
Aprendizaje estadístico interpretable: una aplicación al precio de inmuebles en Montevideo

Ignacio Alvarez-Castro

Natalia da Silva

Leonardo Moreno

Andrés Sosa

Serie Documentos de Trabajo

Preprint para Nº1/22

Febrero, 2022

Universidad de la República.
Facultad de Ciencias Económicas y de Administración,
Instituto de Estadística (IESTA)

Montevideo, Uruguay.



Esta obra está bajo una Licencia Creative Commons Atribución - NoComercial - CompartirIgual 4.0 Internacional.

Forma de citación sugerida para este documento:

Ignacio Alvarez-Castro, Natalia da Silva, Leonardo Moreno, Andrés Sosa (2022). *Aprendizaje estadístico interpretable: una aplicación al precio de inmuebles en Montevideo* (Serie Documentos de Trabajo; Nº1/22). Montevideo: Universidad de la República. Facultad de Ciencias Económicas y de Administración, Instituto de Estadística.

<https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/10518>

Ignacio Alvarez-Castro ¹; Natalia da Silva ² Leonardo Moreno ³; Andrés Sosa ⁴
*Departamento de Métodos Cuantitativos, Instituto de Estadística, Facultad de Ciencias
Económicas y de Administración, Universidad de la República*

Resumen

Las técnicas de aprendizaje estadístico supervisado se han desarrollado a lo largo del tiempo en muy diversas líneas de investigación y en una gran variedad de problemas aplicados. Estos métodos principalmente se emplean con el fin de predecir su respuesta a nuevos datos en problemas complejos debido a su gran flexibilidad en comparación a los modelos clásicos. Sin embargo, mucho de estos métodos se deno minan de caja negra debido a que no son claras las relaciones que se generan entre las variables en las funciones estimadas. En este sentido, el aprendizaje estadístico interpretable se ha vuelto un área de investigación muy activa en los últimos años. El objetivo de este documento es describir algunas de las técnicas más utilizadas del aprendizaje estadístico interpretable. Con el fin de comprender en mayor profundidad las técnicas propuestas se realiza una aplicación económica con datos de precio de oferta de inmuebles en Montevideo. Se aplican varios modelos estadísticos supervisados de regresión para predecir el precio de los mismos y se calculan algunas medidas de interpretabilidad para ellos que permiten analizar el efecto de un conjunto de variables explicativas en el precio del inmueble.

Palabras clave: ALE; Aprendizaje estadístico interpretable; Aprendizaje supervisado; ICE; PD-plot; Precio de oferta de inmuebles.

CÓDIGOS JEL: C32, C33, C81, D22, O47, Q14.

Clasificación MSC2010: 62-07; 62P20 .

¹ *email:* ignacio.alvarez@fcea.edu.uy , ORCID:<https://orcid.org/0000-0003-1633-2432>

² *email:* natalia.dasilva@fcea.edu.uy, ORCID: <https://orcid.org/0000-0002-6031-7451>

³ *email:* leonardo.moreno@fcea.edu.uy, ORCID: <https://orcid.org/0000-0003-1630-1361>

⁴ *email:* andres.sosa@fcea.edu.uy, ORCID: <https://orcid.org/0000-0002-6007-4373>

1. Introducción

Los métodos de aprendizaje estadístico se basan en la idea de entrenar un sistema, *un modelo*, que sea capaz de aprender el comportamiento de fenómenos a partir de un subconjunto de datos observados. Es frecuente el uso de estas técnicas en problemas complejos ya que su flexibilidad permite mejorar el comportamiento predictivo y capturar relaciones que no son capturadas con otros métodos clásicos. Estas técnicas son utilizadas en una amplia variedad de problemas en diversas áreas de estudio. Sin embargo, la aplicación de métodos de aprendizaje estadístico para analizar datos económicos y sociales es de uso relativamente reciente debido a diferentes causas (Grimmer *et al.*, 2021). Uno de los motivos para el rezago relativo se debe a la escasa *interpretabilidad* de los resultados que se obtienen. Los métodos de aprendizaje estadístico muchas veces se describen como buenos para predecir pero no para explicar el fenómeno de interés. En los últimos años se ha generado un fuerte impulso en la investigación sobre aprendizaje estadístico interpretable (Molnar *et al.*, 2020; Du *et al.*, 2019; Rudin *et al.*, 2021) y su utilización en distintas aplicaciones (Azodi *et al.*, 2020).

No hay consenso en la definición de *interpretabilidad* en el aprendizaje estadístico (Murdoch *et al.*, 2019), pero en términos generales es posible asociarlo al hecho de caracterizar o describir el efecto de algunas variables explicativas incluidas en el modelo sobre la variable de respuesta. El problema de la interpretabilidad puede ser atacado principalmente de dos maneras: model-specific y model-agnostic. En la primera, la idea fundamental es utilizar modelos estadísticos (o simplificaciones de otros modelos) que contengan propiedades de interpretación de los resultados, entre ellos se destacan los modelos lineales y los árboles de decisión. En este trabajo el foco se centra en la segunda, la cual es más general que la anterior debido a que involucra herramientas que no se basan en ningún modelo estadístico particular para explicar el comportamiento de las predicciones individuales.

El objetivo del documento de trabajo es describir algunas de las técnicas más utilizadas en la interpretación de los modelos de aprendizaje estadístico. Con el fin de obtener un mayor entendimiento se presenta un problema de aplicación económica de relevancia como lo es la valuación de inmuebles en una ciudad. En tal sentido, contar con una modelización espacial del precio de un conjunto de inmuebles en una determinada región es un problema de alto interés para diversas entidades financieras y gubernamentales. Estos modelos pueden ser una herramienta a considerar en la toma de decisiones tanto a nivel político como empresarial (Osland, 2010; Case *et al.*, 2004). En este trabajo dos fuentes de información serán utilizadas: información espacial en referencia a locación de la propiedad y un conjunto de variables intrínsecas al inmueble que se denomina información hedónica. La información espacial del inmueble se basa en la geo-referenciación del mismo. Es conocido en la literatura la influencia de su ubicación en el valor del mismo (Kiel y

Zabel, 2008). La contextualización de la propiedad al barrio, el acceso a servicios, el valor de inmuebles cercanos, la distancia a la playa y los índices de violencia de la zona son aspectos asociados a la ubicación del inmueble que influyen directamente en el precio. Por otro lado se consideran aquellas variables propias de cada inmueble (por ejemplo: número de dormitorios, superficie del inmueble, antigüedad de la propiedad, si cuenta o no con cochera), que se denominan en la literatura variables hedónicas (Sirmans *et al.*, 2005). En términos generales, muchos modelos propuestos en la literatura parten del supuesto que el precio del inmueble puede ser descompuesto linealmente como suma del precio de sus atributos, es decir se descompone el precio mediante un modelo lineal. Una ventaja de este tipo de modelos es la fácil explicación del efecto de las covariables consideradas en el modelo en términos de sus coeficientes estimados. Sin embargo es menos frecuente el uso de modelos de regresión no lineales y en particular métodos de aprendizaje estadístico para modelar este tipo de problemas (Goyeneche *et al.*, 2017; Limsombunchai, 2004; Park y Bae, 2015; Yoo *et al.*, 2012).

Este documento se estructura de la siguiente manera: en la Sección 2 se presenta una revisión bibliográfica de algunos métodos gráficos de aprendizaje estadístico interpretable. En la Sección 3 se introduce la aplicación económica y se presenta una descripción de los datos utilizados. En la Sección 4 se muestran los determinantes de precios ofertados que surgen de aplicar los métodos descritos previamente a los datos presentados en la Sección 3. Finalmente, en la Sección 5 se realiza una síntesis del trabajo y se expresan posibles líneas de trabajo futuro. Este documento de trabajo fue elaborado por el grupo de investigación de Modelización Estadística para el Análisis de Datos (MESTAD) creado en 2019 en el Instituto de Estadística (IESTA).

2. Métodos para aprendizaje estadístico interpretable

En los últimos años se ha generado un fuerte desarrollo en la interpretabilidad del aprendizaje estadístico (Linardatos *et al.*, 2021). Una simple búsqueda en *Google Scholar* muestra más de 32000 resultados en los últimos 5 años referentes a esta línea de investigación.

