

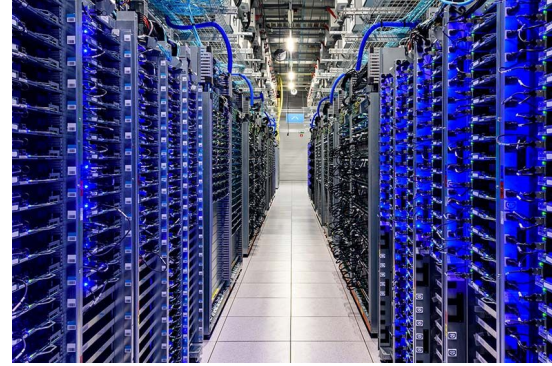


Introducción a MLOps

ML2 UBA

Qué es la nube?

La nube son computadoras que están en un data center las cuales son alquiladas a los usuarios, y además vienen con servicios incluidos e.g. cloudsql

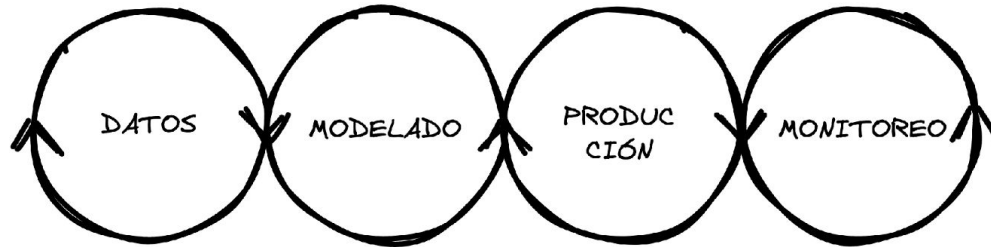


Qué es MLOps?

“MLOps o ML Ops es un conjunto de prácticas que tiene como objetivo implementar y mantener modelos de aprendizaje automático en producción de manera confiable y eficiente”

- Wikipedia

MLOPS



Qué es MLOps?

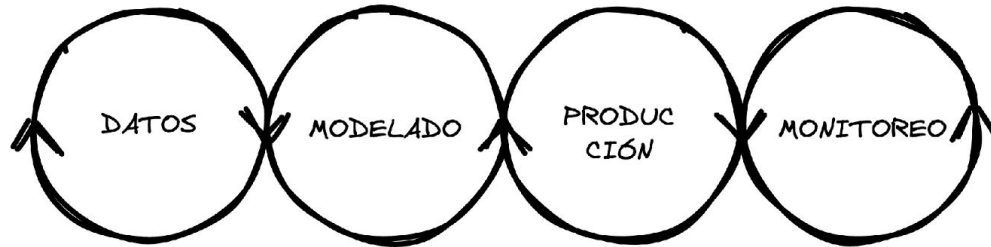
“MLOps o ML Ops es un conjunto de prácticas que tiene como objetivo implementar y mantener modelos de aprendizaje automático en producción de manera confiable y eficiente”

- Wikipedia

“Todo aquello que hace al machine learning que no son notebooks”

- Apu

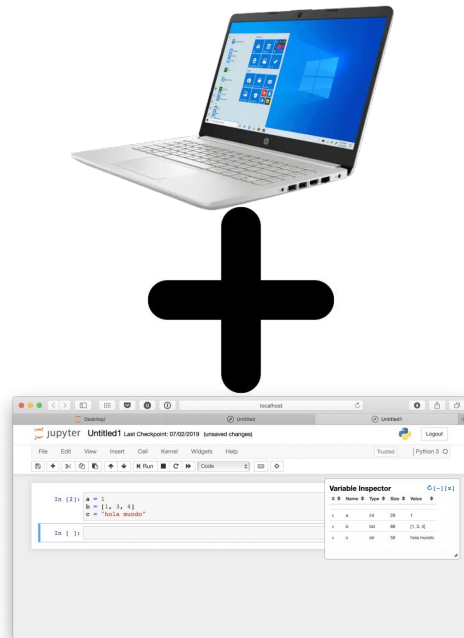
MLOPS



Cómo se hacían las cosas antes?

