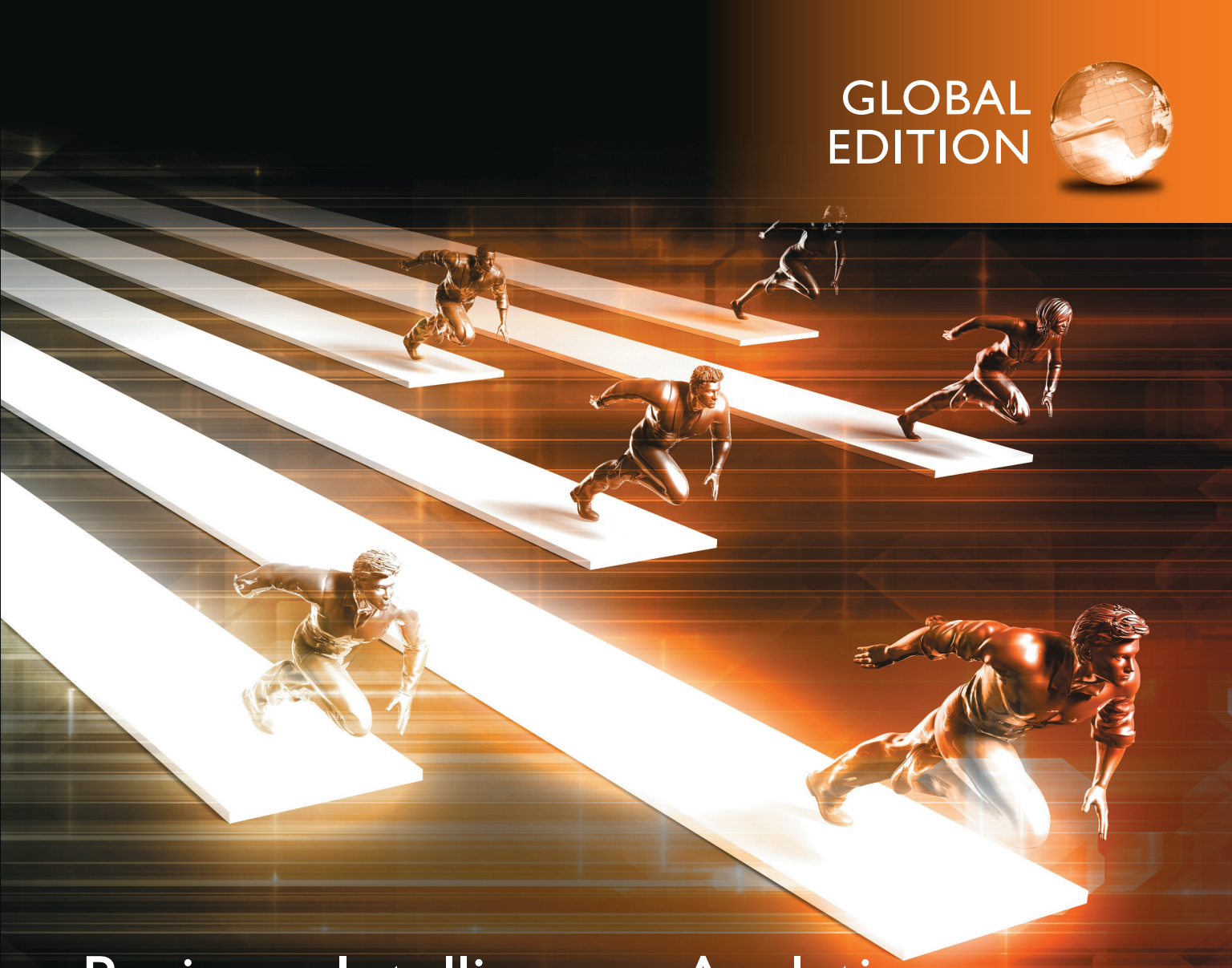
GLOBAL EDITION



Business Intelligence, Analytics, and Data Science

A Managerial Perspective

FOURTH EDITION

Ramesh Sharda • Dursun Delen • Efraim Turban

Pearson

FOURTH EDITION
GLOBAL EDITION

BUSINESS INTELLIGENCE, ANALYTICS, AND DATA SCIENCE:

A Managerial Perspective

Ramesh Sharda
Oklahoma State University

Dursun Delen
Oklahoma State University

Efraim Turban
University of Hawaii

With contributions to previous editions by

J. E. Aronson
The University of Georgia

Ting-Peng Liang
National Sun Yat-sen University

David King
JDA Software Group, Inc.

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

VP Editorial Director: Andrew Gilfillan
Senior Portfolio Manager: Samantha Lewis
Content Development Team Lead: Laura Burgess
Content Developer: Stephany Harrington
Program Monitor: Ann Pulido/SPi Global
Editorial Assistant: Madeline Houpt
Project Manager, Global Edition: Sudipto Roy
Acquisitions Editor, Global Edition: Tahnee Wager
Senior Project Editor, Global Edition: Daniel Luiz
Managing Editor, Global Edition: Steven Jackson
Senior Manufacturing Controller, Production, Global Edition: Trudy Kimber
Product Marketing Manager: Kaylee Carlson
Project Manager: Revathi Viswanathan/Cenveo Publisher Services
Text Designer: Cenveo® Publisher Services
Cover Designer: Lumina Datamatics, Inc.
Cover Art: kentoh/Shutterstock
Full-Service Project Management: Cenveo Publisher Services
Composition: Cenveo Publisher Services

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page within text.

Microsoft and/or its respective suppliers make no representations about the suitability of the information contained in the documents and related graphics published as part of the services for any purpose. All such documents and related graphics are provided as is without warranty of any kind. Microsoft and/or its respective suppliers hereby disclaim all warranties and conditions with regard to this information, including all warranties and conditions of merchantability, whether express, implied or statutory, fitness for a particular purpose, title and non-infringement.

In no event shall Microsoft and/or its respective suppliers be liable for any special, indirect or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortious action, arising out of or in connection with the use or performance of information available from the services. The documents and related graphics contained herein could include technical inaccuracies or typographical errors. Changes are periodically added to the information herein. Microsoft and/or its respective suppliers may make improvements and/or changes in the product(s) and/or the program(s) described herein at any time. Partial screen shots may be viewed in full within the software version specified.

Microsoft® Windows®, and Microsoft Office® are registered trademarks of the Microsoft Corporation in the U.S.A. and other countries. This book is not sponsored or endorsed by or affiliated with the Microsoft Corporation.

