LISTADO DE MODELOS PROBADOS

# obs: all the models loaded are the models that can be used in gurobi-py (this codes are generated to train models that the ds can undertand completly

# its behavior to add as constraint in optimization model

LISADO DE MODELOS

- regresión lineal

- árbol decisión

- gradient boosting regressor

- random forest

- xgboost

- mlp

<https://gurobi-machinelearning.readthedocs.io/en/stable/auto_generated/gurobi_ml.sklearn.gradient_boosting_regressor.html>

PASOS EN EL CÓDIGO:

- Se obtienen los datos

- Se hace una limpieza básica

Detallar

Finalizar con Split train y test

- Se entrenan una serie de modelos. Modelos que acepta gurobi ml. Este repo forma parte de un proyecto más grande que busca conectar modelos regressor con gurobi y estos códigos buscan identificar el mejor modelo

Detallar modelos

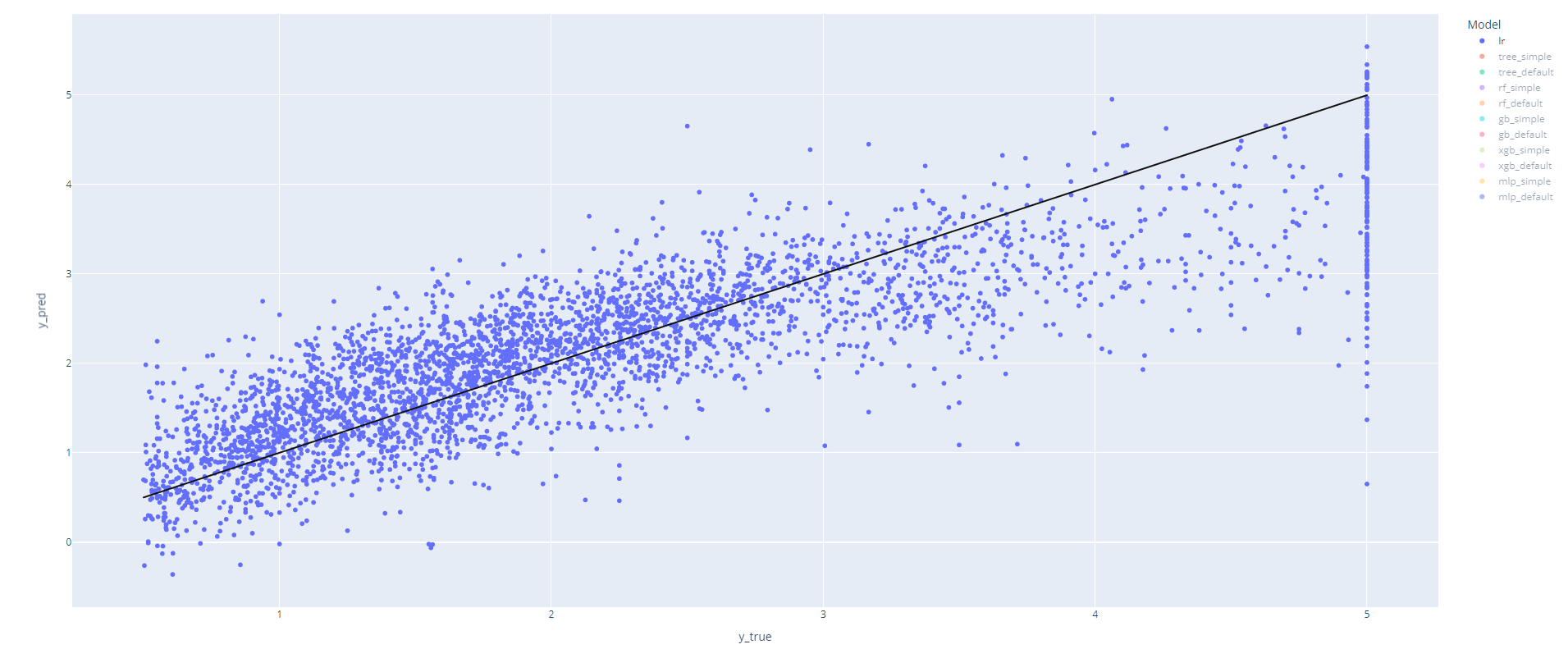
- Se evalúan los modelos bajo diferentes pespectivas

