



CIENCIA DE DATOS, FINANZAS LOCALES Y GESTIÓN TERRITORIAL

Modelos valuatorios, simulaciones y código abierto para la gestión fiscal y el financiamiento urbano

HUÁSCAR EGUINO
JUAN PABLO CARRANZA



CIENCIA **DE DATOS,** FINANZAS **LOCALES** Y GESTIÓN **TERRITORIAL**

Modelos valuatorios, simulaciones y código abierto
para la gestión fiscal y el financiamiento urbano

HUÁSCAR EGUINO
JUAN PABLO CARRANZA

**Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo**

Eguino, Huáscar.

Ciencia de datos, finanzas locales y gestión territorial: modelos de valuación, simulaciones y código abierto para la gestión fiscal y el financiamiento urbano / Huáscar Eguino, Juan Pablo Carranza.

p. cm. — (Monografía del BID ; 1297)

Incluye referencias bibliográficas.

1. Real property and taxation-Latin America. 2. Real property and taxation-Caribbean Area. 3. Cadastres-Latin America. 4. Cadastres-Caribbean Area. 5. City planning-Latin America. 6. City planning-Caribbean Area.

7. Artificial intelligence-Latin America. 8. Artificial intelligence-Caribbean Area. I. Carranza, Juan Pablo.

II. Banco Interamericano de Desarrollo. División de Gestión Fiscal. III. Título. IV. Serie.

IDB-MG-1297

Clasificaciones JEL: C21, C63, H71, R52, R31

Palabras clave: valuación inmobiliaria, finanzas locales, catastro fiscal, administración tributaria, planificación urbana, captura de valor, inteligencia artificial, impuesto a la propiedad

Copyright © 2025 Banco Interamericano de Desarrollo (BID). Esta obra se encuentra sujeta a una licencia Creative Commons CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>). Se deberá cumplir los términos y condiciones señalados en el enlace URL y otorgar el respectivo reconocimiento al BID.

En alcance a la sección 8 de la licencia indicada, cualquier mediación relacionada con disputas que surjan bajo esta licencia será llevada a cabo de conformidad con el Reglamento de Mediación de la OMPI. Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la Comisión de las Naciones Unidas para el Derecho Mercantil (CNUDMI). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID no están autorizados por esta licencia y requieren de un acuerdo de licencia adicional.

Nótese que el enlace URL incluye términos y condiciones que forman parte integral de esta licencia.

Las opiniones expresadas en esta obra son exclusivamente de los autores y no necesariamente reflejan el punto de vista del BID, de su Directorio Ejecutivo ni de los países que representa.



Banco Interamericano de Desarrollo
1300 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org

El Sector de Instituciones para el Desarrollo fue responsable de la producción de la publicación.

Colaboradores externos:

Coordinación de la producción editorial: Sarah Schineller (A&S Information Partners, LLC)

Revisión editorial: Santiago Arévalo

Diagramación: Sara María Ochoa Botero

ÍNDICE



AGRADECIMIENTOS 8

SIGLAS Y ACRÓNIMOS.....9

SOBRE LOS AUTORES.....10

INTRODUCCIÓN 11



CAPÍTULO 1.

VALUACIÓN INMOBILIARIA Y GESTIÓN DE LOS GOBIERNOS

SUBNACIONALES.....16

1.1. Impuesto a la propiedad inmobiliaria.....17

1.2. Financiamiento urbano basado en instrumentos de captura de valor21

1.3. Planificación y gestión urbana.....22



CAPÍTULO 2.

EVOLUCIÓN DE LAS TÉCNICAS DE VALUACIÓN

MASIVA AUTOMATIZADA Y OPORTUNIDADES

DE LA INTELIGENCIA ARTIFICIAL 23

2.1. ¿Qué es un proceso de valuación masiva automatizada?24

2.2. Nivel de aplicación del modelo de valuación automatizada26

2.3. Situación en América Latina y el Caribe27

2.4. Diferencias entre valuaciones masivas y tasaciones individuales.....28

2.5. Evolución de las técnicas aplicadas en valuaciones masivas automatizadas....29

2.6. Ventajas y desventajas de las diferentes técnicas para realizar valuaciones
masivas automatizadas.....34

2.7. ¿Cómo aprende un algoritmo?36



CAPÍTULO 3.

VALUACIÓN INMOBILIARIA COMO INSTRUMENTO DE GESTIÓN PÚBLICA.....

3.1. Administración tributaria del impuesto a la propiedad.....	41
3.2. Instrumentos de financiamiento urbano basados en los impuestos y las contribuciones por mejoras.....	45
3.2.1. Financiamiento basado en el incremento del impuesto a la propiedad inmobiliaria.....	46
3.2.2. Contribuciones por mejoras.....	49
3.3. Planificación y gestión urbana.....	51
3.3.1. Impacto de los cambios de las normas urbanísticas en el potencial constructivo.....	51
3.3.2. Evaluación de proyectos de desarrollo urbano.....	53
3.3.3. Impacto de los riesgos climáticos en el valor de las propiedades.....	55



CAPÍTULO 4.

CAJA DE HERRAMIENTAS PARA REALIZAR AVALÚOS MASIVOS.....

4.1. Cálculo de variables de distancia: Función calcular_dist().....	60
4.2. Cálculo de variables de entorno: Función calcular_entorno().....	64
4.3. Cálculo de variables ráster: Función calcular_raster().....	67
4.4. Homogeneización de la muestra: Función calcular_vut().....	70
4.5. Identificación de observaciones atípicas: Función eliminar_outliers().....	75
4.6. Entrenamiento de modelos: Función entrenar_modelo().....	76
4.6.1. Caso de aplicación: Estimación del valor de los inmuebles en la ciudad de Cali, Colombia.....	79
4.7. Valuación de intervenciones urbanas: Función simular_escenario().....	89
4.7.1. Caso de aplicación: Valuación del impacto de una nueva vía principal en Medellín, Colombia.....	89
4.7.2. Caso de aplicación: Valuación del impacto de un nuevo espacio verde en Belo Horizonte, Brasil.....	94
4.7.3. Caso de aplicación: Estimación de las externalidades generadas por la disminución del riesgo de incendio en Córdoba, Argentina.....	98



CAPÍTULO 5.

LINEAMIENTOS DE ACCIÓN Y CONSIDERACIONES FINALES.....

REFERENCIAS.....

106



LISTA DE CUADROS

Cuadro R1.1.1. Aumento de ingresos por eficiencia administrativa.....	19	Cuadro R3.3.1. Proyectos financiados con contribuciones por valorización en Bogotá.....	50
Cuadro 2.1. Las técnicas de <i>machine learning</i> en los modelos de valuación masiva	35	Cuadro 4.1. Estructura del capítulo 4: procesos, funciones y casos de aplicación	59



LISTA DE GRÁFICOS

Gráfico 1.1. Ecuación de ingresos del impuesto a la propiedad.....	18	Gráfico 3.1. Métodos y técnicas de valuación masiva de inmuebles	43
Gráfico 2.1. Proceso de implementación de un modelo de valuación automatizada	26	Gráfico 3.2. Financiamiento por incremento del impuesto a la propiedad	46
Gráfico 2.2. Semivariograma teórico y empírico para una muestra de inmuebles en venta en Cali, Colombia	31	Gráfico 4.1. Propuesta de proceso para realizar valuaciones masivas.....	58
Gráfico 2.3. Evolución de las técnicas utilizadas en valuaciones masivas automatizadas, período 1960-2015	32	Gráfico 4.2. Comparación entre el valor por metro cuadrado original y el homogeneizado en la ciudad de Córdoba, Argentina	74
Gráfico 2.4. Árbol de clasificación y regresión para una muestra de ejemplo en Belo Horizonte, Brasil	38	Gráfico 4.3. Flujo del proceso de entrenamiento de algoritmos.....	78



LISTA DE MAPAS

Mapa R3.2.1. Área de incidencia del proyecto de tranvía Carrera 80 en Medellín, Colombia48	Mapa 4.9. Distribución espacial de la muestra de inmuebles según el valor del metro cuadrado en Cali, Colombia (en US\$ por m ²)..... 79
Mapa 4.1. Vías primarias de la ciudad de La Plata, Argentina..... 62	Mapa 4.10. Definición del área de estudio: Polígono administrativo de Cali, Colombia.....80
Mapa 4.2. Resultado de la función <code>calcular_dist()</code> para vías secundarias en La Plata, Argentina..... 63	Mapa 4.11. Variables independientes utilizadas en el entrenamiento del modelo para estimar el valor de los inmuebles en la ciudad de Cali, Colombia.....84
Mapa 4.3. Resultado de la función <code>calcular_entorno()</code> para estimar el porcentaje cubierto por espacios verdes en Belo Horizonte, Brasil 65	Mapa 4.12. Distribución espacial del valor de los inmuebles en Cali, Colombia (en US\$ por m ²)86
Mapa 4.4. Resultado de la función <code>calcular_entorno()</code> para estimar la cantidad de espacios verdes en Belo Horizonte, Brasil..... 66	Mapa 4.13. Distribución espacial el valor por metro cuadrado de los inmuebles a nivel de barrios en Cali, Colombia (en US\$ por m ²)87
Mapa 4.5. Resultado de la función <code>calcular_raster()</code> para estimar el porcentaje edificado por píxel en Medellín, Colombia..... 68	Mapa 4.14. Distribución espacial de la muestra de valores de suelo en la ciudad de Medellín, Colombia (en US\$ por m ²).....90
Mapa 4.6. Resultado de la función <code>calcular_raster()</code> para estimar el porcentaje edificado en el entorno en Medellín, Colombia..... 69	Mapa 4.15. Estimación el valor de la tierra en el área urbana de Medellín, Colombia (en US\$ por m ²)..... 91
Mapa 4.7. Cálculo del valor homogéneo: Distribución espacial del valor por metro cuadrado en la ciudad de Córdoba, Argentina (en US\$ por m ²)..... 72	Mapa 4.16. Simulación de una nueva vía primaria en Medellín, Colombia..... 92
Mapa 4.8. Distribución en función del valor del metro cuadrado homogeneizado en la ciudad de Córdoba, Argentina (en US\$ por m ²)..... 73	Mapa 4.17. Efecto esperado de una nueva vía primaria en Medellín, Colombia..... 93
	Mapa 4.18. Estimación del valor del suelo en Belo Horizonte, Brasil (en US\$ por m ²)..... 95

Mapa 4.19. Simulación de un nuevo espacio verde en Belo Horizonte, Brasil.....96

Mapa 4.20. Efecto esperado de un nuevo espacio verde en Belo Horizonte, Brasil (en US\$ por m²).....97

Mapa 4.21. Estimación del valor de la tierra en el área metropolitana de la ciudad de Córdoba, Argentina (en US\$ por m²).....99

Mapa 4.22. Simulación de la eliminación de un polígono para reducir el riesgo de incendio en la ciudad de Córdoba, Argentina100

Mapa 4.23. Efecto esperado de la reducción del riesgo de incendio en Córdoba, Argentina (en US\$ por m²).....101



LISTA DE **RECUADROS**

Recuadro 1.1. Impacto de la mejora de la eficiencia administrativa en la recaudación19

Recuadro 1.2. Uso de inteligencia artificial para la actualización de los valores catastrales.....20

Recuadro 3.1. Ejemplo de aplicación de valuación masiva de inmuebles en la administración del impuesto a la propiedad: Prefeitura municipal de Fortaleza, Brasil.....44

Recuadro 3.2. Ejemplo de financiamiento basado en el incremento del impuesto a la propiedad: Proyecto de tranvía Carrera 80 en Medellín, Colombia48

Recuadro 3.3. Ejemplo de aplicación de contribuciones por mejora en Bogotá, Colombia.....50

Recuadro 3.4. Ejemplo de uso de modelos de valuación automatizada en el análisis de cambios en normas urbanísticas de Curitiba, Brasil.....52

Recuadro 3.5. Ejemplo de evaluación de proyectos de desarrollo urbano con base en la valorización inmobiliaria en Bogotá, Colombia54

Recuadro 3.6. Ejemplo de análisis de los impactos de los riesgos climáticos sobre el valor inmobiliario: Estimación del efecto de la reducción del riesgo de incendio en Córdoba, Argentina.....56

Recuadro 4.1. Nociones básicas sobre representaciones cartográficas60

Recuadro 4.2. Componentes de una función en R.....60

AGRADECIMIENTOS

Los autores desean expresar su gratitud a Diego Erba, Axel Radics y Andrea López Luzuriaga por la revisión del manuscrito y el aporte de sugerencias, que han elevado la calidad editorial de este documento.

Además, manifiestan su agradecimiento a Andrés Abarca Jiménez, por los comentarios realizados en torno de la gestión de riesgos urbanos, y a Mario Piumetto, por la gestión del Mapa de Valores de América Latina, que pone a disposición datos abiertos sobre el valor de mercado del suelo en toda la región desde hace casi 10 años.

Los autores agradecen también la colaboración de la Alcaldía de Cali en la provisión de datos de su observatorio inmobiliario; las gestiones realizadas por Mario Sanginés y Silvia Salamanca para obtener esta información, y el trabajo de la Infraestructura de Datos Espaciales de la Provincia de Córdoba (IDECOR) en la publicación de datos abiertos y la ejecución de estudios territoriales, como el Mapa de Riesgo de Incendios utilizado en esta publicación.

SIGLAS Y ACRÓNIMOS

ACB	análisis costo-beneficio
AVM	modelo de valuación automatizada (<i>automated valuation model</i>)
ALC	América Latina y el Caribe
API	interfaz de programación de aplicaciones (<i>application programming interface</i>)
BID	Banco Interamericano de Desarrollo
BT	base tributaria
CART	árbol de clasificación y regresión (<i>classification and regression tree</i>)
IA	inteligencia artificial
MAPE	error relativo promedio en valor absoluto (<i>mean absolute percentage error</i>)
ML	aprendizaje automático (<i>machine learning</i>)
OCDE	Organización para la Cooperación y el Desarrollo Económicos
OSM	OpenStreetMap
TC	tasa de cobertura
TI	tasa impositiva
TIF	financiamiento por incremento de impuestos (<i>tax increment financing</i>)
TOD	proyectos de desarrollo basados en el transporte (<i>transport oriented development</i>)
TR	tasa de recaudación
TV	tasa de valoración

SOBRE **LOS AUTORES**

HUÁSCAR EGUINO

Economista especializado en política y gestión fiscal, finanzas subnacionales, cambio climático y desarrollo territorial. Cuenta con más de 25 años de experiencia en la provisión de asistencia técnica y el asesoramiento a varios gobiernos nacionales y subnacionales en la mayoría de los países de América Latina y el Caribe. Fue coordinador de la agenda del Banco Interamericano de Desarrollo (BID) en temas de política fiscal y acción climática, descentralización fiscal y gestión subnacional, y la Iniciativa de Ciudades Sostenibles, entre otras acciones. Actualmente, trabaja como consultor de las divisiones de Política Fiscal y de Cambio Climático del BID, asesor de una serie de gobiernos de la región en temas de política fiscal verde y en la elaboración de productos de conocimiento. Tiene una maestría en Desarrollo Local y Regional, en el Instituto de Estudios Sociales (Erasmus University Rotterdam, Países Bajos); estudios de posgrado en el Instituto Tecnológico de Massachusetts (MIT) y la Universidad de Harvard, y certificaciones en política fiscal y finanzas públicas del Fondo Monetario Internacional. Ha publicado más de 25 documentos con el BID y estudios en prestigiosas revistas especializadas.

JUAN PABLO CARRANZA

Es economista, doctor en Política Pública y magíster en Administración Pública. Coordina el Programa de Ciencia de Datos de la Universidad Nacional de Córdoba, en Argentina, donde también se desempeña como profesor e investigador del Instituto de Investigación y Formación en Administración Pública (IIFAP) de la Facultad de Ciencias Sociales. Ha coordinado el equipo de ciencia de datos en Infraestructura de Datos Espaciales de la Provincia de Córdoba (IDECOR), organismo a cargo del diseño y la implementación del proceso metodológico del avalúo masivo realizado en la provincia de Córdoba (Argentina) entre 2017 y 2023. En el marco de ese proceso se implementó el uso de algoritmos de aprendizaje automático (ML, por sus siglas en inglés) para estimar el valor catastral y fiscal de la tierra urbana y rural a niveles de mercado. En esa línea, ha realizado consultorías sobre innovación en las valuaciones inmobiliarias y tecnologías aplicadas a la gestión urbana en diferentes países de la región, como Bolivia, Brasil, Colombia y Perú, además de Argentina, para instituciones como el Programa de las Naciones Unidas para el Desarrollo (PNUD), el Banco de Desarrollo de América Latina y el Caribe (CAF) y el BID. Es autor de numerosos trabajos científicos sobre temas relacionados con las valuaciones y la economía urbana y regional, y ha dictado una importante cantidad de conferencias en eventos internacionales vinculados con esta temática.



INTRODUCCIÓN

El trabajo de investigación propuesto subraya el modo en que la ciencia de datos y los modelos de valuación automatizada (AVM, por sus siglas en inglés) masiva, apoyadas por la inteligencia artificial (IA), constituyen herramientas fundamentales para modernizar la gestión fiscal y urbana en América Latina y el Caribe (ALC).¹ Este documento trasciende el ámbito técnico usual de las valuaciones inmobiliarias, frecuentemente acotado a las finanzas públicas locales, para resaltar su importancia en el diseño de instrumentos de financiamiento urbano, la optimización de la planificación y gestión territorial, y la resiliencia urbana como respuesta a eventos climáticos extremos.

A partir de reconocer un déficit de capacidades institucionales que opera como restricción para la implementación de los procesos de valuación masiva, se propone una serie de herramientas que apuntan a facilitar su realización y se incluyen los respectivos códigos abiertos. Estas herramientas sistematizan un conjunto complejo de procesos e instrumentos que van desde la generación de información territorial hasta el entrenamiento de algoritmos de aprendizaje automático (ML, por sus siglas en inglés) para la ejecución de avalúos masivos, tanto de la tierra como de las propiedades edificadas.²

El documento contiene un capítulo que funciona como guía paso a paso y lleva al lector por diferentes ejemplos de aplicación de las herramientas propuestas, para determinar el valor de los inmuebles o para identificar el valor público generado por intervenciones estatales, como nuevas vías de movilidad, nuevos espacios verdes o infraestructura destinada a atenuar el impacto de eventos climáticos extremos.

La valuación inmobiliaria como pilar de la gestión pública subnacional

La relevancia de disponer de valuaciones inmobiliarias actualizadas y próximas a los valores de mercado reside en su incidencia directa sobre tres ámbitos clave de la gestión de los gobiernos subnacionales: la administración del impuesto a la propiedad, el financiamiento urbano basado en la captura de valor y la planificación y gestión urbana. A pesar de este potencial, en ALC persisten desafíos significativos, como los rezagos catastrales y la subutilización del valor inmobiliario en las políticas públicas. En este contexto, el uso de técnicas avanzadas, como los AVM y la IA, representa una oportunidad para superar estas limitaciones, al reducir costos y tiempos de valuación, y mejorar la precisión de los resultados.

-
- 1** En este documento se usan los términos valuación y tasación inmobiliaria de manera indistinta. En ambos casos, se trata de los procesos empleados para determinar el valor de las propiedades inmuebles.
 - 2** El aprendizaje automático o *machine learning* es un subconjunto de lo que hoy se conoce como inteligencia artificial, que permite a un sistema aprender automáticamente y mejorar su desempeño a partir de la experiencia (datos).

Fortalecimiento de la administración tributaria del impuesto a la propiedad

La investigación aplicada demuestra que el impuesto a la propiedad inmobiliaria posee un considerable potencial de movilización de ingresos en la región que no se aprovecha plenamente. La ecuación de ingresos ilustra que el desempeño de este impuesto depende de variables de política tributaria (determinación de la base tributaria y la tasa impositiva) y de variables de administración tributaria (tasa de cobertura, tasa de valuación y tasa de recaudación). En este sentido, se sostiene que las mejoras en las variables de administración tributaria, que dependen de la capacidad institucional y las herramientas de gestión locales, pueden generar incrementos sustanciales en los ingresos.

