Data Mining for Business Analytics. Concepts, Techniques, and Applications with XLMiner. 3rd Edition

Description:

Praise for the Second Edition

"full of vivid and thought-provoking anecdotes... needs to be read by anyone with a serious interest in research and marketing."

Research Magazine

"Shmueli et al. have done a wonderful job in presenting the field of data mining – a welcome addition to the literature."

ComputingReviews.com

"Excellent choice for business analysts...The book is a perfect fit for its intended audience."

Keith McCormick, Consultant and Author of SPSS Statistics For Dummies, Third Edition and SPSS Statistics for Data Analysis and Visualization

"extremely well organized, clearly written and introduces all of the basic ideas quite well."

Robert L. Phillips, Professor of Professional Practice, Columbia Business School

Data Mining for Business Analytics: Concepts, Techniques, and Applications in Microsoft® Office Excel® with XLMiner®, Third Edition presents an applied approach to data mining and predictive analytics with clear exposition, hands-on exercises, and real-life case studies. Readers will work with all of the standard data mining methods using the Microsoft® Office Excel® add-in XLMiner® to develop predictive models and learn how to obtain business value from Big Data.

- Real-world examples to build a theoretical and practical understanding of key data mining methods
- End-of-chapter exercises that help readers better understand the presented material
- Data-rich case studies to illustrate various applications of data mining techniques
- Completely new chapters on social network analysis and text mining
- A companion site with additional data sets, instructors material that include solutions to exercises and case studies, and Microsoft PowerPoint® slides

Data Mining for Business Analytics: Concepts, Techniques, and Applications in Microsoft® Office Excel® with XLMiner®, Third Edition is an ideal textbook for upper-undergraduate and graduate-level courses as well as professional programs on data mining, predictive modeling, and Big Data analytics. The new edition is also a unique reference for analysts, researchers, and practitioners working with predictive analytics in the fields of business, finance, marketing, computer science, and information technology.

Galit Shmueli, PhD, is Distinguished Professor at National Tsing Hua University’s Institute of Service Science. She has designed and instructed data mining courses since 2004 at University of Maryland, Statistics.com, The Indian School of Business, and National Tsing Hua University, Taiwan. Professor Shmueli is known for her research and teaching in business analytics, with a focus on statistical and data mining methods in information systems and healthcare. She has authored over 70 journal articles, books, textbooks and book chapters.

Peter C. Bruce is President and Founder of the Institute for Statistics Education at ©company website He has written multiple journal articles and is the developer of Resampling Stats software. He is the author of Introductory Statistics and Analytics: A Resampling Perspective, also published by Wiley.
Nitin R. Patel, PhD, is Chairman and cofounder of Cytel, Inc., based in Cambridge, Massachusetts. A Fellow of the American Statistical Association, Dr. Patel has also served as a Visiting Professor at the Massachusetts Institute of Technology and at Harvard University. He is a Fellow of the Computer Society of India and was a professor at the Indian Institute of Management, Ahmedabad for 15 years.

Contents:

Foreword xvii
Preface to the Third Edition xix
Preface to the First Edition xxii
Acknowledgments xxiv

PART I PRELIMINARIES

CHAPTER 1 Introduction 3
1.1 What is Business Analytics? 3
1.2 What is Data Mining? 5
1.3 Data Mining and Related Terms 5
1.4 Big Data 6
1.5 Data Science 7
1.6 Why Are There So Many Different Methods? 8
1.7 Terminology and Notation 9
1.8 Road Maps to This Book 11

Order of Topics 12

CHAPTER 2 Overview of the Data Mining Process 14
2.1 Introduction 14
2.2 Core Ideas in Data Mining 15
2.3 The Steps in Data Mining 18
2.4 Preliminary Steps 20
2.5 Predictive Power and Overfitting 26
2.6 Building a Predictive Model with XLMiner 30
2.7 Using Excel for Data Mining 40
2.8 Automating Data Mining Solutions 40

Data Mining Software Tools (by Herb Edelstein) 42

Problems 45

PART II DATA EXPLORATION AND DIMENSION REDUCTION

CHAPTER 3 Data Visualization 50
3.1 Uses of Data Visualization 50
3.2 Data Examples 52
Example 1: Boston Housing Data 52
Example 2: Ridership on Amtrak Trains 53
3.3 Basic Charts: Bar Charts, Line Graphs, and Scatter Plots 53
Distribution Plots 54
Heatmaps: Visualizing Correlations and Missing Values 57
3.4 Multi-Dimensional Visualization 58
Adding Variables 59
Manipulations 61
Reference: trend line and labels 64
Scaling up to Large Datasets 65
Multivariate Plot 66
Interactive Visualization 67
3.5 Specialized Visualizations 70
Visualizing Networked Data 70
Visualizing Hierarchical Data: Treemaps 72
Visualizing Geographical Data: Map Charts 73
3.6 Summary: Major Visualizations and Operations, by Data Mining Goal 75
Prediction 75
Classification 75
Time Series Forecasting 75
Unsupervised Learning 76
Problems 77
CHAPTER 4 Dimension Reduction 79
4.1 Introduction 79
4.2 Curse of Dimensionality 80
4.3 Practical Considerations 80
Example 1: House Prices in Boston 80
4.4 Data Summaries 81
4.5 Correlation Analysis 84
4.6 Reducing the Number of Categories in Categorical Variables 85
4.7 Converting A Categorical Variable to A Numerical Variable 86
4.8 Principal Components Analysis 86
Example 2: Breakfast Cereals 87
Principal Components 92
Normalizing the Data 93
Using Principal Components for Classification and Prediction 94
4.9 Dimension Reduction Using Regression Models 96
4.10 Dimension Reduction Using Classification and Regression Trees 96
Problems 97

PART III PERFORMANCE EVALUATION
CHAPTER 5 Evaluating Predictive Performance 101
5.1 Introduction 101
5.2 Evaluating Predictive Performance 102
Benchmark: The Average 102
Prediction Accuracy Measures 103
5.3 Judging Classifier Performance 106
Benchmark: The Naive Rule 107
Class Separation 107
The Classification Matrix 107
Using the Validation Data 109
Accuracy Measures 109
Cutoff for Classification 110
Performance in Unequal Importance of Classes 114
Asymmetric Misclassification Costs 116
5.4 Judging Ranking Performance 119
5.5 Oversampling 123
Problems 129

PART IV PREDICTION AND CLASSIFICATION METHODS
CHAPTER 6 Multiple Linear Regression 134
6.1 Introduction 134
6.2 Explanatory vs. Predictive Modeling 135
6.3 Estimating the Regression Equation and Prediction 136
Example: Predicting the Price of Used Toyota Corolla Cars 137

6.4 Variable Selection in Linear Regression 141
Reducing the Number of Predictors 141
How to Reduce the Number of Predictors 142
Problems 147

CHAPTER 7 k-Nearest Neighbors (kNN) 151
7.1 The k-NN Classifier (categorical outcome) 151
Determining Neighbors 151
Classification Rule 152
Example: Riding Mowers 152
Choosing k 154
Setting the Cutoff Value 154
7.2 k-NN for a Numerical Response 156
7.3 Advantages and Shortcomings of k-NN Algorithms 158
Problems 160

CHAPTER 8 The Naive Bayes Classifier 162
8.1 Introduction 162
Example 1: Predicting Fraudulent Financial Reporting 163
8.2 Applying the Full (Exact) Bayesian Classifier 164
8.3 Advantages and Shortcomings of the Naive Bayes Classifier 172
Advantages and Shortcomings of the naive Bayes Classifier 172
Problems 176

CHAPTER 9 Classification and Regression Trees 178
9.1 Introduction 178
9.2 Classification Trees 179
Example 1: Riding Mowers 180
9.3 Measures of Impurity 183
9.4 Evaluating the Performance of a Classification Tree 187
Example 2: Acceptance of Personal Loan 188
9.5 Avoiding Overfitting 192
Stopping Tree Growth: CHAID 192
Pruning the Tree 193

9.6 Classification Rules from Trees 198

9.7 Classification Trees for More Than two Classes 198

9.8 Regression Trees 198

Prediction 199

Measuring Impurity 200

Evaluating Performance 200

9.9 Advantages and Weaknesses of a Tree 200

9.10 Improving Prediction: Multiple Trees 202

Problems 205

CHAPTER 10 Logistic Regression 209

10.1 Introduction 209

10.2 The Logistic Regression Model 211

Example: Acceptance of Personal Loan 212

Model with a Single Predictor 214

Estimating the Logistic Model from Data 215

Interpreting Results in Terms of Odds 218

10.3 Evaluating Classification Performance 219

Variable Selection 220

10.4 Example of Complete Analysis: Predicting Delayed Flights 222

Data Preprocessing 224

Model Fitting and Estimation 224

Model Interpretation 226

Model Performance 226

Variable Selection 227

10.5 Appendix: Logistic Regression for Profiling 231

Appendix A: Why Linear Regression Is Problematic for a Categorical Response 231

Appendix B: Evaluating Explanatory Power 233

Appendix C: Logistic Regression for More Than Two Classes 235

Problems 239

CHAPTER 11 Neural Nets 242

11.1 Introduction 242
Boosting 286

Advantages and Weaknesses of Ensembles 286

13.2 Uplift (Persuasion) Modeling 287

A-B Testing 287

Uplift 288

Gathering the Data 288

A Simple Model 289

Modeling Individual Uplift 290

Using the Results of an Uplift Model 292

13.3 Summary 292

Problems 293

PART V MINING RELATIONSHIPS AMONG RECORDS

CHAPTER 14 Association Rules and Collaborative Filtering 297

14.1 Association Rules 297

Discovering Association Rules in Transaction Databases 298

Example 1: Purchases of Phone Faceplates 298

Generating Candidate Rules 298

The Apriori Algorithm 301

Selecting Strong Rules 301

Data Format 303

The Process of Rule Selection 304

Interpreting the Results 306

Rules and Chance 306

Example 2: Rules for Similar Book Purchases 308

14.2 Collaborative Filtering 310

Data Type and Format 311

Example 3: Netflix Prize Contest 311

User-Based Collaborative Filtering: People Like You 312

Item-Based Collaborative Filtering 315

Advantages and Weaknesses of Collaborative Filtering 316

Collaborative Filtering vs. Association Rules 316

14.3 Summary 318
CHAPTER 15 Cluster Analysis 324

15.1 Introduction 324

Example: Public Utilities 326

15.2 Measuring Distance Between Two Observations 328

Euclidean Distance 328

Normalizing Numerical Measurements 328

Other Distance Measures for Numerical Data 329

Distance Measures for Categorical Data 331

Distance Measures for Mixed Data 331

15.3 Measuring Distance Between Two Clusters 332

15.4 Hierarchical (Agglomerative) Clustering 334

Single Linkage 335

Complete Linkage 335

Average Linkage 336

Centroid Linkage 336

Dendrograms: Displaying Clustering Process and Results 337

Validating Clusters 339

Limitations of Hierarchical Clustering 340

15.5 Non–hierarchical Clustering: The k–Means Algorithm 341

Initial Partition into k Clusters 342

Problems 346

PART VI FORECASTING TIME SERIES

CHAPTER 16 Handling Time Series 351

16.1 Introduction 351

16.2 Descriptive vs. Predictive Modeling 352

16.3 Popular Forecasting Methods in Business 353

Combining Methods 353

16.4 Time Series Components 354

Example: Ridership on Amtrak Trains 354

16.5 Data Partitioning and Performance Evaluation 358
19.2 Directed vs. Undirected Networks 416
19.3 Visualizing and analyzing networks 418
Graph Layout 418
Adjacency List 421
Adjacency Matrix 422
Using Network Data in Classification and Prediction 422
19.4 Social Data Metrics and Taxonomy 423
Node-Level Centrality Metrics 423
Egocentric Network 424
Network Metrics 425
19.5 Using Network Metrics in Prediction and Classification 427
Link Prediction 427
Entity Resolution 427
Collaborative Filtering 428
Advantages and Disadvantages 431
Problems 434
CHAPTER 20 Text Mining 436
20.1 Introduction 436
20.2 The Spreadsheet Representation of Text: Bag-of-Words 437
20.3 Bag-of-Words vs. Meaning Extraction at Document Level 437
20.4 Preprocessing the Text 438
Tokenization 439
Text Reduction 439
Presence/Absence vs. Frequency 440
Term Frequency – Inverse Document Frequency (TF-IDF) 441
From Terms to Concepts: Latent Semantic Indexing 441
Extracting Meaning 441
20.5 Implementing data mining methods 442
20.6 Example: Online Discussions on Autos and Electronics 442
Importing and Labeling the Records 443
Tokenization 444
Fax Order Form
To place an order via fax simply print this form, fill in the information below and fax the completed form to 646-607-1907 (from USA) or +353-1-481-1716 (from Rest of World). If you have any questions please visit http://www.researchandmarkets.com/contact/

Order Information
Please verify that the product information is correct.

Product Name: Data Mining for Business Analytics. Concepts, Techniques, and Applications with XLMiner. 3rd Edition
Web Address: http://www.researchandmarkets.com/reports/3609933/
Office Code: SCPLZAIN

Product Format
Please select the product format and quantity you require:

<table>
<thead>
<tr>
<th>Quantity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard Copy (Hard Back):</td>
<td>USD 123 + USD 28 Shipping/Handling</td>
</tr>
</tbody>
</table>

* Shipping/Handling is only charged once per order.

Contact Information
Please enter all the information below in BLOCK CAPITALS

Title: Mr ☐ Mrs ☐ Dr ☐ Miss ☐ Ms ☐ Prof ☐
First Name: 
Last Name: 
Email Address: *
Job Title: 
Organisation: 
Address: 
City: 
Postal / Zip Code: 
Country: 
Phone Number: 
Fax Number: 

* Please refrain from using free email accounts when ordering (e.g. Yahoo, Hotmail, AOL)
Payment Information

Please indicate the payment method you would like to use by selecting the appropriate box.

☐ Pay by credit card: You will receive an email with a link to a secure webpage to enter your credit card details.

☐ Pay by check: Please post the check, accompanied by this form, to:
Research and Markets,
Guinness Center,
Taylors Lane,
Dublin 8,
Ireland.

☐ Pay by wire transfer: Please transfer funds to:

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account number</td>
<td>833 130 83</td>
</tr>
<tr>
<td>Sort code</td>
<td>98-53-30</td>
</tr>
<tr>
<td>Swift code</td>
<td>ULSBIE2D</td>
</tr>
<tr>
<td>IBAN number</td>
<td>IE78ULSB98533083313083</td>
</tr>
<tr>
<td>Bank Address</td>
<td>Ulster Bank, 27-35 Main Street, Blackrock, Co. Dublin, Ireland.</td>
</tr>
</tbody>
</table>

If you have a Marketing Code please enter it below:

Marketing Code: __________________________________________

Please note that by ordering from Research and Markets you are agreeing to our Terms and Conditions at http://www.researchandmarkets.com/info/terms.asp

Please fax this form to:

(646) 607-1907 or (646) 964-6609 - From USA
+353-1-481-1716 or +353-1-653-1571 - From Rest of World