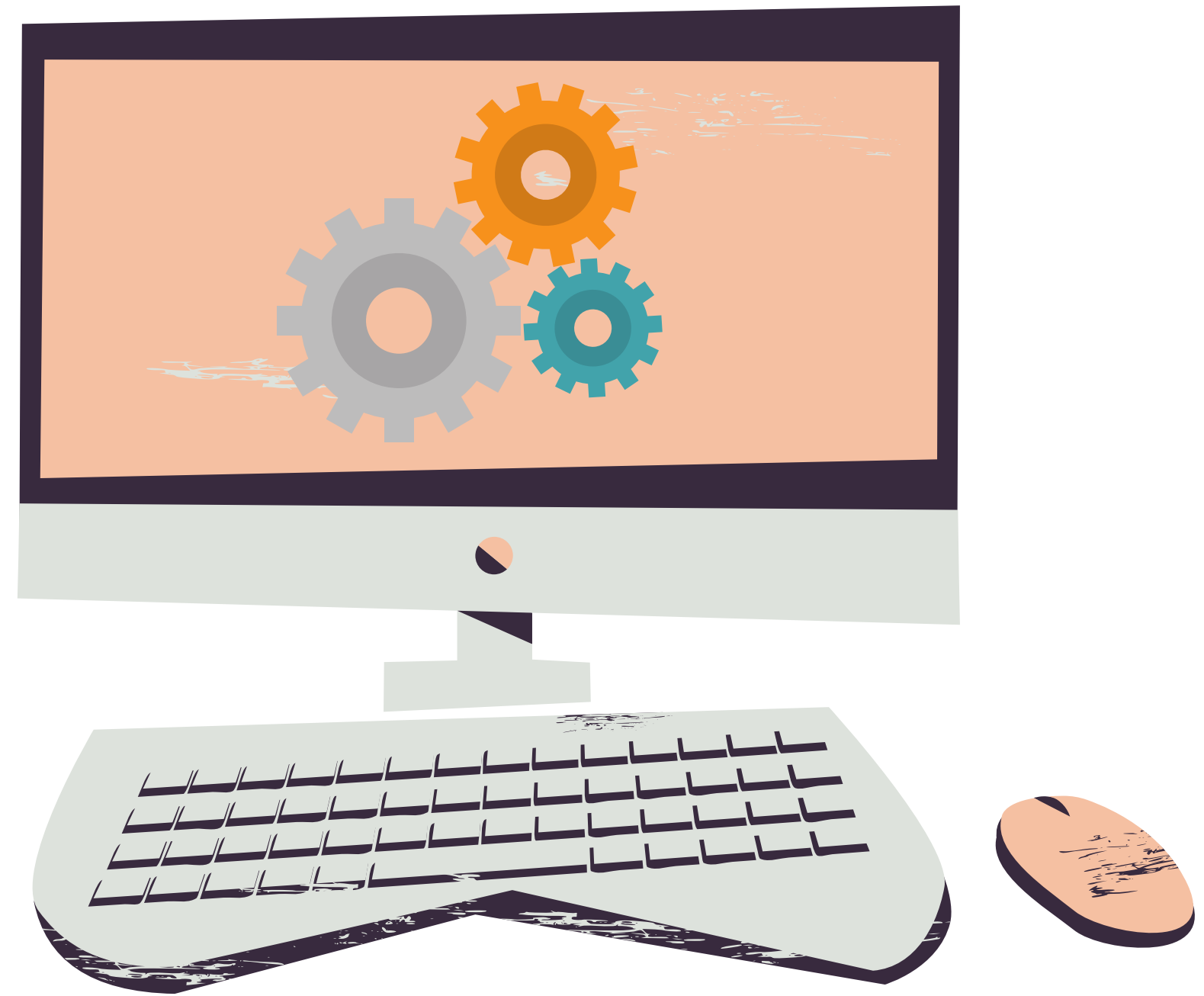


Machine Learning Pipeline

Python Panamá



Agenda

01

Machine Learning

Aprendizaje supervisado

Aprendizaje no supervisado

03

Práctiquemos

Ejemplo en Python

02

Machine Learning Pipeline

Pre-procesamiento de datos

Modelos



Sobre mi

Ingeniera de telecomunicaciones y científica de datos.

