

Índice

1	Introducción	1
2	Datos y Estadísticas descriptivas	1
2.1	Datos	1
2.2	Estadísticas descriptivas	3
3	Modelos y Resultados	4
3.1	Mejor modelo	5
3.2	Otros modelos	6
3.2.1	Regresión lineal	6
3.2.2	Elastic Net	7
3.2.3	Random Forest	7
3.2.4	Redes Neuronales	7
4	Conclusiones	8

Índice de cuadros

1	Estadísticas descriptivas de las principales variables de la base de datos de entrenamiento	3
2	Modelos de Predicción	4

Índice de figuras

1	Ubicación geográfica de las propiedades	2
2	Missing values en la base de entrenamiento por categoría de variable	2
3	Histograma del precio del inmueble en millones de pesos por tipo de vivienda	4
4	Validación cruzada espacial XGBoost	6

1. Introducción

La oferta actual de *Big Data* nos permite tener grandes conjuntos de información sobre distintas medidas económicas, un ejemplo de esto es el mercado inmobiliario. Si bien hace algunos años solo podíamos saber características intrínsecas de los bienes inmuebles, ahora mediante el análisis de datos espaciales podemos incluir atributos adicionales que un consumidor evalúa sobre una vivienda.

El número de amenidades que rodean el hogar resulta de gran importancia; una hipótesis que surge es que un individuo está dispuesto a pagar más dinero por una vivienda cerca de un área comercial que una ubicada más lejos, aún con las mismas características. El número de supermercados aledaños, la distancia a una estación de policía, o las rutas de transporte público son factores subjetivos que los clientes de bienes inmuebles toman en cuenta al momento de hacer su compra, además de las variables tradicionales como el área, estrato y número de cuartos. Un posible problema que puede surgir con los nuevos conjuntos de información es el sobreajuste del modelo. Así mismo determinar qué factores son los más relevantes para la predicción del precio de una vivienda se convierte en un desafío.

Panagiotis y Tsadiras (2023) muestran que el mercado inmobiliario ya es sujeto del uso de la inteligencia artificial, generando búsquedas más precisas para los usuarios. Por lo tanto, un estimativo de costo es crucial para dar un precio promedio de un bien que se adapte a las necesidades de una persona o familia. Al igual que en nuestro caso de estudio, los autores utilizan las características base de la propiedad y amenidades (79 variables). La novedad del artículo es que usan como variable explicativa las oportunidades laborales en Iowa, y su métrica de evaluación es el RMSLE. El artículo evalúa varios tipos de algoritmos, entre los cuales se encuentra: Support Vector Machine, Kernel Ridge, Gradient Boosting, Lasso, Random Forest, XGB, LGBM, Averaged models y Voting Regressor. Dentro de los modelos vistos en clase destaca el modelo lineal con Lasso, y el Gradient Boosting dentro de los no lineales. Por el contrario, Random Forest tiene el peor comportamiento.

Debido a que Random Forest tiene el peor resultado en el anterior artículo, se indagó sobre este modelo en particular. Por su parte, Adetunji et al. (2022) utilizan el algoritmo de Random Forest para la predicción de los precios de las viviendas en Boston, con 14 características de estas. La clave de estos autores fue aumentar la muestra mediante un bootstrap. Para construir el árbol y reducir la necesidad computacional se usó una submuestra aleatoria sin reemplazo, y se formaron 500 nodos. La medida de precisión que usaron los autores fue el MAE y el RMSE. La lección sobre este artículo es que para realizar la proyección con arboles es necesario hacer algunos ajustes a la muestra como es el caso del bootstrap.

Finalmente el artículo de Chou y Truong (2022) muestra el estudio de predicción de precios en la zona metropolitana de Taipei (Taiwan). Al igual que los anteriores trabajos, este usa la teoría hedónica para la predicción de precios; es decir, toma en cuenta las características del hogar y amenidades alrededor de este. Además de los modelos lineales, los autores hacen uso de redes neuronales y CARTS. Algo clave es que en la limpieza de datos los autores son claros en que eliminan los datos outliers de bienes con un costo excesivo (algunos están por encima de cuatro veces la media). El mejor resultado para varias métricas fue el modelo híbrido de redes neuronales artificiales y Bagging.

