

CIENCIA DE DATOS APLICADA - MINE 4101

William Yesid Toro González Código: 201112526

Lina Maria Ojeda Amaya Código: 202112324

Resumen ejecutivo – Apoyo a una empresa de servicios inmobiliarios

Resumen Ejecutivo

Entendimiento y preparación de los datos:

Este proyecto desarrolla un modelo de predicción de precios de apartamentos en Bogotá para la empresa HabitAlpes, con el propósito de automatizar y optimizar el proceso de valoración inmobiliaria. A partir de un dataset real con características estructurales, geográficas y administrativas de miles de inmuebles, se construyó una solución de analítica avanzada capaz de estimar el precio de venta con alta precisión, interpretabilidad y valor económico para la operación.

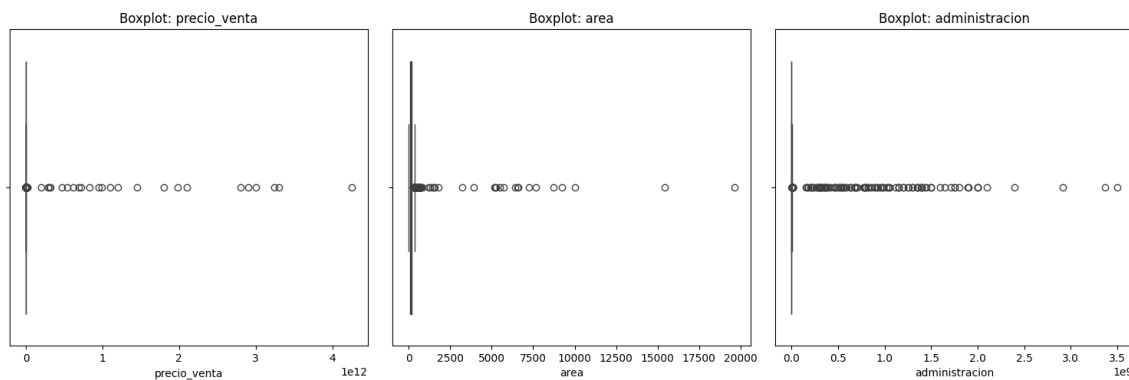


Figura 1. Exploración de variables numéricas dataset

El proceso inició con una depuración detallada de los datos, tratamiento de valores faltantes, categorización de variables poco frecuentes y estandarización de columnas numéricas.

Entrenamiento del modelo de Machine Learning:

Se enriqueció la información incorporando nuevas variables geográficas mediante la fórmula de Haversine, así como una característica derivada del precio promedio por metro cuadrado de los 50 vecinos más cercanos, calculada con BallTree. Aunque estas transformaciones añadieron estructura espacial al modelo, su contribución al desempeño fue un 1% comparado con los resultados obtenidos por XGboost que no fue entrenado con estas transformaciones de incorporación lo cual evidencia que las coordenadas latitud/longitud ya capturan la mayor parte de la variabilidad espacial relevante.

Análisis cuantitativo de resultados del modelo:

Se implementaron y compararon dos modelos principales: CatBoost Regressor y XGBoost Regressor. Ambos alcanzaron métricas destacadas, con R^2 cercanos al 90%, MAE alrededor de 64 millones de pesos y un error relativo (MAPE) cercano al 10%. Sin embargo, XGBoost obtuvo el mejor desempeño, mostrando un equilibrio superior entre error, estabilidad y generalización entre los conjuntos de validación y prueba, lo cual lo posiciona como el modelo final recomendado.

Análisis cualitativo de resultados del modelo:

Se aplicaron técnicas de interpretabilidad global y local a través de SHAP y LIME sobre el mejor modelo. Los resultados identificaron como variables más influyentes el área construida, administración, ubicación espacial (latitud y longitud), número de parqueadero.

