Data Science and Big Data Analytics. Discovering, Analyzing, Visualizing and Presenting Data

Description:

Data Science and Big Data Analytics

Discovering, Analyzing, Visualizing and Presenting Data

Data Science and Big Data Analytics is about harnessing the power of data for new insights. The book covers the breadth of activities, methods and tools that Data Scientists use. The content focuses on concepts, principles and practical applications that are relevant to any industry and technology environment, and the learning is supported and explained with illustrative examples using open-source software.

This book will help you:

- Become a contributor on a data science team
- Deploy a structured lifecycle approach to data analytics problems
- Apply appropriate analytic techniques and tools to analyze big data
- Learn how to tell a compelling story with data to drive business action
- Prepare for EMC ProvenTM Professional Data Scientist certification

EMC Proven Professional is a leading education and certification program in the IT industry, providing comprehensive coverage of information storage technologies, virtualization, cloud computing, data science/big data analytics, and more

Being Proven means investing in yourself and formally validating your expertise!

This book prepares you for the Data Science Associate exam E20–007 leading to EMC Proven Professional Data Science Associate (EMCDSA) certification.

Visit [here](http://www.researchandmarkets.com/reports/2741514/) for details.

Contents:

Introduction xvii

Chapter 1 Introduction to Big Data Analytics 1

1.1 Big Data Overview 2

1.1.1 Data Structures 5

1.1.2 Analyst Perspective on Data Repositories 9

1.2 State of the Practice in Analytics 11

1.2.1 BI Versus Data Science 12

1.2.2 Current Analytical Architecture 13

1.2.3 Drivers of Big Data 15

1.2.4 Emerging Big Data Ecosystem and a New Approach to Analytics 16

1.3 Key Roles for the New Big Data Ecosystem 19

1.4 Examples of Big Data Analytics 22
2.8.1 Phase 1: Discovery  54
2.8.2 Phase 2: Data Preparation  55
2.8.3 Phase 3: Model Planning 56
2.8.4 Phase 4: Model Building 56
2.8.5 Phase 5: Communicate Results  58
2.8.6 Phase 6: Operationalize 59
Summary 60
Exercises  61
Bibliography 61

Chapter 3     Review of Basic Data Analytic Methods Using R 63
3.1 Introduction to R 64
3.1.1 R Graphical User Interfaces  67
3.1.2 Data Import and Export 69
3.1.3 Attribute and Data Types 71
3.1.4 Descriptive Statistics  79
3.2 Exploratory Data Analysis  80
3.2.1 Visualization Before Analysis 82
3.2.2 Dirty Data 85
3.2.3 Visualizing a Single Variable  88
3.2.4 Examining Multiple Variables 91
3.2.5 Data Exploration Versus Presentation  99
3.3 Statistical Methods for Evaluation 101
3.3.1 Hypothesis Testing 102
3.3.2 Difference of Means 104
3.3.3 Wilcoxon Rank-Sum Test 108
3.3.4 Type I and Type II Errors  109
3.3.5 Power and Sample Size 110
3.3.6 ANOVA 110
Summary 114
Exercises  114
Bibliography115

Chapter 4     Advanced Analytical Theory and Methods: Clustering 117
6.2.1 Use Cases 179
6.2.2 Model Description 179
6.2.3 Diagnostics 181
6.3 Reasons to Choose and Cautions 188
6.4 Additional Regression Models 189
Summary 190
Exercises 190

Chapter 7  Advanced Analytical Theory and Methods: Classification 191
7.1 Decision Trees 192
7.1.1 Overview of a Decision Tree 193
7.1.2 The General Algorithm 197
7.1.3 Decision Tree Algorithms 203
7.1.4 Evaluating a Decision Tree 204
7.1.5 Decision Trees in R 206
7.2 Naive Bayes 211
7.2.1 Bayes Theorem 212
7.2.2 Naive Bayes Classifier 214
7.2.3 Smoothing 217
7.2.4 Diagnostics 217
7.2.5 Naive Bayes in R 218
7.3 Diagnostics of Classifiers 224
7.4 Additional Classification Methods 228
Summary 229
Exercises 230

Bibliography 231

Chapter 8  Advanced Analytical Theory and Methods: Time Series Analysis 233
8.1 Overview of Time Series Analysis 234
8.1.1 Box-Jenkins Methodology 235
8.2 ARIMA Model 236
8.2.1 Autocorrelation Function (ACF) 236
8.2.2 Autoregressive Models 238
8.2.3 Moving Average Models 239
8.2.4 ARMA and ARIMA Models 241
8.2.5 Building and Evaluating an ARIMA Model 244
8.2.6 Reasons to Choose and Cautions 252
8.3 Additional Methods 253
Summary 254
Exercises 254

Chapter 9  Advanced Analytical Theory and Methods: Text Analysis 255
9.1 Text Analysis Steps 257
9.2 A Text Analysis Example 259
9.3 Collecting Raw Text 260
9.4 Representing Text 264
9.5 Term Frequency Inverse Document Frequency (TFIDF) 269
9.6 Categorizing Documents by Topics 274
9.7 Determining Sentiments 277
9.8 Gaining Insights 283
Summary 290
Exercises 290

Bibliography 291

Chapter 10  Advanced Analytics Technology and Tools: MapReduce and Hadoop 295
10.1 Analytics for Unstructured Data 296
10.1.1 Use Cases 296
10.1.2 MapReduce 298
10.1.3 Apache Hadoop 300
10.2 The Hadoop Ecosystem 306
10.2.1 Pig 306
10.2.2 Hive 308
10.2.3 HBase 311
10.2.4 Mahout 319
10.3 NoSQL 322
Summary 323
Exercises 324
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 Science and Big Data Analytics. Discovering, Analyzing, Visualizing and Presenting Data
Web Address: http://www.researchandmarkets.com/reports/2741514/
Office Code: SCBRW5P8

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

Quantity
Hard Copy (Hard Back): ☐ USD 99 + USD 29 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