Ostbayerische Technische Hochschule
Amberg-Weiden

# Project Work Machine Learning
# Winter 2025

Prof. Dr. Patrick Levi

3. Januar 2026

Start: 05.01.2026 (via Moodle)
Deadline: 12.01.2026, 12:00:00 (via Moodle)

## General

### General Requirements

- Upload your complete solution until the specified deadline via Moodle.
- Any documentation part or code parts created with AI tools must be specified, what tool was used and to what purpose the tool was used.
- Plagiarism: If your solution copies contains parts copied from any resource including solutions of other people, all solutions involved in the plagiarism (the copy and the master solution) will be graded with 5.0 (fail).

## Grading Criteria

### General

- Your solution solves the task and has sufficient quality.
- Your solution is well founded and well justified. Explain your solution in the documentation.
- Do not limit yourself just to techniques from the lecture but also research other possible approaches to find the best way to solve the project. Include current knowledge in the field and the current literature
- Your solution is efficient and effective (do the right things, do the things right).
- Your solution exceeds the quality obtained by AI tools when they are asked to solve the task.
- Your solution demonstrates a deep understanding of the problem.

### Code

- Code must be written in Python (except otherwise specified)
- Your code is efficient, understandable, and written in a way that is not error-prone.

- Wherever possible, use available Python packages. Restrictions might be specified in the project description.
- Your code must run on the computers in the GPU lab (DC 1.07).

**Documentation**

- Your documentation must be a PDF. The use of LaTeX is recommended, but not obligatory. However, the documentation shall be in a proper report format (see grading). Markdown files or Jupyter notebooks hardly fulfill good report criteria.
- Your documentation presents your solution. Avoid unnecessary information in the documentation.
- It must be written in a way that another AI master student, who is not an expert in the field of the project task, could follow what you did and why you did it.
- It must be well-structured and written in proper language.
- Tables and figures shall be on point, clear, and concise.
- Each step in your solution must be well justified in the documentation.
- List all your references, use a proper scientific citation standard.

# Machine Learning Challenge – Winter 2025

In Machine Learning, challenges are usual. ML practitioners like taking part into online ML challenges, companies like to send out small challenges in advance to a job interview, when hiring data scientists, ML engineers, or similar positions.

In the following, you find a small ML challenge consisting of two parts. Please read the requirements carefully and stick to the instructions.

## General Instructions

Receive your individual dataset from Moodle. First, register for an exam id, then download the dataset **corresponding to your exam ID**. Your solution will be accepted only if it is for your dataset with your ID.

Your dataset consists of three files: ”dataset_‹ID›.csv” contains the training features, ”target_‹ID›.csv” contains the training targets and ”EVAL_‹ID›.csv” contains the evaluation features.

### Documentation

Your documentation is just a lean write-up, but it must be clear, well structured, and understandable to an external person with basic machine learning knowledge. Limit to 1-3 pages for part 1, 2-4 pages for part 2. Focus on a short description of your method and why you chose it. You do not have to add a full data anaylsis to the documentation. However, if it helps presenting your results, you are free to do so. Citations must follow scientific standards.

### Deliverables

Please note that every part has specific deliverables. Make sure to upload **all required deliverables** to Moodle and stick to the **required naming convention**. The files are evaluated by a pipeline. Wrong file names, headers, or formats cause, as in real ML applications, errors in the evaluation pipeline. A submission that cannot be evaluated automatically may be downgraded.

Please upload the following deliverables to Moodle:

- Documentation as PDF
- ”EVAL_target01_‹ID›.csv” (see part 1)
- ”framework_‹ID›.py” (see part 2)
- Source code files in a separate folder.

## Work packages

### Part 1

Predict the continuous variable ”target01”. Try to get the best-possible prediction from the data.

Submit a file ”EVAL_target01_‹ID›.csv” containing the predictions on the evaluation dataset ”EVAL_‹ID›.csv” in the right order. Use the header ”target01”. Describe and justify your approach in the write-up.
Specific criteria: Prediction quality, soundness of the approach, correct realization and description of the approach.

### Part 2

The second target (”target02”) is known to depend only on a few features. Furthermore, it is known that a few simple rules exist to predict it accurately. The rules are generally formulated like this: if condition1 holds, then target02 is calculated according to calculation1, else if condition2 holds, then target02 is calculated according to calculation2, else if ... . Only few conditions and calculations are required. The customer’s experts describe the conditions and calculations as ”simple”.
The customer plans to implement the prediction for ”target02” directly on an edge device. Unfortunately, the edge device may not execute machine learning routines. They developed a simple programming interface working with basic comparisons and executing numerical functions. This reflects their above-mentioned expectation of how ”target02” is related to the features. Figure out, on which features ”target02” depends and how.

The deliverable for this part is the implementation of the conditions and calculations according to the customer’s programming framework to predict ”target02”. You find a file called ”framework.py” in the project. Implement the conditions and calculations at the marked spots, rename the file to ”framework_‹ID›.py” and submit this file via Moodle.
Note, you may not import further packages and the file must be running. You do not need to submit predictions.

Use the write-up to provide evidence for your result. In case it is incomplete or wrong, this may help gaining points. If possible, limit it to 4 pages or less.
Task-specific criteria: Software is correct, follows specifications and runs, correctness of the conditions and calculations (prediction quality on evaluation set), the approach to obtain the conditions and calculations.