- 1) Alguien me daba los datos
- 2) Los cargaba en mi notebook y usaba jupyter para hacer análisis y modelos
- 3) Me generaba un modelo y tenía métricas en test

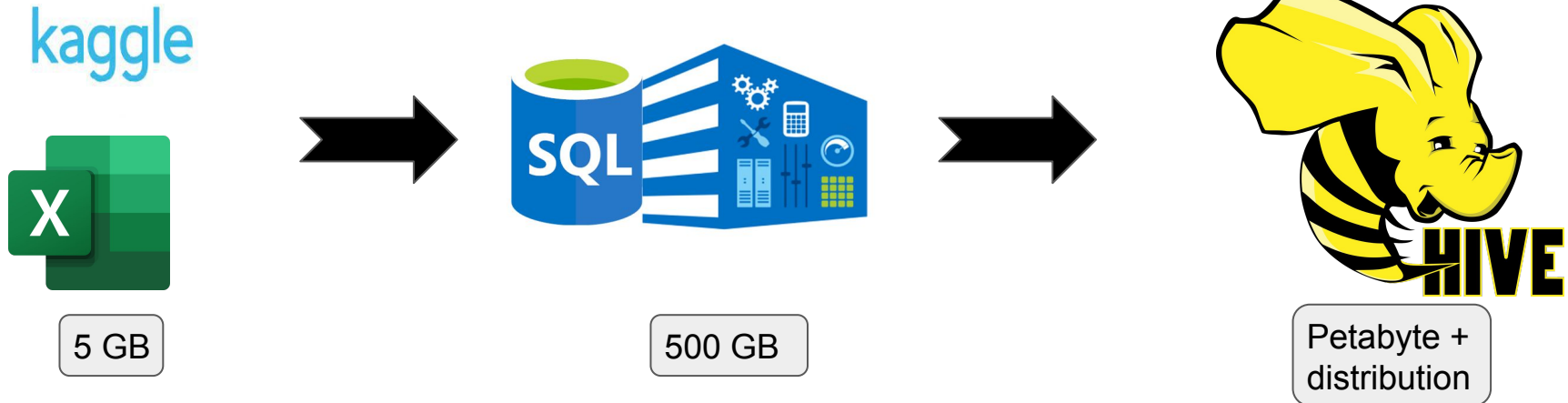
kaggle



Qué problemas tiene esto?

1) Storage

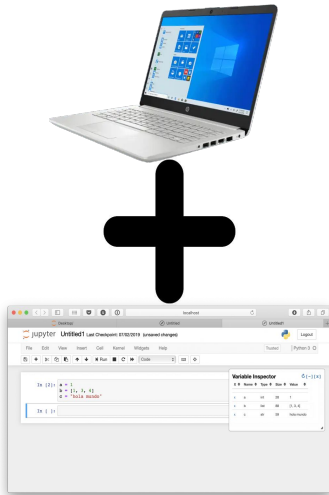
- Un excel normal alrededor de los 5gb se rompe
- En un excel no es posible actualizar los datos en tiempo real
- En un excel no hay una fuente centralizada de información (cada uno tiene su propia versión)
- Etc...



Qué problemas tiene esto?

2) Cómputo

- Los modelos de ML tardan mucho en entrenar
- Yo no quiero comprar una GPU para entrenar un modelo de ML y no usarla mas



Qué problemas tiene esto?

3) Puesta en producción

- En muchas aplicaciones se necesita tener el modelo disponible para ser usado de manera instantánea e.g. Fraude con tarjeta de crédito.
No puede haber una persona cada vez que se hace una transacción ejecutando celdas de una notebook

Examples

```
import numpy as np
from sklearn.linear_model import LinearRegression
X = np.array([[1, 1], [1, 2], [2, 2], [2, 3]])
#  $y = 1 * x_0 + 2 * x_1 + 3$ 
y = np.dot(X, np.array([1, 2])) + 3
reg = LinearRegression().fit(X, y)
reg.score(X, y)

reg.coef_

reg.intercept_

reg.predict(np.array([[3, 5]]))
```



