

Problem Set #3

Sergio Díaz, Gabriel Perdomo, David Salamanca

1 de diciembre de 2024

Ver repositorio de Github

Introducción

El mercado inmobiliario, particularmente nos enfocamos en el barrio de Chapinero, se caracteriza por una asimetría en el acceso a la información sobre las propiedades. Esta desigualdad en el conocimiento puede distorsionar los precios y alejarlos de su valor objetivo, en especial cuando se busca realizar compras a gran escala y no se puede dedicar suficiente tiempo a valorar tradicionalmente cada inmueble. Este proyecto tiene como objetivo desarrollar un modelo predictivo que permita estimar los precios de las propiedades en Chapinero con mayor precisión, optimizando así las decisiones de compra.

El análisis parte de una lección clave aprendida del caso de Zillow, donde la falta de modelos adecuados para estimar precios inmobiliarios llevó a errores significativos en la valoración de propiedades. En ese caso, la sobreestimación de precios resultó en pérdidas financieras de aproximadamente USD 500 millones, destacando la importancia de contar con herramientas precisas para este tipo de predicciones.

Los datos utilizados en este proyecto provienen de Properati, una plataforma que recopila información estructural, de localización y descriptiva de las propiedades. Adicionalmente, se incorporan variables derivadas para capturar interacciones relevantes, como la relación entre la proximidad a puntos de interés (parques, transporte público, centros financieros) y características específicas de los inmuebles. Estas variables permiten abordar el problema de manera detallada y modelar mejor los factores que influyen en los precios.

Para alcanzar este objetivo, se implementan y comparan diversos modelos

predictivos, incluidos elastic net, random forest, regresiones lineales, modelos de boosting y redes neuronales. Cada técnica se evalúa en términos de su capacidad para capturar patrones significativos, predecir con precisión y explicar cómo las interacciones entre variables afectan los precios de las propiedades. Este enfoque busca encontrar modelos predictivos robustos que permitan tomar decisiones financieras mas informadas en proyectos de inversion inmobiliaria.

falta poner cual fue el mejor modelo

Datos

Descripción de los datos

El conjunto de datos contiene 38,644 registros relacionados con propiedades en Bogotá, Colombia. Cada registro incluye información diversa como características físicas de las propiedades (número de habitaciones, baños, área), atributos geográficos (latitud, longitud, cercanía a servicios) y texto descriptivo (título y descripción).

- Variables clave: **price** (precio), **rooms** (número de habitaciones), **bathrooms** (número de baños), **surface_total** (área total).
- Variables geográficas: **lat** (latitud), **lon** (longitud), y distancias a puntos de interés como hospitales, parques y estaciones de transporte.
- Texto descriptivo: **title** y **description**.

El análisis muestra que hay valores faltantes significativos en variables críticas como **surface_total** (79.7 %) y **rooms** (47.3 %), lo que requerirá estrategias de imputación o eliminación en análisis posteriores. Además, la variable **barrio** tiene un 95 % de valores faltantes, lo que la hace poco útil en el análisis actual.

Estadísticas descriptivas

A continuación, se presentan las estadísticas descriptivas para algunas variables:

Variable	Media (M)	Desv. Est. (M)	Mín. (M)	25 % (M)	50 % (M)	Máx. (M)
price	654.53	311.42	300	415	560	1650
surface_total	0.15	0.27	0.02	0.08	0.12	17.14
rooms	3	1.37	1	2	3	11
distance_to_financial_center	6.21	2.22	0.46	4.63	5.90	14.81

Cuadro 1: Estadísticas descriptivas de las variables principales. Los valores están expresados en millones (M) para precios y distancias, y en miles (K) para áreas.

La tabla de estadísticas descriptivas muestra un resumen de las variables clave del conjunto de datos. El precio de las propiedades (**price**) tiene un promedio de 654.53 millones de COP, con un rango que varía desde 300 millones hasta 1,650 millones de COP, lo cual refleja una alta variabilidad en el mercado inmobiliario de Bogotá.

