

Assignment for:
Applied data analysis and machine learning
Classical machine learning

The data set

Subset of the data set described in this paper:

Beata Strack, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore, "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records," BioMed Research International, vol. 2014, Article ID 781670, 11 pages, 2014.

Data for 5000 patients suffering from diabetes, including general information (age group, sex, etc.) and results of lab tests. The final column is the class to be predicted: whether the patient had to be hospitalised again in the future.

The task

Analyse the data set to obtain a model predicting the class of each patient.

The final submission must contain:

1. A report (PDF or Word), with a synthesis of:
 - a. Data exploration / pre-processing
 - b. Results of the classification task.

Note: the report doesn't need to be long. Probably around 5 pages should be enough. Just make sure that all important topics are covered.

2. The Python code used to obtain the results.

Please submit the assignment via the Aula Digital.

Deadline

April 7th, 2025.