Como se establece en la Sección 1, no existe consenso sobre una definición de interpretabilidad ni como hacer una evaluación comparativa (Doshi-Velez y Kim, 2017). Sin embargo, es posible establecer en términos generales que la interpretabilidad de un modelo es el grado en que una persona puede entender la causa de la decisión (Miller, 2019). La interpretabilidad se concentra en comprender que es lo que hizo un modelo o debería haber hecho. Tener modelos de aprendizaje estadístico interpretables es importante ya

que permite comprender en mayor profundidad las relaciones entre la respuesta del modelo (o decisión a tomar) y las distintas variables predictoras. También pueden brindar información sobre cuando el modelo de predicción utilizado puede fallar y de ser necesario modificarlo en base a ello. A su vez contar con métodos de aprendizaje interpretables puede ser importante para detectar sesgos en los modelos y al ser interpretable puede ser explicada y auditada la respuesta. Los modelos interpretables son o se perciben como más confiables por eso además su importancia.

Si es posible asegurar que un modelo de aprendizaje estadístico puede explicar decisiones es más sencillo de chequear las siguientes características (Doshi-Velez y Kim, 2017):

- **Justicia:** Tener predicciones imparciales, sin favorecer grupos.
- **Privacidad:** Proteger información sensible en los datos.
- **Fiabilidad:** Asegurar que pequeños cambios en los datos no impliquen grandes cambios en la predicción.
- **Causalidad:** Chequear que solamente relaciones causales son seleccionadas.
- **Confianza:** Es sencillo para humanos confiar en un sistema que explica sus decisiones que uno que no (black box).

Es posible establecer dos maneras de estudiar la interpretabilidad en los modelos de aprendizaje. Por un lado, utilizar modelos que logren en sus resultados ser interpretables (model-specific), por otro lado aplicando técnicas que permitan analizar los resultados de cualquier modelo posterior a ser entrenado (model-agnostic). A su vez, los métodos intrínsecamente interpretables pueden ser resumidos de acuerdo a su resultado, estadísticas de resumen para las variables explicativas, visualizaciones que resumen las variables explicativas, características intrínsecas al modelo como ser la estructura de un árbol o los parámetros estimados en un modelo lineal. Otros resultados son los datos, en esta categoría entran los modelos que generan datos para hacer el modelo interpretable (explicación contrafactual).

Es posible establecer que las diferentes técnicas de interpretabilidad tienen distintos alcances, entre ellas se encuentran, la transparencia del método (explica como el algoritmo crea el modelo), la interpretación global del modelo (como el modelo entrenado hace predicciones y como las partes afectan la predicción), la interpretación local para una sola predicción (razón por la cual el modelo hace predicciones para una observación particular) y la interpretación local para un grupo de predicciones (razón por la cual el modelo hace predicciones específicas para un grupo de observaciones). Vale destacar también que dentro del campo de la interpretabilidad se analiza la sensibilidad del modelo a cambios en

las variables de entrada. Si bien, en este contexto, la teoría tiene puntos de contacto con el análisis de sensibilidad de modelos de simulación matemáticos la metodología empleada difiere en un caso y en el otro (Molnar *et al.*, 2020).

En esta sección se seleccionan algunas de las técnicas más utilizadas para interpretar modelos estadísticos que no son específicas sino que se pueden utilizar para todos los modelos de aprendizaje (model-agnostic). El foco está principalmente en técnicas que presentan efectos promedio e individuales de las variables. Entre ellas se destacan: el gráfico de dependencia parcial (-PD-plot-, (Friedman, 2001)), el gráfico marginal (marginal plot -M-plot-), la esperanza condicional individual (Individual Conditional Expectation -ICE-, Goldstein *et al.* (2015)) y efectos locales acumulados (accumulated local effects -ALE-, Apley y Zhu (2020)).

El marco matemático teórico es el adecuado para el desarrollo de un modelo de aprendizaje supervisado. El objetivo de todos los modelos de aprendizaje es aproximar $E(Y|X = x) \approx f(x)$ donde $X = (X_1, X_2, \dots, X_q)$ es un vector de q -predictores, Y es la variable respuesta y f es el modelo ajustado que predice el escalar Y como función de X . Vamos a asumir que nos interesa comprender el efecto principal en la dependencia del valor $f(x)$ en un conjunto de predictores determinado por el conjunto $S \in \{1, 2, \dots, q\}$. Los resultados se obtienen mediante los datos de entrenamiento del modelo que consisten en N observaciones de $\{x^{(i)}, y^{(i)}\}_{i=1}^N$ donde $x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_q^{(i)})$.

2.1. Gráfico de dependencia parcial (PD-plot)

Esta primer técnica es la base fundamental del aprendizaje estadístico interpretable. Esto se debe a que la estimación es muy intuitiva y presenta una interpretación causal en los resultados de cualquier modelo de aprendizaje. El gráfico de dependencia parcial es un método que describe el cambio de la variable de respuesta en un modelo en función del efecto marginal de una o más variables de interés (conjunto índice S) al promediar los efectos de las demás variables predictoras (conjunto C).

En la ecuación 1 se define la función de dependencia parcial.

$$\begin{aligned} f_{S,PD}(x_S) &= E_{X_C}[f(x_S, X_C)] \\ &= \int f(x_S, x_C) dP(x_C) = \int p_{X_C}(x_C) f(x_S, x_C) d(x_C), \end{aligned} \quad (1)$$

donde se utiliza la función de distribución de x_C . Esto resulta en una función que depende únicamente de las variables en el conjunto S y de las interacciones con otras variables

incluidas. En este caso, estamos interesados en el subconjunto de predictores S , pero vale destacar que cada elección de predictores tiene su propia función de dependencia parcial f_S , que brinda el promedio de la función f cuando x_S es fijo y x_C varía sobre su distribución $dP(x_C)$.

La función de dependencia parcial se estima como el promedio en los datos de entrenamiento como se muestra en la Ecuación (2).

$$\hat{f}_{S,PD}(x_S) = \frac{1}{n} \sum_{i=1}^N f(x_S, x_C^{(i)}). \quad (2)$$

Es importante destacar, que en el gráfico de dependencia parcial se supone que las variables explicativas en C no están correlacionadas con las variables en S . En caso contrario, lo que sucede en la estimación es que al promediar en las observaciones se realiza una ponderación equiprobable en puntos en los cuales el comportamiento es muy distinto según el nivel de correlación que presenten las variables.

2.2. Gráficos marginales (M-plot)

El gráfico marginal es otro método de interpretabilidad similar en concepción al gráfico de dependencia parcial. En este método, se busca mejorar el comportamiento en el caso que exista correlación entre las variables en el conjunto S y el conjunto C . Esto se realiza mediante la densidad condicional (y no la densidad marginal) en las variables del conjunto C . Mediante esta acción en la estimación de la función se intenta evitar el uso de puntos que tienen probabilidad muy baja de ocurrencia.

En la ecuación 3 se define la función gráfico marginal

$$f_{S,M}(x_S) = E_{X_C}[f(X_S, X_C)|X_S = x_S] = \int p_{x_C|x_S}(x_C|x_S) f(x_S, x_C) dx_C. \quad (3)$$

La estimación de $f_{S,M}$ se define en la Ecuación (4)

$$\hat{f}_{S,M}(x_S) = \frac{1}{n(x_S)} \sum_{i \in N(x_S)} f(x_S, x_C^{(i)}), \quad (4)$$

donde $N(x_S) \subset \{1, 2, \dots, q\}$ es el conjunto de índices i para el cual $x_S^{(i)}$ se encuentra en un entorno de x_S y $n(x_S)$ es la cantidad de observaciones en el entorno elegido.

Es importante observar que dentro de los posibles inconvenientes que presenta este método se encuentra la posibilidad de que en los resultados se incorporen los efectos compartidos en todas las variables correlacionadas en el efecto de una única variable.

2.3. Esperanza condicional individual (ICE)

Esta técnica es diferente a los dos anteriores debido a que el resultado está formado por las N gráficas que se obtienen de los datos de entrenamiento. En esta técnica se pretende reflejar la respuesta del modelo como función de las variables x_S condicionada a una observación de la variable x_C . Se intenta capturar la dependencia en la variable respuesta en un conjunto de variables.

