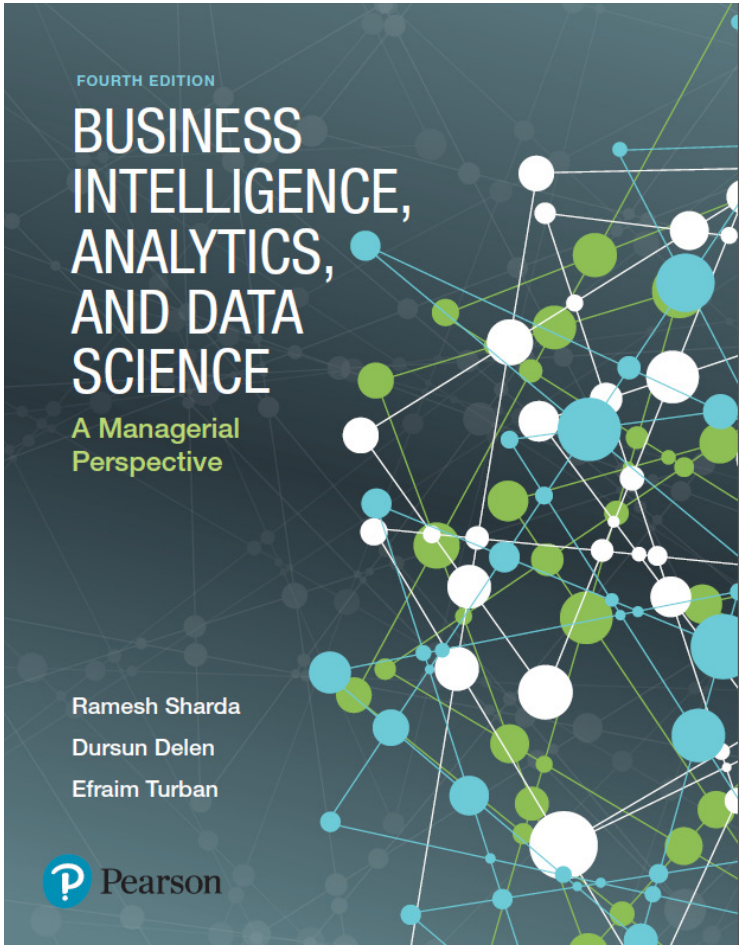
Business Intelligence, Analytics, and Data Science: A Managerial Perspective

Fourth Edition



Chapter 4 – Part B

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

Pearson

Data Mining Applications (1 of 4)

- Customer Relationship Management
  - Maximize return on marketing campaigns
  - Improve customer retention (churn analysis)
  - Maximize customer value (cross-, up-selling)
  - Identify and treat most valued customers
- Banking & Other Financial
  - Automate the loan application process
  - Detecting fraudulent transactions
  - Maximize customer value (cross-, up-selling)
  - Optimizing cash reserves with forecasting

Data Mining Applications (2 of 4)

- Retailing and Logistics
  - Optimize inventory levels at different locations
  - Improve the store layout and sales promotions
  - Optimize logistics by predicting seasonal effects
  - Minimize losses due to limited shelf life
- Manufacturing and Maintenance
  - Predict/prevent machinery failures
  - Identify anomalies in production systems to optimize the use manufacturing capacity
  - Discover novel patterns to improve product quality

Data Mining Applications (3 of 4)

- Brokerage and Securities Trading
  - Predict changes on certain bond prices
  - Forecast the direction of stock fluctuations
  - Assess the effect of events on market movements
  - Identify and prevent fraudulent activities in trading
- Insurance
  - Forecast claim costs for better business planning
  - Determine optimal rate plans
  - Optimize marketing to specific customers
  - Identify and prevent fraudulent claim activities

Data Mining Applications (4 of 4)

- Computer hardware and software
- Science and engineering
- Government and defense
- Homeland security and law enforcement
- Travel, entertainment, sports
- Healthcare and medicine
- Sports,… virtually everywhere…

Application Case 4.3

Predictive Analytic and Data Mining Help Stop Terrorist Funding

Questions for Discussion

1. How can data mining be used to fight terrorism? Comment on what else can be done beyond what is covered in this short application case.
2. Do you think data mining, although essential for fighting terrorist cells, also jeopardizes individuals' rights of privacy?

Data Mining Process

- A manifestation of the best practices
- A systematic way to conduct DM projects
- Moving from Art to Science for DM project
- Everybody has a different version
- Most common standard processes:
  - CRISP-DM (Cross-Industry Standard Process for Data Mining)
  - SEMMA (Sample, Explore, Modify, Model, and Assess)
  - KDD (Knowledge Discovery in Databases)

Data Mining Process: CRISP-DM (1 of 2)

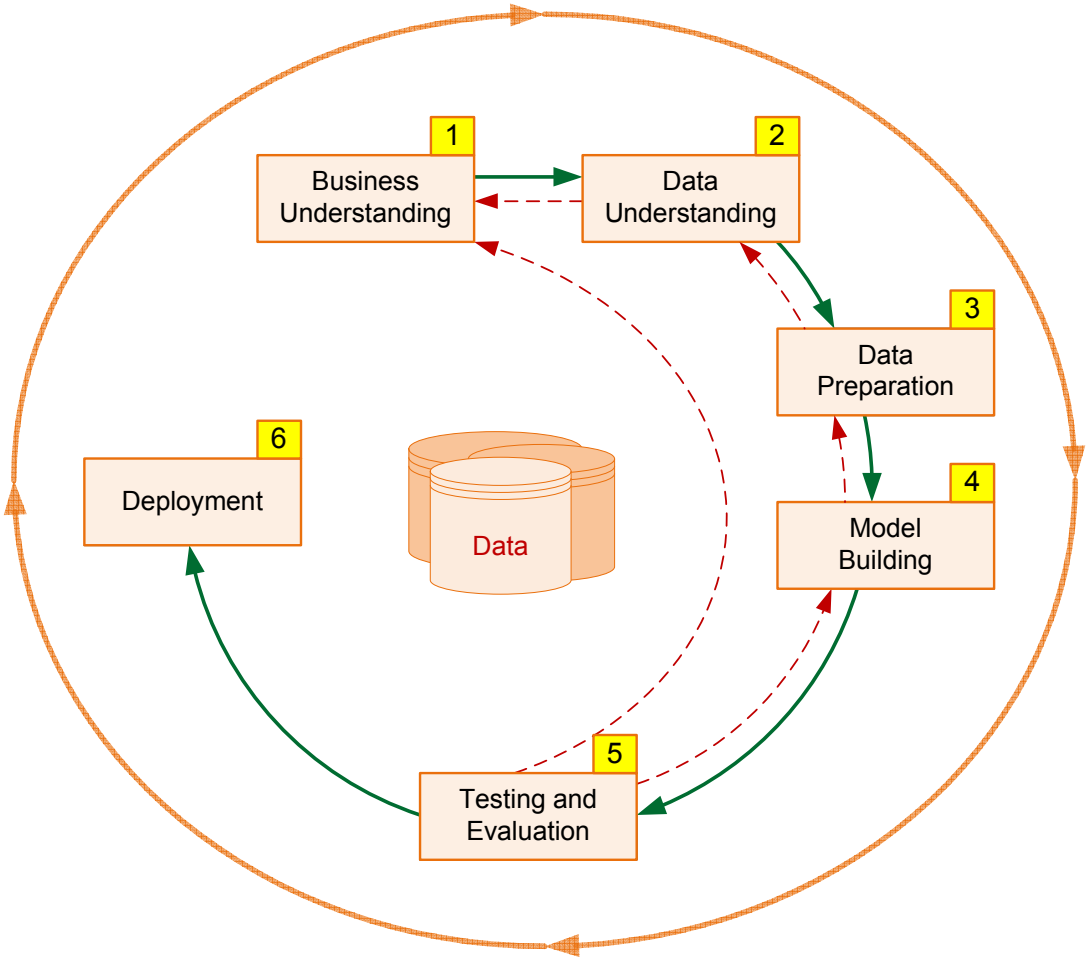
- Cross Industry Standard Process for Data Mining
- Proposed in 1990s by a European consortium
- Composed of six consecutive phases
  - Step 1: Business Understanding
  - Step 2: Data Understanding
  - Step 3: Data Preparation
  - Step 4: Model Building
  - Step 5: Testing and Evaluation
  - Step 6: Deployment