Me gusta la fotografía, leer y programar.

Tengo un blog en Medium "Al mal tiempo, buena data".

Machine learning

Las técnicas de machine learning permiten reconocer patrones entre los datos para realizar una clasificación o predicción basados en las características comunes encontradas.

Aplicaciones:

- Recomendación
- Diagnóstico médico
- Reconocimiento facial



Machine Learning

Conjunto de algoritmos entrenados con datos que simulan la forma de aprendizaje humana.



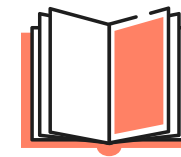
Datos

- Recolección
- Selección
- Análisis



Modelos

- Selección
- Entrenamiento
- Prueba



Implementación

Resultado final

- Modelo
- Dashboard

Aprendizaje supervisado

El algoritmo se entrena con datos de entrada y salida conocidos. Permite encontrar una relación entre las variables

Clasificación:

Regresión logística

Support Vector Machine

Predicción:

Regresión lineal

Aprendizaje no supervisado

Todo el proceso de modelado se realiza sobre un conjunto de datos de entrada. Permite identificar patrones comunes en los datos

Clustering:

k-means

PCA

Machine learning Pipeline

Flujo de trabajo que permite una vez definido el problema realizar ingesta, preparación y selección de datos, implementación y evaluación de modelos.



Pre-procesamiento de datos

- Ingesta de datos
- Exploración de datos (EDA)
- Preparación de datos



Implementación de modelos

- Definir variable objetivo
- Selección de variables
- Entrenamiento del modelo
- Evaluación del modelo

Pre-procesamiento de datos

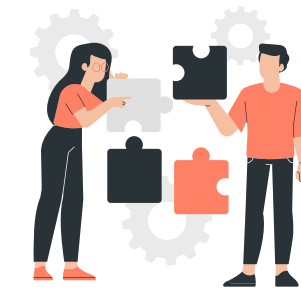
El buen desempeño de tu modelo dependerá en gran medida de la calidad de tus datos.



Ingesta de datos



Exploración de
variables



Preparación de
datos

Implementación de modelos

Jugar con los parámetros del modelo nos permite explorar y encontrar un buen modelo que se ajuste a nuestro problema, sin ajustarse a nuestros datos actuales (overfitting) o no obtener buenas métricas.