Para cada observación en el conjunto de entrenamiento, la gráfica $\hat{f}_{S,ICE}^{(i)}(x_S)$ se obtiene al variar en $i \in \{1, 2, \dots, N\}$ el valor $x_S^{(i)}$ mientras las variables x_C permanecen constantes. Es decir

$$\hat{f}_{S,ICE}^{(i)} = f(x_S, x_C^{(i)}).$$

Para cada observación i , es posible que las gráficas difieran en niveles lo que dificulta la interpretación. Por lo cual, se define de manera muy similar el ICE-centrado como solución a este problema, donde se centran las curvas a cierto nivel en la variable explicativa como se muestra en la Ecuación (5).

$$\hat{f}_{cent,S,ICE}^{(i)} = \hat{f}_{S,ICE}^{(i)} - \mathbf{1}\hat{f}_{S,ICE}^{(i)}(x_S^*, x_C^{(i)}), \quad (5)$$

donde el valor $(x_S^*, x_C^{(i)})$ se representa como el caso base para cada gráfica.

En este método, cada gráfica representa el valor del modelo al modificar los datos en las variables del conjunto C . Vale la pena destacar que si promediamos todas las gráficas de las curvas del método esperanza condicional individual, obtenemos el gráfico de dependencia parcial. Es posible establecer que este método es en esencia muy similar al gráfico de dependencia parcial, pero que permite observar interacciones al visualizar las N gráficas (cuestión que en el promedio es posible que se disuelvan los efectos).

2.4. Efectos acumulados locales (ALE-plot)

La última técnica que se describe es la de efectos acumulados locales. Esta técnica vuelve a tener mucha relación con el gráfico de dependencia parcial en el cual se pretende solucionar (de otra manera que el gráfico marginal) el problema de correlación entre las variables. En general, es posible establecer que en los efectos acumulados locales se divide la q dimensiones en hipercubos que definen las variables incluidas en S y se calcula la diferencia en la predicción del modelo al considerar los extremos del hipercubo de dimensión S . Estas diferencias se acumulan y luego se centran para obtener la gráfica de efectos acumulados locales (ALE-plot). Es decir

$$h_{S,ALE}(x_S) = \int_{z_{0,1}}^{x_{S,1}} \int_{z_{0,2}}^{x_{S,2}} \cdots \int_{z_{0,d}}^{x_{S,d}} E \left[\frac{\partial^d f(x_S, x_C)}{\partial X_{S,1} \partial X_{S,2} \cdots \partial X_{S,d}} \middle| X_{S,1} = z_1, X_{S,2} = z_2, \dots, X_{S,d} = z_d \right] dz_1 dz_2 \cdots dz_d$$

para posteriormente centrarse,

$$f_{S,ALE}(x_S) = h_{S,ALE}(x_S) - CONS(x_{S,1}, x_{S,2}, \dots, x_{S,d}),$$

tal que $f_{S,ALE}(X_S)$ tiene media cero con respecto a la distribución marginal X_S y el efecto ALE-plot marginal de cada $x_{S,i}$ en $f_{S,ALE}(X_S)$ son todas cero.

En este método, la estimación de la función marginal (no centrada) es dada por

$$\hat{h}_{x_S,ALE}(x_S) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i: x_j^{(i)} \in N_j(k)} \left[f(z_{k,j}, x_{\setminus j}^{(i)}) - f(z_{k-1,j}, x_{\setminus j}^{(i)}) \right].$$

La estimación del efecto marginal de una variable desarrollado en la ecuación puede ser generalizado al efecto conjunto de d variables. Por ejemplo para analizar el efecto conjunto de dos variables x_j y x_l el estimador (no centrado) es definido por

$$\hat{h}_{\{j,l\},ALE}(x_j, x_l) = \sum_{k=1}^{k_j(x_j)} \sum_{m=1}^{k_l(x_l)} \frac{1}{n_{\{j,l\}}(k,m)} \sum_{\{i: \mathbf{x}_{i,\{j,l\}} \in N_{\{j,l\}}(k,m)\}} \Delta_f^{\{j,l\}}(K, k, m, \mathbf{x}_{i,\{j,l\}}),$$

con $(x_j, x_l) \in (z_{0,j}, z_{K,j}] \times (z_{0,l}, z_{K,l}]$ y

$$\Delta_f^{\{j,l\}}(K, k, m, \mathbf{x}_{i,\{j,l\}}) = f(z_{k,j}^K, z_{m,l}^K, \mathbf{x}_{i,\setminus\{j,l\}}) - f(z_{k-1,j}^K, z_{m,l}^K, \mathbf{x}_{i,\setminus\{j,l\}}) - f(z_{k,j}^K, z_{m-1,l}^K, \mathbf{x}_{i,\setminus\{j,l\}}) + f(z_{k-1,j}^K, z_{m-1,l}^K, \mathbf{x}_{i,\setminus\{j,l\}}).$$

Es importante destacar que los gráficos de efectos acumulados locales pueden tornarse un poco inestables cuando el número de intervalos es grande. Es posible ampliar el número de intervalos al suavizar la predicción pero resignando parte de la complejidad del problema. En esta situación no existe una solución perfecta para establecer el número de intervalos: si el número es demasiado pequeño, es posible que el gráfico no sea muy preciso en cambio si el número es demasiado grande el gráfico puede volverse inestable.

3. Precios de oferta de inmuebles en Montevideo

El problema económico seleccionado para la aplicación de las técnicas de aprendizaje estadístico interpretable es la valuación de la oferta de inmuebles en Montevideo, Uruguay. El objetivo es predecir el precio de oferta de los apartamentos en Montevideo mediante la utilización de varios métodos de aprendizaje estadístico y calcular medidas que permitan interpretar los resultados para los distintos métodos utilizados. La aplicación de algunos métodos de aprendizaje y las distintas técnicas de interpretabilidad permiten una mayor comprensión del problema. Entre otras cosas nos permiten identificar si hay o no diferencias en la interpretación según los métodos utilizados y la performance predictiva de ellos.

La base de datos utilizada cuenta con la oferta de apartamentos en la ciudad de Montevideo publicados en la web de *Mercado Libre* durante el período Febrero 2018 a Enero 2019 (Picardo, 2019). Esta base cuenta con 92832 observaciones y 116 variables. A modo de simplificar el problema se selecciona únicamente los apartamentos del barrio Pocitos ya que es el de mayor número de publicaciones en el período analizado (15350 registros). En la Tabla 1 se describen las variables utilizadas en los modelos predictivos elaborada en base a (Picardo, 2019). Todas las variables son las originales menos `ldistancia_playa` y `amenities` que son variables construídas. La variable `ldistancia_playa` contiene información espacial, representada por la distancia que se encuentra el apartamento a la playa que se considera a priori una variable importante en el precio del apartamento. A su vez se incluyen las variables hedónicas que parecen relevante incluir en el estudio y presentan buena calidad en su registro (`ascensores`, `baños`, `condición`, `dormitorios`, `expensas`, `garage` y `lsup_constru`). La variable de respuesta utilizada es el precio por metros cuadrados ofertado (`preciom2`) en dólares del apartamento. La variable `ammenities` cuenta la cantidad de comodidades que tiene el apartamento, las variables seleccionadas para este cálculo fueron 18 de 75 variables. La selección se basó en ajustar un bosque aleatorio para la respuesta y las 75 variables de comodidad como predictoras (tiene o no cochera, tiene o no baño en suite, tiene o no aire acondicionado entre otras) y en base la importancia de las variables (Breiman, 2001a) se seleccionan las 18 más importantes.

