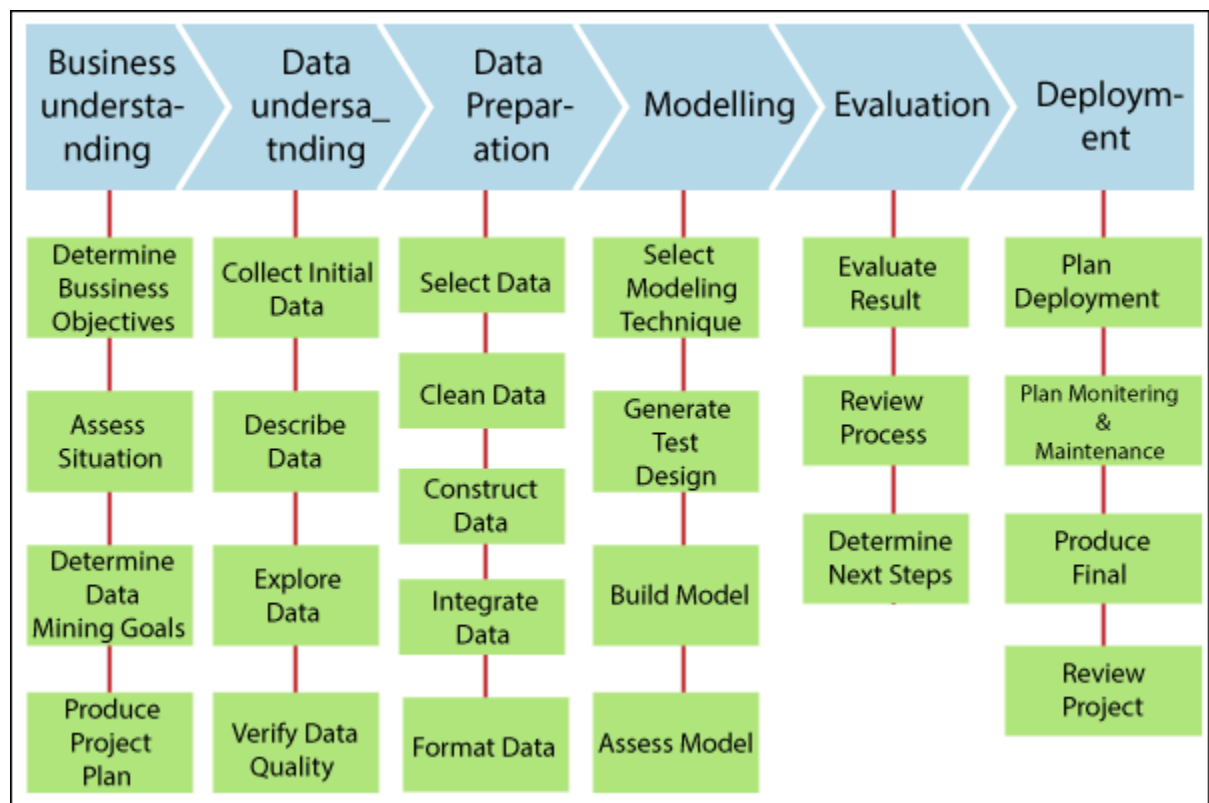


### Data Mining Implementation Process

Many different sectors are taking advantage of data mining to boost their business efficiency, including manufacturing, chemical, marketing, aerospace, etc. Therefore, the need for a conventional data mining process improved effectively. Data mining techniques must be reliable, repeatable by company individuals with little or no knowledge of the data mining context. As a result, a cross-industry standard process for data mining (CRISP-DM) was first introduced in 1990, after going through many workshops, and contribution for more than 300 organizations.

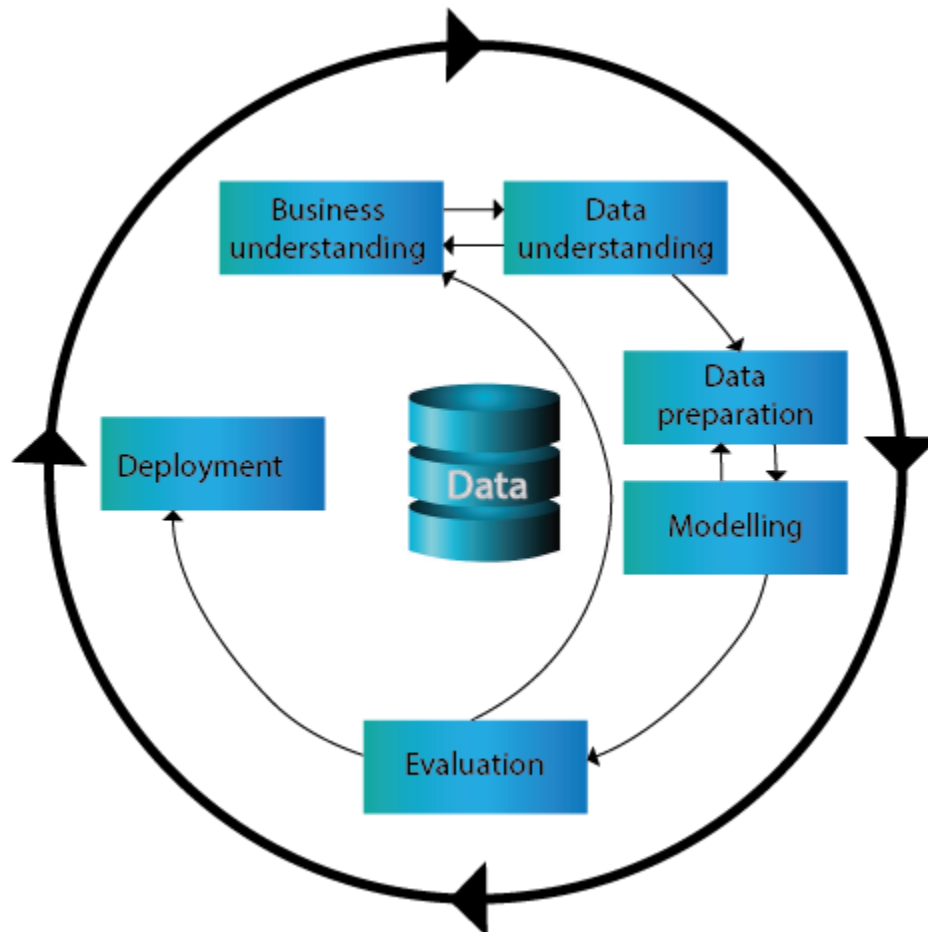
Data mining is described as a process of finding hidden precious data by evaluating the huge quantity of information stored in data warehouses, using multiple data mining techniques such as Artificial Intelligence (AI), Machine learning and statistics.



Let's examine the implementation process for data mining in details:

The Cross-Industry Standard Process for Data Mining (CRISP-DM)

Cross-industry Standard Process of Data Mining (CRISP-DM) comprises of six phases designed as a cyclical method as the given figure:



1. Business understanding:

It focuses on understanding the project goals and requirements from a business point of view, then converting this information into a data mining problem afterward a preliminary plan designed to accomplish the target.

**Tasks:**

- Determine business objectives

- Access situation
- Determine data mining goals
- Produce a project plan

**Determine business objectives:**

- It Understands the project targets and prerequisites from a business point of view.
- Thoroughly understand what the customer wants to achieve.
- Reveal significant factors, at the starting, it can impact the result of the project.

**Access situation:**

- It requires a more detailed analysis of facts about all the resources, constraints, assumptions, and others that ought to be considered.

**Determine data mining goals:**

- A business goal states the target of the business terminology. For example, increase catalog sales to the existing customer.
- A data mining goal describes the project objectives. For example, It assumes how many objects a customer will buy, given their demographics details (Age, Salary, and City) and the price of the item over the past three years.

**Produce a project plan:**

- It states the targeted plan to accomplish the business and data mining plan.
- The project plan should define the expected set of steps to be performed during the rest of the project, including the latest technique and better selection of tools.

**2. Data Understanding:**

Data understanding starts with an original data collection and proceeds with operations to get familiar with the data, to data quality issues, to find better insight in data, or to detect interesting subsets for concealed information hypothesis.

**Tasks:**

- Collects initial data
- Describe data
- Explore data
- Verify data quality

**Collect initial data:**

- It acquires the information mentioned in the project resources.
- It includes data loading if needed for data understanding.
- It may lead to original data preparation steps.
- If various information sources are acquired then integration is an extra issue, either here or at the subsequent stage of data preparation.

**Describe data:**

- It examines the "gross" or "surface" characteristics of the information obtained.
- It reports on the outcomes.

**Explore data:**

- Addressing data mining issues that can be resolved by **querying**, **visualizing**, and **reporting**, including:
  - Distribution of important characteristics, results of simple aggregation.
  - Establish the relationship between the small number of attributes.
  - Characteristics of important sub-populations, simple statical analysis.

- It may refine the data mining objectives.
- It may contribute or refine the information description, and quality reports.
- It may feed into the transformation and other necessary information preparation.

