

Primera entrega de proyecto

POR:

Alejandro Herrera Rivera

Juan José Isaza

MATERIA:

Introducción a la inteligencia artificial

PROFESOR:

Raul Ramos Pollan



UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

MEDELLÍN 2023

1. Planteamiento del problema

En el competitivo mercado del alquiler de propiedades tanto los huéspedes como los anfitriones siempre tienen la inquietud de saber ¿cómo determinar el precio adecuado para una propiedad dada? Los anfitriones desean establecer tarifas competitivas que atraigan a huéspedes potenciales, maximicen sus ingresos y reflejen con precisión el valor de su propiedad. Por otro lado, los huéspedes buscan ofertas atractivas que se ajusten a su presupuesto mientras satisfacen sus necesidades de alojamiento.

Sin embargo, la fijación de precios en este mercado dinámico y diverso es una tarea complicada. Los precios pueden variar significativamente según la ubicación, el tipo de propiedad, el número de habitaciones, las comodidades ofrecidas y otros factores.

El trabajo consta de la creación de un modelo de predicción de precios para propiedades de Airbnb en Londres, utilizando datos que incluyen información sobre la ubicación, características de las propiedades y descripciones de los anuncios.

2. Dataset

El dataset a utilizar proviene de una competencia de kaggle en la cual se proporcionan datos de las propiedades de Airbnb en Londres. El dataset está compuesto por un conjunto de archivos .csv que proporcionan la información requerida tales como la descripción de las variables, los datos de entrenamiento y los datos de prueba.

El archivo que contiene los datos de entrenamiento es nombrado como *train.csv* y contiene la siguiente información:

- **name** = Nombre de la propiedad en Airbnb
- **summary** = Resumen de sus características
- **space** = Descripción de sus áreas
- **description** = Resumen de sus características
- **experiences_offered** = Experiencias ofrecidas
- **neighborhood_overview** = Descripción del barrio
- **transit** = Descripción de la locación y el transporte
- **house_rules** = Reglas de la propiedad
- **picture_url** = Link de la imagen
- **host_id** = Identificador del anfitrión
- **host_since** = Fecha desde que es anfitrión
- **host_response_time** = Tiempo de respuesta del anfitrión
- **host_response_rate** = Índice de respuesta del anfitrión
- **host_is_superhost** = El anfitrión es superanfitrión (V o F)

- **host_total_listings_count** = Recuento de listados del anfitrión
- **host_has_profile_pic** = El anfitrión tiene foto de perfil (V o F)
- **host_identity_verified** = Identidad del anfitrión verificada (V o F)
- **neighbourhood** = Barrio
- **neighbourhood_cleansed** = Distrito
- **zipcode** = Código postal
- **latitude** = Latitud
- **longitude** = Longitud
- **property_type** = Tipo de propiedad
- **room_type** = Tipo de habitación
- **accommodates** = #Capacidad
- **bathrooms** = #Baños
- **bedrooms** = #Habitaciones
- **beds** = #Camas
- **bed_type** = Tipo de cama
- **amenities** = Servicios
- **price** = Precio (Variable objetivo)
- **guests_included** = #Huéspedes incluidos en el precio
- **review_scores_rating** = Puntuación del huésped
- **review_scores_accuracy** = Precisión en las puntuaciones
- **review_scores_cleanliness** = Puntuación a la limpieza
- **review_scores_checkin** = Puntuación experiencia en el check-in
- **review_scores_communication** = Puntuación comunicación con el anfitrión
- **review_scores_location** = Puntuación a la ubicación de la propiedad
- **review_scores_value** = Puntuación a la relación calidad-precio
- **cancellation_policy** = Política de cancelación del anfitrión
- **reviews_per_month** = Reviews por mes
- **listing_id** = Id listado propiedad

3. Métricas

La métrica de evaluación principal del modelo es el RMSE (Root Mean Squared Error). Se utiliza para medir la precisión de un modelo de regresión al evaluar cuán cerca están las predicciones del modelo de los valores reales.

$$RECM = \sqrt{\frac{\sum (O_i - E_i)^2}{n}}$$

donde O_i son los valores observados;

E_i son los valores esperados;

Σ es una letra griega llamada sigma que representa la «sumatoria»; y

n es el tamaño de la muestra (el número de observaciones).

La métrica de negocio está basada en que el valor predicho para la propiedad se ajuste adecuadamente a lo que ofrece, ya que sería injusto para una persona pagar un valor grande por un servicio que realmente no cumple con las necesidades que los clientes requieran. Además permite establecer un precio adecuado para un anfitrión que desea ingresar a la aplicación ofreciendo un lugar.

Además se tendrá en cuenta la métrica R cuadrado que permite saber en cuanto el modelo se ajusta a los datos reales.

4. Desempeño

Por la razón anterior se espera que el RMSE sea lo más cercano a cero y un R cuadrado cercano a 1, es decir que el modelo sea capaz de tener un buen ajuste para predecir con precisión en relación con los valores reales. Indicando que es capaz de aprender de las características de las propiedades para definirles un precio ajustado y relacionado a lo que ofrece el servicio.

Finalmente para airbnb es importante tener precios competitivos, ofrecer una grata experiencia y aumentar la confianza en su servicio, además un buen ajuste del valor de los servicios permite comodidad tanto para los huéspedes como para los anfitriones.

5. Bibliografía

- georgvelev. (2023). ADAMS SoSe23. Kaggle.
<https://kaggle.com/competitions/adams-bose23>