
Predictive model of housing prices as a real estate commercial strategy

Páez Iván¹, Niño Melissa², Bonilla Jessica³, Manco Luis⁴

Facultad de Economía de la Universidad de los Andes

Resumen:

Este documento presenta un modelo de predicción de precios de vivienda en Bogotá (Colombia), diseñado para maximizar la cantidad de propiedades adquiridas por el mercado de compra y venta de propiedades en el barrio de Chapinero en Bogotá, invirtiendo la menor cantidad de dinero posible. Los resultados obtenidos con el modelo de *Regresión Lineal* presentaron el mejor desempeño en la predicción con un MAE de COP \$1.3, el cual logró reducir aproximadamente 1.3% el MAPE. En la sección II de este documento se presentan los datos utilizados, mientras que en la sección III se aborda el modelo de predicción y se detallan sus principales resultados. Para obtener más detalles y replicar el estudio, se proporciona acceso al repositorio de GitHub https://github.com/iapaezg/BD_LM_02

Palabras clave:

Precio de la vivienda,
Predicción,
Bogotá – Colombia,
Modelos de predicción,

I. Introducción

Considerando el impacto macroeconómico que llega a tener la fluctuación de precios de la vivienda en países en desarrollo como en Colombia, existe una necesidad inmediata de diseñar sistemas predictivos más efectivos para estos activos, que permitan a inversores, acreedores y a gobiernos mejorar sus decisiones y actividades de compra y préstamo (Lahmiri et al., 2023). Para este tipo de análisis se tienen dos tipos de abordajes, los que se correlacionan características internas con la vivienda y los que se correlacionan con la ubicación y el entorno de la misma (Chen et al., 2023), siendo estos los que en la actualidad tienen mayor acogida, por la oportunidad de aprovechar datos abiertos y modelos de machine learning y big data, para obtener modelos más robustos y precisos (Wu et al., 2022).

El estrato, condición de entrega y estado constructivo los cuales afectan el precio de manera lineal, mientras que el área, las distancias a parques, vías y estaciones de Transmilenio presentan resultados no lineales (Delgado, Melo & Azcarate, 2021). Además, otras características físicas como el número de habitaciones, número de baños y las zonas privadas y comunes las cuales juegan un papel fundamental en la conformación del precio de la vivienda (Desormeaux, D. y F. Piguillem, 2003) así como años de antigüedad de la construcción (García, Upegui, & Velásquez, 2019), o si la vivienda es nueva o usada (Mora, 2016), o tiene licencia de construcción o no (Clavijo & Muñoz, 2005), lo cual influye en la decisión de los vendedores al establecer un precio.

Diferentes estudios han buscado determinar la relación estadísticamente representativa entre estas variables de ubicación y entorno, encontrando que variables como la cercanía con espacios verdes y la

1 Biólogo. Universidad Nacional de Colombia. ia.paez10@uniandes.edu.co

2 Abogada. Universidad del Rosario. i.nino@uniandes.edu.co

3 Administradora Ambiental. Fundación Universitaria Empresarial de la Cámara de Comercio de Bogotá
jl.bonillacl@uniandes.edu.co

4 Ingeniero Ambiental. Universidad Distrital Francisco José de Caldas. l.manco@uniandes.edu.co

visibilidad de los mismos desde la propiedad afectan el precio de esta (Wu et al., 2022). Estos estudios son insumos para el estudio del bienestar y el impacto política pública, por ejemplo, de la equidad en la accesibilidad del transporte para el trabajo y el estudio (Guzman et al., 2017) o el análisis de la variación de precios por entrada de restricciones vehiculares en ciudades con altos índices de atascamiento vehicular como Bogotá (Lyu, 2022).

Para estos estudios se emplean diferentes modelos con varias formas funcionales y sistemas predictivos. Lahmiri et al (2023) compararon la exactitud de modelos de tres sistemas, regresión de vectores de soporte (aprendizaje automático), regresión de procesos gaussianos (modelos estadísticos) y boosting de árboles de regresión (inteligencia artificial), siendo estos últimos los que generaron mayor exactitud en la predicción, siendo congruente con lo expuesto por Bola et al (2022) quienes identificaron un nivel de error del 5% en modelos de inteligencia artificial similares como lo son Random Forest.

Sin embargo, toda la literatura expuesta destaca que todo modelo empleado para el análisis de precios de vivienda debe ser ajustado a las condiciones específicas de la ciudad en la que quiera ser desarrollado, pues todos presentan limitaciones de precisión y categorización de escenas, al tener una disponibilidad diferenciada de datos lo cual restringe los estudios a algunas zonas del área de estudio, imposibilitando en la mayoría de los casos las variaciones temporales en los precios de la vivienda. Adicionalmente, se carece de otro tipo de información cada vez más específica como la calidad de la infraestructura cercana. No es lo mismo tener varios colegios cercanos, pero ninguno con buena calidad (Wu et al., 2022).