Aquí es donde la IA y las nuevas técnicas de análisis de datos cobran particular importancia, puesto que contribuyen a reducir la brecha entre los valores de mercado y los valores catastrales de las propiedades. Experiencias en ciudades como Córdoba (Argentina) y Fortaleza (Brasil) ponen en evidencia este potencial. En Córdoba, la aplicación de algoritmos, como Random Forest, permitió determinar que el valor catastral de los inmuebles representaba solo el 60% del valor de mercado y, de esta manera, proyectar un incremento del 45% en el impuesto a los bienes inmuebles con un ajuste catastral. De forma similar, en Fortaleza, el AVM indicó que el valor catastral era apenas el 40% del valor de mercado, con lo cual la utilización de los nuevos valores implicaba un incremento superior al 100% en la recaudación. Estos casos resaltan el impacto directo y cuantificable que el uso de los AVM y la IA puede tener en la optimización fiscal local.

Instrumentos de financiamiento urbano basados en la captura de valor

La investigación también valida la relación positiva entre las intervenciones públicas y el aumento del valor de las propiedades, la cual proporciona el fundamento para emplear instrumentos de captura de valor o plusvalía. Estos mecanismos, como las contribuciones por mejoras, los cargos por derechos de construcción, los esquemas de reajuste de terrenos y el financiamiento por incremento de impuestos (TIF, por sus siglas en inglés), requieren conocer el valor actual de los inmuebles y estimar el incremento de su valor posintervención. La capacidad de los AVM para realizar estas estimaciones de forma objetiva y con mayor precisión facilita la aplicación de estos instrumentos y la generación de recursos adicionales para financiar infraestructura urbana y otros proyectos. El uso de estos instrumentos de financiamiento se ilustra con referencias a las experiencias de Medellín y Bogotá, en Colombia.

Usos de la valuación inmobiliaria en la planificación y gestión urbana

De igual forma, disponer de información actualizada y desagregada del valor de los inmuebles tiene un valor relevante para establecer una gestión y planificación urbana más transparente, eficiente y oportuna. En esta materia, la presente investigación muestra cómo el AVM permite introducir cambios en los siguientes aspectos:

- **La evaluación de intervenciones y proyectos urbanos.** Los AVM posibilitan simular el impacto de nuevos desarrollos urbanos en el valor de los inmuebles, lo cual contribuye a analizar la viabilidad fiscal y financiera de los proyectos de infraestructura y seleccionar los que proporcionan mayor rentabilidad. Ejemplos como la valoración del impacto de una nueva vía principal en Medellín o de un nuevo espacio verde en Belo Horizonte permiten comprobar esta aplicación.
- **El fortalecimiento de las normativas de uso de suelo.** Con los modelos de tasación masiva es posible identificar áreas subutilizadas y simular el impacto económico de cambios en las normas de uso de suelo y construcción, lo cual es crucial para elaborar políticas de densificación y optimización del uso del suelo. El empleo de estas técnicas se ilustra mediante referencias a la experiencia de Curitiba, Brasil.

Gestión de riesgos climáticos y ambientales

Otra aplicación novedosa de la IA y las nuevas técnicas de análisis de datos es la capacidad de simular o calcular las pérdidas en el valor de los inmuebles como resultado de eventos climáticos o ambientales extremos, lo cual es vital para la resiliencia urbana. Un caso de aplicación en Córdoba, Argentina, demuestra cómo la eliminación simulada del riesgo de incendio puede generar un aumento significativo del precio de la tierra en el área intervenida, lo que, a su vez, se puede usar como insumo para la evaluación costo-beneficio de las inversiones destinadas a reducir los riesgos de los eventos climáticos extremos.

Evolución de las técnicas de valuación masiva y oportunidades de la inteligencia artificial

La presente publicación documenta la evolución metodológica de las técnicas de valuación masiva, desde regresiones lineales hasta algoritmos de ML, como Random Forest y Gradient Boosting. Estos algoritmos permiten procesar grandes volúmenes de datos con alta precisión, incluso en contextos de información incompleta o dispersa, de modo que se reducen costos y tiempos de valuación. Merece destacarse que los AVM con ML tienen un desempeño progresivamente más eficiente a medida que se integra una mayor cantidad de datos, lo cual contribuye a una actualización costo-eficiente de los catastros.

Asimismo, la investigación aborda las limitaciones de estas técnicas, como la “caja negra” de algunos algoritmos de ML, que dificulta la reconstrucción del proceso de generación de valor, y los desafíos relacionados con la falta de capacidades técnicas locales. Se enfatiza que la implementación tecnológica debe ir acompañada de procesos de formación, institucionalización y construcción de confianza en la ciudadanía, así como de políticas de transparencia y datos abiertos que les aporten legitimidad a estos procesos.

Lineamientos de acción y consideraciones finales

La guía concluye con una serie de lineamientos de acción estratégicos para los gobiernos subnacionales, que derivan directamente de la investigación y las experiencias prácticas. Estos lineamientos incluyen:

- **Invertir en datos e infraestructura tecnológica:** Fortalecer observatorios inmobiliarios, sistemas de información geográfica y catastros digitales.
- **Robustecer las capacidades técnicas:** Capacitar al personal en estadística, programación, IA y gestión de datos espaciales.
- **Modernizar los marcos normativos:** Adaptar las regulaciones para permitir la adopción de nuevas tecnologías en la gestión tributaria/catastral sin comprometer la transparencia.
- **Promover la transparencia y la comunicación:** Informar a la ciudadanía cómo se tasan las propiedades y cómo se utilizan estos valores en las políticas públicas para generar confianza y reducir posibles resistencias a su aplicación.
- **Establecer alianzas estratégicas:** Colaborar con universidades y centros de investigación para desarrollar capacidades y fomentar la innovación en el uso de la IA en la gestión fiscal subnacional y el desarrollo urbano.



CAPÍTULO 1.

Valuación inmobiliaria y gestión de los gobiernos subnacionales

HUÁSCAR EGUINO Y JUAN PABLO CARRANZA



La valuación inmobiliaria es una pieza central en la gestión de los gobiernos subnacionales porque sirve de base para determinar el impuesto a la propiedad inmobiliaria, sustenta la aplicación de instrumentos de financiamiento basados en la valorización del suelo y constituye un elemento de gran utilidad para la planificación y el desarrollo urbanos.

A pesar de su importancia, la estimación y actualización de los valores de las propiedades inmobiliarias tropieza con múltiples obstáculos, como las restricciones institucionales y normativas para su uso, los elevados costos de la actualización catastral, la falta de capacidades para la aplicación de las nuevas técnicas de valuación masiva basadas en el análisis de datos y la inteligencia artificial, la resistencia de los contribuyentes y la falta de decisión de las autoridades políticas responsables, entre otros. El resultado de estas limitaciones es que las tasaciones inmobiliarias se encuentran por debajo del valor de mercado, lo cual conlleva una facturación impositiva que no refleja el valor real de las propiedades, restringe el potencial tributario del impuesto inmobiliario y genera un tratamiento tributario regresivo de los contribuyentes (Maloney et al., 2024).³

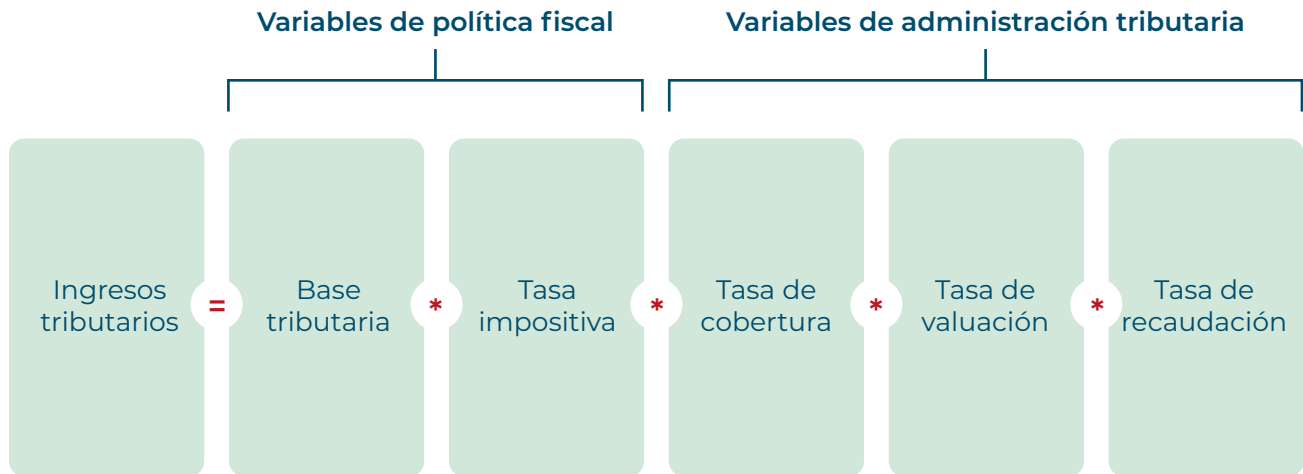
Como se describe en los capítulos siguientes, el uso de la inteligencia artificial (IA) facilita la actualización y estimación de los valores de las propiedades al reducir los costos y los tiempos requeridos. A su vez, contar con valores actualizados de los inmuebles sienta las bases para: **i)** aprovechar mejor el potencial tributario del impuesto a la propiedad y el impuesto a la transferencia de bienes inmuebles, **ii)** aplicar un conjunto de instrumentos de financiamiento urbano basados en la valorización inmobiliaria y **iii)** mejorar la planificación y gestión urbana, como la gestión de los riesgos climáticos. En las secciones que siguen, se explica la importancia y utilidad de disponer de valuaciones inmobiliarias próximas a los valores de mercado en cada uno de estos ámbitos de aplicación.

1.1. Impuesto a la propiedad inmobiliaria

Varios estudios coinciden en que los impuestos sobre la propiedad inmobiliaria tienen un gran potencial para movilizar ingresos en todos los países de América Latina y el Caribe (ALC). En efecto, si se compara la recaudación promedio del impuesto a la propiedad inmobiliaria en ALC (el 0,4% del producto interno bruto [PIB]) con la de los países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) (el 1,1% del PIB), se observa un amplio espacio para la mejora (Barreix et al., 2023).

No obstante, a fin de fortalecer el desempeño del impuesto a la propiedad, es importante entender que este depende de un conjunto de variables de política fiscal y gestión tributaria. La relación de estas variables y el impuesto a la propiedad inmobiliaria se traduce en la llamada ecuación de ingresos del impuesto a la propiedad, mediante la cual se definen las variables de política fiscal y gestión tributaria que inciden sobre este impuesto (gráfico 1.1).

3 Las propiedades de mayor valor suelen ser tasadas por debajo del valor de mercado, lo cual aumenta la regresividad.

Gráfico 1.1. Ecuación de ingresos del impuesto a la propiedad

Fuente: Elaboración propia, con base en Kelly, R. 2013. Making the Property Tax Work. International Center for Public Policy. Documento de trabajo No. 1311. Atlanta, GA: IcePP; Kelly, R., R. White y A. Anand. 2020. Property Tax Diagnostic Manual. Washington, D.C.: Banco Mundial.

Dentro del primer grupo de variables se encuentran las que dependen de la política tributaria. Concretamente, este grupo comprende las disposiciones de políticas que establecen la base tributaria (BT), conformada por la totalidad de las propiedades inmobiliarias sujetas al pago de impuesto, y la tasa o alícuota impositiva (TI) que se aplica a la base para determinar el monto del impuesto. El segundo grupo incluye las variables relacionadas con la administración tributaria: la tasa de cobertura (TC), que mide el número de propiedades registradas en el catastro de propiedades inmobiliarias respecto del total de propiedades sujetas a tributación; la tasa de valoración (TV), que define la relación entre el valor monetario de las propiedades registradas en el catastro y su valor de mercado, y la tasa de recaudación (TR), que refleja el vínculo entre los ingresos efectivamente recaudados y el valor de todas las facturas emitidas.

Como se observa en la ecuación (gráfico 1.1), la base tributaria y la tasa impositiva solo se pueden modificar mediante reformas a la política fiscal o ajustes del código tributario, que en algunos países de ALC no son responsabilidad exclusiva del nivel subnacional de gobierno. En cambio, las variables de administración tributaria (TC, TV y TR) generalmente dependen de las capacidades técnicas e institucionales de las administraciones subnacionales. Por eso son particularmente atractivas para los gobiernos responsables de este impuesto, ya que el desempeño de la administración tributaria obedece a normativas, procesos y decisiones que, en muchos casos, se hallan en la órbita del sistema administrativo y, por lo tanto, están fuera del complejo marco institucional que caracteriza la definición de las políticas tributarias del impuesto a la propiedad. En esta materia, también es importante destacar la heterogeneidad que se observa en ALC respecto de las responsabilidades institucionales sobre el impuesto inmobiliario o responsabilidades institucionales sobre el impuesto inmobiliario o sobre los sistemas de administración tributaria y de catastros,

un aspecto que complejiza aún más la implementación de acciones encaminadas a mejorar la recaudación del impuesto a la propiedad (Radics et al., 2022).⁴

El recuadro 1.1 presenta un ejemplo del impacto que tendría la mejora de la eficiencia administrativa del impuesto a la propiedad sobre la recaudación.

Recuadro 1.1. Impacto de la mejora de la eficiencia administrativa en la recaudación

La mejora de la eficiencia administrativa por sí sola puede incrementar de manera notable los ingresos provenientes del impuesto a la propiedad (cuadro R1.1.1).

Cuadro R1.1.1. Aumento de ingresos por eficiencia administrativa

Ratios administrativas	Escenario base	Escenario 1	Escenario 2
Tasa de cobertura	0,75	0,85	0,95
Tasa de valuación	0,50	0,70	0,90
Tasa de recaudación	0,45	0,55	0,85
Nivel de eficiencia ^a	0,17	0,33	0,73

^a Nivel de eficiencia es igual a $TC \cdot TV \cdot TR$.

Escenario base: Solo el 75% de las propiedades sujetas al impuesto han sido identificadas e incorporadas al catastro. Además, los valores de las propiedades están desactualizados y solo reflejan el 50% del valor de mercado. La administración tributaria es inefectiva y únicamente recauda el 45% de lo que factura. En este escenario, el gobierno local apenas logra recaudar el 17% de lo que podría cobrar si tuviese una mejor administración tributaria del impuesto inmobiliario.

Escenarios 1 y 2: En estos dos escenarios se muestra el impacto sobre la recaudación del impuesto a la propiedad que tendrían la ampliación de la cobertura del catastro (TC), la reducción de la brecha entre los valores catastrales y los valores de mercado (TV), y la mejora de la recaudación (TR). En el primer escenario, el nivel de eficiencia casi se duplicaría (de 0,17 a 0,33) y en el segundo caso se lograría cobrar el 73% de lo establecido por la política tributaria.

En otras palabras, los impactos de una mayor eficiencia en la administración del impuesto inmobiliario pueden ser de gran magnitud, por eso es importante que los gobiernos locales mejoren de manera simultánea sus sistemas de catastro, sus sistemas de valuación inmobiliaria y la gestión de cobro.

Fuente: Elaboración propia con base en Grote M. y J. Wen. 2024. How to Design and Implement Property Tax Reforms. Fondo Monetario Internacional. Nota práctica No. 2024/006. Washington, D.C.: FMI.

⁴ Si bien en la mayoría de los países de la región la gestión del impuesto a la propiedad inmobiliaria, o su equivalente, se delega a los gobiernos locales, la recaudación permanece centralizada en Chile, Guyana, Haití, Panamá y República Dominicana. En Chile y Panamá existen mecanismos de distribución de las responsabilidades sobre este impuesto entre las municipalidades. En el caso de Argentina, este impuesto se recauda a nivel provincial en casi todo el país.

En este contexto, resulta relevante determinar de qué manera se pueden utilizar la inteligencia artificial (IA) y las nuevas técnicas de análisis de datos para fortalecer el desempeño de las administraciones tributarias locales y, en especial, cómo se pueden usar para reducir la brecha entre los valores catastrales de las propiedades y los valores de mercado. De acuerdo con un reciente estudio de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), de un total de 16 países que realizaron mejoras en el impuesto a la propiedad, 9 concentraron sus esfuerzos en la modernización de los métodos de valuación de los inmuebles con técnicas de inteligencia artificial y de análisis de datos, lo cual demuestra la significativa importancia de este tema (OCDE, 2022).

En ALC también se observa un creciente interés en los temas de tasación inmobiliaria y mejora de la recaudación local, como indican la reciente aprobación del programa de fortalecimiento de la gestión fiscal subnacional de Perú, que incluye un componente para la mejora de la información catastral (BID, 2024); las valiosas experiencias de diversos gobiernos subnacionales que ya aplican modelos de valuación masiva en sus administraciones tributarias, y la realización de nuevos estudios sobre el uso de la IA y el análisis de datos en la actualización de los valores catastrales (Eguino y Erba, 2024).

Dos experiencias actuales de aplicación de las técnicas de IA en la valuación masiva de inmuebles las proporcionan las ciudades de Córdoba, en Argentina (Piumetto et al., 2019), y Fortaleza, en Brasil (Ferreira de Oliveira et al., 2024; Carranza et al., 2022). Estas dos ciudades cuentan con una importante trayectoria en el establecimiento de observatorios del mercado inmobiliario y en el uso de técnicas de IA para ajustar los valores catastrales de los inmuebles y, consecuentemente, aumentar la recaudación del impuesto a la propiedad (recuadro 1.2).

Recuadro 1.2. Uso de inteligencia artificial para la actualización de los valores catastrales

Municipio de Córdoba, Argentina. La ciudad de Córdoba cuenta con un observatorio del mercado inmobiliario que permite dar seguimiento a la dinámica del mercado de bienes inmuebles. Con base en la información de este observatorio, el gobierno municipal realizó la valuación masiva de los inmuebles de la ciudad. Para lograrlo, aplicó varios algoritmos de IA y finalmente seleccionó el denominado Random Forest. El modelo permitió determinar que el valor catastral de los inmuebles representaba, en promedio, el 60% del valor de mercado ($TV=0,6$), lo cual hizo posible ajustar los valores catastrales e incrementar el impuesto a los bienes inmuebles en un 45%.

Prefeitura municipal de Fortaleza, Brasil. La ciudad de Fortaleza tiene una importante tradición de dar seguimiento al comportamiento del mercado inmobiliario mediante el Observatorio Urbano de Valores Inmobiliarios. A partir de la información del observatorio, el gobierno municipal aplicó cinco algoritmos de aprendizaje automático (*machine learning* [ML]) para estimar los valores de mercado de los bienes inmuebles y finalmente eligió el algoritmo LightGBM. El modelo de valuación automatizada (AVM, por sus siglas en inglés) marca que el valor catastral de los inmuebles es solo el 40% del valor de mercado ($TV=0,4$), lo cual indica que el impuesto a la propiedad tiene el potencial de crecer más del 100%.

Fuente: Eguino H. y D. Erba. 2024. Mercado inmobiliario e impuesto predial: aplicaciones de técnicas de valuación masiva. Washington, D.C.: BID.

1.2. Financiamiento urbano basado en instrumentos de captura de valor

La relación positiva entre las intervenciones públicas y el valor de la propiedad está ampliamente documentada en la literatura científica, sobre todo cuando el gobierno proporciona servicios de infraestructura, modifica la zonificación y la densidad urbana, o introduce mejoras medioambientales. A su vez, el incremento del valor de la propiedad que generan un proyecto de inversión o una modificación normativa puede ser parcialmente capturado mediante una tasa o un impuesto, a fin de crear recursos adicionales para el gobierno local. Este es el fundamento de los instrumentos de captura de valor (o captura de plusvalía) como las contribuciones por mejoras, los cargos por derechos de construcción, los esquemas de reajustes urbanísticos y los instrumentos de financiamiento por incremento de impuestos (TIF, por sus siglas en inglés) (Blanco Blanco et al., 2016) (López-Ghio, Salazar Echavarría y Schloeter, 2020).

