NOVA IMS
Information Management School

2

INTRODUCTION TO CRISP-DM

Machine Learning for Marketing

Acreditações e Certificações

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa

Why use a standard methodological process

- Framework to record and replicate projects
- Assists project planning and management
- Encourage best practices and the obtention of better results
- Provides a base for new practitioners:
  - Demonstrates the maturity of DM
  - Reduces dependency of ”experts”

CRISP-DM

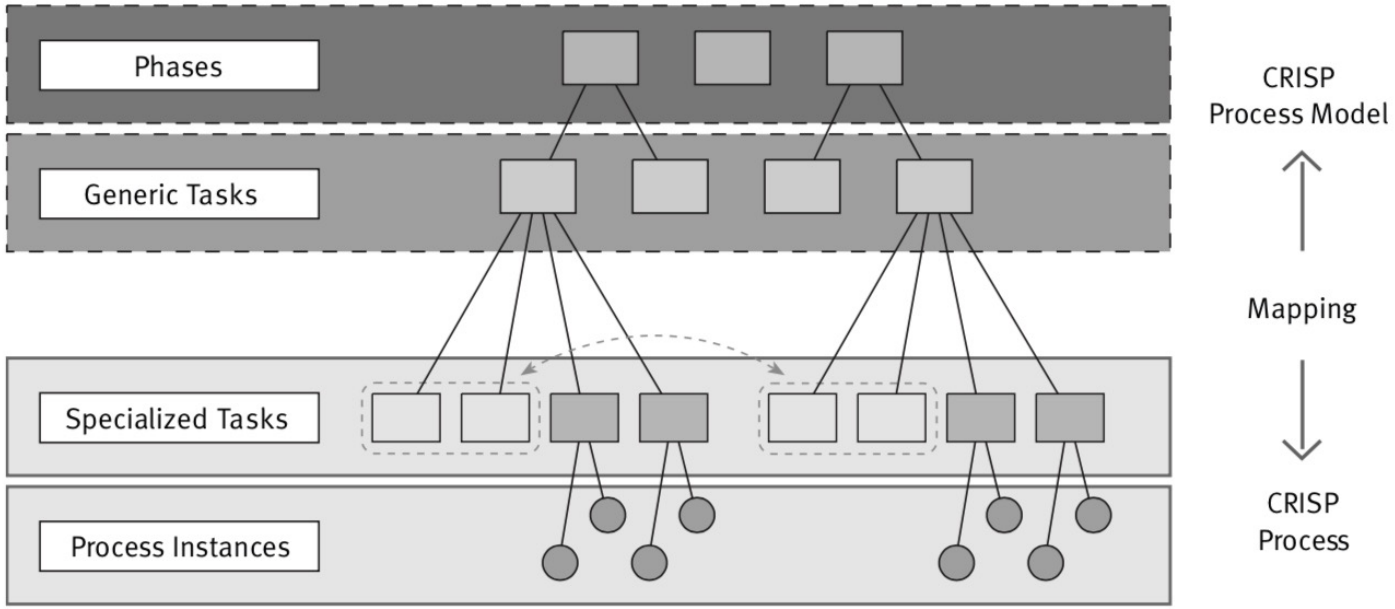
- Applies not only to DM projects, but also to Text Mining, Statistic, and Descriptive and Predictive Analytics
- Used in academy and by DM practitioners
- Non-proprietary
- Tool neutral
- Focus both on the application and the technical perspectives
- Most-often used process model

CRISP-DM 1.0

Step-by-step data mining guide

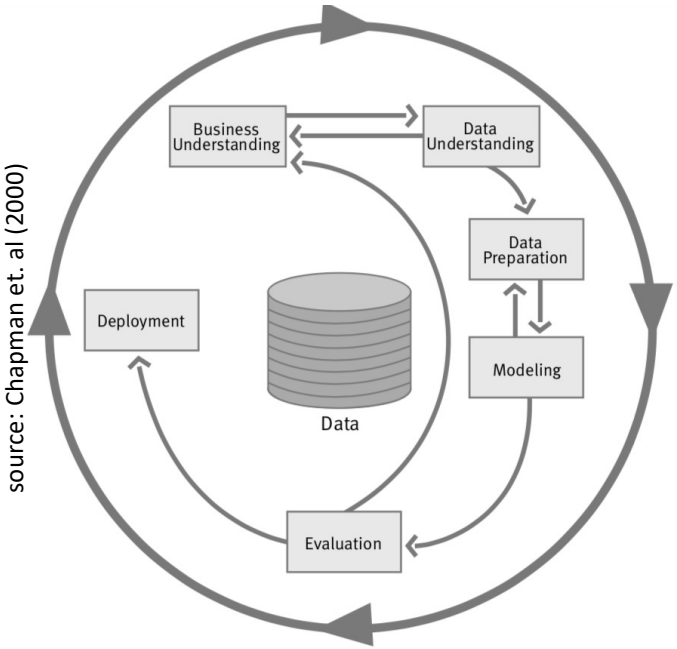
Pete Chapman (NCR), Julian Clinton (SPSS), Randy Kerber (NCR),
Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler),
Colin Shearer (SPSS) and Rüdiger Wirth (DaimlerChrysler)

