**MSE\_DDM501\_MLOPS**

Trần Trọng Hiếu – MSE23185

**Báo cáo**

Các bước thực hiện dự án MLFlow

1. Xây dựng pipeline huấn luyện và tracking với MLflow

- Tạo file models/mlflow\_models.py để sinh dữ liệu giả lập (synthetic) với scikit-learn.

- Cài đặt huấn luyện model Logistic Regression, log các tham số và metric lên MLflow.

- Thực hiện tuning tham số cho Logistic Regression (C).

- Thêm huấn luyện và tuning cho Random Forest (n\_estimators).

- Thêm huấn luyện và tuning cho XGBoost (n\_estimators).

- Đăng ký (register) model tốt nhất lên MLflow Model Registry.

- Lưu model tốt nhất ra thư mục outputs /best\_model/.

2. Xây dựng Flask API phục vụ dự đoán

- Tạo file app/app.py để load model đã lưu và xây dựng API `/predict`.

- Xây dựng route `/` trả về giao diện web đơn giản (HTML form).

- Đảm bảo API nhận input 20 features, trả về prediction và probability.

3. Xây dựng giao diện web đơn giản

- Tạo file app/template/index.html cho phép nhập 20 features và submit lên API.

- Hiển thị kết quả dự đoán và xác suất trên giao diện.

4. Đóng gói các thành phần với Docker

- Viết dockerfile.train để build image cho job training.

- Viết dockerfile.app để build image cho Flask API.

- Viết dockerfile.mlflowui để build image cho MLflow UI.

- Viết docker-compose.yml để orchestration các service: mlflow-ui, train, app.

5. Thiết lập CI/CD với GitLab

- Viết pipeline trong .gitlab-ci.yml để build các Docker image.

- Thiết lập job test: chạy train và test API trong môi trường CI.

- Đảm bảo pipeline chỉ chạy trên branch `main`.

6. Tài liệu và hoàn thiện

- Viết README.md hướng dẫn sử dụng, cấu trúc dự án, các bước chạy thử.

- Đảm bảo các file requirements, env, conda.yaml được sinh ra đầy đủ khi export model.

- Kiểm tra lại quyền ghi cho các volume Docker (`mlruns`, `outputs`).