

QUIZ 3 – Ciencia de Datos Aplicada

Estimador de precio de venta de apartamentos de HabitAlpes

Introducción

- **Contexto de la Empresa:** HabitAlpes es una startup colombiana que ofrece un portafolio de servicios de consultoría en vivienda, incluyendo avalúo de propiedades, remodelación, arrendamiento, y sugerencias de reclasificación de estrato en Bogotá.
- **Objetivo del Proyecto:** HabitAlpes implementar una herramienta de estimación de precios de venta de apartamentos en Bogotá, que permita a los propietarios conocer el valor sugerido de sus inmuebles basándose en características específicas y la ubicación. Esta herramienta ayudará a HabitAlpes a ofrecer precios de venta competitivos en el mercado y a optimizar la efectividad de sus servicios de avalúo.

Implementación del Modelo

- **Datos:** La empresa ha recopilado datos de apartamentos en Bogotá en los últimos dos meses, incluyendo variables relevantes, los cuales se encuentran constatados en un archivo de tipo CSV.
- **Selección de características:** Se eligen las variables más relevantes para estimar el precio de venta, como *area*, *parqueaderos*, *habitaciones*, *estrato*, y otras que capturan servicios y características del apartamento.
- **Transformación de la variable *antigüedad*:** Se aplica un diccionario de mapeo para convertir categorías de antigüedad en valores numéricos, lo cual es clave para preparar los datos para el modelo de regresión.
- **Filtrado de datos:** Se seleccionan solo las propiedades en venta que son apartamentos, lo cual se alinea con el objetivo del modelo.
- **Limpieza de valores nulos:** Los registros con valores nulos en las características seleccionadas o en el precio de venta (variable objetivo) son eliminados para garantizar datos completos para el entrenamiento.
- **Eliminación de registros duplicados:** Mantener registros duplicados, aún sin considerarse un problema de calidad de los datos, ocasionará que durante el proceso de aprendizaje el algoritmo otorgue de forma errónea un peso mayor a estos registros duplicados.
- **Análisis de correlación:** Se exploran relaciones entre las variables de entrada y el precio de venta para validar la relevancia de las características.
- **Particionamiento del conjunto de datos en entrenamiento y prueba:** El conjunto de datos se dividió en entrenamiento y prueba (80/20) para validar la precisión del modelo y evitar el sobreajuste.
- **Modelo utilizado – Regresión lineal:** Se seleccionó una regresión lineal múltiple por su capacidad de modelar la relación entre múltiples variables predictoras y el precio objetivo, proporcionando interpretabilidad y precisión en un contexto de precios de apartamentos.

- **Entrenamiento y Evaluación:**

Al implementar la etapa de entrenamiento y evaluación se estuvo haciendo seguimiento al Error Medio Absoluto (MAE) y la Raíz del Error Cuadrático Medio (RMSE). Durante el desarrollo del modelo se vislumbró la posibilidad de relaciones no lineales entre la variable objetivo y las variables de entrada, por lo que se hizo una regresión de grado dos para indagar si existían interacciones entre las variables que el modelo inicial no estaba teniendo en cuenta. De esto se encontró, revisando el MAE y la RMSE, que las relaciones entre las variables eran más lineales que polinómicas.

Por otro lado, al hacer la revisión del error cometido por nuestro modelo se hicieron ajustes con respecto a los *outliers* en los datos para evitar subestimaciones y sobrestimaciones por parte de este.

- **Resultados:**

- Los resultados de las métricas obtenidos para ambos conjuntos de datos (train y test) fueron:

- **MAE:**

Train: 57416542.19396984

Test: 57348010.79346638

- **RMSE:**

Train: 76140844.37901548

Test: 75705093.76613835

El MAE final obtenido indica, en promedio, el error promedio de las predicciones de precio en COP. El RMSE sugiere la desviación promedio en COP entre las predicciones y los valores reales. Dado que el RMSE es cercano al MAE, el modelo mantiene una precisión consistente sin grandes discrepancias en la mayoría de los casos, lo cual es favorable para ofrecer estimaciones confiables a los clientes de HabitAlpes.

