

Myra's teamprojects KMD – Summer term 2023

PREDMA1a: Building a collection of public domain datasets for predictive maintenance tasks

Goal of this teamproject is to collect public domain datasets with multivariate time series that are appropriate for following predictive maintenance tasks, namely estimation of Remaining Useful Life, survival analysis and anomaly detection.

The collection must encompass following functionalities:

1. Computation and visualization of value distributions for all the variables
2. Correlation analysis and visualizations *
3. Mechanisms for the inspection of individual time series and of the whole multivariate time series to assist in the detection of anomalies through a human user
4. Mechanisms for the segmentation of time series (segment size and location to be given as parameters) *
5. Mechanisms to compute and store the outlieriness value of each segment *
6. Mechanisms for storing and querying the data

* Task requires literature collection and justified selection of methods

PREDMA 1b: Testbed and evaluation of the toolbox of PREDMA 1a

Goal of this teamproject is to establish a testbed for the evaluation of the PREDMA 1a toolbox. The tasks are:

1. Establish the testbed functionalities through interviews with the user
2. Establish a test workflow
3. Develop evaluation mechanisms and incorporate them to the testbed
4. Build a demonstrator
5. Build a documentation for an independent team that will do the evaluation

These two teamprojects run together and are evaluated together. The output of PREDMA 1a consists of software, documentation and a report that contains the description of the toolbox and the description of the justified selection of methods. The output of PREDMA 1b consists of software, documentation, the evaluation report for PREDMA 1a, and a report that contains the description of the interviews, the test workflow, the evaluation mechanisms and the evaluation results.

MIMICLIT: Literature survey on the MIMIC datasets collections

Goal of this teamproject is to perform a literature survey on scientific publications that use the MIMIC III dataset and followup datasets. The survey must follow the PRISMA methodology to answer following questions:

- Which scientific tasks are evaluated on MIMIC?
 - What are the commonalities and differences among these tasks?
 - How to summarize them? [PRISMA synthesis]
- What are the data filtering and data preparation procedures applied on MIMIC?
 - What are the commonalities and differences among these procedures?
 - How to summarize them? [PRISMA synthesis]
- What are the evaluation criteria used on the scientific tasks evaluated on MIMIC?
 - What are the commonalities and differences among these criteria?
 - How to summarize them? [PRISMA synthesis]
- What are the main findings of scientific research that uses MIMIC?
 - What are the commonalities and differences among these findings?
 - How to summarize them? [PRISMA synthesis]

The output of this teamproject is a literature survey to be submitted to a scientific journal.