**Big-Five Factor Markers**

Mateo González

Alejandro Uribe

June 2020

**Table of Contents**

[1. Introduction 1](#_Toc42270843)

[2. Data Mining Process 1](#_Toc42270844)

[2.1. Business Understanding 1](#_Toc42270845)

[2.2. Data Understanding 1](#_Toc42270846)

[2.3. Data Preparation 1](#_Toc42270847)

[2.4. Modelling 2](#_Toc42270848)

[2.4.1. Choice of Algorithm 2](#_Toc42270849)

[2.4.2. Training 2](#_Toc42270850)

[2.5. Evaluation 2](#_Toc42270851)

[2.5.1. Choice of Metrics 2](#_Toc42270852)

[2.5.2. Testing 2](#_Toc42270853)

[3. Conclusion 2](#_Toc42270854)

[4. References 2](#_Toc42270855)

**Table of Figures**

[Figure 1. CRISP-DM Methodology. 1](#_Toc42271388)

**List of Tables**

No table of figures entries found.

# Introduction

# Data Mining Process

*Cross-Industry Standard Process for Data Mining* (CRISP-DM) is the most commonly used methodology for analytics, data mining and data science projects which provides an overview of the life cycle of a data mining project. [1]. CRISP-DM methodology guides the data scientist through a small set of phases which cover all possible data mining situations. The Figure 1 shows the CRISP-DM Methodology.

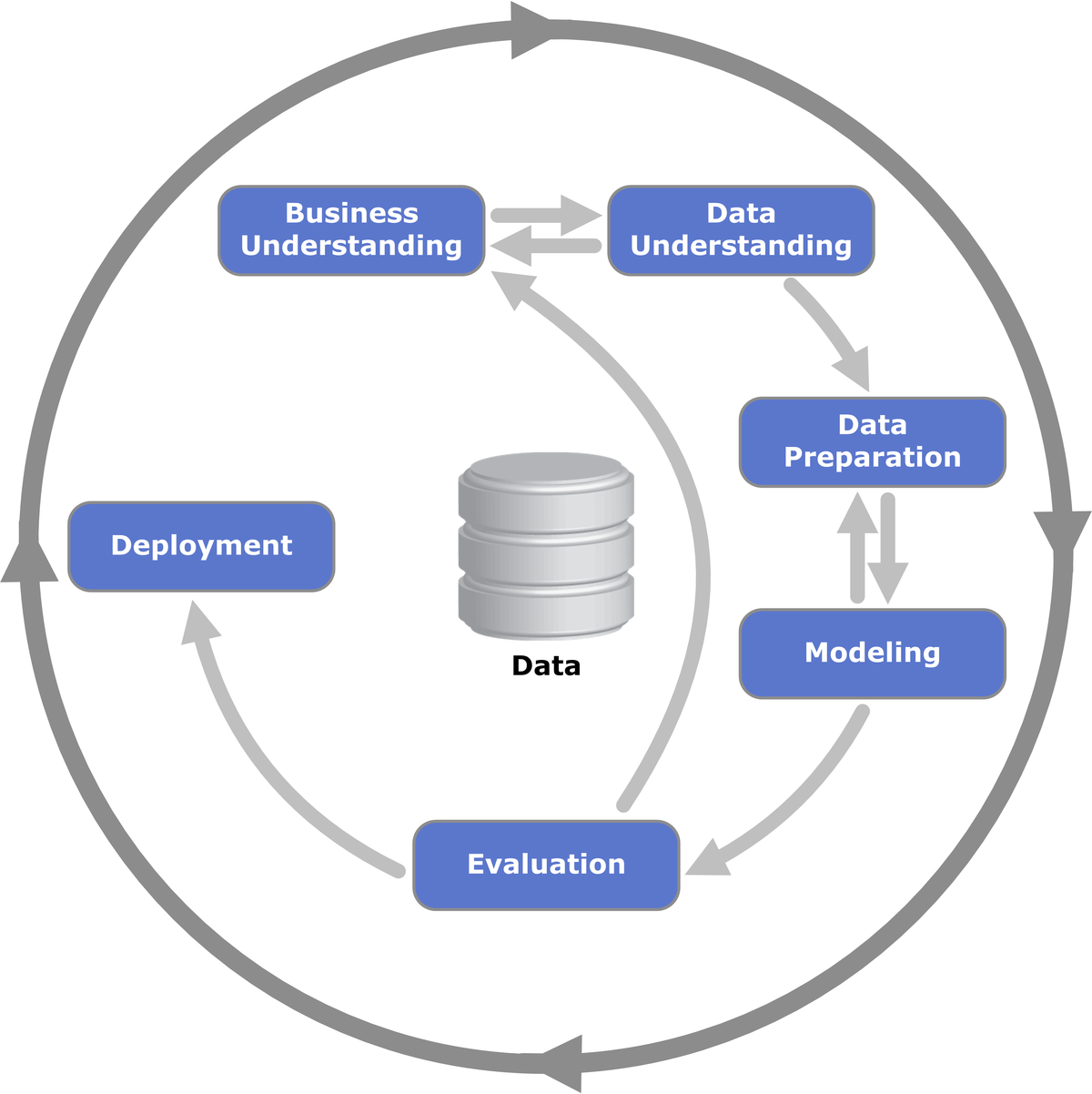


Figure 1. CRISP-DM Methodology.

Source: [2].

The application of this methodology can trigger new business questions which can benefit the data mining experience.

## Business Understanding

The first phase is the most important, a clear understanding of the problem

## Data Understanding

The success of a data mining project depends on the data available, its quality and the thorough understanding of the same by the team. Focusing on this phase can greatly reduce future data mining effort.

The first approach

## Data Preparation

This phase can represent 80% of the spent in the project, is aimed to detect the errors and inconsistencies in the dataset, these values can be either ignored or replaced.

The fig illustrates an approach to data cleansing [3].

## Modelling

Here comes the “fun part”, since there is a wide range of algorithms or techniques which can apply for the given problem. Multiple algorithms will be tried

### Choice of Algorithm

### Training

## Evaluation

### Choice of Metrics

### Testing

# Conclusion

­­

# References

[1] M. F. Hornick, E. Marcadé, and S. Venkayala, “Chapter 3 - Data Mining Process,” in *The Morgan Kaufmann Series in Data Management Systems*, M. F. Hornick, E. Marcadé, and S. B. T.-J. D. M. Venkayala, Eds. Burlington: Morgan Kaufmann, 2007, pp. 51–83.

[2] “File:CRISP-DM Process Diagram.png - Wikimedia Commons.” [Online]. Available: https://commons.wikimedia.org/wiki/File:CRISP-DM\_Process\_Diagram.png. [Accessed: 06-Jun-2020].

[3] K. K. Al-jabery, T. Obafemi-Ajayi, G. R. Olbricht, and D. C. Wunsch II, “2 - Data preprocessing,” K. K. Al-jabery, T. Obafemi-Ajayi, G. R. Olbricht, and D. C. B. T.-C. L. A. to D. A. in B. A. Wunsch II, Eds. Academic Press, 2020, pp. 7–27.