

Modelo de predicción de precios de hoteles de alojamiento

FUENTE DE DATOS: PLATAFORMA DE
AIRBNB

AUTOR: MAURO RODRÍGUEZ

OBJETIVOS

El objetivo será entrenar un algoritmo que nos permita predecir los precios de los alojamientos en Estados Unidos a partir de los datos que provee la plataforma de Airbnb dedicada a la oferta de alojamientos. A través de la misma, es posible que anfitriones y huéspedes pueden valorarse mutuamente, para ofrecer referencias.

CONTEXTO

Obtuvimos un dataset que contiene 74111 registros de propiedades, con sus características y respectivos precios de alojamiento. Nuestro proyecto buscará generar un modelo que pueda predecir los precios con precisión y que determinar qué variables son relevantes para explicar los resultados.

MODELOS SELECCIONADOS

Dado que nuestra variable objetivo es numérica utilizaremos técnicas de regresión. Hemos seleccionado 3 algoritmos de ensambles basados en árboles de decisión: Random Forest, Gradient Boosting y XGBoost.

Si bien los tres modelos dieron resultados similares en el conjunto de testeo, Random Forest tendió a sobreajustar más que los modelos boosting. Estos últimos tendieron a generalizar mejor.

INSIGHTS PRELIMINARES

Lo que nos devuelven los modelos y la exploración realizada en el EDA, es que la variable más predominante para explicar los precios de los alquileres es el tipo de lugar que se alquila. Alquilar un casa entera es más caro que alquilar una habitación. También influyen los inmuebles con más de un baño y las ciudades en donde se asientan (San Francisco parece ser más cara que las demás consideradas)

RESULTADOS

RANDOM FOREST

MAE: u\$d 0.33

GRADIENT BOOSTING

MAE: u\$d 0.32

XGBOOST

MAE: u\$d 0.32