En general, todos estos instrumentos consideran que se conoce el valor de mercado de las propiedades inmobiliarias y que es posible estimar su incremento como resultado de una intervención urbana realizada por el sector público, por el sector privado o por un cambio normativo. De manera sintética, los instrumentos de captura de plusvalía funcionan de la siguiente manera:

- Las contribuciones por mejoras son un cargo a los propietarios de bienes inmuebles que se benefician de una mejora en la infraestructura local. Las contribuciones por mejoras pueden ser *ex ante*, es decir, antes de la construcción de la obra, o *ex post*. Por su parte, el monto para recaudar se puede definir sobre el costo del proyecto que se financiará o sobre la valorización inmobiliaria que este producirá.
- Los cargos por derechos de construcción abarcan una amplia gama de aportaciones urbanísticas en forma de dinero, obras o terrenos, que se estipulan o acuerdan como contraprestación por el derecho de realizar nuevos desarrollos urbanos.
- Los esquemas de reajuste de terrenos permiten incluir parcelas individuales dentro del área de aplicación de un plan parcial, lo cual incide en su valorización.
- El financiamiento por incremento de impuestos habilita el uso del incremento futuro de los impuestos a la propiedad producido por una intervención urbana, para financiar los costos de esta.

El capítulo 3 presenta una descripción más detallada sobre este tipo de instrumentos.

1.3. Planificación y gestión urbana

Contar con información sobre el valor de los inmuebles también es útil para mejorar la planificación y gestión urbana. En efecto, a partir del conocimiento del valor de los inmuebles es posible mejorar la forma en que se evalúan las intervenciones urbanas, optimizar el uso del suelo, conocer el riesgo de las infraestructuras frente a eventos climáticos o ambientales, y evaluar las pérdidas o los daños en caso de desastres. A continuación, se explican brevemente estas aplicaciones:

- **Evaluación de intervenciones urbanas.** El análisis costo-beneficio (ACB) es un poderoso instrumento para seleccionar proyectos que proporcionan la mayor rentabilidad financiera y socioeconómica. Una forma de evaluar los proyectos urbanos es determinar su impacto sobre la valorización inmobiliaria, que, a su vez, permite analizar su viabilidad fiscal y financiera. Concretamente, una vez que se cuenta con un AVM, es posible simular el impacto de un nuevo desarrollo urbano y establecer su efecto sobre el valor de los inmuebles en el área de influencia. Por ejemplo, los proyectos de desarrollo basados en el transporte (TOD, por sus siglas en inglés) suelen tener un importante impacto en el mercado inmobiliario del área de influencia. Al mismo tiempo, el incremento del valor de los inmuebles genera la posibilidad de aumentar el cobro del impuesto inmobiliario y produce un beneficio privado que acrecienta la rentabilidad financiera de los proyectos en el área de influencia.
- **Mejora de las normativas de uso de suelos.** Los modelos de valuación o tasación masiva también permiten identificar las áreas urbanas subutilizadas y simular el impacto económico que tendría un cambio en las normas de uso de suelos y construcción. Esto es particularmente útil en ciudades que siguen una política de densificación como forma de reducir los costos de las infraestructuras y los servicios urbanos.
- **Mejora de la gestión de riesgos climáticos y ambientales.** Otra de las aplicaciones de los modelos de valuación masiva consiste en estimar o calcular las pérdidas en el valor de los inmuebles provocadas por eventos climáticos o ambientales extremos y simular el impacto que tendrían las intervenciones orientadas a reducir los riesgos. Esta aplicación tiene especial relevancia en zonas urbanas sujetas a huracanes, inundaciones, deslizamientos, incendios y otros fenómenos físicos que generan pérdidas económicas y humanas, y afectan el mercado inmobiliario.

En los siguientes capítulos se explica en detalle la forma de integrar la valuación inmobiliaria en las finanzas y la planificación de los gobiernos locales. Para cada una de estas aplicaciones se describen los pasos que se deben seguir, ejemplos de uso y los códigos en R correspondientes.



CAPÍTULO 2. Evolución de las técnicas de valuación masiva automatizada y oportunidades de la inteligencia artificial

JUAN PABLO CARRANZA Y HUÁSCAR EGUINO



El presente capítulo sintetiza las características esenciales de los modelos de valuación automatizada (AVM) y las compara con las técnicas de valuación tradicionales, para resaltar sus principales ventajas y desventajas desde una óptica de política pública. Se incluyen las técnicas que han sido más utilizadas para la realización de este tipo de procedimientos, como las regresiones lineales (junto con sus desarrollos más recientes elaborados desde la estadística espacial), la interpolación geostatística y las novedosas técnicas de aprendizaje automático (ML). Para cada una de estas técnicas se hace un recuento de sus beneficios y sus dificultades, con especial atención en los aspectos técnicos y tecnológicos, pero también en su aplicación por parte de las administraciones públicas subnacionales. Finalmente, se proporciona un ejemplo simple del proceso de aprendizaje de un algoritmo de ML, para explicar la lógica de funcionamiento de esta clase de técnicas.

2.1. ¿Qué es un proceso de valuación masiva automatizada?

Un AVM consiste en la aplicación de un programa informático de base matemática o estadística para elaborar una estimación del valor de mercado de los inmuebles a partir de información sobre sus características, su ubicación y las condiciones del mercado (IAAO, 2022). Es decir, se trata de un proceso automatizado que parte de la observación de datos de mercado para interpolar el valor de mercado de las propiedades inmuebles a un área determinada según un modelo matemático o estadístico. Esta interpolación del valor se realiza con la ayuda de información territorial que describe las características de la localización y el entorno de los inmuebles. Por lo general, en estos modelos el valor de mercado de la propiedad inmobiliaria recibe el nombre de **variable dependiente** y el conjunto de información territorial que describe las características de la localización y/o el entorno del inmueble se denomina **variables independientes**.

El proceso de relevamiento de datos de mercado y la calidad de la muestra son fundamentales para el uso de los AVM. Una buena muestra requiere no solo determinar la veracidad de los datos, sino también su **distribución espacial**. Existen varias técnicas para analizar este último aspecto, aunque, en general, se considera que una muestra distribuida de manera homogénea en el espacio es de mayor calidad que una muestra en la que existen grandes partes del área de estudio sin observaciones. En esta etapa se debe registrar el precio y la localización de los inmuebles, así como sus **características intrínsecas**: superficie, ubicación en la cuadra, situación jurídica (cuenta con escritura o no), si es una oferta inmobiliaria o un inmueble efectivamente vendido, entre otras. Asimismo, es importante contar con algún instrumento de sistematización de la información relevada, a fin de que los datos no permanezcan aislados en diferentes registros y puedan ser procesados de forma ordenada. A tal efecto, hay distintas alternativas de sistemas de información geográfica, tanto de acceso libre (QGIS o gvSIG, por ejemplo) como *software* propietario o aplicaciones *online* que permiten gestionar este tipo de información.

También es vital tratar correctamente la muestra de mercado, ya que esto facilita la comparación de los inmuebles con diferentes características intrínsecas. En ese sentido, algunas fuentes de heterogeneidad que dificultan el análisis son: **i)** inmuebles de tamaño, antigüedad o estado de conservación distintos; **ii)** parcelas que varían en forma, dimensiones o ubicación en la cuadra. Para afrontar este problema, habitualmente los catastros utilizan una serie de tablas de coeficientes que permiten descontar los efectos particulares de cada inmueble y expresar su valor por metro cuadrado en términos de un inmueble típico o mediano. La dificultad que tiene este enfoque es que, por lo general, la determinación de los coeficientes no guarda

relación con la dinámica del mercado inmobiliario. El capítulo 4 presenta una serie de herramientas novedosas para dar cuenta de estos aspectos y homogeneizar los valores por metro cuadrado mediante la aplicación de técnicas de estadística espacial, en línea con la propuesta metodológica realizada por Cerino et al. (2020) y por Carranza (2024).

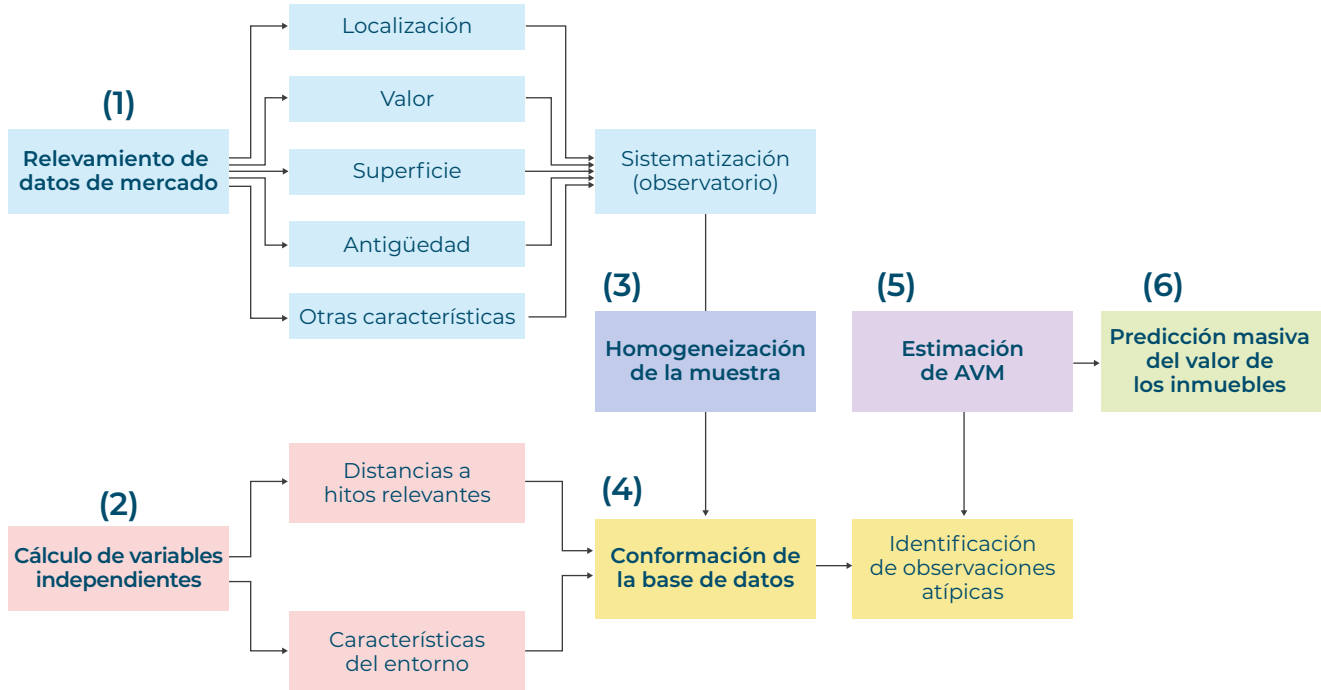
Con relación a la generación de información sobre las **variables independientes**, en general se requiere personal calificado en el manejo de sistemas de información geográfica o de lenguajes de programación. A grandes rasgos, el cálculo de variables independientes se puede agrupar en las que informan sobre la **posición relativa** de cada observación respecto de diferentes hitos urbanos de interés (como avenidas o edificios públicos, entre otros) y las que reportan las **características del entorno** de cada observación (como la cantidad de comercios en el entorno o la altura promedio de las edificaciones en el vecindario, entre otras). El capítulo 4 proporciona un conjunto de herramientas orientadas a facilitar la generación de este tipo de información y disminuir las barreras a la implementación de los AVM.

Corresponde resaltar que el cálculo de las variables independientes se debe realizar para toda el área de estudio y no solo en las localizaciones que atañen a la muestra de mercado, puesto que las variables independientes operan como el vehículo que utilizan los AVM para interpolar las relaciones matemáticas entre estas y el valor de mercado al resto del espacio.

Una vez que se ha conformado la muestra de mercado, se ha verificado que los datos que contiene se encuentran completos con relación a las características relevadas para cada inmueble, y que son comparables y se han calculado las variables independientes, se está en condiciones de realizar un análisis de **detección de observaciones atípicas**. Esta estrategia se puede implementar de diversas maneras y en todas ellas se compara el valor de mercado de cada observación con el precio observado en el entorno y/o con las características del vecindario. En el capítulo 4 se presentan algunas herramientas que sistematizan el abordaje de este fenómeno.

Finalmente, con la muestra depurada de observaciones atípicas y con las variables independientes calculadas para cada observación, se procede a la **estimación de un AVM**. Existen muchas técnicas disponibles para construir un modelo de este tipo, que se pueden reunir en cuatro grupos: modelos geoestadísticos, modelos econométricos, modelos de aprendizaje automático (*machine learning* [ML]) y modelos de aprendizaje profundo (*deep learning*). Todas estas técnicas tienen ventajas y desventajas (se discuten en la sección 2.6), pero la lógica que domina el proceso es esencialmente la misma: a partir de la interacción del valor de mercado en diferentes localizaciones y las variables independientes utilizadas, se interpolan estas relaciones al resto del espacio urbano donde se dispone de información sobre la distribución espacial de las variables independientes, pero se desconoce el valor de mercado. Esta etapa del proceso es crítica en términos del capital humano necesario para implementarla, ya que, por lo común, las capacidades técnicas requeridas no se hallan en el ámbito de las administraciones públicas locales. Esta limitación ha sido un factor clave en el atraso que exhiben las valuaciones inmobiliarias en América Latina y el Caribe. En el capítulo 4 se proporciona un conjunto de herramientas que apuntan a aliviar esta restricción operativa y dotar de un flujo de trabajo ordenado para la estimación de un AVM mediante técnicas de ML.

El gráfico 2.1 esquematiza el proceso de implementación de un AVM en los términos detallados en los párrafos precedentes.

Gráfico 2.1. Proceso de implementación de un modelo de valuación automatizada

Fuente: Elaboración propia.
 AVM: modelo de valuación automatizada, por su sigla en inglés.

2.2. Nivel de aplicación del modelo de valuación automatizada

En 2019, el Lincoln Institute of Land Policy y la International Association of Assessing Officers (IAAO) realizaron una encuesta para conocer el nivel de aplicación de modelos de valuación automatizada (AVM) en diferentes jurisdicciones (Bidanset y Rakow, 2022). El relevamiento se efectuó entre los más de 8.000 miembros de la IAAO, quienes mayormente se ubican en Australia, Canadá, Estados Unidos e Irlanda. Si bien se halló más propensión a valorar los resultados de la aplicación de AVM en las jurisdicciones más grandes y con ciclos de valuación más cortos, solo un 16% de los encuestados afirmó utilizar este tipo de metodologías, lo cual permitió establecer un espacio de oportunidades para la modernización de las políticas públicas de valuación inmobiliaria.

En una extensa revisión de la bibliografía científica sobre la inserción de los procesos de valuación automatizada en diferentes campos, El Jaouhari et al. (2024) destacan su papel transformador en la racionalización de los procesos de valoración, la mejora de la eficiencia del mercado inmobiliario y el apoyo a la toma de decisiones basada en datos en distintos sectores, como el desarrollo inmobiliario, la administración del suelo y la política fiscal.

En línea con lo anterior, actualmente la utilización de este tipo de modelos para la predicción del valor de los inmuebles trasciende los ámbitos de la política pública fiscal o de desarrollo urbano. A medida que la efectividad de este tipo de estrategias metodológicas crece, cada vez más actores del sector privado se involucran en su uso. Se trata, principalmente, de empresas vinculadas con los sectores de seguros y desarrollo inmobiliario, que canalizan una gran cantidad de capital.

En 2019, la Asociación de Banqueros Hipotecarios de Estados Unidos (MBA, por sus siglas en inglés), resaltaba que la normativa y las políticas limitaban la implementación de modelos valuatorios y señalaba que algunas de esas restricciones se podían superar con la aplicación de AVM, la creciente capacidad para procesar grandes volúmenes de información de manera rápida y con mayor precisión, la generación de información estadística adicional relativa a una tasación y la implementación de procesos automatizados y basados en reglas que pueden minimizar el sesgo humano (MBA, 2019). Estas observaciones de la MBA continúan siendo relevantes en la actualidad. En este contexto, no es extraño que los sectores más dinámicos en la implementación del AVM sean los que se encuentran fuera de las administraciones públicas, como las empresas de desarrollo inmobiliario, los bancos y las compañías de seguros, ya que no están sujetos a un cuerpo normativo diseñado específicamente para su aplicación. Esta situación pone de relieve la característica integral de política pública que conlleva un proceso de valuación masiva ejecutado desde el Estado: para implementar un AVM, no alcanza con resolver las restricciones técnicas y los déficit de capacidades de las administraciones estatales, sino que también será necesario trabajar sobre los marcos normativos para modernizarlos y hacerlos permeables a las innovaciones técnicas y metodológicas.

2.3. Situación en América Latina y el Caribe

En América Latina y el Caribe (ALC) existen escasos ejemplos de realización de avalúos masivos mediante la aplicación de modelos estadísticos o matemáticos. Como se detalló en el capítulo 1, se destacan los casos de los gobiernos subnacionales de Córdoba (Argentina) y Fortaleza (Brasil), que aplican técnicas de ML sobre una muestra de inmuebles en venta o vendidos. También se puede mencionar el caso de Bogotá (Colombia), que realiza anualmente un avalúo catastral a partir de un censo inmobiliario. Sin embargo, independientemente de las consideraciones metodológicas para la ejecución de avalúos en estas y otras pocas ciudades de ALC, la situación general en la región muestra que las valuaciones están desactualizadas (Maloney et al., 2024; Bueno Villegas et al., 2022; Bonet et al., 2014).

En la región, la responsabilidad institucional para el registro y mantenimiento de toda la información relacionada con los inmuebles, que incluye las valuaciones, siempre recayó en los catastros. Tradicionalmente, los modelos de datos catastrales comprenden información con un elevado nivel de detalle, por lo cual el mantenimiento y la actualización de esta información es un proceso lento y costoso, que deriva en la desactualización de la base de datos. Ante esta situación, en general los catastros recurrieron a la definición de zonas catastrales homogéneas, con un mismo valor por metro cuadrado, sin considerar sus particularidades intrazonales. De esta manera, la tensión entre la necesidad de información detallada de los modelos de datos catastrales y la ausencia de metodologías de valuación rápida y masiva se resolvió mediante la estimación de valuaciones a escalas mucho menos minuciosas. En este escenario, la disyuntiva entre la pulsión al detalle en el registro de la información catastral y la imprecisión de las valuaciones por zonas

homogéneas dominó por mucho tiempo las estrategias de valuación de los catastros latinoamericanos, lo cual contribuyó a su paulatina desactualización.

En un estudio realizado para el Banco Interamericano de Desarrollo (BID) en 10 provincias argentinas se observó un atraso generalizado en las valuaciones inmobiliarias (Piumetto et al., 2022). El documento destaca algunas situaciones, como la de Corrientes, donde la valuación fiscal urbana representaba el 0,1% de los valores de mercado, y las de Mendoza y Río Negro, donde esa relación se ubicaba en torno del 30%. Pero, en general, la investigación encuentra que en la mayor parte de las provincias las valuaciones se ubican por debajo del 10% de los valores de mercado. Por su parte, en México se advierte un escenario similar, con atrasos de más de 15 años en la actualización de las valuaciones en el 40% de los inmuebles, lo cual contradice, incluso, las leyes nacionales (Bueno Villegas et al., 2022).

Esta situación es transversal a diferentes países de América Latina y el Caribe. Una investigación efectuada en 200 jurisdicciones de la región halló que, en términos generales, las valuaciones están desactualizadas y representan porcentajes inferiores al 50% de los valores de mercado (De Cesare, 2016).

De igual manera, otros estudios han medido el potencial de recaudación perdido en la gestión del impuesto inmobiliario (Bonet et al., 2014; Eguino y Erba, 2020). Independientemente de las diferencias metodológicas, los resultados revelan un potencial recaudatorio desaprovechado, que encuentra en la desactualización de las valuaciones una de sus principales explicaciones (Castro et al., 2014).

El escenario, entonces, es propicio para implementar iniciativas de innovación pública que lleven a una modernización en la determinación de las valuaciones fiscales y catastrales, mediante la aplicación de AVM como parte de una política integral que, además, apunte a la actualización de los marcos normativos.

2.4. Diferencias entre valuaciones masivas y tasaciones individuales

El principal atractivo de los modelos de valuación automatizada consiste en facilitar el avalúo inmobiliario a gran escala, aunque a expensas de resignar un mayor detalle en los resultados. Si bien el proceso de tasación tradicional está sujeto a consideraciones normativas y múltiples recomendaciones de las organizaciones de tasadores, es cuasiartesanal y, debido a su nivel de detalle, emplea un tiempo considerable para determinar el valor de mercado de cada propiedad inmueble. Por lo tanto, la utilización de los modelos de valuación inmobiliaria a gran escala se convierte en un elemento central de la estrategia de actualización catastral de las administraciones locales. Como se mencionó en apartados anteriores, la utilidad de estos modelos también es notoria para las entidades bancarias que proporcionan financiamiento hipotecario, las compañías de seguros, los inversionistas en grandes proyectos de desarrollo inmobiliario, los planificadores urbanos y los responsables de las finanzas locales. Aunque sus márgenes de error son más elevados que los de las tasaciones individuales, los modelos de valuación automatizada proveen información masiva y actualizada sobre el valor de mercado de los inmuebles, que es un insumo útil para dinamizar los procesos administrativos, ya que constituye una base sobre la que luego se pueden hacer tasaciones particulares más detalladas.

