Discovering Knowledge in Data. An Introduction to Data Mining. 2nd Edition.
Wiley Series on Methods and Applications in Data Mining

Description: The field of data mining lies at the confluence of predictive analytics, statistical analysis, and business intelligence. Due to the ever-increasing complexity and size of data sets and the wide range of applications in computer science, business, and health care, the process of discovering knowledge in data is more relevant than ever before.

This book provides the tools needed to thrive in today’s big data world. The author demonstrates how to leverage a company’s existing databases to increase profits and market share, and carefully explains the most current data science methods and techniques. The reader will learn data mining by doing data mining. By adding chapters on data modelling preparation, imputation of missing data, and multivariate statistical analysis, Discovering Knowledge in Data, Second Edition remains the eminent reference on data mining.

Contents:

PREFACE xi

CHAPTER 1 AN INTRODUCTION TO DATA MINING 1
1.1 What is Data Mining? 1
1.2 Wanted: Data Miners 2
1.3 The Need for Human Direction of Data Mining 3
1.4 The Cross-Industry Standard Practice for Data Mining 4
1.4.1 Crisp-DM: The Six Phases 5
1.5 Fallacies of Data Mining 6
1.6 What Tasks Can Data Mining Accomplish? 8
1.6.1 Description 8
1.6.2 Estimation 8
1.6.3 Prediction 10
1.6.4 Classification 10
1.6.5 Clustering 12
1.6.6 Association 14
References 14
Exercises 15

CHAPTER 2 DATA PREPROCESSING 16
2.1 Why do We Need to Preprocess the Data? 17
2.2 Data Cleaning 17
7.9 Application of k–Nearest Neighbor Algorithm Using IBM/SPSS Modeler 160
The R Zone 162
Exercises 163
Hands–On Analysis 164

CHAPTER 8 DECISION TREES 165
8.1 What is a Decision Tree? 165
8.2 Requirements for Using Decision Trees 167
8.3 Classification and Regression Trees 168
8.4 C4.5 Algorithm 174
8.5 Decision Rules 179
8.6 Comparison of the C5.0 and Cart Algorithms Applied to Real Data 180
The R Zone 183
References 184
Exercises 185
Hands–On Analysis 185

CHAPTER 9 NEURAL NETWORKS 187
9.1 Input and Output Encoding 188
9.2 Neural Networks for Estimation and Prediction 190
9.3 Simple Example of a Neural Network 191
9.4 Sigmoid Activation Function 193
9.5 Back–Propagation 194
9.5.1 Gradient Descent Method 194
9.5.2 Back–Propagation Rules 195
9.5.3 Example of Back–Propagation 196
9.6 Termination Criteria 198
9.7 Learning Rate 198
9.8 Momentum Term 199
9.9 Sensitivity Analysis 201
9.10 Application of Neural Network Modeling 202
The R Zone 204
References 207
Exercises 207

Hands-On Analysis 207

CHAPTER 10 HIERARCHICAL AND k-MEANS CLUSTERING 209

10.1 The Clustering Task 209

10.2 Hierarchical Clustering Methods 212

10.3 Single-Linkage Clustering 213

10.4 Complete-Linkage Clustering 214

10.5 k-Means Clustering 215

10.6 Example of k-Means Clustering at Work 216

10.7 Behavior of MSB, MSE, and PSEUDO-F as the k-Means Algorithm Proceeds 219

10.8 Application of k-Means Clustering Using SAS Enterprise Miner 220

10.9 Using Cluster Membership to Predict Churn 223

The R Zone 224

References 226

Exercises 226

Hands-On Analysis 226

CHAPTER 11 KOHONEN NETWORKS 228

11.1 Self-Organizing Maps 228

11.2 Kohonen Networks 230

11.2.1 Kohonen Networks Algorithm 231

11.3 Example of a Kohonen Network Study 231

11.4 Cluster Validity 235

11.5 Application of Clustering Using Kohonen Networks 235

11.6 Interpreting the Clusters 237

11.6.1 Cluster Profiles 240

11.7 Using Cluster Membership as Input to Downstream Data Mining Models 242

The R Zone 243

References 245

Exercises 245

Hands-On Analysis 245

CHAPTER 12 ASSOCIATION RULES 247

12.1 Affinity Analysis and Market Basket Analysis 247
12.1.1 Data Representation for Market Basket Analysis 248
12.2 Support, Confidence, Frequent Itemsets, and the a Priori Property 249
12.3 How Does the a Priori Algorithm Work? 251
12.3.1 Generating Frequent Itemsets 251
12.3.2 Generating Association Rules 253
12.4 Extension from Flag Data to General Categorical Data 255
12.5 Information–Theoretic Approach: Generalized Rule Induction Method 256
12.5.1 J–Measure 257
12.6 Association Rules are Easy to do Badly 258
12.7 How Can We Measure the Usefulness of Association Rules? 259
12.8 Do Association Rules Represent Supervised or Unsupervised Learning? 260
12.9 Local Patterns Versus Global Models 261

The R Zone 262
References 263
Exercises 263

Hands–On Analysis 264

CHAPTER 13 IMPUTATION OF MISSING DATA 266
13.1 Need for Imputation of Missing Data 266
13.2 Imputation of Missing Data: Continuous Variables 267
13.3 Standard Error of the Imputation 270
13.4 Imputation of Missing Data: Categorical Variables 271
13.5 Handling Patterns in Missingness 272

The R Zone 273
Reference 276
Exercises 276

Hands–On Analysis 276

CHAPTER 14 MODEL EVALUATION TECHNIQUES 277
14.1 Model Evaluation Techniques for the Description Task 278
14.2 Model Evaluation Techniques for the Estimation and Prediction Tasks 278
14.3 Model Evaluation Techniques for the Classification Task 280
14.4 Error Rate, False Positives, and False Negatives 280
14.5 Sensitivity and Specificity 283
14.6 Misclassification Cost Adjustment to Reflect Real–World Concerns 284
14.7 Decision Cost/Benefit Analysis 285
14.8 Lift Charts and Gains Charts 286
14.9 Interweaving Model Evaluation with Model Building 289
14.10 Confluence of Results: Applying a Suite of Models 290

The R Zone 291
Reference 291
Exercises 291
Hands–On Analysis 291

APPENDIX: DATA SUMMARIZATION AND VISUALIZATION 294
INDEX 309
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: Discovering Knowledge in Data. An Introduction to Data Mining. 2nd Edition. Wiley Series on Methods and Applications in Data Mining
Web Address: http://www.researchandmarkets.com/reports/2174342/
Office Code: SCPL621U

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

| Quantity | Hard Copy (Hard Back): | USD 98 + USD 28 Shipping/Handling |

* 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:

Account number 833 130 83
Sort code 98-53-30
Swift code ULSBIE2D
IBAN number IE78ULSB98533083313083
Bank Address Ulster Bank,
27-35 Main Street,
Blackrock,
Co. Dublin,
Ireland.

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