

Laboratorio 3

- MLFlow y Feast -

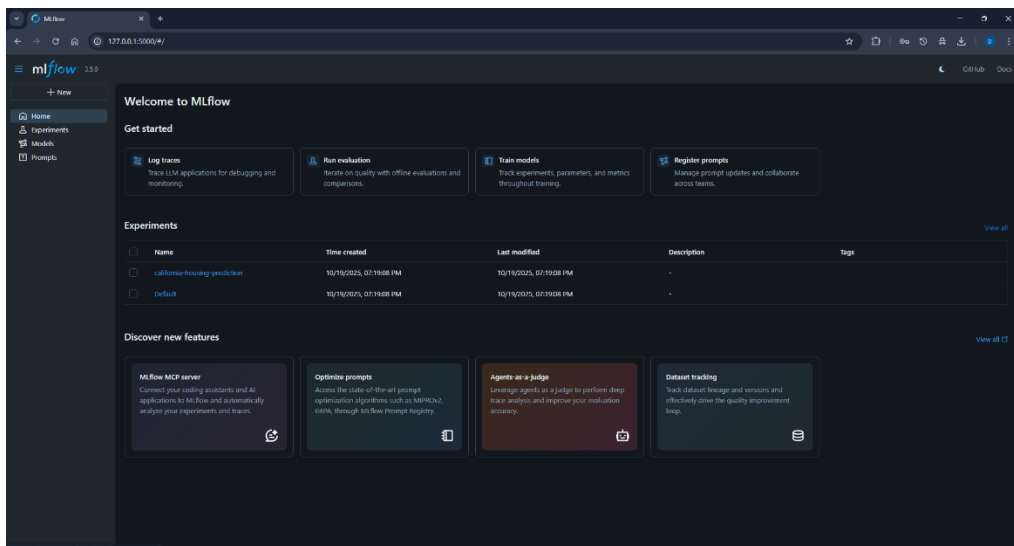
Diego Alberto Leiva Pérez

Guatemala, 19 de octubre de 2025

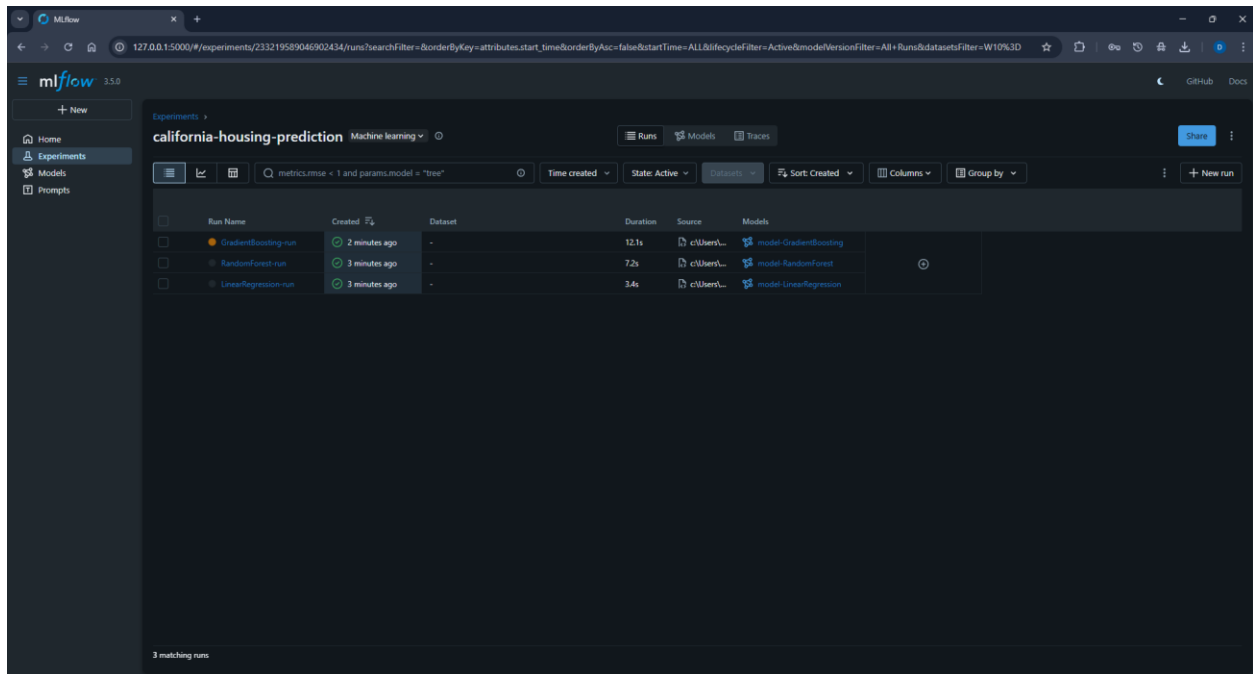
Lanzamiento de interfaz con comando

```
Miniconda - mlflow ui --port 5000 X + v
(mlops-lab3) C:\Users\diego\Documents\UVG\10mo Semestre\MachineLearning\MLOps-Lab3>mlflow ui --port 5000
c:\Users\diego\miniconda3\envs\mlops-lab3\Lib\site-packages\mlflow\gateway\config.py:454: PydanticDeprecatedS
ic V2.0 to be removed in V3.0. See Pydantic V2 Migration Guide at https://errors.pydantic.dev/2.12/migration/
class Route(ConfigModel):
[MLFlow] Security middleware enabled with default settings (localhost-only). To allow connections from other
INFO: Uvicorn running on http://127.0.0.1:5000 (Press CTRL+C to quit)
INFO: Started parent process [27660]
INFO: Started server process [20504]
INFO: Waiting for application startup.
INFO: Started server process [28788]
INFO: Started server process [4848]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Application startup complete.
INFO: Started server process [16420]
INFO: Waiting for application startup.
INFO: Application startup complete.
```