Tabla 1: Variables utilizadas en los modelos predictivos

Variable	Descripción	Criterios considerados en la limpieza
<code>preciom2</code>	precio en USD solicitado en metros cuadrados	Se eliminan observaciones con precios como sucesiones de números ("1234", etc.) o números con dígitos repetidos ("1111", "9999", etc.).
<code>ammanities</code>	Cantidad de comodidades seleccionadas entre 18	En base a la medida de importancia se seleccionaron las 18 variables más importantes para predecir la respuesta usando 75 variables de comodidad como ser, terraza, dormitorio en suite entre otros.
<code>ascensores</code>	cantidad	Se redujo a valores entre 0 y 2. Más de 2, en 2.
<code>baños</code>	número de baños	Se redujo a valores entre 1 y 3. Los valores mayores a 3 se asignan a 3
<code>condición</code>	condición del inmueble (nuevo/usado)	Se hizo una asignación de los inmuebles que indicaban "sin especificar". Si la variable antigüedad era menor a 1 año, se asigna como nuevo, de lo contrario se asigna a "usado".
<code>dormitorios</code>	número de dormitorios	Se transformó valores como "monoambiente", "no tiene", etc. a 0; "más de 4", a 5. Se redujo a valores entre 0 (monoambientes) y 3 y se toma como factor. Los mayores a 3 se asignan al valor 3. Pocos valores que quedan con NA, se imputan a 0 y 1 luego de chequear coherencia (cruzado con superficie construida y precio).
<code>expensas</code>	Gastos expensas	Valor numérico que significa pago mensuales en pesos
<code>garage</code>	tiene (si/no)	Se cruza con la variable cocheras (si cochera es mayor a uno, tiene garage). NA se imputa como "No".
<code>ldistancia_playa</code>	log de la mínima distancia a una playa montevideana	Mínima distancia euclidiana entre el punto en que se encuentra el inmueble (<i>centroide</i>) y el punto que identifica a una playa de Montevideo.
<code>lsup_constru</code>	log de superficie en m2	Se realiza una limpieza que imputa NA a valores de tres cifras y más que repiten dígitos y son mayores a 300 ("333", "444", "999", "1111", etc.). Se limpian valores que repiten dígitos menores a "300" ("222", "111") luego de verificarlos. Se imputa NA a los valores menores a 9m2 y mayores a 2000m2. Se cruza con variable <i>sup_total</i> , luego de su limpieza para completar algunos NA.

Para comprender en mejor manera el problema de valuación y la base de datos correspondiente se presenta un breve exploración en ellos en algunas de las variables seleccionadas para los modelos predictivos. En la Figura 1 se muestra el mapa de Montevideo con los datos que se utilizan para el análisis del barrio Pocitos donde el color representa el logaritmo del precio. Es posible observar que los apartamentos más cerca a la costa tienen un rango de precios más alto representados en rojo. Se ven algunos apartamentos que están mal etiquetados como Pocitos que deberían ser sacados para el análisis.

En la Figura 2 se presenta un histograma bidimensional entre el precio en metros cuadrados de los apartamentos en dólares y el logaritmo de la superficie construida en metros cuadrados. La intensidad del color en cada hexágono representa la cantidad de observaciones, los más frecuentes se representan con colores más claros (celeste) y los más oscuros

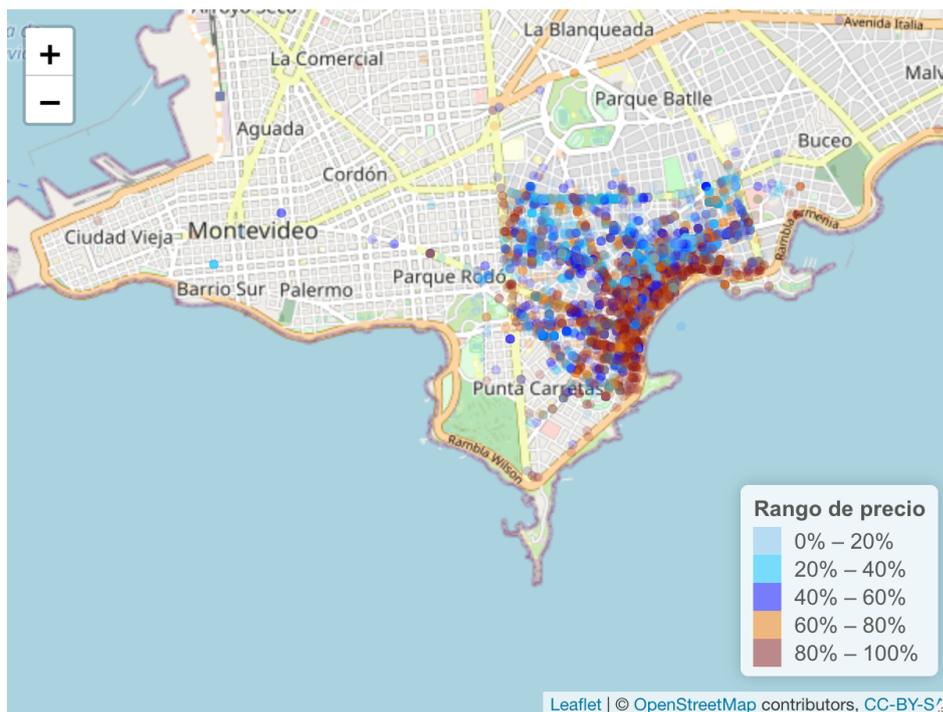


Figura 1: Mapa de Montevideo donde los puntos representan las publicaciones de oferta de apartamentos etiquetadas por el vendedor como Pocitos y el color es el logaritmo del precio en dólares.

representan valores menos frecuentes. Es posible observar una relación lineal, negativa, y moderada entre el precio en metros cuadrados y el logaritmo de superficie construida en metros cuadrados de los apartamentos. La mayoría de las observaciones se encuentran entre 3.5 y 5 (lsup_constru) y entre 2000 y 4000 que corresponden a apartamentos entre 30 y 150 metros cuadrados aproximadamente con un precio que va entre 2000 y 4000 dólares aproximadamente el metro cuadrado. A su vez existen algunos datos atípicos en los valores más bajos de ambas variables.

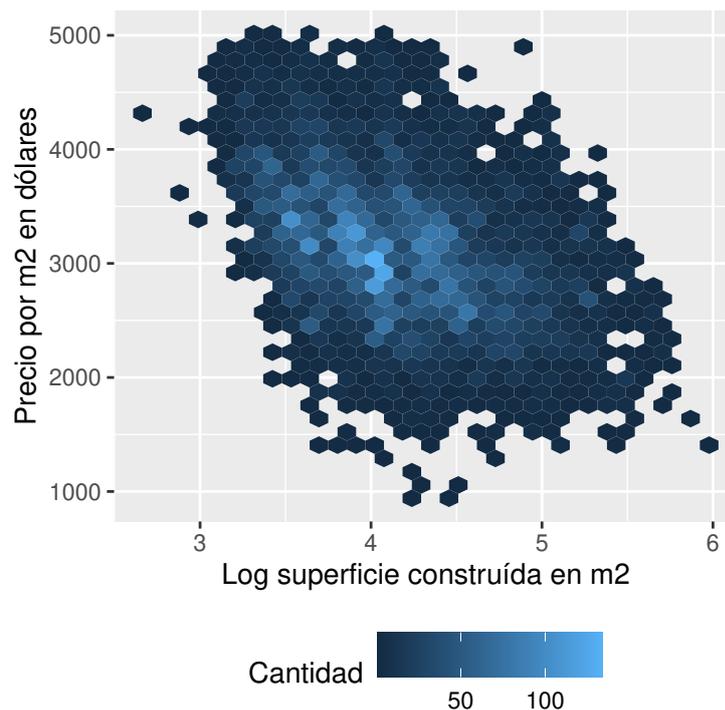


Figura 2: Histograma en dos dimensiones (precio por metro cuadrado en dólares y logaritmo de la superficie construida en metros cuadrados), la intensidad del color representa el número de observaciones en cada hexágono

En la Figura 3 se presenta nuevamente un histograma bidimensional entre la variable de respuesta y el logaritmo de la distancia a la playa en kilómetros. En este caso no se ve una relación clara entre las variables o al menos si hay una relación no es posible considerarla como lineal. Para finalizar esta breve visualización de los datos disponibles, en la Figura 4 se analiza la cantidad de dormitorios y su relación con el precio por metro cuadrado del inmueble. Existe una asociación negativa entre el número de cuartos y el precio del inmueble por metro cuadrado. Hay una diferencia en el valor mediano del precio por metro

cuadrados de 500 dólares entre un apartamento sin dormitorios (aproximadamente 3500 dólares) y uno con 3 dormitorios (aproximadamente 3000 dólares).

