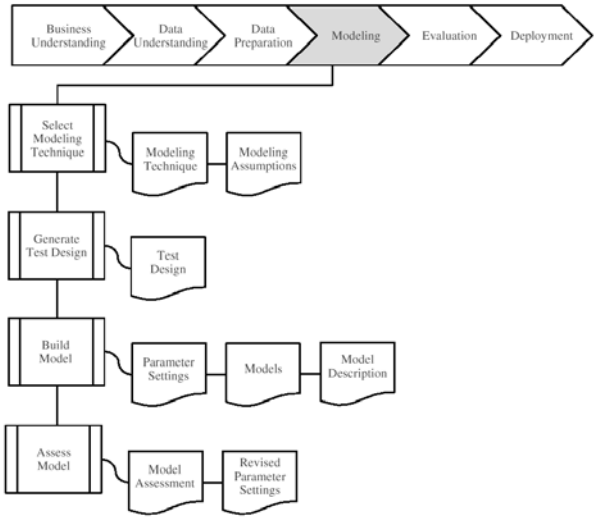
VIII Modeling (Summary)



VIII Modeling (Summary)

VIII.1 Select Modeling Technique

Task
As the first step in modeling, select the actual modeling technique, e.g., decision tree building with C4.5 or neural network generation with back propagation.

Outputs
This output refers to the actual modeling technique which is used.

Modeling Assumptions
Many modeling techniques make specific assumptions on the data, e.g., all attributes have uniform distributions, no missing values allowed, class attribute must be symbolic etc.

VIII Modeling (Summary)

VIII.2 Generate Test Design

Task
Before building a model, we need to generate a procedure or mechanism how to test the model's quality and validity, e.g., error rates for classification tasks.

Output
Description of the intended plan for training, testing, and evaluating the models.

VIII Modeling (Summary)

VIII.3 Build Model

Task
Run the modeling tool on the prepared data set to create one or more models.

Outputs
- Listing of the parameters and their chosen value, along with rationale for the choice of parameter settings.
- Models produced by the modeling tool, not a report.
- Model Description:Description of the resultant model, report on the interpretation of the models and any difficulties encountered with their meanings.

VIII.4 Assess Model

Task

- The data mining engineer **interprets** the models according to his domain knowledge, data mining success criteria, and the desired test design.
- This task **interferes with the subsequent evaluation phase**. Whereas the data mining engineer judges the success of the application of modeling and discovery techniques more technically, he contacts business analysts and domain experts later in order to discuss the data mining results in the business context. This task only considers models whereas the evaluation phase also takes into account all other results which were produced in the course of the project.
- The data mining engineer tries to **rank the results**. He assesses the models according to the evaluation criteria. As far as possible he also takes into account business objectives and business success criteria.
- In most data mining projects, the data mining engineer applies a single technique more than once or generates data mining results with different alternative techniques. In this task, he also **compares all results** according to the evaluation criteria.

VIII.4 Assess Model (cont'd)

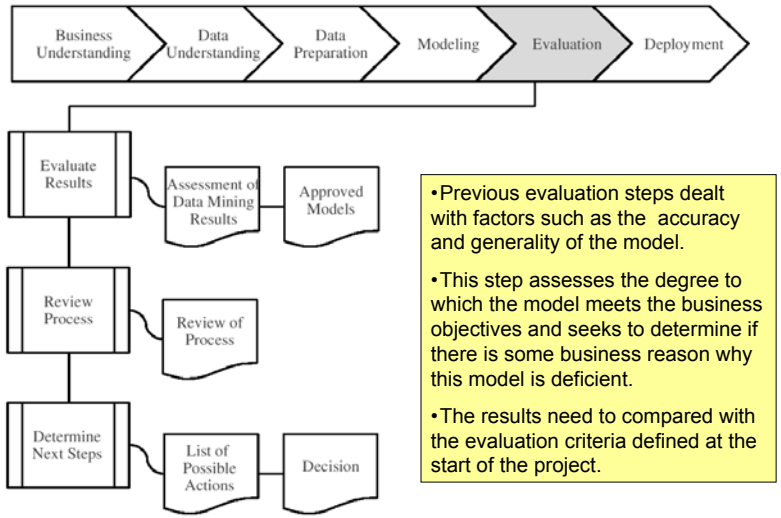
Activities

- Evaluate result w.r.t. evaluation criteria
- Test result according to a test strategy (e.g.: Train & Test, Cross-validation, bootstrapping etc.)
- Compare evaluation results and interpretation
- Create ranking of results w.r.t. success and evaluation criteria
- Select best models
- Interpret results in business terms (as far as possible at this stage)
- Check plausibility of model
- Check impacts for data mining goal
- Check model against given knowledge base to see if the discovered information is novel and useful
- Check reliability of result
- Analyse potentials for deployment of each result
- If there is a verbal description of the generated model (e.g. via rules), assess the rules; are they logical, are they feasible, are there too many or too few, do they offend common sense?
- Assess results