Pearson Education Limited
KAO Two
KAO Park
Harlow
CM17 9NA
United Kingdom

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsonglobaleditions.com

© Pearson Education Limited 2018

The rights of Ramesh Sharda, Dursun Delen, and Efraim Turban to be identified as the authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Authorized adaptation from the United States edition, entitled Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th edition, ISBN 978-0-13-463328-2, by Ramesh Sharda, Dursun Delen, and Efraim Turban, published by Pearson Education © 2018.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a license permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

ISBN 10: 1-292-22054-6
ISBN 13: 978-1-292-22054-3

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

10 9 8 7 6 5 4 3 2 1
14 13 12 11 10

Typeset in ITC Garamond Std-Lt by Cenveo Publisher Services.

Printed and bound by Vivar, Malaysia.

Brief Contents

Preface 19
About the Authors 25

Chapter 1 An Overview of Business Intelligence, Analytics, and Data Science 29
Chapter 2 Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization 79
Chapter 3 Descriptive Analytics II: Business Intelligence and Data Warehousing 153
Chapter 4 Predictive Analytics I: Data Mining Process, Methods, and Algorithms 215
Chapter 5 Predictive Analytics II: Text, Web, and Social Media Analytics 273
Chapter 6 Prescriptive Analytics: Optimization and Simulation 345
Chapter 7 Big Data Concepts and Tools 395
Chapter 8 Future Trends, Privacy and Managerial Considerations in Analytics 443

Glossary 493
Index 501

This page intentionally left blank

Contents

Preface 19
About the Authors 25

Chapter 1 **An Overview of Business Intelligence, Analytics, and Data Science** 29

1.1 OPENING VIGNETTE: Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics 30
1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics 37
1.3 Evolution of Computerized Decision Support to Analytics/Data Science 39
1.4 A Framework for Business Intelligence 41
Definitions of BI 42
A Brief History of BI 42
The Architecture of BI 42
The Origins and Drivers of BI 42
▸ **APPLICATION CASE 1.1 Sabre Helps Its Clients Through Dashboards and Analytics** 44
A Multimedia Exercise in Business Intelligence 45
Transaction Processing versus Analytic Processing 45
Appropriate Planning and Alignment with the Business Strategy 46
Real-Time, On-Demand BI Is Attainable 47
Developing or Acquiring BI Systems 47
Justification and Cost–Benefit Analysis 48
Security and Protection of Privacy 48
Integration of Systems and Applications 48
1.5 Analytics Overview 48
Descriptive Analytics 50
▸ **APPLICATION CASE 1.2 Silvaris Increases Business with Visual Analysis and Real-Time Reporting Capabilities** 50
▸ **APPLICATION CASE 1.3 Siemens Reduces Cost with the Use of Data Visualization** 51
Predictive Analytics 51
▸ **APPLICATION CASE 1.4 Analyzing Athletic Injuries** 52
Prescriptive Analytics 52
Analytics Applied to Different Domains 53
▸ **APPLICATION CASE 1.5 A Specialty Steel Bar Company Uses Analytics to Determine Available-to-Promise Dates** 53
Analytics or Data Science? 54

1.6 Analytics Examples in Selected Domains 55
Analytics Applications in Healthcare—Humana Examples 55
Analytics in the Retail Value Chain 59
1.7 A Brief Introduction to Big Data Analytics 61
What Is Big Data? 61
▸ **APPLICATION CASE 1.6 CenterPoint Energy Uses Real-Time Big Data Analytics to Improve Customer Service** 63
1.8 An Overview of the Analytics Ecosystem 63
Data Generation Infrastructure Providers 65
Data Management Infrastructure Providers 65
Data Warehouse Providers 66
Middleware Providers 66
Data Service Providers 66
Analytics-Focused Software Developers 67
Application Developers: Industry Specific or General 68
Analytics Industry Analysts and Influencers 69
Academic Institutions and Certification Agencies 70
Regulators and Policy Makers 71
Analytics User Organizations 71
1.9 Plan of the Book 72
1.10 Resources, Links, and the Teradata University Network Connection 73
Resources and Links 73
Vendors, Products, and Demos 74
Periodicals 74
The Teradata University Network Connection 74
The Book's Web Site 74
Chapter Highlights 75
Key Terms 75
Questions for Discussion 75
Exercises 76
References 77

Chapter 2 Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization 79

2.1 OPENING VIGNETTE: SiriusXM Attracts and Engages a New Generation of Radio Consumers with Data-Driven Marketing 80
2.2 The Nature of Data 83
2.3 A Simple Taxonomy of Data 87
▸ **APPLICATION CASE 2.1 Medical Device Company Ensures Product Quality While Saving Money** 89

2.4 The Art and Science of Data Preprocessing 91
▸ **APPLICATION CASE 2.2 Improving Student Retention with Data-Driven Analytics** 94
2.5 Statistical Modeling for Business Analytics 100
Descriptive Statistics for Descriptive Analytics 101
Measures of Centrality Tendency (May Also Be Called Measures of Location or Centrality) 102
Arithmetic Mean 102
Median 103
Mode 103
Measures of Dispersion (May Also Be Called Measures of Spread Decentrality) 103
Range 104
Variance 104
Standard Deviation 104
Mean Absolute Deviation 104
Quartiles and Interquartile Range 104
Box-and-Whiskers Plot 105
The Shape of a Distribution 106
▸ **APPLICATION CASE 2.3 Town of Cary Uses Analytics to Analyze Data from Sensors, Assess Demand, and Detect Problems** 110
2.6 Regression Modeling for Inferential Statistics 112
How Do We Develop the Linear Regression Model? 113
How Do We Know If the Model Is Good Enough? 114
What Are the Most Important Assumptions in Linear Regression? 115
Logistic Regression 116
▸ **APPLICATION CASE 2.4 Predicting NCAA Bowl Game Outcomes** 117
Time Series Forecasting 122
2.7 Business Reporting 124
▸ **APPLICATION CASE 2.5 Flood of Paper Ends at FEMA** 126
2.8 Data Visualization 127
A Brief History of Data Visualization 127
▸ **APPLICATION CASE 2.6 Macfarlan Smith Improves Operational Performance Insight with Tableau Online** 129
2.9 Different Types of Charts and Graphs 132
Basic Charts and Graphs 132
Specialized Charts and Graphs 133
Which Chart or Graph Should You Use? 134
2.10 The Emergence of Visual Analytics 136
Visual Analytics 138
High-Powered Visual Analytics Environments 138