AWS Lambda

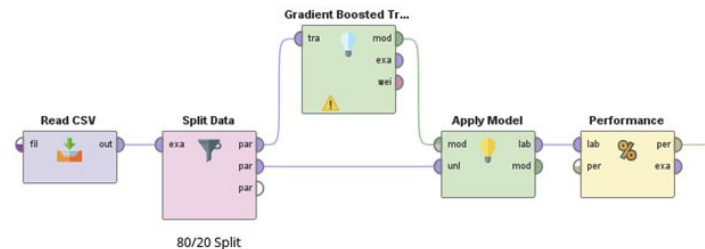
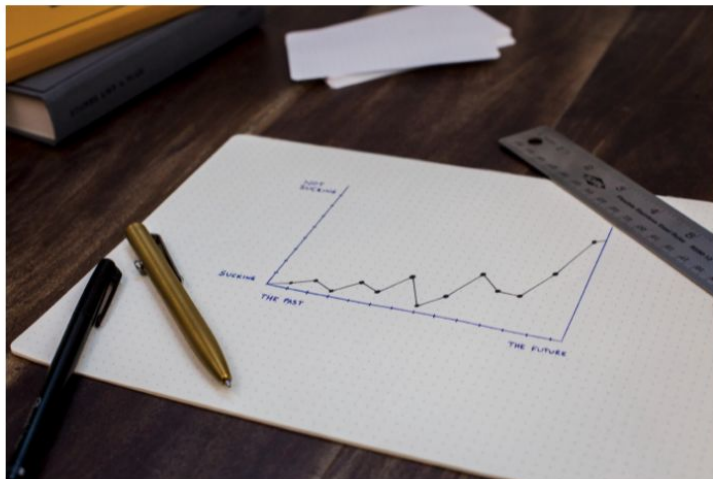
Para todo esto y para resolver otros problemas se creó el concepto de MLOps



Martin Schmitz, PhD

Mar 4 · 3 min read · [Listen](#)

There is No Place for model.fit() Data Scientists



A simple workflow in [RapidMiner](#) fitting a Gradient Boosted Tree with an 80/20 Split Validation

| *Why is my work better than the AutoML work?*

| *I can hire data scientists, but it's hard to find good ones.*

[Link](#)

Modelado / Training

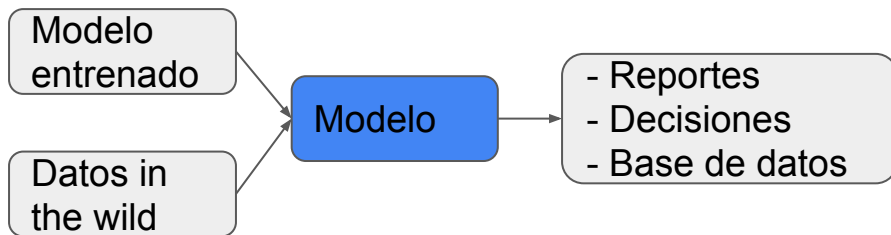
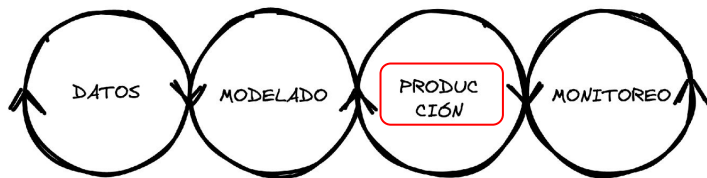
MLOPS



- Es lo que se da en ML1 o cualquier otro curso de DS
- Es la parte más cara del proceso
- Librerías:
sklearn, Tensorflow, Keras, Prophet, Statsmodels, etc...
- Herramientas:
[Servers](#) + scripts de python
[Databricks](#)
[Vertex IA](#)
Etc ...

Producción / Serving

MLOPS



- En esta parte ya se tiene un modelo entrenado y se lo disponibiliza para su uso

e.g. Un brasilero hizo una compra en Kenia y mi modelo tiene que decidir si se aprueba o no.

- Herramientas fundamentales

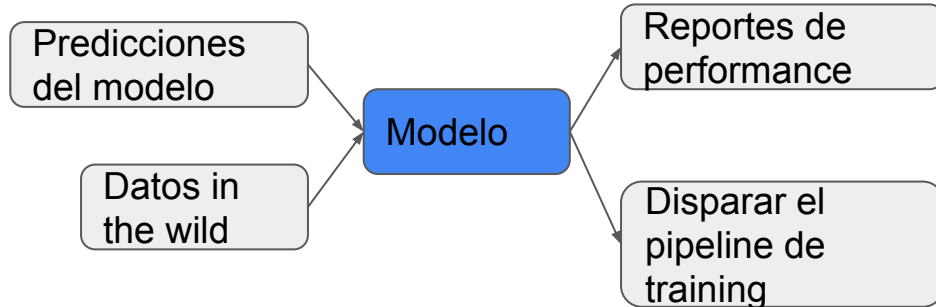
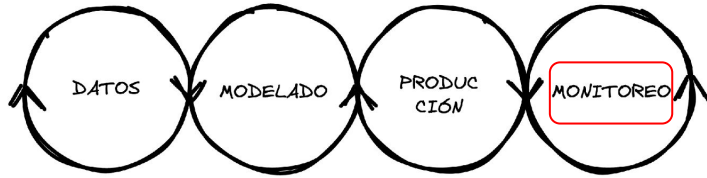
Conección en tiempo real a los datos.

Un modelo disponible.

- Herramientas:
[Servers](#) + scripts de python
[Databricks](#)
[Vertex IA](#)
[MLflow](#)

Monitoreo / Testing

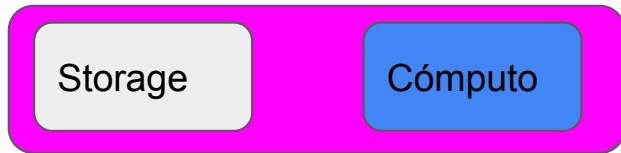
MLOPS



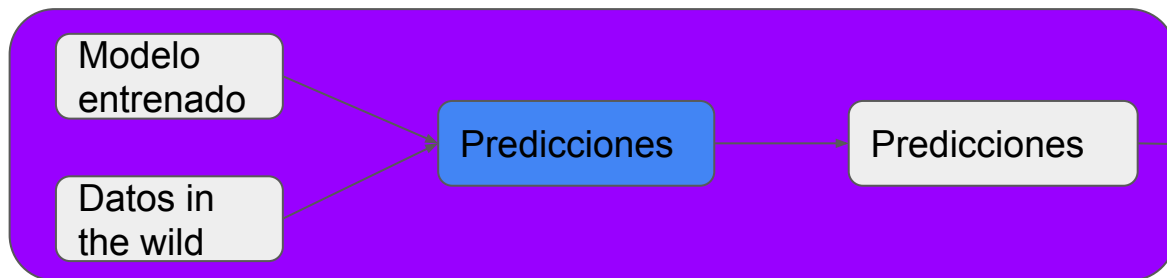
- Ahora hay que comparar la predicción que obtuvimos con la realidad final

e.g. El brasileiro llamó y se quejó porque quería comprar algo y no pasó la tarjeta.

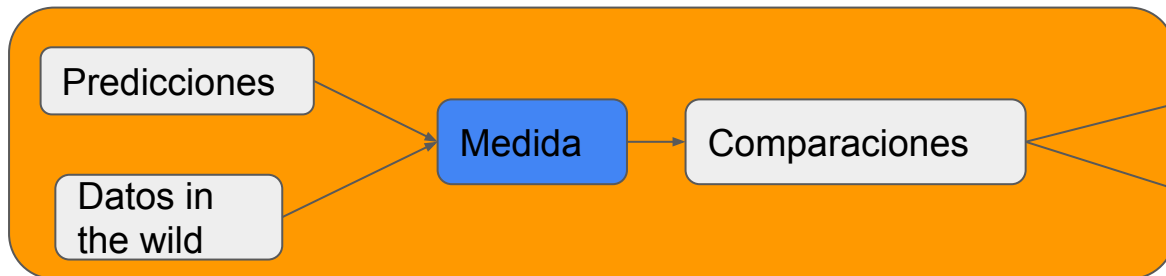
- Herramientas:
Bases de datos con las predicciones pasadas
Servers + scripts de python
[Databricks](#)
[Vertex IA](#)
[MLflow](#)
[Cloud functions](#)