01

Definir variable objetivo

02

Selección de variables

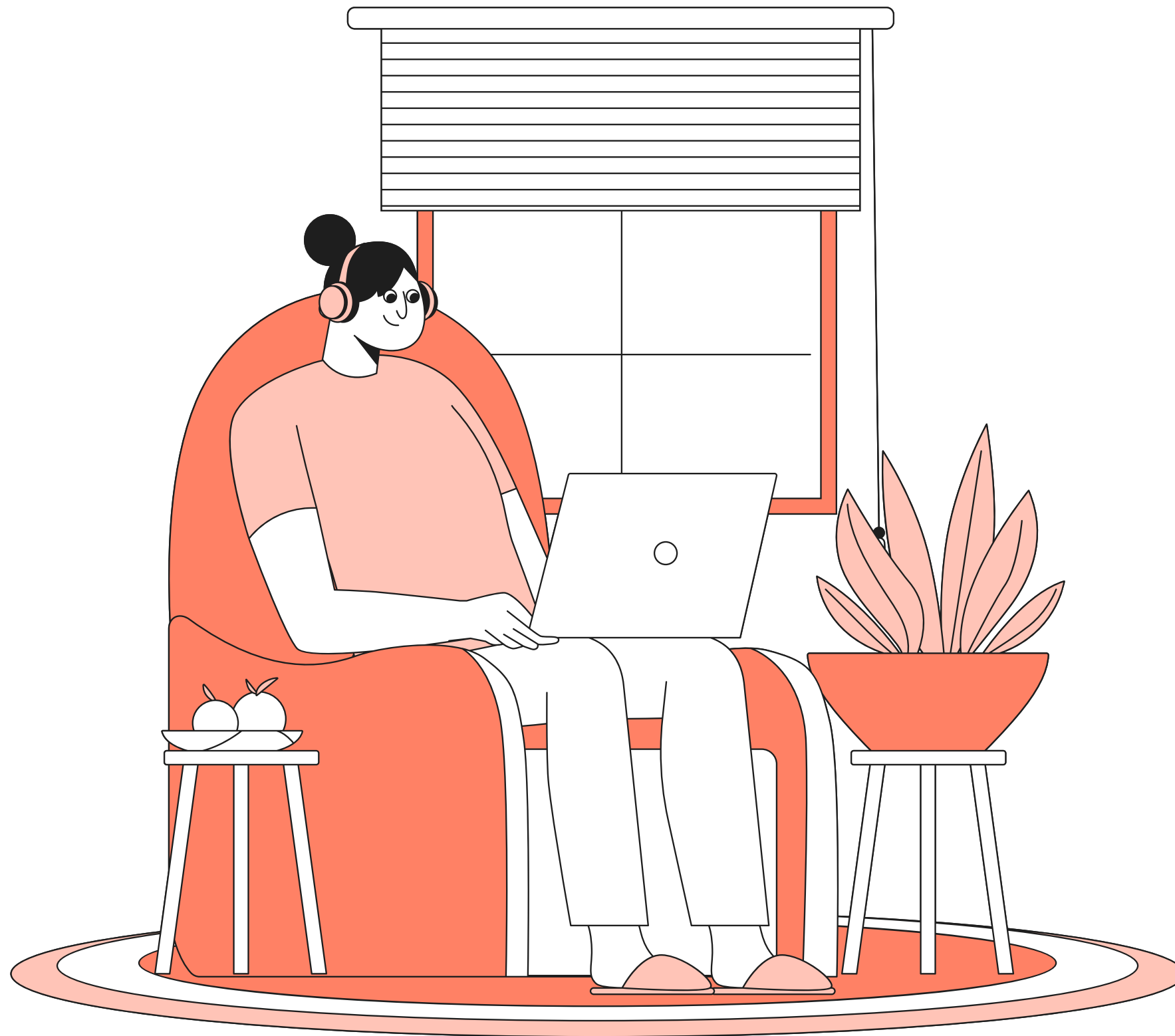