2.11 Information Dashboards 143
▸ APPLICATION CASE 2.7 Dallas Cowboys Score Big with Tableau and Teknion 144
Dashboard Design 145
▸ APPLICATION CASE 2.8 Visual Analytics Helps Energy Supplier Make Better Connections 145
What to Look for in a Dashboard 147
Best Practices in Dashboard Design 147
Benchmark Key Performance Indicators with Industry Standards 147
Wrap the Dashboard Metrics with Contextual Metadata 147
Validate the Dashboard Design by a Usability Specialist 148
Prioritize and Rank Alerts/Exceptions Streamed to the Dashboard 148
Enrich the Dashboard with Business-User Comments 148
Present Information in Three Different Levels 148
Pick the Right Visual Construct Using Dashboard Design Principles 148
Provide for Guided Analytics 148
Chapter Highlights 149
Key Terms 149
Questions for Discussion 150
Exercises 150
References 152

Chapter 3 **Descriptive Analytics II: Business Intelligence and Data Warehousing 153**

3.1 OPENING VIGNETTE: Targeting Tax Fraud with Business Intelligence and Data Warehousing 154
3.2 Business Intelligence and Data Warehousing 156
What Is a Data Warehouse? 157
A Historical Perspective to Data Warehousing 158
Characteristics of Data Warehousing 159
Data Marts 160
Operational Data Stores 161
Enterprise Data Warehouses (EDW) 161
Metadata 161
▸ APPLICATION CASE 3.1 A Better Data Plan: Well-Established TELCOs Leverage Data Warehousing and Analytics to Stay on Top in a Competitive Industry 161
3.3 Data Warehousing Process 163
3.4 Data Warehousing Architectures 165
Alternative Data Warehousing Architectures 168
Which Architecture Is the Best? 170

3.5 Data Integration and the Extraction, Transformation, and Load (ETL) Processes 171

Data Integration 172

▸ **APPLICATION CASE 3.2 BP Lubricants Achieves BIGS Success** 172

Extraction, Transformation, and Load 174

3.6 Data Warehouse Development 176

▸ **APPLICATION CASE 3.3 Use of Teradata Analytics for SAP Solutions Accelerates Big Data Delivery** 177

Data Warehouse Development Approaches 179

Additional Data Warehouse Development Considerations 182

Representation of Data in Data Warehouse 182

Analysis of Data in Data Warehouse 184

OLAP versus OLTP 184

OLAP Operations 185

3.7 Data Warehousing Implementation Issues 186

Massive Data Warehouses and Scalability 188

▸ **APPLICATION CASE 3.4 EDW Helps Connect State Agencies in Michigan** 189

3.8 Data Warehouse Administration, Security Issues, and Future Trends 190

The Future of Data Warehousing 191

3.9 Business Performance Management 196

Closed-Loop BPM Cycle 197

▸ **APPLICATION CASE 3.5 AARP Transforms Its BI Infrastructure and Achieves a 347% ROI in Three Years** 199

3.10 Performance Measurement 201

Key Performance Indicator (KPI) 201

Performance Measurement System 202

3.11 Balanced Scorecards 203

The Four Perspectives 203

The Meaning of Balance in BSC 205

3.12 Six Sigma as a Performance Measurement System 205

The DMAIC Performance Model 206

Balanced Scorecard versus Six Sigma 206

Effective Performance Measurement 207

▸ **APPLICATION CASE 3.6 Expedia.com's Customer Satisfaction Scorecard** 208

Chapter Highlights 209

Key Terms 210

Questions for Discussion 210

Exercises 211

References 213

Chapter 4 **Predictive Analytics I: Data Mining Process, Methods, and Algorithms** 215

4.1 OPENING VIGNETTE: Miami-Dade Police Department Is Using Predictive Analytics to Foresee and Fight Crime 216

4.2 Data Mining Concepts and Applications 219

▸ **APPLICATION CASE 4.1 Visa Is Enhancing the Customer Experience While Reducing Fraud with Predictive Analytics and Data Mining** 220

Definitions, Characteristics, and Benefits 222

How Data Mining Works 223

▸ **APPLICATION CASE 4.2 Dell Is Staying Agile and Effective with Analytics in the 21st Century** 224

Data Mining versus Statistics 229

4.3 Data Mining Applications 229

▸ **APPLICATION CASE 4.3 Bank Speeds Time to Market with Advanced Analytics** 231

4.4 Data Mining Process 232

Step 1: Business Understanding 233

Step 2: Data Understanding 234

Step 3: Data Preparation 234

Step 4: Model Building 235

▸ **APPLICATION CASE 4.4 Data Mining Helps in Cancer Research** 235

Step 5: Testing and Evaluation 238

Step 6: Deployment 238

Other Data Mining Standardized Processes and Methodologies 238

4.5 Data Mining Methods 241

Classification 241

Estimating the True Accuracy of Classification Models 242

▸ **APPLICATION CASE 4.5 Influence Health Uses Advanced Predictive Analytics to Focus on the Factors That Really Influence People's Healthcare Decisions** 249

Cluster Analysis for Data Mining 251

Association Rule Mining 253

4.6 Data Mining Software Tools 257

▸ **APPLICATION CASE 4.6 Data Mining Goes to Hollywood: Predicting Financial Success of Movies** 259

4.7 Data Mining Privacy Issues, Myths, and Blunders 263

▸ **APPLICATION CASE 4.7 Predicting Customer Buying Patterns—The Target Story** 264

Data Mining Myths and Blunders 264

Chapter Highlights 267

Key Terms 268

Questions for Discussion 268

Exercises 269

References 271

Chapter 5 **Predictive Analytics II: Text, Web, and Social Media Analytics** 273

5.1 OPENING VIGNETTE: Machine versus Men on Jeopardy!: The Story of Watson 274

5.2 Text Analytics and Text Mining Overview 277

▸ **APPLICATION CASE 5.1 Insurance Group Strengthens Risk Management with Text Mining Solution** 280

5.3 Natural Language Processing (NLP) 281

▸ **APPLICATION CASE 5.2 AMC Networks Is Using Analytics to Capture New Viewers, Predict Ratings, and Add Value for Advertisers in a Multichannel World** 283

5.4 Text Mining Applications 287

Marketing Applications 287

Security Applications 287

▸ **APPLICATION CASE 5.3 Mining for Lies** 288

Biomedical Applications 290

Academic Applications 292

▸ **APPLICATION CASE 5.4 Bringing the Customer into the Quality Equation: Lenovo Uses Analytics to Rethink Its Redesign** 292

5.5 Text Mining Process 294

Task 1: Establish the Corpus 295

Task 2: Create the Term–Document Matrix 295

Task 3: Extract the Knowledge 297

▸ **APPLICATION CASE 5.5 Research Literature Survey with Text Mining** 299

5.6 Sentiment Analysis 302

▸ **APPLICATION CASE 5.6 Creating a Unique Digital Experience to Capture the Moments That Matter at Wimbledon** 303

