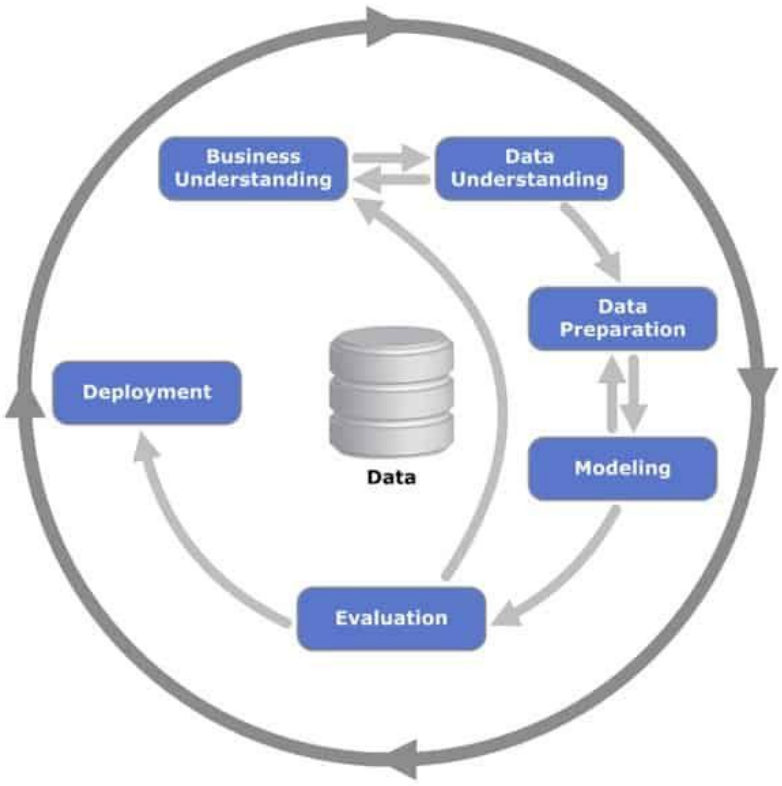
Vinoothna Bavireddy

CRISP DM

Introduction:

CRISP DM stands for Cross-industry standard process for data mining. The lifecycle of CRISP DM has six different phases, where each iterative phase has its own defined tasks and deliverables such as documentation and reports. The arrows between the phases represent the direction of action and to indicate that there is a back and forth action happening in between these phases to achieve better results. The outer circle represents the continuous learning process involved in this process.



CRISP-DM process diagram (Wikimedia Commons, 2012).

History:

It was developed by representatives from SPSS, Teradata, Daimler, NCR, and OHRA in 1996 to standardize a data mining process across industries. CRISP-DM has not been built in a theoretical, academic manner working from technical principles, nor by elite committees behind closed doors. The success of CRISP-DM is based on its ability to be implemented in data mining practice of a real world scenario.

Methodology:

The methodology involved in a CRISP DM process follows a hierarchical breakdown which involves 4 levels of abstraction as shown in the figure below. The four levels :

1.phase: Categorized into different phases which is further broken down into several secondary tasks
2.generic task: It is called generic because it is intended to cover all generic situations for data mining which is both complete and stable.
3.specialized task: This level describes how the generic tasks are to be performed. For example Transforming the data from Categorical to numerical.
4.process instance: Is a result of an actual data mining engagement which is specific to an engagement rather than a generic action.