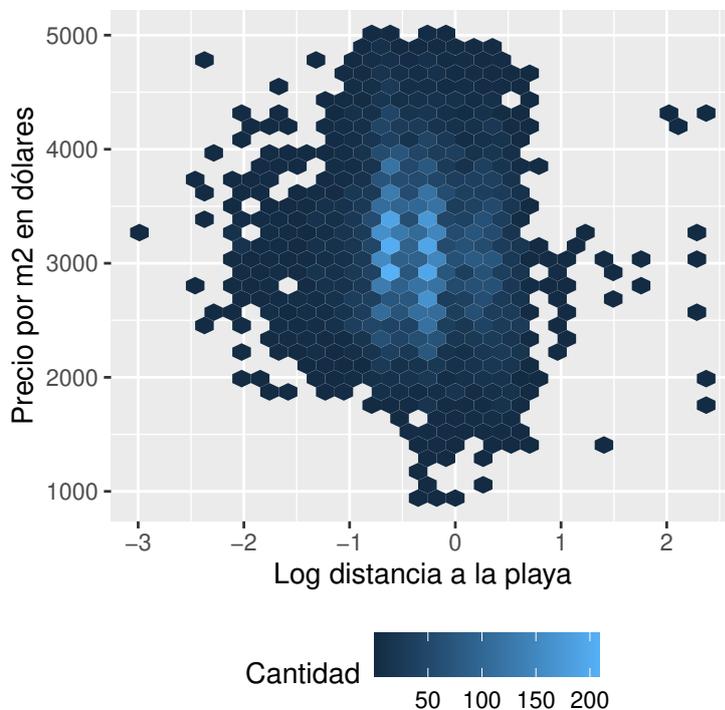


Figura 3: Histograma en dos dimensiones (precio por metro cuadrado en dólares y logaritmo de la distancia a la playa en kilómetros), la intensidad del color representa el número de observaciones en cada hexágono

4. Determinantes del precio ofertado de inmuebles

En esta sección se presentan los resultados en la valuación de los apartamentos en el barrio Pocitos. En 4.1 se presentan los resultados de ajustar varios modelos predictivos, en los cuales se trabaja con un modelo de la forma

$$E(y_i|x_i) = f(x_i),$$

donde y_i representa el precio ofertado por metro cuadrado en dólares de un apartamento en el barrio Pocitos, y $x_i = (x_{i1}, \dots, x_{iq})^\top$ las variables explicativas que se consideran posibles determinantes del precio. Más precisamente, en el trabajo se utilizan cuatro

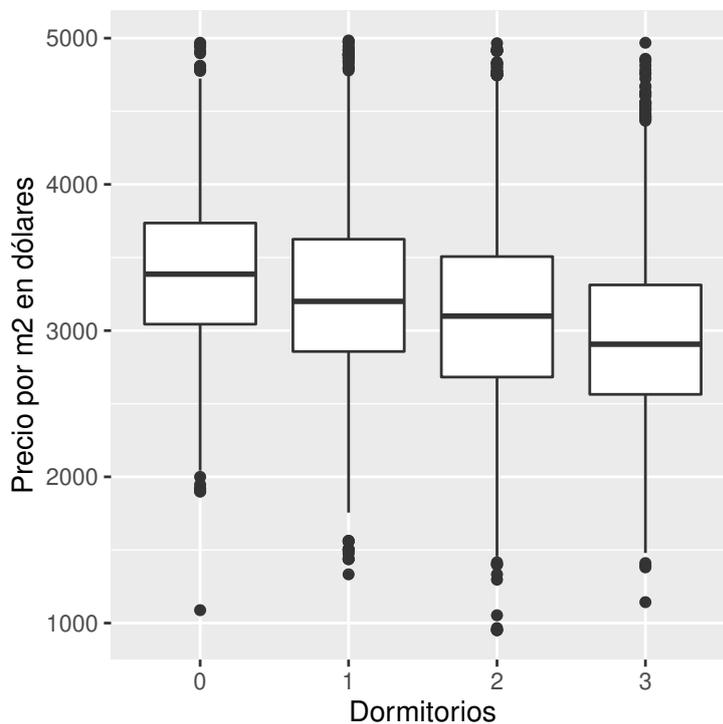


Figura 4: Gráfico de caja del precio por metro cuadrado en dólares y el número de dormitorios .

modelos para estimar $f(\cdot)$: modelo lineal aditivo (modelo hedónico), modelo de árbol de regresión, modelo de bosques aleatorios y modelo de redes neuronales. En lo que resta de la sección, los resultados de cada modelo se comparan y se presentan los resultados de las técnicas de interpretación descritas en el trabajo. Estos resultados permiten caracterizar el efecto de alguna variable de interés sobre los precios de los apartamentos.

4.1. Modelos predictivos

Se utilizan cuatro modelos para estimar la esperanza condicional del precio por metro cuadrado ofertado. En todos los casos las variables explicativas utilizadas son las descritas en la Tabla 1.

Los modelos estadísticos se estiman mediante la biblioteca `caret` (Kuhn, 2021) que es de utilidad para seleccionar los parámetros de *ajuste* con validación cruzada, facilita la comparación del desempeño y permite uniformizar la sintaxis de código. En la Tabla 2 se muestra un código de ejemplo para ajustar un modelo con `caret`. El argumento `method`

permite al usuario elegir entre los distintos modelos: por ejemplo `method = lm` estima un modelo lineal; `method = rpart` estima un árbol de regresión; `method = ranger` estima un bosque aleatorio (en base al paquete `ranger` ya que es una alternativa computacionalmente más rápida para estimar estos modelos) y `method = nnet` estima una red neuronal de una sola capa.

Tabla 2: Código de `caret`

```
train_control <- trainControl(method="cv", number=10)
objeto.md <- train(formula = preciom2 ~ . ,
                  data = training,
                  trControl = train_control,
                  method = 'MODELO')
```

Adicionalmente al tener una sintaxis unificada, el código anterior realiza una selección automática de los principales parámetros de ajuste en cada algoritmo. En el caso del árbol se ajusta el costo-complejidad, para el bosque aleatorio se ajusta la cantidad de predictores seleccionados y la regla utilizada en cada partición, para el caso de la red neuronal son ajustados los parámetros de tamaño y decaimiento.

En la Tabla 3 se observan medidas para evaluar la performance predictiva de los métodos de estimación ajustados. Se presenta, para cada modelo, la raíz cuadrada del error cuadrático medio (RMSE), el R^2 y error absoluto medio (MAE). Las tres medidas se calculan en base al método de validación cruzada con 10 grupos. Adicionalmente se presenta el error porcentual absoluto medio (MAPE) en el conjunto de validación.

Tabla 3: Medidas para evaluar la performance predictiva

	modelo	RMSE	Rsquared	MAE	MAPE
1	lineal	512.37	0.31	402.49	13.59
2	arbol	563.66	0.16	437.91	14.90
3	bosque	365.88	0.65	253.94	8.60
4	neuronal	599.38	0.06	473.06	16.01

Los resultados del comportamiento predictivo muestran al bosque aleatorio como el modelo con mejor desempeño (8.6% de error en la predicción del precio ofertado) y al modelo de red neuronal como el peor (16% de error en promedio). Esto sugiere que tanto el modelo lineal como el árbol de regresión no son suficientemente flexibles para capturar la relación entre la respuesta y los predictores. Es necesario tener cuidado con el mal

resultado en el modelo de red neuronal. Es posible que al utilizar la biblioteca `nnet` redes neuronales más complejas puedan capturar mejor el problema de interés, sin embargo a los efectos de este trabajo es suficiente con los modelos presentados y su desempeño para poder enfocarnos en la interpretación de los resultados. Es claro que modelos con mayor precisión son deseables y modelos que no representan precisamente el problema no son confiables y generan sospechas en el análisis posterior (Breiman, 2001b) que sería necesario considerar.

4.2. Importancia de las variables explicativas

Un primer paso para interpretar los efectos de las variables explicativas sobre la respuesta es medir su importancia, en este sentido variables predictoras más importantes indican que tienen mayor peso en el valor estimado.

