



## PROYECTO FINAL – DIPLOMADO EN MACHINE LEARNING CLOUD (MÓDULO 4: MLOPS)

### Objetivo General

Que los estudiantes diseñen, entrenen, desplieguen y documenten un proyecto completo de Machine Learning en la nube, aplicando Azure ML Designer, AutoML, SDK v2, Pipelines, MLflow y Azure Cognitive Services, demostrando las competencias adquiridas en el módulo.

### Lineamientos del Proyecto

#### 1. DATASET DE ELECCIÓN PROPIA

- El dataset debe ser público y de libre acceso (Kaggle, UCI, Open Data, etc.).
- Puede ser tabular.
- El problema debe pertenecer a clasificación o regresión.
- El dataset debe tener al menos 10.000 registros y requerir preprocesamiento.

#### 2. PIPELINE DE ENTRENAMIENTO

- Construcción del pipeline con Azure ML SDK v2.
- Paso de ingesta y limpieza de datos.
- Entrenamiento de modelos (mínimo 2 algoritmos).
- Evaluación y comparación de métricas.
- Registro del modelo en Azure ML Registry.
- Uso opcional de Designer o AutoML para comparación.

#### 3. DESPLIEGUE

- El modelo final debe ser desplegado como un endpoint en Azure:
  - \* Online Endpoint (tiempo real).
  - \* Batch Endpoint (predicciones masivas).
- Deben probar su endpoint con Postman o script en Python.

#### 4. INTEGRACIÓN CON COGNITIVE SERVICES

- Integrar al menos un Cognitive Service en el proyecto (ejemplo: Decision o Text Analytics para tabulares).
- La integración puede ser:
  - \* Preprocesamiento de datos.
  - \* Enriquecimiento del modelo.
  - \* Servicio complementario en el despliegue.



## 5. DOCUMENTACIÓN

- Notebook o Reporte en PDF con:
  - \* Justificación del dataset elegido.
  - \* Diseño del pipeline (incluyendo DAG).
  - \* Comparación de modelos y métricas.
  - \* Evidencia del despliegue (capturas de endpoints y pruebas).
  - \* Integración de Cognitive Services.
- Video demo (5-7 min) mostrando:
  - \* Pipeline en Azure ML.
  - \* Ejecución de pruebas al endpoint.
  - \* Uso del Cognitive Service integrado.



## Evaluación (100%)

Componente	Ponderación
Selección y justificación del dataset	10%
Pipeline en Azure ML SDK/Designer/AutoML	25%
Registro y comparación de modelos	15%
Despliegue en endpoint (Online/Batch)	20%
Integración con Cognitive Services	20%
Documentación y presentación final	10%



## Entregables

1. Repositorio en GitHub/Drive con:
  - Código fuente (notebooks, scripts).
  - Archivos YAML de componentes/pipelines.
  - Reporte final en PDF.
2. Video Demo (link en el repo).
3. Endpoint funcional (si la suscripción lo permite; caso contrario, simulación local documentada).



## **Requerimientos mínimos del Proyecto (10–15)**

1. Dataset (Data Asset)
2. Data Cleaning / Preprocesamiento
3. Feature Engineering
4. Split Data
5. Entrenamiento del Modelo (mínimo 2 algoritmos)
6. Evaluación de Modelos
7. Selección del Best Model
8. Integración con Cognitive Services
9. Registro del Modelo en Azure ML
10. Pipeline en Azure ML (SDK v2)
11. Deploy del Modelo (Online/Batch Endpoint)
12. Pruebas de Consumo (API REST con Postman o Python)
13. MLflow (Tracking y Registry)
14. Documentación Técnica
15. Demo Final en Video