Accounts for ~85% of total project time

Data Mining Process: SEMMA

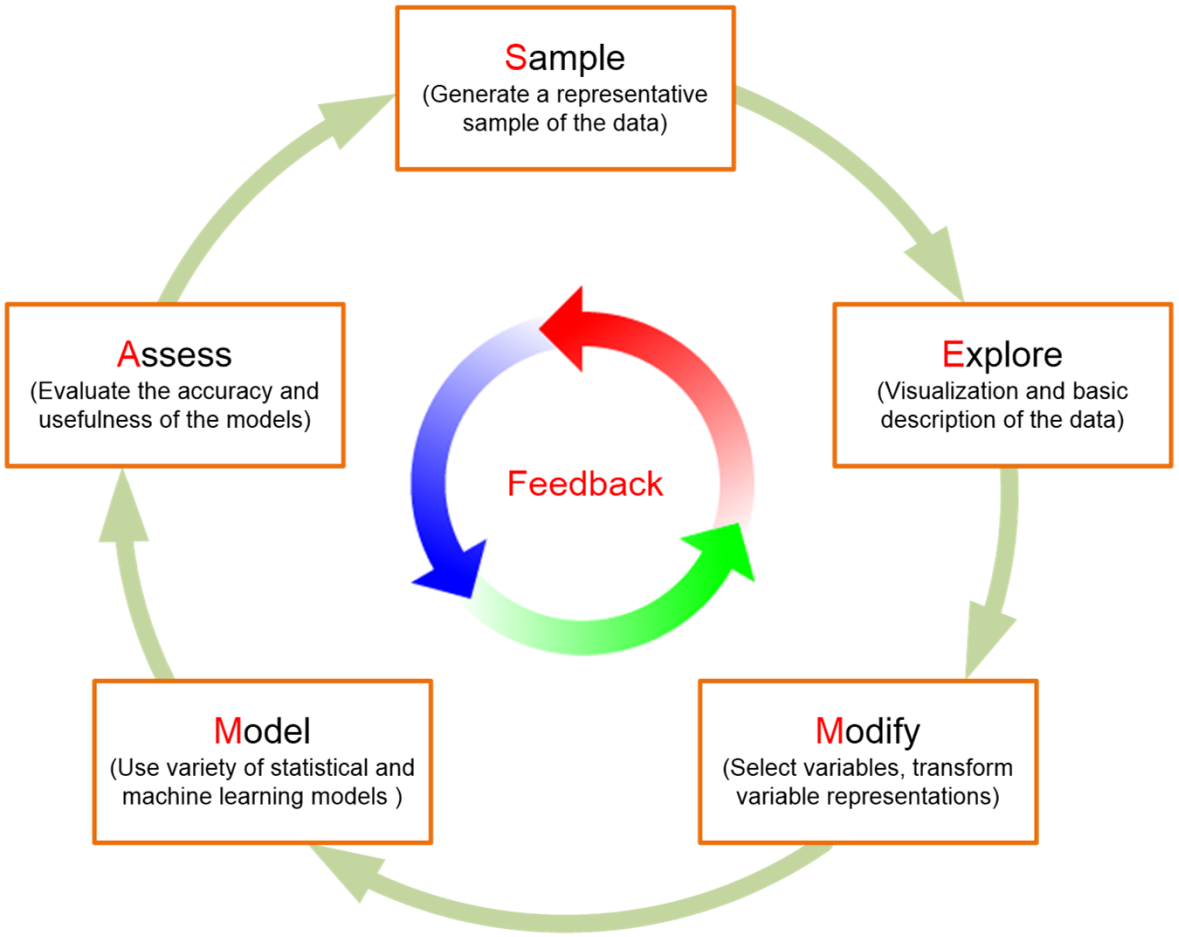
- FIGURE 4.3 The Six-Step CRISP-DM Data Mining Process →

- The process is highly repetitive and experimental (DM: art versus science?)



