MLOps project

The aim of the project is to simulate the realworld process of deploying machine learning models, using the concepts that we have discussed during the classes.

Project Deliverables

Report with maximum of 6 pages:

- Explain why you choose that data and what you try to achive. Define **your sucess metrics.**
- Project planning: Divide your project in several sprints with clear goals, e.g., Sprint 1: 3 days, the goal was to explore the data, Sprint 2: 1 day, the goal was to build feature engineeing pipeline, etc.
- **Results and conclusions** from data exploration and data modelling (plots, feature importance, explainability).
- Since this is a proof of concept, discuss how this would be implemented in production and what are the **advantages of the technologies used, risks and possible mitigations**, e.g. “we are using only Pandas, so if there the amount of data sclales up, our pipeline will no be efficient. We propose more x weeks to build in Spark, as a mitigation solution.
- **List of the packages and versions** used for the project.

Code for generating your pipeline:

- Use Kedro organization and modular code. You can keept your exploration notebooks in your appropriate folder.
- Try to include the following components in your pipeline:
    1. **Unit data tests**: you can use one of the tools from the class or your own solutions, but it is important to have several asserts for the data quality.
    2. MLflow for **experimentation and model versioning**.
    3. **Data drift evaluation**: if you build a pipeline to test a sample of data of your strongest model, include this component as well. You can play with your sample if you want to generate drift, or see how the mtetrics would change if drift happens.
    4. Try to **build tests** for your relevant functions and pipelines.

In the end, everyond should be able to run you pipeline and to produce the same results. Projects will be graded based on the **quality of the report, code and creativity shown** for using the technologies.

Deadline 16/06/2023 with the report, zip of the code with a sample of data to run or a Git link.