Una ventaja adicional de los AVM con relación a las tasaciones individuales es que al abarcar la totalidad del mercado inmobiliario local permiten obtener una estimación del error esperado en los resultados. Esto es imposible de lograr en las tasaciones individuales que trabajan caso a caso sin realizar una comparación sistemática con la estructura del mercado inmobiliario.⁵ Conocer el error esperado en los resultados es un insumo crítico para las autoridades locales que emprenden un proceso de avalúo masivo, puesto que se convierte en una medida del costo social o político al que se exponen al implementar este tipo de política pública.

Por todo lo expuesto, los AVM no suelen ser considerados sustitutos de las tasaciones individuales, sino un complemento que aporta información a una escala mucho más amplia, pero con un nivel de detalle menor.

2.5. Evolución de las técnicas aplicadas en valuaciones masivas automatizadas

Las primeras aproximaciones científicas a las valuaciones masivas aplicaron técnicas estadísticas, como la regresión lineal múltiple.⁶ A mediados de la década de 1960, y a partir del análisis pionero de Alonso (1964) centrado en el uso del suelo, se dio una expansión de las investigaciones sobre el precio de los inmuebles producidas desde la incipiente economía urbana. Este tipo de investigaciones se apoyaba en modelos teóricos que analizaban las decisiones de localización de las personas y/o de las empresas en función de la interacción entre la demanda de inmuebles y los desplazamientos al trabajo en un contexto de maximización de la utilidad o de minimización del costo de producción mediante la proximidad al área central de la ciudad. El instrumental de análisis estocástico que generalmente aplicaban estas investigaciones consistía en modelos lineales ajustados por mínimos cuadrados ordinarios. Como consecuencia de este *boom* en torno del análisis del mercado inmobiliario como escenario para la localización de personas y empresas, se comenzaron a observar iniciativas públicas para determinar de forma masiva el precio de los inmuebles apoyadas, también, en esta clase de enfoques estadísticos.

La utilización de técnicas de regresión lineal múltiple, que comúnmente se aplican por medio de modelos de precios hedónicos, permite **explicar** el precio de los inmuebles a partir de una fórmula impuesta de antemano por el investigador. Con la estimación de los parámetros de la regresión se puede conocer la influencia de diferentes variables independientes y su contribución a la formación del valor de mercado de los inmuebles. Si bien los parámetros estimados por la regresión lineal son insesgados, tienen una elevada varianza cuando se emplean para predecir el valor de los inmuebles fuera de la muestra (Kleinberg et al., 2015). En ese sentido, aunque la técnica de regresión lineal proporciona una estimación de las variables que explica las valuaciones obtenidas, su **nivel de error es elevado** en los entornos donde la cobertura de datos muestrales es escasa.

-
- 5 Al margen de que algunas técnicas de tasación implican la comparación con inmuebles cercanos, no hay un análisis estructural respecto del resto de los precios observados en el mercado.
 - 6 Bidanset y Rakow (2022) resaltan que la primera mención a un AVM se encuentra en una tesis realizada por George Haas en 1922, donde se propone utilizar una regresión lineal para tasar las tierras rurales aptas para cultivo con mayor precisión a efectos del impuesto sobre bienes inmuebles y la concesión de préstamos.

Un problema adicional de los modelos lineales estimados por mínimos cuadrados ordinarios es la existencia de **sesgos de dependencia espacial** (Anselin, 1988), principalmente en los residuos del modelo de regresión por mínimos cuadrados ordinarios. En estas situaciones, los residuos positivos tienden a aglomerarse en determinadas partes del espacio geográfico, al igual que los negativos. En la práctica, esto implica que, si se usan estos modelos para la valuación inmobiliaria, diferentes sectores de una ciudad serán subvaluados, en tanto que otros serán sobrevaluados. En consecuencia, cuando se aplican las valuaciones se producen problemas de equidad relevantes, por ejemplo, para determinar las bases imponibles del impuesto inmobiliario.

A mediados de la década de 1970 se comenzaron a usar en las valuaciones masivas diferentes técnicas de interpolación geoestadística, cuyo desarrollo venía potenciando hacía algunos años el sector minero en la detección de diferentes yacimientos en el subsuelo. Estas técnicas tomaron impulso con el trabajo pionero de Krige (1951) e implicaron un cambio metodológico respecto de las regresiones lineales, ya que el valor de los inmuebles pasaba a explicarse solo por la proximidad a observaciones muestrales y no por relaciones de causalidad con distintas variables independientes. Las más utilizadas son algunas variantes de Kriging, cuyas principales características se analizan más adelante.

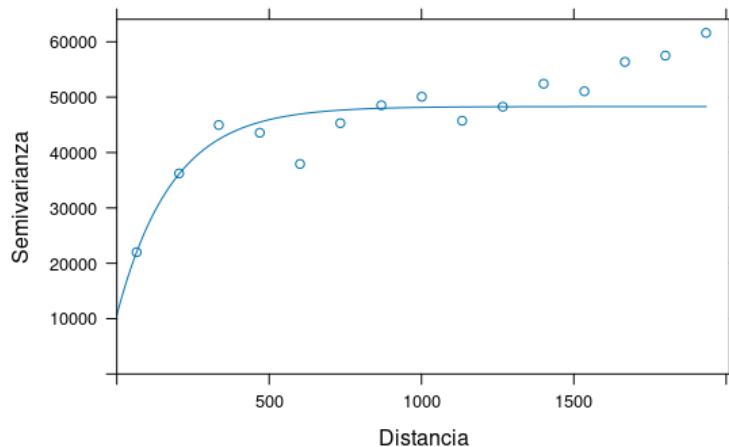
Por lo tanto, la estimación de valuaciones con técnicas de interpolación geoestadística tiene la ventaja de que no precisa otras variables explicativas al margen de las **coordenadas geográficas** de cada dato.⁷ Así, luego de la localización geográfica de cada dato muestral, se interpola el valor de mercado al resto del espacio bajo estudio. Esta característica reduce drásticamente el tiempo (y el presupuesto) destinado a la recolección y el procesamiento de información territorial para generar variables independientes. Sin embargo, se trata de técnicas que necesitan **muestras grandes y distribuidas de forma relativamente homogénea** en el espacio, por lo cual resulta difícil aplicarlas en espacios urbanos donde el mercado inmobiliario es poco dinámico.

La lógica de un modelo geoestadístico como Kriging Ordinario requiere la estimación de un semivariograma, que resume la estructura de autocorrelación espacial de los datos muestrales. Esto se logra mediante la utilización de un plano cartesiano donde el eje de las abscisas mide la distancia euclidiana entre pares de observaciones muestrales y el eje de las ordenadas mide la semivarianza del valor de los inmuebles (la mitad de la varianza de las diferencias entre puntos separados por una distancia constante). Una vez ubicados los datos muestrales en este plano, se estima el modelo teórico que mejor se ajusta a la estructura espacial de la muestra. En el gráfico 2.2 se expone un semivariograma construido para una muestra de inmuebles edificados en venta en la ciudad de Cali, Colombia. En la distribución de los datos en este plano se puede apreciar que los inmuebles localizados más cerca entre sí tienen una semivarianza menor (son más similares) que aumenta con la distancia hasta alcanzar un comportamiento asintótico. Esto implica que los datos presentan autocorrelación espacial, entendida como una situación en la que el valor de una observación está influido por los valores observados en el vecindario. El modelo teórico que mejor se ajusta a los datos utilizados en este ejemplo es un modelo exponencial y está representado en el plano por la línea continua que corta a la nube de puntos. Este semivariograma teórico provee los valores de tres parámetros

⁷ Existen algunas variantes de estas técnicas que sí utilizan otras variables independientes además de las coordenadas geográficas, como *cokriging* o *regression kriging*.

relevantes (*nugget*, *partial sill* y *range*) que se emplean para realizar una combinación lineal que devuelve el valor por metro cuadrado de los inmuebles en la totalidad del área de estudio.

Gráfico 2.2. Semivariograma teórico y empírico para una muestra de inmuebles en venta en Cali, Colombia



Fuente: Elaboración propia.

Hacia mediados de la década de 1990 se comenzaron a aplicar técnicas que combinaban los modelos lineales con los modelos geoestadísticos. En concreto, esta alternativa metodológica, conocida como *regression kriging* (Odeh, McBratney y Chittleborough, 1995), consiste en la realización de un modelo lineal y la posterior interpolación de los residuos del modelo de regresión para eliminar el sesgo por dependencia espacial del análisis. En la misma época, se observó una relativa proliferación de técnicas de estadística espacial (Anselin, 1988). Estas técnicas son una evolución de los modelos de regresión lineal múltiple por mínimos cuadrados ordinarios que apunta a eliminar los sesgos por dependencia espacial. Esta estrategia permite dar cuenta de la autocorrelación espacial inherente al fenómeno, donde el valor que adquiere una determinada variable de índole espacial puede ser función del valor que asumen las observaciones vecinas, del valor de las variables dependientes en el entorno o de variables omitidas en el entorno (efecto que canaliza la autocorrelación espacial a través del término de error).

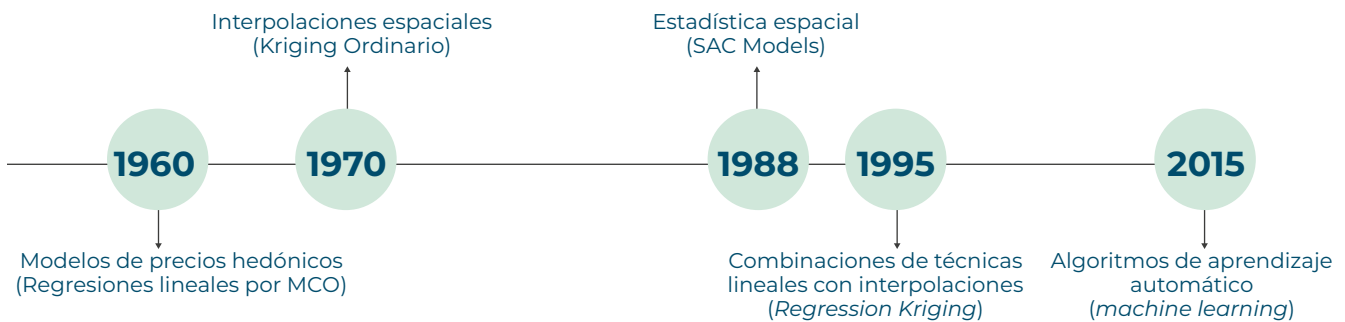
El insumo básico para la internalización de los efectos de la dependencia espacial dentro de los modelos lineales radica en la definición de una estructura de vecindario a partir de una matriz no estocástica de pesos espaciales W_{ij} . Esta matriz permite incorporar a los parámetros estimados los efectos de la autocorrelación espacial de diferentes términos del modelo de regresión. Los supuestos aplicados sobre la estructura de esta matriz son críticos y condicionan de manera significativa los resultados. La matriz de vecindario es cuadrada, con una dimensión igual al tamaño de la muestra. Cada elemento de esta matriz asume el valor 1 si la observación i es vecina de la observación j y el valor 0 en otro caso (por definición, se asume que ninguna observación puede ser vecina de sí misma, por lo cual la diagonal principal de la matriz es igual a cero). Posteriormente, la matriz se estandariza por filas, con lo cual es posible ponderar el efecto del vecindario según la distancia inversa entre cada par de observaciones (es decir, cada elemento puede dividirse por la distancia euclidiana entre ambas observaciones o por alguna otra medida de distancia).

Esta última estrategia permite darles mayor peso a las observaciones más próximas y menor peso a las observaciones más lejanas.

Si bien la aplicación de los modelos de estadística espacial constituyó un avance notable para el tratamiento de estos fenómenos geográficos, su uso no permeó hacia las políticas públicas de actualización de las valuaciones catastrales o fiscales y quedó acotado a la academia.

Ya entrado el siglo actual, la irrupción de técnicas modernas de ML impulsó algunas iniciativas de aplicación de algoritmos en la realización de avalúos masivos (gráfico 2.3). Este tipo de técnicas permite trabajar simultáneamente sobre los problemas de la estadística clásica y la geoestadística, sin resignar sus principales ventajas. Con la utilización de algoritmos para predecir el precio de los inmuebles es posible obtener estimaciones confiables a partir de muestras de tamaño relativamente pequeño, lo cual conforma una ventaja en sectores donde la cobertura muestral no es intensa (o es nula).

Gráfico 2.3. Evolución de las técnicas utilizadas en valuaciones masivas automatizadas, período 1960-2015



Fuente: Elaboración propia.
MCO: mínimos cuadrados ordinarios; SAC: Spatial Autoregressive Confused.

En un metaestudio realizado en 2019, Wang y Li (2019) analizaron 104 publicaciones sobre valuaciones masivas automatizadas aplicadas desde 2000 y observaron un crecimiento significativo a partir de 2017, principalmente vinculado a la aplicación de modelos de ML.

El menú de algoritmos que conforma el universo de técnicas de ML es muy amplio. A grandes rasgos, en su aplicación acotada a la predicción del comportamiento de una variable continua a partir de una muestra,⁸ hay dos grandes categorías que pueden agrupar a múltiples algoritmos según sus características lógicas: los **modelos basados en árboles** y los **modelos basados en redes neuronales**. Existen otros algoritmos que han sido aplicados a este tipo de problemas, pero por sus particularidades no pueden ser incluidos en una categoría específica junto a otros algoritmos similares (modelos basados en máquinas de soporte, modelos basados en vecinos más cercanos en el hiperplano, modelos lineales penalizados o modelos bayesianos, entre otros).

8 Conocidos formalmente como problemas predictivos de regresión con una estrategia supervisada.

En la academia existe consenso sobre un desempeño relativamente más eficiente de los modelos basados en árboles para la predicción del valor de los inmuebles, en particular del algoritmo Random Forest (Breiman, 2001) en alguna de sus variantes.⁹

A nivel regional, el desarrollo de este tipo de investigación es aún incipiente. Carranza et al. (2022) implementaron un proceso de valuación masiva de la tierra a precios de mercado mediante la utilización de datos abiertos y algoritmos basados en árboles en la ciudad de Fortaleza (Brasil). En Argentina, Carranza et al. (2019) aplicaron algoritmos de la misma naturaleza metodológica, pero con información catastral, para predecir el precio de la tierra en la ciudad de San Francisco, en Córdoba (Argentina). Eguino y Erba (2024) usaron modelos de ML para predecir el valor de mercado de la tierra urbana en las ciudades de Córdoba (Argentina), Corrientes (Argentina), Fortaleza (Brasil), Manizales (Colombia) y Benito Juárez (México). En tanto, Cerino et al. (2021) propusieron abordajes similares para predecir el precio de la tierra urbana en un corredor turístico de las sierras de Córdoba (Argentina). Por su parte, Córdoba et al. (2021) implementaron un proceso predictivo similar, pero aplicado al precio de la tierra rural en la provincia de Córdoba (Argentina). El algoritmo utilizado permitió, además de la estimación de un mapa de precios de mercado de la tierra rural, la construcción de un mapa con la distribución espacial de la calidad predictiva del modelo empleado, lo cual hizo posible identificar sectores donde los resultados no tienen la calidad deseada. De esta manera, se puede reforzar la muestra de forma recursiva en dichos espacios, para elevar la calidad de los estándares en los resultados finales.

En todos los casos mencionados, el proceso metodológico se basa en el relevamiento de una muestra de datos de lotes en venta o vendidos, la generación de variables independientes de índole territorial con potencial de explicar la formación del precio de la tierra, el entrenamiento de diferentes algoritmos de ML, la evaluación de la capacidad predictiva de cada uno de ellos medida fuera de la muestra y la elección del mejor algoritmo para realizar una predicción del precio por metro cuadrado en toda el área urbana. Este proceso permite obtener un mapa detallado del valor de mercado de los inmuebles a partir de la utilización de muestras relativamente pequeñas y con una elevada calidad en los resultados.

Jafary et al. (2024) analizaron distintas estrategias metodológicas para predecir el precio de la tierra en el área metropolitana de la ciudad australiana de Melbourne. Al contrario que en los estudios citados antes, en este caso el objetivo era examinar las diferencias en la calidad de las predicciones entre métodos de aprendizaje automático y métodos de **aprendizaje profundo**. Esta distinción es importante porque la inteligencia artificial es un campo amplio que reúne técnicas de ML o aprendizaje automático, que permiten extraer conocimiento de un conjunto de datos mediante algoritmos relativamente sencillos, y del más recientemente desarrollado aprendizaje profundo, que requiere menos intervención humana, debido a que las características de un conjunto de datos se extraen de forma automática. Por lo general, este último tipo de desarrollo se considera más preciso, pero necesita un poder de cómputo exponencialmente más elevado. Los investigadores compararon el desempeño de una red neuronal profunda con diferentes algoritmos de ML basados en árboles y hallaron una diferencia muy elevada a favor de estos últimos. En consecuencia,

⁹ Para más información, véanse los siguientes documentos: Carranza et al. (2022); Gao et al. (2022); Cerino et al. (2021); Córdoba et al. (2021); Worachairungreung, Thanakunwutthirot y Ninsawat (2021); Kim et al. (2021); Zhang et al. (2021); Ma et al. (2020); Hong, Choi y Kim (2020); Yilmazer y Kocaman (2020); Derdouri y Murayama (2020); Carranza et al. (2019).

concluyen que, a pesar de los enormes avances logrados en el uso de métodos de aprendizaje profundo para tareas de clasificación basadas en textos e imágenes, los modelos de ML siguen mostrando resultados de mayor calidad en determinados escenarios que incluyen datos tabulares y problemas de regresión.

Quizás la única desventaja de la aplicación de algoritmos de ML es la **imposibilidad de reconstruir el proceso** de generación de valor. Esto obedece a la naturaleza de “cajas negras” de este tipo de métodos. En contraposición a las técnicas de interpolación, donde el valor predicho se explica por proximidad a los datos de la muestra, o las técnicas de regresión lineal, en las que el valor de los inmuebles se determina a partir de una forma funcional impuesta de antemano, en los algoritmos de ML **el modelo surge de los datos**. Esta capacidad de capturar localmente las relaciones no lineales entre la variable dependiente y las variables independientes proporciona una ventaja comparativa en la calidad de los resultados, pero no permite reconstruir de forma global estas interrelaciones o establecer los factores causales. Esta limitación puede ser determinante en entornos legales-formales, lo cual afecta el potencial de estas técnicas innovadoras para traducirse en políticas públicas.

2.6. Ventajas y desventajas de las diferentes técnicas para realizar valuaciones masivas automatizadas

Con base en lo expresado en el apartado 2.5, se puede argumentar que existen ventajas y desventajas asociadas a cada una de las estrategias para determinar valuaciones a gran escala. Como se ha señalado, los modelos lineales permiten reconstruir una estructura lógica para explicar la contribución de las variables independientes en la formación del precio de los inmuebles. De esta forma es posible obtener una fórmula para detallar, caso por caso, cómo se conforma la valuación, lo cual constituye una gran ventaja en entornos formales-legales como los que afrontan las administraciones públicas en el trato tributario con los contribuyentes. Pero los modelos lineales generalizan mal hacia afuera de la muestra, lo cual los hace susceptibles de incurrir en errores significativos en los sectores donde la cobertura de datos relevados no ha sido adecuada. Lamentablemente, en problemas predictivos orientados a determinar el valor de los inmuebles hay sectores en los que el mercado inmobiliario no tiene el dinamismo suficiente y es frecuente encontrar partes de las ciudades escasamente representadas en la muestra.

Además, generar variables independientes implica articular mucha información territorial que puede ser costosa de recopilar o bien la administración pública puede carecer de las capacidades técnicas para procesarla. Esta restricción se puede superar con la utilización de técnicas de interpolación como Kriging Ordinario, que solo requiere las coordenadas geográficas de los datos muestrales para realizar predicciones en el espacio bajo estudio. Sin embargo, al igual que los modelos lineales, este tipo de técnicas necesita muestras de un tamaño relativamente grande y distribuidas de forma homogénea en el espacio, por lo cual genera errores considerables en sectores que, debido a la dinámica inmobiliaria, no están representados de manera correcta en la muestra.

Los modelos de ML, en tanto, si bien generalizan mejor hacia afuera de la muestra, con resultados más robustos en sectores poco cubiertos, tienen una complejidad significativa y requieren capacidades técnicas para su implementación que, por lo general, no se encuentran en las administraciones públicas locales. Asi-

mismo, este tipo de técnicas posee el potencial de amplificar sesgos subyacentes en los datos empleados, lo cual puede derivar en un tratamiento inequitativo a los contribuyentes.

Por lo tanto, la elección de la técnica de valuación masiva automatizada tiene importantes implicancias en la implementación de las políticas públicas. Con base en esta conclusión emergen las siguientes preguntas:

¿Los formuladores de políticas necesitan estimaciones que reflejen la dinámica del mercado inmobiliario, aunque no sean muy detalladas y minuciosas a nivel individual, o precisan que los resultados tengan un grado de detalle significativo para que los contribuyentes entiendan los elementos que inciden en las valuaciones?

¿La administración pública dispone de presupuesto para generar las variables independientes a efectos de estimar modelos lineales o de ML, o prefiere destinar los recursos disponibles a intensificar el relevamiento de datos de mercado y aplicar una técnica geoestadística?

¿La administración pública local tiene las capacidades necesarias para implementar modelos de ML o cuenta con apoyo de universidades e instituciones de la sociedad civil para suplir potenciales déficits de capacidades técnicas?

Estas preguntas son indicativas de que no existe la “mejor técnica” para aplicar a los estudios de valuación. Por el contrario, su elección es una decisión de política pública, que debe considerar cuestiones relacionadas con presupuesto, capacidades estatales y regulaciones gubernamentales. En lo que sigue, esta guía apunta a generar conocimiento y herramientas para realizar valuaciones masivas con técnicas de ML de una forma relativamente simple, a fin de reducir las barreras para su implementación por parte de los gobiernos locales de América Latina y el Caribe (cuadro 2.1). No obstante, en el análisis de su implementación se debe tener en cuenta el resto de los factores mencionados antes, que hacen al espectro de sus posibilidades de aplicación.

Cuadro 2.1. Las técnicas de *machine learning* en los modelos de valuación masiva

Ventajas	Desventajas
<p>Tradeoff entre sesgo y varianza: En consecuencia, mayor precisión.</p> <p>Escalabilidad: Mejor desempeño a medida que se suma información.</p> <p>Replicabilidad: Posibilidad de aplicar un modelo entrenado en otra ciudad.</p> <p>Flexibilidad: Sin supuestos <i>a priori</i> sobre el tamaño de la muestra, o la distribución estadística o geográfica de los datos.</p> <p>Adaptabilidad: Capacidad de reflejar las variaciones caóticas del precio de los inmuebles en entornos urbanos segregados.</p>	<p>Complejidad: Frecuente déficit de capacidades estatales para su implementación.</p> <p>Cajas negras: Dificultades para explicar detalladamente el proceso que lleva a determinar un precio en particular.</p> <p>Ampliación de sesgos: Posibilidad de aumentar los sesgos contenidos en los datos, si estos no se tratan correctamente antes de ingresar a los modelos (por ejemplo, ingresar una muestra de inmuebles ofrecidos en el mercado sin descontar el margen de negociación).</p>

Fuente: Elaboración propia.

2.7. ¿Cómo aprende un algoritmo?

El proceso de aprendizaje de un algoritmo puede describirse como la construcción de un modelo matemático a partir de datos. En el caso de un AVM, los datos hacen referencia a una muestra de inmuebles en venta o vendidos, con el valor y las características correspondientes. A cada dato se le agrega información relativa a su localización y a las particularidades de su entorno. Esta información recibe el nombre de variables independientes. El valor de mercado de cada inmueble, en tanto, constituye lo que se conoce como variable dependiente.

El modelo que emerge de los datos se utiliza para identificar patrones, relaciones o reglas que permiten hacer predicciones o tomar decisiones sobre datos nuevos. Para los datos nuevos solo se cuenta con información sobre las variables independientes y se desconoce su valor de mercado.

El objetivo del algoritmo es, entonces, minimizar un error (o maximizar un acierto) al ajustar los datos a una función matemática o modelo. A través de un proceso iterativo, el algoritmo ajusta los parámetros del modelo para reducir el error mediante técnicas de optimización. El modelo se evalúa sobre un subconjunto de los datos originales, previamente separados de la muestra, o mediante un proceso de validación cruzada.¹⁰

Por lo tanto, a partir del análisis iterativo de los datos de mercado, el algoritmo “aprende” las relaciones entre el valor de mercado y las variables independientes y puede replicar estas relaciones para “predecir” el precio de los inmuebles en toda el área de estudio.

La lógica de uno de los algoritmos más simples, conocido como “árbol de clasificación y regresión” (CART, por sus siglas en inglés) (Breiman et al., 1984), que sirve de base a una amplia gama de algoritmos más complejos, permite ilustrar el concepto. El algoritmo CART tiene la misión de **dividir recursivamente** los datos de la muestra de mercado en dos subconjuntos que deben ser lo más diferentes posible entre sí y lo más similares hacia su interior. Para lograr esto, el algoritmo evalúa todas las variables independientes, en todo su rango de variación, e identifica valores que actúan como umbrales a partir de los cuales se generan estos dos subconjuntos. Como se indicó, cada uno de estos subconjuntos debe cumplir simultáneamente dos condiciones: **i)** maximizar la varianza del valor de mercado entre los dos grupos, **ii)** minimizar la varianza del valor hacia dentro de cada grupo. El proceso de separación de la muestra continúa de manera iterativa y genera cada vez más subgrupos.

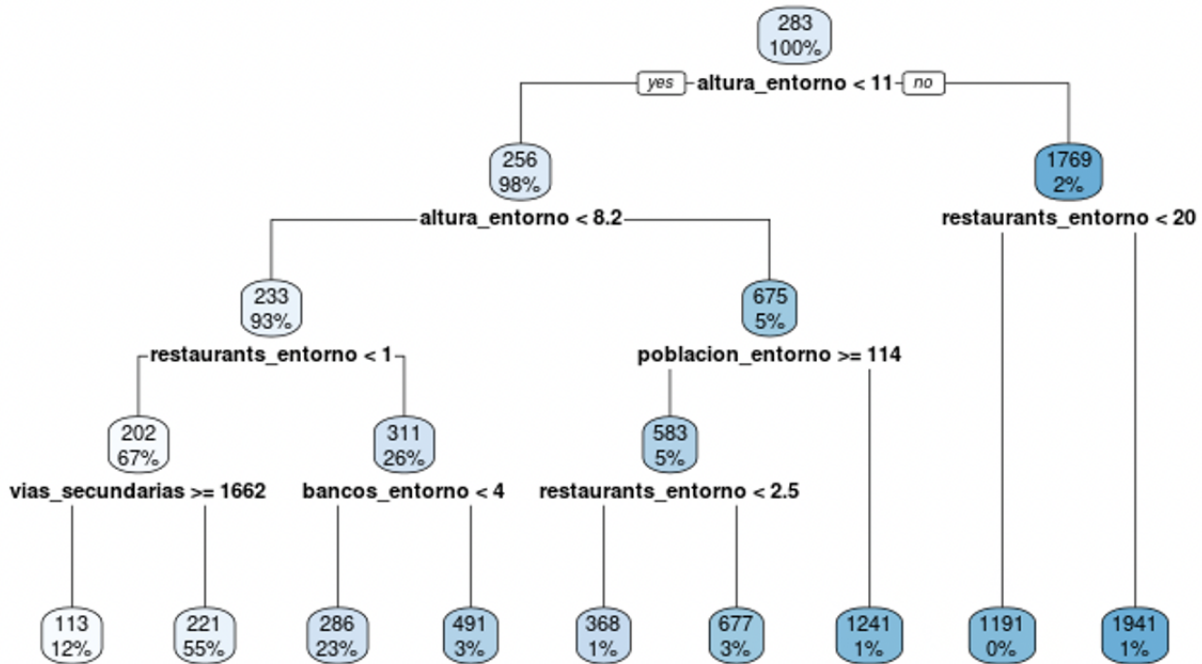
10 A grandes rasgos, un proceso de validación cruzada implica la eliminación recursiva de subconjuntos de la muestra para validar allí el desempeño del modelo. Por ejemplo, se separa aleatoriamente la muestra original en 10 subconjuntos de igual tamaño, se retira el primero de estos subconjuntos y se entrena el algoritmo con los nueve subconjuntos restantes. Luego, el desempeño del modelo se valida en el subconjunto extraído inicialmente. A continuación, se ingresa el subconjunto extraído y se retira el siguiente subconjunto, y el proceso continúa de manera iterativa hasta que todos los subconjuntos han sido validados fuera de la muestra original. Los detalles de los procesos de validación del error se tratan en el capítulo 4.

El gráfico 2.4 presenta la lógica de funcionamiento del algoritmo CART para un conjunto de datos de ejemplo correspondientes a la ciudad de Belo Horizonte, en Brasil. La muestra tiene un total de 7.041 lotes baldíos urbanos, sobre los que se conoce el valor de mercado de la tierra (variable dependiente), a los que se agregan las siguientes variables independientes:

- **altura_entorno:** Altura promedio de las edificaciones en un entorno de 500 metros.
- **bancos_entorno:** Cantidad de bancos en un entorno de 500 metros.
- **edificaciones_entorno:** Porcentaje del espacio ocupado por edificaciones en un entorno de 500 metros.
- **hoteles_entorno:** Cantidad de hoteles en un entorno de 500 metros.
- **plazas_entorno:** Proporción del espacio ocupado por espacios verdes (plazas o parques) en un entorno de 500 metros.
- **poblacion_entorno:** Suma de la población en un entorno de 500 metros.
- **restaurants_entorno:** Cantidad de restaurantes en un entorno de 500 metros.
- **vias_primarias:** Distancia en metros lineales a la vía primaria más cercana.
- **vias_secundarias:** Distancia en metros lineales a la vía secundaria más cercana.

Como se observa en el gráfico 2.4, el total de la muestra tiene un precio promedio de US\$283 por metro cuadrado. La misión del algoritmo consiste en analizar los rangos de variación de todas las variables independientes para identificar en cuál de ellas y a partir de qué valor es posible obtener dos subconjuntos muy diferentes entre sí, y muy similares, en términos del valor de la variable dependiente en cada uno de los elementos que los integran. El mejor corte para dividir este conjunto de datos, en términos de la suma de cuadrados de los residuos de cada espacio, se da en la variable que informa sobre la altura edificada en el entorno en un valor de 11 metros. A partir de este corte se conforman dos subconjuntos. Por un lado, todas las observaciones que tienen una altura edificada en el entorno menor de 11 metros (que incluye el 98% de las observaciones), con un precio promedio por metro cuadrado igual a US\$256. Por otro lado, todas las observaciones con una altura edificada en el entorno superior o igual a 11 metros (el 2% de la muestra), con un precio promedio igual a US\$1.769 por metro cuadrado. Este proceso continúa de manera iterativa hasta que se conforman las nueve hojas finales del árbol. Cada punto en el que se subdivide la muestra de manera recursiva recibe el nombre de “nodo de partición”.

Gráfico 2.4. Árbol de clasificación y regresión para una muestra de ejemplo en Belo Horizonte, Brasil



Fuente: Elaboración propia.

La estructura del árbol estimado permite realizar predicciones para toda el área urbana, de manera que todos los lotes urbanos recorran el camino correspondiente por el árbol según los valores de cada una de las variables independientes. De esta forma, por ejemplo, una parcela localizada en un sector donde la altura edificada en el entorno es de 9 metros, la población en el entorno es igual a 1.210 personas y los restaurantes en el entorno son 10 tendrá un valor por metro cuadrado predicho igual a US\$677. Si todas las parcelas de la ciudad se someten al mismo procedimiento, se obtiene la predicción del valor de mercado de la tierra para toda el área urbana.

El principal problema de este algoritmo es que lleva a la división del territorio urbano en nueve zonas de valor homogéneo. Es decir, a todas las parcelas que cumplen con los requisitos indicados en la estructura del árbol se les imputará el mismo valor por metro cuadrado. En términos técnicos, esta situación indica que el algoritmo se desempeña con un **sesgo elevado**.

De esta manera, por ejemplo, todas las parcelas urbanas que se encuentren en una zona con una altura efectiva en el entorno menor de 8,2 metros, más de un restaurante en el entorno y menos de cuatro bancos en el entorno tendrán un valor predicho igual a US\$286 por metro cuadrado. Esto implica que grandes partes de la ciudad tendrán como resultado la misma valuación, lo cual genera problemas de equidad relevantes en el caso de que estos valores se utilicen como bases imponibles para el impuesto a los inmuebles u otros gravámenes.

Además de desempeñarse con un elevado error evaluado fuera de la muestra, debido a la definición de grandes partes del espacio urbano con el mismo valor predicho, el proceso descrito es propenso a generar un **exceso de ajuste** (*overfitting*) en las predicciones, ya que el árbol resultante podría ser demasiado complejo y, en el extremo, cada hoja final del árbol coincidiría con un dato de la muestra. En consecuencia, el resto de las parcelas de la ciudad sería forzado a encajar en alguna de estas hojas, sin considerar las particularidades del mercado inmobiliario que no están siendo observadas en la muestra.

Para resolver estos problemas surge la idea de ensamblaje, que consiste en combinar diferentes modelos para reducir el sesgo y/o la varianza. Dos de los métodos de ensamblaje más utilizados son *bagging* (Breiman, 1996) y *boosting* (Freund, 1990; Schapire, 1990).

El algoritmo Random Forest (Breiman, 2001) se basa en generar varios árboles de regresión a partir de la misma muestra mediante una técnica de remuestreo conocida como *bootstrapping*, que consiste en retirar de la muestra una porción de los datos seleccionados de manera aleatoria y reemplazarlos por la replicación aleatoria de otros datos que permanecen en la muestra. Random Forest combina las predicciones de estos árboles a través del promedio condicional de las predicciones de cada modelo. Esta técnica se denomina agregación *bootstrap* o *bagging* y permite reducir el sesgo y la varianza del modelo original CART al mismo tiempo (Hastie, Tibshirani y Friedman, 2009). Además, Random Forest arroja predicciones más precisas al disminuir la correlación entre árboles. Esto se consigue con la selección aleatoria de un subconjunto de variables independientes para considerar en cada nodo de partición. Los hiperparámetros ajustados en el entrenamiento del algoritmo Random Forest son, entonces, el número de árboles, la cantidad mínima de observaciones en cada nodo y el número de variables independientes contempladas en cada nodo de partición de cada uno de los árboles.

Existen algunas variantes del algoritmo Random Forest. Una de ellas es conocida como Quantile Regression Forest (Meinshausen, 2006) y ha demostrado ser particularmente apta para la predicción del valor de los inmuebles (Carranza et al., 2022; Córdoba et al., 2021). Se trata de una técnica que combina las predicciones de los árboles por cuantiles, con lo cual produce una estimación empírica de la probabilidad de que la variable dependiente tenga un valor inferior o igual a un umbral determinado (según el valor de las variables independientes).

Por su parte, el algoritmo Gradient Boosting Model (Friedman, 2001) utiliza la técnica *boosting* (Freund, 1990; Schapire, 1990), combinada con *gradient descent* (Bottou, Curtis y Nocedal, 2018), en lugar de ensamblar las predicciones de diferentes árboles mediante *bagging*. La diferencia entre *bagging* y *boosting* se halla en la técnica de remuestreo aplicada en cada árbol. En *bagging* cada observación tiene la misma probabilidad de replicarse en la base de datos, mientras que en *boosting* se ponderan en función del error con el que fueron predichas. Cuanto mayor es el error, mayor es la probabilidad de que una observación permanezca en la muestra de árboles posteriores, a fin de enfatizar los casos difíciles. Además de ensamblar modelos mediante *boosting*, este algoritmo aplica el descenso de gradiente, donde los errores se consideran como un vector con dirección y no simplemente como una magnitud. Este vector apunta a una dirección en la que se minimiza la función de pérdida del algoritmo.

En la actualidad existe una amplia variedad de algoritmos de ML, como algoritmos basados en árboles (algunos de los cuales se han comentado brevemente aquí), redes neuronales de distinto tipo y técnicas particulares como máquinas de soporte vectorial, algoritmos de reducción de la dimensionalidad, entre muchos otros. El análisis que se presenta en los capítulos siguientes se centra en la aplicación de algoritmos basados en árboles, pero esto no invalida que se puedan alcanzar resultados muy buenos con algoritmos de naturaleza diferente. En el capítulo 4 se describe la secuencia lógica de un AVM completo, con aplicaciones realizadas para la ciudad de Cali, en Colombia; se revisan las principales características de cada etapa del proceso y se proveen herramientas de acceso abierto para replicar el modelo en otras ciudades de la región.



CAPÍTULO 3.

Valuación inmobiliaria como instrumento de gestión pública

HUÁSCAR EGUINO Y JUAN PABLO CARRANZA



Este capítulo presenta ejemplos de aplicación de los valores de los bienes inmuebles en cuatro ámbitos de políticas y gestión públicas: **i)** la administración del impuesto a la propiedad inmobiliaria, **ii)** los instrumentos de financiamiento urbano basados en la valorización inmobiliaria, **iii)** la planificación y la gestión urbanas, y **iv)** la evaluación de los impactos de los riesgos físicos en el precio de los inmuebles. Para facilitar la explicación de cada una de estas aplicaciones, el capítulo sigue la siguiente secuencia: en primer término, se hace una breve reseña de los fundamentos teóricos de la aplicación, luego se describen los principales pasos o etapas que se deben seguir y, finalmente, se presenta un ejemplo de uso.

3.1. Administración tributaria del impuesto a la propiedad

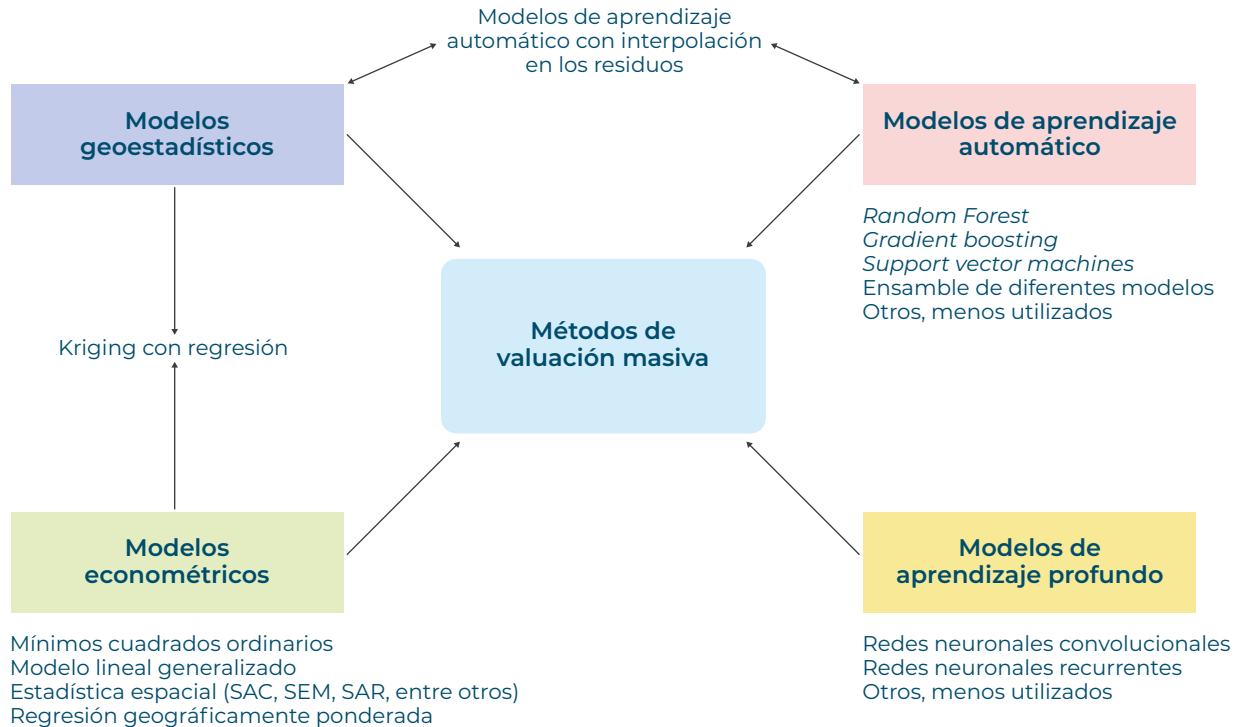
El valor de los inmuebles es la variable que permite relacionar el sistema de catastro y el sistema de administración tributaria, dos sistemas de gestión de gran importancia para los gobiernos subnacionales. Esto se debe a que el sistema de catastro incorpora tres tipos de datos: la información legal de la propiedad, la información de las características físicas de los inmuebles y la información económica, que incluye, esencialmente, el valor de los inmuebles. Por su parte, el sistema de administración del impuesto a la propiedad incluye la base tributaria (que corresponde al valor catastral de los inmuebles), la aplicación de las tasas impositivas a la base tributaria para determinar el monto del impuesto, y la gestión de los procesos de facturación, cobranza, fiscalización y atención al contribuyente. En este sentido, como el valor de los inmuebles es la variable que permite enlazar los sistemas de catastro y de administración tributaria, se convierte en un producto, en el primer caso, y en un insumo, en el segundo. Esta es la razón técnica por la cual si las valuaciones o tasaciones catastrales de las propiedades no reflejan correctamente la dinámica del mercado inmobiliario resulta imposible aprovechar el potencial tributario que tiene el impuesto a la propiedad inmobiliaria, además de inducir una serie de inequidades en la política fiscal.