Four level breakdown



source: Chapman et. al (2000)

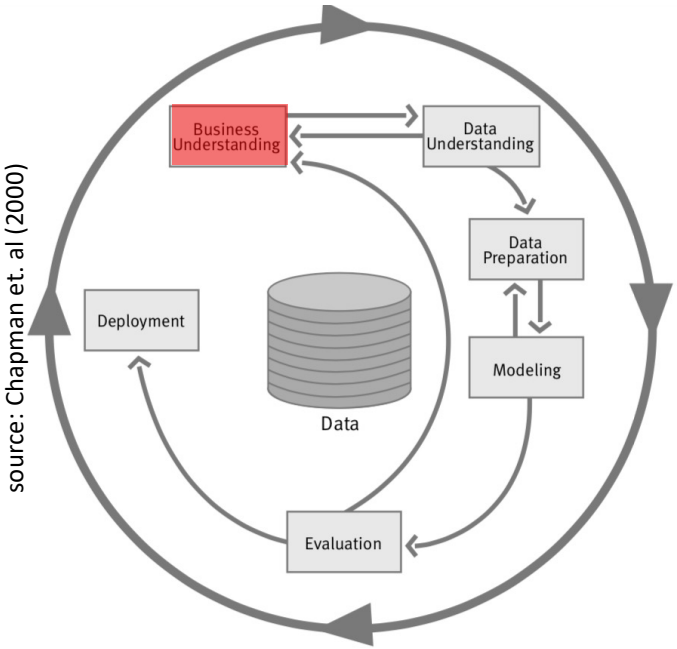
Phases



source: Chapman et. al (2000)

- Six phases
- Phase's sequence is not rigid
- Moving back and forward is always necessary
- The outcome of each phase, determines the next phase/task
- Arrows indicate the most important and frequent dependencies between phases