Sentiment Analysis Applications 306

Sentiment Analysis Process 308

Methods for Polarity Identification 310

Using a Lexicon 310

Using a Collection of Training Documents 311

Identifying Semantic Orientation of Sentences and Phrases 312

Identifying Semantic Orientation of Documents 312

5.7 Web Mining Overview 313

Web Content and Web Structure Mining 315

5.8 Search Engines 317

Anatomy of a Search Engine 318

1. Development Cycle 318

2. Response Cycle 320
Search Engine Optimization 320
Methods for Search Engine Optimization 321
▸ **APPLICATION CASE 5.7 Understanding Why Customers Abandon Shopping Carts Results in a $10 Million Sales Increase** 323
5.9 Web Usage Mining (Web Analytics) 324
Web Analytics Technologies 325
Web Analytics Metrics 326
Web Site Usability 326
Traffic Sources 327
Visitor Profiles 328
Conversion Statistics 328
5.10 Social Analytics 330
Social Network Analysis 330
Social Network Analysis Metrics 331
▸ **APPLICATION CASE 5.8 Tito's Vodka Establishes Brand Loyalty with an Authentic Social Strategy** 331
Connections 334
Distributions 334
Segmentation 335
Social Media Analytics 335
How Do People Use Social Media? 336
Measuring the Social Media Impact 337
Best Practices in Social Media Analytics 337
Chapter Highlights 339
Key Terms 340
Questions for Discussion 341
Exercises 341
References 342

Chapter 6 Prescriptive Analytics: Optimization and Simulation 345

6.1 OPENING VIGNETTE: School District of Philadelphia Uses Prescriptive Analytics to Find Optimal Solution for Awarding Bus Route Contracts 346
6.2 Model-Based Decision Making 348
Prescriptive Analytics Model Examples 348
▸ **APPLICATION CASE 6.1 Optimal Transport for ExxonMobil Downstream through a DSS** 349

Identification of the Problem and Environmental Analysis 350
Model Categories 350
▸ **APPLICATION CASE 6.2 Ingram Micro Uses Business Intelligence Applications to Make Pricing Decisions** 351
6.3 Structure of Mathematical Models for Decision Support 354
The Components of Decision Support Mathematical Models 354
The Structure of Mathematical Models 355
6.4 Certainty, Uncertainty, and Risk 356
Decision Making under Certainty 356
Decision Making under Uncertainty 357
Decision Making under Risk (Risk Analysis) 357
6.5 Decision Modeling with Spreadsheets 357
▸ **APPLICATION CASE 6.3 Primary Schools in Slovenia Use Interactive and Automated Scheduling Systems to Produce Quality Timetables** 358
▸ **APPLICATION CASE 6.4 Spreadsheet Helps Optimize Production Planning in Chilean Swine Companies** 359
▸ **APPLICATION CASE 6.5 Metro Meals on Wheels Treasure Valley Uses Excel to Find Optimal Delivery Routes** 360
6.6 Mathematical Programming Optimization 362
▸ **APPLICATION CASE 6.6 Mixed-Integer Programming Model Helps the University of Tennessee Medical Center with Scheduling Physicians** 363
Linear Programming Model 364
Modeling in LP: An Example 365
Implementation 370
6.7 Multiple Goals, Sensitivity Analysis, What-If Analysis, and Goal Seeking 372
Multiple Goals 372
Sensitivity Analysis 373
What-If Analysis 374
Goal Seeking 374
6.8 Decision Analysis with Decision Tables and Decision Trees 375
Decision Tables 376
Decision Trees 377
6.9 Introduction to Simulation 378
Major Characteristics of Simulation 378
▸ **APPLICATION CASE 6.7 Syngenta Uses Monte Carlo Simulation Models to Increase Soybean Crop Production** 379
Advantages of Simulation 380
Disadvantages of Simulation 381
The Methodology of Simulation 381

Simulation Types 382
Monte Carlo Simulation 383
Discrete Event Simulation 384
▸ **APPLICATION CASE 6.8 Cosan Improves Its Renewable Energy Supply Chain Using Simulation** 384
6.10 Visual Interactive Simulation 385
Conventional Simulation Inadequacies 385
Visual Interactive Simulation 385
Visual Interactive Models and DSS 386
Simulation Software 386
▸ **APPLICATION CASE 6.9 Improving Job-Shop Scheduling Decisions through RFID: A Simulation-Based Assessment** 387
Chapter Highlights 390
Key Terms 390
Questions for Discussion 391
Exercises 391
References 393

Chapter 7 Big Data Concepts and Tools 395

7.1 OPENING VIGNETTE: Analyzing Customer Churn in a Telecom Company Using Big Data Methods 396
7.2 Definition of Big Data 399
The "V"s That Define Big Data 400
▸ **APPLICATION CASE 7.1 Alternative Data for Market Analysis or Forecasts** 403
7.3 Fundamentals of Big Data Analytics 404
Business Problems Addressed by Big Data Analytics 407
▸ **APPLICATION CASE 7.2 Top Five Investment Bank Achieves Single Source of the Truth** 408
7.4 Big Data Technologies 409
MapReduce 409
Why Use MapReduce? 411
Hadoop 411
How Does Hadoop Work? 411
Hadoop Technical Components 412
Hadoop: The Pros and Cons 413
NoSQL 415
▸ **APPLICATION CASE 7.3 eBay's Big Data Solution** 416
▸ **APPLICATION CASE 7.4 Understanding Quality and Reliability of Healthcare Support Information on Twitter** 418

7.5 Big Data and Data Warehousing 419
Use Cases for Hadoop 419
Use Cases for Data Warehousing 420
The Gray Areas (Any One of the Two Would Do the Job) 421
Coexistence of Hadoop and Data Warehouse 422
7.6 Big Data Vendors and Platforms 423
IBM InfoSphere BigInsights 424
▸ **APPLICATION CASE 7.5 Using Social Media for Nowcasting the Flu Activity** 426
Teradata Aster 427
▸ **APPLICATION CASE 7.6 Analyzing Disease Patterns from an Electronic Medical Records Data Warehouse** 428
7.7 Big Data and Stream Analytics 432
Stream Analytics versus Perpetual Analytics 434
Critical Event Processing 434
Data Stream Mining 434
7.8 Applications of Stream Analytics 435
e-Commerce 435
Telecommunications 435
▸ **APPLICATION CASE 7.7 Salesforce Is Using Streaming Data to Enhance Customer Value** 436
Law Enforcement and Cybersecurity 437
Power Industry 437
Financial Services 437
Health Sciences 437
Government 438
Chapter Highlights 438
Key Terms 439
Questions for Discussion 439
Exercises 439
References 440

Chapter 8 Future Trends, Privacy and Managerial Considerations in Analytics 443

8.1 OPENING VIGNETTE: Analysis of Sensor Data Helps Siemens Avoid Train Failures 444
8.2 Internet of Things 445
▸ **APPLICATION CASE 8.1 SilverHook Powerboats Uses Real-Time Data Analysis to Inform Racers and Fans** 446
▸ **APPLICATION CASE 8.2 Rockwell Automation Monitors Expensive Oil and Gas Exploration Assets** 447
IoT Technology Infrastructure 448