En el presente estudio se emplea un abordaje similar al de Soltani et al (2022) en el cual se consideran variables internas y externas de las viviendas de Bogotá (Colombia) bajo la premisa de obtener un modelo de predicción que permita adquirir la mayor cantidad de propiedades en la localidad de Chapinero, invirtiendo la menor cantidad de dinero posible. Este modelo se basa en las variables más determinantes y utilizadas en el país para el mercado de compraventa de propiedades.

Las variables clave consideradas en el modelo, fueron obtenidas de bases de datos abiertas, usando estrategias de extracción de datos mediante expresiones regulares. Dentro de las variables consideradas se están el área de la vivienda en metros cuadrados, características de la vivienda: número de baños, habitaciones; si tiene o no terrazas, patios, balcones, ascensor y garajes, así como las distancias a parques, hospitales, universidades y estaciones de transporte. Se evaluaron cuatro modelos predictivos, incluyendo una *regresión lineal simple*, *Elasticnet*, *Random Forest* y *árboles*.

Los resultados muestran que el modelo *Regresión Lineal* presenta el mejor desempeño en la predicción de precios de vivienda en la localidad de Chapinero, con un Error absoluto medio (MAE) de COP \$1.3, el cual logró reducir aproximadamente a (MAPE) 1.3% el error de predicción del modelo. Los resultados obtenidos con este modelo de predicción de precios de vivienda son relevantes debido a su simplicidad, actualización continua y precisión en la estimación del costo de la vivienda.

II. Datos

La base de datos utilizada contiene información sobre los precios de venta de 48.930 propiedades ubicadas en la ciudad de Bogotá durante los años 2019, 2020 y 2021. Estos datos fueron proporcionados por Properati (plataforma web de anuncios clasificados para venta y alquiler de propiedades) y se utilizaron técnicas de normalización de texto para extraer características físicas relevantes que influyen en el precio de los inmuebles. Para ello, se identificaron patrones básicos y se utilizaron expresiones regulares para buscar términos como "garaje", "parqueadero", "ascensor", "terrazza", "balcón", "habitación(es)", "baño(s)", "área" y "metro(s)" en la descripción de la base de datos y con esto extraer las características de cada inmueble, como lo puede ser cuantas habitaciones posee. En la Tabla 1 se describen las variables de interés.

En el caso de las variables que presentaban valores faltantes, se utilizó información extraída del texto para imputar valores en variables como el número de baños, número de habitaciones y área total. Sin embargo, después de la imputación, las variables: área total, ascensor y número de habitaciones aún presentaban datos faltantes. En la variable *área total* se reemplazaron los valores faltantes por el promedio de las observaciones, mientras que en las otras dos variables se emplea la mediana de los datos como estrategia de imputación. Estas observaciones faltantes representaban menos del 1% del total de observaciones, después de lo cual se obtuvieron 48.924 observaciones finales.

Tabla 1. Variables de interés

| Variable | Descripción |
|-------------------------------|--|
| <i>Tipo propiedad</i> | Variable categórica, “Casa”, “Apartamento”. |
| <i>Área</i> | Variable continua medida en metros cuadrados |
| <i>Baños</i> | Variable discontinua, de conteo |
| <i>Habitaciones</i> | Variable discontinua, de conteo |
| <i>Deposito / bodega</i> | Variable dicotómica, 1 = cuenta con deposito 0 = lo contrario |
| <i>Ascensor</i> | Variable dicotómica, 1 = cuenta con ascensor 0 = lo contrario |
| <i>Exterior</i> | Variable dicotómica, 1 = cuenta con terraza, balcón y/o patio 0 = lo contrario |
| <i>Parqueadero</i> | Variable dicotómica, 1 = cuenta garaje o parqueadero 0 = lo contrario |
| <i>Parques</i> | Variable continua, representa la distancia mínima (cercanía) a parques medida en metros |
| <i>Hospitales</i> | Variable continua, representa la distancia mínima (cercanía) a hospitales medida en metros |
| <i>Estación de transporte</i> | Variable continua, representa la distancia mínima (cercanía) a estaciones de transporte medida en metros |
| <i>Universidades</i> | Variable continua, representa la distancia mínima (cercanía) a universidades medida en metros |

