



códigofacilito

— Agenda

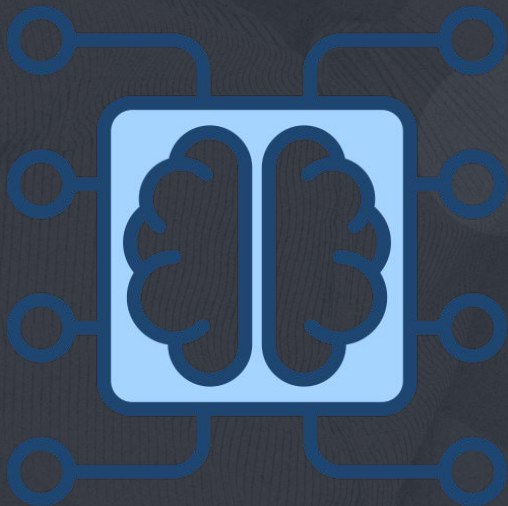
- Repaso de Modelos Machine Learning
- Trabajo con entornos de Azure Machine Learning
- AutoML, Jupyter, MLflow y más

Irving Uribe - ML Engineer / Data
Tech Lead / AI profesor





Introducción a los modelos de Machine Learning



[Complemento] ¿Cómo funciona el aprendizaje automático?

Es un **paradigma** de programación cuya tarea es calcular los **parámetros** de un **modelo** para obtener las **salidas deseadas** según los ejemplos de entrada.

Programación Tradicional



Ejemplos:

$$D = A + B - C$$

$$y = x1 - x2(x3)$$

$$\text{out} = \max(0, \text{inp})$$

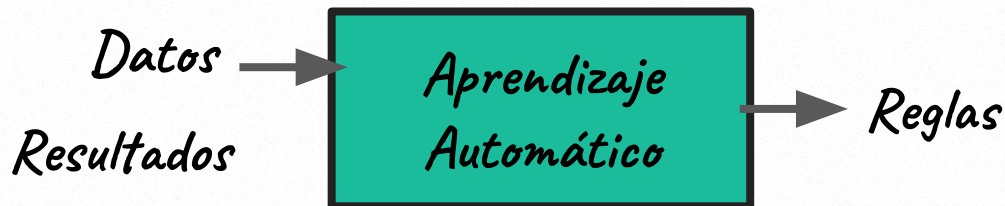
Aprendizaje Automático

Ejemplos:

E1: $[0 \ 1 \ 1 \ 0] \rightarrow \text{"Gato"}$

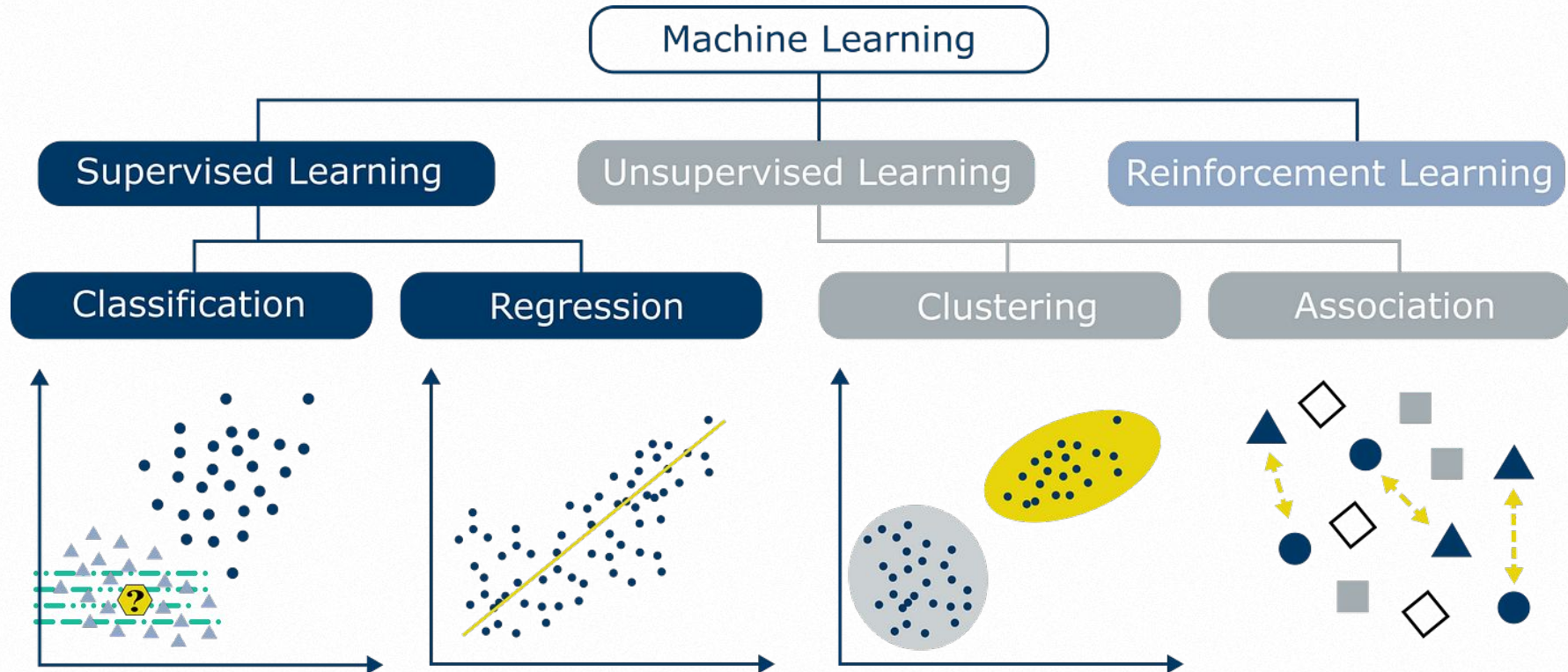
E2: $[1 \ 0 \ 0 \ 1] \rightarrow \text{"Perro"}$

Resultado: Modelo



[Complemento] ¿Cómo se clasifica el machine learning?

Existen **3 tipos principales** de algoritmos de Machine Learning según su **tipo de aprendizaje**, supervisado, no supervisado y por refuerzos.





[Complemento] Entender mi problema

> Una empresa quiere automatizar el filtrado de correos electrónicos para identificar mensajes no deseados (spam).

Opciones de modelos:

1. **Regresión logística:** Modelo simple y eficiente para clasificación binaria.
2. **Árboles de decisión:** Ofrecen interpretabilidad y manejan interacciones entre variables.
3. **Redes neuronales profundas:** Alta precisión, pero requieren más datos y recursos.



"Si estuvieran implementando este sistema en una empresa pequeña con recursos limitados, ¿qué modelo elegirían y por qué?"



[Complemento] Entender mi problema

> Una compañía eléctrica quiere predecir la demanda horaria de energía para optimizar sus recursos.

Opciones de modelos:

1. Regresión lineal: Fácil de implementar, pero puede ser demasiado simple para datos complejos.
2. ARIMA: Modelo clásico para series de tiempo, enfocado en datos estacionarios.
3. LSTM (Long Short-Term Memory): Modelo avanzado basado en redes neuronales para capturar patrones a largo plazo.



Pregunta al público:

Si la predicción debe ser muy precisa para evitar sobrecostos, pero el entrenamiento del modelo toma mucho tiempo, ¿cuál elegirían y por qué?"



Trabajos con entornos en Azure Machine Learning

<https://learn.microsoft.com/es-es/training/modules/work-environments-azure-machine-learning/>

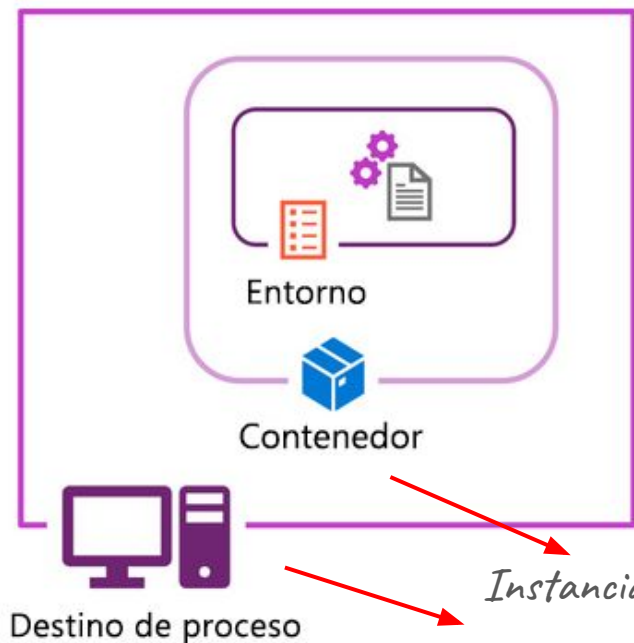




[Repaso] Entorno de Azure Machine Learning



El entorno especifica los **paquetes instalados** disponibles para el código. La administración de estos entornos generalmente se hace por medio de **pip** o **conda**.



*Instancia que contiene las dependencias necesarias para la ejecución
(Compute Target) Ligado a un sistema Operativo*

Python

```
from azure.ai.ml.entities import Environment

env_docker_conda = Environment(
    image="mcr.microsoft.com/azureml/openmpi3.1.2-ubuntu18.04",
    conda_file="./conda-env.yml",
    name="docker-image-plus-conda-example",
    description="Environment created from a Docker image plus Conda environment.",
)
ml_client.environments.create_or_update(env_docker_conda)
```

yml

```
name: basic-env-cpu
channels:
  - conda-forge
dependencies:
  - python=3.7
  - scikit-learn
  - pandas
  - numpy
  - matplotlib
```




[Repaso] Entorno de Azure Machine Learning

> **Entorno mantenido:** Son aquellos administrados por Azure, tienen nombres y dependencias específicas “listas” para su uso.

- **Entorno Personalizado:** Como en este ejemplo, el usuario puede definir cuáles son las dependencias y versiones específicas que requiere.

yml

```
name: basic-env-cpu
channels:
  - conda-forge
dependencies:
  - python=3.7
  - scikit-learn
  - pandas
  - numpy
  - matplotlib
```

Python

```
from azure.ai.ml import command

# configure job
job = command(
    code="./src",
    command="python train.py",
    environment="docker-image-plus-conda-example:1",
    compute="aml-cluster",
    display_name="train-custom-env",
    experiment_name="train-custom-env"
)

# submit job
returned_job = ml_client.create_or_update(job)
```

Python

```
from azure.ai.ml.entities import Environment

env_docker_conda = Environment(
    image="mcr.microsoft.com/azureml/openmpi3.1.2-ubuntu18.04",
    conda_file="./conda-env.yml",
    name="docker-image-plus-conda-example",
    description="Environment created from a Docker image plus Conda environment.",
)

ml_client.environments.create_or_update(env_docker_conda)
```



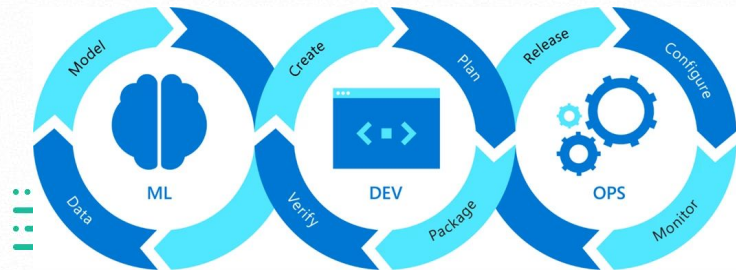
[Duda] ¿Qué medidas de seguridad se pueden aplicar a los entornos en Azure ML para proteger datos y modelos?

➤ Seguridad en los Datos:

- Azure cifra los datos en automático.
- Tokenización (Proceso desarrollado por el usuario)
- Configurar cuentas con RBAC (Role-Based Access Control) para limitar quién puede acceder a los datos

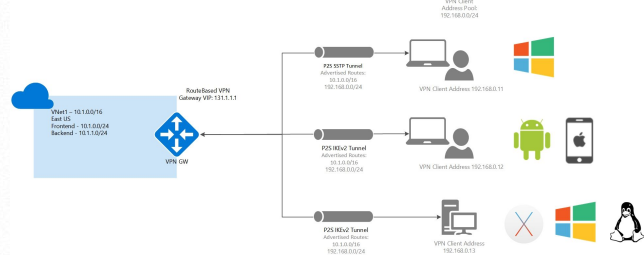
● Seguridad en Modelos:

- Uso de MLOps para control de versiones
- Disponibilización segura con los endpoints administrados por Azure



● Seguridad en el cómputo:

- Uso de Azure VPN Gateway
- Usar Azure Kubernetes Service (AKS) para tener un entorno de cómputo aislado
- Uso de identidades administradas en Azure





Búsqueda del mejor modelo de clasificación con Aprendizaje Automático – Auto ML

<https://learn.microsoft.com/es-es/training/modules/work-environments-azure-machine-learning/>

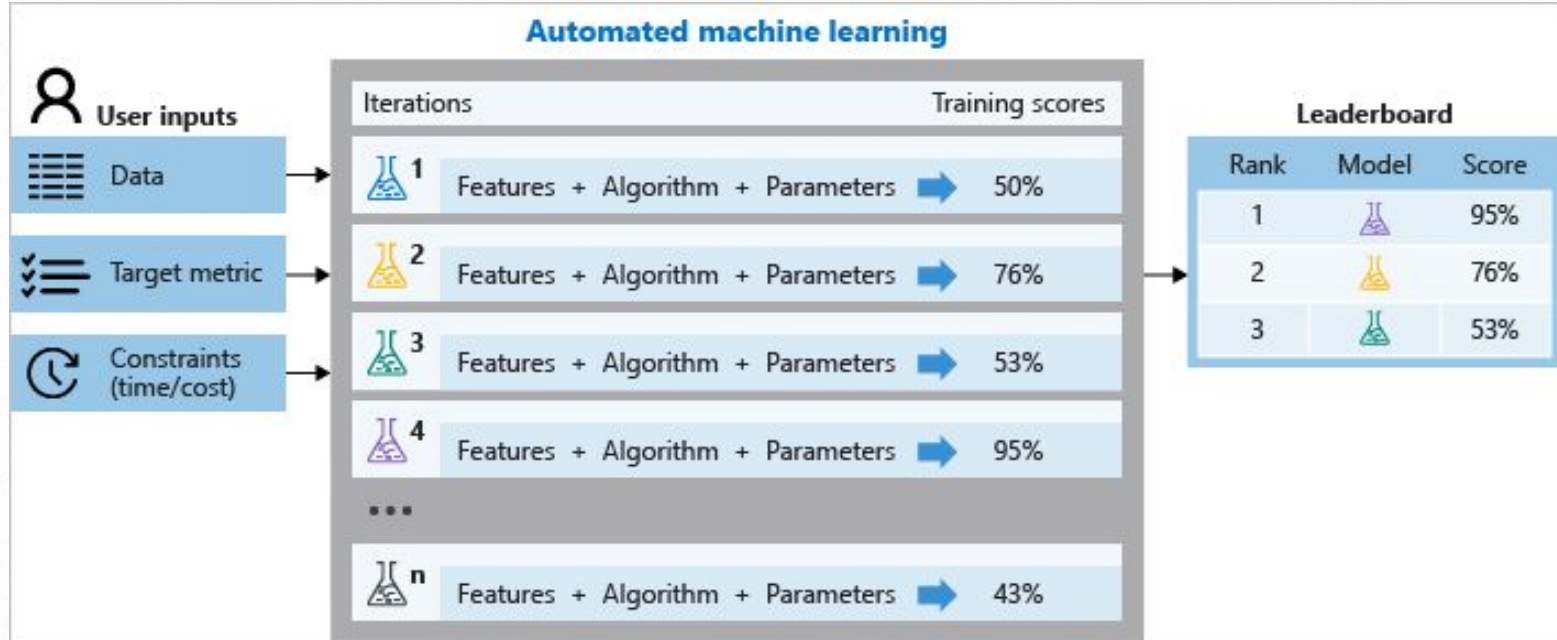




[Repaso] ¿Qué es el Aprendizaje automático automatizado de Azure (Auto-ML)?



Servicio de Azure que permite, de manera *automatizada* crear y *comparar* modelos de ML, agilizando así el desarrollo





[Repaso] Comparativa de modelos

Algorithm name	Explained	Responsible AI	Accuracy ↓
VotingEnsemble	View explanation		0.95300
StackEnsemble			0.95280
MaxAbsScaler, LightGBM			0.95180
MaxAbsScaler, XGBoostClassifier			0.95180
MaxAbsScaler, ExtremeRandomTrees			0.83740

```
from azure.ai.ml import automl

# configure the classification job
classification_job = automl.classification(
    compute="InstanciaAutoMLTest",
    experiment_name="AnálisisSentimientosExperimento", # Nombre que le quieres dar al experimento
    training_data=training_data_input,
    target_column_name="sentiment",
    primary_metric="accuracy",
    n_cross_validations=5,
    enable_model_explainability=True
)

# Nombre del cluster de cómputo
# Data / Dataset de input
# Columna objetivo de la clasificación
# Métrica de evaluación
# Número de validaciones cruzadas
# Permite las explicaciones de las predicciones
```



[Repaso] Pipeline de auto-entrenamiento de modelos



```
1 from azure.ai.ml.constants import AssetTypes
2 from azure.ai.ml import Input
3 import mltable
4
5 # Crea un Dataset basado en los archivos en la carpeta de datos local
6 nombre_tabla = "Análisis_Sentimientos"
7 version = "1"
8
9 # Cargamos los datos desde el Asset de datos y los leemos como DataFrame
10 data_asset = ml_client.data.get(nombre_tabla, version=version)
11 tbl = mltable.load(f'azureml://{data_asset.id}')
12 df = tbl.to_pandas_dataframe()
13 df
```

2.- Lectura de la Data

1.- Conexión con el workspace

```
1 # Librerías para conexión con el Workspace desde cuenta Azure
2 from azure.identity import DefaultAzureCredential, InteractiveBrowserCredential
3 from azure.ai.ml import MLClient
4
5 try:
6     # Se intenta obtener el token por defecto desde el workspace actual
7     credential = DefaultAzureCredential()
8     credential.get_token("https://management.azure.com/.default")
9 except Exception as ex:
10     # En caso de que NO funcionen las credenciales por defecto, se abrirá el
11     # InteractiveBrowserCredential (para autenticarte desde el navegador)
12     credential = InteractiveBrowserCredential()
```

```
1 # Cree una instancia cliente para administrar el Worspace
2 ml_client = MLClient.from_config(credential=credential)
```

3.- Ejecución de Job de Entrenamiento

```
1 from azure.ai.ml import automl
2
3 # configure the classification job
4 classification_job = automl.classification(
5     compute="InstanciaAutoMLTest",
6     experiment_name="AnálisisSentimientosExperiment",
7     training_data=training_data_input,
8     target_column_name="sentiment",
9     primary_metric="accuracy",
10    n_cross_validations=5,
11    enable_model_explainability=True
12 )
```




[Duda] ¿Cómo se configuran y administran los entornos en Azure Machine Learning para experimentos de ML?

The screenshot displays the Azure AI Machine Learning Studio interface. The top navigation bar is blue and contains the text "Azure AI | Machine Learning Studio" on the left, and a series of icons (clock, bell, gear, speech bubble, question mark, smiley face) and a subscription status "Suscripción de Azure 1 handson_AutoML" on the right. A green arrow icon is visible on the far left.

The left sidebar is a vertical navigation menu with the following items: "All workspaces", "Home" (selected), "Model catalog", "Authoring" (with sub-items: "Notebooks", "Automated ML", "Designer", "Prompt flow"), "Assets" (with sub-items: "Data", "Jobs", "Components", "Pipelines", "Environments", "Models", "Endpoints").

The main content area is titled "handson_AutoML" and features a "Generative AI with Prompt flow" section. This section contains four cards, each with a 3D abstract image and a description:

- Multi-Round Q&A on Your Data**: "Create a chatbot that uses LLM and data from your own indexed files to ground multi-round question and answering capabilities in enterprise chat scenarios." Buttons: "Start", "Clone".
- Q&A on Your Data**: "Use LLM and data from your own indexed files to ground multi-round question and answering capabilities." Buttons: "Start", "Clone".
- Web Classification**: "Use LLM to classify URLs into multiple categories." Buttons: "Start", "Clone".
- Chat with Wikipedia**: "Create a chatbot that leverages V to ground the responses." Buttons: "Start", "Clone".

Below this section is a "Generative AI models" section, which includes a "View all" link and four model cards:

- CodeLlama-7b-hf**: "Text generation"
- CodeLlama-7b-Python-hf**: "Text generation"
- CodeLlama-7b-Instruct-hf**: "Text generation"
- Meta-Llama-3.1-8B-In**: "Chat completion"

At the bottom, there is a "Notebook samples" section with a "View all" link.



Seguimiento del entrenamiento de modelos en cuadernos de Jupyter Notebook con MLflow

<https://learn.microsoft.com/es-es/training/modules/track-model-training-jupyter-notebooks-mlflow/>

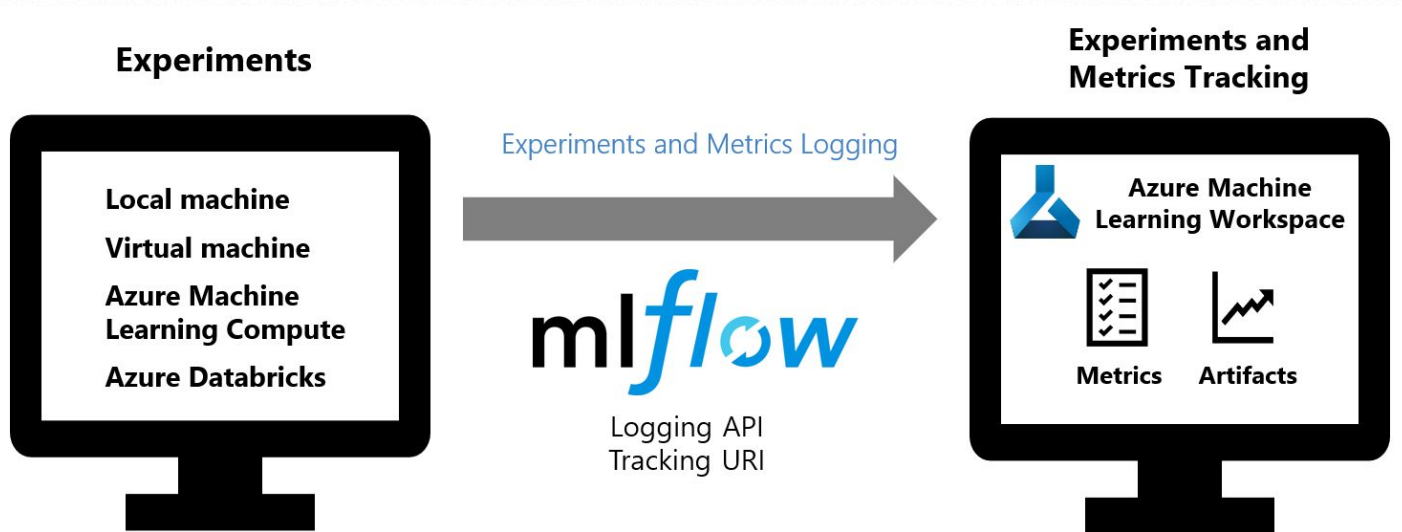




[Repaso] ¿Qué es Azure ML - Flow?



Herramienta que permite la **reproducibilidad** y **track** de todo el **entrenamiento** del modelo

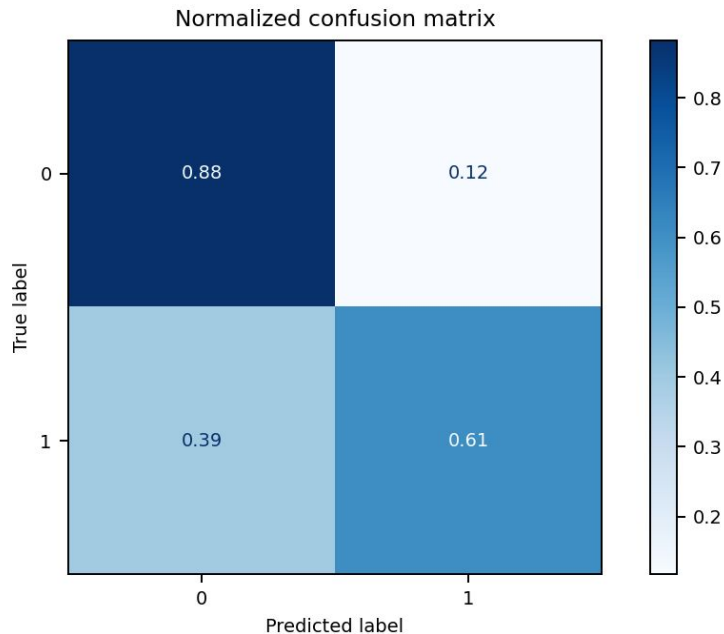
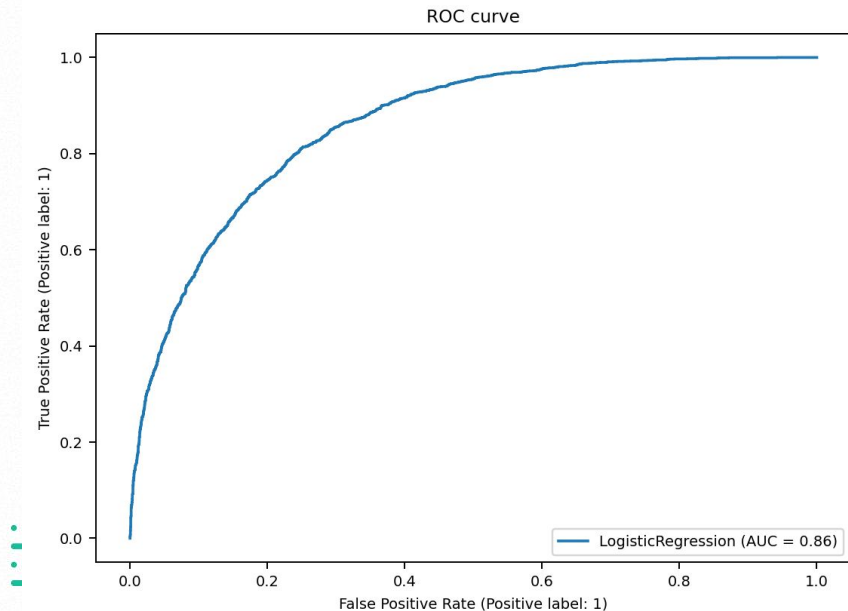




¿Qué puedo medir con ML Flow?



Puedes encontrar el .pkl del modelo entrenado, métricas como la AUC, ROC, o matriz de confusión





[Duda] ¿Cómo funciona MLflow?



Microsoft > John's testing working > Jobs > car-classification > best-model-training

best-model-training Completed

Overview **Metrics** Images Child jobs Outputs + logs Code Explanations (preview) Fairness (preview) Monitoring

Refresh Cancel Create custom chart View as... Current view: *John's view Save view Edit view

Select metrics

Select to view as visualization or table of the data

Search

☐ Select all

Custom chart(s) (3)

- ☒ val_acc vs val_loss ...
- ☒ val_loss ...
- ☒ acc ...

Metrics (7)

- ☐ acc
- ☒ best_score
- ☒ restored_epoch
- ☒ stopped_epoch
- ☐ val_acc
- ☐ val_loss
- ☒ wait_count

best_score

0.9332024

restored_epoch

17

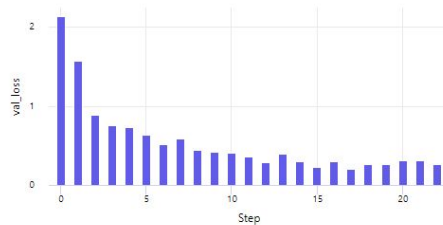
stopped_epoch

22

wait_count

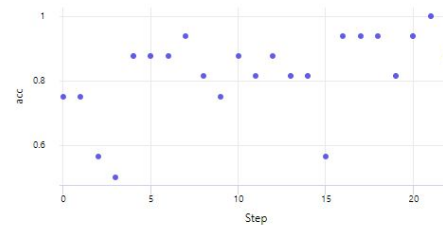
5

val_loss



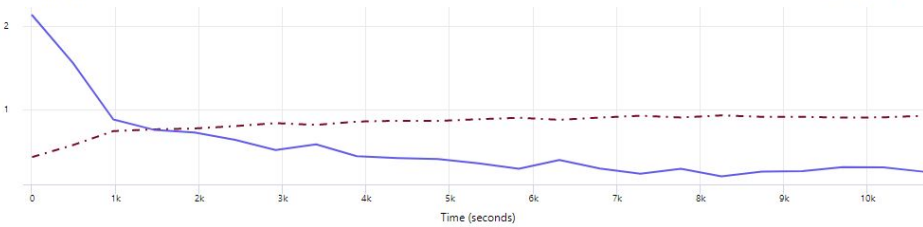
val_loss

acc



acc

val_acc vs val_loss



val_acc val_loss