Interfaz de MLFlow



Experimentos

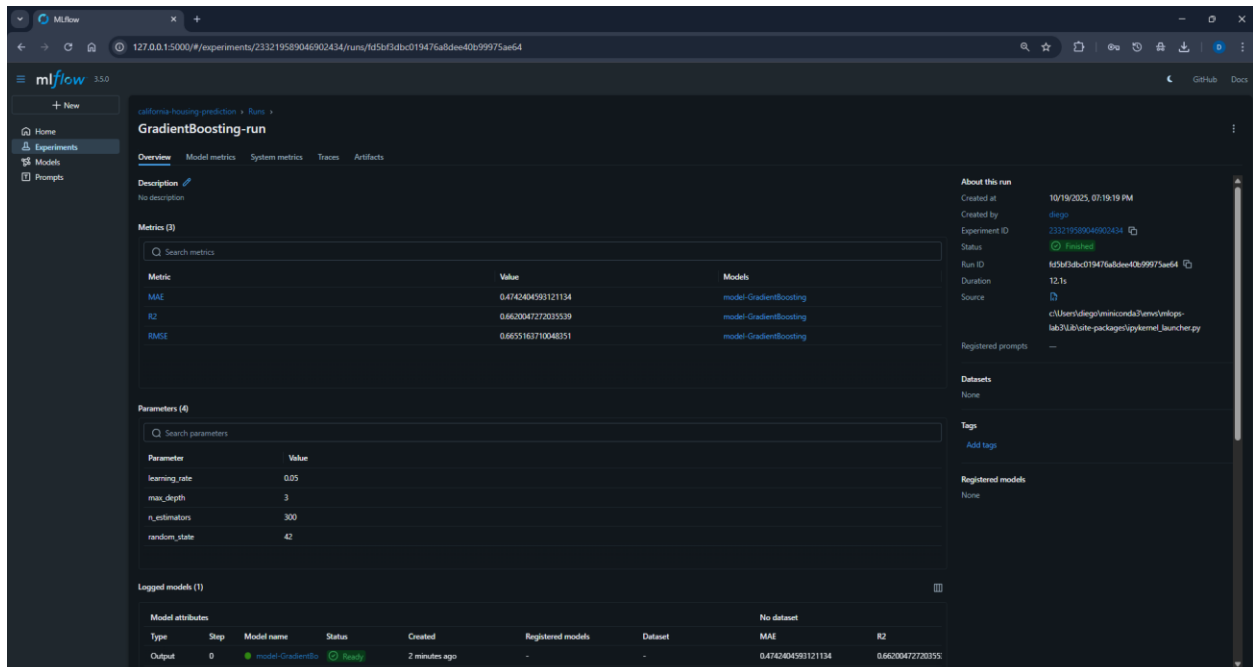


The screenshot shows the mlflow Experiments page for the 'california-housing-prediction' experiment. The interface includes a sidebar with navigation links (Home, Experiments, Models, Prompts) and a main content area. The main area displays a table of runs with columns: Run Name, Created, Dataset, Duration, Source, and Models. Three runs are listed: GradientBoosting-run, RandomForest-run, and LinearRegression-run. The GradientBoosting-run is highlighted as the selected run.