Es común que los modelos de aprendizaje estadístico tengan medidas de importancia de las variables específicas del propio modelo, y al mismo tiempo, existen medidas de importancia que pueden ser aplicadas a cualquier modelo. El paquete `caret` permite obtener medidas de importancia de los predictores para cada modelo. En el caso del modelo lineal, utiliza el valor absoluto del estadístico t asociado a cada covariable; para el árbol de regresión la importancia es medida con la disminución que produce cada variable en el error cuadrático medio; en el bosque aleatorio se utiliza el error por permutación de cada predictor y en el caso de la red neuronal se utiliza una combinación lineal de los ponderadores en valor absoluto. En todos los casos, las medidas se escalan para que la variable más importante en cada modelo tenga valor 100, esto facilita la comparación del orden de covariables entre los diferentes modelos estimados.

En la Figura 5 se observa la importancia (mediante su escala) en cada modelo. Se aprecia como la superficie construida es la variable más relevante para todos los modelos menos en el árbol de regresión. En el caso del bosque aleatorio parece que la única variable relevante es justamente la superficie construida. Por otro lado, la red neuronal muestra `expensas` y `amenities` como variables importantes además de la superficie.

4.3. Técnicas interpretables: Efecto promedio de las variables

Las medidas de importancia permiten construir una jerarquía de las variables utilizadas en términos de su poder explicativo en un modelo concreto. Sin embargo, no dicen nada sobre el efecto que cada variable tiene sobre la respuesta. Para caracterizar el efecto de las variables explicativas se utilizan las técnicas desarrolladas en la sección 2, al ser aplicado al problema de predicción de precio ofertado por metro cuadrado en dólares de

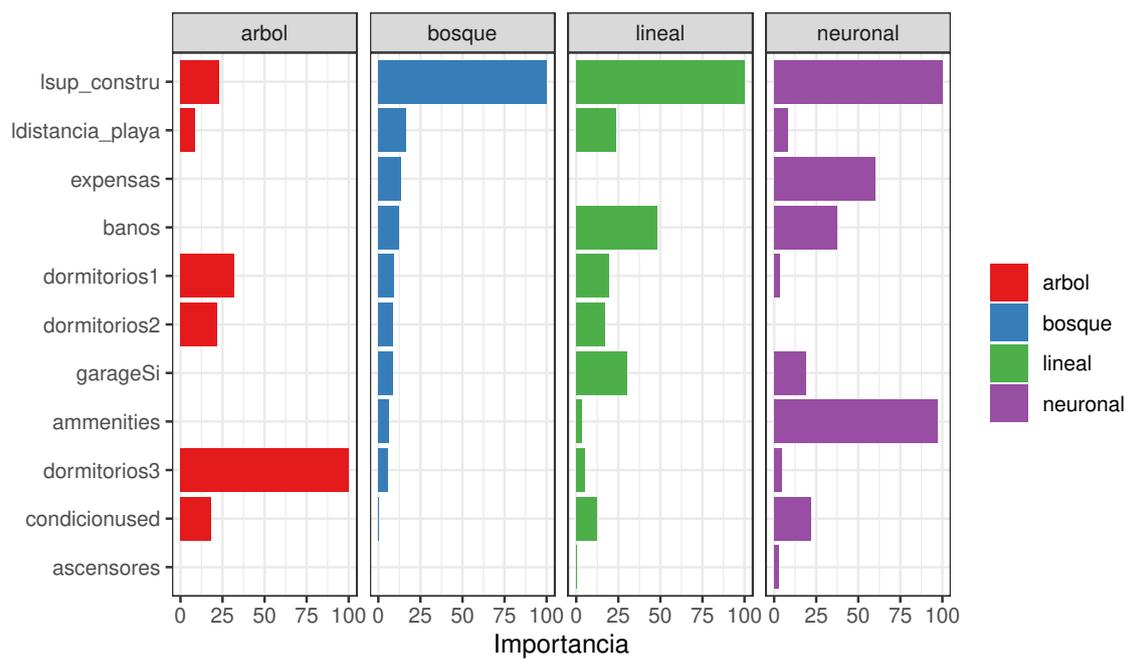


Figura 5: Importancia de variables. Cada panel representa un modelo, y en el eje y se presentan las variables m'as importantes. La longitud de cada barra representa la importancia de la variable como explicativa del precio por metro cuadrado.

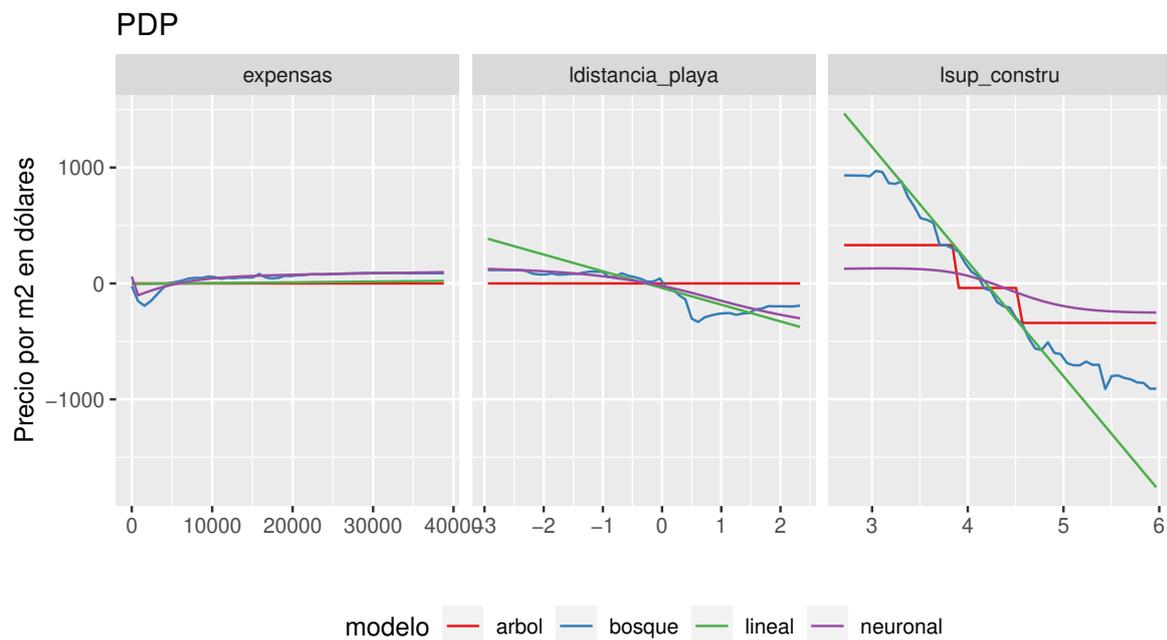


Figura 6: Gráficos de dependencia parcial. Cada panel corresponde a una variable explicativa de interés el color de las líneas representa el método de estimación.

apartamentos en el barrio Pocitos.

En la Figura 6 se muestran los gráficos de dependencia parcial de expensas mensuales (**expensas**), distancia a la playa en logaritmo (**ldistancia_playa**) y superficie construida en logaritmo (**lsup_constru**), para los cuatro modelos estimados. Se pueden apreciar las similitudes y diferencias en el efecto de una misma variable en cada modelo.

La superficie construida es la variable más relevante, esto ya se aprecian con las medidas de importancia pero también se evidencia en que el gráfico de dependencia parcial tiene mayor variabilidad para esta variable. Sin embargo, se puede ver que tanto para la red neuronal como para el árbol de regresión esta variable no es muy relevante, en cambio en el modelo lineal y el bosque aleatorio si se muestra un efecto negativo importante de la superficie construida sobre su precio. En el caso del bosque aleatorio se aprecia que en los extremos, donde no hay muchas observaciones, el gráfico de dependencia parcial no tiene variaciones, a diferencia de lo que ocurre en el modelo lineal que es una recta con pendiente igual al coeficiente estimado.

