Manual 3 – Jupyter Notebooks

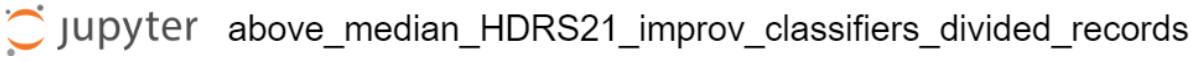
In this manual there will be a brief explanation on the 5 Jupyter notebooks attached.

This notebook contains the following parts:

- creating count table out of the abstractions Json folder.
- merging it with a y label.
- creating a binary label for above/below median.
- defining the function filter_by_percentile that filters middle percentile of distribution.
- running a loop for different models that uses filter by percentile and leave_one_out validation.
- Running second loop for calculating metrics and saving them in a DataFrame.
- Saving results.

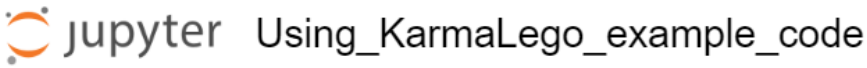
Notes:

- The process can be done on different abstraction methods since they are supposed to be in a conventional format for KarmaLego usage.
- The label can be altered to a different label.
- The models and hyperparameters can and should also be changed and tested for optimal results.



This notebook is essentially the same as the one above with minor changes, as follows:

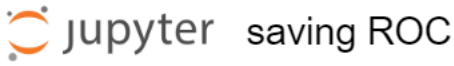
- Json files merging is connected to records dividing using the function "divide_entities_and_merge" from "json_file_func" (explained further in Manual 2)
- The "leave one out" validation is replaced with "leave one subject out" ensuring no data leakage.



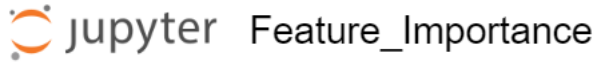
This notebook contains an algorithm used for running KarmaLego from Json abstraction folder to final Dataframe used for modeling. It contains the following steps:

- Sampling the data to save computing power usage.
- Merging all sampled Json files.
- Creating multi KarmaLego config file.
- Running multi KarmaLego on merged file.
- Creating single KarmaLego config files for all the full records.

- Running single KarmaLego in loop for all full records.
- Merging results to single .csv file containing horizontal support and mean duration for all subjects.
- Merging the file with label to final file that can be used for modeling.



This notebook is made for creating and saving ROC plot of top performing model, including AUC calculation and Knee Values based on the maximum of Youden's J value.



This notebook contains the calculation and presentation of the feature importance method elaborately explained in the final report.