Este trabajo busca predecir los precios de casas y apartamentos en venta de Bogotá, en su mayoría de la localidad de Chapinero. Los principales resultados de nuestro estudio muestran que el algoritmo más preciso para predecir los precios de oferta de las propiedades es *XGBoost*, seguido del algoritmo de Redes Neuronales, en ambos casos se encontró que los modelos más complejos con más variables descriptivas y componentes principales son los que mejores resultados arrojan. En este [repositorio](#) de GitHub se encuentran los datos utilizados para la predicción, así como los modelos y resultados que fueron utilizados para la competencia de *Kaggle*.

2. Datos y Estadísticas descriptivas

2.1. Datos

La base de datos de entrenamiento cuenta con 38.644 observaciones de propiedades en Bogotá, mientras que la base de testeo cuenta con 10.286 observaciones, de las cuales el 97% se encuentran ubicadas en Chapinero. Por su parte, la base de entrenamiento tiene sus propiedades en distintas localidades y solo el 11% en Chapinero

(figura 1). Las dos bases de datos contienen 16 variables, con la particularidad de que la base de testeo no tiene ningún precio.

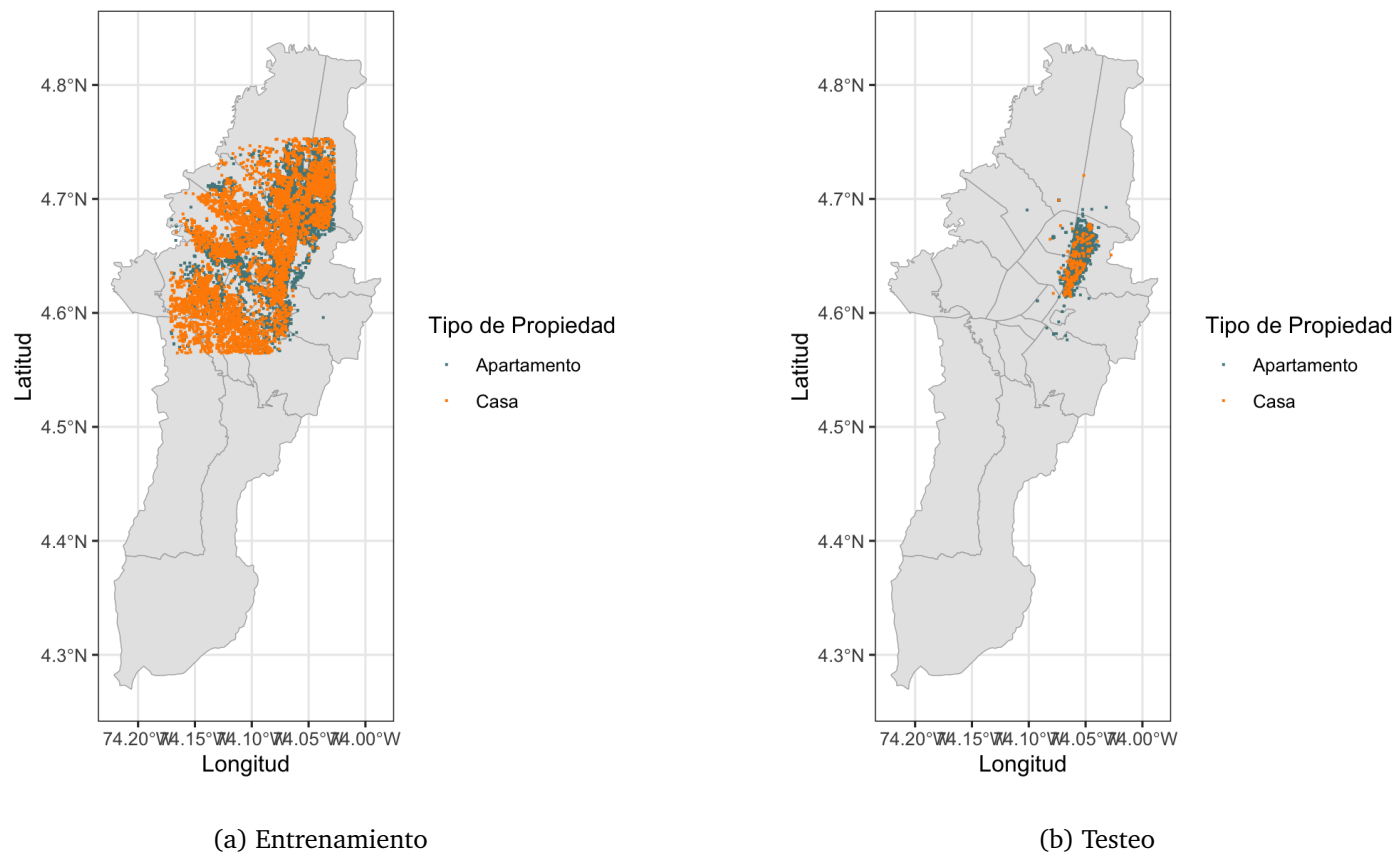


Figura 1: Ubicación geográfica de las propiedades

Esta base inicial cuenta con *missing values*. En particular, para la base de entrenamiento en el área total de la vivienda con el 80 % de los datos con *missing*, el área cubierta (78 %), habitaciones (47 %) y baños (26 %). La representación visual se puede observar en la gráfica 2.

Figura 2: *Missing values* en la base de entrenamiento por categoría de variable



Para solucionar los *missing values* se agrega por tipo de vivienda (casa o apartamento) y barrio, esta última variable se utilizó con georreferenciación de **Datos Abiertos Bogotá**. Para las variables “habitaciones” y “baños” se utiliza la moda dado que el número de habitaciones o baños se representa mediante un número entero, mientras

que para las variables “área total” y “área cubierta” se asignó el área promedio por barrio.

En vista de que este es un ejemplo real de Bogotá y existen paquetes de análisis espacial y datos públicos georreferenciados, se procedió a capturar nuevas variables que generen una variación sobre los precios de las viviendas. Para esto, se utilizaron fuentes de Datos Abiertos Bogotá y de [Open Street Maps \(OMP\)](#).