Training



Inferencia

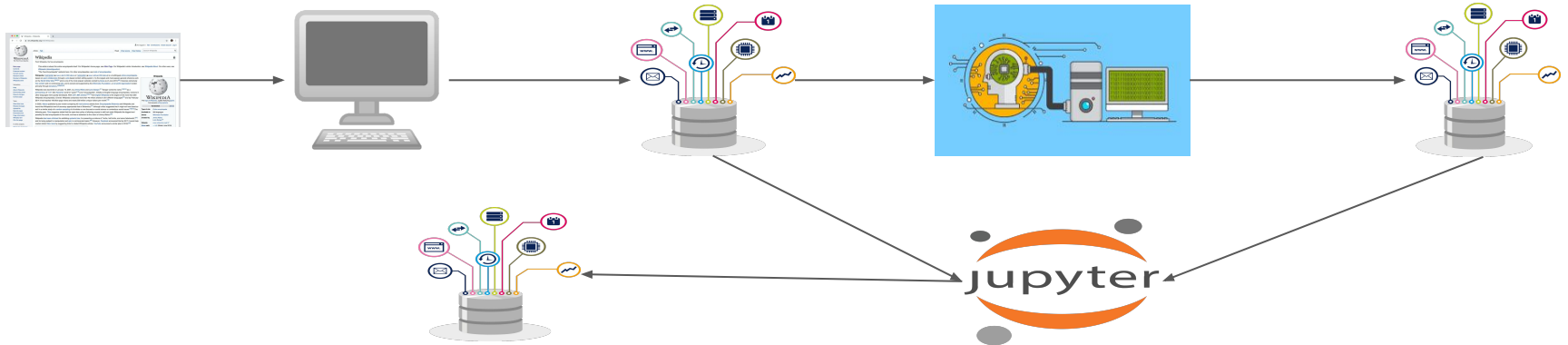
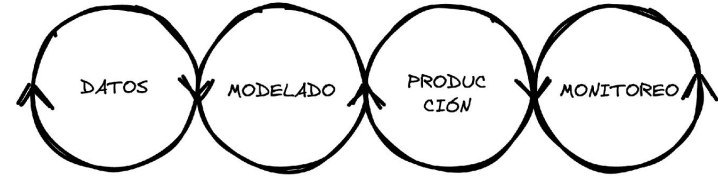


Monitoreo

MLOps first proposal

- 1) levantar un server
- 2) Scrapear alguna página web
- 3) Guardar los datos en una base
- 4) Entrenar un modelo de ML
- 5) Hacer predicciones día a día con lo que vayamos scrapeando
- 6) Guardar las predicciones en otra base de datos
- 7) Comparar las predicciones con la realidad

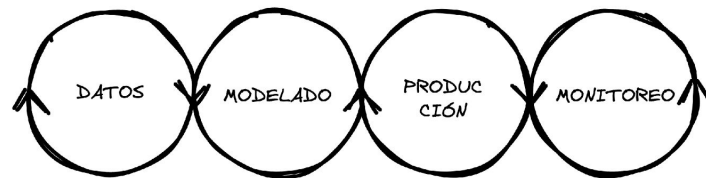
MLOPS



MLOps second proposal

- 1) levantar un bucket (levantar un bucket)
- 2) Hacer una función lambda
- 3) Disparar un modelo pre entrenado (ver [ejemplos](#))
- 4) Guardar los resultados en un S3

MLOps



Fundamental step:

Levantar y manejar servidores

- [Cuenta de GCP](#)
- [Smarty](#)

Comandos comunes

- ls (ver files)
- pwd (ver path)
- cd (cambiar directorio)
- mkdir (crear carpeta)
- rm (remover file)
- rm -rf (remover folder)
- vim (editor de texto)
- python (ejecutar intérprete)
- mv (mover files)
- cp (copy)

To Do:

- Entrar al home
- Crear una carpeta con tu nombre
- Subir un script de python
- Hacer la conexión a la base de datos
- Levantar esos datos, hacer una limpieza
- Subirlos a la base a una nueva tabla

Fundamental step:

Levantar y manejar servidores

- [Cuenta de GCP](#)
- [Smartty](#)

Comandos comunes

- `sudo apt update`
- `sudo apt install postgresql`
- `sudo -i -u postgres`
- `psql -h 34.70.163.208 -p 5432 -d postgres -U postgres`
- `pkOpbodE53I1zzgN`
- `\copy (select * from public.persona limit 1) to 'persona.csv' csv;`
- `gsutil`

To Do:

- Hacer un script que me calcule ciertas medidas
- Subirlo
- Hacer la conexión
- procesarlos
- Subirlos en una nueva tabla