""About the Authors""; ""About the Reviewers""; ""www.PacktPub.com""; ""Table of Contents""; ""Preface""; ""Chapter 1: Data Understanding""; ""Introduction""; ""Using an empty aggregate to evaluate sample size""; ""Evaluating the need to sample from the initial data""; ""Using CHAID stumps when interviewing an SME""; ""Using a single cluster K-means as an alternative to anomaly detection""; ""Using an @NULL multiple Derive to explore missing data""; ""Creating an outlier report to give to SMEs""</p> <p>""Detecting potential model instability early using the Partition node and Feature Selection""""Chapter 2: Data Preparation a€? Select""; ""Introduction""; ""Using the Feature Selection node creatively to remove, or decapitate, perfect predictors""; ""Running a Statistics node on anti-join to evaluate potential missing data""; ""Evaluating the use of sampling for speed""; ""Removing redundant variables using correlation matrices""; ""Selecting variable using the CHAID modeling node""; ""Selecting variables using the Means node""</p> <p>""Selecting variables using single-antecedent association rules""""Chapter 3: Data Preparation a€? Clean""; ""Introduction""; ""Binning scale variables to address missing data""; ""Using a full data model/partial data model approach to address missing data""; ""Imputing in-stream mean or median""; ""Imputing missing values randomly from uniform or normal distributions""; ""Using random imputation to match a variable's distribution""; ""Searching for similar</p>

records using a neural network for inexact matching""; ""Using neuro-fuzzy searching to find similar names""
""Producing longer Soundex codes""""Chapter 4: Data Preparation a€? Construct""; ""Introduction""; ""Building transformations with multiple Derive nodes""; ""Calculating and comparing conversion rates"";
""Grouping categorical values""; ""Transforming high skew and kurtosis variables with a multiple Derive node""; ""Creating flag variables for aggregation""; ""Using Association Rules for interaction detection/feature creation""; ""Creating time-aligned cohorts"";
""Chapter 5: Data Preparation a€? Integrate and Format"";
""Introduction""
""Speeding up merge with caching and optimization settings""""Merging a look-up table""; ""Shuffle-down (nonstandard aggregation)"";
""Cartesian product merge using key-less merge by key""; ""Multiplying out using Cartesian product merge, user source, and derive dummy"";
""Changing large numbers of variable names without scripting"";
""Parsing nonstandard dates""; ""Parsing and performing a conversion on a complex stream""; ""Sequence processing""; ""Chapter 6: Selecting and Building a Model""; ""Introduction""; ""Evaluating balancing with the Auto Classifier""
""Building models with and without outliers""

Sommario/riassunto

This is a practical cookbook with intermediate-advanced recipes for SPSS Modeler data analysts. It is loaded with step-by-step examples explaining the process followed by the experts. If you have had some hands-on experience with IBM SPSS Modeler and now want to go deeper and take more control over your data mining process, this is the guide for you. It is ideal for practitioners who want to break into advanced analytics.