En concreto, la estimación de los valores de mercado de los inmuebles con el modelo de valuación sistematizada (AVM) y su aplicación en la administración tributaria sigue el siguiente proceso:

- **Etapas 1.** Establecer una base de datos con información de las transacciones inmobiliarias registradas en la jurisdicción bajo estudio. Por lo general, estas bases de datos se construyen con información que proviene de dos fuentes: **i)** los informes de las entidades que hacen seguimiento del mercado inmobiliario, como institutos de estadística, centros de investigación, asociaciones de desarrolladores urbanos, corredores inmobiliarios y observatorios del mercado inmobiliario; **ii)** información recopilada por medio de minería de datos de las transacciones inmobiliarias disponibles en los sitios web especializados. En esta etapa es recomendable recabar la información de los valores del suelo y de las construcciones por separado. Para obtener información sobre el valor del metro construido se pueden usar las tablas de valores que recopilan las cámaras de constructores o desarrolladores locales.
- **Etapas 2.** Depurar la base de datos. Una vez que se cuenta con la base de datos, se procede a depurarla por medio de protocolos o algoritmos que permiten identificar y eliminar datos anómalos o incompletos.

- **Etapa 3.** Estimar el valor del suelo con modelos econométricos, geoestadísticos o de inteligencia artificial. A modo ilustrativo, el gráfico 3.1 presenta las principales metodologías y técnicas para realizar estas estimaciones: modelos geoestadísticos, modelos econométricos, modelos de aprendizaje automático y modelos de aprendizaje profundo.
- **Etapa 4.** Estimar el potencial tributario del impuesto inmobiliario. Una vez que se cuenta con los resultados del modelo de valuación automatizada (AVM) del suelo, se le adiciona el valor de las construcciones. De esa forma se obtiene la base tributaria sobre la que se aplicará la tasa impositiva. El potencial tributario es el resultado de aplicar la tasa impositiva vigente a la base tributaria.
- **Etapa 5.** Ajustar los valores catastrales de acuerdo con los resultados del AVM y la política tributaria vigente. Una vez que se tiene la estimación del potencial tributario, se procede a ajustar los valores catastrales según la tipología de bienes establecida en las normativas locales. En esta etapa se aplican las disposiciones de la política tributaria sobre, entre otros, los siguientes aspectos: **i)** uso de tasas diferenciadas, **ii)** exenciones y deducciones y **iii)** métodos de depreciación de la propiedad. Como los ajustes de los valores catastrales pueden ser significativos y generar resistencia en los contribuyentes, es recomendable adoptar una estrategia de ajuste de los valores que responda a la dinámica de la economía local (por ejemplo, se pueden realizar ajustes progresivos) y/o las preferencias de los responsables de la política tributaria local.

Gráfico 3.1. Métodos y técnicas de valuación masiva de inmuebles



Fuente: Elaboración propia.

SAC: Spatial Autoregressive Confused model; SAR: Spatial Autoregressive model; SEM: Spatial Error model.

Recuadro 3.1. Ejemplo de aplicación de valuación masiva de inmuebles en la administración del impuesto a la propiedad: Prefeitura municipal de Fortaleza, Brasil

La ciudad de Fortaleza, capital del Estado de Ceará, se encuentra en la región noreste de Brasil y tiene una población de más de 2,6 millones de habitantes, lo cual la convierte en la quinta ciudad más importante del país. La Secretaría Municipal de Finanzas de Fortaleza (Sefin) tiene bajo su responsabilidad la administración del catastro y del impuesto a la propiedad, así como la actualización del valor de los inmuebles. En 2021 se realizó una actualización de la base tributaria y una estimación del potencial tributario de acuerdo con el proceso descrito para estimar los valores de mercado de los inmuebles mediante un modelo de valuación automatizada y aplicarlos en la administración tributaria. En concreto, el proceso funcionó de la siguiente manera:

Establecimiento de la base de datos con información de las transacciones inmobiliarias. El municipio tiene un observatorio urbano de valores inmobiliarios (OUV) que funciona desde 2013. A finales de 2020, el OUV contaba con una base de datos de 90.000 observaciones, que constituyeron la base de datos del proyecto de estimación del potencial tributario del impuesto a la propiedad. De esta base se seleccionaron las observaciones de venta de lotes baldíos, que sumaron cerca del 10% del total. Paralelamente, se estableció la base de datos del valor de las construcciones con información proveniente de la cámara de construcción local.

Depuración de la base de datos. Se procedió a depurar la base de datos, para lo cual se eliminaron las observaciones con información incompleta o anómala, tanto del total de transacciones como del número de lotes baldíos vendidos. Este proceso dio como resultado una reducción de la muestra cercana al 10%.

Estimación del valor del suelo con modelos de IA. Se aplicaron cinco algoritmos de aprendizaje automático (ML) (Random Forest, Extra-Trees, XGBoost, CATBoost y LightGBM) y se infirió el valor de mercado de cada inmueble registrado en el catastro municipal. La muestra se dividió aleatoriamente en un 80% para entrenamiento y un 20% para prueba. Las medidas de desempeño de los cinco modelos de ML resultaron muy cercanas, aunque el modelo LightGBM presentó un nivel ligeramente superior, por lo cual se utilizó para estimar el potencial tributario y elaborar el nuevo mapa de valores del suelo.

Estimación del potencial tributario del impuesto inmobiliario. Una vez obtenidas las estimaciones de los valores del suelo y de las construcciones, se procedió a determinar el potencial tributario del impuesto a la propiedad. En el caso de Fortaleza, esta estimación se hizo por tipología de viviendas, según lo establecido en el código tributario. El potencial se definió como la suma de todos los valores inmobiliarios de la ciudad, estimados con la información de mercado y los modelos descritos en el punto anterior. Los resultados indican que la tasa de valuación era del 40% de lo observado en el mercado y que la recaudación del impuesto a la propiedad podría crecer un 145%.

Ajuste de los valores catastrales y aplicación de la política tributaria vigente. Finalmente, se procedió a ajustar los valores catastrales a partir de las estimaciones del valor potencial de mercado de los bienes inmuebles y se determinó el monto del impuesto en función de todas las exenciones, deducciones o rebajas establecidas en el código tributario. Las autoridades decidieron ajustar los valores catastrales solo un 70% respecto del valor de mercado, en consideración de la potencial resistencia de los contribuyentes frente a ajustes de mayor magnitud.

Fuente: Eguino H. y D. Erba. 2024. Mercado inmobiliario e impuesto predial: aplicaciones de técnicas de valuación masiva. Washington, D.C.: BID.

3.2. Instrumentos de financiamiento urbano basados en los impuestos y las contribuciones por mejoras

En un contexto de crecientes restricciones en las finanzas públicas, los instrumentos de captura de valor pueden ser una opción atractiva para financiar la demanda de infraestructuras locales. En efecto, la captura del incremento del valor de las propiedades es un medio efectivo para que los gobiernos subnacionales obtengan ingresos adicionales para financiar las mejoras realizadas en la infraestructura y los servicios públicos. Además, la variedad y la versatilidad de los instrumentos de captura de valor los hacen suficientemente flexibles como para ser aplicados en diversos marcos institucionales y contextos socioeconómicos.

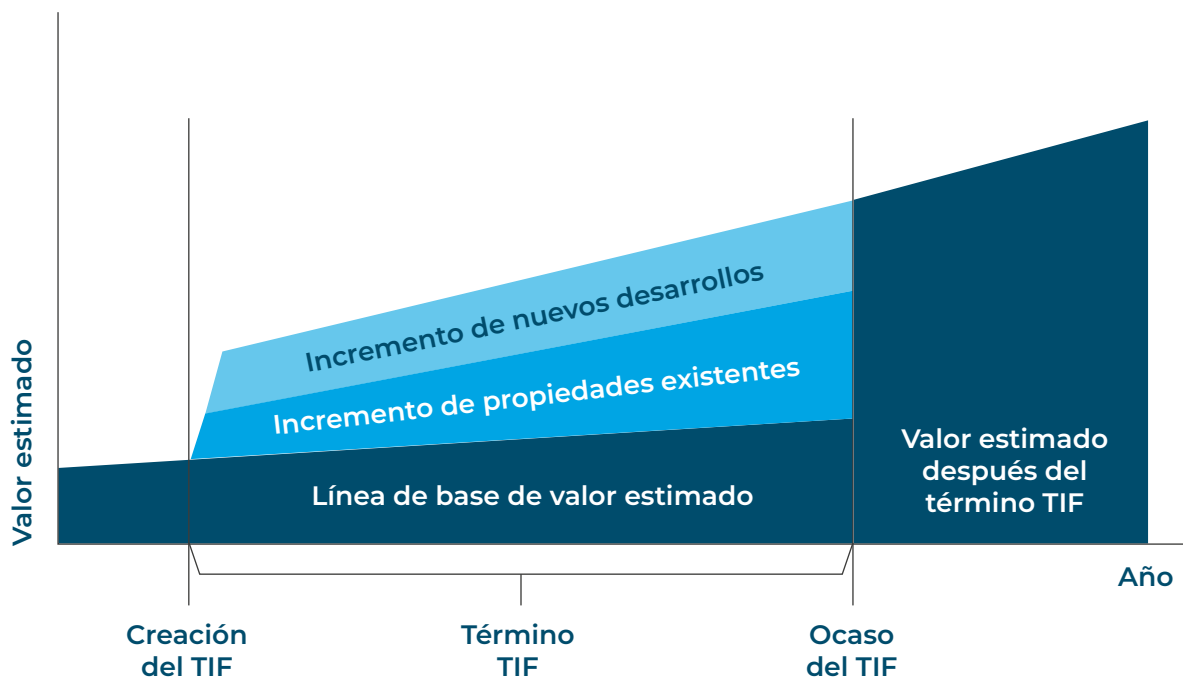
Un aspecto importante para que el uso de este tipo de instrumentos resulte exitoso es que los propietarios conozcan los incrementos del valor de los inmuebles y tengan información sobre los costos y los beneficios asociados a las intervenciones públicas. En otras palabras, es esencial que los propietarios sepan el valor de mercado de sus propiedades, antes y después de un proyecto de mejoras, y que las estimaciones realizadas sobre los incrementos de los valores de los inmuebles reflejen correctamente y de manera objetiva la dinámica del mercado.

Los instrumentos de captura de la valorización inmobiliaria se dividen en dos categorías. La primera agrupa las herramientas de tipo fiscal y comprende el incremento de los impuestos por una intervención pública (TIF, por sus siglas en inglés) y las contribuciones por mejoras. La segunda categoría abarca los instrumentos urbanísticos, como las contribuciones de los desarrolladores o promotores inmobiliarios, la venta de derechos de desarrollo, la gestión estratégica del patrimonio inmobiliario y los reajustes regulatorios. En esta sección se describen las herramientas de tipo fiscal (OCDE, 2022).

3.2.1. Financiamiento basado en el incremento del impuesto a la propiedad inmobiliaria

Este instrumento permite destinar el incremento del impuesto que produce una intervención urbana al pago de un crédito o de un financiamiento. Se utiliza ampliamente en países con sistemas de financiamiento municipal avanzados, como Estados Unidos y Reino Unido. También se registran casos de uso más recientes en otras regiones del mundo, incluso en algunos países de América Latina y el Caribe (López-Ghio, Salazar Echavarría y Schloeter, 2020; Blanco Blanco, Fretes Cibils y Muñoz Miranda, 2016). La aplicación de la herramienta de financiamiento basado en el incremento de impuestos es relativamente sencilla y se basa en una serie de pasos o etapas claramente establecidos (gráfico 3.2).

Gráfico 3.2. Financiamiento por incremento del impuesto a la propiedad



Fuente: López-Ghio, R, C. Salazar Echavarría y L. Schloeter. 2020. Tax Increment Financing for Urban Projects: An Alternative to Fund Infrastructure in Latin American Cities? The Case of Medellín's Carrera 80 Tram Project. Washington, D.C.: BID.
TIF: financiamiento por incremento de impuestos, por sus siglas en inglés.

A continuación, se explica, de manera simplificada, cada una de las etapas de la aplicación de la herramienta de financiamiento basado en el incremento de impuestos, para luego presentar un caso práctico.

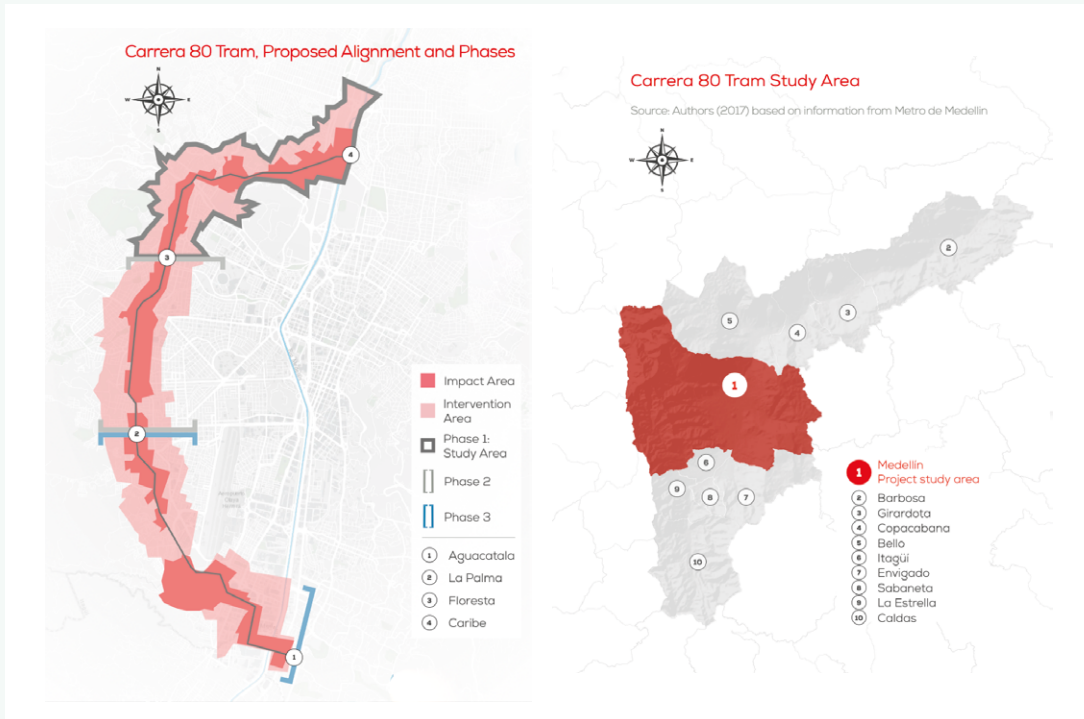
- **Etapa 1.** Definir el proyecto urbano y su zona de intervención o incidencia. El promotor del proyecto debe definir su alcance y la zona de incidencia del proyecto, también llamada “distrito TIF”. Dentro del distrito TIF, el valor catastral que se usará como contrafactual o línea de base corresponde al valor de mercado de las propiedades. Este valor se puede ajustar en función de la tendencia del mercado inmobiliario o la dinámica económica local. La determinación de la extensión real de la zona de incidencia puede presentar algunos inconvenientes, puesto que se trata de un proceso que a menudo es subjetivo, en el que se fija el área de manera discreta o sin considerar la dinámica del mercado inmobiliario local.¹¹ Esto puede generar resistencias en los contribuyentes. En el capítulo 4 se describen algunas herramientas que contribuyen a establecer la extensión de estas zonas de manera objetiva.
- **Etapa 2.** Estimar el impacto del proyecto. Una vez que se cuenta con el valor catastral de base y se establece su incremento esperado, se procede a estimar el impacto del nuevo proyecto. En los casos de aplicación más elaborados, estas estimaciones se hacen con modelos de valuación masiva y se simula el impacto de las nuevas intervenciones. El impacto del proyecto corresponde a la diferencia entre el valor de las propiedades con proyecto y sin él. Es importante que la metodología aplicada para determinar el impacto del proyecto sobre el valor de las propiedades sea bien comunicada y tenga legitimidad en el universo de contribuyentes involucrados. En el capítulo 4 se desarrolla una serie de instrumentos que apuntan a facilitar esta tarea.
- **Etapa 3.** Definir la estrategia de financiamiento del proyecto con base en el incremento de impuestos. Esta etapa consiste en usar el flujo de impuestos adicionales que generará el proyecto para definir la capacidad de endeudamiento y de pago del gobierno subnacional, a fin de poder acceder al financiamiento. Como parte de la estrategia de financiamiento se debe precisar el período de vigencia del distrito TIF, que usualmente está asociado a las condiciones del financiamiento (en especial, a los períodos de amortización y gracia).

¹¹ Por ejemplo, un proyecto urbano de modernización de luminarias en las calles puede estar sujeto a un sector claramente delimitado, pero las externalidades que se generan en términos de una mayor seguridad pueden ser absorbidas por el valor de los inmuebles en un sector más amplio que el beneficiado de forma directa.

Recuadro 3.2. Ejemplo de financiamiento basado en el incremento del impuesto a la propiedad: Proyecto de tranvía Carrera 80 en Medellín, Colombia

Definición del proyecto urbano y su zona de incidencia. El proyecto del tranvía Carrera 80 consiste en una nueva línea de tránsito de aproximadamente 14 km de longitud con 20 estaciones a lo largo del recorrido (mapa R3.2.1). La línea del tranvía se planificó como un sistema de transporte público de última generación, que transformaría el corredor occidental de Medellín con un desarrollo inmobiliario considerable. El proyecto incorporó nuevos espacios abiertos, equipamientos públicos, líneas peatonales y ciclovías. La naturaleza de la intervención presentaba oportunidades sin precedentes para el uso de mecanismos de captura de valor, lo cual convertía al proyecto en una opción ideal para financiar mediante el instrumento de financiamiento por incremento de impuestos.

Mapa R3.2.1. Área de incidencia del proyecto de tranvía Carrera 80 en Medellín, Colombia



Estimación del impacto del proyecto. Se realizó un estudio del mercado inmobiliario para determinar el impacto del proyecto del tranvía en su área de incidencia (un radio de 500 metros). Los resultados indicaron que el proyecto daría cabida a, por lo menos, 3,1 millones de m² de nuevos desarrollos entre 2020 y 2044, con el consiguiente incremento de los ingresos del municipio por impuestos a la propiedad inmobiliaria.