Con la fuente de Datos Abiertos Bogotá, además del barrio, se extrajo información de localidad y estrato. En específico, los datos de barrio y localidad permitieron hacer una mejor imputación de datos de las variables originales. No obstante, estas dos variables sirven parcialmente para la predicción, dado que las localidades y barrios de la muestra de testeo están subrepresentadas en la muestra de entrenamiento.

Por su parte, de OPM se extrajeron dos variables que representan amenidades de la vivienda. De acuerdo con el consejo de Bogotá, en 2023 se cometieron 147.666 hurtos a personas en la capital. La seguridad es un factor relevante para los bogotanos, y por ende podría ayudar a determinar el precio de una vivienda. Para plasmar esta variable, se utilizó como proxy la distancia más cercana a un CAI o estación de policía. Por otra parte, se anexó la distancia al parque más cercano, desde su centroide.

Por otro lado, la descripción de los bienes inmuebles a la venta permitió agregar al análisis variables extraídas del texto cuya información es valiosa para la predicción del precio de las viviendas. Se agregó la variable *número de pisos* de la vivienda y una *dummy* que es igual a 1 cuando en la descripción aparece “zona T o zona G”.

Finalmente, se extrajeron características difíciles de observar que pocas viviendas tienen como, por ejemplo, si la vivienda cuenta con área de BBQ, con depósito, chimenea, gimnasio, entre otros. Estas características poco usuales podrían ayudar a generar una variabilidad en el precio de viviendas con características similares, pero con estos diferenciales. Para este ejercicio se eliminaron *stop words* de la descripción y de su resultado; no se tuvieron en cuenta aquellas palabras que tuviesen entradas nulas (no se mencionaran en la descripción) del 90 %, lo que representó la creación de 71 variables nuevas.

2.2. Estadísticas descriptivas

Luego de consolidar los datos de la base de entrenamiento y testeo, se presentan las estadísticas descriptivas de la base de entrenamiento en la tabla 1. Allí se muestra la media, la desviación estándar, valores mínimos y máximos.

Tabla 1: Estadísticas descriptivas de las principales variables de la base de datos de entrenamiento

