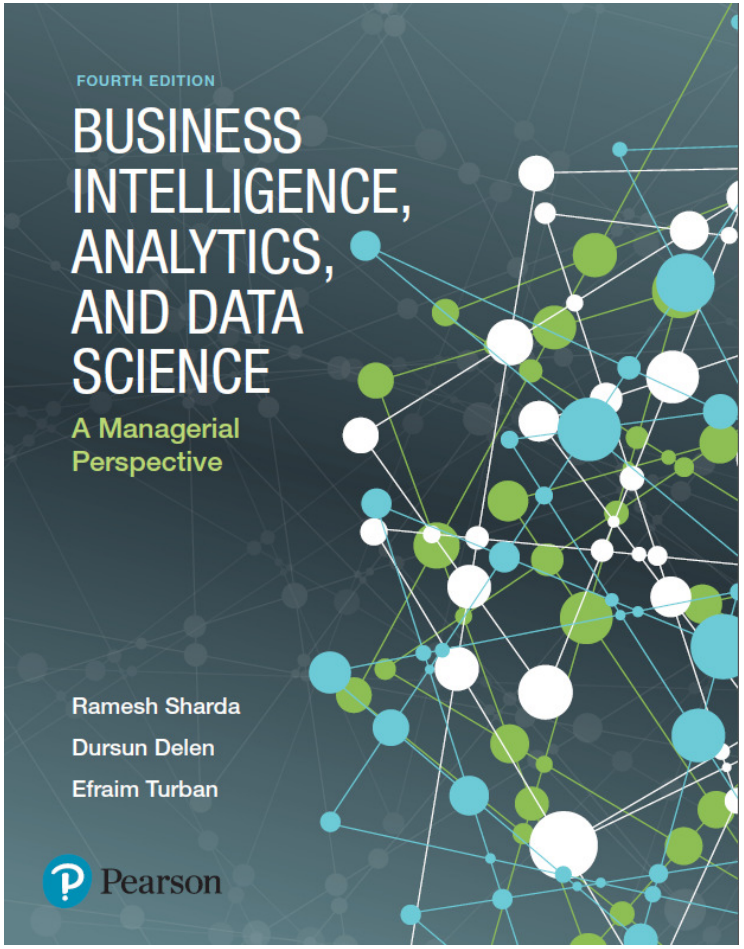
Business Intelligence, Analytics, and Data Science: A Managerial Perspective

Fourth Edition



Chapter 4 – Part A

Predictive Analytics I: Data Mining Process, Methods, and Algorithms

Pearson

Learning Objectives (1 of 2)

4.1 Define data mining as an enabling technology for business analytics

4.2 Understand the objectives and benefits of data mining

4.3 Become familiar with the wide range of applications of data mining

4.4 Learn the standardized data mining processes

4.5 Learn different methods and algorithms of data mining

Learning Objectives (2 of 2)

4.6 Build awareness of the existing data mining software tools

4.7 Understand the privacy issues, pitfalls, and myths of data mining

OPENING VIGNETTE Miami-Dade Police Department Is Using Predictive Analytics to Foresee and Fight Crime (1 of 3)

- Predictive analytics in law enforcement
  - Policing with less
  - New thinking on cold cases
  - The big picture starts small
  - Success brings credibility
  - Just for the facts
  - Safer streets for smarter cities

OPENING VIGNETTE Miami-Dade Police Department Is Using Predictive Analytics to Foresee and Fight Crime (2 of 3)

Discussion Questions

1. Why do law enforcement agencies and departments like Miami-Dade Police Department embrace advanced analytics and data mining?
2. What are the top challenges for law enforcement agencies and departments like Miami-Dade Police Department? Can you think of other challenges (not mentioned in this case) that can benefit from data mining?

OPENING VIGNETTE Miami-Dade Police Department Is Using Predictive Analytics to Foresee and Fight Crime (3 of 3)

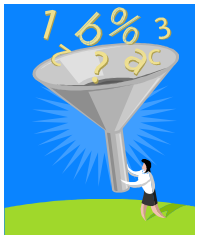
Discussion Questions (continued)

3. What are the sources of data that law enforcement agencies and departments like Miami-Dade Police Department use for their predictive modeling and data mining projects?
4. What type of analytics do law enforcement agencies and departments like Miami-Dade Police Department use to fight crime?
5. What does “the big picture starts small” mean in this case? Explain.

Data Mining Concepts and Definitions
Why Data Mining?

- More intense competition at the global scale.
- Recognition of the value in data sources.
- Availability of quality data on customers, vendors, transactions, Web, etc.
- Consolidation and integration of data repositories into data warehouses.
- The exponential increase in data processing and storage capabilities; and decrease in cost.
- Movement toward conversion of information resources into nonphysical form.

Definition of Data Mining



- The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data stored in structured databases. -- *Fayyad et al., (1996)*
- Keywords in this definition: Process, nontrivial, valid, novel, potentially useful, understandable.
- Data mining: a misnomer?
- Other names: knowledge extraction, pattern analysis, knowledge discovery, information harvesting, pattern searching, data dredging,…

Data Mining Is a Blend of Multiple Disciplines

- FIGURE 4.1 Data Mining Is a Blend of Multiple Disciplines

Application Case 4.1

Visa Is Enhancing the Customer Experience While Reducing Fraud with Predictive Analytics and Data Mining

Questions for Discussion

1. What challenges were Visa and the rest of the credit card industry facing?
2. How did Visa improve customer service while also improving retention of fraud?
3. What is in-memory analytics, and why was it necessary?

Data Mining Characteristics & Objectives

- Source of data for DM is often a consolidated data warehouse (not always!).
- DM environment is usually a client-server or a Web-based information systems architecture.
- Data is the most critical ingredient for DM which may include soft/unstructured data.
- The miner is often an end user.
- Striking it rich requires creative thinking.
- Data mining tools' capabilities and ease of use are essential (Web, Parallel processing, etc.).

How Data Mining Works

- DM extract patterns from data
  - Pattern? A mathematical (numeric and/or symbolic) relationship among data items
- Types of patterns
  - Association
  - Prediction
  - Cluster (segmentation)
  - Sequential (or time series) relationships

Application Case 4.2

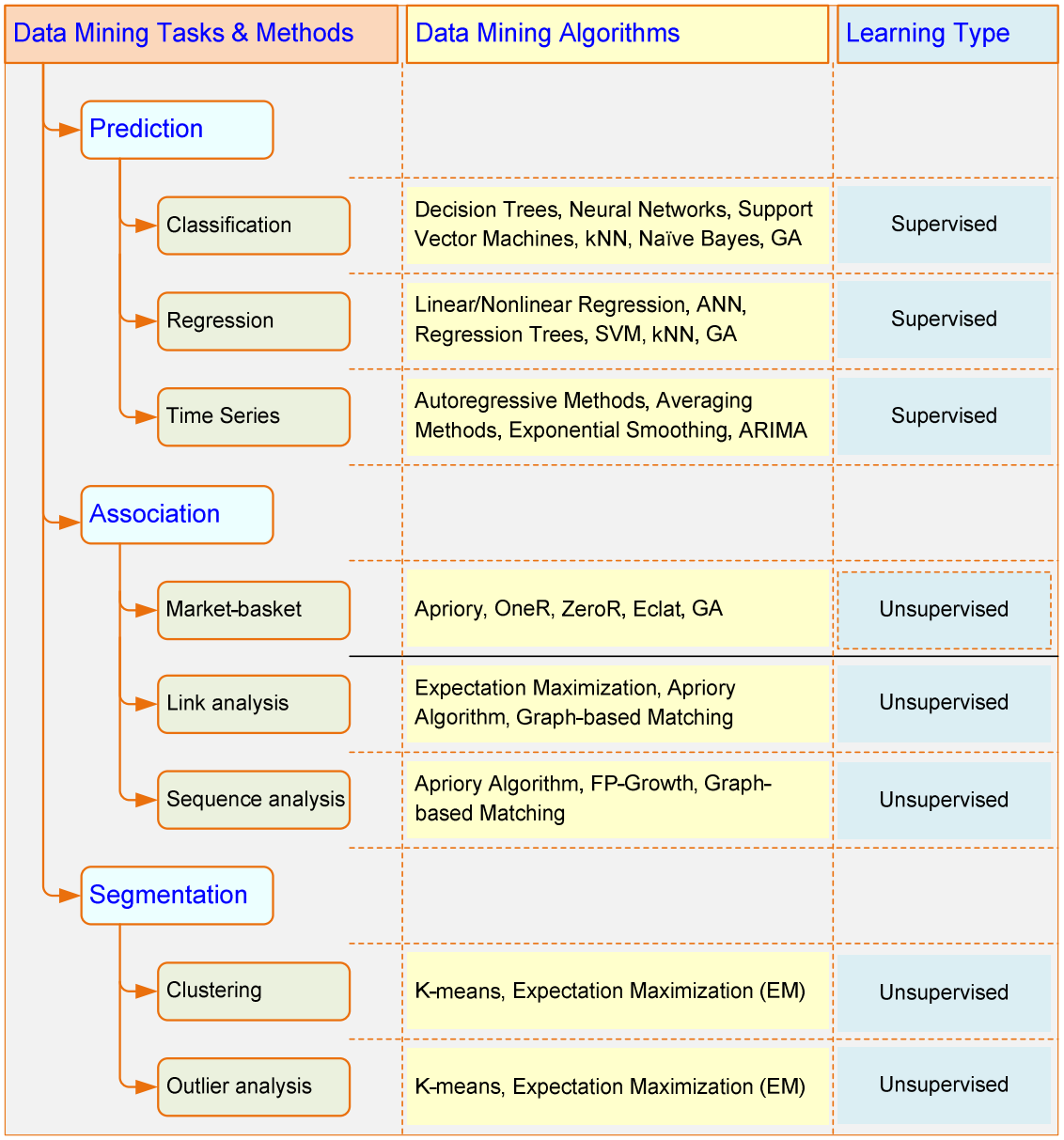
Dell Is Staying Agile and Effective with Analytics in the 21st Century

Questions for Discussion

1. What was the challenge Dell was facing that led to their analytics journey?
2. What solution did Dell develop and implement? What were the results?
3. As an analytics company itself, Dell has used its service offerings for its own business. Do you think it is easier or harder for a company to taste its own medicine? Explain.

A Taxonomy for Data Mining

- FIGURE 4.2 A Simple Taxonomy for Data Mining Tasks, Methods, and Algorithms



Other Data Mining Patterns/Tasks

- Time-series forecasting
  - Part of the sequence or link analysis?
- Visualization
  - Another data mining task?
  - Covered in Chapter 3
- Data Mining versus Statistics
  - Are they the same?
  - What is the relationship between the two?