Ostbayerische Technische Hochschule
Amberg-Weiden

Project Work Machine Learning Summer 2025

Prof. Dr. Patrick Levi

2. Mai 2025

Start: 05.05.2025 (via Moodle)
Deadline: 04.06.2025, 23:59:59 (via Moodle)

General

General Requirements

- Your solution of this project work consists of a documentation (approx. 10-15 pages) and code files.
- 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.
- Hand in a zip file containing your documentation, the filled-in and signed declaration at the end of this project description for each team member, and your code.
- 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 team mates (for individual parts) or of other teams, all solutions involved in the plagiarism (the copy and the master solution) will be graded with 5.0 (fail).

Team Requirements

- The following project is subdivided into two parts, a team part and an individual part. The individual part must be done individually by each team member!
- You are free to do the whole project work individually. You do not have to work in teams.
- You must select a team via Moodle until the start date above. Team changes are not possible after this date. However, you are free to decide to do the task individually at any time. Just hand in an individual solution.
- Therefore, please indicate the authorship in your final solution (see below, requirements on team authorship).

- In case of a team breakup during the project, each team member has to hand in an individual solution. New teams may not be formed.
- In case you work in a team:
  - The team size is limited to two students.
  - Every student has to hand in a complete solution (code and documentaiton for team part and individual part).
  - Please indicate the correct authorship in your documentation (e.g. ”This chapter was authored by name team member 1 and name team member 2.” for team parts). The same applies for code. In your code, indicate the authorship in the docstrings of your classes and methods.
  - In the project specification you will find some more specific team requirements.
  - **Grading will be individual.**

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 well structured (packages, classes, methods, ...), easy to read, understandable, and there are sufficient comments in the code. Uncommented code will be down-graded.
- In addition to comments in the code, every function must contain an appropriate docstring. You can follow the NumPy docstring guide: `https://numpydoc.readthedocs.io/en/latest/format.html`. Notice, a Sphinx documentation is not required.
- As a rule of thumb: The more complex the function, the more comments are required in the code.
- 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 presents your solution. Avoid unnecessary information in the documentation. It is not intended to be a protocol of your progress, but a result report.
- 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 follow a scientific writing standard. Take scientific papers as a template.
- It must be well-structured and written in proper language.
- Tables and figures shall be on point, clear, and consise.
- Each step in your solution must be well justified in the documentation.
- List all your references, use a proper scientific citation standard.

Project – Predictive Maintenance for Wind Turbines

Imagine the following situation: You are working for a company developing machine learning solutions for predictive maintenance. Your customer needs a predictive solution for wind turbines. Your task is to prototype and compare two approaches and write a short report for the customer.

You get data for wind farms along with a descirption, in the paper [1]. You work on data for wind farm C only. You will find failure event descriptions coming with the dataset. The dataset itself contains various recordings of sensor data.
The customer requires a solution that

- Predicts all the failure events for all turbines as early in advance as possible.
- While it keeps false alarm rate as low as possible.

Work packages

- WP 1 - Data analysis: Analyze the data to select features to use for the machine learning part in WP 2. Find appropriate methods of feature selection from the data, provide arguments for your features.
- WP 2 - Design, implement, run, evaluate, and compare two approaches: One based on supervised learning and a second one based on clustering. Use both approaches to predict failures ready in time. For supervised learning use a meaningful target variable. Fulfill the customer requirements and evaluate compliance. Use your features from WP 1, develop ideas how to optimze your approach, define good training and test sets, and demonstrate the validity of your approaches. This part is a challenge, you may find many ways to solve it. Avoid doing the same things as others, including your team mate. Which approach (supervised or unsupervised) do you prefer and why?

Present your results in a short report in a structured, scientific way. Select meaningful plots, figures, textual descriptions to describe your approach and your results. Consider your reader (the customer) has principle knowledge of machine learning but has never looked into the dataset in detail.

Make sure, that individual and team parts are appropriately marked as such (e.g. ”authored by”).

Team work restrictions

- WP 1 can be done in a team of two students. WP 2 must be done individually.
- The ideas and investigations to realize WP 2 must be individual.

Please note that grading will be more strict than in winter due to the similarity of both tasks.

Literatur

[1] Christian Gück, Cyriana M. A. Roelofs, and Stefan Faulstich, CARE to Compare: A Real-World Benchmark Dataset for Early Fault Detection in Wind Turbine Data, arXiv:2404.10320v2 (2024)

Ostbayerische Technische Hochschule
Amberg-Weiden

Academic Integrity Declaration for the Project Work

Summer 2025
Prof. Dr. Patrick Levi

Surname, First Name:

Student Number:

Academic integrity declaration for examinations

Herewith I declare that we created this project work by ourself. Individual parts are marked in code and documentation correctly. I declare that I created individual parts of this project work by myself. All used material and references are decleared in the project work.

Place, date, signature