A pesar de la alta volatilidad de los precios de las viviendas, la siguiente gráfica de la distribución de precios da una perspectiva que aclara un poco más el tipo de propiedades. En particular, se evidencia que la gran masas de apartamento se concentra en el rango de los 300 a los 800 millones. Este hecho es relevante puesto que puede indicar que la volatilidad se presenta por las propiedad más caras, en la cola derecha, que distorsionan las estadísticas a primera vista. Del mismo modo, es particularmente problemático cuando no hay información en variables relevantes debido a que solo con las demás características, dos casas que no tengan suficiente información, que están en lados opuestos de la distribución, van a parecer "iguales" pero con diferencias del doble del precio si no se controla ese factor.

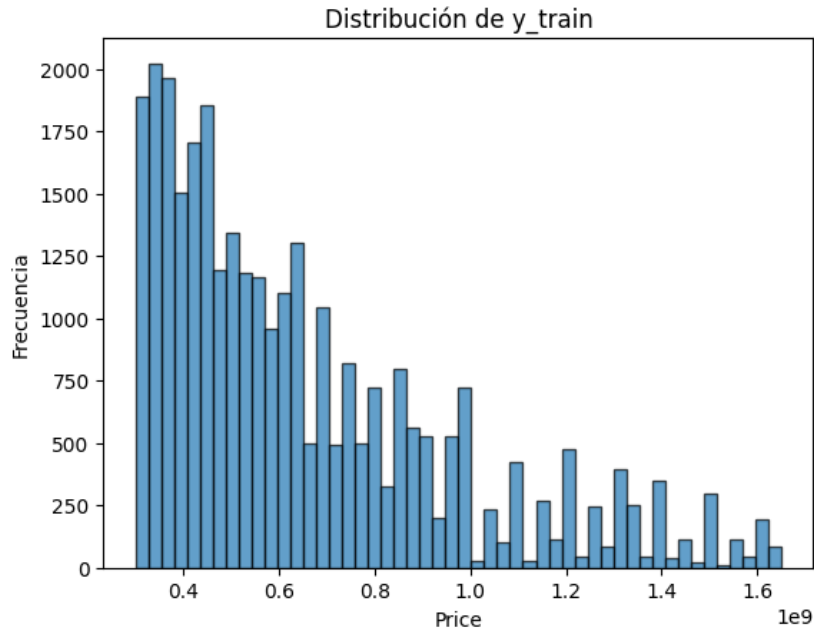


Figura 1: Distribución de las propiedades en Bogotá.

La superficie total (**surface_total**), que representa el tamaño de las propiedades, tiene un promedio de 153.95 m^2 , aunque algunas propiedades alcanzan hasta $17,137 \text{ m}^2$, lo que indica la inclusión de propiedades excepcionales (mismo indicio de una distribución desigual que jalona la volatilidad). Por otro lado, las propiedades tienen en promedio 3 habitaciones (**rooms**) y 2.88 baños (**bathrooms**), con un rango que varía desde 1 hasta 11 habitaciones y hasta 13 baños, destacando una diversidad significativa en los tamaños y configuraciones. Finalmente, la distancia al centro financiero (**distance.to.financial.center**) tiene un promedio de 6.21 kilómetros, lo que sugiere que muchas propiedades están ubicadas dentro de zonas relativamente accesibles para actividades económicas clave. Estas estadísticas resaltan la heterogeneidad de las propiedades y el impacto que variables como el tamaño, la ubicación y las características internas tienen sobre su precio.

Análisis geográfico

En los mapas a continuación, se visualiza la distribución de las propiedades en Bogotá y su proximidad al centro financiero:

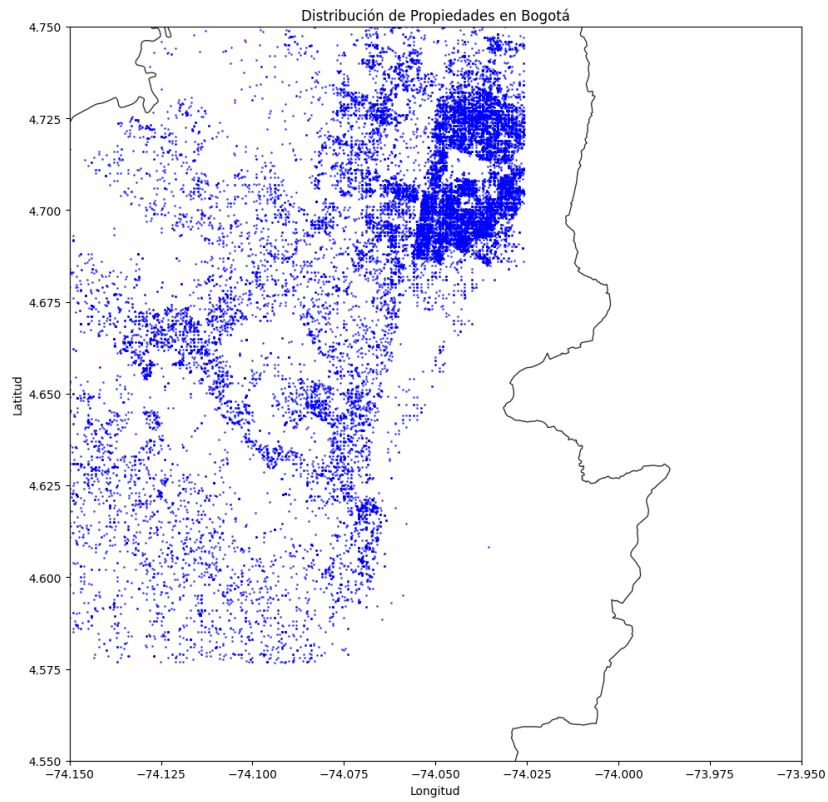


Figura 2: Distribución de las propiedades en Bogotá.

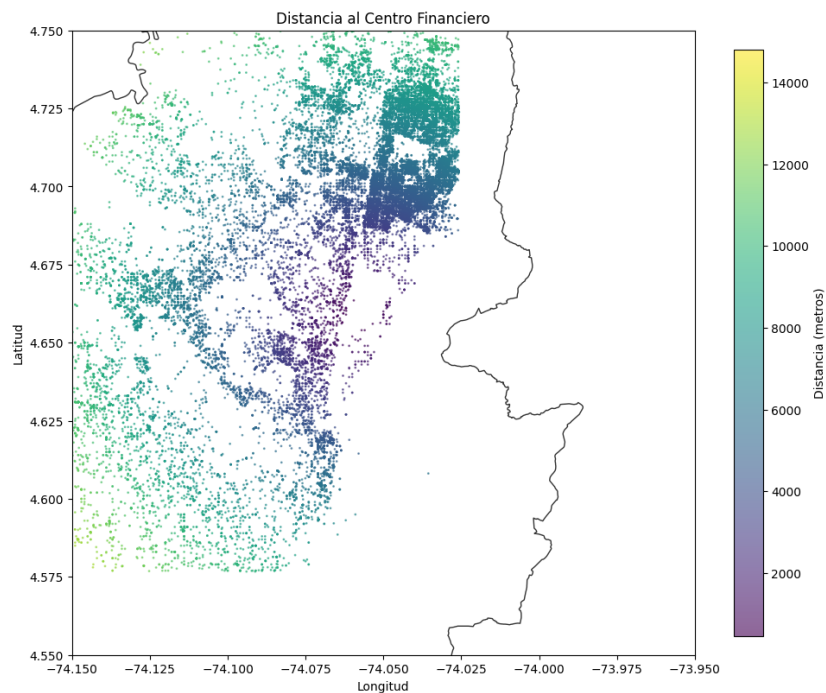


Figura 3: Proximidad de las propiedades al centro financiero.

Interpretación:

Los mapas geográficos una visión detallada de la distribución de las propiedades en Bogotá y su relación con la proximidad a puntos de interés clave. En el primer mapa, se observa una alta densidad de propiedades en áreas centrales de la ciudad, particularmente en zonas como Chapinero, que es reconocida actividad económica y oferta residencial variada.

El segundo mapa resalta la relación entre el precio de las propiedades y su proximidad al centro financiero. Las propiedades más cercanas al centro financiero, que se encuentran dentro de un rango de aproximadamente 5 kilómetros, tienden a ser más valiosas debido a su accesibilidad a lugares de trabajo y servicios financieros. En contraste, las propiedades más alejadas muestran una disminución en su valor percibido, reflejando la importancia económica de la ubicación en mercados urbanos. Este análisis visual refuerza la noción de que tanto la densidad de propiedades como su cercanía a puntos estratégicos tienen un impacto significativo en los precios del mercado.