VIII.4 Assess Model (cont'd)

Outputs

- **Model Assessment**: Summarizes results of this task, lists qualities of generated models (e.g., in terms of accuracy), and ranks their quality in relation to each other.
- **Revised Parameter Settings**: According to the model assessment, revise parameter settings and tune them for the next run in task build model. Iterate model building and assessment until you strongly believe that you found the *best* model(s).



- 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.
- The results need to compared with the evaluation criteria defined at the start of the project.

IX.1 Evaluate Results

Data mining results cover **models** which are necessarily related to the original business objectives, **and all other findings** which are not necessarily related to the original business objectives but might also unveil additional challenges, information, or hints for future directions.

Activities

- Understand the data mining result
- Interpret the results in terms of the application
- Check impacts for data mining goal
- Check the data mining result against the given knowledge base to see if the discovered information is novel and useful
- Evaluate and assess result w.r.t. business success criteria
- Compare evaluation results and interpretation
- Create ranking of results w.r.t. business success criteria
- Check impacts of result for initial application goal

IX.1 Evaluate Results (cont'd)

Output

Assessment of Data Mining Results w.r.t. Business Success Criteria:
Summarizes assessment results in terms of business success criteria including a final statement whether the project already meets the initial business objectives.

Approved Models:
After model assessment w.r.t. business success criteria, you eventually get approved models if the generated models meet the selected criteria.

IX.2 Review Process

Task

- At this point the resultant model appears to be satisfactory and appears to satisfy business needs. It is now appropriate to make 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.

- At this stage, the Process Review takes on the form of a **Quality Assurance Review**. E.g., did we correctly build the model? Did we only use attributes that we were allowed to use and that are available for future analysis?

IX.2 Review Process (cont'd)

Activities

- Analyse data mining process
- Identify failures
- Identify misleading steps
- Identify possible alternative actions, unexpected paths in the process

Output
Summarizes the process review and gives hints for activities that have been missed and/or should be repeated.

IX.3 Determine Next Steps

Task

According to the assessment results and the process review, the project **decides how to proceed** at this stage. The project needs to decide whether to finish this project and **move onto deployment**, or whether to initiate **further iterations** or whether to set up **new data mining projects**.

Activities for Possible Actions

- Analyse potential for deployment of each result
- Estimate potential for improvement of current process
- Check remaining resources to determine if they allow additional process iterations (or whether additional resources can be made available)
- Recommend alternative continuations
- Refine process plan

Output Possible Actions

A list of possible further actions along with the reasons for and against each option.

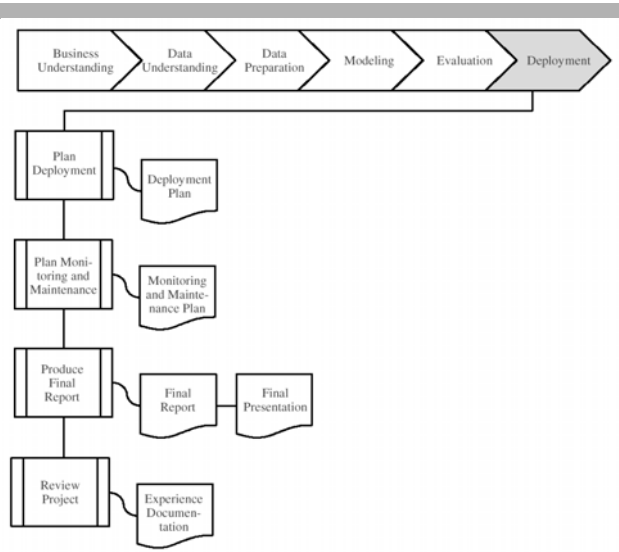
IX.3 Determine Next Steps

Activities Decision

- Rank the possible actions
- Select one of the possible actions
- Document reasons for the choice

Output Decision

Describes the decision as to how to proceed along with the rationale.



