

Machine learning pipelines

A short introduction to Metaflow



¿Qué es un pipeline?

- Flujo de información de un punto a otro
- Sufre transformaciones en el proceso
- Dividido en pasos que comparten información entre ellos

¿Metaflow?

- Biblioteca de Python para crear flujos de trabajo (*workflows*)
- Es de código abierto
- Desarrollada y usada en **NETFLIX**

Alternativas

- Airflow
- Luigi
- Prefect
- Kubeflow

En dónde se ejecuta Metaflow



Local

Ideal para comenzar y cargas de trabajo pequeñas.


Fácil de crear y ejecutar.



Cloud

Compatible con AWS, Azure, GCP y Kubernetes.


Requiere configuración extra.



¿Qué uso en mi trabajo?

Databricks

Usamos un framework similar al de
Metaflow, pero con diferente “backend”



DEMO

tcs.dev/metaflow-pipelines

Respaldo: github.com/fferegrino/cf-mlops/tree/metaflow-pipelines

P&R

Más información

- Metaflow: docs.metaflow.org
- Outerbounds: outerbounds.com/metaflow/
- Libro: Effective Data Science Infrastructure