RFID Sensors 448
Fog Computing 451
IoT Platforms 452
▸ **APPLICATION CASE 8.3 Pitney Bowes Collaborates with General Electric IoT Platform to Optimize Production** 452
IoT Start-Up Ecosystem 453
Managerial Considerations in the Internet of Things 454
8.3 Cloud Computing and Business Analytics 455
Data as a Service (DaaS) 457
Software as a Service (SaaS) 458
Platform as a Service (PaaS) 458
Infrastructure as a Service (IaaS) 458
Essential Technologies for Cloud Computing 459
Cloud Deployment Models 459
Major Cloud Platform Providers in Analytics 460
Analytics as a Service (AaaS) 461
Representative Analytics as a Service Offerings 461
Illustrative Analytics Applications Employing the Cloud Infrastructure 462
▸ **MD Anderson Cancer Center Utilizes Cognitive Computing Capabilities of IBM Watson to Give Better Treatment to Cancer Patients** 462
▸ **Public School Education in Tacoma, Washington, Uses Microsoft Azure Machine Learning to Predict School Dropouts** 463
▸ **Dartmouth-Hitchcock Medical Center Provides Personalized Proactive Healthcare Using Microsoft Cortana Analytics Suite** 464
▸ **Mankind Pharma Uses IBM Cloud Infrastructure to Reduce Application Implementation Time by 98%** 464
▸ **Gulf Air Uses Big Data to Get Deeper Customer Insight** 465
▸ **Chime Enhances Customer Experience Using Snowflake** 466
8.4 Location-Based Analytics for Organizations 467
Geospatial Analytics 467
▸ **APPLICATION CASE 8.4 Indian Police Departments Use Geospatial Analytics to Fight Crime** 469
▸ **APPLICATION CASE 8.5 Starbucks Exploits GIS and Analytics to Grow Worldwide** 470
Real-Time Location Intelligence 471
▸ **APPLICATION CASE 8.6 Quiznos Targets Customers for Its Sandwiches** 472
Analytics Applications for Consumers 472

8.5 Issues of Legality, Privacy, and Ethics 474
Legal Issues 474
Privacy 475
Collecting Information about Individuals 475
Mobile User Privacy 476
Homeland Security and Individual Privacy 476
Recent Technology Issues in Privacy and Analytics 477
Who Owns Our Private Data? 478
Ethics in Decision Making and Support 478
8.6 Impacts of Analytics in Organizations: An Overview 479
New Organizational Units 480
Redesign of an Organization through the Use of Analytics 481
Analytics Impact on Managers' Activities, Performance, and Job Satisfaction 481
Industrial Restructuring 482
Automation's Impact on Jobs 483
Unintended Effects of Analytics 484
8.7 Data Scientist as a Profession 485
Where Do Data Scientists Come From? 485
Chapter Highlights 488
Key Terms 489
Questions for Discussion 489
Exercises 489
References 490
Glossary 493
Index 501

This page intentionally left blank

Preface

Analytics has become the technology driver of this decade. Companies such as IBM, SAP, IBM, SAS, Teradata, SAP, Oracle, Microsoft, Dell and others are creating new organizational units focused on analytics that help businesses become more effective and efficient in their operations. Decision makers are using more computerized tools to support their work. Even consumers are using analytics tools, either directly or indirectly, to make decisions on routine activities such as shopping, health/healthcare, travel, and entertainment. The field of business intelligence and business analytics (BI & BA) has evolved rapidly to become more focused on innovative applications for extracting knowledge and insight from data streams that were not even captured some time back, much less analyzed in any significant way. New applications turn up daily in healthcare, sports, travel, entertainment, supply-chain management, utilities, and virtually every industry imaginable. The term *analytics* has become mainstream. Indeed, it has already evolved into other terms such as data science, and the latest incarnation is deep learning and Internet of Things.

This edition of the text provides a managerial perspective to business analytics continuum beginning with descriptive analytics (e.g., the nature of data, statistical modeling, data visualization, and business intelligence), moving on to predictive analytics (e.g., data mining, text/web mining, social media mining), and then to prescriptive analytics (e.g., optimization and simulation), and finally finishing with Big Data, and future trends, privacy, and managerial considerations. The book is supported by a Web site (pearsonglobaleditions.com/sharda) and also by an independent site at dssbibook.com. We will also provide links to software tutorials through a special section of the Web sites.

The purpose of this book is to introduce the reader to these technologies that are generally called *business analytics* or *data science* but have been known by other names. This book presents the fundamentals of the techniques and the manner in which these systems are constructed and used. We follow an EEE approach to introducing these topics: **Exposure**, **Experience**, and **Exploration**. The book primarily provides **exposure** to various analytics techniques and their applications. The idea is that a student will be inspired to learn from how other organizations have employed analytics to make decisions or to gain a competitive edge. We believe that such **exposure** to what is being done with analytics and how it can be achieved is the key component of learning about analytics. In describing the techniques, we also introduce specific software tools that can be used for developing such applications. The book is not limited to any one software tool, so the students can **experience** these techniques using any number of available software tools. Specific suggestions are given in each chapter, but the student and the professor are able to use this book with many different software tools. Our book's companion Web site will include specific software guides, but students can gain **experience** with these techniques in many different ways. Finally, we hope that this **exposure** and **experience** enable and motivate readers to **explore** the potential of these techniques in their own domain. To facilitate such **exploration**, we include exercises that direct them to Teradata University Network and other sites as well that include team-oriented exercises where appropriate. We will also highlight new and innovative applications that we learn about on the book's Web site.

Most of the specific improvements made in this fourth edition concentrate on four areas: reorganization, new chapters, content update, and a sharper focus. Despite the many changes, we have preserved the comprehensiveness and user friendliness that have made the text a market leader. Finally, we present accurate and updated material that is not available in any other text. We next describe the changes in the fourth edition.

What's New in the Fourth Edition?

With the goal of improving the text, this edition marks a major reorganization of the text to reflect the focus on business analytics. This edition is now organized around three major types of business analytics (i.e., descriptive, predictive, and prescriptive). The new edition has many timely additions, and the dated content has been deleted. The following major specific changes have been made.

