Relevant links:

- operation
    - (in case link does not work: https://github.com/REMLA24-Team-5/operation/tree/a1)
- model training
    - (in case link does not work: https://github.com/REMLA24-Team-5/Model-Training/tree/a1)

Comments for A1:

Below we specify all the things we have implemented for each section. The model training repository contains all the code (link above).

Project best practices:

- we have split the code into different files and structured the repository using the data science template of cookie cutter
- Poetry allows for the installation of the correct dependencies.
- the dataset is automatically downloaded as part of the dvc pipeline
- exploratory code has been removed from Python files and is only visible in the `notebooks` folder
- the README includes the steps as to how to use our repository
- all decisions are properly documented in the README

Pipeline Management:

- the pipeline can be run with `dvc repro`
- a remote storage has been configured using Google Drive
- DVC is used to report metrics and to keep track of different experiments/models.
- Different metrics are reported beyond model correctness

Code Quality

- The project follows code quality best practices
- Pylint yields a perfect score and was properly configured.
- Configuration decisions follow ML conventions and are properly motivated in the Readme file
- Both dslinter (extension of pylint) and Bandit are configured and code quality information is displayed in the Readme
- In the README we critically analyse the linter rules, and propose changes that fit the ML project