- Métricas

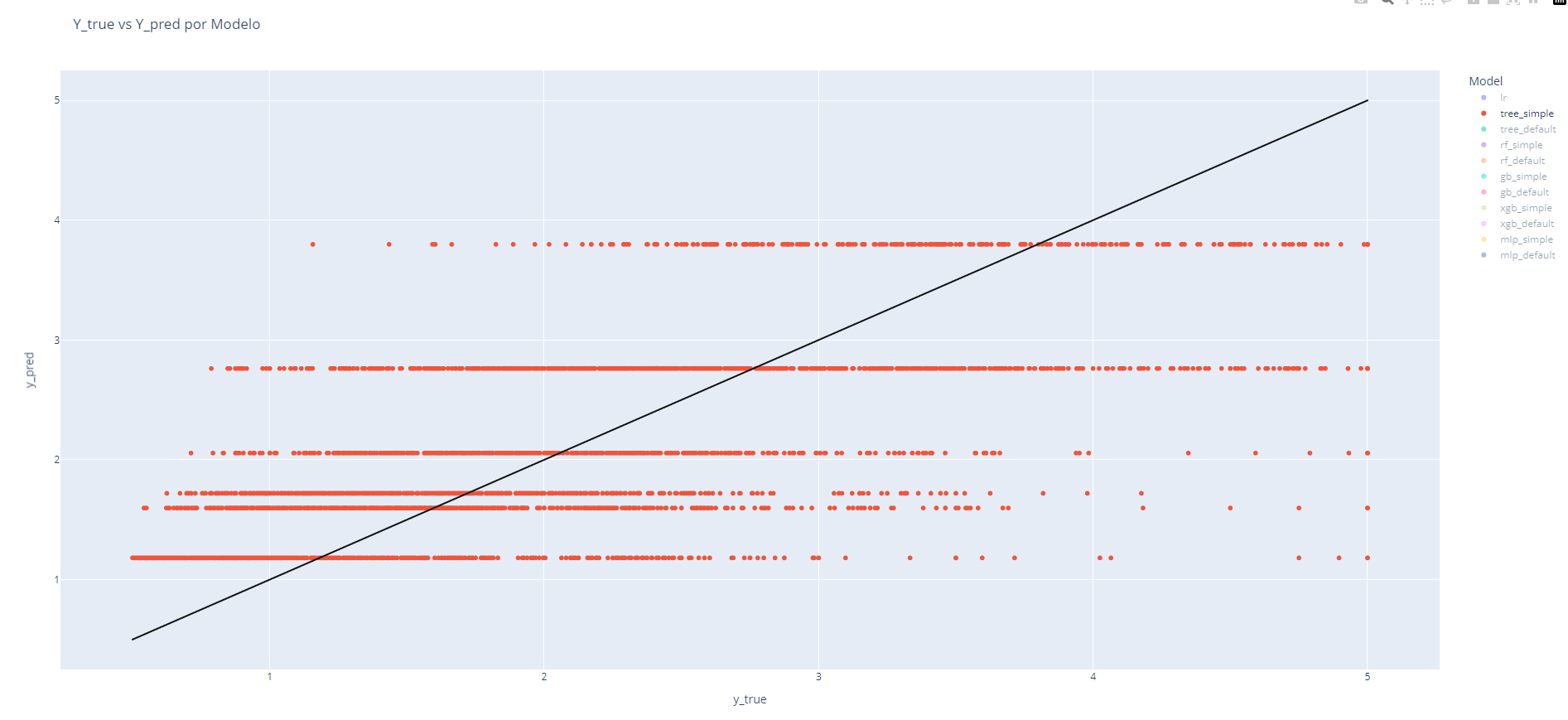
- plot y\_true vs y\_pred

EJEMPLO COMPARANDO VISDUALIZAICONES EN DATOS TEST

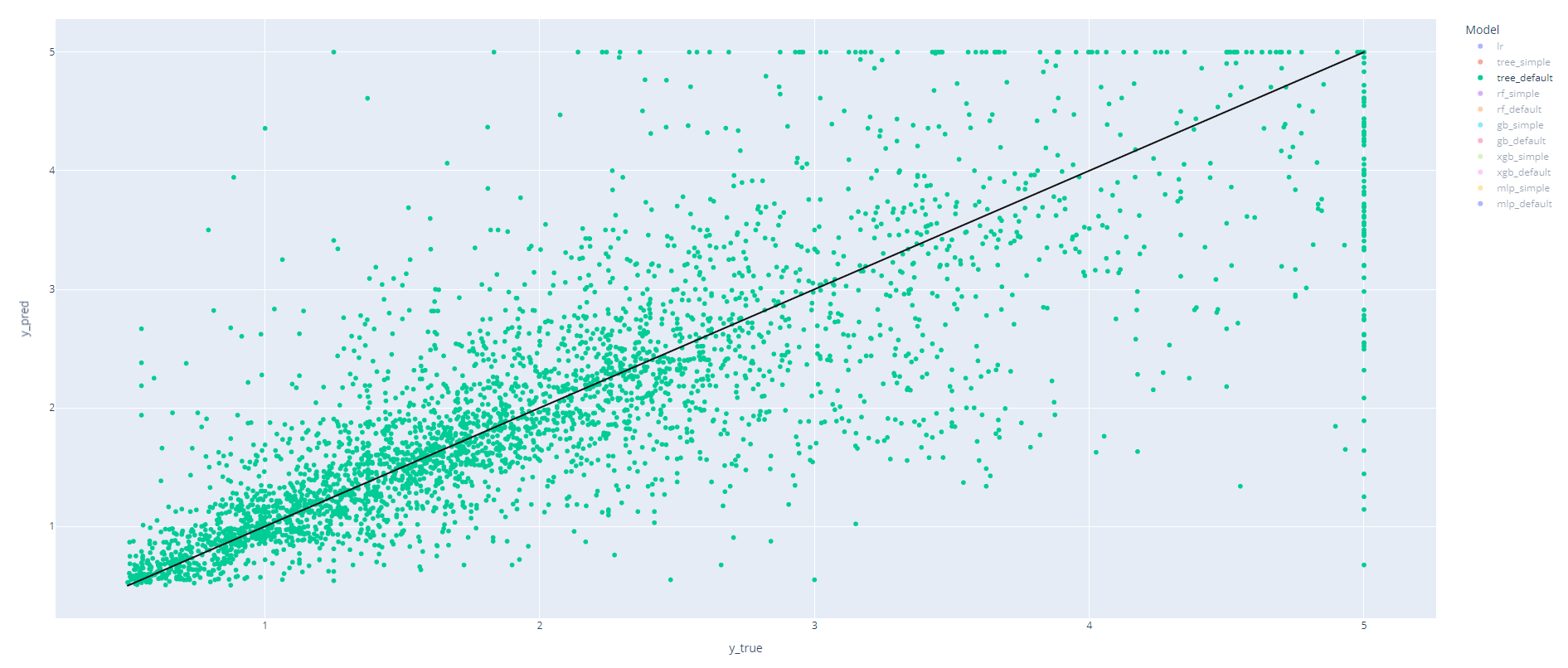
Regresión lineal



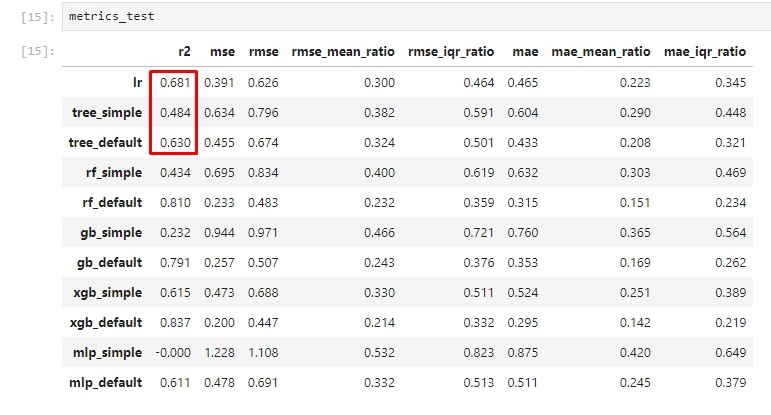
Arbol de decisión. Se queda pegado prediciendo un valor



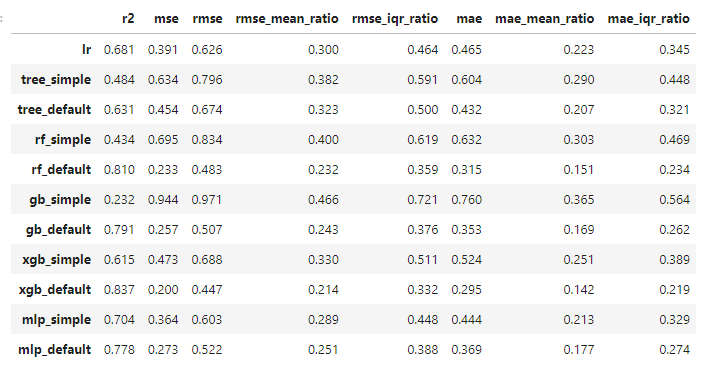
Arbol de decisión 2



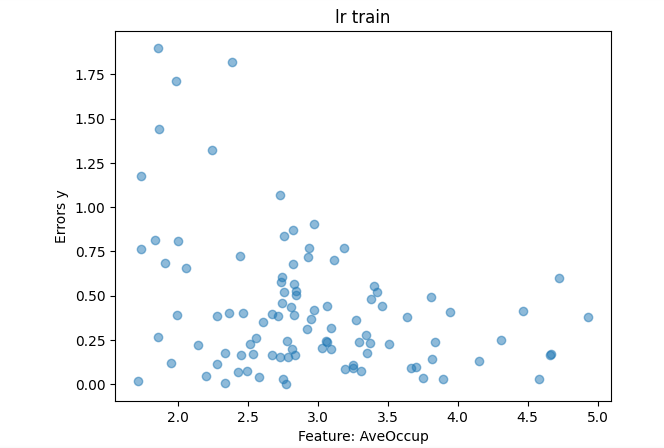
Visualmente, claramente el árbol de decisión es peor. Pero, mirando solo las métricas, los valores no son tan diferentes de la regresión lineal



En las metrics scaler se puede observar que escalar los datos no afecta en las métricas, excepto en las redes neuronales donde es sabido que es necesario scalar los datos entre 0 y 1, donde los valores mejoran considerablemente



Para ver los errores, aquí se ve que cuando AveOccup toma valores bajos los errores de la predicción son mayores



**MUY IMPORTANTE: COMO SE PUEDE VER HAY TRANSFORMACIONES DE LOS DATOS QUE TIENE QUE IR DENTRO DEL PKL DEL MODELO**

Si hay una feature del modelo cuyo valor se relaciona con el resto de restricciones por ejemplo una feature que toma valores entre 10 y 20. Es necesario que si el modelo hace un scaler de los datos, este scaler esté dentro de los datos, porque sino los datos no van a coincidir con el orden de magnitud.

Pero otro lado, otras transforciones como por ejemplo suavizado de las señales de un sensor, no es necesario esté dentro del modelo porque se puede asumir que todas las variables de decisión del optimizador también están suavizadas.