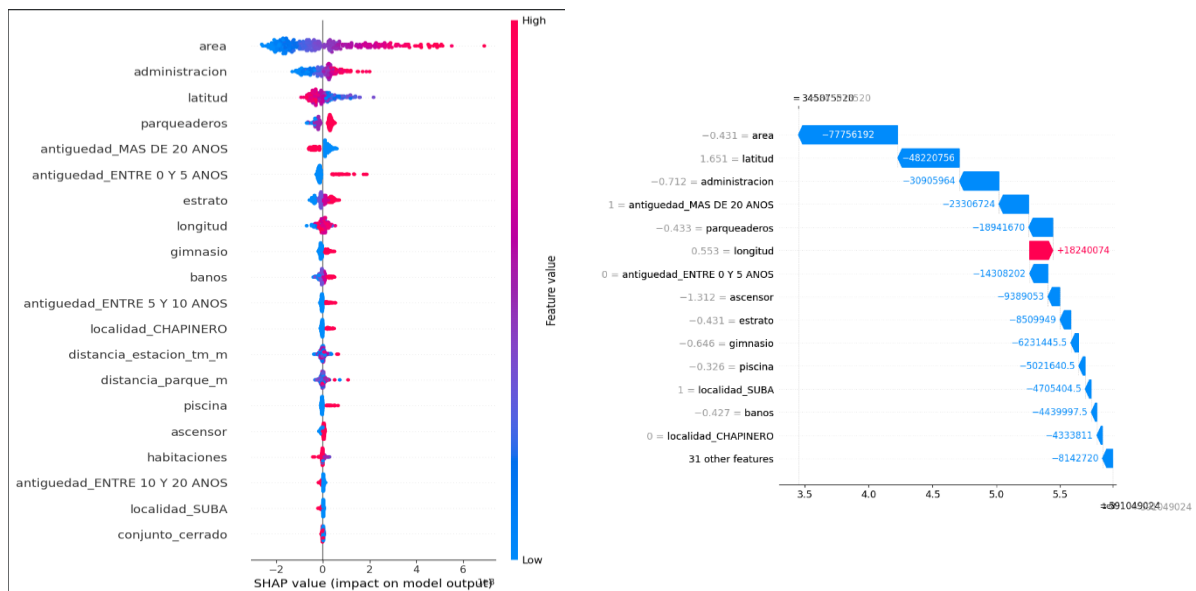


Figura 2. Representación valores SHAP y LIME predicción Xgboost

Generación de valor:

La automatización de los avalúos mediante el modelo de regresión genera un impacto financiero altamente favorable para HabitAlpes. El tiempo requerido por un perito se reduce de 6 a 1 hora, lo que implica un ahorro directo de **\$47.500 COP por valoración** y más de **\$2,7 millones mensuales** en eficiencia para un volumen de 500 avalúos. Incluso considerando el costo asociado a errores donde un 31% de los casos requiere verificación presencial por un valor de \$57.000 COP y los gastos de infraestructura en la nube, el balance mensual sigue siendo positivo. Con un costo de desarrollo cercano a **\$25 millones COP** y una operación que continúa generando ahorros sostenidos, la inversión se recupera en aproximadamente **7 meses**, logrando un ROI mensual del **15,6%**, lo que

confirma que el modelo no solo es financieramente rentable, sino estratégico para la operación desde sus primeras etapas de uso.

Insights y conclusiones:

En conclusión, la solución desarrollada no solo demuestra viabilidad técnica, sino también relevancia operativa y rentabilidad financiera. XGBoost se consolida como el modelo más adecuado para la tarea, y la herramienta resultante ofrece a HabitAlpes una base sólida para avanzar hacia una plataforma de avalúos automatizados, escalable, interpretable y alineada con las dinámicas reales del mercado inmobiliario.

Si bien el modelo presenta mayor precisión en propiedades de alto valor, se identifica oportunidad para expandir su aplicabilidad mediante técnicas de balanceo que capturen la heterogeneidad de segmentos medios.

Para mantener esta ventaja competitiva, se recomienda establecer un ciclo de evaluación mensual que permita capitalizar las dinámicas de mercado en evolución constante, mediante la detección temprana de cambios en los patrones de compraventa, particularmente en los sectores exclusivos de Bogotá donde el modelo actualmente exhibe su máximo desempeño.