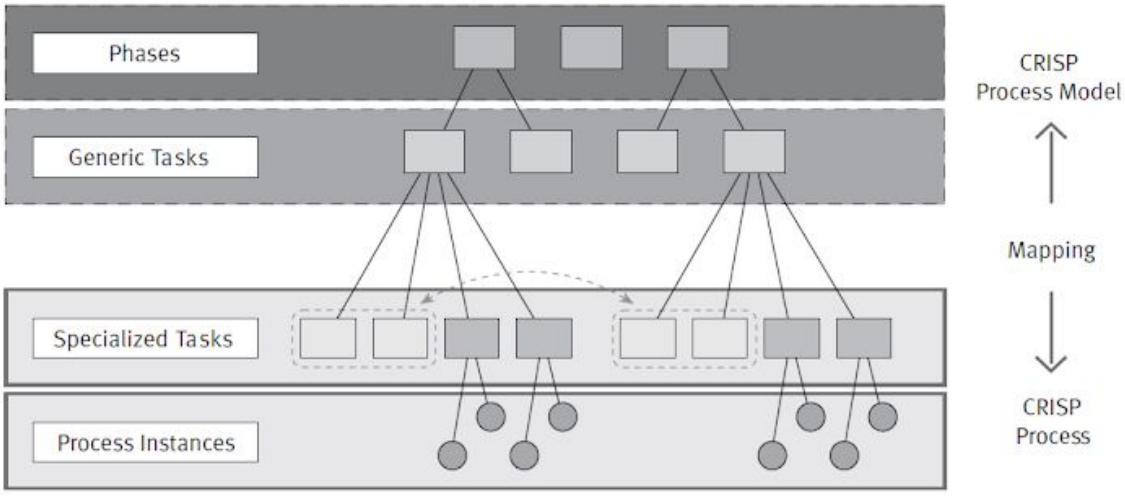


Figure 2: Four-level Breakdown of CRISP methodology[1]

Model:
The different stages of the life cycle include:
1.Business Understanding: stating the objectives of the research question or problem statement, assessing the issue at hand, to determine the goals for data mining and to draft a plan.
2.Data Understanding: Identifying the source of the data, exploring and understanding the data and to verify the quality of data.
3.Data Preparation (generally, the most time-consuming phase): This includes steps like SELECTING, CLEANING (dealing with Null values), TRANSFORM and FORMATTING to fit the use case.
4.Modeling: selecting your modeling technique, building and testing the model happens in this phase.
5.Evaluation: This step is crucial in addressing your objectives by evaluating the results obtained and to check if there is anything that needs to be addressed.
6.Deployment: This step involves implementing/ deploying the result into practice. This is not the end step and will require continuous review to achieve best results.

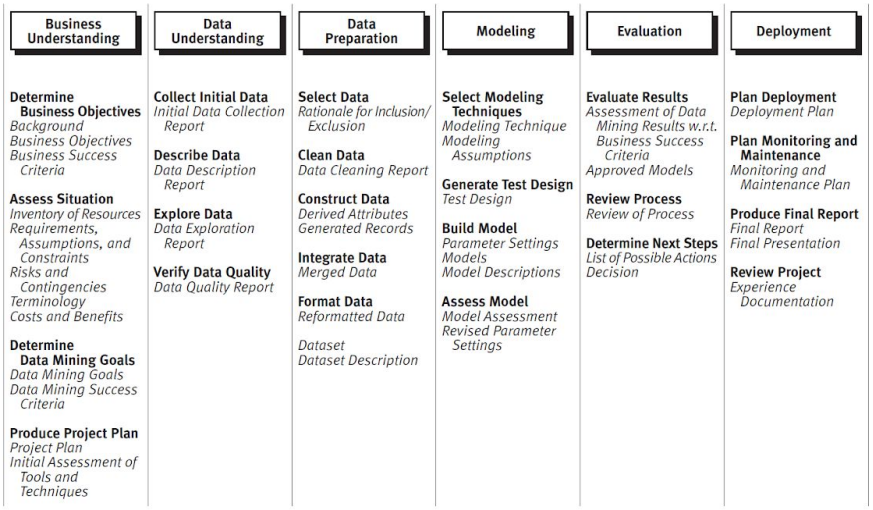


Figure 3: The generic tasks are in bold and outcomes of that step are represented in italics[2]

PROs for CRISP DM:

1. The biggest advantage of CRISP DM is that it is generalizable to many situations even for those that are not data mining related.
2.CRISP DM approach offers strong guidance and hence improves performance
3.Can be adopted easily without much training.
4.The first step of the cycle helps understand business needs thereby helping to solve the problem at hand by aligning technical work with the business side of things.
5.The iterative process ensures smooth deployment helping individuals to transition into maintenance and operations quickly.
6.CRISP DM follows an Agile approach giving a deeper understanding of data and imbibing the knowledge from previous cycles to the next there by creating room for flexibility.

CONS for CRISP DM:

1. Heavy documentation for data collection.
2. Has not been updated since its creation. So, may not be applicable for modern day problems
3. Does not encourage teamwork and collaboration and is not a true project management methodol

References:

1. *Crisp-dm*. (2020, August 5). Data Science Project Management. https://www.datascience-pm.com/crisp-dm-2/
2. Pete Chapman (NCR), Julian Clinton (SPSS), Randy Kerber (NCR). et.al (2000). *Step by step Data Mining Guide*. Predictive Analytics & Machine Learning Global Services in Training, Consulting and Enablement by TMA, LLC. https://the-modeling-agency.com/crisp-dm.pdf