Proyecto Ecobici-AutoML

# Estructura avanzada del proyecto Ecobici-AutoML

Este documento describe la estructura avanzada y organización de carpetas del proyecto Ecobici-AutoML, así como la función de cada componente hasta la versión estable del 25 de octubre de 2025.

## 1. Estructura de carpetas

📁 data/  
 ├── raw/ → Datos originales descargados de la API GCBA.  
 ├── interim/ → Datos intermedios durante la limpieza y unificación.  
 ├── curated/ → Parquets listos para modelado (nivel Silver).  
📁 notebooks/  
 ├── 01\_ingest.ipynb → Adquisición y unificación de fuentes.  
 ├── 02\_preprocess.ipynb → Limpieza, selección y generación de features.  
 ├── 03A\_PyCaret\_AutoML.ipynb → Entrenamiento con PyCaret.  
 ├── 03B\_H2O\_AutoML.ipynb → Entrenamiento con H2O.  
 ├── 03C\_FLAML\_AutoML.ipynb → Entrenamiento con FLAML.  
 ├── 04A\_BatchEval.ipynb → Evaluación por horizonte.  
 ├── 04C\_Deploy.ipynb → Despliegue de predicciones.  
📁 models/  
 ├── 03D/best\_model.pkl → Modelo final seleccionado.  
📁 reports/  
 ├── backtest\_metrics\_by\_horizon.csv  
 ├── error\_by\_hour\_and\_horizon.csv  
📁 docs/ → Documentación (README, FAA\_avanzado, Explicacion\_Flujo, ONBOARDING).  
📁 src/ → Scripts productivos (train.py, predict\_batch.py, utils.py).  
📁 app/ → Dashboard Streamlit (mapa de calor y panel de estaciones).

## 2. Entorno y configuración

El entorno base se gestiona mediante conda:  
 conda env create -f environment.yml  
 conda activate ecobici\_automl  
  
Validación del entorno:  
 make verify-env  
  
Makefile incluye objetivos para:  
- repo-tree → genera docs/dev/repo\_tree.txt  
- mlflow-ui → inicia MLflow en localhost:5001

## 3. Flujo general

1️⃣ Ingesta de datos (01\_ingest)  
2️⃣ Preprocesamiento y feature engineering (02\_preprocess)  
3️⃣ Modelado AutoML (03A, 03B, 03C)  
4️⃣ Evaluación comparativa y selección (03D)  
5️⃣ Generación batch y despliegue (04A, 04C)  
6️⃣ Dashboard final (feat/dashboard\_clean\_2025-10-25)

## 4. Guía para nuevos usuarios

1. Clonar el repositorio.  
2. Crear el entorno con environment.yml.  
3. Ejecutar `make verify-env` para validar dependencias.  
4. Correr los notebooks en orden (01 → 04).  
5. Iniciar el dashboard con:  
 streamlit run app/dashboard.py