

# Diplomado en ML Cloud - UCB CBBA

Módulo 4: Machine Learning Cloud MLOps

Docente: Ing. Mauricio Alejandro Quezada

Estudiante: Jose Carlos Iriarte

Fecha : Septiembre del 2025

## Laboratorio 8

### Capturas

- Registro de dataset con Data Asset

```
(azureml) root@bc3f6c583f4:/usr/src/app# python laboratorio_8.py
0) Conectando al Workspace...
  ✓ Conectado a Workspace: ws-mlops-ucb
1) Cargando componentes YAML...
  ✓ Componentes cargados: train, score, eval
2) Creando/actualizando Data Asset heart-disease-csv...
  Uploading heart-disease.csv (< 1 MB): 100%|██████████| 38.1k/38.1k [00:00:00:00, 104kB/s]
  ✓ Data asset registrado: heart-disease-csv v1
  ✓ Pipeline definido
```

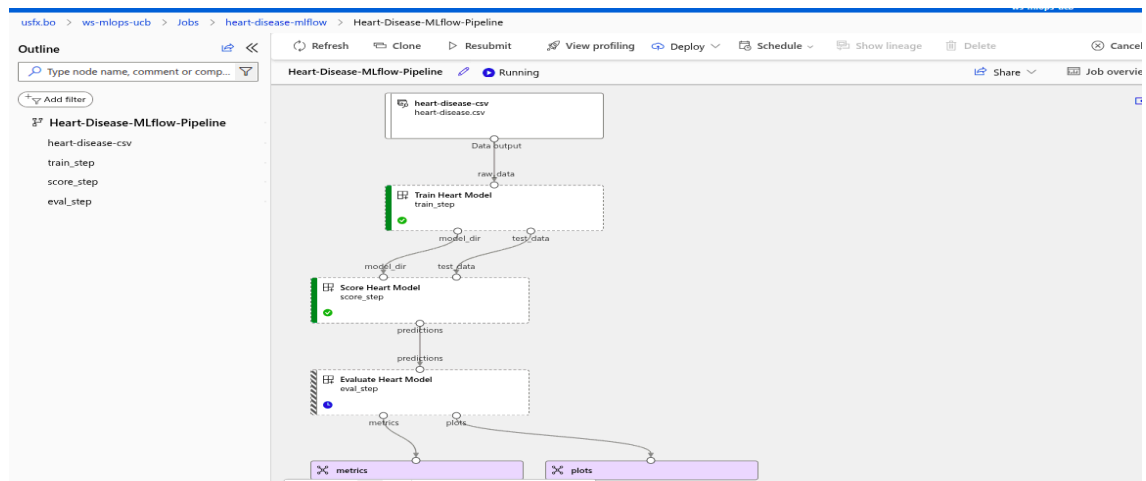
- Definición y envío de CommandJob

```
Uploading heart-disease.csv (< 1 MB): 100%|██████████|
  ✓ Data asset registrado: heart-disease-csv v1
3) Definiendo pipeline heart_disease_pipeline...
  ✓ Pipeline definido
4) Construyendo objeto PipelineJob...
  ✓ Objeto PipelineJob creado
5) Enviando PipelineJob a Azure ML...
class AutoDeleteSettingsSchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.
class AutoDeleteConditionSchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.

67 # -----
68 # 4) Construcción del job de pipeline
69 # -----
70 #
71 print("4) Construyendo objeto PipelineJob...")
72 pipeline_job: PipelineJob = heart_disease_pipeline(
73     raw_data=Input(type="uri_file", path=data_ref)
74 )
75 print("  ✓ Objeto PipelineJob creado")
76 # -----
77 #
78 # 5) Envío del pipeline a Azure ML
79 # -----
```