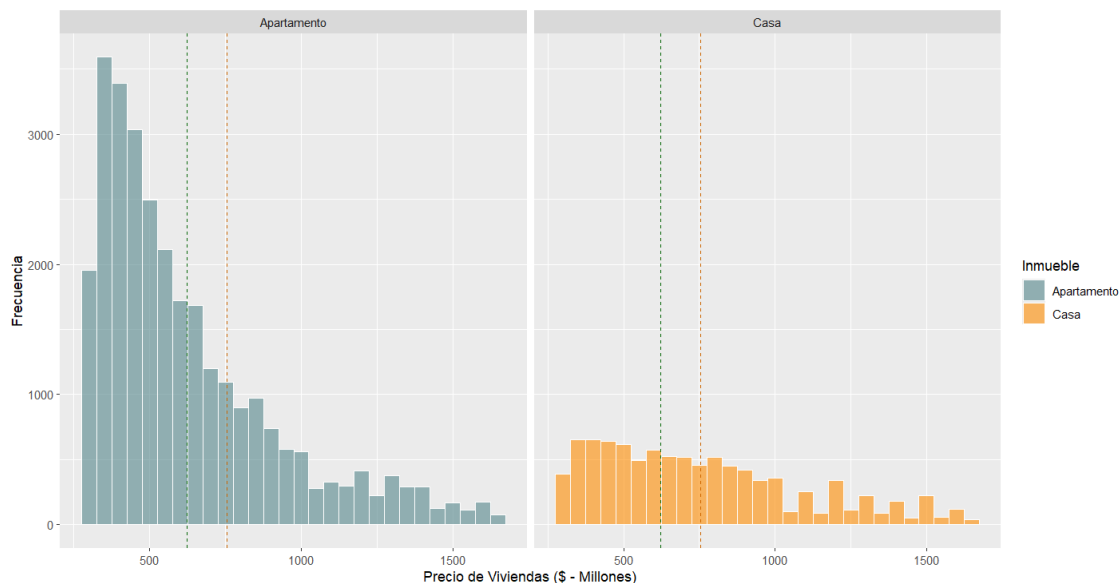
Variable	N	Media	Desv. Est.	Min	Máx
Dormitorios	38,644	3	2	0	11
Precio por m ² (millones)	38,644	5	2	0	40
Distancia a la policía	38,644	722	368	2	2,165
Distancia al parque	38,644	160	101	1	3,345
Estrato	38,644	4	1	0	6
Habitaciones	38,644	3	1	1	11
Baños	38,644	3	1	1	13
Área total	38,644	143	178	16	17,137
Área cubierta	38,644	138	70	2	1,336
Número de pisos	38,644	1	1	1	15

A partir de la imputación de datos se evidencia que el área total de la vivienda tiene una mayor desviación estándar, así como un valor máximo superior, con respecto a área cubierta. En ese sentido, área cubierta parece ser una variable más realista de las propiedades. El trade-off que se paga por usar esta variable -en lugar del área total- es que se pierde información sobre áreas extra del bien inmueble, como es el caso de balcones o zonas de

BBQ, por ejemplo. Sin embargo, se asume que los predios que cuentan con estas adiciones serán una minoría.

Por otro lado, el precio se comporta de la manera esperada. La figura 3 muestra el precio de las viviendas en millones de pesos dividido por apartamentos y casas. Las viviendas más económicas tienden a tener una mayor frecuencia -similar a la distribución de ingresos de los colombianos que muestra la GEIH-, mientras que son limitadas las viviendas que exceden las dos desviaciones estándar. En la capital se observa que los apartamentos son el tipo de vivienda con mayor cuota en el mercado inmobiliario. Otro factor para destacar es que el precio medio de las casas en Bogotá es 21 % mayor que el precio promedio de los apartamentos.

Figura 3: Histograma del precio del inmueble en millones de pesos por tipo de vivienda



3. Modelos y Resultados

En esta sección se describe el mejor modelo predictivo entregado en la competencia de Kaggle, junto con los detalles de su estimación y la selección de hiper-parámetros relevantes. Se utilizaron cinco algoritmos diferentes para predecir el precio de las viviendas: Regresión Lineal, Elastic Net, Random Forest, XGBoost y Redes Neuronales. Estos modelos fueron evaluados en términos de su capacidad predictiva, medida por el Mean Absolute Error (MAE), y sus puntajes se presentan en la Tabla 2. El modelo basado en el algoritmo de XGBoost obtuvo el mejor desempeño, con un MAE de 248,837,015. A continuación, se detalla el proceso de estimación de este modelo, incluyendo las variables utilizadas, la selección de hiper-parámetros, y la validación cruzada espacial. Asimismo, se comparan los resultados del modelo XGBoost con los obtenidos por los otros cuatro modelos presentados en Kaggle, proporcionando una visión comprensiva de su rendimiento relativo.