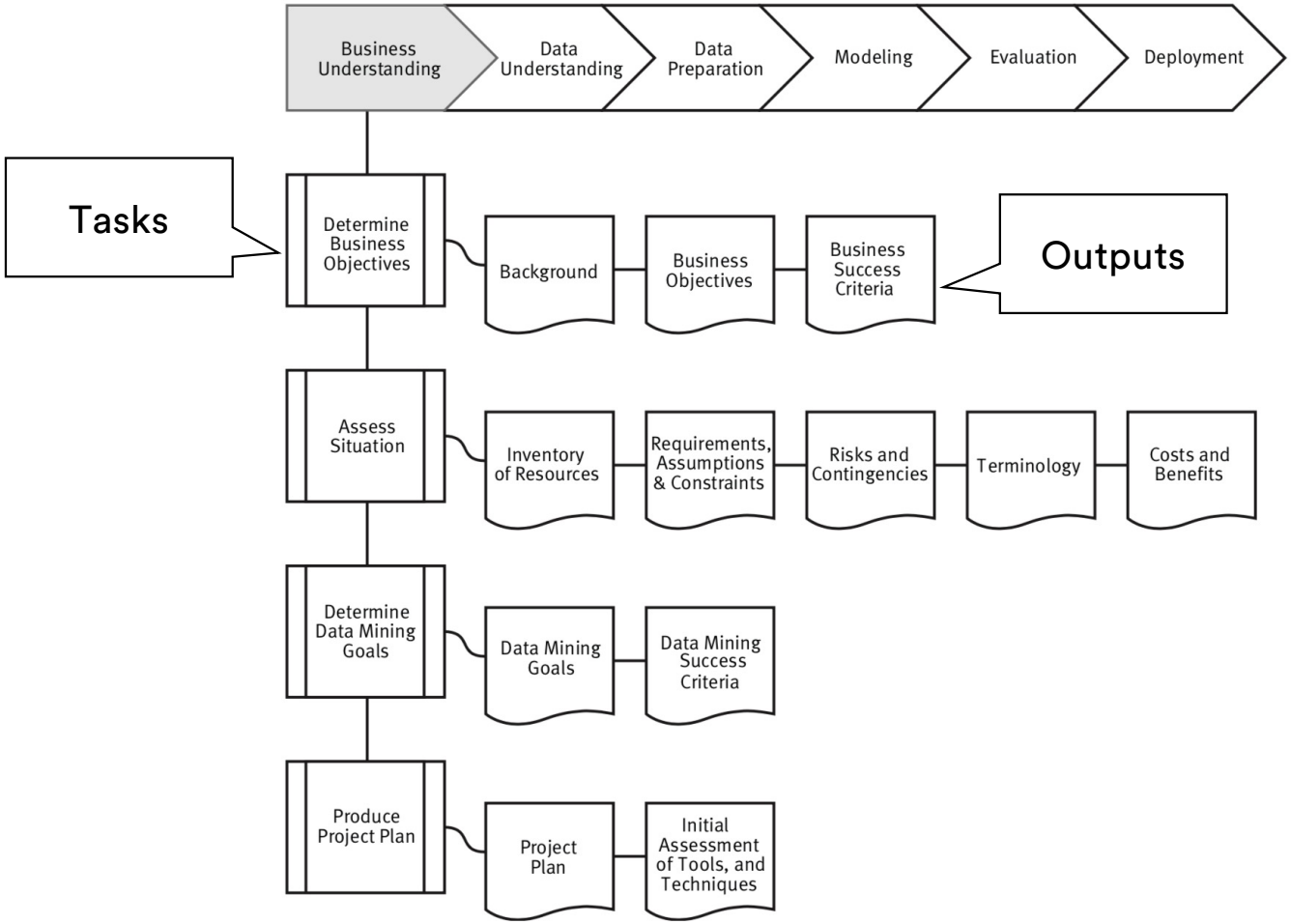
Phase: Business understanding



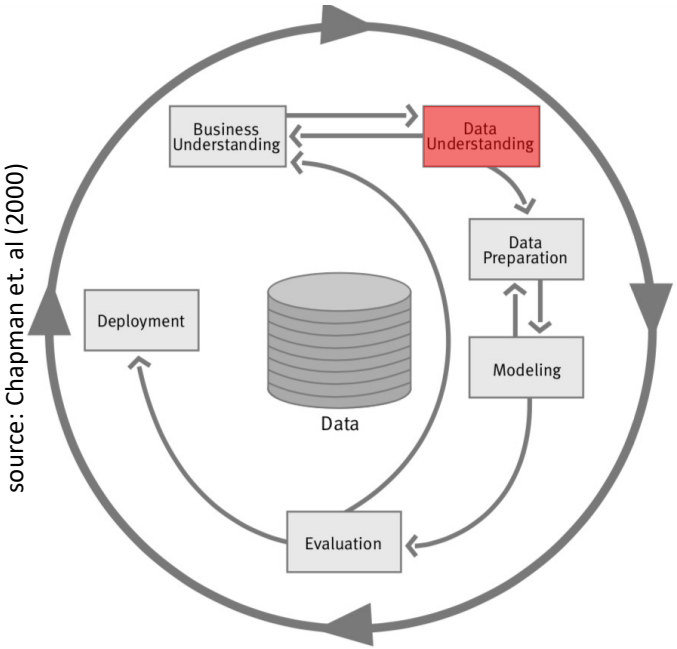
source: Chapman et. al (2000)

- Determine business objectives
  - Background
  - Business objectives
  - Business success criteria
- Assess situation
  - Resources
  - Requirements
  - Risks and contingencies
- Determine data mining goals and success criteria
- Produce project plan

Business understanding



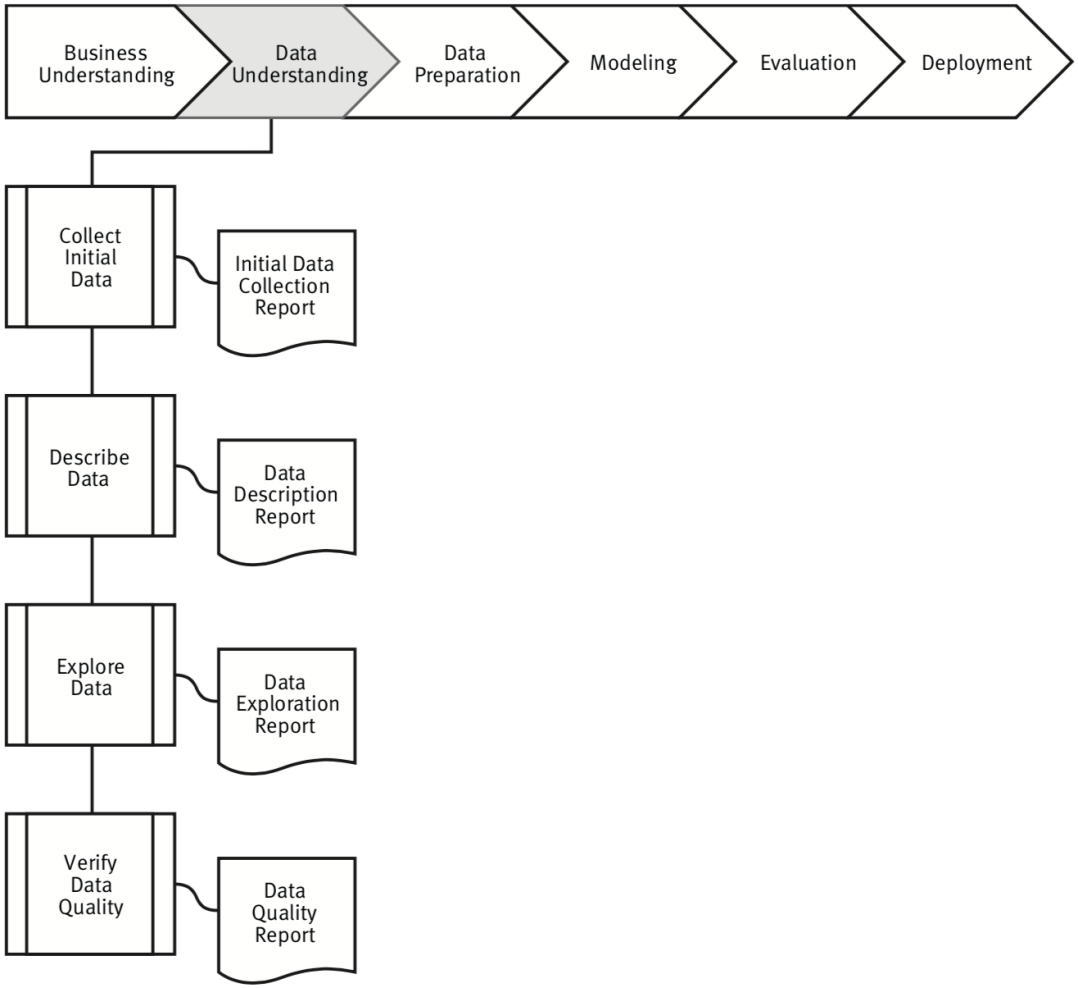
Phase: Data understanding



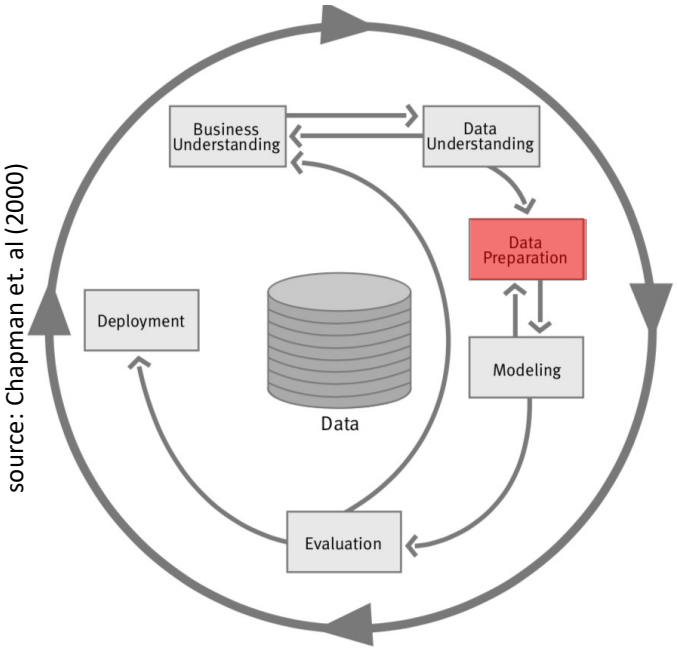
source: Chapman et. al (2000)

