

# MACHINE LEARNING

Project work



Markus Mayer  
markus.mayer@th-deg.de

## OVERVIEW

The project work in general:

- ▶ Two parts:
  - ▶ Presentation
  - ▶ Written work
- ▶ Presentations take place on friday, 2023-06-30 (reserved in case needed: 2023-07-07)
- ▶ Written work deadline: Thursday, 2023-07-06
- ▶ Deadline for forming groups / dataset opinion: 2023-05-13
- ▶ Recommended: Groups of 2
- ▶ Grading:  $\sim 1/3$  presentation,  $\sim 2/3$  written work
- ▶ 60 points in total, 30 required to pass (4.0 grade)

## TOPICS

- ▶ Find a suitable dataset on [kaggle](#)
- ▶ The data should allow for **classification** experiments
- ▶ Perfect solutions with traditional ML methods should NOT be present (or at least none that can be copy-pasted and have a complete evaluation).
- ▶ The license/rights of the dataset should be clear and in the hands of the person who posted it to kaggle.
- ▶ You may work on images, or .csv files or anything else.
- ▶ Take care: You'll need to motivate your choice.
- ▶ Don't make your choice too hard to work on (e.g. a too large dataset is counterproductive), nor too easy (you'll need some samples to do proper ML).
- ▶ Choosing a proper dataset is part of your project.
- ▶ No dataset should be used twice! Talk to your colleagues!

## PRESENTATION (1)

- ▶ The presentations take place on 2023-06-30 or 2023-07-07.
- ▶ In presence. On both dates. **Absence requires a doctors certificate** (if you want to pass).
- ▶ 10 minutes + 5min questions.
- ▶ **The presentations have to be handed in via iLearn on 2023-06-28 latest.** A single PDF or PowerPoint (take care of the file size).

## PRESENTATION (2)

- ▶ Contents:
  - ▶ Topic explanation and problem motivation
  - ▶ Preliminary method
  - ▶ Preliminary results
  - ▶ Outlook
- ▶ Grading:
  - ▶ 10 points presentation style and form (e.g. oral presentation, slide design).
  - ▶ 10 points familiarity with the topic (motivation of topic, answering questions to methodology etc.)
  - ▶ NOT: Perfect, full fledged solution. But the general way must be clear and a part of the results should already be computed.

## WRITTEN WORK

- ▶ Outline:
  - ▶ Introduction
  - ▶ Data
  - ▶ Method
  - ▶ Results and discussion
  - ▶ Summary
- ▶ 12-18 pages in the LaTeX template provided, including figures.
- ▶ Be concise and selective.
- ▶ Grading:
  - ▶ 10 points form (correct citations at proper places, correct links, e.g. to dataset, proper captions for figures...)
  - ▶ 10 points environment and problem motivation (e.g. introduction and data section, linkage to related work etc.)
  - ▶ 20 points evaluation method and discussion (usage of our toolset, parameter selection, construction of the evaluation, presentation of results, conclusions made)

## WRITTEN WORK: REQUIRED DETAILS (1)

- ▶ **Form:** Use the LaTeX template. Correct citations. Dataset is linked. Figures have captions that are explanatory and need as little knowledge about the text as possible. 12-18 pages including figures, no font sizes changed. Use of scientific language (e.g. no “I have...”). Graphs have to be interpretable in grayscale printing.
- ▶ **Introduction:** Data origins. Used in scientific works? What did others find out? State the problem. Why is the problem interesting?
- ▶ **Data:** Explanatory plots or example samples from the data. Motivation of problem statement. Exclusion/inclusions. Separation into train and test data.

## WRITTEN WORK: REQUIRED DETAILS (2)

- ▶ **Method:** Describe all steps of the classification evaluation. A few words on preliminary tests performed. Parameters that are studied. Order of experiments. Selection of meaningful performance measures.
- ▶ **Results and discussion:** Classification experiment results in the order described in “Method”. What do we learn from the results? Is there a surprise?
- ▶ **Summary:** Everything (from introduction to discussion) summed up in a half to at max. one page (no word more).
- ▶ Hand in via iLearn task. A single PDF! No code etc.
- ▶ I don't look at the written works before the deadline (but specific questions are of course valid)!



## GRADING

Category	Maximum
Pres. Style and Form	10
Pres. Familiarity with Topic	10
Written Form	10
Written Environment and Motivation	10
Written Method and Discussion	20

- ▶ At maximum 6 points added to the score due to voluntary exercises
- ▶ Temporary mark =  $(70 - \text{score}) / 10$
- ▶ Temporary mark rounded down to X.0, X.3, X.7 grades

## WORKING GROUPS AND PROJECT START

- ▶ You can form groups of 2 (which is recommended). No more than 2.
- ▶ Groups have to be announced to me **in an eMail** to make the respective iLearn entries.
- ▶ **Deadline for group announcement: 2023-05-13.**
- ▶ If I did not receive a group eMail by the deadline, you're expected to work alone. Same regulations apply. No iLearn changes after the deadline.
- ▶ Also: You may **send me your chosen dataset** until the beforementioned deadline (as a link in an eMail) to check if it is valid (at first glance). By that you also avoid the usage of a duplicate dataset.

## FINAL NOTES

- ▶ The project work is ongoing work *throughout the semester*.
- ▶ Starting 1 week before the presentation most likely yields suboptimal results.
- ▶ The written work does not have to be finished before the presentation (e.g. the feedback from the presentation can still influence your written work).
- ▶ If you have questions: Ask anytime! (eMail, in the lecture, on iLearn etherpads)
- ▶ If you need support: First ask your colleagues. If there are open questions or no solution is found - see above.

## MORE FINAL NOTES

- ▶ **You may only use the tools that we discuss in the lecture** - I need comparability to have a fair grading.
- ▶ Why the “Perfect solutions with traditional ML methods should NOT be present” rule? The project work is an **exam**. If I find *almost everything* you did in a Kaggle notebook/discussion, there are two possibilities:
  - ▶ You cited the source correctly and stated that someone else did the same evaluations before. That’s OK - to a certain extend. Without me seeing that there is any original work in the project report, a good grade (or even passing), is not possible.
  - ▶ You (somehow - however that may happen...) forget to cite the source of the evaluations. Then you created a duplicate. A duplicate automatically leads to a “not passed” grade.

So this rule is present for your own safety and to allow for a lot of good, original works that are a joy to read and grade!