Mining Massive Data (SS2025)

Pen & Paper 4

Availability date: 11.05.2025
Exercise session: 22.05.2025 (attendance is mandatory)
Checkmark & upload date: 21.05.2025, 23:00

Number of tasks: 4
Maximum achievable points: 30 points

Submission & Grading

This pen & paper exercise is graded. You have to indicate which tasks you have solved and upload your solutions by the *checkmark & upload deadline* (see the header) on the course's Moodle webpage [3]. To be awarded the points for these tasks, you have to be present in the respective pen & paper session, be ready to present your solutions, and, if you are selected, provide a concise and comprehensible presentation of your solution which demonstrates your understanding of the problem & solution. The selection process will mainly be uniformly at random. **If you fail to be present in the session or are not able to explain your solution, you will not be awarded any points for the whole pen & paper session.**

Because of the large number of students enrolled in the course, we do not provide individual feedback on your solutions. If you want feedback on your solutions, contact us via email (see the contact details below).

Questions

If you have any questions, please ask them in the respective Moodle forum. If you don't get an answer from your colleagues or us within 48 hours, reach out to us via email (ensure that you start the subject of your email with [MMD25]):

- Timothée Schmude (timothee.schmude@univie.ac.at)

Tasks

Task 1: Conceptual Foundations

(maximum achievable points: 10 points)

(a) **[2 point(s)]** Define 'explanation' and 'justification', state their differences, and apply both definitions to a self-chosen example.

(b) **[4 point(s)]** For the following explanations methods, provide a description, contrast them, and indicate how each relates to Speith et al. 2022's taxonomy wrt scope, stage, and result:

a LIME vs. SHAP

b Partial Dependence Plots (PDP) vs. Accumulated Local Effects (ALE)

(c) **[4 point(s)]** In your own words, describe the four key principles of everyday explanations as stated by Miller 2019. Provide two examples, one of a white-box and one of black-box model, and use them to discuss the principles' limitations and where they might interfere.

Task 2: Explanation Method Selection

(maximum achievable points: 5 points)

[5 point(s)] You're working on a regression model that calculates whether patients are at risk of cardiac diseases within the next year. The model is trained on electronic health data and uses both categorical and numerical features. The final system will be deployed in a hospital as a decision-making assistance tool.

- Choose a non-technical stakeholder and describe at least 5 stakeholder attributes that might be relevant for the choice of explanation, create 5 why-questions that this stakeholder might have
- Draw a concept map or flowchart that documents how you would select an explainability method for this stakeholder (you can take inspiration but not reproduce the Retzlaff 2024 flowchart). Consider model type, stakeholder needs, and explanation scope, format, presentation, and evaluation.
- Justify your choices and explain how your method meets the explanation quality criteria of a self-chosen reference framework from the literature, briefly state why you chose this evaluation framework.

Task 3: Explanation critique

(maximum achievable points: 5 points)

Find an example of an explanation from a real-world ML application. This can be from a research paper, blog post by a tech company, case study, etc. Do not use the exact explanations from the lecture.

(a) [1 point(s)] Summarize the application, the explanation method used, the audience of the explanations, and the evaluation (if any).

(b) [2 point(s)] Discuss if the method was appropriate for the model and audience, if any assumptions or limitations were overlooked, and if another method might have worked equally well or better.

(c) [2 point(s)] If you would change something about the explanation method, what would it be and why?

Base your reasoning for (b) and (c) on the references given in the lectures or provide own references.

Task 4: Implementing explanations

(maximum achievable points: 10 points)

For this task you will implement explanations. Find the open dataset "**Predict Students' Dropout and Academic Success**" on the Machine Learning Repository: tinyurl.com/4fxerra4

The dataset includes information on academic background, demographics, and social-economic factors. Its intended use was the prediction of student dropouts and using these predictions to put measures in place that reduce dropouts.

(a) [2 point(s)] Briefly describe the dataset in your own words. Create three stakeholder personas and questions that these stakeholders might have about the data.

(b) [5 point(s)] Answer these questions using either a white-box model or an explainable black-box model. Be sure to adapt the explanations to the stakeholders and describe this process (or, if not possible, how you *would* adapt them).

(c) [3 point(s)] Describe how you created these explanations and how you would evaluate them in a real-world setting.

To solve this task, you can use existing explainability libraries such as interpretML (interpret.ml/docs/index.html). Please also make sure that you document all three tasks so that they are presentable in the P&P session (slides, visualizations, etc.).