- ***New organization.*** The book recognizes three types of analytics: descriptive, predictive, and prescriptive, a classification promoted by INFORMS. Chapter 1 introduces BI and analytics with an application focus in many industries. This chapter also includes an overview of the analytics ecosystem to help the user explore all the different ways one can participate and grow in the analytics environment. It is followed by an overview of statistics, importance of data, and descriptive analytics/visualization in Chapter 2. Chapter 3 covers data warehousing and data foundations including updated content, specifically data lakes. Chapter 4 covers predictive analytics. Chapter 5 extends the application of analytics to text, Web, and social media. Chapter 6 covers prescriptive analytics, specifically linear programming and simulation. It is totally new content for this book. Chapter 7 introduces Big Data tools and platforms. The book concludes with Chapter 8, emerging trends and topics in business analytics including location analytics, Internet of Things, cloud-based analytics, and privacy/ethical considerations in analytics. The discussion of an analytics ecosystem recognizes prescriptive analytics as well.
- ***New chapters.*** The following chapters have been added:

  Chapter 2. ***Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization*** This chapter aims to set the stage with a thorough understanding of the nature of data, which is the main ingredient for any analytics study. Next, statistical modeling is introduced as part of the descriptive analytics. Data visualization has become a popular part of any business reporting and/or descriptive analytics project; therefore, it is explained in detail in this chapter. The chapter is enhanced with several real-world cases and examples (75% new material).

  Chapter 6. ***Prescriptive Analytics: Optimization and Simulation*** This chapter introduces prescriptive analytics material to this book. The chapter focuses on optimization modeling in Excel using linear programming techniques. It also introduces the concept of simulation. The chapter is an updated version of material from two chapters in our DSS book, 10th edition. For this book it is an entirely new chapter (99% new material).

  Chapter 8. ***Future Trends, Privacy and Managerial Considerations in Analytics*** This chapter examines several new phenomena that are already changing or are likely to change analytics. It includes coverage of geospatial analytics, Internet of Things, and a significant update of the material on cloud-based analytics. It also updates some coverage from the last edition on ethical and privacy considerations (70% new material).
- ***Revised Chapters.*** All the other chapters have been revised and updated as well. Here is a summary of the changes in these other chapters:

  Chapter 1. ***An Overview of Business Intelligence, Analytics, and Data Science*** This chapter has been rewritten and significantly expanded. It opens with a new vignette covering multiple applications of analytics in sports. It introduces the three types of analytics as proposed by INFORMS: descriptive, predictive, and prescriptive analytics. A noted earlier, this classification is used in

guiding the complete reorganization of the book itself (earlier content but with a new figure). Then it includes several new examples of analytics in healthcare and in the retail industry. Finally, it concludes with significantly expanded and updated coverage of the analytics ecosystem to give the students a sense of the vastness of the analytics and data science industry (about 60% new material).

Chapter 3. *Descriptive Analytics II: Business Intelligence and Data Warehousing* This is an old chapter with some new subsections (e.g., data lakes) and new cases (about 30% new material).

Chapter 4. *Predictive Analytics I: Data Mining Process, Methods, and Algorithms* This is an old chapter with some new content organization/flow and some new cases (about 20% new material).

Chapter 5. *Predictive Analytics II: Text, Web, and Social Media Analytics* This is an old chapter with some new content organization/flow and some new cases (about 25% new material).

Chapter 7. *Big Data Concepts and Analysis* This was Chapter 6 in the last edition. It has been updated with a new opening vignette and cases, coverage of Teradata Aster, and new material on alternative data (about 25% new material).

- ***Revamped author team.*** Building on the excellent content that has been prepared by the authors of the previous editions (Turban, Sharda, Delen, and King), this edition was revised primarily by Ramesh Sharda and Dursun Delen. Both Ramesh and Dursun have worked extensively in analytics and have industry as well as research experience.
- ***Color print!*** We are truly excited to have this book appear in color. Even the figures from previous editions have been redrawn to take advantage of color. Use of color enhances many visualization examples and also the other material.
- ***A live, updated Web site.*** Adopters of the textbook will have access to a Web site that will include links to news stories, software, tutorials, and even YouTube videos related to topics covered in the book. This site will be accessible at dssbibook.com.
- ***Revised and updated content.*** Almost all the chapters have new opening vignettes that are based on recent stories and events. In addition, application cases throughout the book have been updated to include recent examples of applications of a specific technique/model. New Web site links have been added throughout the book. We also deleted many older product links and references. Finally, most chapters have new exercises, Internet assignments, and discussion questions throughout.
- ***Links to Teradata University Network (TUN).*** Most chapters include new links to TUN (teradatauniversitynetwork.com).
- ***Book title.*** As is already evident, the book's title and focus have changed substantially.
- ***Software support.*** The TUN Web site provides software support at no charge. It also provides links to free data mining and other software. In addition, the site provides exercises in the use of such software.

The Supplement Package: www.pearsonglobaleditions.com/sharda

A comprehensive and flexible technology-support package is available to enhance the teaching and learning experience. The following instructor and student supplements are available on the book's Web site, pearsonglobaleditions.com/sharda:

- ***Instructor's Manual.*** The Instructor's Manual includes learning objectives for the entire course and for each chapter, answers to the questions and exercises at the end

of each chapter, and teaching suggestions (including instructions for projects). The Instructor's Manual is available on the secure faculty section of pearsonglobaleditions.com/sharda.

- ***Test Item File and TestGen Software.*** The Test Item File is a comprehensive collection of true/false, multiple-choice, fill-in-the-blank, and essay questions. The questions are rated by difficulty level, and the answers are referenced by book page number. The Test Item File is available in Microsoft Word and in TestGen. Pearson Education's test-generating software is available from www.pearsonglobaleditions.com/sharda. The software is PC/MAC compatible and preloaded with all the Test Item File questions. You can manually or randomly view test questions and drag-and-drop to create a test. You can add or modify test-bank questions as needed.
- ***PowerPoint slides.*** PowerPoint slides are available that illuminate and build on key concepts in the text. Faculty can download the PowerPoint slides from pearsonglobaleditions.com/sharda.

Acknowledgments

Many individuals have provided suggestions and criticisms since the publication of the first edition of this book. Dozens of students participated in class testing of various chapters, software, and problems and assisted in collecting material. It is not possible to name everyone who participated in this project, but our thanks go to all of them. Certain individuals made significant contributions, and they deserve special recognition.