X.1 Plan Deployment

Task

This task takes the evaluation results and **concludes a strategy for deployment** of the data mining result(s) **into the business**.

Activities

- Develop and evaluate alternative plans for deployment
- Identify possible problems when deploying the data mining results (pitfalls of the deployment)

Output

Deployment Plan: Summarizes deployment strategy including necessary steps and how to perform them.

X.2 Monitoring and Maintenance

Task

- Monitoring and maintenance are important issues if the data mining result becomes part of the day-to-day business and its environment.
- A careful preparation of a **maintenance strategy** helps to avoid unnecessarily long periods of incorrect usage of data mining results.
- In order to monitor the deployment of the data mining result(s), the project needs a detailed plan on the **monitoring process**. This plan takes into account the specific type of deployment.

X.2 Monitoring and Maintenance (cont'd)

Activities

- Check for dynamic aspects (i.e. what things could change in the environment?)
- When should the data mining result or model not be used any more? Identify criteria (validity, threshold of accuracy, new data, change in the application domain, etc.)? What should happen if the model or result can no longer be used? (Update model, set up new data mining project, etc.)
- Will the business objectives of the use of the model change over time? – fully document the initial problem the model was attempting to solve
- Develop Monitoring and Maintenance Plan

Output Monitoring and Maintenance Plan:
Summarizes monitoring and maintenance strategy including necessary steps and how to perform them.

X.3 Produce Final Report

Task
At the end of the project, the project leader and his team write up a final report. It depends on the deployment plan, if this report is only a **summary of the project and its experiences**, or if this report is a **final presentation** of the data mining result(s).

X.3 Produce Final Report (cont'd)

Activities Final Report

- Identify what reports are needed (slide presentation, management summary, detailed findings, explanation of models ....)
- Analyse how well initial data mining goals have been met
- Identify target groups for report
- Outline structure and contents of report(s)
- Select findings to be included in the reports
- Write a report

Output Final Report
At the end of the project, there will be (at least one) final report where all the threads will be brought together. As well as identifying the results obtained, the report should also describe the process, show which costs have been incurred, define any deviations from the original plan, describe implementation plans and make any recommendations for future work. The actual detailed content of the report depends very much on the audience for the particular report.

X.3 Produce Final Report (cont'd)

Activities Final Presentation

- Decide on target group for final presentation (will they already have received final report?)
- Select which items from final report should be included in final presentation

Output Final Presentation

As well as a Final Report, it may be necessary to make a Final Presentation to summarize the project – maybe to the management sponsor, for example. The Presentation will normally contain a subset of the information contained in the Final Report, but structured in a different way.

X.4 Review Project

Task Assess **what went right** and **what went wrong**, what was done well and **what needs to be improved**.

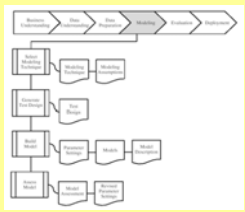
Activities

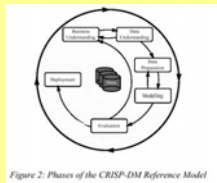
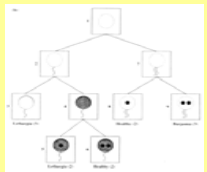
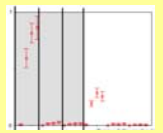
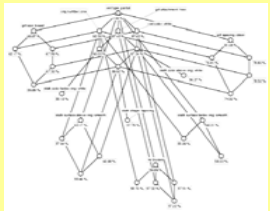
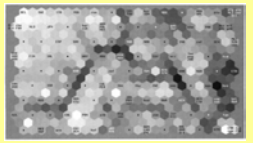
- Interview all significant people involved in the project and ask them about their experiences during the project
- If end users in the business work with the data mining result(s), interview them - are they satisfied? What could have been done better? Do they need additional support?
- Summarise feedback and write the experience documentation
- Analyse the process (things that worked well, mistakes made, lessons learned,...)
- Document the specific data mining process (How can the results and the experience of applying the model be fed back into the process?)
- Abstract from details to make the experience useful for future projects.

X.4 Review Project (cont'd)

Output Experience Documentation

Summarizes important experiences made during the project. For example, pitfalls, misleading approaches or hints for selecting the best-suited data mining techniques in similar situations could be part of this documentation. In ideal projects, experience documentation covers also any reports that have been written by individual project members during the project phases and their tasks.





The End