Definición de la estrategia de financiamiento. Finalmente, se estimó cuántos ingresos adicionales por impuestos sobre la propiedad recibiría la ciudad de las propiedades inmobiliarias nuevas y revalorizadas. Se estableció que el municipio de Medellín podría recaudar casi US\$571 millones en ingresos incrementales por impuestos a la propiedad entre 2020 y 2044. Además, se concluyó que el incremento de los ingresos por impuestos sobre la propiedad podría respaldar entre US\$88 millones y US\$224 millones en deuda, según las condiciones crediticias, que cubrirían entre el 21% y el 55% de los gastos de capital (*capex*) de la primera fase del proyecto del tranvía.

Fuente: López-Ghio, R, C. Salazar Echavarría y L. Schloeter. 2020. Tax Increment Financing for Urban Projects: An Alternative to Fund Infrastructure in Latin American Cities? The Case of Medellín's Carrera 80 Tram Project. Washington, D.C.: BID.

3.2.2. Contribuciones por mejoras

Las contribuciones por mejoras son una carga obligatoria impuesta a los propietarios de un grupo de bienes inmuebles seleccionados para financiar, en su totalidad o parcialmente, el costo de una mejora o de servicios específicos (Borrero Ochoa, 2011). En América Latina y el Caribe (ALC) se conocen como contribuciones por mejoras o contribuciones de valorización.

Las contribuciones por mejoras se pueden recaudar de forma *ex ante* o *ex post*, en función de las necesidades de financiamiento del proyecto. El objetivo de las contribuciones por mejoras, como en cualquier otro instrumento de captura de plusvalía, es que los gobiernos capten los incrementos del valor de las propiedades resultantes de las inversiones públicas y usen los recursos recibidos para cubrir los costos del proyecto.

Al igual que en el caso del financiamiento basado en el incremento del impuesto a la propiedad inmobiliaria, uno de los principales retos para la aplicación de las contribuciones por mejoras es determinar el área de impacto, definida como la zona de influencia o el espacio geográfico donde el proyecto provoca cambios en el valor de las propiedades. Otro desafío importante es estimar los incrementos del valor de las propiedades en las zonas de intervención o incidencia de un proyecto. Esas tareas tienen relevancia crucial para que las contribuciones se distribuyan correctamente entre los beneficiarios del proyecto. Las dos se pueden realizar mediante modelos de valuación masiva y técnicas de análisis de datos.

El procedimiento para aplicar las contribuciones por mejoras mediante AVM comprende las siguientes etapas:

- **Etapas 1.** Definir el área de influencia del proyecto y dividirla en zonas homogéneas con características similares en cuanto a normativa urbanística, estructura espacial y uso del suelo, entre otras. Esta información permitirá distribuir correctamente los costos del proyecto entre los propietarios de los inmuebles ubicados en la zona de influencia.
- **Etapas 2.** Utilizar modelos de valuación de las propiedades para establecer un escenario con la situación actual y otro con los valores esperados a partir de la implementación del proyecto. La diferencia entre los valores con proyecto y sin él permite establecer el valor máximo de la contribución por valorización.
- **Etapas 3.** Determinar los costos del proyecto en las zonas de influencia. Esta información proviene de los estudios de factibilidad económica y financiera del proyecto. Por lo general, se dividen en costos de inversión y de operación, y la contribución por mejora solo se aplica a una parte de los costos de inversión.
- **Etapas 4.** Definir el monto de las contribuciones por mejoras que se impondrá a los propietarios de los inmuebles ubicados en el área de influencia. Estas tasas se calculan con base en tres parámetros: **i)** el monto de la inversión requerida por el proyecto (etapa 3), **ii)** el valor de las propiedades que puede atribuirse al proyecto (etapa 2) y **iii)** la asequibilidad del gravamen (es decir, la capacidad de pago de los propietarios).

Recuadro 3.3. Ejemplo de aplicación de contribuciones por mejora en Bogotá, Colombia

Bogotá, la capital de Colombia, es una ciudad de 7,9 millones de habitantes, con una superficie de 1.587 km², situada en una sabana plana de la cordillera de los Andes. La administración de las contribuciones por mejoras es responsabilidad del Instituto de Desarrollo Urbano (IDU), que también se encarga de definir los proyectos que se financiarán con este instrumento. La contribución por mejoras (o contribución por valorización, como se conoce en Colombia) se aplica a todas las propiedades afectadas por el proyecto (o conjunto de proyectos). Las etapas que se siguen para implementar las contribuciones por valorización son las siguientes:

Determinación del área de influencia. Para recaudar la tasa de contribución por mejoras, el IDU establece el área de influencia, es decir, la zona en la que el proyecto de construcción (por ejemplo, una carretera) genera beneficios. Los criterios para establecer las áreas de influencia incluyen variables como la proximidad y la accesibilidad a la nueva infraestructura o el nuevo servicio, entre otras.

Medición de los beneficios del proyecto. Los impactos del proyecto sobre el valor de las propiedades (beneficio) se estiman mediante variables como la mayor cercanía a las nuevas infraestructuras o los nuevos servicios (por ejemplo, la distancia a la carretera), la accesibilidad a los servicios en el área de influencia u otras variables específicas asociadas a las particularidades del proyecto.

Distribución de la contribución. En general, la proporción del costo del proyecto que se distribuye entre los propietarios beneficiarios la define el gobierno responsable. Luego, se establece el monto de la contribución de cada propietario en función de la valorización que tendrá su propiedad como resultado del proyecto.

El cuadro R3.3.1 presenta un listado de proyectos que fueron financiados mediante contribuciones por valorización en la ciudad de Bogotá.

Cuadro R3.3.1. Proyectos financiados con contribuciones por valorización en Bogotá

Proyectos	Fecha de aprobación	Fecha de recaudación	Monto recaudado (Millones de US\$) ^a
Contribución general	1993	1993	106,2
Formar Ciudad - Fase 1	1995	1996-1998	351,9
Formar Ciudad - Fase 2	2001	2002	55,9
Acuerdo 180 Valorización local - Fase 1	2005	2007-2010	260,2
Valorización local - Fase 2	2005	2009	265,7
Valorización local - Fase 3	2005	2012	262,1 ^b
Valorización local - Fase 4	2005	2015	85,5 ^c

^a Pesos colombianos convertidos a la tasa de cambio promedio del 2009: US\$1 = Col\$ 2.000.

^{b, c} Monto que se espera recaudar.

Fuente: Elaboración propia con base en Borrero Ochoa O. 2011. Betterment Levy in Colombia: Relevance, Procedures, and Social Acceptability. *Land Lines*, 23(2):14-19.

3.3. Planificación y gestión urbana

Existen múltiples aplicaciones de los modelos de valuación inmobiliaria orientados a la gestión urbana, en el ámbito público y en el privado. Como se adelantó en el capítulo 2, en el ámbito privado, los modelos de valuación masiva automatizada son ampliamente utilizados por las compañías de seguros de bienes inmuebles y los desarrolladores urbanos, para establecer las primas que deben pagar los propietarios o para determinar el impacto de un nuevo proyecto en el *stock* inmobiliario. El uso de los AVM por parte del sector público también es muy significativo, aunque su difusión en los países menos desarrollados es escasa. A continuación, se describen las aplicaciones más conocidas.

3.3.1. Impacto de los cambios de las normas urbanísticas en el potencial constructivo

Los modelos de valuación masiva permiten simular el impacto de un cambio en las normas de uso de suelos y de construcción sobre el valor de los inmuebles. Esto es particularmente relevante en ciudades que siguen una política de densificación como forma de reducir los costos de las infraestructuras y los servicios urbanos.

Esta aplicación consiste en usar las técnicas de los AVM para evaluar el impacto de cambios regulatorios sobre el potencial constructivo y, consecuentemente, sobre el *stock* inmobiliario del área sujeta a cambios normativos. De manera simplificada, los pasos que se siguen para evaluar el impacto de los cambios de las normas urbanísticas en el potencial constructivo son las siguientes:

- **Etapas 1.** Establecer el área sujeta al cambio normativo. Esto incluye determinar la zona afectada por la modificación normativa (por ejemplo, zona urbana o semiurbana) y definir el objetivo de gestión urbana que se pretende alcanzar (densificación, preservación, crecimiento sostenible, etc.).
- **Etapas 2.** Definir y detallar el cambio regulatorio. Implica establecer y precisar los cambios normativos en la densidad permitida, la altura máxima, los usos autorizados o restringidos, y los coeficientes de ocupación y de utilización del suelo, entre otros.
- **Etapas 3.** Analizar el impacto del cambio normativo en el potencial constructivo. Estimar el impacto del cambio normativo sobre el valor actual y potencial de los inmuebles, para lo cual se pueden usar modelos de valuación automatizada.
- **Etapas 4.** Monetizar (o vender) el potencial constructivo para financiar nuevas infraestructuras y servicios públicos en las zonas de desarrollo o en otras zonas priorizadas por el gobierno local.

Recuadro 3.4. Ejemplo de uso de modelos de valuación automatizada en el análisis de cambios en normas urbanísticas de Curitiba, Brasil

El financiamiento urbano basado en cambios regulatorios es una estrategia con la cual los gobiernos municipales generan recursos para realizar inversiones a partir de las modificaciones realizadas en las normativas urbanísticas. En Brasil hay varios ejemplos de aplicación de este tipo de instrumentos, como es el caso de Curitiba.

El pago por los permisos de derechos de construcción (*outorga onerosa do direito de construir*, OODC) es un instrumento que les permite a los propietarios de inmuebles edificar por encima de los índices urbanísticos establecidos, para lo cual deben realizar una solicitud y un pago al municipio. El proceso de solicitud y pago por el derecho de construir con reformas regulatorias funciona de la siguiente manera:

1. Identificación del área de intervención y cambio regulatorio requerido por el proyecto. En el caso de la construcción de un edificio de uso mixto (comercial y residencial) en un terreno ubicado en la zona central, donde el coeficiente de aprovechamiento permite la construcción de hasta 1,0 veces el área del terreno, el cambio regulatorio solicitado consiste en aumentar el coeficiente de uso (por ejemplo, hasta 2,5 veces) para viabilizar el proyecto.

2. Aprobación del cambio regulatorio y cálculo/pago de la contrapartida (etapas 3 y 4). El desarrollador presenta una solicitud de modificación regulatoria, que es evaluada en función de la normativa de desarrollo urbano vigente en el municipio. Una vez que se aprueba el cambio regulatorio, el municipio estima el monto de la contrapartida financiera que deberá pagar el desarrollador, de acuerdo con los siguientes pasos: **i)** determina el área adicional que será construida a partir del cambio regulatorio, **ii)** establece el valor de mercado del metro cuadrado construido y **iii)** aplica el coeficiente o la tasa al valor de mercado de superficie prevista en el proyecto, y ejecuta el cobro.

En el caso de Curitiba, los fondos recaudados a través de la OODC se destinan a mejoras de la infraestructura urbana, como ampliación de vías, mejoras en el transporte público y desarrollo de áreas verdes.

Fuente: Elaboración propia con base en Furtado F. y C. Maleronka. 2023. A outorga onerosa e a gestão pública dos direitos de construir. En: Diálogos para uma Política Nacional de Desenvolvimento Urbano: desafios do financiamento urbano e da governança interfederativa: volume 2. Brasília: IPEA.

3.3.2. Evaluación de proyectos de desarrollo urbano

El análisis costo-beneficio (ACB) es un poderoso instrumento para seleccionar proyectos que proporcionan la mayor rentabilidad financiera y socioeconómica. Una forma de evaluar los proyectos de desarrollo urbano es determinar su impacto sobre la valorización inmobiliaria. En concreto, es posible usar las técnicas de los AVM para simular el impacto (*ex ante* o *ex post*) de un proyecto de desarrollo urbano sobre la apreciación/depreciación de los inmuebles ubicados en el área de influencia. Este tipo de evaluaciones siempre es complementario al ACB y resulta especialmente útil para establecer la viabilidad o el impacto de los proyectos de renovación urbana, rehabilitación de áreas deterioradas, transporte público, mejoramiento de barrios y desarrollos urbanos orientados al transporte.¹²

La evaluación de proyectos de desarrollo urbano con técnicas de AVM se basa en comparar el valor de los inmuebles antes y después de la intervención planeada. De esa forma, se puede estimar la variación en el valor inmobiliario que genera el proyecto (beneficio) y compararla con el monto de las inversiones previstas (costo). Si la evaluación es previa a la realización del proyecto (evaluación *ex ante*), este se considera rentable si el valor inmobiliario generado es superior al monto de la inversión prevista. Si la evaluación es posterior a la implementación del proyecto (evaluación *ex post* o de impacto), el análisis permite establecer si las inversiones provocaron los incrementos del *stock* inmobiliario previstos en la fase de diseño.

La evaluación de proyectos urbanos con modelos de valuación masiva atraviesa las siguientes etapas:

- **Etapas 1.** Establecer el área de incidencia del proyecto y desarrollar un AVM con técnicas econométricas, geoestadísticas, modelos de aprendizaje automático o modelos de aprendizaje profundo.
- **Etapas 2.** Definir el escenario sin proyecto (contrafactual). Consiste en estimar el valor del *stock* inmobiliario en la zona de intervención, en función de las tendencias de crecimiento o contracción del mercado de bienes inmuebles antes de la realización del proyecto. Este escenario proporciona el valor del *stock* inmobiliario que será afectado por el proyecto.
- **Etapas 3.** Evaluar el impacto del proyecto en el valor del *stock* inmobiliario. Implica determinar la apreciación inmobiliaria que el proyecto generará en su área de influencia. Usualmente, para esto se compara el valor del *stock* inmobiliario en el área de intervención antes del proyecto con el que tiene después. La relación entre el valor inmobiliario generado por el proyecto y el monto de la inversión constituye el indicador de rentabilidad del proyecto.

¹² Las evaluaciones de proyectos urbanos que se basan en la valorización inmobiliaria son particularmente útiles para las empresas desarrolladoras y los gobiernos locales, pero no sustituyen el ACB, que es la técnica correcta para determinar la rentabilidad del proyecto desde la perspectiva socioeconómica o financiera.

Recuadro 3.5. Ejemplo de evaluación de proyectos de desarrollo urbano con base en la valorización inmobiliaria en Bogotá, Colombia

El sistema de buses de tránsito rápido de Bogotá (Transmilenio), Colombia, es un caso destacado de evaluación *ex post* del impacto de nuevas infraestructuras en los valores inmobiliarios. El proyecto, iniciado en 2000, fue diseñado con el objetivo de mejorar la movilidad urbana mediante el uso de autobuses de alta capacidad en carriles de tránsito dedicados o exclusivos. El alcance del proyecto incluyó la construcción de un sistema BRT (*bus rapid transit*) de más de 100 km en toda la ciudad de Bogotá.

La evaluación del proyecto consistió en estimar el impacto del sistema Transmilenio en el valor inmobiliario en las zonas de influencia. La evaluación se realizó con varias técnicas de análisis que incluyeron modelos de valuación masiva, econometría espacial, análisis de diferencias en diferencias, entre otras. Los resultados permitieron establecer que el proyecto tuvo los siguientes efectos:

- **Impacto positivo sobre la valorización de los inmuebles.** Las propiedades ubicadas a menos de 500 metros de la estación sujeta a la evaluación aumentaron su valor entre un 6% y un 20%. Las variables independientes fueron la accesibilidad a la estación del BRT, la fase de implementación del proyecto y las condiciones del contexto (nivel socioeconómico del área, entre otras).
- **Impacto positivo en las zonas de menores ingresos.** Los inmuebles ubicados en zonas de menores ingresos registraron una mayor apreciación relativa, de entre el 15% y el 20%, en función de la cercanía a la estación.
- **Efecto anticipado.** La suba en los valores de las propiedades empezó antes de la construcción y continuó durante las obras. Se observó que las expectativas de mejor accesibilidad se incorporaron rápidamente en el valor de los inmuebles.
- **Efecto derrame.** Los impactos del proyecto en el valor de los inmuebles se extendieron a zonas ubicadas a más de 500 metros de las estaciones, debido a la mejor accesibilidad a las estaciones de transporte, los menores tiempos de viaje y una mayor presión para el desarrollo urbano en áreas cercanas al BRT.

Fuente: Elaboración propia con base en Rodríguez D. y E. Vergel. 2013. BRT-Oriented Development in Quito and Bogotá. Lincoln Institute of Land Policy. Documento de trabajo No. WP14DR1. Cambridge: Lincoln Institute of Land Policy; Perdomo-Calvo J., C. Mendoza-Álvarez, J. Mendieta-López y A. Baquero-Ruiz. 2007. Study of the Effect of the TransMilenio Mass Transit Project on the Value of Properties in Bogotá, Colombia. Lincoln Institute of Land Policy. Documento de trabajo No. WP07CA1. Cambridge: Lincoln Institute of Land Policy.

3.3.3. Impacto de los riesgos climáticos en el valor de las propiedades

Otra de las aplicaciones de los AVM consiste en estimar y cuantificar las pérdidas en el valor de las propiedades ocasionadas por los eventos climáticos o ambientales extremos, y simular el impacto que tendrían las intervenciones orientadas a reducir los riesgos. Estas aplicaciones son particularmente importantes en zonas urbanas sujetas a huracanes, inundaciones, deslizamientos, incendios y otros fenómenos físicos que generan pérdidas económicas y humanas, y afectan el valor de los inmuebles (García et al., 2019).

Si bien este tipo de análisis de riesgos es bastante conocido en el mundo de las compañías aseguradoras, su aplicación es menos frecuente en el sector público, en especial en los gobiernos municipales, que también pueden beneficiarse de su uso. En efecto, la evaluación de impactos inmobiliarios de los eventos climáticos o ambientales extremos (o de los desastres físicos) no solo sirve para cuantificar posibles pérdidas económicas, sino también para analizar la sostenibilidad fiscal de los gobiernos subnacionales y planificar las inversiones en resiliencia urbana.

De manera sintética, la implementación de este tipo de análisis recorre los siguientes pasos:

- **Etapas 1.** Definir el riesgo cuyo impacto se quiere evaluar. Es fundamental tener claridad sobre el tipo de riesgo que se quiere evaluar y conocer las limitaciones de los modelos y la información disponible.
- **Etapas 2.** Recopilar información del mercado inmobiliario y los riesgos frente a eventos climáticos extremos en el área de análisis. Esto implica contar con información actual y prospectiva sobre el mercado inmobiliario y mapas de riesgo para las principales fuentes de amenazas a las que está expuesta la zona de análisis.
- **Etapas 3.** Desarrollar los modelos que relacionen los factores de riesgo y el valor de los inmuebles. Si bien estos modelos pueden ser de alta complejidad, debido a la multiplicidad de posibles riesgos y a la propia dinámica de crecimiento urbano, es recomendable simplificarlos a fin de otorgar mayor importancia a las amenazas climáticas con más recurrencia o relevancia en la zona de análisis. Por ejemplo, se puede modelar solo el impacto de los riesgos de inundaciones o de incendios.
- **Etapas 4.** Analizar el impacto de los riesgos climáticos sobre el valor de los inmuebles. Este análisis se basa en la construcción de escenarios que simulan los impactos de eventos climáticos de distinta intensidad. Los escenarios permiten establecer el modo en que los factores de riesgo impactan sobre el valor de las propiedades y lo que sucedería con dichos valores si se aplicaran estrategias para eliminarlos o mitigarlos. En otras palabras, este tipo de modelos permite evaluar el impacto que tendría sobre el valor de los inmuebles la implementación de estrategias como prohibir la construcción en áreas de riesgo, realizar inversiones en proyectos de mitigación o establecer incentivos fiscales como los créditos tributarios para gastos dirigidos a la reducción de los factores de riesgo.

Como este tipo de aplicaciones es muy novedoso en la región, no se encontraron experiencias concretas, salvo las desarrolladas por los autores, a modo de ejemplo, en Córdoba, Argentina.

Recuadro 3.6. Ejemplo de análisis de los impactos de los riesgos climáticos sobre el valor inmobiliario: Estimación del efecto de la reducción del riesgo de incendio en Córdoba, Argentina

Determinación del problema. Se trata de establecer el impacto de la reducción del riesgo de incendio sobre el valor de los inmuebles en el área metropolitana de la ciudad de Córdoba.

Establecimiento de la base de datos. Se empleó información del observatorio del mercado inmobiliario y mapas de riesgos de incendios elaborados por el gobierno provincial. En este caso, se cuenta con una muestra de 1.139 observaciones del mercado inmobiliario debidamente georreferenciada. La variable dependiente es el valor de mercado de la tierra urbana y las variables independientes incluyen el nivel de riesgo de incendios e información sobre las características de los inmuebles y su entorno. Un 8,5% de la muestra corresponde a lotes ubicados dentro de áreas de riesgo elevado de incendios.