First, we appreciate the efforts of those individuals who provided formal reviews of the first through third editions (school affiliations as of the date of review):

Ann Aksut, Central Piedmont Community College
Bay Arinze, Drexel University
Andy Borchers, Lipscomb University
Ranjit Bose, University of New Mexico
Marty Crossland, MidAmerica Nazarene University
Kurt Engemann, Iona College
Badie Farah, Eastern Michigan University
Gary Farrar, Columbia College
Jerry Fjermestad, New Jersey Institute of Technology
Christie M. Fuller, Louisiana Tech University
Martin Grossman, Bridgewater State College
Jahangir Karimi, University of Colorado, Denver
Huei Lee, Eastern Michigan University
Natalie Nazarenko, SUNY Fredonia
Joo Eng Lee-Partridge, Central Connecticut State University
Gregory Rose, Washington State University, Vancouver
Khawaja Saeed, Wichita State University
Kala Chand Seal, Loyola Marymount University
Joshua S. White, PhD, State University of New York Polytechnic Institute
Roger Wilson, Fairmont State University
Vincent Yu, Missouri University of Science and Technology
Fan Zhao, Florida Gulf Coast University

We also appreciate the efforts of those individuals who provided formal reviews of this text and our other DSS book—*Business Intelligence and Analytics: Systems for Decision Support*, 10th Edition, Pearson Education, 2013.

Second, several individuals contributed material to the text or the supporting material. Susan Baskin of Teradata and Dr. David Schrader provided special help in identifying new TUN and Teradata content for the book and arranging permissions for the same. Dr. Dave Schrader contributed the opening vignette for the book. This vignette also included material developed by Dr. Ashish Gupta of Auburn University and Gary Wilkerson of the University of Tennessee–Chattanooga. It will provide a great introduction to analytics. We also thank INFORMS for their permission to highlight content from *Interfaces*. We also recognize the following individuals for their assistance in developing this edition of the book: Pankush Kalgotra, Prasoon Mathur, Rupesh Agarwal, Shubham Singh, Nan Liang, Jacob Pearson, Kinsey Clemmer, and Evan Murlette (all of Oklahoma State University). Their help for this edition is gratefully acknowledged. Teradata Aster team, especially Mark Ott, provided the material for the opening vignette for Chapter 7. Aster material in Chapter 7 is adapted from other training guides developed by John Thuma and Greg Bethardy. Dr. Brian LeClaire, CIO of Humana Corporation led with contributions of several real-life healthcare case studies developed by his team at Humana. Abhishek Rathi of vCreaTek contributed his vision of analytics in the retail industry. Dr. Rick Wilson's excellent exercises for teaching and practicing linear programming skills in Excel are also gratefully acknowledged. Matt Turck agreed to let us adapt his IoT ecosystem material. Ramesh also recognizes the copyediting assistance provided by his daughter, Ruchy Sharda Sen. In addition, the following former PhD students and research colleagues of ours have provided content or advice and support for the book in many direct and indirect ways:

Asil Oztekin, Universality of Massachusetts-Lowell
Enes Eryarsoy, Sehir University
Hamed Majidi Zolbanin, Ball State University
Amir Hassan Zadeh, Wright State University
Supavich (Fone) Pengnate, North Dakota State University
Christie Fuller, Boise State University
Daniel Asamoah, Wright State University
Selim Zaim, Istanbul Technical University
Nihat Kasap, Sabanci University

Third, for the previous edition, we acknowledge the contributions of Dave King (JDA Software Group, Inc.). Other major contributors to the previous edition include J. Aronson (University of Georgia), who was our coauthor, contributing to the data warehousing chapter; Mike Goul (Arizona State University), whose contributions were included in Chapter 1; and T. P. Liang (National Sun Yet-Sen University, Taiwan), who contributed material on neural networks in the previous editions. Judy Lang collaborated with all of us, provided editing, and guided us during the entire project in the first edition.

Fourth, several vendors cooperated by providing case studies and/or demonstration software for the previous editions: Acxiom (Little Rock, Arkansas), California Scientific Software (Nevada City, California), Cary Harwin of Catalyst Development (Yucca Valley, California), IBM (San Carlos, California), DS Group, Inc. (Greenwich, Connecticut), Gregory Piatetsky-Shapiro of KDnuggets.com, Gary Lynn of NeuroDimension Inc. (Gainesville, Florida), Palisade Software (Newfield, New York), Promised Land Technologies (New Haven, Connecticut), Salford Systems (La Jolla, California), Sense Networks (New York, New York), Gary Miner of StatSoft, Inc. (Tulsa, Oklahoma), Ward Systems Group, Inc. (Frederick, Maryland), Idea Fisher Systems, Inc. (Irving, California), and Wordtech Systems (Orinda, California).

Fifth, special thanks to the Teradata University Network and especially to Susan Baskin, Program Director; Hugh Watson, who started TUN; and Michael Goul, Barb Wixom, and Mary Gros for their encouragement to tie this book with TUN and for providing useful material for the book.

Finally, the Pearson team is to be commended: Samantha Lewis, who has worked with us on this revision and orchestrated the color rendition of the book; and the production team, Ann Pulido, and Revathi Viswanathan and staff at Cenveo, who transformed the manuscript into a book.

We would like to thank all these individuals and corporations. Without their help, the creation of this book would not have been possible.

R.S.

D.D.

E.T.

Global Edition Acknowledgments

For his contributions to the content of the Global Edition, Pearson would like to thank Bálint Molnár (Eötvös Loránd University, Budapest), and for their feedback, Daqing Chen (London South Bank University), Ng Hu (Multimedia University, Malaysia), and Vanina Torlo (University of Greenwich).

Note that Web site URLs are dynamic. As this book went to press, we verified that all the cited Web sites were active and valid. Web sites to which we refer in the text sometimes change or are discontinued because companies change names, are bought or sold, merge, or fail. Sometimes Web sites are down for maintenance, repair, or redesign. Most organizations have dropped the initial "www" designation for their sites, but some still use it. If you have a problem connecting to a Web site that we mention, please be patient and simply run a Web search to try to identify the new site. Most times, the new site can be found quickly. We apologize in advance for this inconvenience.

About the Authors

Ramesh Sharda (MBA, PhD, University of Wisconsin–Madison) is the Vice Dean for Research and Graduate Programs, Watson/ConocoPhillips Chair, and a Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University (OSU). He cofounded and directed OSU's PhD in Business for the Executives Program. About 200 papers describing his research have been published in major journals, including *Operations Research*, *Management Science*, *Information Systems Research*, *Decision Support Systems*, and the *Journal of MIS*. He cofounded the AIS SIG on Decision Support Systems and Knowledge Management (SIGDSA). Dr. Sharda serves on several editorial boards, including those of *Decision Sciences Journal*, *Decision Support Systems*, and *ACM Data Base*. He has authored and edited several textbooks and research books and serves as the coeditor of several Springer book series (Integrated Series in Information Systems, Operations Research/Computer Science Interfaces, and Annals of Information Systems) with Springer. He is also currently serving as the Executive Director of the Teradata University Network. His current research interests are in decision support systems, business analytics, and technologies for managing information overload.

Dursun Delen (PhD, Oklahoma State University) is the Spears Endowed Chair in Business Administration, Patterson Foundation Endowed Chair in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University (OSU). Prior to his academic career, he worked for a privately owned research and consultancy company, Knowledge Based Systems Inc., in College Station, Texas, as a research scientist for 5 years, during which he led a number of decision support and other information systems–related research projects funded by several federal agencies including the Department of Defense (DoD), National Aeronautics and Space Administration (NASA), National Institute for Standards and Technology (NIST), Ballistic Missile Defense Organization (BMDO), and Department of Energy (DOE). Dr. Delen has published more than 100 peer-reviewed articles, some of which have appeared in major journals like *Decision Sciences*, *Decision Support Systems*, *Communications of the ACM*, *Computers and Operations Research*, *Computers in Industry*, *Journal of Production Operations Management*, *Artificial Intelligence in Medicine*, *International Journal of Medical Informatics*, *Expert Systems with Applications*, and *IEEE Wireless Communications*. He recently authored/coauthored seven textbooks in the broad areas of business analytics, data mining, text mining, business intelligence, and decision support systems. He is often invited to national and international conferences for keynote addresses on topics related to data/text mining, business analytics, decision support systems, business intelligence, and knowledge management. He served as the General Cochair for the Fourth International Conference on Network Computing and Advanced Information Management (September 2–4, 2008, in Seoul, South Korea) and regularly chairs, tracks, and minitracks at various information systems and analytics conferences. He is currently serving as Editor-in-Chief, Senior Editor, Associate Editor, or Editorial Board Member for more than a dozen academic journals. His research and teaching interests are in data and text mining, business analytics, decision support systems, knowledge management, business intelligence, and enterprise modeling.

Efraim Turban (MBA, PhD., University of California, Berkeley) is a Visiting Scholar at the Pacific Institute for Information System Management, University of Hawaii. Prior to this,

he was on the staff of several universities, including City University of Hong Kong; Lehigh University; Florida International University; California State University, Long Beach; Eastern Illinois University; and the University of Southern California. Dr. Turban is the author of more than 100 refereed papers published in leading journals, such as *Management Science*, *MIS Quarterly*, and *Decision Support Systems*. He is also the author of 20 books, including *Electronic Commerce: A Managerial Perspective* and *Information Technology for Management*. He is also a consultant to major corporations worldwide. Dr. Turban's current areas of interest are Web-based decision support systems, social commerce, and collaborative decision making.

Business Intelligence, Analytics, and Data Science

A Managerial Perspective

This book deals with a collection of computer technologies that support managerial work—essentially, decision making. These technologies have had a profound impact on corporate strategy, performance, and competitiveness. Collectively, these technologies are called *business intelligence*, *business analytics,* and *data science*. Although the evolution of the terms is discussed, these names are also used interchangeably. This book tells stories of how smart people are employing these techniques to improve performance, service, and relationships in business, government, and non-profit worlds.

This page intentionally left blank

CHAPTER

1

An Overview of Business Intelligence, Analytics, and Data Science

LEARNING OBJECTIVES

- Understand the need for computerized support of managerial decision making
- Recognize the evolution of such computerized support to the current state—analytics/data science
- Describe the business intelligence (BI) methodology and concepts
- Understand the different types of analytics and see selected applications
- Understand the analytics ecosystem to identify various key players and career opportunities

The business environment (climate) is constantly changing, and it is becoming more and more complex. Organizations, both private and public, are under pressures that force them to respond quickly to changing conditions and to be innovative in the way they operate. Such activities require organizations to be agile and to make frequent and quick strategic, tactical, and operational decisions, some of which are very complex. Making such decisions may require considerable amounts of relevant data, information, and knowledge. Processing these, in the framework of the needed decisions, must be done quickly, frequently in real time, and usually requires some computerized support.

This book is about using business analytics as computerized support for managerial decision making. It concentrates on the theoretical and conceptual foundations of decision support, as well as on the commercial tools and techniques that are available. This book presents the fundamentals of the techniques and the manner in which these systems are constructed and used. We follow an EEE approach to introducing these topics: **Exposure**, **Experience**, and **Exploration**. The book primarily provides exposure to various analytics techniques and their applications. The idea is that a student will be inspired to learn from how other organizations have employed analytics to make decisions or to gain a competitive edge. We believe that such **exposure** to what is being done with analytics and how it can be achieved is the key component of learning about analytics. In describing the techniques, we also give examples of specific software tools that can be

used for developing such applications. The book is not limited to any one software tool, so students can **experience** these techniques using any number of available software tools. We hope that this exposure and experience enable and motivate readers to explore the potential of these techniques in their own domain. To facilitate such **exploration**, we include exercises that direct the reader to Teradata University Network (TUN) and other sites that include team-oriented exercises where appropriate.

This introductory chapter provides an introduction to analytics as well as an overview of the book. The chapter has the following sections:

1.1 Opening Vignette: Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics 30
1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics 37
1.3 Evolution of Computerized Decision Support to Analytics/Data Science 39
1.4 A Framework for Business Intelligence 41
1.5 Analytics Overview 48
1.6 Analytics Examples in Selected Domains 55
1.7 A Brief Introduction to Big Data Analytics 61
1.8 An Overview of the Analytics Ecosystem 63
1.9 Plan of the Book 72
1.10 Resources, Links, and the Teradata University Network Connection 73

1.1 OPENING VIGNETTE: Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics

The application of analytics to business problems is a key skill, one that you will learn in this book. Many of these techniques are now being applied to improve decision making in all aspects of sports, a very hot area called sports analytics. Sports analytics is the art and science of gathering data about athletes and teams to create insights that improve sports decisions, such as deciding which players to recruit, how much to pay them, who to play, how to train them, how to keep them healthy, and when they should be traded or retired. For teams, it involves business decisions such as ticket pricing, as well as roster decisions, analysis of each competitor's strengths and weaknesses, and many game-day decisions.

Indeed, sports analytics is becoming a specialty within analytics. It is an important area because sports is a big business, generating about $145B in revenues each year, plus an additional $100B in legal and $300B in illegal gambling, according to Price Waterhouse.[1] In 2014, only $125M was spent on analytics (less than 0.1% of revenues). This is expected to grow at a healthy rate to $4.7B by 2021.[2]

[1] "Changing the Game: Outlook for the Global Sports Market to 2015," Price Waterhouse Coopers Report, appears at https://www.pwc.com/gx/en/hospitality-leisure/pdf/changing-the-game-outlook-for-the-global-sports-market-to-2015.pdf. Betting data from https://www.capcredit.com/how-much-americansspend-on-sports-each-year/.

[2] "Sports Analytics Market Worth $4.7B by 2021," Wintergreen Research Press Release, covered by PR Newswire at http://www.prnewswire.com/news-releases/sports-analytics-market-worth-47-billion-by-2021-509869871.html, June 25, 2015.