



DML-004 – GUÍA DE LABORATORIO 10

Añadiendo MLflow al pipeline de Heart Disease (SDK v2)

1) Objetivos

- Integrar MLflow Tracking al pipeline existente (8 pasos).
- Registrar parámetros, métricas y artefactos por componente.
- Visualizar y comparar runs en Azure ML Studio.
- (Opcional) Registrar el modelo entrenado en Model Registry (sin desplegar).

2) Prerrequisitos

- Haber completado el Lab 9 con el pipeline de 8 componentes.
- Workspace de Azure ML operativo (compute existente).

3) Cambios mínimos al proyecto

- Reutilizar la estructura del Lab 9.
- Solo se agregan dependencias y código de MLflow en train, score y eval.

3.1 Estructura

```
heart-pipeline/  
├─ components/  
│   ├── select_cols/ ... (sin MLflow)  
│   ├── impute/     ... (sin MLflow)  
│   ├── encode/     ... (sin MLflow)  
│   ├── scale/      ... (sin MLflow)  
│   ├── split/      ... (sin MLflow)  
│   ├── train_lr/   ← añade MLflow  
│   ├── score/      ← opcional: MLflow "best-effort"  
│   └─ eval/        ← añade MLflow (métricas + artefactos)  
├─ envs/  
│   └─ conda.yml    ← agrega azureml-mlflow  
├─ data/  
└─ pipelines/
```



3.2 Environment (envs/conda.yml)

name: heart-mlops

channels: [conda-forge, defaults]

dependencies:

- python=3.10
- pip
- pandas
- numpy
- scikit-learn
- mlflow
- matplotlib
- pip:
 - azure-identity
 - azure-ai-ml
 - azureml-mlflow

4) Integración de MLflow por componente

- train_lr: log_param, log_metric, log_model, opcional register_model.
- score: opcional log_artifact de predicciones.
- eval: log métricas, curvas ROC/PR como series, artifacts (json/png).

5) Pipeline (SDK v2)

- Reutilizar el mismo DAG de 8 pasos del Lab 9.
- Asignar experiment_name y tags para identificar runs.

6) Ejecución

- 1. Subir/actualizar componentes y conda.yml.
- 2. Ejecutar pipeline.
- 3. Ver métricas y artifacts en Azure ML Studio.
- 4. (Opcional) Ver modelo en Model Registry.



7) Entregables

- Captura de Metrics con métricas y curvas.
- Captura de Artifacts (metrics.json, plots).
- Captura del DAG.
- (Opcional) Modelo en Registry.
- Informe corto comparando runs.

8) Troubleshooting

- Error UnsupportedModelRegistryStoreURI: falta azureml-mlflow.
- Runs sin métricas: falta nested=True.
- Curvas no aparecen: loguear series con step.
- Artefactos no visibles: revisar rutas y log_artifact.