Run Name	Created	Dataset	Duration	Source	Models
GradientBoosting-run	2 minutes ago	-	12.1s	c:\Users\...	model-GradientBoosting
RandomForest-run	3 minutes ago	-	7.2s	c:\Users\...	model-RandomForest
LinearRegression-run	3 minutes ago	-	3.4s	c:\Users\...	model-LinearRegression

Model Registry

GradientBoosting



The screenshot shows the mlflow Model Registry page for the 'GradientBoosting-run' model. The interface includes a sidebar with navigation links (Home, Experiments, Models, Prompts) and a main content area. The main area displays a table of metrics (MAE, R2, RMSE) and parameters (learning_rate, max_depth, n_estimators, random_state). The 'About this run' section on the right provides details about the run, including the created time, created by, experiment ID, status, run ID, duration, source, and registered prompts.

Metric	Value	Model
MAE	0.4742404593121134	model-GradientBoosting
R2	0.662004727263539	model-GradientBoosting
RMSE	0.6653163710048351	model-GradientBoosting

Parameter	Value
learning_rate	0.05
max_depth	3
n_estimators	300
random_state	42

About this run

- Created at: 10/19/2025, 07:19:19 PM
- Created by: diego
- Experiment ID: 233219589045902434
- Status: Finished
- Run ID: fd5d3dbcd019476a8dee40b99975ae64
- Duration: 12.1s
- Source: c:\Users\diego\miniconda7\envs\mlflow\lib\site-packages\ipykernel_launcher.py
- Registered prompts: —
- Datasets: None
- Tags: Add tags
- Registered models: None

Model attributes	Type	Step	Model name	Status	Created	Registered models	Dataset	MAE	R2
Output	0	model-GradientBo	Ready	2 minutes ago	-	-	0.4742404593121134	0.662004727263539	

RandomForest

california-housing-prediction > Run > RandomForest-run

Overview Model metrics System metrics Traces Artifacts

Description No description

Metrics (3)

Metric	Value	Models
MAE	0.4615835214036077	model-RandomForest
R2	0.6764184594911478	model-RandomForest
RMSE	0.651173721255115	model-RandomForest

Parameters (4)

Parameter	Value
max_depth	12
n_estimators	300
n_jobs	-1
random_state	42

Logged models (1)

Model attributes							No dataset	
Type	Step	Model name	Status	Created	Registered models	Dataset	MAE	R2
Output	0	model-RandomFor	Ready	3 minutes ago	-	-	0.4615835214036077	0.67641845949114

About this run

Created at 10/19/2025, 07:19:12 PM
Created by diego
Experiment ID 233219589046902434
Status Finished
Run ID 27f66a02dfe5401d8f1462d1381b0bdc
Duration 7.2s
Source c:\Users\diego\miniconda3\envs\mlflow-lab\lib\site-packages\ipykernel_launcher.py
Registered prompts ---
Datasets None
Tags [Add tags](#)
Registered models None

LinearRegression

california-housing-prediction > Run > LinearRegression-run

Overview Model metrics System metrics Traces Artifacts

Description No description

Metrics (3)

Metric	Value	Models
MAE	0.53550466964532	model-LinearRegression
R2	0.5817648126077724	model-LinearRegression
RMSE	0.7403101678182196	model-LinearRegression

Parameters (2)

Parameter	Value
fit_intercept	True
n_jobs	-1

Logged models (1)

Model attributes							No dataset	
Type	Step	Model name	Status	Created	Registered models	Dataset	MAE	R2
Output	0	model-LinearRegr	Ready	4 minutes ago	-	-	0.53550466964532	0.58176481260777