- Collect initial data
- Describe data
- Explore data
- Verify data quality

Data understanding



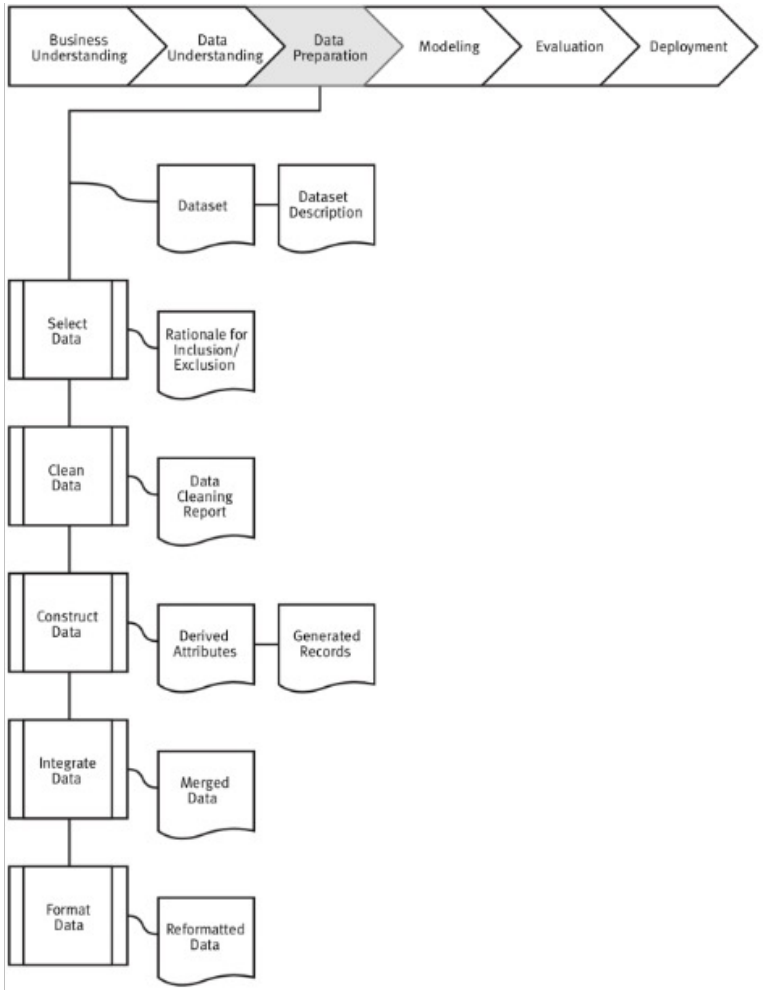
Phase: Data preparation



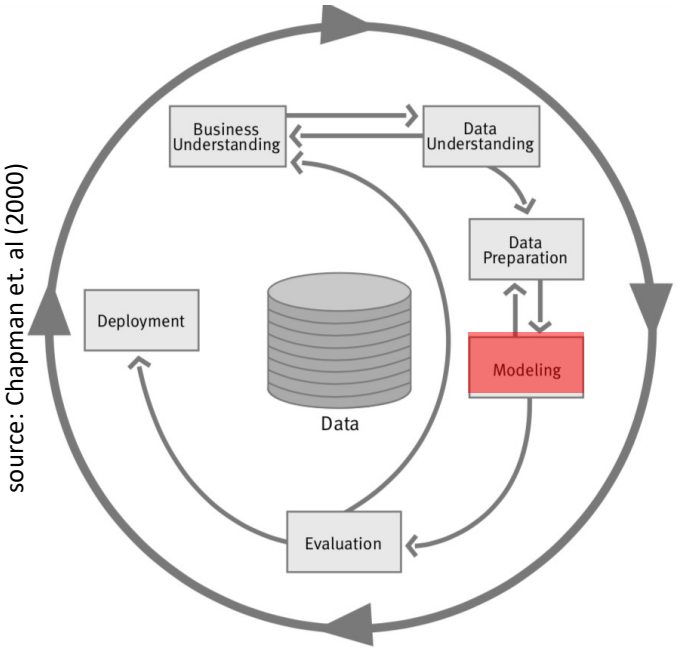
source: Chapman et. al (2000)