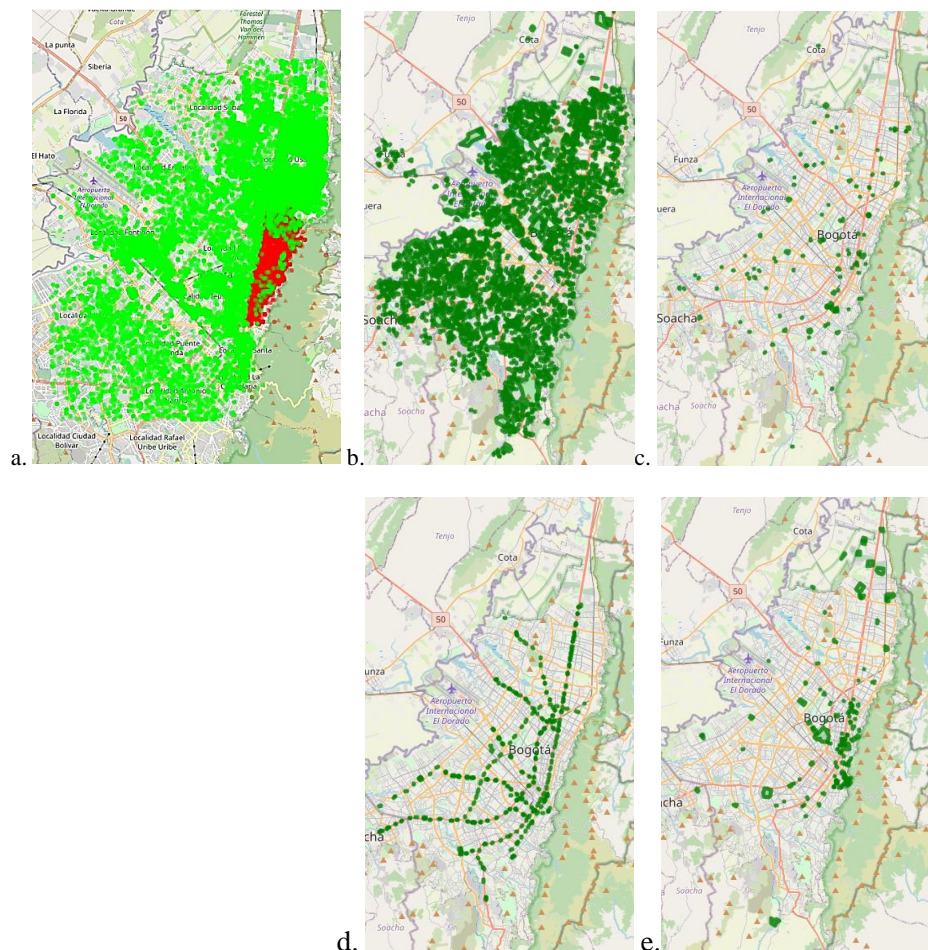
En la **¡Error! No se encuentra el origen de la referencia.** (sección a) se muestran las áreas utilizadas en el análisis, divididos en dos conjuntos de datos, una muestra de entrenamiento (área verde) y una muestra de prueba (área roja). La muestra de entrenamiento incluye todas las localidades de Bogotá, excepto Chapinero, y representa el 79% de los datos totales. Por otro lado, la muestra de prueba corresponde únicamente a la localidad de Chapinero y abarca el 21% restante de los datos. La muestra de entrenamiento está representada en verde y consta de 38.638 observaciones, mientras que la muestra de prueba se muestra en rojo y está compuesta por 10.286 observaciones.

Los datos de entrenamiento fueron utilizados para ajustar los modelos de predicción y los datos de prueba se emplearon para evaluar su desempeño. Esto permitirá medir de manera objetiva la capacidad de los modelos para predecir correctamente los resultados. El tipo de propiedades fue considerado como parte de los modelos, para la muestra de entrenamiento se contaban con un total de 29.173 apartamentos (59,6%) y 9.465 casas (19,7%), mientras que para el caso de la muestra de prueba se distribuyeron en 10.012 apartamentos (20,4%) y 274 casas (0,56%).

a. Análisis descriptivo de los datos

Se pudo observar la distribución geográfica de algunas de las variables en la Figura 1. (a) La localidad de chapinero la cual es el área de prueba se encuentra en el área hola oriental de la ciudad alejada de lugares de alta relevancia hoy en la ciudad como lo pueden ser los principales ingresos vehiculares coma el aeropuerto, teniendo como mayor elemento diferenciador la cercanía a los cerros orientales. (b) como única variable distribuida de manera similar a lo largo de Bogotá coma se encuentran los parques, está la ubicación de parques, hola para los cuales no se consideró su área. (c) en cuanto a universidades se identifica que la localidad de chapinero a diferencia del resto de la ciudad cuenta con mayor concentración de estos establecimientos lo cual podría generar una desviación de los modelos al ser probados al igual que (d) el acceso más estaciones de transporte público y (e) hospitales.

Figura 1. Distribución de a. Viviendas de entrenamiento y prueba, b. Parques, c. Universidades, d. Estaciones de transporte y hospitales.



