

Groups

Groups with two elements

Goals

The goal of this project is to apply the knowledge acquired as part of this course on a real problem. After the project, you should be able to:

- Develop a data mining project by following an appropriate methodology
- Clearly understand a few preprocessing methods and learning algorithms
- Clearly understand model evaluation methodologies and a few evaluation measures
- Present the results of the project in a clear and synthetic way

Problem definition

You should identify a data analytics problem that is of your particular interest. The data involved should:

- Include qualitative and quantitative predictive attributes
- More than 2 classes, if it is a classification problem

If you cannot identify a problem, you should look for one in data repositories such as the UCI ML Repository (<https://www.kaggle.com/>).

You should identify one descriptive or predictive problem that could be addressed with the chosen dataset.

The work should be done using RapidMiner, R language and/or Python

The evaluation is based on one report

The report can be written in Portuguese or English and should follow the CRISP-DM methodology, phases 1 to 5.

Delivery date: 01/12/2022.

The evaluation criteria for the report are:

- Problem understanding: 20%
- Data identification and understanding: 20%
- Data preparation, modeling and evaluation including the diversity of algorithms used: 40%
- General quality of the report: 20%
- There will be a penalty of 1 value per each day of delay