Correlaciones entre variables

La imagen a continuación muestra las correlaciones entre el precio (**price**) y las demás variables:

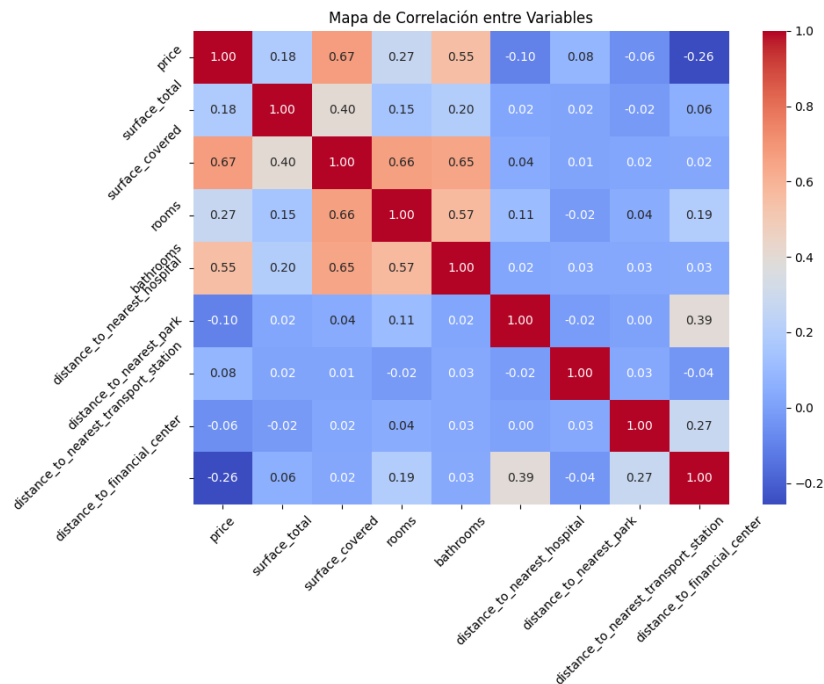


Figura 4: Tabla de Correlaciones.

Interpretación:

- Las variables con mayor correlación positiva con el precio son **surface_covered** (0.67) y **bathrooms** (0.55), indicando que propiedades más grandes y con más baños tienden a ser más valiosas.
- La distancia al centro financiero tiene una correlación negativa notable (-0.26), mostrando que la cercanía a zonas económicas clave es un determinante importante del precio.
- Variables como la distancia a parques tienen una correlación positiva baja (0.08), sugiriendo una menor influencia en los precios.

Relación entre variables y precios

El precio de una propiedad está influenciado por una combinación de factores estructurales, de tamaño y de ubicación, que reflejan las preferencias

y necesidades de los compradores. Las variables analizadas tienen diferentes grados de impacto en el precio, según su relación con la funcionalidad, la calidad de vida y el acceso a servicios clave.

La proximidad a zonas económicas como el centro financiero tiene una correlación negativa significativa con el precio ($-0,26$), indicando que las propiedades más cercanas a estas áreas suelen ser más valiosas debido a la accesibilidad para actividades laborales y comerciales. Estas propiedades tienden a atraer a compradores con mayor capacidad adquisitiva, especialmente trabajadores y ejecutivos que buscan minimizar el tiempo de desplazamiento.

La cercanía a parques muestra una correlación positiva baja ($0,08$), pero sigue siendo relevante. Los parques incrementan el valor percibido de las propiedades al ofrecer espacios recreativos, áreas verdes y una mejor calidad de vida, especialmente para familias o personas interesadas en actividades al aire libre. Aunque su impacto es menor que otras variables, sigue siendo un factor atractivo en decisiones de compra.

La distancia a estaciones de transporte tiene una correlación negativa débil ($-0,057$), reflejando una relación menos directa. Sin embargo, la accesibilidad al transporte público es crucial para compradores que dependen de estos sistemas para sus desplazamientos diarios. Propiedades cercanas a estaciones son más convenientes y, por ende, más valoradas por quienes buscan conectividad sin necesidad de un vehículo.