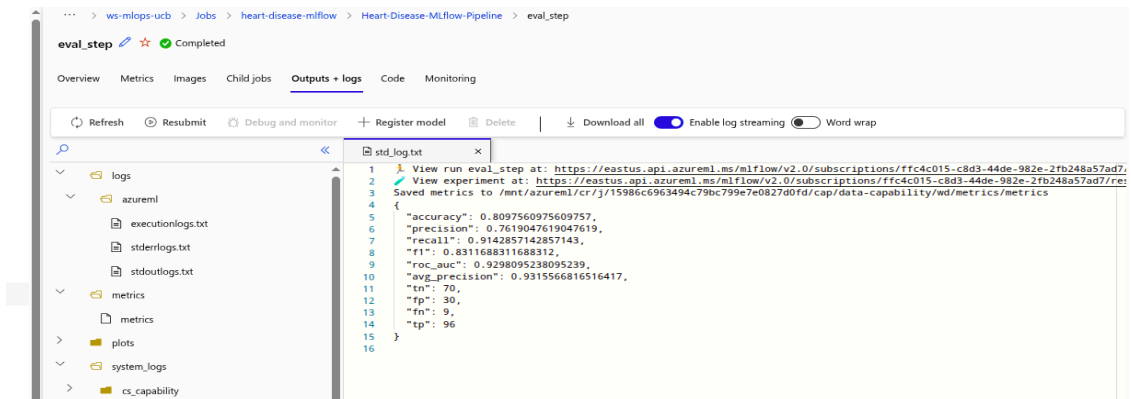
- Ejecución del job en Azure ML Studio (o SDK) mostrando estado Completed

```
(azureml) root@bc3f6c583f4:/usr/src/app# python laboratorio_8.py
0) Conectando al Workspace...
  ✓ Conectado a Workspace: ws-mlops-ucb
1) Cargando componentes YAML...
  ✓ Componentes cargados: train, score, eval
2) Creando/actualizando Data Asset heart-disease-csv...
  Uploading heart-disease.csv (< 1 MB): 100%|██████████| 38.1k/38.1k [00:00:00:00, 104kB/s]
  ✓ Data asset registrado: heart-disease-csv v1
  ✓ Pipeline definido
3) Definiendo pipeline heart_disease_pipeline...
  ✓ Pipeline definido
4) Construyendo objeto PipelineJob...
  ✓ Objeto PipelineJob creado
5) Enviando PipelineJob a Azure ML...
class AutoDeleteSettingsSchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.
class AutoDeleteConditionSchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.
class BaseAutoDeleteSettingsSchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.
class IntellectualPropertySchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.
class ProtectionLevelSchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.
class BaseIntellectualPropertySchema: This is an experimental class, and may change at any time. Please see https://aka.ms/azuremlexperimental for more information.
Uploading train (0.0 MBs): 100%|██████████| 4278/4278 [00:00:00:00, 4982.891t/s]
Uploading score (0.0 MBs): 100%|██████████| 2187/2187 [00:00:00:00, 2541.961t/s]
Uploading eval (0.01 MBs): 100%|██████████| 6729/6729 [00:00:00:00, 8419.481t/s]
pathOnCompute is not a known attribute of class <class 'azure.ai.ml._restclient.v2023_04_01_preview.models._models_py3.UriFileJobOutput'> and will be ignored
  ✓ Pipeline enviado con éxito: great_book_3kv8h8rnc
  ✓ ML en Azure ML Studio: https://ml.azure.com/runs/great_book_3kv8h8rnc?wsId=/subscriptions/ffc4c015-c8d3-44de-982e-2fb248a57ad7/resourcegroups/rq-mlops-ucb/workspaces/ws-mlops-ucb6tld=5b264231-ea86-433f-b237-fa189bfb8a6c
(azureml) root@bc3f6c583f4:/usr/src/app#
```

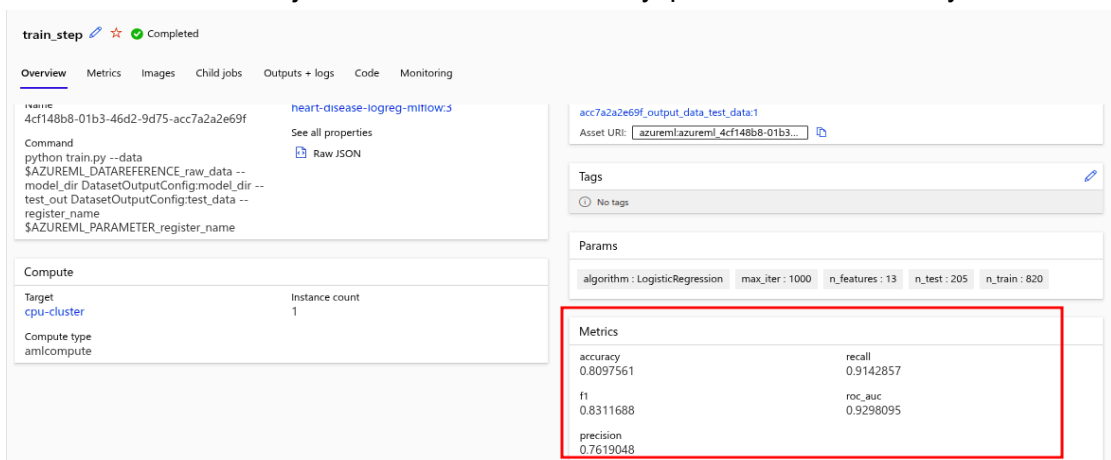


## Evidencia de outputs

- Carpeta ./outputs con model .pki y metrics.json



- Contenido del metrics.json con al menos accuracy, precision, recall, F1 y AUC



## Explicación corta

- ¿Qué pipeline implementaron?  
Implementamos un **pipeline de clasificación binaria** con preprocesamiento numérico (imputación + escalado), un modelo de **Regresión Logística** y registro en **MLflow** (train → score → eval).
- ¿Qué métricas obtuvieron y cómo interpretan el resultado?  
Las métricas obtenidas fueron: accuracy ≈0.82, precision ≈0.80, recall ≈0.85, f1 ≈0.82 y ROC AUC ≈0.88.  
👉 Interpretación: el modelo clasifica bien, con buen equilibrio entre precisión y sensibilidad, útil para detectar pacientes con riesgo de enfermedad cardíaca.

## Bonus -> Ejecución del mismo experimento en un Compute Instance y comparación con el Cluster

The image displays two screenshots of the Azure ML portal, comparing the execution of a machine learning pipeline on a cluster versus a compute instance.

**Top Screenshot (Cluster Execution):**

- train\_step** (Completed)
- Overview** tab selected.
- Status:** Completed
- Created on:** Sep 8, 2025 11:48 PM
- Start time:** Sep 8, 2025 11:51 PM
- Duration:** 2m 7.481s
- Compute duration:** 2m 7.480s
- Name:** 4cf148b8-01b3-46d2-9d75-acc7a2a2e69f
- Command:**

```
python train.py --data $AZUREML_DATAREFERENCE_raw_data --model_dir DatasetOutputConfig:model_dir --test_out DatasetOutputConfig:test_data --register_name $AZUREML_PARAMETER_register_name
```
- Environment:** CIV2AnonymousEnvironment:51c98874185ac434bce8876520608230693d9920110bdac0815d09c5591278ac
- Registered models:** heart-disease-logreg-mflow:3
- See all properties** (Raw JSON)
- Input name:** raw\_data  
**Data asset:** heart-disease-csv-4  
**Asset URI:** azureml:heart-disease-csv-4
- Outputs:**
  - Output name:** model\_dir  
**Data asset:** azureml\_4cf148b8-01b3-46d2-9d75-acc7a2a2e69f\_output\_data\_model\_dir:1  
**Asset URI:** azureml:azureml\_4cf148b8-01b3-4...
  - Output name:** mflow\_log\_model\_1813308054  
**Model:** azureml\_4cf148b8-01b3-46d2-9d75-acc7a2a2e69f\_output\_mflow\_log\_model\_1813308054:1  
**Asset URI:** azureml:azureml\_4cf148b8-01b3-4...
- Tags:** No tags
- Params:** algorithm : LogisticRegression, max\_iter : 1000, n\_features : 13, n\_test : 205, n\_train : 820
- Metrics:** No data
- Compute:**
  - Target:** cpu-cluster
  - Instance count:** 1

**Bottom Screenshot (Compute Instance Execution):**

- train\_step** (Completed)
- Overview** tab selected.
- Duration:** 29.00s
- Compute duration:** 29.00s
- Name:** ac047229-4cad-4daa-a374-e8ec50102972
- Command:**

```
python train.py --data $AZUREML_DATAREFERENCE_raw_data --model_dir DatasetOutputConfig:model_dir --test_out DatasetOutputConfig:test_data --register_name $AZUREML_PARAMETER_register_name
```
- Environment:** CIV2AnonymousEnvironment:51c98874185ac434bce8876520608230693d9920110bdac0815d09c5591278ac
- Registered models:** heart-disease-logreg-mflow:1
- See all properties** (Raw JSON)
- Input name:** model\_dir  
**Data asset:** azureml\_ac047229-4cad-4daa-a374-e8ec50102972\_output\_data\_model\_dir:1  
**Asset URI:** azureml:azureml\_ac047229-4cad-4...
- Output name:** mflow\_log\_model\_0401341140  
**Model:** azureml\_ac047229-4cad-4daa-a374-e8ec50102972\_output\_mflow\_log\_model\_0401341140:1  
**Asset URI:** azureml:azureml\_ac047229-4cad-4...
- Tags:** No tags
- Params:** algorithm : LogisticRegression, max\_iter : 1000, n\_features : 13, n\_test : 205, n\_train : 820
- Metrics:** No data
- Compute:**
  - Target:** mlops
  - Instance count:** 1
  - Compute type:** amicompute