Tabla 2: Modelos de Predicción

No. Modelo	Algoritmo	Puntaje Kaggle
1	Regresión Lineal	259,082,291
2	Elastic Net	283,469,139
3	Random Forest	255,325,981
4	XGBoost	248,837,015
5	Redes Neuronales	253,544,306

3.1. Mejor modelo

El modelo con el mejor poder predictivo se estimó utilizando el algoritmo de *XGBoost*, donde se obtuvo un MAE de 248,837,015 en Kaggle y un MAE de entrenamiento de 100,712,716. Las variables que se utilizaron para entrenar este modelo fueron:

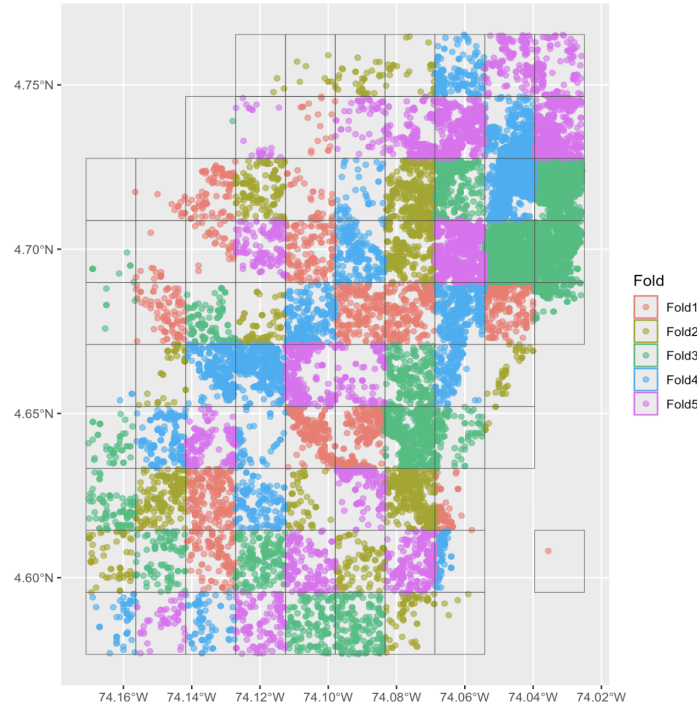
$$\begin{aligned} \text{Precio}_i = & \beta_0 + \beta_1 \text{Habitaciones}_i + \beta_2 \text{Baños}_i + \beta_3 \text{Área_total}_i \\ & + \beta_4 \text{Área_cubierta}_i + \beta_5 \text{Número_pisos}_i + \beta_6 \text{Zona_t_g}_i \\ & + \beta_7 \text{Estrato}_i + \beta_8 \text{Habitaciones}_i^2 + \beta_9 \text{Área_total}_i^2 \\ & + \beta_{10} \text{Área_cubierta}_i^2 + \beta_{11} \text{Apartamento}_i \\ & + \beta_{12} \text{Casa}_i + \gamma D_i + \phi P_i + \theta X_i + \epsilon_i \end{aligned} \quad (1)$$

Donde D_i son 10 dummies de localidad, P_i son 71 variables de componentes principales (PCA) y X_i son 71 variables de texto creadas a partir de la descripción del inmueble. La elección de estas variables se determinó de acuerdo a 3 factores relevantes que afectan directamente el precio de las viviendas:

- i) **Tamaño y capacidad:** Todas las variables sobre área total, área construida, número de habitaciones, baños, pisos y tipo de propiedad (apartamento o casa) afectan directamente el precio de los bienes raíces, pues entre mayor tamaño y capacidad, se espera que mayor sea el precio. Además, acá incluimos algunos términos cuadráticos que nos permiten capturar relaciones no lineales. Por ejemplo, el impacto adicional de una habitación extra puede disminuir a medida que el número total de habitaciones aumenta.
- ii) **Ubicación:** Las variables de estrato, zona T-G y localidad nos permite clasificar las viviendas según su ubicación geográfica y nivel socioeconómico de la zona, lo cual impacta directamente al precio de las viviendas, teniendo en cuenta que de acuerdo a la zona se pueden tener diferencias significativas en cuanto a servicios, infraestructura, seguridad y atractivo general.
- iii) **Características descriptivas:** Las variables de texto creadas a partir del análisis de las descripciones de las viviendas junto con las variables del análisis de componentes principales (PCA) aplicado a las características textuales, nos permiten capturar información valiosa sobre características específicas de las propiedades como por ejemplo si cuentan con chimenea o balcón, además de brindarnos información a través de las variables PCA sobre la variabilidad más significativa de las descripciones textuales que pueden influir en el precio.

