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

Description:
Data Science and Big Data Analytics

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 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
Chapter 2  Data Analytics Lifecycle 25

2.1 Data Analytics Lifecycle Overview 26

2.1.1 Key Roles for a Successful Analytics Project 26

2.1.2 Background and Overview of Data Analytics Lifecycle 28

2.2 Phase 1: Discovery 30

2.2.1 Learning the Business Domain 30

2.2.2 Resources 31

2.2.3 Framing the Problem 32

2.2.4 Identifying Key Stakeholders 33

2.2.5 Interviewing the Analytics Sponsor 33

2.2.6 Developing Initial Hypotheses 35

2.2.7 Identifying Potential Data Sources 35

2.3 Phase 2: Data Preparation 36

2.3.1 Preparing the Analytic Sandbox 37

2.3.2 Performing ETLT 38

2.3.3 Learning About the Data 39

2.3.4 Data Conditioning 40

2.3.5 Survey and Visualize 41

2.3.6 Common Tools for the Data Preparation Phase 42

2.4 Phase 3: Model Planning 42

2.4.1 Data Exploration and Variable Selection 44

2.4.2 Model Selection 45

2.4.3 Common Tools for the Model Planning Phase 45

2.5 Phase 4: Model Building 46

2.5.1 Common Tools for the Model Building Phase 48

2.6 Phase 5: Communicate Results 49

2.7 Phase 6: Operationalize 50

2.8 Case Study: Global Innovation Network and Analysis (GINA) 53
4.1 Overview of Clustering 118
4.2 K–means 118
4.2.1 Use Cases 119
4.2.2 Overview of the Method 120
4.2.3 Determining the Number of Clusters 123
4.2.4 Diagnostics 128
4.2.5 Reasons to Choose and Cautions 130
4.3 Additional Algorithms 134
Summary 135
Exercises 135
Bibliography 136

Chapter 5  Advanced Analytical Theory and Methods: Association Rules 137
5.1 Overview 138
5.2 Apriori Algorithm 140
5.3 Evaluation of Candidate Rules 141
5.4 Applications of Association Rules 143
5.5 An Example: Transactions in a Grocery Store 143
5.5.1 The Groceries Dataset 144
5.5.2 Frequent Itemset Generation 146
5.5.3 Rule Generation and Visualization 152
5.6 Validation and Testing 157
5.7 Diagnostics 158
Summary 158
Exercises 159
Bibliography 160

Chapter 6  Advanced Analytical Theory and Methods: Regression 161
6.1 Linear Regression 162
6.1.1 Use Cases 162
6.1.2 Model Description 163
6.1.3 Diagnostics 173
6.2 Logistic Regression 178
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.

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Data Science and Big Data Analytics. Discovering, Analyzing, Visualizing and Presenting Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office Code</td>
<td>SCH35R7G</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Hard Copy (Hard Back)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>USD 99 + USD 29 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**

<table>
<thead>
<tr>
<th>Title:</th>
<th>Mr ☐ Mrs ☐ Dr ☐ Miss ☐ Ms ☐ Prof ☐</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Name:</td>
<td></td>
</tr>
<tr>
<td>Email Address: *</td>
<td></td>
</tr>
<tr>
<td>Job Title:</td>
<td></td>
</tr>
<tr>
<td>Organisation:</td>
<td></td>
</tr>
<tr>
<td>Address:</td>
<td></td>
</tr>
<tr>
<td>City:</td>
<td></td>
</tr>
<tr>
<td>Postal / Zip Code:</td>
<td></td>
</tr>
<tr>
<td>Country:</td>
<td></td>
</tr>
<tr>
<td>Phone Number:</td>
<td></td>
</tr>
<tr>
<td>Fax Number:</td>
<td></td>
</tr>
</tbody>
</table>

* 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