**Verify data quality:**

- It examines the data quality and addressing questions.

**3. Data Preparation:**

- It usually takes more than 90 percent of the time.
- It covers all operations to build the final data set from the original raw information.
- Data preparation is probable to be done several times and not in any prescribed order.

**Tasks:**

- Select data
- Clean data
- Construct data
- Integrate data
- Format data

**Select data:**

- It decides which information to be used for evaluation.
- In the data selection criteria include significance to data mining objectives, quality and technical limitations such as data volume boundaries or data types.
- It covers the selection of characteristics and the choice of the document in the table.

**Clean data:**

- It may involve the selection of clean subsets of data, inserting appropriate defaults or more ambitious methods, such as estimating missing information by modeling.

**Construct data:**

- It comprises of Constructive information preparation, such as generating derived characteristics, complete new documents, or transformed values of current characteristics.

**Integrate data:**

- Integrate data refers to the methods whereby data is combined from various tables, or documents to create new documents or values.

**Format data:**

- Formatting data refer mainly to linguistic changes produced to information that does not alter their significance but may require a modeling tool.

**4. Modeling:**

In modeling, various modeling methods are selected and applied, and their parameters are measured to optimum values. Some methods gave particular requirements on the form of data. Therefore, stepping back to the data preparation phase is necessary.

**Tasks:**

- Select modeling technique
- Generate test design
- Build model
- Access model

**Select modeling technique:**

- It selects the real modeling method that is to be used. For example, decision tree, neural network.
- If various methods are applied, then it performs this task individually for each method.

#### **Generate test Design:**

- Generate a procedure or mechanism for testing the validity and quality of the model before constructing a model. For example, in classification, error rates are commonly used as quality measures for data mining models. Therefore, typically separate the data set into train and test set, build the model on the train set and assess its quality on the separate test set.

#### **Build model:**

- To create one or more models, we need to run the modeling tool on the prepared data set.

#### **Assess model:**

- It interprets the models according to its domain expertise, the data mining success criteria, and the required design.
- It assesses the success of the application of modeling and discovers methods more technically.
- It Contacts business analytics and domain specialists later to discuss the outcomes of data mining in the business context.

#### **5. Evaluation:**

- At the last of this phase, a decision on the use of the data mining results should be reached.
- It evaluates the model efficiently, and review the steps executed to build the model and to ensure that the business objectives are properly achieved.

- The main objective of the evaluation is to determine some significant business issue that has not been regarded adequately.
- At the last of this phase, a decision on the use of the data mining outcomes should be reached.

**Tasks:**

- Evaluate results
- Review process
- Determine next steps

**Evaluate results:**

- It assesses the degree to which the model meets the organization's business objectives.
- It tests the model on test apps in the actual implementation when time and budget limitations permit and also assesses other data mining results produced.
- It unveils additional difficulties, suggestions, or information for future instructions.

**Review process:**

- The review process does a more detailed evaluation of the data mining engagement to determine when there is a significant factor or task that has been somehow ignored.
- It reviews quality assurance problems.

**Determine next steps:**

- It decides how to proceed at this stage.
- It decides whether to complete the project and move on to deployment when necessary or whether to initiate further iterations or set up new data-mining initiatives. It includes resources analysis and budget that influence the decisions.



## 6. Deployment:

### **Determine:**

- Deployment refers to how the outcomes need to be utilized.

### **Deploy data mining results by:**

- It includes scoring a database, utilizing results as company guidelines, interactive internet scoring.
- The information acquired will need to be organized and presented in a way that can be used by the client. However, the deployment phase can be as easy as producing. However, depending on the demands, the deployment phase may be as simple as generating a report or as complicated as applying a repeatable data mining method across the organizations.

### **Tasks:**

- Plan deployment
- Plan monitoring and maintenance
- Produce final report
- Review project

### **Plan deployment:**

- To deploy the data mining outcomes into the business, takes the assessment results and concludes a strategy for deployment.
- It refers to documentation of the process for later deployment.

### **Plan monitoring and maintenance:**

- It is important when the data mining results become part of the day-to-day business and its environment.
- It helps to avoid unnecessarily long periods of misuse of data mining results.

- It needs a detailed analysis of the monitoring process.

**Produce final report:**

- A final report can be drawn up by the project leader and his team.
- It may only be a summary of the project and its experience.
- It may be a final and comprehensive presentation of data mining.

**Review project:**

- Review projects evaluate what went right and what went wrong, what was done wrong, and what needs to be improved.