03

Entrenamiento del modelo

04

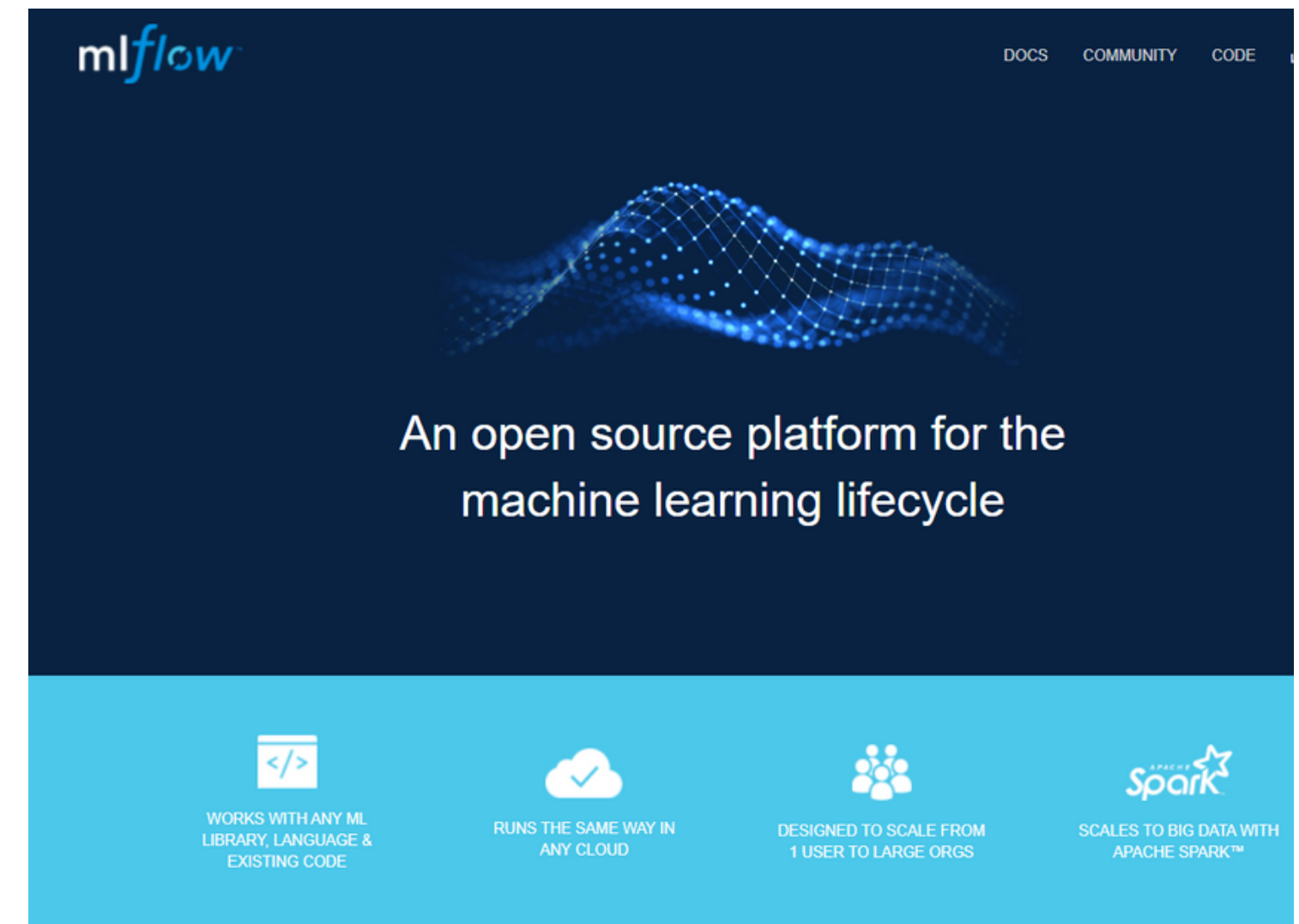