Variables numéricas / continuas

Inicialmente, la información detallada sobre las estadísticas descriptivas de las variables de interés se puede verificar en el anexo 1.

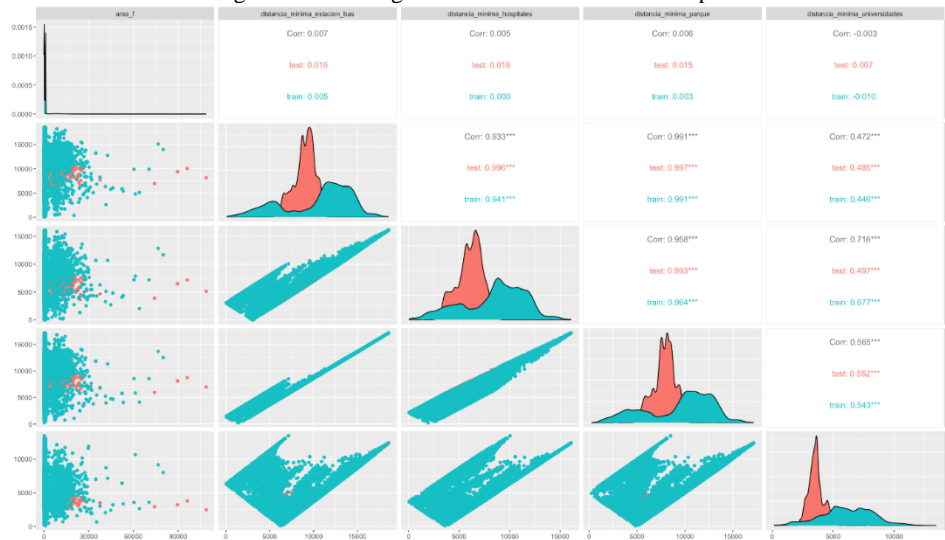
En este anexo se puede observar que en promedio la muestra de entrenamiento tiene una media para el área de $896 \pm 2990 \text{ m}^2$ mientras que la muestra de prueba tiene una media de $951 \pm 2601 \text{ m}^2$, y se verifica de acuerdo con la Figura 2. En las diagonales de esta figura, se puede observar la distribución de los datos, en la “escalera” inferior se detalla la relación de las variables entre sí, mientras que en la “escalera” superior se encuentran los valores del índice de correlación para la muestra total y para cada submuestra (color azul representa los datos de entrenamiento y el naranja los de prueba).

La variable área no presenta normalidad en ninguna de las dos muestras, por lo cual no cumplen este supuesto. Así mismo, se puede verificar que no hay correlación con las diferentes variables de distancia al lugar más cercano (hospital, universidad, parque, Transmilenio).

Por otra parte, las variables de distancia mínima entre sí muestran una correlación significativa para ambas muestras y para la muestra total, lo cual puede representar colinealidad entre las variables cuando se describan los modelos empleados. Ninguna de estas variables presenta una distribución normal, por

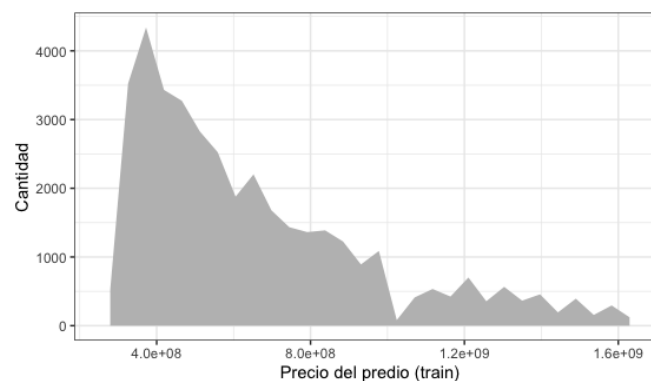
lo cual no cumplen el supuesto desde un análisis gráfico, por ejemplo, la muestra de entrenamiento presenta una distribución bimodal.

Figura 2. Análisis gráfico de las estadísticas descriptivas



Los valores de la variable respuesta precio del predio presentan una media de 654 millones de pesos con una desviación estándar de 311 millones de pesos aproximadamente. Como se puede determinar la Figura 3, la distribución de la variable no es normal con el 50% de los datos por debajo de los 560 millones de pesos, para el área de entrenamiento del modelo (En rojo en la Figura 1a).

