Week 1

ToDo:

all: Choose any dataset you like and do the following

3.1 Description of the original data Where was it obtained (web page link), description of the problem on hands: target of classification, original number of examples, original number of columns meaning and kind of values, impact of missing values, etc

We will choose on the next session the one we think is more appropriate

Tommaso:

Eric:

Alex:

Dataset: **Predict students' dropout and academic success**

Rows: 4424

Columns: 36

Webpage: <https://archive-beta.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success>

He:

Pol:

Done:

Tommaso:First analysis to see how the models performs with …

Eric:

Alex:

He:

Pol: First analysis to see how the models performs with KNN

Week 2

To do:

Description of the original data

Description of preprocessing of data

Evaluation criteria of data mining models

Tommaso :Execution of different machine learning methods:

* *KNN*
* *…*

Eric: Missing data, motivation description

Alex: data source presentatio, bi-variate

He: data structure and metadata, boxplots, histogram, univariete analysis

Pol: Evaluation criteria of data mining models (documentation), Meta-learning algorithms

Done:

Tommaso:

Eric:

* Missing data almost finished
* Motivation and description of the problem

Alex:

He:

* 5/05 -> all boxplot and histogram for univariate analysis plotted. Ready to analyze, explain the plots, also explain the outliers.

Pol:

Week 3

To do:

Tommaso:

Eric:

Alex:

He:

Pol:

Done:

Tommaso:

Eric:

Alex:

He:

Pol: