

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

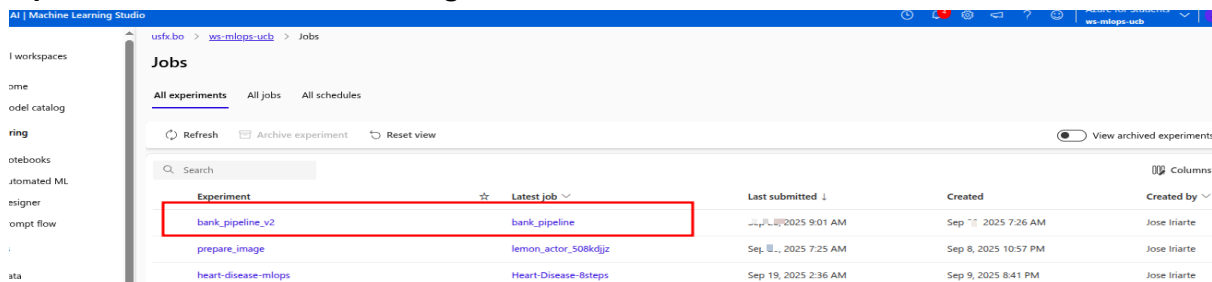
Proyecto Final

Verificar que el dataset cumpla con los requisitos con una cantidad de datos mayor a 10k de registros

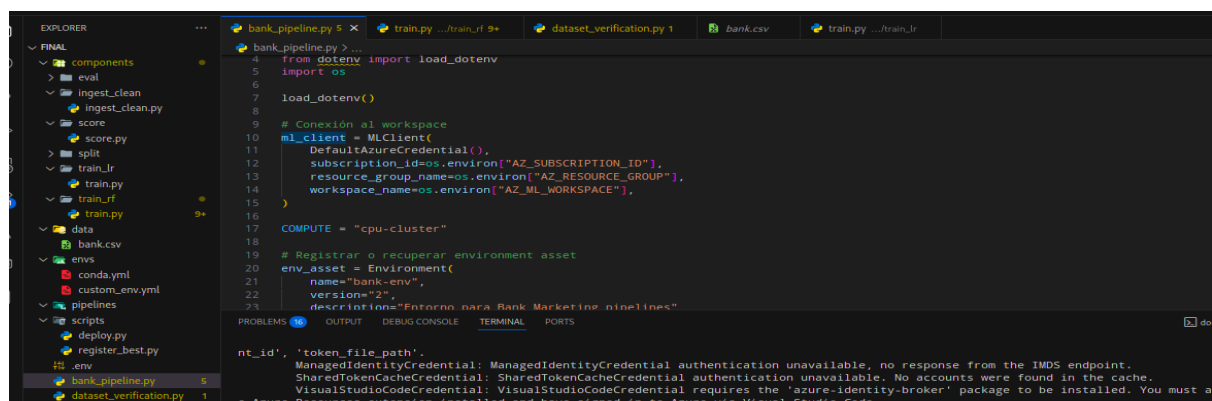


Enlace del dataset: <https://www.kaggle.com/datasets/henriqueyamahata/bank-marketing>

Pipeline de entrenamiento + Cognitive Services



| Experiment | Latest job | Last submitted | Created | Created by |
|---------------------|-----------------------|----------------------|----------------------|--------------|
| bank_pipeline_v2 | bank_pipeline | Sep 19, 2025 9:01 AM | Sep 20, 2025 7:26 AM | Jose Iriarte |
| prepare_image | lemon_actor_S08kdjz | Sep 19, 2025 7:25 AM | Sep 8, 2025 10:57 PM | Jose Iriarte |
| heart-disease-mlops | Heart-Disease-8tsteps | Sep 19, 2025 2:36 AM | Sep 9, 2025 8:41 PM | Jose Iriarte |



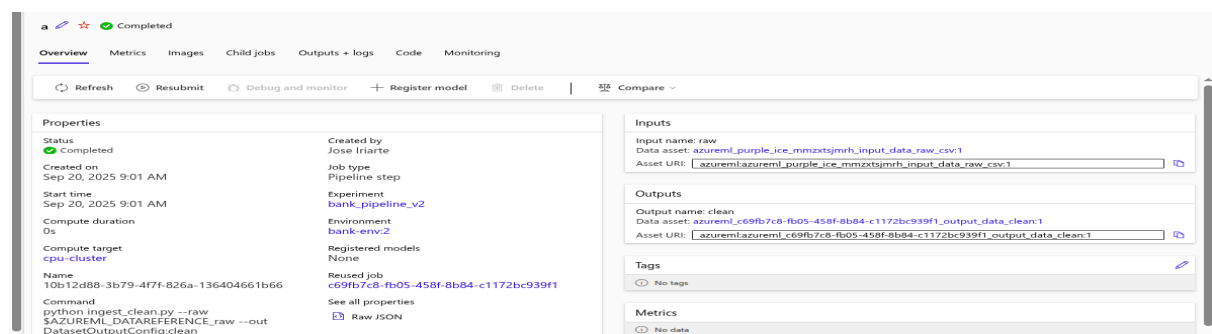
```
from dotenv import load_dotenv
import os
load_dotenv()

# Conexión al workspace
ml_client = MLClient(
    DefaultAzureCredential(),
    subscription_id=os.environ["AZ_SUBSCRIPTION_ID"],
    resource_group_name=os.environ["AZ_RESOURCE_GROUP"],
    workspace_name=os.environ["AZ_ML_WORKSPACE"],
)

COMPUTE = "cpu-cluster"

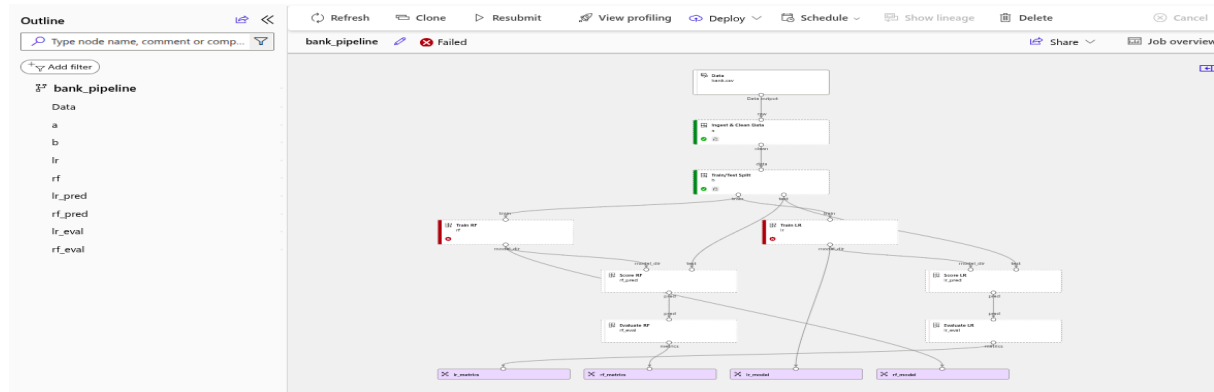
# Registrar o recuperar environment asset
env_asset = Environment(
    name="bank-env",
    version="2",
    descriptions="Entorno para Bank Marketing pipelines"
)

ml_client.register_environment_asset(
    env_asset,
    token_file_path="token_file_path".
)
```



| Properties | Inputs | Outputs |
|---|--|---|
| Status: Completed Created on: Sep 20, 2025 9:01 AM Start time: Sep 20, 2025 9:01 AM Compute duration: 0s Compute target: cpu-cluster Name: 10b12d88-3b79-4f7f-826a-136404661b66 Command: python ingest_clean.py --raw \$AZUREML_DATAREFERENCE_raw --out DatasetOutputConfig:clean | Input name: raw Data asset: azureml_purple_ice_mmxxtsjmth_input_data_raw_csv:1 Asset URI: azureml:azureml_purple_ice_mmxxtsjmth_input_data_raw_csv:1 | Output name: clean Data asset: azureml_c69b7c8-b05-458f-8b84-c1172bc939f1_output_data_clean:1 Asset URI: azureml:azureml_c69b7c8-b05-458f-8b84-c1172bc939f1_output_data_clean:1 |

Capturas del DAG



Conclusiones del Proyecto Bank Marketing en Azure ML

Diseño, entrenamiento, despliegue y documentación de un proyecto completo de Machine Learning en la nube usando Azure ML, demostrando competencias en MLOps aplicado al dataset de marketing bancario.

Resultados

- Dataset: 41,188 registros de clientes bancarios procesados
- Algoritmos evaluados: Logistic Regression, Random Forest

Pipeline Incompleto

Componentes exitosos:

- Ingesta y Limpieza de Datos – Manejo de valores “unknown” y nulos con imputación.
- Feature Engineering – Codificación categórica (OneHotEncoder) y escalado de variables numéricas.
- Model Training – Comparación entre Logistic Regression y Random Forest.

Competencias

MLOps Técnico

- Construcción y orquestación de pipelines en Azure ML SDK v2.
- Preprocesamiento robusto con imputación, encoding y escalado.
- Evaluación comparativa y automática de múltiples algoritmos.
- Registro y versionado de modelos en Azure ML Model Registry.

Arquitectura Cloud

- Configuración de compute clusters para entrenamiento y scoring.
- Manejo de environments y dependencias mediante conda.yml.
- Preparación de endpoints online (real-time) y batch (predicciones masivas).

Buenas Prácticas

- Separación modular de código: ingest, split, train, score, eval.
- Configuración centralizada de parámetros en YAML.
- Testing de componentes de pipeline antes de ejecución final.
- Documentación técnica detallada del flujo de datos, componentes y despliegue.

