

CRISP-DM stage one – business understanding

Determine business objectives

Task

The first stage of the CRISP-DM process is to understand what the customer wants to accomplish from a business perspective. Customers often have competing objectives and constraints that must be properly balanced. The analyst's goal is to uncover important factors that could influence the outcome of the project. Neglecting this step can mean that a great deal of effort is put into producing the right answers to the wrong questions.

Outputs

1. **Set objectives** – describe the customer's primary objective from a business perspective. There may also be other related business questions that the customer would like to address. For example, the primary business goal might be to keep current customers by predicting when they are prone to move to a competitor. Related business questions might be "Does the channel used affect whether customers stay or go?" or "Will lower ATM fees significantly reduce the number of high-value customers who leave?"
2. **Produce project plan** – describe the plan for achieving the data mining and business goals. The plan should specify the steps to be performed during the rest of the project, including the initial selection of tools and techniques.
3. **Business success criteria** – describe the criteria for a successful outcome to the project from the business point of view. This could be specific and measurable, for example reduction of customer churn to a certain level, or it might be general and subjective, such as "give useful insights into the relationships." In the latter case, it needs to be clear who it is that makes the subjective judgment

Assess situation

Task

This task involves more detailed fact-finding about all of the resources, constraints, assumptions, and other factors that should be considered in determining the data analysis goal and project plan. In the previous task, your objective is to quickly get to the crux of the situation. Here, you want to expand upon the details

Outputs

1. **Inventory of resources** – list the resources available to the project, including personnel (business experts, data experts, technical support, data mining experts), data (fixed extracts, access to live, warehoused, or operational data),

computing resources (hardware platforms), and software (data mining/ Machine Learning tools, other relevant software).

2. **Requirements, assumptions and constraints** – list all requirements of the project, including schedule of completion, comprehensibility and quality of results, and security, as well as legal issues. As part of this output, make sure that you are allowed to use the data. List the assumptions made by the project. These may be assumptions about the data that can be verified data mining/Machine Learning but may also include non-verifiable assumptions about the business related to the project. It is particularly important to list the latter if it will affect the validity of the results. List the constraints on the project. These may be constraints on the availability of resources but may also include technological constraints such as the size of data set that it is practical to use for modelling.
3. **Risks and contingencies** – list the risks or events that might delay the project or cause it to fail. List the corresponding contingency plans, what action will be taken if these risks or events take place.
4. **Terminology** – compile a glossary of terminology relevant to the project. This may include two components: (1) A glossary of relevant business terminology, which forms part of the business understanding available to the project. Constructing this glossary is a useful “knowledge elicitation” and education exercise. (2) A glossary of data mining/ Machine Learning terminology illustrated with examples relevant to the business problem in question.
5. **Costs and benefits** – construct a cost-benefit analysis for the project which compares the costs of the project with the potential benefits to the business if it is successful. The comparison should be as specific as possible. For example, use monetary measures in a commercial situation.

Determine data mining/Machine Learning goals

Task

A business goal states objectives in business terminology. A data mining/Machine Learning goal states the project objective in technical terms. For example, the business goal might be “Increase catalogue sales to existing customers.” A data mining/Machine Learning goal might be “Predict how many widgets a customer will buy, given their purchases over the past three years, demographic information (age, salary, city, etc.), and the price of the item.”

Outputs

1. **Business success criteria** – describe the intended outputs of the project that enable the achievement of the business objectives.
2. **Data mining / Machine Learning success criteria** – define the criteria for a successful outcome to the project in technical terms—for example, a certain level of predictive accuracy or a propensity-to-purchase profile with a given degree of “lift.” As with business success criteria, it may be necessary to describe these in subjective terms, in which case the person or persons making the subjective judgment should be identified.

Produce project plan

Task

Describe the intended plan for achieving the data mining/Machine Learning goals and thereby achieving the business goals. The plan should specify the steps to be performed during the rest of the project, including the initial selection of tools and techniques.

Outputs

1. **Project plan** – list the stages to be executed in the project, together with their duration, resources required, inputs, outputs, and dependencies. Where possible, make explicit the large-scale iterations in the data mining process—for example, repetitions of the modelling and evaluation phases. As part of the project plan, it is also important to analyse dependencies between time schedule and risks. Mark results of these analyses explicitly in the project plan, ideally with actions and recommendations if the risks are manifested. Note: the project plan contains detailed plans for each phase. Decide at this point which evaluation strategy will be used in the evaluation phase. The project plan is a dynamic document in the sense that at the end of each phase, a review of progress and achievements is necessary, and a corresponding update of the project plan is recommended. Specific review points for these updates are part of the project plan.
2. **Initial assessment of tools and techniques** – at the end of the first phase, an initial assessment of tools and techniques should be performed. Here, for example, you select a data mining tool that supports various methods for different stages of the process. It is important to assess tools and techniques early in the process since the selection of tools and techniques may influence the entire project.