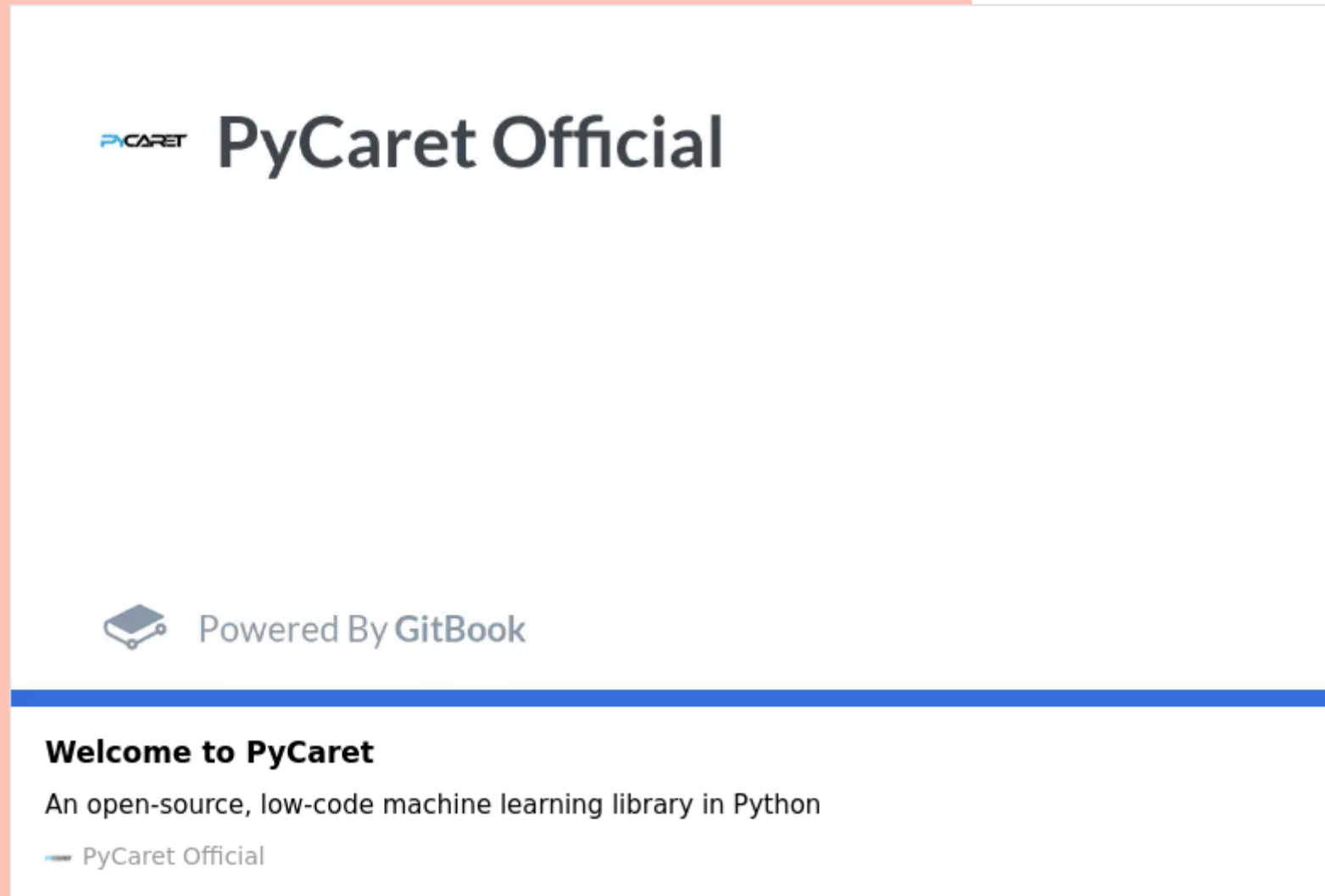
Evaluación del modelo