La Figura 7 presenta los gráficos de la técnica de efectos acumulados para las mismas

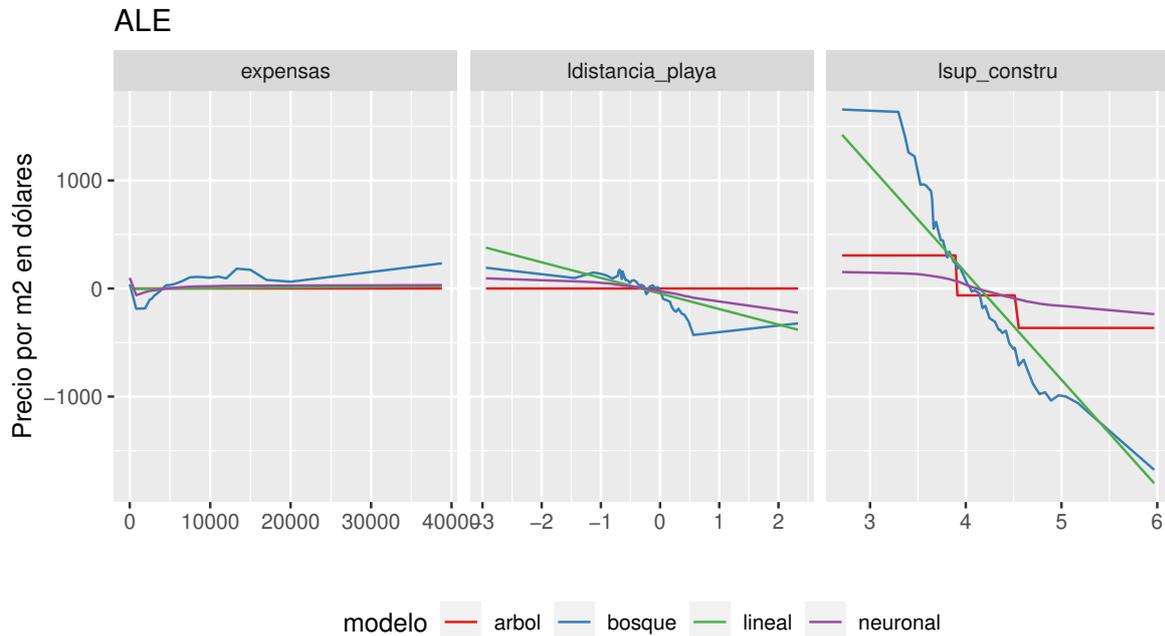


Figura 7: Gráficos efectos locales acumulados. Cada panel corresponde a una variable explicativa de interés el color de las líneas representa el método de estimación.

variables que se presentaron en la Figura 6. En términos globales los patrones que se observan en los efectos acumulados son similares a los que se ven en los gráficos de dependencia parcial. La mayor diferencia entre las distintos ALE-plot y los PD-plot es el efecto de la variable **expensas** en el bosque aleatorio, en ambos casos se observa un efecto negativo sobre el precio ofertado al pasar de expensas nulas a positivas y luego el PD-plot parece no mostrar otro efecto mientras en el ALE-plot sugiere que apartamentos con expensas más grandes tienden a tener precios de oferta más altos. De todos modos este efecto parece ser pequeño.

4.4. Técnicas interpretables: Efectos individuales de las variables

Los gráficos en 4.3 muestran los efectos *promedio* de las variables explicativas en la muestra. Una de las técnicas alternativas expresadas en la sección 2 establece la importancia de visualizar las curvas individuales que componen el gráfico de dependencia parcial, para explorar otros patrones en los efectos que no sea únicamente el valor medio.

La Figura 8 se presentan las gráficas de la técnica esperanza condicional individual para cada modelo de la variable superficie construida, que es la variable más importante en la predicción del precio ofertado por metro cuadrado para todos los modelos menos para el árbol de regresión. Se observa que los resultados difieren sustantivamente para cada modelo utilizado. Por un lado, el árbol de regresión parece completamente insensible a efectos individuales ya que todas las curvas están superpuestas (resultado que sorprende pero puede deberse a que la variable `lsup_constru` no es la más importante en este modelo), similarmente el modelo lineal muestra que todas las curvas individuales tienen la misma pendiente y la única diferencia entre ellas es la ordenada en el origen.

Por otro lado, los paneles que corresponden al bosque aleatorio y a la red neuronal sugieren patrones más interesantes. El modelo de red neuronal muestra que existe un efecto negativo de la superficie construida sobre el precio ofertado por metro cuadrado pero únicamente para un subconjunto de apartamentos, también hay un conjunto de observaciones en las que la variable superficie no parece tener efecto sobre el precio. En el caso del bosque aleatorio, el efecto de superficie parece ser negativo para todos los apartamentos pero con diferencias, para algunos parece haber una región de superficie en la que el efecto es positivo (los más chicos) y para otros el efecto parece ser negativo en todo momento. En cualquier caso, la relevancia de este gráfico es la indicación de posibles patrones a explorar, en este caso sería interesante poder caracterizar los subgrupos de apartamentos con efectos de superficie distintos según alguna de las demás características.

Es importante recordar que tanto los árboles de regresión como modelos lineales son métodos *intrínsecamente* interpretables. En estos casos no es necesario tener un métodos específicos para comprender el efecto de las variables explicativas sobre la respuesta. Los bosques aleatorios y las redes neuronales son modelos con mayor flexibilidad para ajustar relaciones pero al mismo tiempo con menor interpretabilidad, ya que no es directo entender como afectan las variables explicativas a la respuesta. En este sentido, no es sorprendente que un gráfico como la Figura 8 sea más revelador para los métodos que no son intrínsecamente interpretables.

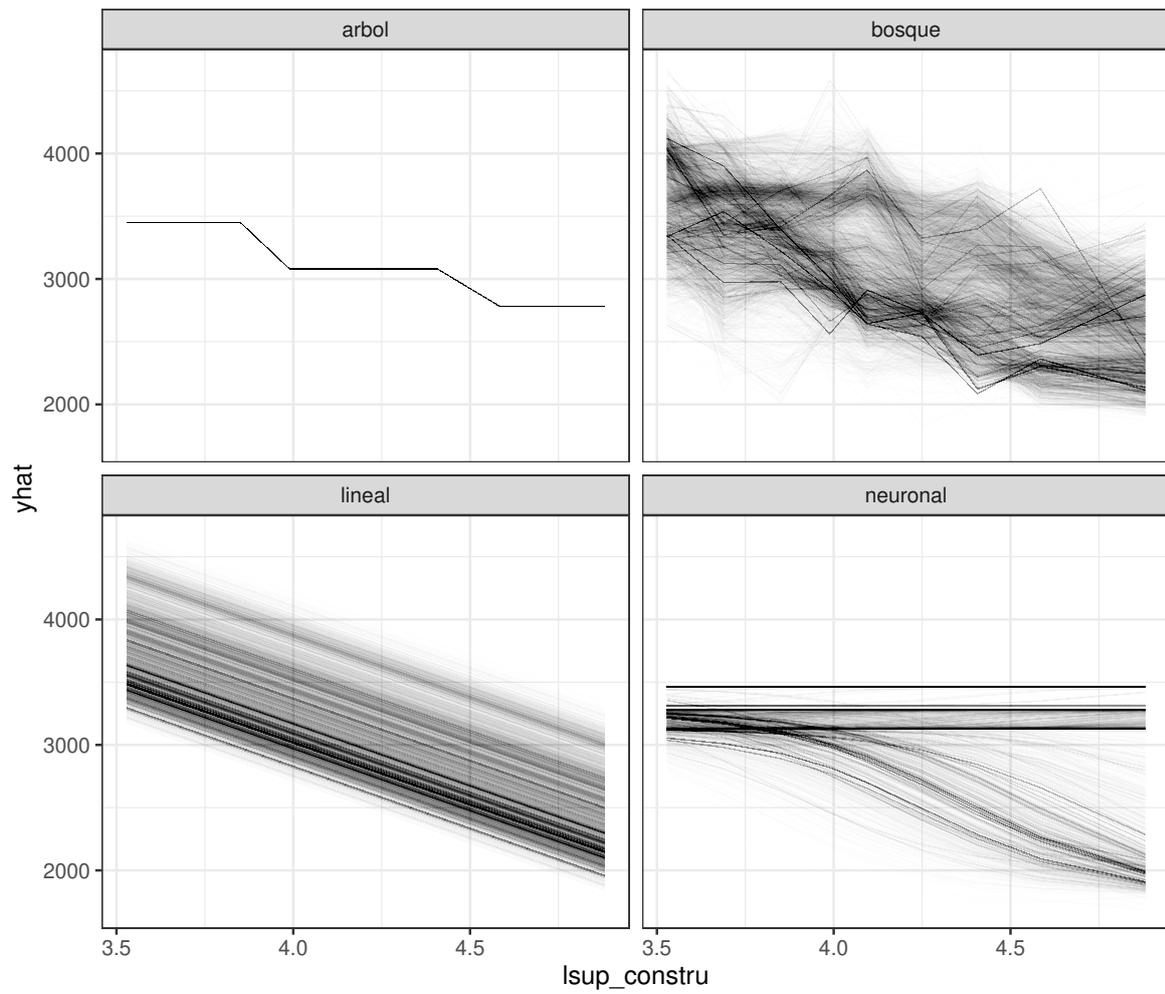


Figura 8: Gráficos de esperanza condicional individual para superficie construída

5. Comentarios finales

En este trabajo se presentaron algunas de las técnicas más utilizadas para la interpretación de los resultados obtenidos en modelos de aprendizaje estadístico. El subconjunto de técnicas mencionadas entran dentro de las que se llaman *model-agnostic* ya que no son específicas para un modelo sino de uso general, que a su vez pueden ser utilizadas también en modelos que son intrínsecamente interpretables (*model-specific*). Adicionalmente se presentó un problema aplicado de interés económico como es la valuación de inmuebles. Se trabajó con información disponible en la web de Mercado Libre de oferta de apartamentos seleccionando como caso de estudio el barrio Pocitos. En este caso se aplicaron cuatro modelos predictivos (modelo lineal, árboles de regresión, bosques aleatorios y redes neuronales) para la variable de respuesta que es el precio de oferta por metro cuadrado de apartamentos en Pocitos.

Una vez estimados los modelos se obtuvieron medidas de importancia de las variables, se calculó el PD-plot y el ALE-plot para los cuatro modelos ajustados y para tres variables seleccionadas (`expensas`, `ldistancia_playa` y `lsup_constru`). Para todos los modelos, menos el árbol, la variable superficie construída (`lsup_constru`) resultó ser la más importante y en algunos casos como en el bosque la única relevante. En cuanto al PD-plot la superficie construída es la que muestra más variabilidad entre las tres variables presentadas, sin embargo para la red neuronal y el árbol no parece tener un efecto importante. Para el bosque y el modelo de regresión lineal presenta un efecto negativo entre la superficie construída y el precio. En el ALE-plot los resultados son similares al PD-plot, la mayor diferencia se da en la variable `expensas` en el bosque aleatorio, en ambos casos se observa un efecto negativo sobre el precio ofertado al pasar `expensas` nulas a positivas y luego el PD-plot parece no mostrar otro efecto mientras en el ALE-plot sugiere que apartamentos con `expensas` más grandes tienden a tener precios de oferta más altos. De todos modos este efecto parece ser pequeño.

Tanto PD-plot como ALE-plot muestran los efectos promedios de las variables explicativas, para tener medidas de los efectos individuales se calculó a su vez los gráficos de la técnica ICE. En este caso solamente se hizo el análisis de la variable superficie construída. Los resultados son bastante distintos para cada uno de los modelos estimados. El árbol de regresión por ejemplo parece ser insensible a esta variable lo cual se explica porque la misma no es importante para este método. Para el modelo lineal todas las curvas presentan pendiente negativa difiriendo únicamente en la ordenada en el origen. Por otro lado, los paneles que corresponden al bosque aleatorio y red neuronal sugieren patrones más interesantes. Para la red neuronal existen patrones diferenciados para dos grupos uno de ellos con efecto negativo de la variable y otro sin efecto. En cambio para el caso del bosque todo el efecto son principalmente negativos.

En varios aspectos de los presentados en el documento de trabajo es necesario profundizar el estudio. Las dos direcciones básicas para exploración de futuras líneas de investigación corresponden a la implementación de mejoras en el problema de valuación y a una mayor profundidad en los métodos para dotar de interpretabilidad a los modelos estimados.

En relación a la aplicación de valuación de inmuebles, sería deseable utilizar datos de toda la ciudad de Montevideo y no solo de un barrio específico. Esto puede ser un desafío para incorporar efectos de grupos (los distintos barrios) en modelos como bosques aleatorios o redes neuronales. También sería muy interesante combinar los datos de precios ofertados con datos de transacciones efectivamente realizadas, aquí hay un problema estadístico de enlace de registros (falta de datos). Por otro lado, el ajuste de los distintos modelos de aprendizaje estadístico puede ser mejorado utilizando algoritmos automáticos como los implementados en la biblioteca `h2o` que es más eficiente que `caret` y también permite automatizar varias etapas necesarias para el modelado teniendo a su vez una amplia variedad de métodos implementados. Finalmente, en cuanto a los métodos de interpretabilidad se pueden buscar modificaciones a las propuestas planteadas, por ejemplo trabajar con una versión continua del ALE, o relacionar las curvas individuales con alguna medida que indique su distancia al centro de los datos de entrenamiento.

Referencias Bibliográficas

- Apley, D. W. y Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):1059–1086.
- Azodi, C. B., Tang, J., y Shiu, S.-H. (2020). Opening the black box: Interpretable machine learning for geneticists. *Trends in genetics*, 36(6):442–455.
- Breiman, L. (2001a). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.
- Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.
- Case, B., Clapp, J., Dubin, R., y Rodriguez, M. (2004). Modeling spatial and temporal house price patterns: A comparison of four models. *The Journal of Real Estate Finance and Economics*, 29(2):167–191.
- Doshi-Velez, F. y Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
- Du, M., Liu, N., y Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, 63(1):68–77.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232.
- Goldstein, A., Kapelner, A., Bleich, J., y Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1):44–65.
- Goyeneche, J. J., Moreno, L., y Scavino, M. (2017). Predicción del valor de un inmueble mediante técnicas agregativas. *Serie DT (17/1)*.
- Grimmer, J., Roberts, M. E., y Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, 24:395–419.
- Kiel, K. A. y Zabel, J. E. (2008). Location, location, location: The 3l approach to house price determination. *Journal of Housing Economics*, 17(2):175–190.
- Kuhn, M. (2021). *caret: Classification and Regression Training*. R package version 6.0-90.

- Limsombunchai, V. (2004). House price prediction: hedonic price model vs. artificial neural network. En *New Zealand agricultural and resource economics society conference*, pp. 25–26.
- Linardatos, P., Papastefanopoulos, V., y Kotsiantis, S. (2021). Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1):18.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38.
- Molnar, C., Casalicchio, G., y Bischl, B. (2020). Interpretable machine learning—a brief history, state-of-the-art and challenges. En *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 417–431. Springer.
- Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., y Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080.
- Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, 32(3):289–320.
- Park, B. y Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert systems with applications*, 42(6):2928–2934.
- Picardo, P. (2019). Predicción de precios de la vivienda aprendizaje estadístico con datos de ofertas y transacciones para montevideo. *Tesis de Maestría en Economía, FCEA-UDELAR*.
- Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., y Zhong, C. (2021). Interpretable machine learning: Fundamental principles and 10 grand challenges. *arXiv preprint arXiv:2103.11251*.
- Sirmans, S., Macpherson, D., y Zietz, E. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, 13(1):1–44.
- Yoo, S., Im, J., y Wagner, J. E. (2012). Variable selection for hedonic model using machine learning approaches: A case study in onondaga county, ny. *Landscape and Urban Planning*, 107(3):293–306.

Instituto de Estadística

Serie Documentos de Trabajo



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN

IESTA

INSTITUTO
DE ESTADÍSTICA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Gonzalo Ramirez 1926, Piso 1, Oficina 23 - C.P. 11200 -
Montevideo, Uruguay
Teléfono: (598) 2410 2564
<https://iesta.fcea.udelar.edu.uy/>
Área Publicaciones

Febrero, 2022

DT (01/22)

Instituto de Estadística

Serie Documentos de Trabajo



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN

IESTA 80 1942-2022 INSTITUTO
DE ESTADÍSTICA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Gonzalo Ramirez 1926, Piso 1, Oficina 23 - C.P. 11200 -
Montevideo, Uruguay
Teléfono: (598) 2410 2564
<https://iesta.fcea.udelar.edu.uy/>
Área Publicaciones

Febrero, 2022

Nº 1/22