Model Training Batch Job
CSC 6605 ML Production Systems

Instructions

In most cases, intelligent systems need their models retrained periodically to counteract data distribution shift. Your team will design and implement a model training and evaluation pipeline. You will then evaluate your system using the MLOps maturity model and suggest improvements to advance the maturity of your system. You will submit a written report that covers all three parts of this assignment.

Part I: Model Training and Evaluation Batch Job

Design and implement a batch that will periodically:

1. Grab cleaned and joined data from the database
2. Perform a train-test split
3. Train the feature extraction and model pipeline on the training data
4. Evaluate the model using the testing data
5. Re-train the pipeline using all of the data
6. Deposit the trained pipeline into a PostgreSQL table (that you will design and create) along with evaluation metrics like RMSE. Use a byte or bytea column for storing the serialized pipeline.

Ensure that your code is in your GitHub repository with appropriate documentation, that your pipeline is running in a Docker container on your VM, and that records are being inserted into the table.

In your report, describe your design (including the table schema), explain your decisions, and include screenshots showing the containers running and data in the table.

Part II: MLOps Maturity Evaluation

Review the [Microsoft MLOps maturity model](#) presented in class. The model addresses maturity in four areas (people, model creation, model release, and application integration) with five levels of maturity. In your report, provide an evaluation of the system we implemented and processes we used this term across the four areas. For each area, determine the current maturity level, provide at least one paragraph explaining your assessment, and provide a second paragraph describing an action you can take, technology you can employ, or component you could implement to advance the maturity by one level.

Part III: Class Reflection

Each person should write a reflection on the class that will be included in the overall report. The reflection should cover both your performance in the class as well as the course content, design, and delivery. The reflection responses can be organized by question (with one paragraph per person) or by person (with one paragraph per question). Answer the following questions:

1. What did you learn that was new to you?
2. What went well?
3. What could have gone better?
4. Do you think the course content will be useful to your future endeavors?
5. What wasn't covered in the course that you would like to learn more about?

This reflection is being used to improve the course, so constructive criticism of how I designed and ran the course is welcomed and expected.

Submission Instructions
Submit a PDF of your group's report to Canvas. One person should submit for the entire group.

Rubric

Description	Percentage
Written Report • Report is written in a professional manner using proper grammar and spelling. • Report is a useful standalone document that can be shared with a business partner.	20%
MLOps Evaluation • Maturity levels are given for each of the four areas. • The assigned maturity levels are justified and explained appropriately. • An action for increasing the maturity level for each area is described.	60%
Class Reflection • Each question is answered with a paragraph-length response. • The student performed a meaningful, thoughtful reflection.	20%