Figura 3. Distribución del precio de las viviendas

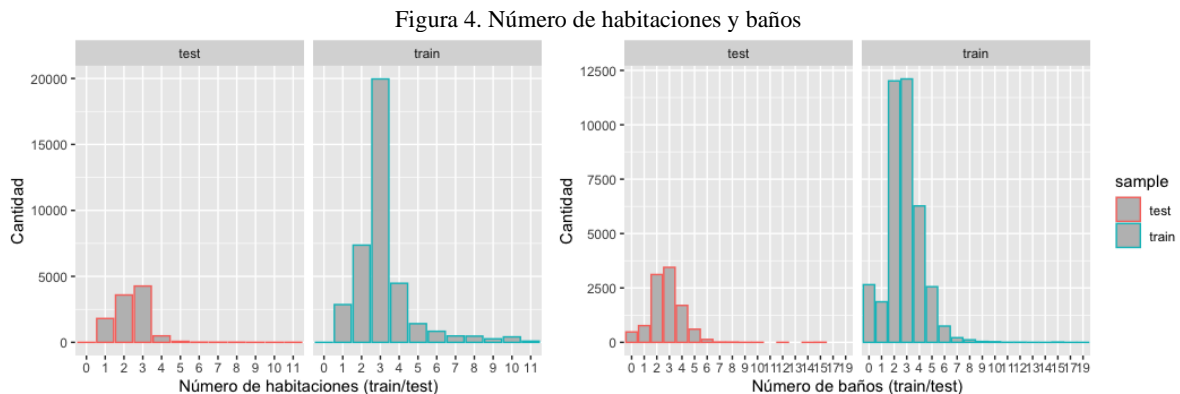


Variables de conteo /discontinuas

Las estadísticas descriptivas de las variables de presencia o ausencia de ascensores, depósitos, parqueaderos y zonas exteriores pueden observarse en el anexo 1. Se puede determinar que en la muestra de entrenamiento el 18% de los predios cuenta con ascensores, el 40% con depósito, cerca del 70% con parqueaderos y 50% con zonas exteriores como balcones, terrazas o patios. Por su parte, las variables en la muestra de prueba presentaron los porcentajes de 21%, 41%, 69% y 47%, respectivamente.

Para el caso de las variables que se generaron a partir de la extracción de datos de las descripciones como el número de baños y habitaciones (ver Figura 4), se puede observar que la moda para el entrenamiento y la prueba es de tres habitaciones mientras que para el número de baños los datos más frecuentes son dos y tres. La mediana para número de baños fue tomada como valor para los datos faltantes en ambas muestras, arrojando un valor de tres baños como mediana. Para el caso de las

habitaciones, la muestra de entrenamiento tuvo como mediana tres habitaciones y dos para la muestra de prueba.



III. Modelo y resultados

Para el ejercicio de predicción de precios de las viviendas en Chapinero se realizaron cuatro modelos. Los enfoques de pronóstico usados fueron *regresión lineal*, *ElasticNet*, *Árboles* y *Random Forest*.

- **Modelo 1:** Se utilizó un modelo de regresión lineal.
- **Modelo 2:** Se aplicó la técnica de ElasticNet utilizando el paquete "caret". Este paquete facilitó el proceso de construcción del modelo al permitir el entrenamiento, evaluación, selección de características, preprocesamiento de datos y ajuste de hiperparámetros.
- **Modelo 3:** Se empleó el enfoque de Random Forest utilizando el paquete "ranger". Se realizaron selecciones de hiperparámetros basadas en validación cruzada (cross-validation), evaluando distintas combinaciones de valores para encontrar la mejor combinación.
- **Modelo 4:** Se utilizó un modelo de árboles con un resampleo en cinco particiones y una profundidad máxima del árbol de 50 nodos terminales.

El modelo seleccionado para la predicción de los precios de la vivienda fue el *modelo regresión lineal simple*, cuyo MAE es COP 1.30, es decir el modelo se equivoca en un peso y 30 centavos; con un error de predicción del 1.3% (ver Tabla 2). Las variables que se emplearon para entrenar el modelo siguen lo descrito por la bibliografía expuesta en la introducción, siendo estas como presencia o no de ascensores, número de baños o habitaciones, tipo de propiedad (casa o apartamento) y la distancia a la estación de transporte más cercana, al igual que hospital y parques.

$$\log(\text{precio}) = f(\text{ascensor}, \text{exterior}, \text{número de baños}, \text{número de habitaciones}, \text{tipo de propiedad}, \text{distancia mínima a parques}, \text{distancia mínima a estación de transporte}, \text{distancia mínima a hospitales}, \text{distancia mínima a universidades})$$

Forma funcional de todos los modelos

En el anexo 2, se presenta la regresión lineal de la ecuación anterior. Los coeficientes de la regresión con una significancia del 1% en la variable independiente, es decir, el logaritmo del precio de la vivienda está relacionado con características como la presencia de ascensor, terraza o balcón, la cantidad de baños (1, 2 o más de 4), el tipo de propiedad y las distancias mínimas a la infraestructura de importancia en relación con cada inmueble. Por ejemplo, el coeficiente asociado al tipo de propiedad tiene una relación positiva con el precio de los inmuebles. En promedio, si un inmueble es una casa en comparación con un apartamento, se observa un incremento del 6.2% en el precio. Esto indica que las casas tienen a tener un valor más alto en relación con los apartamentos en este contexto específico.

