

EARIN - project guidelines

Instructor: Daniel Marczak

General information

The project is done in pairs and has three stages. Maximum number of points from the project is 25. The project timeline, with points for each stage, is as follows:

- **Preliminary assumptions** deadline is **May 12th**. (0-3 points)
- **Midterm solution** should be delivered by **May 29th**. (0-5 points)
- **Final solution** deadline is **June 11th**. (0-17 points)

Delay in handing in each stage results in a penalty of 20% total points available from the stage. After each deadline, we will have project meetings on MS Teams. **The meetings are mandatory to complete the project**, and we will do them in a week following handing in of each stage. Please contact me to arrange the meeting. In case of any questions about projects, either ask me directly on Teams or write on the Projects channel.

Technical details

- Solutions should be implemented in Python (Jupyter Notebooks are allowed).
- Please ensure that your code adheres to basic standards of lean coding in accordance to PEP 8 and add instruction on how to run the code.
- The final report should follow the structure of a scientific paper. Please use template from IEEE / Springer / Nature / NeurIPS etc for the report.
- The code, alongside the reports, should be stored on faculty GitLab. **Please send the links to repository as a submissions for each stage of the project.**

Assessment criteria

Preliminary report (max 3 points)

Preliminary report should be a pdf file containing:

- Description of the task that you are working on, eg: what is the goal of the project and how do you frame task you want to solve (regression, binary classification, multilabel classification).
- Description of the dataset:
 - Does it require any preprocessing or cleaning?
 - Are the labels in the dataset balanced? How do you plan to address the data imbalance in the project?
 - How do you plan to split the data for the project?
- How would you evaluate the performance of your solution? Consider both the metrics and data splits, and look for an upper bound of your solution, eg. what is the best score on Kaggle.
- Description of your solution:
 - Describe briefly usual methods used for the problem.

- Select two or three of them that you want to implement and compare. Add more detailed description of those methods and data preprocessing required for them. Consider how the preprocessing applies to your dataset.

Midterm solution (max 5 points)

At this stage, you should present your progress, eg. what have you already implemented and how it works. The grading of this stage will take into an account:

- Your progress and in coding and evaluation of the method - at least one of selected algorithms should be implemented at this stage. **(1 point)**
- Intermediate results and their presentation. Please add the results to the project report. The results should be presented accordingly, as plots or tables. **Make sure your results are reproducible** by adding seed your code. **(1 point)**
- Presentation and discussion on the challenges, results and findings during the project so far, and potential corrections to the previous assumptions. This should also be included in the report. **(2 points)**
- Discussion on the plans for finishing the project, eg. what experimental results do you want to show in the final report. **(1 point)**.

Final solution (max 17 points)

Final solution should be the code and report in form of scientific paper. The grading of this part will be:

- Submission of the code for your solution and correctness of the code. **(6 points)**
- Final report should contain the information from the previous stages. The grading of report will be as follows:
 - Proper form of the report following the format of scientific paper (introduction, description of the data and algorithms that you will use, intermediate results and insights from the midterm solution, final experimental results, conclusions). **Don't include the code in your report. (2 points)**
 - Experiments, presentation and the discussion of the results. **(5 points)**
 - Comparison of the algorithms that you implemented, based on the results from experimental section. **(2 points)**
 - Final conclusions - which algorithm performed the best, is the solution satisfactory and how could it potentially be improved? **(2 points)**