Por último, variables como la superficie cubierta ($0,67$), el número de baños ($0,55$) y habitaciones ($0,27$) tienen un impacto considerable en los precios, ya que reflejan la funcionalidad y comodidad de la propiedad. Propiedades más grandes, con mayor número de habitaciones y baños, atraen a compradores dispuestos a pagar más por el espacio y la calidad de vida. Esta combinación de factores refuerza el valor percibido y real de las propiedades en un mercado competitivo como Bogotá.

Construcción de variables

Para complementar la base de datos, se calcularon nuevas variables utilizando la información ya existente en la base de datos y fuentes externas. Nos basamos en las distancias a puntos de interés clave en Bogotá para crear 4 nuevas variables: hospitales, parques, estaciones de transporte público y el centro financiero. Estas variables permiten medir la proximidad de cada propiedad a estos puntos estratégicos, los cuales tienen un impacto significativo en su valoración. A continuación, se describe el proceso de construcción:

Primero, se definieron las coordenadas del centro financiero de Bogotá (Calle 72 con Carrera Séptima) como el punto de referencia para medir la distancia de cada propiedad. Utilizando la función `geopy.distance.geodesic`,

se calculó la distancia en metros entre cada propiedad y este punto central. Esta nueva variable se añadió a las bases de entrenamiento y prueba bajo el nombre `distance_to_financial_center`.

Adicionalmente, para calcular las distancias a hospitales, parques y estaciones de transporte público, se utilizaron datos de las ubicaciones previamente recopilados y almacenados en archivos CSV de los cada uno de estos lugares. De la misma forma, usando la ubicación de las propiedades se calculo la distancia a cada uno de los lugares más cercanos dentro de esas tres categorías.

Las variables derivadas de la base de datos, se crearon utilizando técnicas básicas de procesamiento de lenguaje natural para extraer información de las descripciones de las propiedades. La variable `barrio` se generó identificando menciones de nombres de barrios específicos en los campos `title` y `description` mediante coincidencias de texto con una lista predefinida, a partir de ahí se construyen variables dummies para barrios de interes que serian utilizados en el analisis. Asimismo, se incluyó `cerca_virrey`, una variable binaria que toma el valor 1 si la propiedad se encuentra cerca al virrey. Adicionalmente, se contaron menciones de palabras clave relacionadas con parqueaderos y baños en las descripciones para generar las variables `num_parqueaderos` y `num_banos`. Estas transformaciones aprovecharon patrones textuales simples para capturar información relevante del texto descriptivo de cada propiedad.

Modelos y Resultados

Elastic Net

En este análisis, se utilizó un modelo de Elastic Net. El objetivo principal era capturar relaciones lineales y no lineales entre las variables independientes y el precio de las propiedades.

El modelo fue entrenado utilizando un conjunto de datos de entrenamiento que contenía variables como el año de construcción, superficie total, número de habitaciones, tipo de propiedad, proximidad a servicios, y variables dummy para los barrios. Los datos fueron particionados en un conjunto de entrenamiento (70 %) y un conjunto de validación (30 %).

Se utilizó validación cruzada para determinar el valor óptimo de `lambda`, el cual fue encontrado como 229209. El valor de `alpha` fue fijado en 0.5, lo que representa un equilibrio entre Lasso y Ridge.

El modelo alcanzó un **Error Absoluto Medio (MAE)** de 328,209,992 COP en el conjunto de validación. Este resultado indica que el modelo fue capaz de realizar predicciones relativamente precisas en cuanto al precio de las propiedades, aunque hay margen para mejorar la precisión.

Estos resultados muestran que el modelo puede capturar patrones importantes en los datos, pero también reflejan que los precios de las propiedades en Chapinero siguen siendo afectados por factores no completamente modelados, lo que sugiere que otras variables o modelos más complejos podrían ofrecer una mejor precisión en futuras iteraciones.

El modelo de Elastic Net ha demostrado ser una opción viable para predecir precios inmobiliarios, pero aún existe la oportunidad de mejorar su rendimiento mediante la inclusión de más variables o la exploración de técnicas más avanzadas.

Random Forest

El modelo Random Forest es un algoritmo basado en árboles de decisión que crea un conjunto de árboles de manera aleatoria y promedia las predicciones de todos ellos para obtener una predicción final más robusta y generalizable. Los hiperparámetros utilizados en este análisis son los siguientes:

- **Número de árboles (ntree):** Se utilizaron 500 árboles en el modelo. Este valor fue seleccionado para asegurar una buena robustez y evitar el sobreajuste. Un mayor número de árboles generalmente mejora la estabilidad del modelo, pero también incrementa el tiempo de cómputo.
- **Número de variables en cada división (mtry):** El valor de `mtry` se fijó en 3, lo que significa que en cada división de un árbol se consideraron 3 variables aleatorias para determinar la mejor división. Este valor controla la complejidad del modelo y ayuda a evitar el sobreajuste al restringir la cantidad de variables que se prueban en cada división.
- **Importancia de las variables:** Se habilitó el cálculo de la importancia de las variables para identificar qué factores tienen mayor peso en la predicción del precio de las propiedades. Esto ayuda a entender mejor los determinantes del precio y a mejorar la interpretación del modelo.

- **Manejo de valores faltantes (`na.action`):** Se utilizó la opción `na.omit` para eliminar las observaciones con valores faltantes en las variables utilizadas en el modelo.

Entrenamiento del Modelo

El modelo fue entrenado utilizando un conjunto de datos de entrenamiento que contenía variables como el año de construcción, superficie total, número de habitaciones, tipo de propiedad, proximidad a servicios y las variables dummy para los barrios. Los datos fueron particionados en un conjunto de entrenamiento (70 %) y un conjunto de validación (30 %).

Resultados

El modelo de Random Forest alcanzó un **Error Absoluto Medio (MAE)** de 284,073,992 COP en el conjunto de validación. Este resultado indica que el modelo logró una predicción razonablemente precisa, aunque aún existen áreas para mejorar la exactitud de las predicciones, especialmente considerando las características complejas del mercado inmobiliario de Chapinero.

Conclusión

El modelo Random Forest ha demostrado ser útil para predecir los precios de las propiedades en Chapinero, pero los resultados sugieren que se pueden explorar otras mejoras, como la inclusión de nuevas variables o la optimización de los hiperparámetros para reducir aún más el error de predicción.

Redes Neuronales

En este análisis, se utilizó un modelo de redes neuronales para predecir el precio de las propiedades. El objetivo principal era capturar relaciones complejas entre las variables independientes (como tamaño, ubicación y características físicas) y el precio de las propiedades. A continuación, se describen los detalles técnicos del modelo:

Arquitectura del modelo:

- La red neuronal consta de cinco capas densas (**Dense**) con activación ReLU.
- El número de neuronas en cada capa es: 264 (entrada), 128, 64, 32, y 1 (salida para predecir `log_price`).

- Se utilizó una capa de **Dropout** con una tasa del 30 % después de cada capa oculta para prevenir el sobreajuste.
- La salida es una predicción en escala logarítmica (**log_price**), lo cual ayuda a manejar la alta variabilidad de los precios pero también se exploró sin esta transformación y el resultado fue el mismo.

Hiperparámetros utilizados:

- Optimizador: **Adam**, con una tasa de aprendizaje (**learning_rate**) de 0.001.
- Función de pérdida: **mean_squared_error** (MSE).
- Métricas de evaluación: **mean_absolute_error** (MAE).
- Épocas de entrenamiento: 150.
- batch size: 16.

Entrenamiento del modelo: El modelo fue entrenado con datos de entrenamiento (**X_train** y **y_train**) y validado en un conjunto de datos independiente (**X_val** y **y_val**). Durante el entrenamiento, se supervisó la función de pérdida en ambos conjuntos para garantizar un buen desempeño y prevenir el sobreajuste.

Resultados: El modelo alcanzó los siguientes resultados en el conjunto de validación, después de revertir la transformación logarítmica:

- Error cuadrático medio (RMSE): 321,375,717.78 COP.
- Error absoluto medio (MAE): 241,693,067.71 COP.

Estos resultados indican que el modelo no es adecuado para el problema planteado. Los valores altos de RMSE y MAE reflejan que la red neuronal tiene dificultades para capturar las relaciones complejas entre las variables y predecir con precisión los precios. Esto puede deberse a una combinación de factores, como:

- Insuficiente calidad o cantidad de datos relevantes.
- Falta de variables predictoras clave que expliquen mejor la variabilidad de los precios.
- Necesidad de una mayor optimización del modelo o una arquitectura más específica para datos inmobiliarios.

Conclusión: El desempeño del modelo de redes neuronales es insatisfactorio en este caso. Sufre mucho por las variables que tiene una gran cantidad de missing values a que aun imputadas se pierda una gran cantidad de información relevante. así mismo lo que sucede es que la red predice para todos. Pese a normalizar las variables con mayor volatilidad, usar transformaciones para la variable dependiente, cambiar los tamaños de los batch y ajustar la tasa de aprendizaje, no fue posible mejorar los resultados.

Regresión Lineal

El modelo de regresión lineal es una técnica estadística que asume una relación lineal entre las variables independientes y la variable dependiente, en este caso, el precio de las propiedades. Este modelo fue utilizado como línea base para evaluar su capacidad de capturar patrones en los datos. A continuación, se describen los detalles del modelo:

- **Manejo de valores faltantes:** Los valores faltantes en las variables fueron rellenados con la media de cada columna para mantener la consistencia y evitar la pérdida de información relevante.

Entrenamiento del Modelo

El modelo fue entrenado utilizando un conjunto de datos que contenía variables como el año de construcción (`year`), superficie total y cubierta (`surface_total`, `surface_covered`), número de habitaciones (`rooms`, `bedrooms`, `bathrooms`), y proximidad a servicios. Los datos fueron particionados en un conjunto de entrenamiento (80 %) y un conjunto de validación (20 %).

Resultados

El modelo de regresión lineal alcanzó los siguientes resultados en el conjunto de validación:

- **Error Absoluto Medio (MAE):** 256,178512.36 COP.
- **Error Cuadrático Medio (MSE):** 65,437,915.27 COP.

Estos resultados muestran que el modelo logra capturar relaciones lineales básicas entre las características y el precio de las propiedades, pero tiene dificultades para manejar la complejidad del mercado inmobiliario en Chapinero.

Conclusión

El modelo de regresión lineal es útil como referencia inicial para las predicciones inmobiliarias. Sin embargo, los resultados indican que este enfoque es limitado para capturar relaciones más complejas. Se recomienda explorar enfoques más avanzados, como el Boosting o técnicas de aprendizaje no lineal, para mejorar la precisión de las predicciones.

Boosting (XGBoost)

El modelo de Boosting (XGBoost) es una técnica de aprendizaje supervisado basada en la combinación de múltiples modelos débiles (en este caso, árboles de decisión) para generar un modelo robusto. Este enfoque itera secuencialmente, ajustando los errores de los modelos previos, lo que le permite capturar relaciones complejas en los datos. A continuación, se describen los detalles del modelo y su implementación.

- **Número de árboles (`n_estimators`):** Se utilizaron inicialmente 500 árboles. Este valor permite capturar patrones complejos, pero también incrementa el tiempo de cómputo.
- **Tasa de aprendizaje (`learning_rate`):** Se fijó en 0.05 para ajustar el impacto de cada árbol en el modelo final, proporcionando un enfoque más conservador.
- **Profundidad máxima de los árboles (`max_depth`):** Los árboles tienen una profundidad máxima de 6, lo que controla la complejidad de cada árbol y previene el sobreajuste.
- **Submuestreo (`subsample`):** Se utilizó un valor de 0.8, lo que implica que cada árbol se entrena con el 80 % de los datos disponibles.
- **Submuestreo de características (`colsample_bytree`):** Se seleccionó un valor de 0.8, permitiendo al modelo considerar solo el 80 % de las características en cada árbol.
- **Manejo de valores faltantes:** Los valores faltantes fueron rellenos con la media de las columnas correspondientes en el conjunto de entrenamiento.

Entrenamiento del Modelo

El modelo fue entrenado utilizando un conjunto de datos que contenía variables como el año de construcción (`year`), superficie total y cubierta (`surface_total`, `surface_covered`), número de habitaciones (`rooms`, `bedrooms`, `bathrooms`), y proximidad a servicios. Los datos fueron particionados en un conjunto de entrenamiento (80 %) y un conjunto de validación (20 %).

Además, se utilizó un **Grid Search** para optimizar los hiperparámetros del modelo, explorando combinaciones de `n_estimators`, `learning_rate`, `max_depth`, `subsample` y `colsample_bytree`. La mejor combinación encontrada fue:

- `n_estimators`: 500
- `learning_rate`: 0.05
- `max_depth`: 6
- `subsample`: 0.8
- `colsample_bytree`: 0.8

Resultados

El modelo de Boosting (XGBoost) alcanzó los siguientes resultados en el conjunto de validación:

- **Error Absoluto Medio (MAE)**: 250.92 millones de COP.
- **Error Cuadrático Medio (MSE)**: 40,040,250.13 millones de COP.

Tras realizar el Grid Search, los resultados mejoraron ligeramente, con un nuevo MSE de:

- **MSE posterior al Grid Search**: 38,705,675.33 millones de COP.

Sin embargo, esta mejora fue marginal, lo que indica que los hiperparámetros iniciales ya estaban cerca de los valores óptimos. La falta de una mejora significativa puede deberse a la naturaleza de los datos y a la ausencia de variables predictoras clave que expliquen la variabilidad de los precios.

Conclusión

El modelo de Boosting (XGBoost) demostró ser una herramienta robusta para la predicción de precios inmobiliarios en Chapinero, ofreciendo resultados competitivos. Sin embargo, la falta de una mejora sustancial tras la optimización de hiperparámetros sugiere que el modelo puede estar limitado por la calidad y cantidad de los datos disponibles. Se recomienda explorar la inclusión de nuevas variables y datos adicionales para reducir aún más el error de predicción.

Comparación de resultados

Aunque los modelos más complejos, como Random Forest, Redes Neuronales y XGBoost, tienen ventajas en escenarios con relaciones no lineales, el mejor desempeño fue logrado por el modelo de Regresión Lineal. Esto sugiere que las relaciones lineales predominan en los datos y que las interacciones más complejas no fueron suficientemente relevantes para justificar el uso de modelos más avanzados. La Regresión Lineal se presenta, por tanto, como una solución eficaz y fácilmente interpretable para este problema.

Sin embargo, estos resultados también reflejan la necesidad de explorar la inclusión de nuevas variables que capturen de forma más precisa las dinámicas del mercado inmobiliario en Chapinero. Por ejemplo, características adicionales como la calidad de la construcción, acceso a servicios específicos y dinámicas de oferta y demanda podrían proporcionar mayor profundidad al análisis. Asimismo, modelos más complejos como Random Forest o XGBoost podrían beneficiarse de la optimización de hiperparámetros y del aumento en la cantidad y calidad de datos disponibles. Este enfoque combinado podría permitir encontrar un equilibrio entre interpretabilidad y capacidad predictiva, logrando un mejor ajuste y reduciendo aún más los errores de predicción.

Conclusiones y Recomendaciones

El análisis realizado en este trabajo resalta la importancia de utilizar modelos predictivos robustos para estimar los precios de propiedades en mercados inmobiliarios complejos como el de Chapinero. Aunque los modelos más avanzados, como XGBoost y Random Forest, demostraron ser herramientas útiles y con capacidades para capturar relaciones no lineales en los datos, el modelo de Regresión Lineal destacó como el más eficiente en términos de simplicidad y precisión, con un mejor MAE en comparación con los modelos que presentan una mayor complejidad.

Los resultados también muestran que existe un límite en el rendimiento alcanzado debido a la calidad y cantidad de los datos disponibles. La presencia de valores faltantes significativos y la falta de variables relevantes, como calidad de construcción, antigüedad exacta o dinámica de oferta y demanda, limitan la capacidad predictiva de los modelos. Estos factores generan una alta volatilidad en las predicciones, especialmente en los precios más extremos del mercado. La inclusión de esto, puede significar una mejora sustancial para el modelo.

En conclusión, la Regresión Lineal demostró ser una solución confiable y fácilmente interpretable para este problema, pero el potencial de mejora radica en la incorporación de datos adicionales y la optimización de modelos avanzados. Este enfoque podría reducir los errores de predicción y proporcionar herramientas más precisas para la toma de decisiones en el mercado inmobiliario de Chapinero.

Referencias

Apéndice