

CIENCIA DE DATOS APLICADA

QUIZ 003

Diego Alberto Rodríguez Cruz

Este informe presenta el desarrollo de un modelo de Machine Learning para la predicción de precios de venta de apartamentos en Bogotá, Colombia, desarrollado para HabitAlpes. Se describirá cómo funciona el modelo, se determinará su rentabilidad y se ofrecerán insights relevantes que pueden ser utilizados para mejorar la toma de decisiones en la empresa.

Datos

El modelo se basa en un conjunto de datos que incluye diversas características de apartamentos, tales como:

Columna	Descripción
código	Código único que identifica cada apartamento
tipo_propiedad	Tipo de propiedad (apartamento, casa, etc.)
tipo_operacion	Tipo de operación (venta, arriendo, etc.)
precio_venta	Precio de venta del apartamento COP
precio_arriendo	Precio de arriendo del apartamento COP
área	Área del apartamento en metros cuadrados
habitaciones	Número de habitaciones del apartamento
baños	Número de baños del apartamento
administración	Valor de la cuota de administración del apartamento
parqueaderos	Número de parqueaderos disponibles
sector	Sector o zona en la que se encuentra el apartamento
estrato	Estrato socioeconómico del apartamento
antigüedad	Antigüedad del apartamento en años
estado	Estado del apartamento (nuevo, usado)
longitud	Longitud geográfica del apartamento
latitud	Latitud geográfica del apartamento
descripción	Descripción detallada del apartamento
datetime	Fecha y hora de extracción de los datos
jacuzzi	Indica si el apartamento cuenta con jacuzzi
piscina	Indica si el apartamento cuenta con piscina
salon_comunal	Indica si el apartamento cuenta con salón comunal
terrazza	Indica si el apartamento cuenta con terraza
vigilancia	Indica si el apartamento cuenta con vigilancia privada
piso	Número de piso en el que se encuentra el apartamento
closets	Número de closets en el apartamento
chimenea	Indica si el apartamento cuenta con chimenea

permite_mascotas	Indica si se permiten mascotas en el apartamento
gimnasio	Indica si el apartamento cuenta con gimnasio
ascensor	Indica si el edificio cuenta con ascensor
conjunto_cerrado	Indica si el apartamento se encuentra en conjunto cerrado
coords_modified	Coordenadas modificadas del apartamento
localidad	Localidad en la que se encuentra el apartamento
barrio	Barrio en el que se encuentra el apartamento
estacion_tm_cercana	Nombre de la estación de transporte masivo más cercana
distancia_estacion_tm_m	Distancia a la estación de transporte masivo más cercana
is_cerca_estacion_tm	Indica si está cerca de una estación de transporte masivo $\leq 500m$
parque_cercano	Nombre del parque más cercano al apartamento
distancia_parque_m	Distancia al parque más cercano al apartamento en metros
is_cerca_parque	Indica si está cerca de un parque $\leq 500m$
website	Sitio web relacionado a la propiedad
compañía	Compañía o agencia responsable de la propiedad
last_view	Fecha de la última vez que el scraper visito el apartamento
timeline	Historial de precios del apartamento
url	URL del apartamento

En los anteriores datos, se encontraron valores faltantes para algunas columnas, las cuales se describen a continuación

Columna	Cantidad nulos/vacíos
precio_venta	15429
habitaciones	1
banos	1
administración	7925
parqueaderos	1
sector	1641
estrato	1
antigüedad	10
dirección	42219
website	2
last_view	2
datetime	2
url	42219
timeline	13586
estado	796
compañía	4530

precio_arriendo	27184
jacuzzi	2
piso	43013
closets	43013
chimenea	2
permite_mascotas	2
gimnasio	2
ascensor	2
conjunto_cerrado	2
piscina	2
salon_comunal	2
terraza	2
vigilancia	2
barrio	193

Los datos fueron ajustados según sus características, así como se descartaron registros para el desarrollo del modelo.

Dentro de los valores a seleccionar, se tuvo en cuenta que la variable denominada tipo_operacion sea diferente a ARRIENDO dado que al revisar el precio_venta no existe para estas unidades habitacionales, así mismo, se descartaron precios_venta por encima de los 1'850.000.000 con el fin de evitar sesgos en valores atípicos y únicos para precios altos.

Modelo

Se utiliza un modelo de regresión, con el fin de predecir el precio de venta de propiedades en la ciudad de Bogotá. Las variables seleccionadas son:

- precio_venta
- area
- habitaciones
- banos
- sector
- tipo_propiedad
- tipo_operacion
- parqueaderos
- localidad
- barrio
- is_cerca_estacion_tm
- is_cerca_parque

El modelo fue entrenado utilizando el conjunto de datos con las variables seleccionadas, el cual se dividió en conjuntos de entrenamiento y prueba. Se emplearon métricas de rendimiento como el Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) y el

R^2 para evaluar su precisión. El AUC de 0.83 obtenido en la evaluación del modelo sugiere que tiene una buena capacidad de discriminación en la clasificación de los precios.

Rentabilidad del modelo

La implementación del modelo permite reducir el tiempo necesario para llevar a cabo un avalúo de precios de 6 horas a solo 1 hora. Dado que el costo del trabajo de un perito es de 9,500 COP por hora, esto representa un ahorro de:

- Ahorro por avalúo: $6 \text{ horas} \times 9,500 \text{ COP/hora} = 57,000 \text{ COP}$
- Ahorro mensual (500 apartamentos): $500 \times 57,000 \text{ COP} = 28,500,000 \text{ COP}$

Insights

- Identificación de oportunidades de mercado: analizar los datos y las predicciones puede revelar áreas geográficas o segmentos de mercado donde la demanda es mayor o donde hay una escasez de oferta. Esto puede guiar a HabitAlpes en la adquisición de propiedades o en la promoción de desarrollos.
- Personalización de servicios: los resultados del modelo pueden ser utilizados para personalizar los servicios ofrecidos a los clientes, ajustando las recomendaciones basadas en las características específicas de cada propiedad y las preferencias del cliente.