- Select data
- Clean data
- Construct data
- Integrate/merge data
- Format data

Data preparation



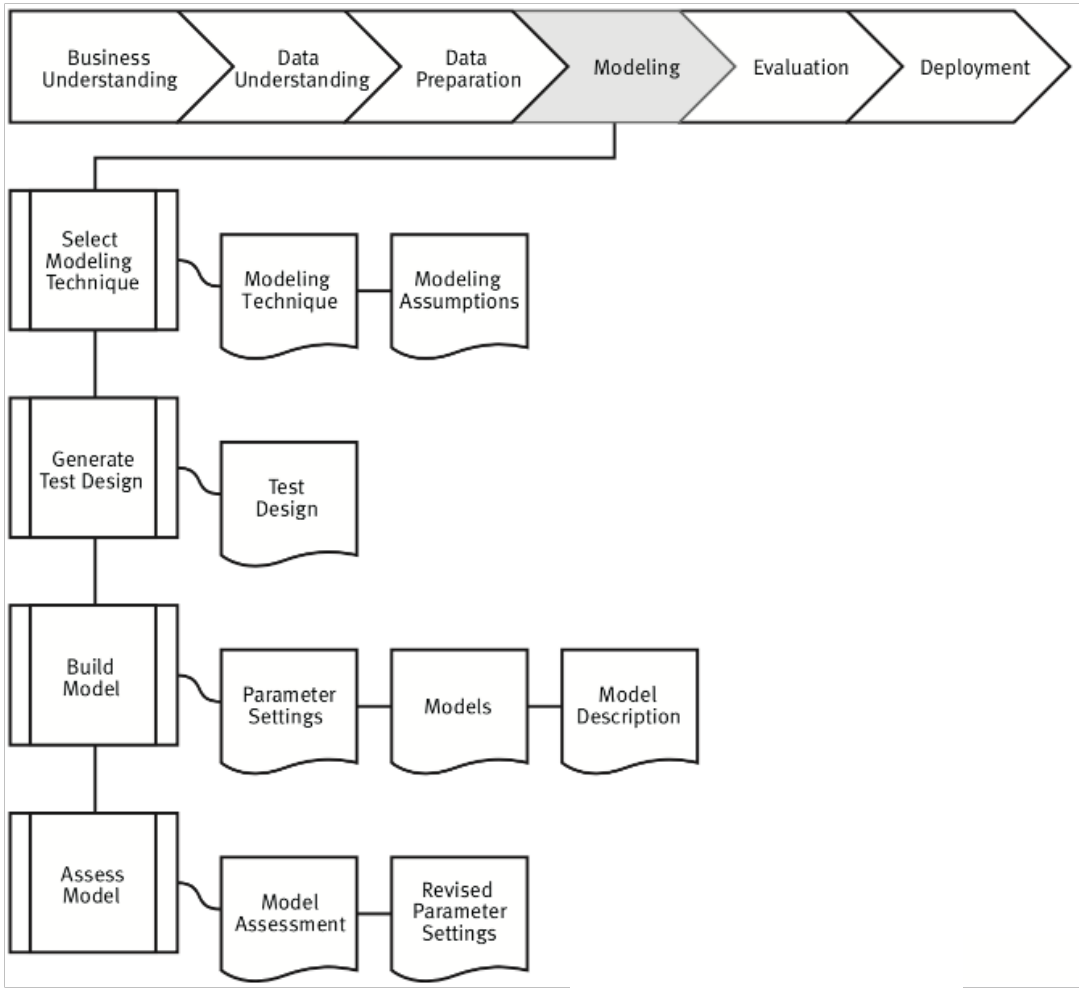
Phase: Modeling



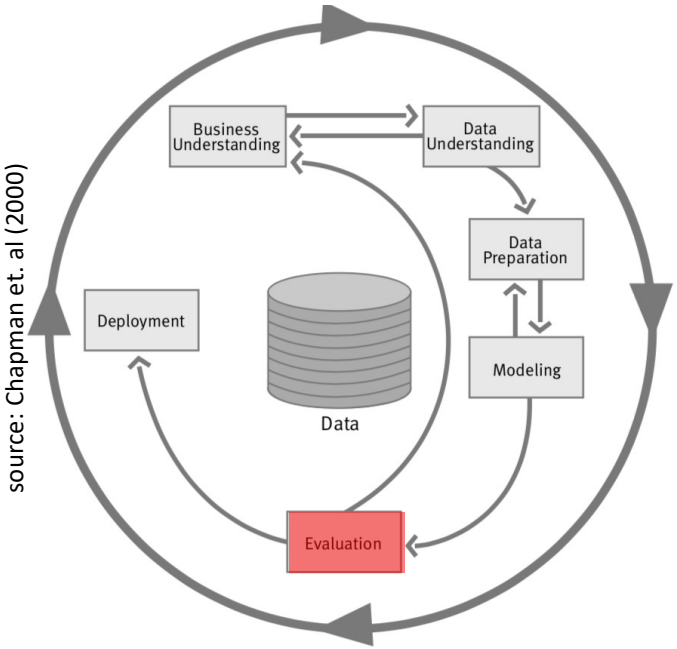
source: Chapman et. al (2000)

- Select modeling techniques
  - Algorithm selection
  - Modeling assumptions
- Generate test design
- Build model
- Assess model

Modeling


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Select Modeling Technique
Modeling Technique
Modeling Assumptions
Generate Test Design
Test Design
Build Model
Parameter Settings
Models
Model Description
Assess Model
Model Assessment
Revised Parameter Settings

