ML Engineer

Case Study

General conditions

The primary purpose of the case study is to showcase your technical skills, rather than your modeling skills.

The estimated time to complete the tasks is approximately 3-4 hours.

Please upload your solutions, such as Python, Docker, and CI/CD files, to a publicly accessible repository of your choice and share it with us one day before the presentation.

If you have any questions, feel free to reach out to me via the following email:

gianni.balistreri@covestro.com

I'll be happy to answer your questions.

Have fun and good luck! ☺

Scenario

Case study description

The AI Engineering team is responsible for operationalizing specialized AI models, which are developed by the Data Scientists, in order to generate economic value for the company.

You are working on a use case with a Data Scientist and a Data Engineer to operationalize a model that predicts whether a room is occupied.

Based on the prediction, devices controlling room temperature, air conditioning, and other systems are enabled or disabled.

Task 1

Model training

- The Data Scientist has defined the following model:
  - Feature Engineering: One-hot encoding of extracted categorical features from the datetime feature
  - Modeling: XGBoost Classifier
  - Hyperparameter Optimization: Random Search
  - Model Evaluation: ROC-AUC

- Your first task (as a MLE):
  - Sketch the individual workflow steps of the training pipeline (either graphically or verbally)
  - Implement a simple version of the training pipeline in python using the attached dataset
  - Structure your code accordingly

  - ATTENTION: Model performance is not the priority! Please focus on the technical aspects of the task.

Task 2

Inference

- Your second task (as a MLE):
  - Build and implement a model endpoint to generate event-driven predictions

Task 3

CI / CD / CT

- Your third task (as a MLE):
  - Build a CI/CD pipeline (tool of your choice) to:
    1. Build, test and push training pipeline steps to a container repository
    2. Build, test and push model endpoint service to a container repository
  - Verbally describe how you would implement continuous training (CT)

Happy developing ☺