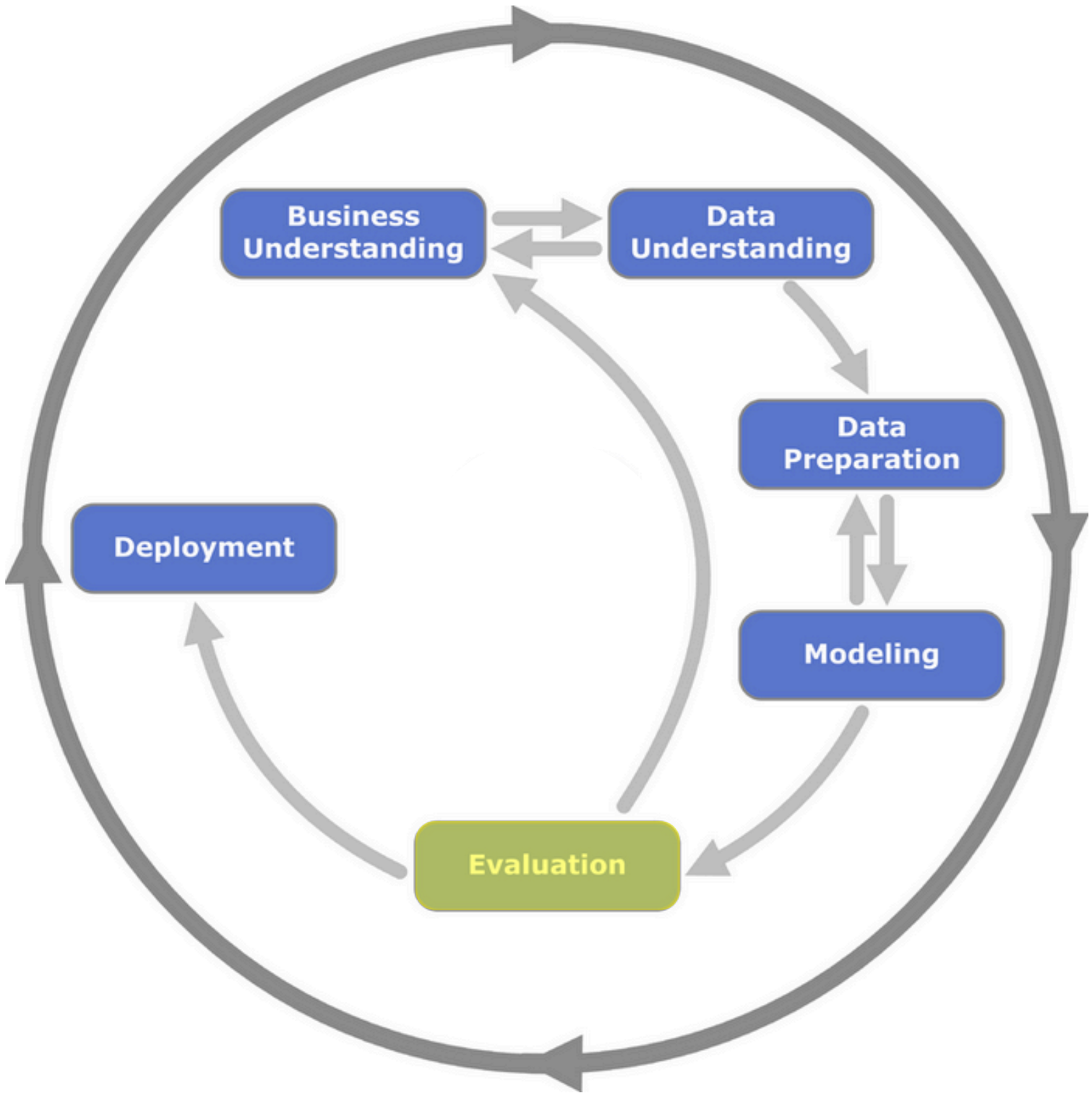
CS490DSC Data Science Capstone Evaluation

Jean Honorio
Purdue University

Important

- Please read this together with the case study
- The case study will discuss a fictitious health insurance company called the Amazing Health Network

CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

Phase 5: Evaluation

- In the previous phase, you have built a model (or models) that appears to have high quality from a data analysis perspective (e.g., accuracy, generalization)
- **In this phase, we thoroughly evaluate the model and review the steps executed to create it, to make sure the model achieves the business objectives**
- A key objective is to determine if there is some important business issue that has not been sufficiently considered
- At the end of this phase, a decision on the use of the data mining results should be reached

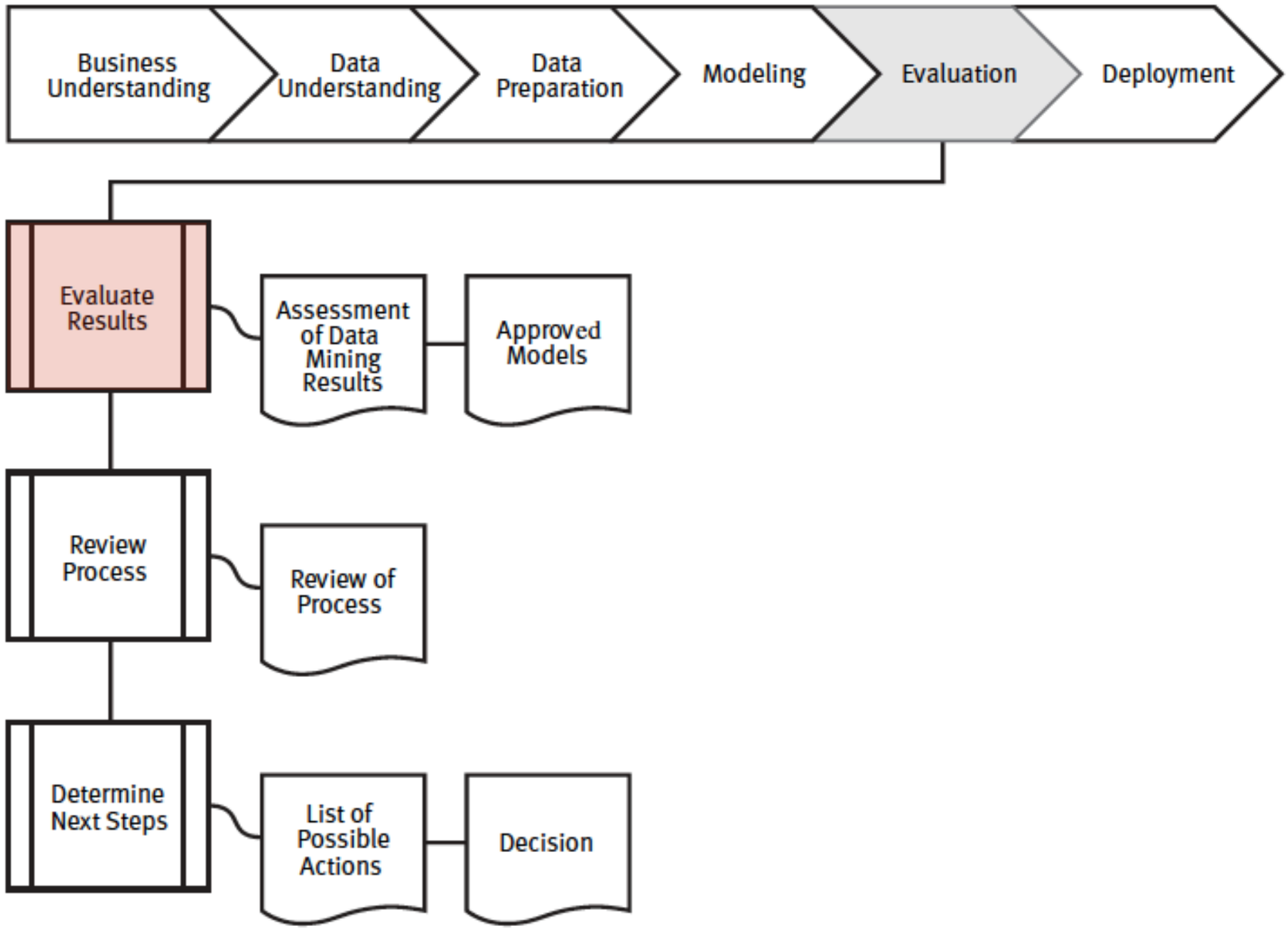
Phase 5: Evaluation

- A good way of defining the total outputs of a data mining project is to use the equation:

RESULTS = MODELS + FINDINGS

- In this equation, we are defining that the total output of the data mining project is
  - not just the models (although they are, of course, important) but also the findings
  - findings = anything (apart from the model) that is important in meeting the objectives of the business or important in leading to new questions, lines of approach, or side effects (e.g., data quality problems uncovered by the data mining exercise)

Phase 5: Evaluation



1. Evaluate results

- Previous evaluation steps dealt with factors such as the accuracy and generality of the model
- This step assesses the degree to which the model meets the business objectives and seeks to determine if there is some business reason why this model is deficient
- Another option is to test the model(s) on test applications in the real application, if time and budget constraints permit
- Moreover, evaluation also assesses other data mining results generated
  - Data mining results involve models that are necessarily related to the original business objectives and all other findings that are not necessarily related to the original business objectives, but might also unveil additional challenges, information, or hints for future directions

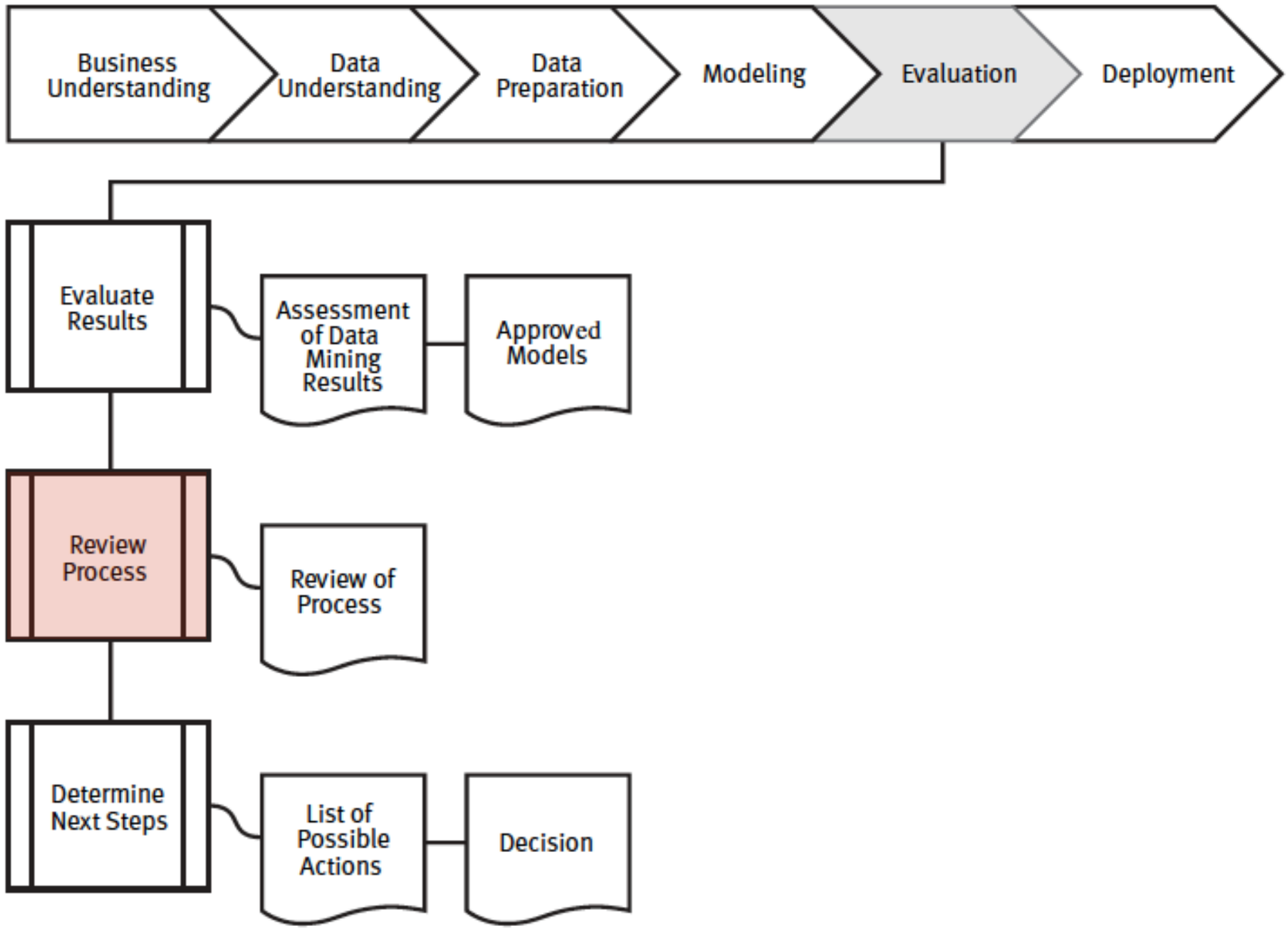
1.1. Assessment of data mining results with respect to business success criteria

- Summarize assessment results in terms of business success criteria, including a final statement regarding whether the project already meets the initial business objectives

1.2. Approved models

- After assessing models with respect to business success criteria
- The generated models that meet the selected criteria become the approved models

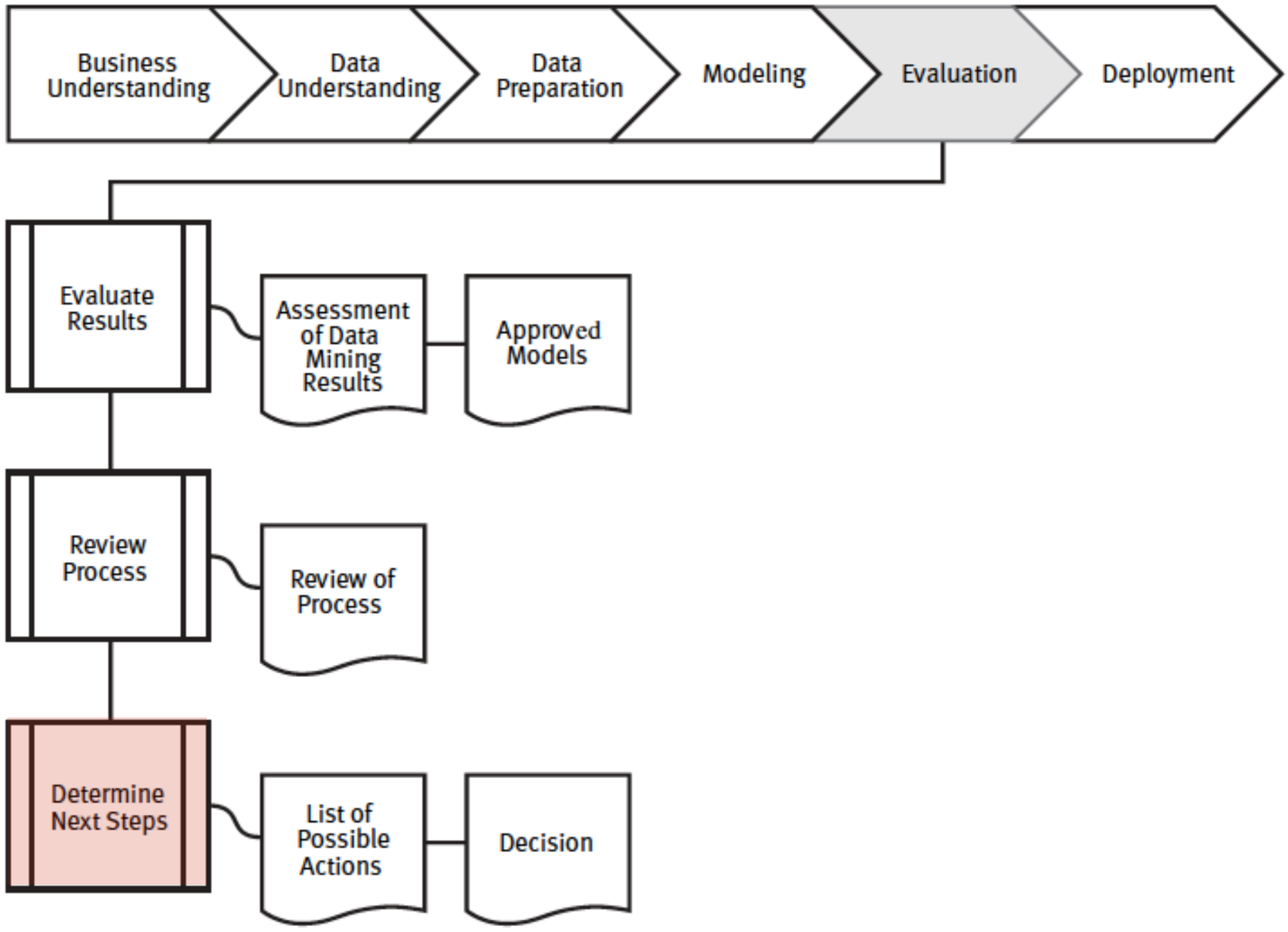
Phase 5: Evaluation



2. Review process

- At this point, the resulting models appear to be satisfactory and to satisfy business needs
- It is now appropriate to do a more thorough review of the data mining engagement in order to determine if there is any important factor or task that has somehow been overlooked
- This review also covers quality assurance issues, for example:
  - Did we correctly build the model?
  - Did we use only the attributes that we are allowed to use and that are available for future analyses?
- Summarize the process review and highlight activities that have been missed and those that should be repeated

Phase 5: Evaluation



3. Determine next steps

- Depending on the results of the assessment and the process review, the project team decides how to proceed
- The team decides whether to
  - finish this project and move on to deployment
  - initiate further iterations
  - or set up new data mining projects
- This task includes analyses of remaining resources and budget, which may influence the decisions

3.1. List of possible actions

- List the potential further actions, along with the reasons for and against each option

3.2. Decision

- Describe the decision as to how to proceed, along with the rationale