FOURTH 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

330 Hudson Street, NY NY 10013

Brief Contents

Preface xix
About the Authors xxv

Chapter 1 **An Overview of Business Intelligence, Analytics, and Data Science 3**
Chapter 2 **Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization 53**
Chapter 3 **Descriptive Analytics II: Business Intelligence and Data Warehousing 127**
Chapter 4 **Predictive Analytics I: Data Mining Process, Methods, and Algorithms 189**
Chapter 5 **Predictive Analytics II: Text, Web, and Social Media Analytics 247**
Chapter 6 **Prescriptive Analytics: Optimization and Simulation 319**
Chapter 7 **Big Data Concepts and Tools 369**
Chapter 8 **Future Trends, Privacy and Managerial Considerations in Analytics 417**

Glossary 467
Index 475

Contents

Preface xix
About the Authors xxv

Chapter 1 **An Overview of Business Intelligence, Analytics, and Data Science** 3

1.1 OPENING VIGNETTE: Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics 4
1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics 11
1.3 Evolution of Computerized Decision Support to Analytics/Data Science 13
1.4 A Framework for Business Intelligence 15
Definitions of BI 16
A Brief History of BI 16
The Architecture of BI 16
The Origins and Drivers of BI 16
▶ **APPLICATION CASE 1.1 Sabre Helps Its Clients through Dashboards and Analytics** 18
A Multimedia Exercise in Business Intelligence 19
Transaction Processing versus Analytic Processing 19
Appropriate Planning and Alignment with the Business Strategy 20
Real-Time, On-Demand BI Is Attainable 21
Developing or Acquiring BI Systems 21
Justification and Cost–Benefit Analysis 22
Security and Protection of Privacy 22
Integration of Systems and Applications 22
1.5 Analytics Overview 22
Descriptive Analytics 24
▶ **APPLICATION CASE 1.2 Silvaris Increases Business with Visual Analysis and Real-Time Reporting Capabilities** 24
▶ **APPLICATION CASE 1.3 Siemens Reduces Cost with the Use of Data Visualization** 25
Predictive Analytics 25
▶ **APPLICATION CASE 1.4 Analyzing Athletic Injuries** 26
Prescriptive Analytics 26
Analytics Applied to Different Domains 27
▶ **APPLICATION CASE 1.5 A Specialty Steel Bar Company Uses Analytics to Determine Available-to-Promise Dates** 27
Analytics or Data Science? 28

1.6 Analytics Examples in Selected Domains 29
Analytics Applications in Healthcare—Humana Examples 29
Analytics in the Retail Value Chain 33
1.7 A Brief Introduction to Big Data Analytics 35
What Is Big Data? 35
▶ **APPLICATION CASE 1.6 CenterPoint Energy Uses Real-Time Big Data Analytics to Improve Customer Service** 37
1.8 An Overview of the Analytics Ecosystem 37
Data Generation Infrastructure Providers 39
Data Management Infrastructure Providers 39
Data Warehouse Providers 40
Middleware Providers 40
Data Service Providers 40
Analytics-Focused Software Developers 41
Application Developers: Industry Specific or General 42
Analytics Industry Analysts and Influencers 43
Academic Institutions and Certification Agencies 44
Regulators and Policy Makers 45
Analytics User Organizations 45
1.9 Plan of the Book 46
1.10 Resources, Links, and the Teradata University Network Connection 47
Resources and Links 47
Vendors, Products, and Demos 48
Periodicals 48
The Teradata University Network Connection 48
The Book's Web Site 48
Chapter Highlights 49
Key Terms 49
Questions for Discussion 49
Exercises 50
References 51

Chapter 2 Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization 53

2.1 OPENING VIGNETTE: SiriusXM Attracts and Engages a New Generation of Radio Consumers with Data-Driven Marketing 54
2.2 The Nature of Data 57
2.3 A Simple Taxonomy of Data 61
▶ **APPLICATION CASE 2.1 Medical Device Company Ensures Product Quality While Saving Money** 63

2.4 The Art and Science of Data Preprocessing 65
► **APPLICATION CASE 2.2 Improving Student Retention with Data-Driven Analytics** 68
2.5 Statistical Modeling for Business Analytics 74
Descriptive Statistics for Descriptive Analytics 75
Measures of Centrality Tendency (May Also Be Called Measures of Location or Centrality) 76
Arithmetic Mean 76
Median 77
Mode 77
Measures of Dispersion (May Also Be Called Measures of Spread Decentrality) 77
Range 78
Variance 78
Standard Deviation 78
Mean Absolute Deviation 78
Quartiles and Interquartile Range 78
Box-and-Whiskers Plot 79
The Shape of a Distribution 80
► **APPLICATION CASE 2.3 Town of Cary Uses Analytics to Analyze Data from Sensors, Assess Demand, and Detect Problems** 84
2.6 Regression Modeling for Inferential Statistics 86
How Do We Develop the Linear Regression Model? 87
How Do We Know If the Model Is Good Enough? 88
What Are the Most Important Assumptions in Linear Regression? 89
Logistic Regression 90
► **APPLICATION CASE 2.4 Predicting NCAA Bowl Game Outcomes** 91
Time Series Forecasting 96
2.7 Business Reporting 98
► **APPLICATION CASE 2.5 Flood of Paper Ends at FEMA** 100
2.8 Data Visualization 101
A Brief History of Data Visualization 101
► **APPLICATION CASE 2.6 Macfarlan Smith Improves Operational Performance Insight with Tableau Online** 103
2.9 Different Types of Charts and Graphs 106
Basic Charts and Graphs 106
Specialized Charts and Graphs 107
Which Chart or Graph Should You Use? 108
2.10 The Emergence of Visual Analytics 110
Visual Analytics 112
High-Powered Visual Analytics Environments 112

2.11 Information Dashboards 117
▶ **APPLICATION CASE 2.7 Dallas Cowboys Score Big with Tableau and Teknion** 118
Dashboard Design 119
▶ **APPLICATION CASE 2.8 Visual Analytics Helps Energy Supplier Make Better Connections** 119
What to Look for in a Dashboard 121
Best Practices in Dashboard Design 121
Benchmark Key Performance Indicators with Industry Standards 121
Wrap the Dashboard Metrics with Contextual Metadata 121
Validate the Dashboard Design by a Usability Specialist 122
Prioritize and Rank Alerts/Exceptions Streamed to the Dashboard 122
Enrich Dashboard with Business-User Comments 122
Present Information in Three Different Levels 122
Pick the Right Visual Construct Using Dashboard Design Principles 122
Provide for Guided Analytics 122
Chapter Highlights 123
Key Terms 123
Questions for Discussion 124
Exercises 124
References 126

Chapter 3 Descriptive Analytics II: Business Intelligence and Data Warehousing 127

3.1 OPENING VIGNETTE: Targeting Tax Fraud with Business Intelligence and Data Warehousing 128
3.2 Business Intelligence and Data Warehousing 130
What Is a Data Warehouse? 131
A Historical Perspective to Data Warehousing 132
Characteristics of Data Warehousing 133
Data Marts 134
Operational Data Stores 135
Enterprise Data Warehouses (EDW) 135
Metadata 135
▶ **APPLICATION CASE 3.1 A Better Data Plan: Well-Established TELCOs Leverage Data Warehousing and Analytics to Stay on Top in a Competitive Industry** 135
3.3 Data Warehousing Process 137
3.4 Data Warehousing Architectures 139
Alternative Data Warehousing Architectures 142
Which Architecture Is the Best? 144

3.5 Data Integration and the Extraction, Transformation, and Load (ETL) Processes 145

Data Integration 146

▶ **APPLICATION CASE 3.2 BP Lubricants Achieves BIGS Success** 146

Extraction, Transformation, and Load 148

3.6 Data Warehouse Development 150

▶ **APPLICATION CASE 3.3 Use of Teradata Analytics for SAP Solutions Accelerates Big Data Delivery** 151

Data Warehouse Development Approaches 153

Additional Data Warehouse Development Considerations 156

Representation of Data in Data Warehouse 156

Analysis of Data in Data Warehouse 158

OLAP versus OLTP 158

OLAP Operations 159

3.7 Data Warehousing Implementation Issues 160

Massive Data Warehouses and Scalability 162

▶ **APPLICATION CASE 3.4 EDW Helps Connect State Agencies in Michigan** 163

3.8 Data Warehouse Administration, Security Issues, and Future Trends 164

The Future of Data Warehousing 165

3.9 Business Performance Management 170

Closed-Loop BPM Cycle 171

▶ **APPLICATION CASE 3.5 AARP Transforms Its BI Infrastructure and Achieves a 347% ROI in Three Years** 173

3.10 Performance Measurement 175

Key Performance Indicator (KPI) 175

Performance Measurement System 176

3.11 Balanced Scorecards 177

The Four Perspectives 177

The Meaning of Balance in BSC 179

3.12 Six Sigma as a Performance Measurement System 179

The DMAIC Performance Model 180

Balanced Scorecard versus Six Sigma 180

Effective Performance Measurement 181

▶ **APPLICATION CASE 3.6 Expedia.com's Customer Satisfaction Scorecard** 182

Chapter Highlights 183

Key Terms 184

Questions for Discussion 184

Exercises 185

References 187

Chapter 4 Predictive Analytics I: Data Mining Process, Methods, and Algorithms 189

4.1 OPENING VIGNETTE: Miami-Dade Police Department Is Using Predictive Analytics to Foresee and Fight Crime 190

4.2 Data Mining Concepts and Applications 193

▸ **APPLICATION CASE 4.1 Visa Is Enhancing the Customer Experience While Reducing Fraud with Predictive Analytics and Data Mining** 194

Definitions, Characteristics, and Benefits 196

How Data Mining Works 197

▸ **APPLICATION CASE 4.2 Dell Is Staying Agile and Effective with Analytics in the 21st Century** 198

Data Mining versus Statistics 203

4.3 Data Mining Applications 203

▸ **APPLICATION CASE 4.3 Predictive Analytic and Data Mining Help Stop Terrorist Funding** 205

4.4 Data Mining Process 206

Step 1: Business Understanding 207

Step 2: Data Understanding 208

Step 3: Data Preparation 208

Step 4: Model Building 209

▸ **APPLICATION CASE 4.4 Data Mining Helps in Cancer Research** 209

Step 5: Testing and Evaluation 212

Step 6: Deployment 212

Other Data Mining Standardized Processes and Methodologies 212

4.5 Data Mining Methods 215

Classification 215

Estimating the True Accuracy of Classification Models 216

▸ **APPLICATION CASE 4.5 Influence Health Uses Advanced Predictive Analytics to Focus on the Factors That Really Influence People's Healthcare Decisions** 223

Cluster Analysis for Data Mining 225

Association Rule Mining 227

4.6 Data Mining Software Tools 231

▸ **APPLICATION CASE 4.6 Data Mining Goes to Hollywood: Predicting Financial Success of Movies** 233

4.7 Data Mining Privacy Issues, Myths, and Blunders 237

▸ **APPLICATION CASE 4.7 Predicting Customer Buying Patterns—The Target Story** 238

Data Mining Myths and Blunders 238

Chapter Highlights 241

Key Terms 242

Questions for Discussion 242

Exercises 243

References 245

Chapter 5 Predictive Analytics II: Text, Web, and Social Media Analytics 247

5.1 OPENING VIGNETTE: Machine versus Men on Jeopardy!: The Story of Watson 248

5.2 Text Analytics and Text Mining Overview 251

▶ **APPLICATION CASE 5.1 Insurance Group Strengthens Risk Management with Text Mining Solution** 254

5.3 Natural Language Processing (NLP) 255

▶ **APPLICATION CASE 5.2 AMC Networks Is Using Analytics to Capture New Viewers, Predict Ratings, and Add Value for Advertisers in a Multichannel World** 257

5.4 Text Mining Applications 261

Marketing Applications 261

Security Applications 261

▶ **APPLICATION CASE 5.3 Mining for Lies** 262

Biomedical Applications 264

Academic Applications 266

▶ **APPLICATION CASE 5.4 Bringing the Customer into the Quality Equation: Lenovo Uses Analytics to Rethink Its Redesign** 266

5.5 Text Mining Process 268

Task 1: Establish the Corpus 269

Task 2: Create the Term–Document Matrix 269

Task 3: Extract the Knowledge 271

▶ **APPLICATION CASE 5.5 Research Literature Survey with Text Mining** 273

5.6 Sentiment Analysis 276

▶ **APPLICATION CASE 5.6 Creating a Unique Digital Experience to Capture the Moments That Matter at Wimbledon** 277

Sentiment Analysis Applications 280

Sentiment Analysis Process 282

Methods for Polarity Identification 284

Using a Lexicon 284

Using a Collection of Training Documents 285

Identifying Semantic Orientation of Sentences and Phrases 286

Identifying Semantic Orientation of Documents 286

5.7 Web Mining Overview 287

Web Content and Web Structure Mining 289

5.8 Search Engines 291

Anatomy of a Search Engine 292

1. Development Cycle 292

2. Response Cycle 293
Search Engine Optimization 294
Methods for Search Engine Optimization 295
▸ **APPLICATION CASE 5.7 Understanding Why Customers Abandon Shopping Carts Results in a $10 Million Sales Increase** 297
5.9 Web Usage Mining (Web Analytics) 298
Web Analytics Technologies 299
Web Analytics Metrics 300
Web Site Usability 300
Traffic Sources 301
Visitor Profiles 302
Conversion Statistics 302
5.10 Social Analytics 304
Social Network Analysis 304
Social Network Analysis Metrics 305
▸ **APPLICATION CASE 5.8 Tito's Vodka Establishes Brand Loyalty with an Authentic Social Strategy** 305
Connections 308
Distributions 308
Segmentation 309
Social Media Analytics 309
How Do People Use Social Media? 310
Measuring the Social Media Impact 311
Best Practices in Social Media Analytics 311
Chapter Highlights 313
Key Terms 314
Questions for Discussion 315
Exercises 315
References 316

Chapter 6 Prescriptive Analytics: Optimization and Simulation 319

6.1 OPENING VIGNETTE: School District of Philadelphia Uses Prescriptive Analytics to Find Optimal Solution for Awarding Bus Route Contracts 320
6.2 Model-Based Decision Making 322
Prescriptive Analytics Model Examples 322
▸ **APPLICATION CASE 6.1 Optimal Transport for ExxonMobil Downstream through a DSS** 323

Identification of the Problem and Environmental Analysis 324
Model Categories 324
► **APPLICATION CASE 6.2 Ingram Micro Uses Business Intelligence Applications to Make Pricing Decisions** 325
6.3 Structure of Mathematical Models for Decision Support 328
The Components of Decision Support Mathematical Models 328
The Structure of Mathematical Models 329
6.4 Certainty, Uncertainty, and Risk 330
Decision Making under Certainty 330
Decision Making under Uncertainty 331
Decision Making under Risk (Risk Analysis) 331
6.5 Decision Modeling with Spreadsheets 331
► **APPLICATION CASE 6.3 American Airlines Uses Should-Cost Modeling to Assess the Uncertainty of Bids for Shipment Routes** 332
► **APPLICATION CASE 6.4 Pennsylvania Adoption Exchange Uses Spreadsheet Model to Better Match Children with Families** 333
► **APPLICATION CASE 6.5 Metro Meals on Wheels Treasure Valley Uses Excel to Find Optimal Delivery Routes** 334
6.6 Mathematical Programming Optimization 336
► **APPLICATION CASE 6.6 Mixed-Integer Programming Model Helps the University of Tennessee Medical Center with Scheduling Physicians** 337
Linear Programming Model 338
Modeling in LP: An Example 339
Implementation 344
6.7 Multiple Goals, Sensitivity Analysis, What-If Analysis, and Goal Seeking 346
Multiple Goals 346
Sensitivity Analysis 347
What-If Analysis 348
Goal Seeking 348
6.8 Decision Analysis with Decision Tables and Decision Trees 349
Decision Tables 350
Decision Trees 351
6.9 Introduction to Simulation 352
Major Characteristics of Simulation 352
► **APPLICATION CASE 6.7 Simulating Effects of Hepatitis B Interventions** 353
Advantages of Simulation 354
Disadvantages of Simulation 355
The Methodology of Simulation 355

Simulation Types 356
Monte Carlo Simulation 357
Discrete Event Simulation 358
► **APPLICATION CASE 6.8 Cosan Improves Its Renewable Energy Supply Chain Using Simulation** 358
6.10 Visual Interactive Simulation 359
Conventional Simulation Inadequacies 359
Visual Interactive Simulation 359
Visual Interactive Models and DSS 360
Simulation Software 360
► **APPLICATION CASE 6.9 Improving Job-Shop Scheduling Decisions through RFID: A Simulation-Based Assessment** 361
Chapter Highlights 364
Key Terms 364
Questions for Discussion 365
Exercises 365
References 367

Chapter 7 Big Data Concepts and Tools 369

7.1 OPENING VIGNETTE: Analyzing Customer Churn in a Telecom Company Using Big Data Methods 370
7.2 Definition of Big Data 373
The "V"s That Define Big Data 374
► **APPLICATION CASE 7.1 Alternative Data for Market Analysis or Forecasts** 377
7.3 Fundamentals of Big Data Analytics 378
Business Problems Addressed by Big Data Analytics 381
► **APPLICATION CASE 7.2 Top Five Investment Bank Achieves Single Source of the Truth** 382
7.4 Big Data Technologies 383
MapReduce 383
Why Use MapReduce? 385
Hadoop 385
How Does Hadoop Work? 385
Hadoop Technical Components 386
Hadoop: The Pros and Cons 387
NoSQL 389
► **APPLICATION CASE 7.3 eBay's Big Data Solution** 390
► **APPLICATION CASE 7.4 Understanding Quality and Reliability of Healthcare Support Information on Twitter** 392

7.5 Big Data and Data Warehousing 393
Use Cases for Hadoop 393
Use Cases for Data Warehousing 394
The Gray Areas (Any One of the Two Would Do the Job) 395
Coexistence of Hadoop and Data Warehouse 396
7.6 Big Data Vendors and Platforms 397
IBM InfoSphere BigInsights 398
► **APPLICATION CASE 7.5 Using Social Media for Nowcasting the Flu Activity** 400
Teradata Aster 401
► **APPLICATION CASE 7.6 Analyzing Disease Patterns from an Electronic Medical Records Data Warehouse** 402
7.7 Big Data and Stream Analytics 406
Stream Analytics versus Perpetual Analytics 408
Critical Event Processing 408
Data Stream Mining 408
7.8 Applications of Stream Analytics 409
e-Commerce 409
Telecommunications 409
► **APPLICATION CASE 7.7 Salesforce Is Using Streaming Data to Enhance Customer Value** 410
Law Enforcement and Cybersecurity 411
Power Industry 411
Financial Services 411
Health Sciences 411
Government 412
Chapter Highlights 412
Key Terms 413
Questions for Discussion 413
Exercises 413
References 414

Chapter 8 Future Trends, Privacy and Managerial Considerations in Analytics 417

8.1 OPENING VIGNETTE: Analysis of Sensor Data Helps Siemens Avoid Train Failures 418
8.2 Internet of Things 419
► **APPLICATION CASE 8.1 SilverHook Powerboats Uses Real-Time Data Analysis to Inform Racers and Fans** 420
► **APPLICATION CASE 8.2 Rockwell Automation Monitors Expensive Oil and Gas Exploration Assets** 421
IoT Technology Infrastructure 422

RFID Sensors 422
Fog Computing 425
IoT Platforms 426
▶ **APPLICATION CASE 8.3 Pitney Bowes Collaborates with General Electric IoT Platform to Optimize Production** 426
IoT Start-Up Ecosystem 427
Managerial Considerations in the Internet of Things 428
8.3 Cloud Computing and Business Analytics 429
Data as a Service (DaaS) 431
Software as a Service (SaaS) 432
Platform as a Service (PaaS) 432
Infrastructure as a Service (IaaS) 432
Essential Technologies for Cloud Computing 433
Cloud Deployment Models 433
Major Cloud Platform Providers in Analytics 434
Analytics as a Service (AaaS) 435
Representative Analytics as a Service Offerings 435
Illustrative Analytics Applications Employing the Cloud Infrastructure 436
▶ **MD Anderson Cancer Center Utilizes Cognitive Computing Capabilities of IBM Watson to Give Better Treatment to Cancer Patients** 436
▶ **Public School Education in Tacoma, Washington, Uses Microsoft Azure Machine Learning to Predict School Dropouts** 437
▶ **Dartmouth-Hitchcock Medical Center Provides Personalized Proactive Healthcare Using Microsoft Cortana Analytics Suite** 438
▶ **Mankind Pharma Uses IBM Cloud Infrastructure to Reduce Application Implementation Time by 98%** 438
▶ **Gulf Air Uses Big Data to Get Deeper Customer Insight** 439
▶ **Chime Enhances Customer Experience Using Snowflake** 440
8.4 Location-Based Analytics for Organizations 441
Geospatial Analytics 441
▶ **APPLICATION CASE 8.4 Great Clips Employs Spatial Analytics to Shave Time in Location Decisions** 443
▶ **APPLICATION CASE 8.5 Starbucks Exploits GIS and Analytics to Grow Worldwide** 444
Real-Time Location Intelligence 445
▶ **APPLICATION CASE 8.6 Quiznos Targets Customers for Its Sandwiches** 446
Analytics Applications for Consumers 446

8.5 Issues of Legality, Privacy, and Ethics 448
Legal Issues 448
Privacy 449
Collecting Information about Individuals 449
Mobile User Privacy 450
Homeland Security and Individual Privacy 450
Recent Technology Issues in Privacy and Analytics 451
Who Owns Our Private Data? 452
Ethics in Decision Making and Support 452
8.6 Impacts of Analytics in Organizations: An Overview 453
New Organizational Units 454
Redesign of an Organization through the Use of Analytics 455
Analytics Impact on Managers' Activities, Performance, and Job Satisfaction 455
Industrial Restructuring 456
Automation's Impact on Jobs 457
Unintended Effects of Analytics 458
8.7 Data Scientist as a Profession 459
Where Do Data Scientists Come From? 459
Chapter Highlights 462
Key Terms 463
Questions for Discussion 463
Exercises 463
References 464
Glossary 467
Index 475