About this run

Created at 10/19/2025, 07:19:08 PM
Created by diego
Experiment ID 233219589046902434
Status Finished
Run ID 714ed3d02a684d0992c3c78228549b4d
Duration 3.4s
Source c:\Users\diego\miniconda3\envs\mlflow-lab\lib\site-packages\ipykernel_launcher.py
Registered prompts ---
Datasets None
Tags [Add tags](#)
Registered models None

Comandos de Feast

```
(mlops-lab3) C:\Users\diego\Documents\UVG\18mo Semestre\MachineLearning\MLOps-Lab3>feast init california_feature_store
Creating a new Feast repository in C:\Users\diego\Documents\UVG\18mo Semestre\MachineLearning\MLOps-Lab3\california_feature_store.

(mlops-lab3) C:\Users\diego\Documents\UVG\18mo Semestre\MachineLearning\MLOps-Lab3>cd california_feature_store
(mlops-lab3) C:\Users\diego\Documents\UVG\18mo Semestre\MachineLearning\MLOps-Lab3\california_feature_store>cd feature_repo

(mlops-lab3) C:\Users\diego\Documents\UVG\18mo Semestre\MachineLearning\MLOps-Lab3\california_feature_store\feature_repo>feast apply
c:\Users\diego\miniconda3\envs\mlops-lab3\lib\site-packages\feast\infra\online_stores\sqlite.py:91: UserWarning: Field name "vector_enabled" in "SqliteOnlineStoreConfig" shadows an attribute in parent "VectorStoreConfig"
  class SqliteOnlineStoreConfig(FeastConfigBaseModel, VectorStoreConfig):
c:\Users\diego\miniconda3\envs\mlops-lab3\lib\site-packages\feast\infra\online_stores\sqlite.py:91: UserWarning: Field name "vector_len" in "SqliteOnlineStoreConfig" shadows an attribute in parent "VectorStoreConfig"
  class SqliteOnlineStoreConfig(FeastConfigBaseModel, VectorStoreConfig):
No project found in the repository. Using project name california_feature_store defined in feature_store.yaml
Applying changes for project california_feature_store
Created project california_feature_store
Created entity district_id
Created feature view california_stats
Created sqlite table california_feature_store_california_stats

(mlops-lab3) C:\Users\diego\Documents\UVG\18mo Semestre\MachineLearning\MLOps-Lab3\california_feature_store\feature_repo>feast materialize-incremental 2025-10-19
c:\Users\diego\miniconda3\envs\mlops-lab3\lib\site-packages\feast\infra\online_stores\sqlite.py:91: UserWarning: Field name "vector_enabled" in "SqliteOnlineStoreConfig" shadows an attribute in parent "VectorStoreConfig"
  class SqliteOnlineStoreConfig(FeastConfigBaseModel, VectorStoreConfig):
c:\Users\diego\miniconda3\envs\mlops-lab3\lib\site-packages\feast\infra\online_stores\sqlite.py:91: UserWarning: Field name "vector_len" in "SqliteOnlineStoreConfig" shadows an attribute in parent "VectorStoreConfig"
  class SqliteOnlineStoreConfig(FeastConfigBaseModel, VectorStoreConfig):
Materializing 1 feature views to 2025-10-18 18:00:00-06:00 into the sqlite online store.
california_stats from 2015-10-22 20:12:23-06:00 to 2025-10-18 18:00:00-06:00:
Bit [00:00, 7it/s]
```

Conclusiones

- El ingreso medio es el predictora dominante para el valor promedio de una vivienda, mientras que habitaciones aporta de forma moderada a la predicción y características como latitud y longitud no tiene influencia alguna sobre la predicción del valor medio de una vivienda en California. La distribución de la variable objetivo es sesgada a la derecha y truncada en 5.0 limitando el techo de predicción.
- La regresión lineal obtuvo un R cuadrado de alrededor de 0.7 al menos 0.15 puntos sobre sus contrapartes de Random Forest y Gradient Boosting, indicando que la regresión captura mejor las interacciones y asimetría del valor medio de una vivienda.
- El uso de tecnologías como MLFlow y Feast, permiten el registro correcto de parámetros, métricas y modelos, facilitando la selección automática del mejor. Así como almacenamiento y recuperación de features de forma consistente para su uso futuro, generando en este caso predicciones coherentes de entre 0.7 y 4.4 cientos de miles del valor medio.