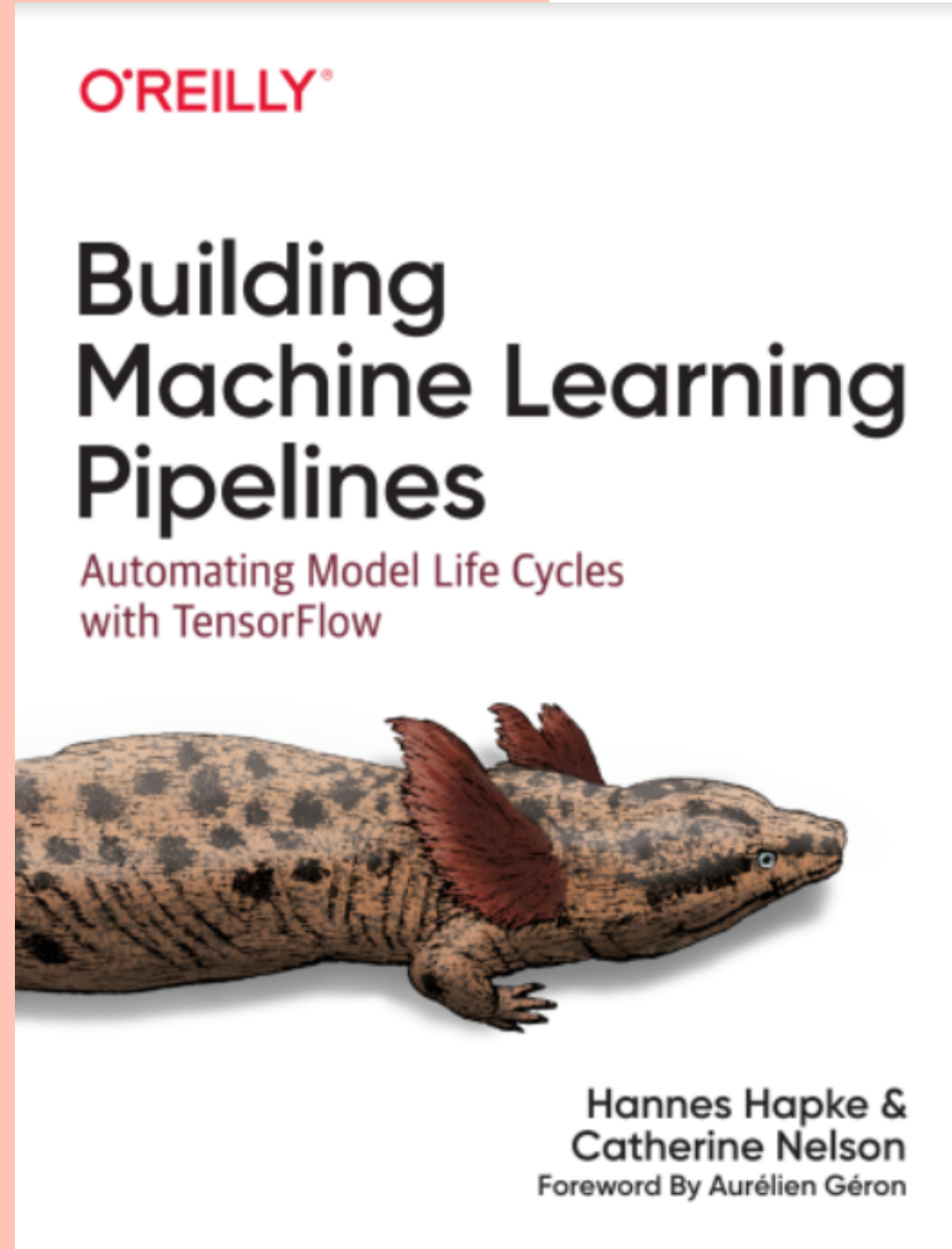


**Manos a la
obra!**

Herramientas



Recursos




coursera
project network

★★★★★ 4.3 (47) | 2.5K Students

Building Machine Learning Pipelines in PySpark MLlib


coursera

Building Machine Learning Pipelines in PySpark MLlib
Complete this Guided Project in under 2 hours. By the ...

 Coursera

The image shows the front cover of the book 'Architecting a Machine Learning Pipeline' by Hannes Hapke & Catherine Nelson. The cover has a dark grey background with a pattern of interlocking puzzle pieces. A white puzzle piece in the center contains the title 'ARCHITECTING A MACHINE LEARNING PIPELINE' in teal and black text. Below the title, it says 'HOW TO BUILD SCALABLE MACHINE LEARNING SYSTEMS'.

Architecting a Machine Learning Pipeline
How to build scalable Machine Learning systems — Part 2/2

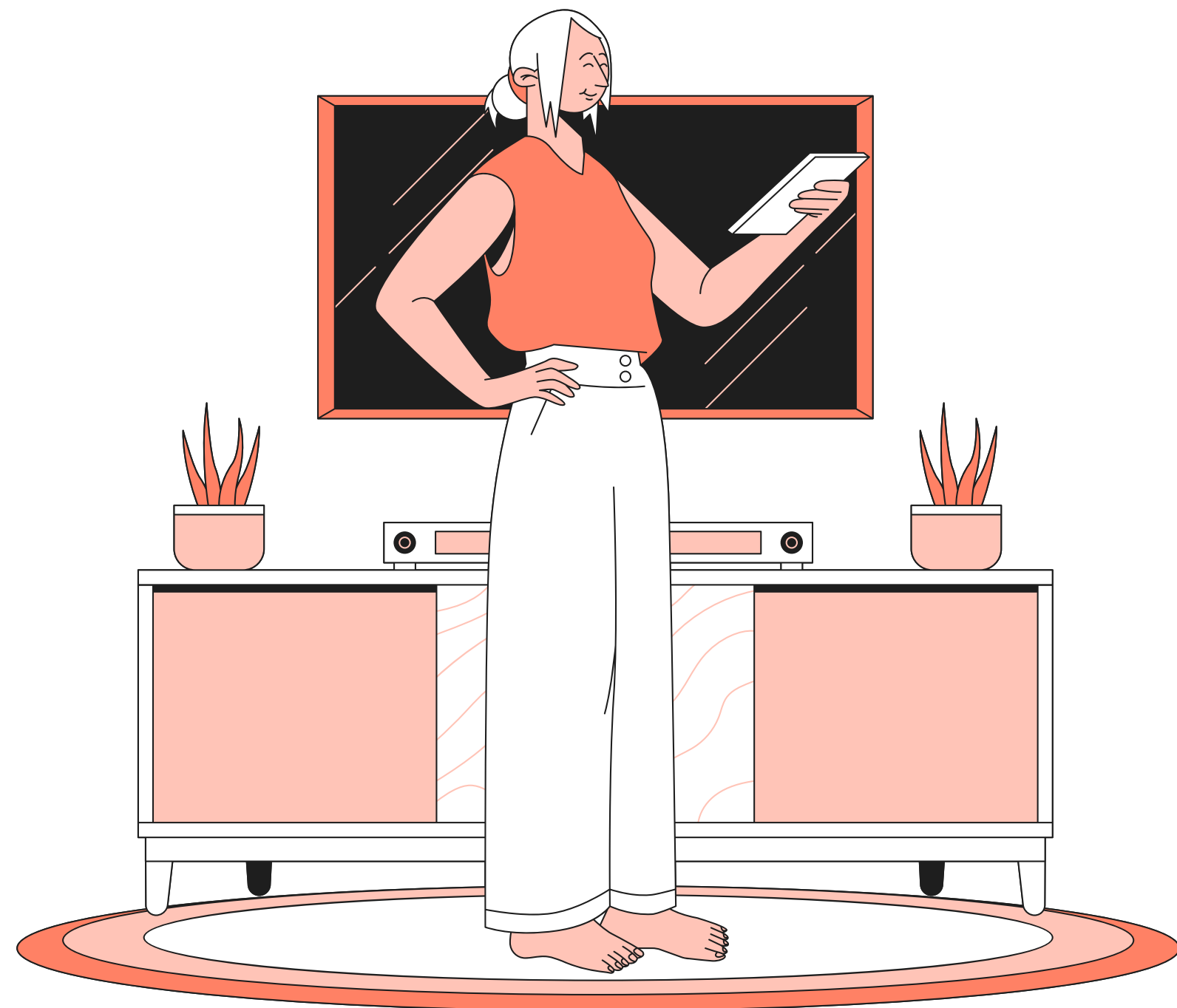
 Medium / Apr 5, 2019



**my own
model**

**me giving training
tips to others**

©skt



Contacto

lauralpezb





¡Gracias!