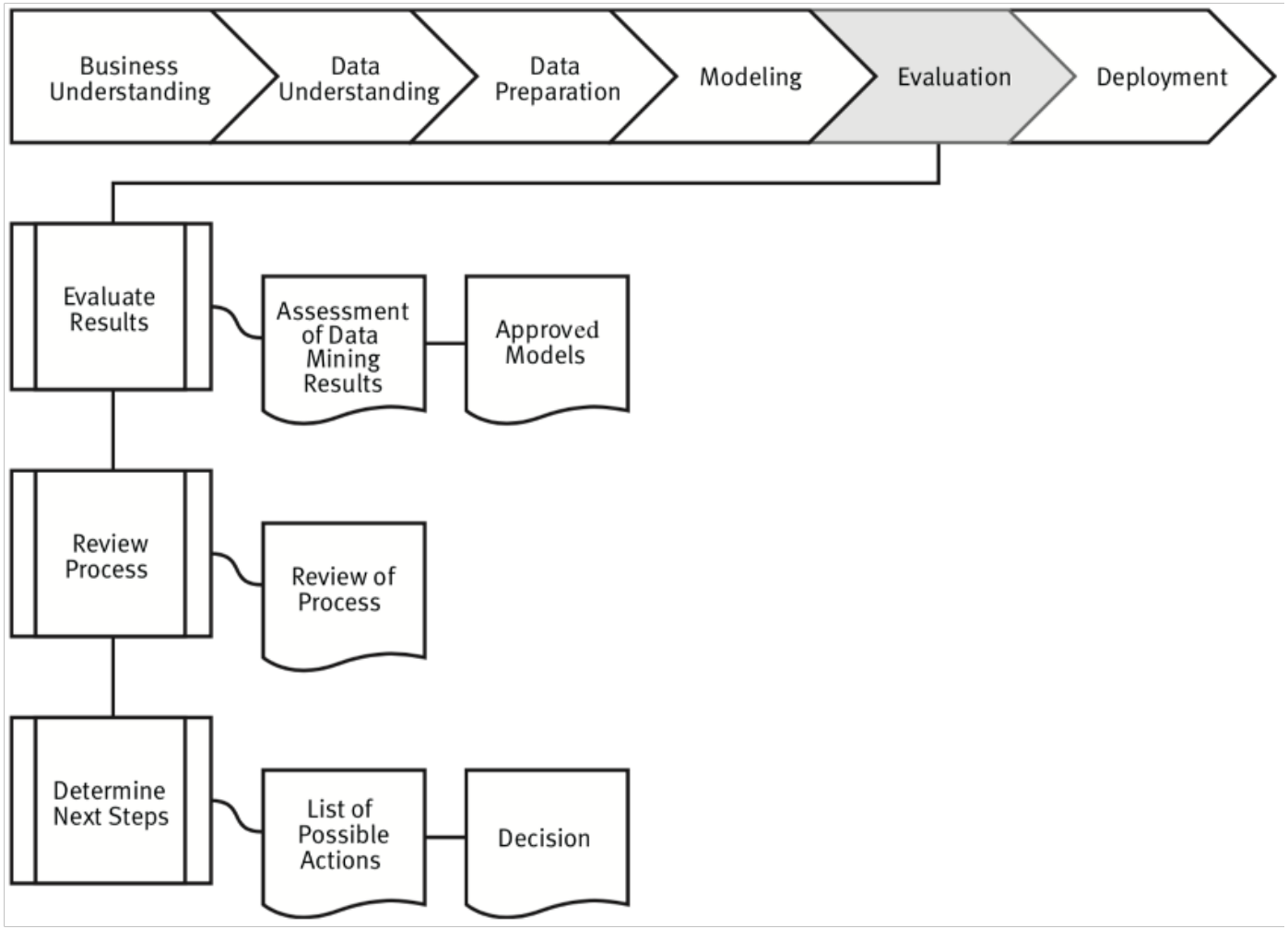
Phase: Evaluation



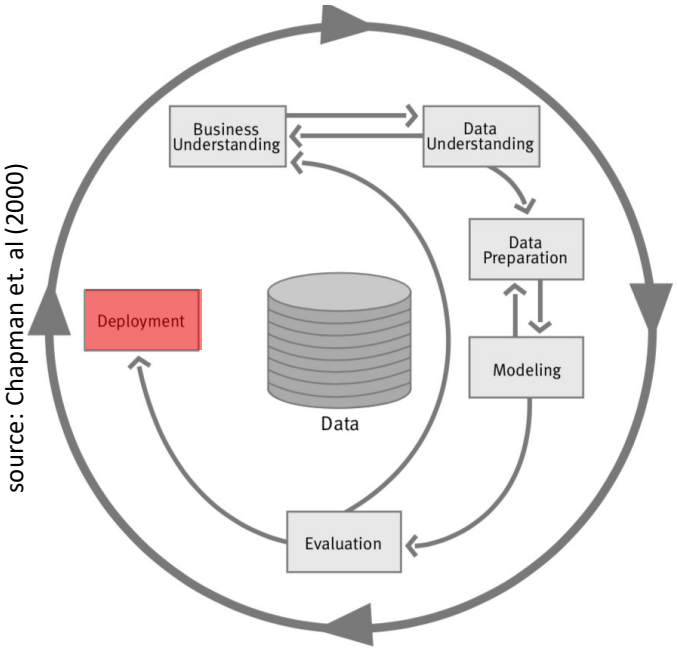
source: Chapman et. al (2000)

- Evaluate results
  - Assess data mining results vs business success criteria
  - Approve model
- Review process
- Determine next steps
  - Production or not?
  - Additional requirements?

Evaluation


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Evaluate Results
Assessment of Data Mining Results
Approved Models
Review Process
Review of Process
Determine Next Steps
List of Possible Actions
Decision