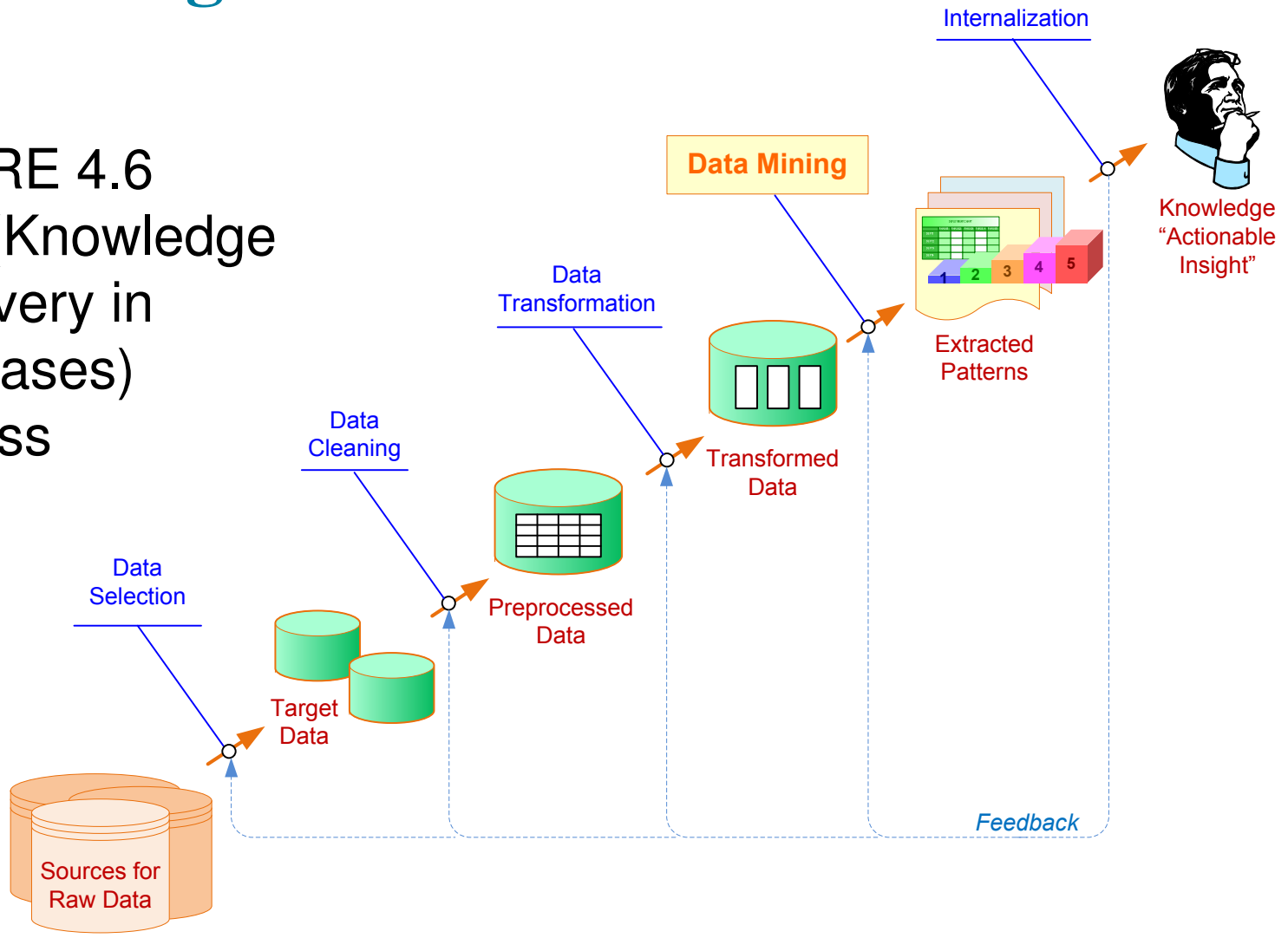
Data Mining Process: CRISP-DM (2 of 2)

- FIGURE 4.5 SEMMA Data Mining Process
- Developed by SAS Institute



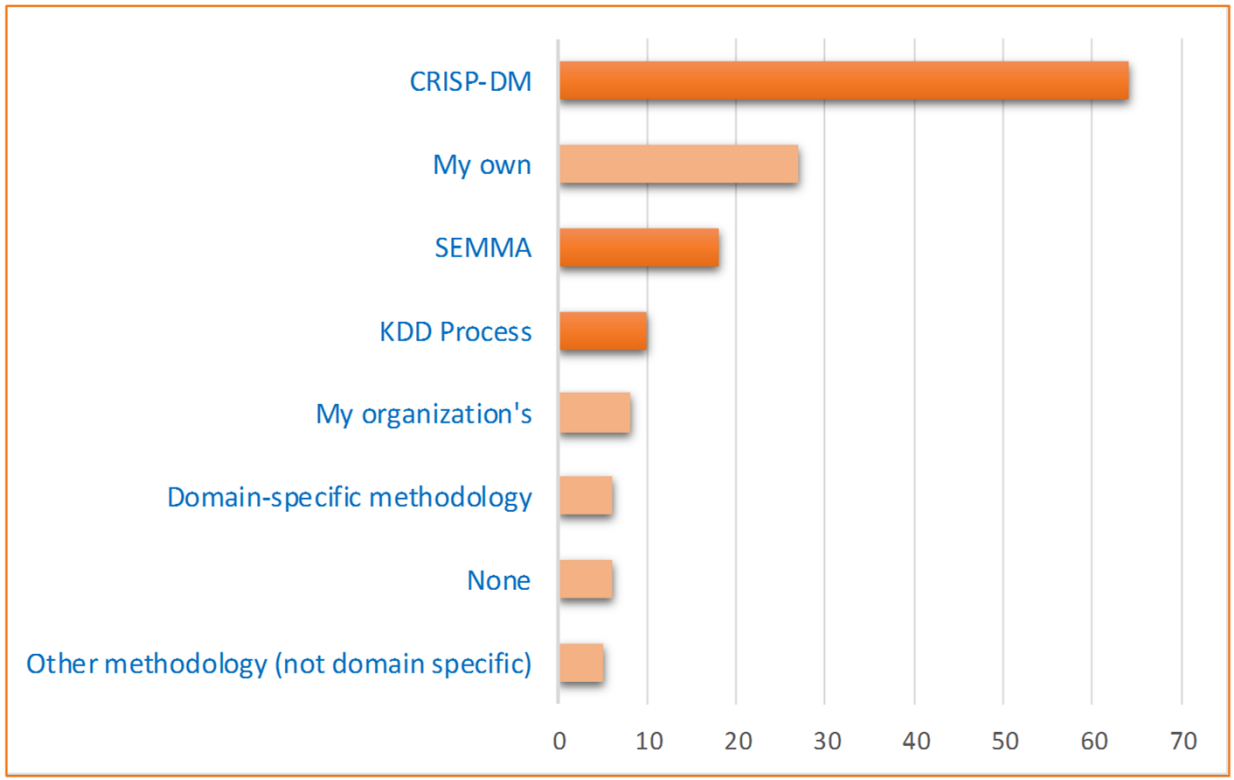
Data Mining Process: KDD

- FIGURE 4.6 KDD (Knowledge Discovery in Databases) Process



Which Data Mining Process is the Best?

- FIGURE 4.7
  Ranking of Data Mining Methodologies/Processes.
  Source: Used with permission from KDnuggets.com.



Application Case 4.4
Data Mining Helps in Cancer Research

Questions for Discussion

1. How can data mining be used for ultimately curing illnesses like cancer?
2. What do you think are the promises and major challenges for data miners in contributing to medical and biological research endeavors?