Se utilizó el paquete "doParallel" del software R, para optimizar los cálculos en memoria e implementar la ejecución paralela utilizando varios núcleos del procesador.

Adicionalmente, se desarrolló un modelo de *ElasticNet* empleando la misma forma funcional usando como hiperparámetro un α de 0.05263158 y λ de 0.001, obteniendo un MAE de \$ 1.30 COP y un MAPE del 1.31%. También se consideraron modelos de *Random Forest* para los cuales se generaron diferentes corridas considerando una validación cruzada de cinco secciones con diferentes profundidades de nodos terminales. Se probaron 10, 30, 50, 70 y 100 nodos. Encontrando que el de mayor ajuste fue el de 50, con un MAE de \$1.22 COP y un MAPE: del 0.97%. Por último, se desarrolló un modelo de árboles considerando un mínimo de 10 nodos, considerando metrys de 2, 4 y 6 encontrando un MAE promedio de \$ 1.28 COP y un MAPE de 1.23% (Toda la comparación de esto modelos se puede ver en la Tabla 2).

Tabla 2. Comparación de ajuste de modelos

| | Modelo 1 | Modelo 2 | Modelo 3 | Modelo 4 |
|------------------|----------|--|---|---|
| Metodología | OLS | ElasticNet | Random forest | Árbol – 50 nodos terminales |
| MAE_test (pesos) | 1.304 | 1.305 | 1.219 | 1.285 |
| MAPE_test (%) | 1.31 | 1,31 | 0,97 | 1,23 |
| Hiperparámetro | NA | α : 0.05263158 λ : 0.001 | máxima profundidad del árbol: 10, 30, 50, 70, 100 | Metry se tomó para el análisis como 2, 4, 6 |
| Nombre en Script | Modelo 5 | Modelo final EN | Random forest | Árbol (50) |

IV. Conclusiones

Luego de correr un total de 10 modelos diferentes, se determinó que las regresiones más simples retornaron valores más cercanos a la realidad al comprobar las predicciones en Kaggle. Aunque modelos más complejos tenían valores de precisión más bajos, las predicciones resultaron en precisión más pobre, lo cual puede tener relación con un sobreajuste del modelo, posible multicolinealidad de las variables de distancia mínima a una infraestructura de interés e inclusive a dependencia espacial. Adicionalmente, dado que la muestra de entrenamiento representaba a viviendas de toda la ciudad de Bogotá, la heterogeneidad de la ciudad puede contribuir a la baja precisión en la predicción.

Es recomendable analizar la teoría sobre la determinación de los precios de las viviendas, dado que la forma funcional predicha puede ser equivocada como resultado de un sobreajuste al modelo. Así mismo, el modelo se puede ver favorecido de otras variables como estrato, inseguridad, número del piso en el que se encuentra el apartamento, entre otras.

Considerando los niveles de error hallados con los modelos planteados, la recomendación del estudio para quienes deseen adquirir propiedades en Chapinero sugiere que el modelo de Regresión lineal es el más preciso para predecir los precios de la vivienda en esta localidad, por lo tanto el modelo predictivo desarrollado puede ayudar actores del mercado inmobiliario a determinar qué propiedades tienen un mayor potencial de ganancias y, por lo tanto, en cuáles debería invertir a través del establecimiento de precios competitivos. Sin embargo, considerando lo reportado en la bibliografía, respecto a las limitaciones de estos modelos, se debe realizar una constante actualización de los datos de entrenamiento para asegurar su exactitud la cual se debe validar además con una lectura del entorno macroeconómico.

V. Anexos

Anexo 1: estadísticas descriptivas

Variables de conteo /discontinuas: Numero de baños y habitaciones

| skim_variable | sample | n_missing | complete_rate | factor.n_unique | factor.top_counts |
|---------------|--------|-----------|---------------|-----------------|--------------------------------------|
| bano_f | test | 0 | 1 | 14 | 3: 3447, 2: 3117, 4: 1697, 1: 767 |
| bano_f | train | 0 | 1 | 18 | 3: 12107, 2: 12016, 4: 6270, 0: 2650 |
| bed_f | test | 0 | 1 | 12 | 3: 4266, 2: 3589, 1: 1807, 4: 490 |
| bed_f | train | 0 | 1 | 12 | 3: 19963, 2: 7368, 4: 4475, 1: 2861 |