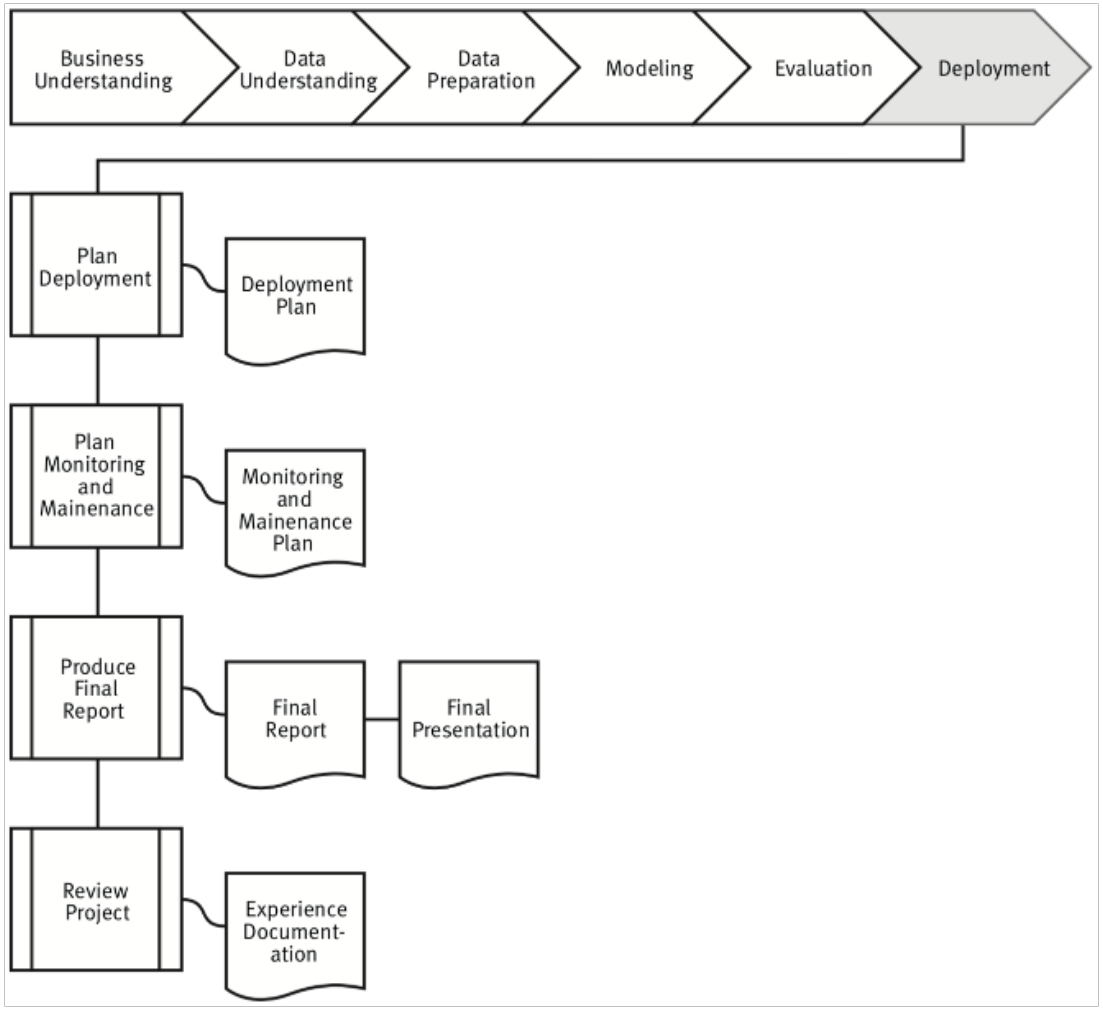
Phase: Deployment



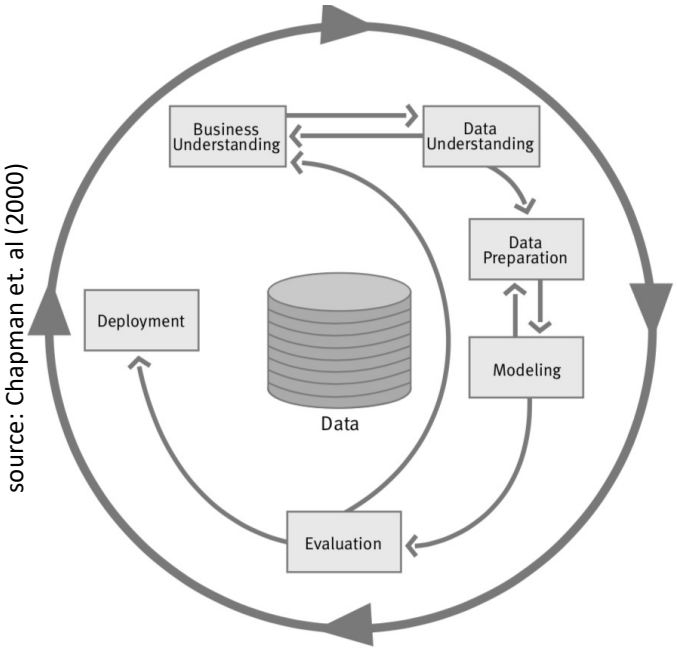
source: Chapman et. al (2000)

- Plan deployment
  - Strategy to deploy the model, including integration in business processes
- Plan monitoring and maintenance
  - Performance assessment
  - Models' update
- Produce final report
- Review project

Deployment


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Plan Deployment
Deployment Plan
Plan Monitoring and Mainenance
Monitoring and Mainenance Plan
Produce Final Report
Final Report
Final Presentation
Review Project
Experience Documentation

Cyclical nature



source: Chapman et. al (2000)

- The model is not “closed” once a project is deployed!!!
- Lessons learned during the project development and from the deployed project can trigger more-focused business questions

Questions?

Machine Learning for Marketing

Acreditações e Certificações

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa