

**UNIVERSIDAD DEL VALLE DE GUATEMALA**  
ML Engineering



## Ejercicio 6

SEBASTIAN ARISTONDO PEREZ 20880  
PABLO DANIEL GONZALEZ RAMOS 20362  
JOSE DANIEL GONZALEZ CARRILLO 20293  
JEYNER DAVID ARANGO RUIZ 201106

**GUATEMALA, 27 de septiembre de 2024**

```
name: Re-train model

on:
  push:
    branches:
      - train # Se ejecutará el pipeline cuando se haga push a la rama 'train'

jobs:
  train:
    runs-on: ubuntu-latest

    steps:
      # Step 1: Check out the repository
      - name: Checkout repository
        uses: actions/checkout@v3

      # Step 2: Set up Python environment
      - name: Set up Python
        uses: actions/setup-python@v4
        with:
          python-version: '3.x' # Cambia por la versión de Python que estés usando

      # Step 3: Install dependencies
      - name: Install dependencies
        run: |
          python -m venv venv
          source venv/bin/activate
          pip install -r requirements.txt

      # Step 4: Run training script
      - name: Train the model
        run: |
          source venv/bin/activate
          python train_model.py # Asegúrate de que el script de entrenamiento sea el correcto

      # Step 5 (Optional): Upload model to S3 or another storage
      - name: Upload model to S3
        env:
          AWS_ACCESS_KEY_ID: ${ secrets.AWS_ACCESS_KEY_ID }
```

```
AWS_SECRET_ACCESS_KEY: ${ secrets.AWS_SECRET_ACCESS_KEY }}  
run: |  
    aws s3 cp path/to/model s3://your-bucket/path/to/save  
--recursive
```

Las herramientas de CI/CD (Integración Continua y Entrega Continua) son esenciales en proyectos de Machine Learning para automatizar y mejorar el ciclo de vida del modelo.

## 1. Automatización del reentrenamiento

Al integrar CI/CD, los modelos se reentrenan automáticamente cuando se realizan cambios en el código o los datos. Esto asegura que los modelos estén siempre actualizados y reduce errores al eliminar tareas manuales.

## 2. Pruebas y validación automatizadas

Los pipelines de CI/CD pueden incluir pruebas automáticas para validar el rendimiento de los modelos antes de desplegarlos en producción. Esto garantiza que solo los modelos que cumplen con criterios específicos lleguen a producción, mejorando la calidad y confiabilidad.

## 3. Despliegue continuo y monitoreo

El despliegue continuo permite que los modelos entrenados se envíen automáticamente a producción. Además, el monitoreo integrado detecta problemas como la deriva de datos, activando alertas o nuevos reentrenamientos de manera automática.

## 4. Reproducibilidad y trazabilidad

CI/CD ayuda a mantener control de versiones sobre los datos y el código, asegurando que los experimentos sean reproducibles y que el rendimiento de los modelos pueda rastrearse a lo largo del tiempo.

En resumen, CI/CD en Machine Learning automatiza procesos clave, mejora la calidad de los modelos y facilita su mantenimiento en producción, haciendo más eficientes los proyectos.