

EDFB: Prep for Hackathon

Duration: 180 minutes

Dataset provided: `df_raw_stratified_sampled.csv`

Instructions

- You will practice the complete machine learning pipeline on a small credit dataset.
- You can decide if you want to work in groups or individually. Note: the actual hackathon is individual (!)
- You can use any programming language for the task (R, Python ...) and you can use any notebook (R Markdown, Google Colab ...)
- Submit a notebook containing code + explanations.

Tasks

- Your task is to explore the data, build one simple classifier, and interpret the results - the same workflow you will follow during the actual hackathon.
- Your the target variable is:
 - `default = 1` → loan default
 - `default = 0` → loan repaid
- Tasks to follow:
 - Explore the data set; inspect its shape & quality.
 - Identify at least *two* interesting feature patterns.
 - Decide whether any data transformation steps are needed.
 - Choose a classification method to use. Elaborate the choice. Note: it is important to put this discussion into perspective of the actual data and use case. Generic arguments on the pros and cons linked to certain classification techniques will be awarded zero points.
 - Evaluate the classifier. Please use metrics and concepts that we have seen in class.

How the hackathon output will be evaluated?

- Does the notebook submitted works without an error?
- Does it contain all tasks?
- Are the technical steps (analysis, transformations, modeling, evaluation) carried out correctly and logically?
- Are explanations, reasoning, and presentation clear, structured, and easy to follow?