Al hacer predicciones utilizando los datos de prueba se obtuvo que aproximadamente el 33.78% de los precios estimados por el modelo tienen un déficit superior a COP 20'000,000 con respecto al precio real del inmueble. De esto se obtiene que en el 66.22% de los casos NO será necesario que el perito haga un avalúo presencial a los apartamentos, dado que se prevé que el cliente lo solicitará solo cuando la estimación esté por debajo de 20 millones del valor del mismo.

Evaluación de Rentabilidad

- **Ahorro en costos relacionados a horas de trabajo de los peritos:**

Al mes...

- **Costo actual:** $\$9,500 * 6 * 500 = \$57,000 * 500 = \$28'500,000$

- **Costo con el modelo:** $(\$57,000 * 500 * 0.3378) + (\$9,500 * 500 * 0.6622) = \$12'772,750$

Ahorro mensual = $\$28'500,000 - \$12',772,750 = \$15'727,250$

$$\text{Ahorro anual} = \$15'727,250 * 12 = \$188'727,000$$

- **Impacto en el Cliente:** El modelo de estimación de precios de HabitAlpes ofrece a los clientes la conveniencia de obtener una valoración rápida y precisa sin necesidad de una visita física, lo que ahorra tiempo y costos. Esta agilidad permite a los clientes tomar decisiones informadas en un mercado competitivo como el de Bogotá, posicionando a HabitAlpes como un aliado que responde a las necesidades actuales del sector inmobiliario.
- **Riesgo de Estimaciones Inexactas:** En aproximadamente el 34% de los casos, cuando el modelo subestima el precio en más de 20 millones, se requiere un avalúo presencial para garantizar la precisión. Aunque esto puede afectar la percepción de fiabilidad del servicio, HabitAlpes puede informar a los clientes sobre la naturaleza probabilística de la estimación, resaltando que el modelo está diseñado para ahorrar tiempo en la mayoría de los casos y reforzar la transparencia explicando que estas verificaciones aseguran estimaciones precisas y reflejan el compromiso de la empresa con la satisfacción del cliente.

Insights y Recomendaciones

- **Principales Insights:** El análisis del modelo reveló que las variables más influyentes en el precio de los apartamentos en Bogotá son el área, el estrato, la cantidad de parqueaderos, el precio de la administración, la antigüedad y la distancia a una estación de Transmilenio. Estos hallazgos permiten a HabitAlpes ajustar sus recomendaciones según las características clave de cada propiedad, ayudando a los clientes a obtener precios competitivos en el mercado. Por ejemplo, el énfasis en el estrato puede guiar a los clientes en sus decisiones de compra o venta en zonas de alta demanda.
- **Recomendaciones para Optimización:** Sugiere que HabitAlpes integre la herramienta de estimación en otros procesos, como asesorías de compra o remodelación, para maximizar su utilidad. Además, podrías recomendar un seguimiento de los resultados del modelo en el tiempo, para ajustar y mejorar la precisión en futuras estimaciones.

Conclusiones

- La implementación del modelo de estimación de precios permite a HabitAlpes reducir significativamente los costos y el tiempo de evaluación, disminuyendo el tiempo de avalúo de 6 a 1 hora en la mayoría de los casos. Esto aumenta la eficiencia operativa, permitiendo procesar más evaluaciones al mes, y ofrece a los clientes un servicio ágil y económico, mejorando la experiencia y el valor añadido.
- Para maximizar los beneficios, se recomienda perfeccionar el modelo a través de ajustes periódicos basados en los resultados obtenidos. Además, HabitAlpes podría integrar esta herramienta en otros servicios, como asesorías de compra o remodelación, ampliando su alcance y utilidad para los clientes.