En cuanto a la optimización del modelo, se llevó a cabo una cuidadosa selección de los hiper-parámetros a través de una búsqueda en grilla. Se utilizaron 500 rondas de *boosting* con una profundidad máxima de 4 para equilibrar la capacidad del modelo de capturar relaciones complejas sin sobreajustarse a los datos. La tasa de aprendizaje fue ajustada a 0.25, permitiendo al modelo aprender a un ritmo moderado. No se aplicó regularización adicional ($\gamma = 0$). Para fomentar la diversidad en los árboles y reducir el sobreajuste, el modelo utilizó el 66 % de las características (*colsample_bytree* = 0.66) y el 40 % de las observaciones (*subsample* = 0.4) en cada ronda de *boosting*. Finalmente, un valor de peso mínimo de las observaciones por nodo hijo (*min_child_weight*) de 50 fue seleccionado para asegurar que los nodos hijos tuvieran un número suficiente de observaciones, evitando nodos poco representativos. Esta combinación de hiper-parámetros permitió construir un modelo robusto y preciso, adecuado para capturar las complejas relaciones presentes en los datos de viviendas.

Figura 4: Validación cruzada espacial XGBoost



Para evaluar la consistencia y robustez del modelo *XGBoost* en la predicción del precio de las viviendas se utilizó un proceso de validación cruzada espacial con 5 pliegues (5-fold cross-validation). Este enfoque asegura que los datos se dividen en bloques espaciales, de modo que cada bloque representa un subconjunto geográficamente cohesivo de los datos. La Figura 4 de validación cruzada espacial muestra la distribución de los datos en cada uno de los cinco pliegues, indicados por diferentes colores.

Los resultados indican una distribución equitativa de los datos entre los diferentes pliegues, con cada color representando un *fold* diferente. Esto asegura que el modelo se entrena y valida en diferentes regiones espaciales, lo que ayuda a mitigar el riesgo de sobreajuste y garantiza que el modelo tenga una buena capacidad de generalización geográfica.

3.2. Otros modelos

3.2.1. Regresión lineal

En primera instancia, se realizaron distintos modelos de regresión lineal, a través de estimación de mínimos cuadrados ordinarios (MCO). Como resultado, se obtuvo el tercer mejor modelo de predicción, con un MAE de 259.082.291 (Kaggle). Los resultados se obtuvieron con el siguiente modelo:

$$\begin{aligned} \text{Precio}_i = & \beta_0 + \beta_1 \text{Área_cubierta}_i + \beta_2 \text{Área_total}_i + \beta_3 \text{Habitaciones}_i + \beta_4 \text{Baños}_i \\ & + \beta_5 \text{Estrato}_i + \beta_6 \text{Número_pisos}_i + \beta_7 \text{Zona_t_g}_i + \beta_8 \text{Distancia_policia}_i \\ & + \beta_9 \text{Distancia_parque}_i + \beta_{10} \text{Longitud}_i + \beta_{11} \text{Latitud}_i + \beta_{12} \text{Localidad}_i + \epsilon_i \quad (2) \end{aligned}$$

En esta aproximación se incluyen las variables iniciales con la imputación por tipo de vivienda y barrio, las variables espaciales, y se utiliza la latitud y la longitud como predictores para incluir el análisis espacial en la regresión lineal. En general, con los datos de entrenamiento la estimación evidencia significancia estadística con un nivel de confianza del 99 % para todos los predictores, excepto para superficie total, habitaciones, zona T y G, y las localidades de Bosa, Ciudad Bolívar y San Cristóbal; reflejando un MAE de 186,363,914.

3.2.2. Elastic Net

Para la estimación de Elastic Net se incluyeron más predictores, dado que su metodología permite llevar a cero los coeficientes de algunos de estos, con el propósito de evitar el sobreajuste del modelo. Para este tipo de algoritmo se obtuvo un mejor resultado con un modelo que implementaba 71 variables extraídas a partir de la descripción de la vivienda, y sus 71 componentes principales correspondientes.

Adicionalmente, incluyó las variables de tipo de propiedad, habitaciones, baños, superficie total, superficie cubierta, número de pisos y zona T-G. Considerando que los precios de las viviendas pueden estar correlacionados por su ubicación, se realiza una validación cruzada espacial con 5 folds (Figura 4). Los hiper-parámetros que obtuvieron el mejor resultado fueron $\alpha = 1$ y $\lambda = 0,1$, es decir, que este modelo refleja una estimación a través de *Lasso* en donde varios coeficientes pueden convertirse en cero.

Por último, esta metodología permitió obtener un MAE de entrenamiento de 180,012,424, pero un MAE de testeo de 283,469,139. Por lo tanto, podría inferirse que el incluir una gran cantidad de predictores, junto con sus componentes principales pudo crear sobreajuste en los datos, reflejando un mejor comportamiento en entrenamiento pero no en prueba.

3.2.3. Random Forest

El tercer modelo con el mejor poder predictivo del precio de las viviendas se estimó a través del algoritmo de *Random Forest*. Se obtuvo un MAE de 255,325,981 en Kaggle y un MAE de entrenamiento de 47,857,225. Para este modelo, se emplearon las mismas variables de la ecuación 1. En cuanto a los hiper-parámetros seleccionados, estos son: el número de predictores en cada división ($mtry = 10$), el número mínimo de observaciones por nodo ($min_n = 2$), y el número de árboles ($trees = 1000$). La elección de estos valores, guiada por la métrica de error absoluto medio (MAE), asegura un equilibrio entre la diversidad de los árboles y la granularidad de los nodos, permitiendo al modelo capturar patrones complejos sin sobreajustarse. La validación cruzada espacial refuerza la capacidad del modelo para generalizar en diferentes regiones geográficas, resultando en predicciones precisas y robustas.

Como podemos ver en la Tabla 2, a pesar de tener la misma especificación que el mejor modelo, tuvo una distancia importante respecto al MAE, esto puede deberse a varias razones, pues *XGBoost* maneja mejor las interacciones complejas y no linealidades gracias a su naturaleza iterativa de *boosting*. Además, incluye parámetros de regularización que previenen el sobreajuste, y permite un control más granular sobre la complejidad del modelo. Adicionalmente, *XGBoost* optimiza directamente la función de pérdida específica para la tarea, maneja mejor los datos desbalanceados y es altamente eficiente en términos de tiempo de entrenamiento y predicción. Estas características le permiten adaptarse mejor a las particularidades del conjunto de datos y mejorar su capacidad predictiva sobre el algoritmo de *Random Forest*.

3.2.4. Redes Neuronales

En esta sección se propusieron dos modelos de red neuronal para la predicción de precios de vivienda. Por un lado, se utilizó la librería *nnet* de R para construir un modelo con una sola capa. La estructura de esta red se puede representar de la siguiente manera:

$$\hat{y} = f(W \cdot X + b) \quad (3)$$

Donde X^1 es el vector de entrada, W es la matriz de pesos y b es el vector de sesgo. Este modelo se entrenó utilizando un conjunto de datos preprocesado que incluyó normalización, creación de variables dummy y eliminación de predictores con varianza cero.

Por otro lado, utilizando la librería *keras* de R y realizando una red compuesta por tres capas, se obtuvo el segundo mejor puntaje:

$$\hat{y} = f(W_3 \cdot g(W_2 \cdot g(W_1 \cdot X + b_1) + b_2) + b_3) \quad (4)$$

¹Las variables empleadas como vector de entrada se corresponden con aquellas utilizadas en el análisis de regresión lineal.

Donde X es el vector de entrada, W_1, W_2, W_3 son las matrices de pesos de cada capa, b_1, b_2, b_3 son los vectores de sesgo de cada capa, g es la función de activación ReLU, f es la función de activación lineal para la capa de salida, y \hat{y} es la predicción del modelo.

Para este modelo, se utilizó la misma preparación de datos que para el modelo de una sola capa, incluyendo normalización, creación de variables dummy y eliminación de predictores con varianza cero. Asimismo, se crearon tres capas ocultas con 64, 32 y 16 unidades con activación ReLU y una capa de salida lineal conjunto a técnicas de *Early Stopping*² y reducción de la tasa de aprendizaje³ durante el entrenamiento para evitar el sobreajuste.

Las predicciones obtenidas permitieron observar un rendimiento superior del modelo de 3 capas utilizando keras, frente al de una sola capa con la librería nnet, con un MAE de: (253, 544, 306) y (332,929,135) respectivamente. No obstante, el rendimiento del modelo de tres capas, fue superado por el rendimiento del modelo XGboost, como se pudo observar al inicio de esta sección.

4. Conclusiones

En el presente trabajo se llevó a cabo un análisis exhaustivo con el objetivo principal de construir un modelo predictivo de los precios de oferta de las viviendas en Bogotá, enfocado principalmente en la localidad de Chapinero. A partir del artículo de Rosen Rosen (1974), se planteó que el precio de las viviendas se ve afectado por sus atributos estructurales, servicios públicos del vecindario y comodidades locales. Por lo tanto, se utilizaron datos espaciales con características y descripciones de propiedades en Bogotá con más de 38 mil observaciones, estos se complementaron con georreferenciación de [Datos Abiertos Bogotá](#) y de [Open Street Maps \(OMP\)](#).

En la búsqueda del mejor modelo predictivo se entrenaron más de 18 modelos empleando diversas técnicas de Machine Learning como Regresión Lineal, *Elastic Net*, *Random Forest*, *XGBoost* y Redes Neuronales. Se encontró que el modelo basado en *XGBoost* demostró ser el más preciso, obteniendo un MAE de 248,837,015 en la competencia de Kaggle. Su éxito se debe a la capacidad de manejar interacciones complejas y no linealidades de manera efectiva, gracias a su naturaleza iterativa de *boosting* y a sus parámetros de regularización que previenen el sobreajuste.

Así mismo, encontramos que los modelos más complejos tenían un mejor *performance* que los más simples, siempre y cuando incluyeran variables del tamaño y capacidad de la vivienda (teniendo en cuenta también distintos tipos de formas funcionales de estas variables), su ubicación (que permita capturar el nivel socioeconómico de la zona en que se encuentra) y características descriptivas de las propiedades, así como lo sugiere el artículo de Rosen Rosen (1974).

La validación cruzada espacial fue crucial para evaluar la robustez de los modelos, asegurando que los datos se dividieran en bloques geográficamente cohesivos y evitando el sobreajuste. Los resultados mostraron que, aunque modelos como Redes Neuronales y *Random Forest* también presentaron un buen desempeño, *XGBoost* optimiza directamente la función de pérdida específica para la tarea y maneja mejor los datos desbalanceados, demostrando ser altamente eficiente en términos de tiempo de entrenamiento y predicción.

En este sentido, se recomienda utilizar modelos complejos con técnicas de regularización, como *XGBoost*, para predicciones de precios de vivienda, e incorporar variables espaciales y socioeconómicas junto con características textuales extraídas de las descripciones de propiedades. Este enfoque integral puede mejorar significativamente la precisión de las predicciones en diferentes contextos geográficos, proporcionando un modelo robusto y preciso para el análisis del mercado inmobiliario.

Referencias

Adetunji, Abigail Bola et al. (2022). «House Price Prediction using Random Forest Machine Learning Technique». En: *Procedia Computer Science* 199. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 2021): Developing Global Digital Economy after COVID-19,

²Se monitoreó el MAE en el conjunto de validación y se detuvo el entrenamiento si el error no mejoraba después de 10 *epochs*.

³Se redujo la tasa de aprendizaje en un factor de 0.1 si el MAE no mejoraba después de 5 *epochs*.

- págs. 806-813. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2022.01.100>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050922001016>.
- Chou Jui-Sheng; Fleshman, Dillon-Brandon y Dinh-Nhat Truong (2022). «Comparison of machine learning models to provide preliminary forecasts of real estate prices». En: *Journal of Housing and the Built Environment*. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 2021): Developing Global Digital Economy after COVID-19. ISSN: 1573-7772. DOI: [10.1007/s10901-022-09937-1](https://doi.org/10.1007/s10901-022-09937-1). URL: <https://link.springer.com/article/10.1007/s10901-022-09937-1>.
- Panagiotis, Konstantinos y Athanasios Tsadiras (2023). «Comparing Machine Learning Techniques for House Price Prediction». En: *Springer Nature Switzerland*. Ed. por Ilias Maglogiannis et al. Artificial Intelligence Applications and Innovations, págs. 292-303.
- Rosen, Sherwin (1974). «Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition». En: *Journal of Political Economy* 82(1), págs. 34-55. URL: <http://www.jstor.org/stable/1830899>.