

1. Record Nr.	UNINA9910795808803321
Autore	Bruce Peter C
Titolo	Data Mining for Business Analytics : Concepts, Techniques, and Applications with XLMiner
Pubbl/distr/stampa	Hoboken : , : John Wiley & Sons, Incorporated, , 2016 ©2016
ISBN	9781118729243 9781118729472
Edizione	[3rd ed.]
Descrizione fisica	1 online resource (549 pages)
Altri autori (Persone)	ShmueliGalit PatelNitin R
Disciplina	005.54
Soggetti	Business--Data processing Electronic books.
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Cover -- Title Page -- Copyright -- Dedication -- Contents -- Foreword -- Preface to the Third Edition -- Preface to the First Edition -- Acknowledgments -- Part I: Preliminaries -- Chapter 1: Introduction -- 1.1 What Is Business Analytics? -- 1.2 What Is Data Mining? -- 1.3 Data Mining and Related Terms -- 1.4 Big Data -- 1.5 Data Science -- 1.6 Why Are There So Many Different Methods? -- 1.7 Terminology and Notation -- 1.8 Road Maps to This Book -- Order of Topics -- Chapter 2: Overview of the Data Mining Process -- 2.1 Introduction -- 2.2 Core Ideas in Data Mining -- Classification -- Prediction -- Association Rules and Recommendation Systems -- Predictive Analytics -- Data Reduction and Dimension Reduction -- Data Exploration and Visualization -- Supervised and Unsupervised Learning -- 2.3 The Steps in Data Mining -- 2.4 Preliminary Steps -- Organization of Datasets -- Sampling from a Database -- Oversampling Rare Events in Classification Tasks -- Preprocessing and Cleaning the Data -- 2.5 Predictive Power and Overfitting -- Creation and Use of Data Partitions -- Overfitting -- 2.6 Building a Predictive Model with XLMiner -- Predicting Home Values in the West Roxbury Neighborhood -- Modeling Process -- 2.7 Using Excel for Data Mining -- 2.8 Automating

Data Mining Solutions -- Data Mining Software Tools: the State of the Market -- Problems -- Part II: Data Exploration and Dimension Reduction -- Chapter 3: Data Visualization -- 3.1 Uses of Data Visualization -- 3.2 Data Examples -- Example 1: Boston Housing Data -- Example 2: Ridership on Amtrak Trains -- 3.3 Basic Charts: Bar Charts, Line Graphs, and Scatter Plots -- Distribution Plots: Boxplots and Histograms -- Heatmaps: Visualizing Correlations and Missing Values -- 3.4 Multidimensional Visualization -- Adding Variables: Color, Size, Shape, Multiple Panels, and Animation. Manipulations: Rescaling, Aggregation and Hierarchies, Zooming, Filtering -- Reference: Trend Line and Labels -- Scaling up to Large Datasets -- Multivariate Plot: Parallel Coordinates Plot -- Interactive Visualization -- 3.5 Specialized Visualizations -- Visualizing Networked Data -- Visualizing Hierarchical Data: Treemaps -- Visualizing Geographical Data: Map Charts -- 3.6 Summary: Major Visualizations and Operations, by Data Mining Goal -- Prediction -- Classification -- Time Series Forecasting -- Unsupervised Learning -- Problems -- Chapter 4: Dimension Reduction -- 4.1 Introduction -- 4.2 Curse of Dimensionality -- 4.3 Practical Considerations -- Example 1: House Prices in Boston -- 4.4 Data Summaries -- Summary Statistics -- Pivot Tables -- 4.5 Correlation Analysis -- 4.6 Reducing the Number of Categories in Categorical Variables -- 4.7 Converting a Categorical Variable to a Numerical Variable -- 4.8 Principal Components Analysis -- Example 2: Breakfast Cereals -- Principal Components -- Normalizing the Data -- Using Principal Components for Classification and Prediction -- 4.9 Dimension Reduction Using Regression Models -- 4.10 Dimension Reduction Using Classification and Regression Trees -- Problems -- Part III: Performance Evaluation -- Chapter 5: Evaluating Predictive Performance -- 5.1 Introduction -- 5.2 Evaluating Predictive Performance -- Benchmark: The Average -- Prediction Accuracy Measures -- Comparing Training and Validation Performance -- Lift Chart -- 5.3 Judging Classifier Performance -- Benchmark: The Naive Rule -- Class Separation -- The Classification Matrix -- Using the Validation Data -- Accuracy Measures -- Propensities and Cutoff for Classification -- Performance in Unequal Importance of Classes -- Asymmetric Misclassification Costs -- Generalization to More Than Two Classes -- 5.4 Judging Ranking Performance. Lift Charts for Binary Data -- Decile Lift Charts -- Beyond Two Classes -- Lift Charts Incorporating Costs and Benefits -- Lift as Function of Cutoff -- 5.5 Oversampling -- Oversampling the Training Set -- Evaluating Model Performance Using a Non-oversampled Validation Set -- Evaluating Model Performance If Only Oversampled Validation Set Exists -- Problems -- Part IV: Prediction and Classification Methods -- Chapter 6: Multiple Linear Regression -- 6.1 Introduction -- 6.2 Explanatory vs. Predictive Modeling -- 6.3 Estimating the Regression Equation and Prediction -- Example: Predicting the Price of Used Toyota Corolla Cars -- 6.4 Variable Selection in Linear Regression -- Reducing the Number of Predictors -- How to Reduce the Number of Predictors -- Problems -- Chapter 7: -Nearest-Neighbors (-NN) -- 7.1 The -NN Classifier (categorical outcome) -- Determining Neighbors -- Classification Rule -- Example: Riding Mowers -- Choosing -- Setting the Cutoff Value -- -NN with More Than Two Classes -- Converting Categorical Variables to Binary Dummies -- 7.2 -NN for a Numerical Response -- 7.3 Advantages and Shortcomings of -NN Algorithms -- Problems -- Chapter 8: The Naive Bayes Classifier -- 8.1 Introduction -- Cutoff Probability Method -- Conditional Probability -- Example 1: Predicting Fraudulent Financial Reporting -- 8.2 Applying the Full (Exact) Bayesian Classifier -- Using the "Assign to

the Most Probable Class" Method -- Using the Cutoff Probability Method -- Practical Difficulty with the Complete (Exact) Bayes Procedure -- Solution: Naive Bayes -- Example 2: Predicting Fraudulent Financial Reports, Two Predictors -- Example 3: Predicting Delayed Flights -- 8.3 Advantages and Shortcomings of the Naive Bayes Classifier -- Problems -- Chapter 9: Classification and Regression Trees -- 9.1 Introduction -- 9.2 Classification Trees. Recursive Partitioning -- Example 1: Riding Mowers -- Measures of Impurity -- Tree Structure -- Classifying a New Observation -- 9.3 Evaluating the Performance of a Classification Tree -- Example 2: Acceptance of Personal Loan -- 9.4 Avoiding Overfitting -- Stopping Tree Growth: CHAID -- Pruning the Tree -- 9.5 Classification Rules from Trees -- 9.6 Classification Trees for More Than two Classes -- 9.7 Regression Trees -- Prediction -- Measuring Impurity -- Evaluating Performance -- 9.8 Advantages, Weaknesses, and Extensions -- 9.9 Improving Prediction: Multiple Trees -- Problems -- Chapter 10: Logistic Regression -- 10.1 Introduction -- 10.2 The Logistic Regression Model -- Example: Acceptance of Personal Loan -- Model with a Single Predictor -- Estimating the Logistic Model from Data: Computing Parameter Estimates -- Interpreting Results in Terms of Odds (for a Profiling Goal) -- 10.3 Evaluating Classification Performance -- Variable Selection -- 10.4 Example of Complete Analysis: Predicting Delayed Flights -- Data Preprocessing -- Model Fitting and Estimation -- Model Interpretation -- Model Performance -- Variable Selection -- 10.5 Appendix: Logistic Regression for Profiling -- Appendix A: Why Linear Regression Is Problematic for a Categorical Response -- Appendix B: Evaluating Explanatory Power -- Appendix C: Logistic Regression for More Than Two Classes -- Problems -- Chapter 11: Neural Nets -- 11.1 Introduction -- 11.2 Concept and Structure of a Neural Network -- 11.3 Fitting a Network to Data -- Example 1: Tiny Dataset -- Computing Output of Nodes -- Preprocessing the Data -- Training the Model -- Example 2: Classifying Accident Severity -- Avoiding Overfitting -- Using the Output for Prediction and Classification -- 11.4 Required User Input -- 11.5 Exploring the Relationship Between Predictors and Response. 11.6 Advantages and Weaknesses of Neural Networks -- Unsupervised Feature Extraction and Deep Learning -- Problems -- Chapter 12: Discriminant Analysis -- 12.1 Introduction -- Example 1: Riding Mowers -- Example 2: Personal Loan Acceptance -- 12.2 Distance of an Observation from a Class -- 12.3 Fisher's Linear Classification Functions -- 12.4 Classification Performance of Discriminant Analysis -- 12.5 Prior Probabilities -- 12.6 Unequal Misclassification Costs -- 12.7 Classifying More Than Two Classes -- Example 3: Medical Dispatch to Accident Scenes -- 12.8 Advantages and Weaknesses -- Problems -- Chapter 13: Combining Methods: Ensembles and Uplift Modeling -- 13.1 Ensembles -- Why Ensembles Can Improve Predictive Power -- Simple Averaging -- Bagging -- Boosting -- Advantages and Weaknesses of Ensembles -- 13.2 Uplift (Persuasion) Modeling -- A-B Testing -- Uplift -- Gathering the Data -- A Simple Model -- Modeling Individual Uplift -- Using the Results of an Uplift Model -- 13.3 Summary -- Problems -- Part V: Mining Relationships among Records -- Chapter 14: Association Rules and Collaborative Filtering -- 14.1 Association Rules -- Discovering Association Rules in Transaction Databases -- Example 1: Synthetic Data on Purchases of Phone Faceplates -- Generating Candidate Rules -- The Apriori Algorithm -- Selecting Strong Rules -- Data Format -- The Process of Rule Selection -- Interpreting the Results -- Rules and Chance -- Example 2: Rules for Similar Book Purchases -- 14.2 Collaborative Filtering -- Data Type

and Format -- Example 3: Netflix Prize Contest -- User-Based Collaborative Filtering: "People Like You" -- Item-Based Collaborative Filtering -- Advantages and Weaknesses of Collaborative Filtering -- Collaborative Filtering vs. Association Rules -- 14.3 Summary -- Problems -- Chapter 15: Cluster Analysis -- 15.1 Introduction. Example: Public Utilities.