Variables numéricas / continuas: precio

| skim_variable | sample | n_missing | complete_rate | numeric.mean | numeric.sd | numeric.p0 | numeric.p25 | numeric.p50 | numeric.p75 | numeric.p100 |
|---------------|--------|-----------|---------------|--------------|-------------|------------|-------------|-------------|-------------|---------------|
| price | test | 10,286 | - | NA | NA | NA | NA | NA | NA | NA |
| price | train | - | 1.00 | 654,510,792 | 311,390,877 | 0 | 415,000,000 | 559,990,000 | 810,000,000 | 1,650,000,000 |

Variables numéricas / continuas: distancia mínima a parques, universidad, hospitales, estaciones de transporte

| skim_variable | sample | n_missing | complete_rate | numeric.mean | numeric.sd | numeric.p0 | numeric.p25 | numeric.p50 | numeric.p75 | numeric.p100 |
|--------------------------------|--------|-----------|---------------|--------------|------------|------------|-------------|-------------|-------------|--------------|
| distancia_minima_estacion_bus | test | 0 | 1 | 8,969 | 1,181 | 3,170 | 8,318 | 9,141 | 9,823 | 14,400 |
| distancia_minima_estacion_bus | train | 0 | 1 | 10,547 | 3,875 | 70 | 7,689 | 11,647 | 13,458 | 18,574 |
| distancia_minima_hospitales | test | 0 | 1 | 6,006 | 1,207 | 1,392 | 5,272 | 6,180 | 6,901 | 11,960 |
| distancia_minima_hospitales | train | 0 | 1 | 8,482 | 3,103 | 27 | 6,393 | 8,999 | 10,795 | 16,101 |
| distancia_minima_parque | test | 0 | 1 | 7,809 | 1,059 | 2,847 | 7,231 | 7,930 | 8,545 | 12,992 |
| distancia_minima_parque | train | 0 | 1 | 9,435 | 3,484 | 246 | 6,390 | 10,308 | 12,104 | 17,167 |
| distancia_minima_universidades | test | 0 | 1 | 3,547 | 588 | 1,345 | 3,181 | 3,522 | 3,842 | 8,654 |
| distancia_minima_universidades | train | 0 | 1 | 6,149 | 2,135 | 132 | 4,833 | 6,162 | 7,687 | 13,467 |

Variables de conteo /discontinuas

Numero de baños y número de habitaciones

| skim_variable | sample | n_missing | complete_rate | numeric_mean | numeric.sd | numeric.p0 | numeric.p25 | numeric.p50 | numeric.p75 | numeric.p100 |
|---------------|--------|-----------|---------------|--------------|------------|------------|-------------|-------------|-------------|--------------|
| area_f | test | 0 | 1 | 896.3 | 2,990.5 | 10 | 114 | 229 | 896.3 | 108,800 |
| area_f | train | 0 | 1 | 951.6 | 2,601.7 | 2 | 110 | 448 | 951.5 | 80,000 |
| asc_f | test | 0 | 1 | 0.2 | 0.4 | 0 | 0 | 0 | 0 | 1 |
| asc_f | train | 0 | 1 | 0.2 | 0.4 | 0 | 0 | 0 | 0 | 1 |
| dep_f | test | 0 | 1 | 0.4 | 0.5 | 0 | 0 | 0 | 1 | 1 |
| dep_f | train | 0 | 1 | 0.4 | 0.5 | 0 | 0 | 0 | 1 | 1 |
| par_f | test | 0 | 1 | 0.7 | 0.5 | 0 | 0 | 1 | 1 | 1 |
| par_f | train | 0 | 1 | 0.7 | 0.5 | 0 | 0 | 1 | 1 | 1 |
| ext_f | test | 0 | 1 | 0.5 | 0.5 | 0 | 0 | 0 | 1 | 1 |
| ext_f | train | 0 | 1 | 0.5 | 0.5 | 0 | 0 | 1 | 1 | 1 |

Anexo 2: Resultados regresión mejor modelo – Regresión lineal

| Dependent variable: | |
|---------------------|----------------------|
| log(price) | |
| asc_f | -0.083*** (0.005) |
| ext_f | 0.084*** (0.004) |
| factor(bano_f) 1 | -0.094*** (0.010) |
| factor(bano_f) 2 | -0.186*** (0.007) |
| factor(bano_f) 3 | 0.002 (0.007) |
| factor(bano_f) 4 | 0.192*** (0.008) |
| factor(bano_f) 5 | 0.319*** (0.010) |
| factor(bano_f) 6 | 0.290*** (0.014) |
| factor(bano_f) 7 | 0.337*** (0.024) |

