Bitácora de “El Arca”

Noe Hsueh y Axel Fridman

Nuestra idea fue empezar con el modelo ya funcional de Pablo y expandirlo con LGB agregando nuevas variables y probando agregar lo qué mejoraba el modelo y descartando lo qué no.

El orden en el qué fuimos probando cosas es el siguiente:

- usar el título

- usar la descripción

- ajustar los estimadores de lgbm considerando trade off entre tiempo de cómputo y error en dev. La qué mejor nos dio fue la de 500 estimadores, pero para modelos más complejos usamos la de 100 por simplicidad y tiempo.

- usar precio promedio de los barrios en vez de barrio

- usar también precio x m2 de un radio de 20

Probamos después combinando modelos: random forest + lgbm y hacer un ensamble estimando coef regresión de ridge. Sin embargo este enfoque no dio frutos. No pudimos analizar porque.

Los errores de análisis qué hicimos se enfocaron en analizar a mano las propiedades con mayor error. En donde un estacionamiento estaba a la venta a demolición y eso es algo inusual, o eran propiedades muy caras e incluso alguna qué (intuimos) el precio lo habían puesto en pesos pero la currency en USD.

Notamos qué a nivel temporal el random forest era lo qué generaba nuestro mayor coste computacional llegando a tardar más de 1:30hs en entrenarse y unos 5 min en predecir. Principalmente por lo pesado qué es analizar la descripción. En contraste el LGB tardaba un décimo del tiempo en todas las tareas.