Ostbayerische Technische Hochschule
Amberg-Weiden

Project Work Machine Learning
Winter 2024

Prof. Dr. Patrick Levi

2. Dezember 2024

Start: 09.12.2024 (via Moodle)
Deadline: 09.01.2025 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 – Wind Turbine Data

In [1] the authors introduce a wind turbine dataset. It contains data recordings on three wind farms (A, B, and C) and information on the state of the corresponding wind turbines. The dataset shall help researchers to examine suitable anomaly detection methods to predict state degradation in wind turbines.

Imagine you want to conduct a reserch work (e.g. a thesis) about machine learning methods applied to this dataset. To be able to do research on the data you first have to analyze it and try some models on it which you can use as benchmarks later.

The goal is to find machine learning models which can predict the occurrence of the events in the dataset reasonably before the event begins.

Consider your work here as the beginning of a research work. You will investigate the dataset and create some benchmark analyses. For your work, you only use data for wind farm B. In the dataset you also find sensor data, labels that mark events, and also some metadata. The goal of the work is pro predict the anomaly events in the data as early as possible.

Work packages

Your project consists of the following work packages (WP).

- WP 1 - Data analysis: Analyze the datasets to understand its structure and contents. You shall be able to work with the dataset afterwards. Among other things, describe the data set, find possible problems, issues, or remarkable aspects which could be relevant for machine learning algorithms later. Create meaningful visualizations of the data. Keep only the most relevant figures in the main text of your documentation.
  Perform a useful preprocessing with respect to the following ML tasks and extract suitable features. Include the metadata into your considerations. Present your analysis and data processing in the documentation. Consider the reader of your documentation does not know the dataset.
  Develop a strategy for the subsequent work packages. Your approach shall be well justified based on relevant scientific literature.
- WP 2 - Clustering: Assess whether a clustering approach can help you subdividing time stamps into (developing) anomalies and normal behavior. Select one clustering algorithm per team member. Measure the success and try to optimize the algorithm. If needed, also adapt your data preprocessing, feature engineering.
- WP 3 - Change point detection: Assess whether you can detect the change of state (normal/anomaly) with a change point detection approach. Use the ruptures package (`https://github.com/deepcharles/ruptures`). Select one detection algorithm per team member. Measure the success and try to optimize the algorithm. If needed, also adapt your data preprocessing and fea-

ture engineering.
Analyse and compare your compare your algorithms from WP2 and WP3.

Your documentation shall present your results in a structured, scientific way. You could use a structure like introduction, relevant literature, method, experiments, results, discussion. Make sure, that individual and team parts are appropriately marked as such (”authored by”), do not write parts from individual tasks (e.g. methods, results) as a team work.

Team requirements

- WP 1 can be done in a team of two students. All other WPs must be done individually.
- If you work in a team, align on the algorithms you analyze in WPs 2 and 3. Make sure that every team member analyzes a different algorithm.

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

Winter 2024
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