| | |
|-------------------|---------------------|
| factor(bano_f) 8 | 0.346*** (0.033) |
| factor(bano_f) 9 | 0.459*** (0.054) |
| factor(bano_f) 10 | 0.577*** (0.064) |
| factor(bano_f) 11 | 0.679*** (0.139) |
| factor(bano_f) 12 | 0.106 (0.128) |
| factor(bano_f) 13 | 0.399* (0.239) |
| factor(bano_f) 14 | -0.335 (0.338) |
| factor(bano_f) 15 | -0.018 (0.090) |
| factor(bano_f) 17 | 0.343 (0.337) |
| factor(bano_f) 19 | 1.119*** (0.338) |
| factor(bed_f) 1 | -0.278 (0.195) |
| factor(bed_f) 2 | -0.073 (0.195) |
| factor(bed_f) 3 | 0.091 (0.195) |
| factor(bed_f) 4 | 0.190 (0.195) |
| factor(bed_f) 5 | 0.138 (0.195) |
| factor(bed_f) 6 | 0.118 (0.195) |
| factor(bed_f) 7 | 0.010 (0.195) |
| factor(bed_f) 8 | -0.004 (0.196) |
| factor(bed_f) 9 | -0.045 (0.196) |

| | |
|--------------------------------|-----------------------------|
| factor (bed_f) 10 | 0.020 (0.196) |
| factor (bed_f) 11 | 0.026 (0.198) |
| property_typeCasa | 0.064*** (0.006) |
| distancia_minima_estacion_bus | -0.0001*** (0.00001) |
| distancia_minima_hospitales | -0.0001*** (0.00000) |
| distancia_minima_parque | 0.0002*** (0.00001) |
| distancia_minima_universidades | -0.0001*** (0.00000) |
| Constant | 20.301*** (0.195) |
| ----- | |
| Observations | 38,638 |
| R2 | 0.404 |
| Adjusted R2 | 0.403 |
| Residual Std. Error | 0.337 (df = 38602) |
| F Statistic | 746.360*** (df = 35; 38602) |
| ===== | |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

VI. Bibliografía

- Delgado, Melo-Marínez, O. O., & Azcarate-Romero, J. S. (2021). Determinantes del precio de la vivienda nueva en Bogotá para el año 2019: una aproximación a través de un modelo semiparamétrico de regresión espacial/Determinants of New Housing Prices in Bogotá for 2019: an Approach Through a Semiparametric Spatial Regression Model. *Ingeniería y ciencia* (Medellín, Colombia), 17(34), 23–. <https://doi.org/10.17230/ingciencia.17.34.2>
- Desormeaux, D. y F. Piguillem (2003). Precios Hedónicos e Índices de Precios de Viviendas. Documento de Trabajo N° 12, Cámara Chilena de la Construcción.
- Clavijo, Janna, M., & Muñoz, S. (2005). La vivienda en Colombia: sus determinantes socioeconómicos y financieros.
- García-Rendón, Díaz, S., Upegui, J. C., & Velásquez, H. (2019). Determinantes del precio de la vivienda nueva en Medellín: Un modelo estructural. *Cuadernos de economía* (Bogotá, Colombia), 38(76), 109–136. <https://doi.org/10.15446/cuad.econ.v37n76.57032>
- Mora León. (2016). Determinantes del precio de vivienda nueva en Colombia del 2004 al 2015 : una comparación de Santiago de Cali con otras capitales de departamento. Universidad de los Andes.

- Bola, A., Noah, O., & Alaba, F. (2022). ScienceDirect House Price Prediction using Random Forest Machine Learning Technique. *Procedia Computer Science*, 199, 806–813.
<https://doi.org/10.1016/j.procs.2022.01.100>
- Chen, Y., Jiao, J., & Farahi, A. (2023). Disparities in affecting factors of housing price : A machine learning approach to the effects of housing status , public transit , and density factors on single-family housing price. *Cities*, 140(November 2022), 104432.
<https://doi.org/10.1016/j.cities.2023.104432>
- Guzman, L. A., Oviedo, D., & Rivera, C. (2017). Assessing equity in transport accessibility to work and study : The Bogotá region. *Journal of Transport Geography*, 58, 236–246.
<https://doi.org/10.1016/j.jtrangeo.2016.12.016>
- Lahmire, S., Bekiros, S., & Avdoulas, C. (2023). A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decision Analytics Journal*, 6(November 2022), 100166. <https://doi.org/10.1016/j.dajour.2023.100166>
- Lyu, X. (2022). Car restriction policies and housing markets. *Journal of Development Economics*, 156(March), 102850. <https://doi.org/10.1016/j.jdevco.2022.102850>
- Soltani, A., Heydari, M., Aghaei, F., & James, C. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities*, 131(August), 103941. <https://doi.org/10.1016/j.cities.2022.103941>
- Wu, C., Du, Y., Li, S., Liu, P., & Ye, X. (2022). Land Use Policy Does visual contact with green space impact housing prices ? An integrated approach of machine learning and hedonic modeling based on the perception of green space. *Land Use Policy*, 115(January 2021), 106048. <https://doi.org/10.1016/j.landusepol.2022.106048>