Desarrollo del modelo y análisis de resultados. Se entrenó un modelo de aprendizaje automático para simular el impacto de una intervención pública consistente en la eliminación de un polígono sujeto a riesgo elevado de incendios. Al realizar la estimación del valor del suelo para toda el área de estudio, una vez retirado de la muestra el polígono de alto riesgo, se observó que el impacto de la eliminación del riesgo de incendio en dicho sector generaba un incremento del valor del suelo igual a US\$6.772.725, lo cual representa un aumento del valor de los inmuebles del 13,31% para toda el área sujeta a la intervención. Este valor sirve de referencia para establecer el monto máximo de las iniciativas y los proyectos de mitigación del riesgo de incendios que se realicen en la zona analizada.

Fuente: Elaboración propia.



CAPÍTULO 4.

Caja de herramientas para realizar avalúos masivos

JUAN PABLO CARRANZA Y HUÁSCAR EGUINO



La caja de herramientas que se presenta en este capítulo proporciona códigos útiles para desarrollar un proceso completo de valuación inmobiliaria a escala masiva. Se trata de una librería elaborada en el lenguaje de programación R (R Core Team, 2025)¹³ llamada **'valutate'** (Carranza, 2025).

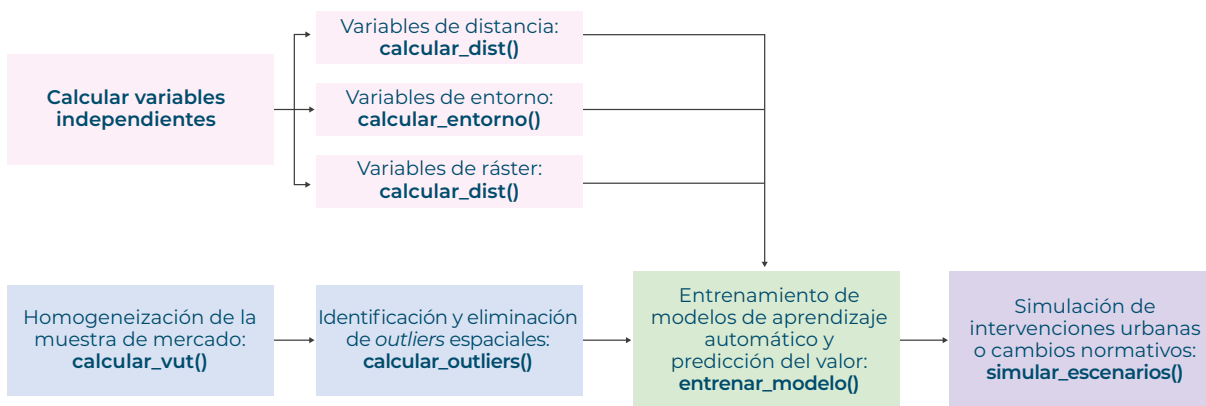
Una librería consiste en un conjunto de funciones que, a partir de una serie de *inputs*, también llamados argumentos, permiten obtener determinados *outputs* o resultados.

Para instalar la librería en el entorno de trabajo de R se debe proceder de la siguiente manera:

```
remotes::install_github("r-valutate/valutate")
```

El gráfico 4.1 esquematiza el proceso propuesto en esta guía para implementar el modelo de valuación automatizada (AVM) y valorar intervenciones urbanas. La presentación de las funciones de la librería sigue la estructura lógica del esquema y se desarrolla con aplicaciones y ejemplos para cada una de ellas.

Gráfico 4.1. Propuesta de proceso para realizar valuaciones masivas



Fuente: Elaboración propia.

El cuadro 4.1 presenta la estructura del capítulo, con las secciones correspondientes a cada etapa del proceso detallado en el gráfico 4.1 y sus casos de aplicación. El capítulo comienza con la explicación y ejemplificación del uso de tres funciones desarrolladas para facilitar el cálculo de variables independientes en el modelo predictivo (procesos en rosado en el gráfico 4.1). Estas funciones permiten calcular variables que resumen, para toda el área de estudio, la distancia a diferentes hitos urbanos o las características del entorno al cuantificar la cantidad de hitos que se encuentran en la proximidad a cada inmueble. Asimismo, el capítulo incluye una función que permite adecuar información calculada previamente a los criterios aplicados en el proceso de valuación diseñado para un área urbana específica. La guía contiene ejemplos de utilización de cada una de estas funciones en diferentes áreas urbanas: Córdoba y La Plata (Argentina), Belo Horizonte (Brasil), Cali y Medellín (Colombia).

13 Para acceder a más información sobre el manejo de datos espaciales en R se sugiere consultar la siguiente publicación: Carranza, J.P. 2025. Ciencia de datos geográficos aplicada a la política pública. Córdoba: UNC. Disponible en: https://carranzajuana.github.io/spatial_data_science/intro_datos_territoriales.html.

Cuadro 4.1. Estructura del capítulo 4: procesos, funciones y casos de aplicación

Proceso	Funciones	Caso	Sección
Cálculo de variables independientes.	calcular_dist() calcular_entorno() calcular_raster()	La Plata, Argentina. Belo Horizonte, Brasil. Medellín, Colombia.	Sección 4.1 Sección 4.2 Sección 4.3
Homogeneización y depuración de la muestra.	calcular_vut() eliminar_outliers()	Córdoba, Argentina.	Sección 4.4 Sección 4.5
Entrenamiento de modelos y predicción del valor de los inmuebles.	entrenar_modelo()	Cali, Colombia.	Sección 4.6
Valuación de intervenciones urbanas.	simular_escenario()	Medellín, Colombia. Belo Horizonte, Brasil. Córdoba, Argentina.	Sección 4.7.1 Sección 4.7.2 Sección 4.7.3

Fuente: Elaboración propia.

Además, se proveen herramientas para eliminar fuentes potenciales de error originadas en la heterogeneidad propia del proceso de relevamiento (por ejemplo, en la comparación de inmuebles próximos, pero con antigüedad o categoría constructiva diferentes) y una función para identificar y eliminar observaciones espacialmente atípicas (procesos en celeste en el esquema del gráfico 4.1). La aplicación de estas funciones se ejemplifica con un caso en el área metropolitana de Córdoba (Argentina).

Una vez que se han calculado las variables independientes con las tres funciones mencionadas en el primer párrafo y se ha depurado la muestra de mercado con las dos funciones mencionadas en el párrafo precedente se está en condiciones de entrenar un modelo para la predicción masiva del valor de los inmuebles (procesos en verde en el gráfico 4.1). A tal efecto, se provee una función y se ejemplifica con el proceso completo para la ciudad de Cali (Colombia).

Finalmente, se proporciona una función para simular escenarios de intervenciones públicas (procesos en morado en el gráfico 4.1). Este paso del proceso es crucial, puesto que permite obtener una estimación económica objetiva del impacto potencial de los beneficios que generan los proyectos urbanos o las decisiones administrativas de los gobiernos, lo cual constituye información de suma utilidad para el diseño, la implementación y la evaluación de políticas públicas basadas en evidencia. Se desarrollan tres ejemplos de este tipo de intervenciones: la estimación de los beneficios producidos por una nueva vía primaria en Medellín (Colombia), la estimación de las externalidades positivas generadas por un nuevo espacio verde en Belo Horizonte (Brasil) y la estimación de los beneficios originados por una reducción del riesgo de incendios forestales en el área metropolitana de Córdoba (Argentina). La función propuesta puede ser utilizada, además, para estimar de forma objetiva el impacto en el valor de los inmuebles de un cambio normativo sobre el perfil constructivo en determinadas zonas, entre otros efectos relevantes para el desarrollo urbano y territorial.

Recuadro 4.1. Nociones básicas sobre representaciones cartográficas

Las representaciones cartográficas son modelos que permiten expresar con símbolos el espacio geográfico. Por lo tanto, estos modelos son simplificaciones simbólicas del territorio.

Los símbolos utilizados para representar el espacio geográfico se pueden agrupar en dos grandes categorías:

- **Datos vectoriales.** Son puntos, líneas y polígonos. De esta manera, por ejemplo, la localización de un comercio se puede representar como un punto; una ruta o avenida se pueden representar como una línea, y el área de una parcela urbana se puede representar como un polígono. Las extensiones de archivos de formato vectorial más utilizadas son .shp, .gpkg, .json o .geojson, entre otras.
- **Datos ráster.** Son matrices de celdas o píxeles en las que cada celda contiene un valor que representa una característica del espacio en esa ubicación específica. Este tipo de datos es especialmente útil para representar fenómenos continuos, como la altitud, la temperatura, la cobertura vegetal o la concentración de contaminantes. Cada celda del ráster tiene una resolución determinada, que define el tamaño del área del territorio que representa. Cuanto menor es la resolución (es decir, cuanto más pequeñas son las celdas), mayor es el nivel de detalle espacial que se puede observar. La extensión de archivos ráster más utilizada es .tif.

Fuente: Elaboración propia.

4.1. Cálculo de variables de distancia: Función `calcular_dist()`

Esta función, que corresponde al proceso resaltado en rosa en el gráfico 4.1, devuelve un ráster con las distancias euclidianas, en metros, desde cada píxel hacia el objeto vectorial más cercano. La función permite calcular variables independientes que resumen, para toda el área de estudio, la distancia al objeto espacial más próximo. De esta manera, por ejemplo, se puede contar con un ráster que informe sobre las distancias a hitos urbanos, como vías primarias o secundarias, plazas, parques, entre otros.

Recuadro 4.2. Componentes de una función en R

En el lenguaje de programación **R**, una *función* es un bloque de código diseñado para realizar una tarea específica. Las funciones permiten reutilizar código y mejorar la legibilidad y modularidad de los programas. Además, facilitan el análisis de datos al automatizar procesos.

Una función en R toma entradas (llamadas **argumentos**), ejecuta un conjunto de instrucciones sobre esas entradas y devuelve un **valor de salida**. Las funciones pueden ser predefinidas (como **mean()**, que sirve para calcular la media de un conjunto de datos; **sum()**, que suma los valores de un conjunto de datos, o **lm()**, que realiza una regresión lineal por mínimos cuadrados ordinarios, entre otras) o pueden ser definidas por el usuario.

Componentes de una función en R:

- Nombre de la función. Es el identificador que se usa para denominar la función.
- Argumentos. Son los valores de entrada que recibe la función. Se colocan entre paréntesis después del nombre y pueden tener valores por defecto.
- Valor de salida. Es el resultado que la función produce.

Fuente: Elaboración propia.

Debido a la lógica del lenguaje de programación R (recuadro 4.2), una función tiene una serie de argumentos, que funcionan como *inputs* necesarios para el cálculo de un *output* determinado. En el caso de la función `calcular_dist()` se requieren cuatro argumentos para lograr el cálculo correspondiente: un polígono que detalle el área de estudio, los objetos hacia los que se quiere calcular la distancia (vías de circulación, plazas, comercios, etc.), la dimensión del ráster¹⁴ resultante y el nombre que se quiere dar a este resultado.

```
calcular_dist(area, objeto, dim, nombre)
```

Argumentos:

- *area*: Polígono que indica el área sobre la cual se ejecutará el cálculo.
- *objeto*: Objeto espacial de la librería **'sf'** (Pebesma, 2018) contra el que se calcularán las distancias.¹⁵
- *dim*: Tamaño de cada píxel del ráster resultante, en metros.
- *nombre*: Nombre de la variable resultante. El ráster será guardado en el entorno de trabajo con este nombre y la extensión .tif.

Para ejemplificar su aplicación, es necesario definir un área de estudio. Se toma como caso la ciudad de La Plata, en Argentina. Para identificar el polígono administrativo de esta ciudad se utilizan las funciones `geo_lite_sf()`, de la librería **'nominatimlite'** (Hernangómez, Lacko y White, 2025), y `st_transform()` y `st_bbox()`, de la librería **'sf'**.

```
library(nominatimlite)
library(sf)
area_de_estudio <- geo_lite_sf(address = "La Plata, Argentina", points_only = F)
bbox <- st_transform(area_de_estudio, 4326) |> st_bbox(bbox)
```

A continuación, se usa la librería **'osmdata'** (Padgham et al., 2017) para extraer información espacial desde la interfaz de programación de aplicaciones (API, por sus siglas en inglés) de OpenStreetMap (OSM). En este caso, se extraen las vías secundarias comprendidas dentro del área de estudio.

```
library(osmdata)
vias_sec <- opq(bbox) |>
  add_osm_feature(key = "highway", value = "secondary") |>
  osmdata_sf()
vias_sec <- vias_sec$osm_lines
```

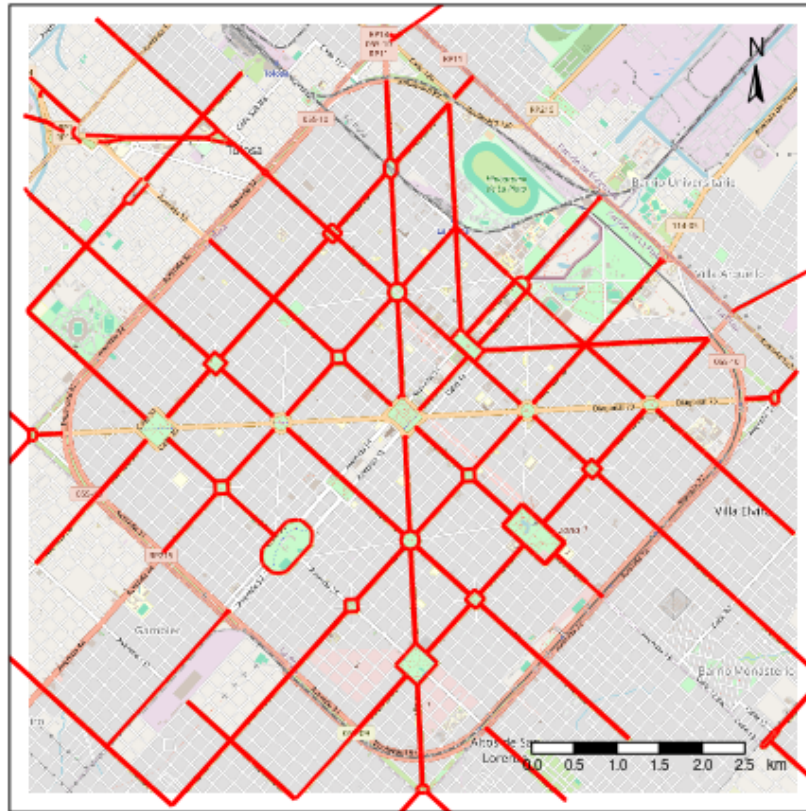
14 Un ráster es una matriz, que en la práctica puede entenderse como una serie de píxeles que indican el valor de una variable sobre el territorio. Se trata de uno de los tipos de objetos espaciales con los que trabajan los sistemas de información geográfica: datos vectoriales (puntos, líneas o polígonos) y datos ráster. La mayor parte de las funciones contenidas en la librería genera como resultado un ráster, puesto que su cálculo presenta una mayor eficiencia en términos de recursos computacionales y la extracción de valores hacia objetos vectoriales es relativamente simple.

15 Debido a sus capacidades, la variedad de funciones que ofrece y la calidad de su desarrollo, **'sf'** es la librería más utilizada en la gestión de información espacial en el lenguaje de programación R. Para acceder a más información, véase el siguiente enlace: <https://r-spatial.github.io/sf/>.

Se utilizan las librerías **'tmap'** (Tennekes, 2018) y **'tmtools'** (Tennekes, 2025) para construir un mapa que permita verificar la información bajada desde la API de OSM. El mapa 4.1 muestra las vías secundarias de la ciudad de La Plata (Argentina), simbolizadas como líneas de color rojo.

```
library(tmap)
library(tmtools)
tm_shape(read_osm(bb(área_de_estudio)), ext = 1.05, type = "waze") +
  tm_rgb() +
  tm_shape(vías_sec) +
  tm_lines(lwd = 2, col = "red") +
  tm_compass(type = "arrow", position=c("right", "top"), size = 1) +
  tm_scale_bar()
```

Mapa 4.1. Vías primarias de la ciudad de La Plata, Argentina

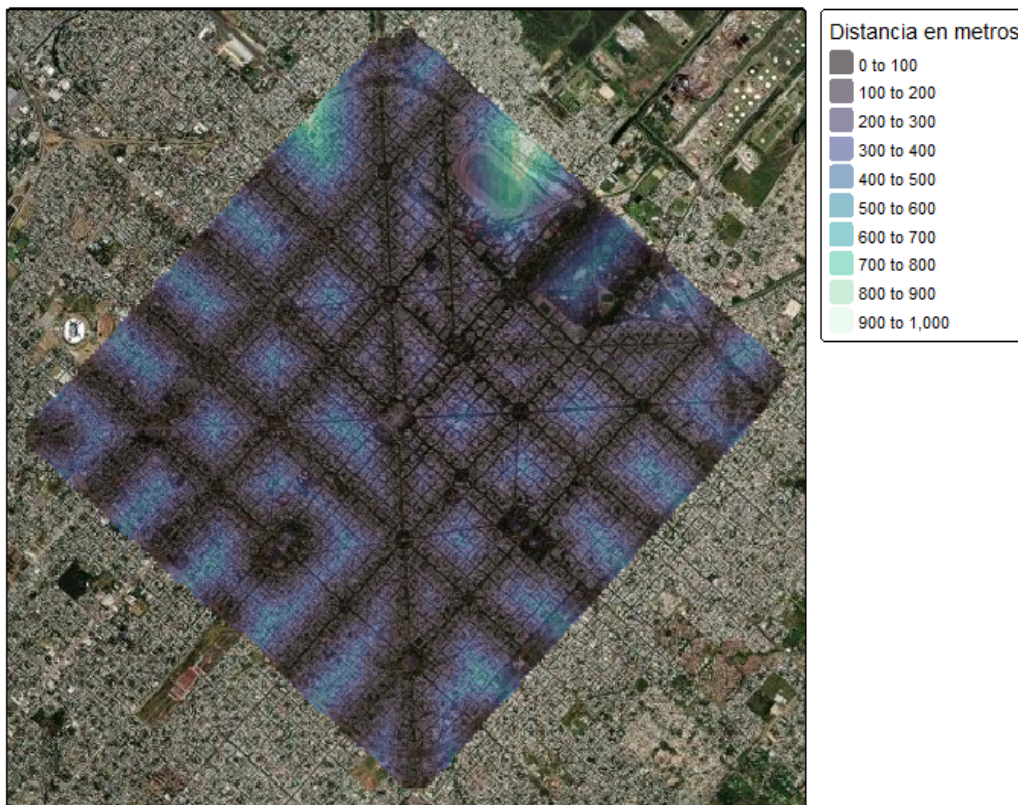


Fuente: Elaboración propia.

Luego, se usa la función `calcular_dist()` de la librería `'valuate'` para elaborar un ráster con la distancia de cada píxel a la vía secundaria más cercana (mapa 4.2).

```
library(valuate)
calcular_dist(area = area_de_estudio,
              objeto = vias_sec,
              dim = 50,
              nombre = "vias_secundarias")
```

Mapa 4.2. Resultado de la función `calcular_dist()` para vías secundarias en La Plata, Argentina



Fuente: Elaboración propia.

El ráster resultante indica que la distancia media a la vía secundaria más próxima en el área de estudio es igual a 200 metros, con un valor mínimo de 0 metros y un máximo de 998 metros.

4.2. Cálculo de variables de entorno: Función `calcular_entorno()`

Esta función, que también corresponde al proceso resaltado en rosa en el gráfico 4.1, devuelve un ráster que resume las características del vecindario a partir de capas geográficas vectoriales de polígonos o de puntos.

```
calcular_entorno(area, objeto, dim, ext, nombre)
```

Argumentos:

- *area*: Polígono que indica el área sobre la cual se ejecutará el cálculo.
- *objeto*: Objeto espacial **'sf'** con geometría del tipo "POLYGON" o "MULTIPOLYGON" sobre el que se calcularán las características del entorno.
- *dim*: Tamaño de cada píxel del ráster resultante, en metros.
- *ext*: Extensión del entorno o radio que define el vecindario, en metros lineales.
- *nombre*: Nombre de la variable resultante. El ráster será guardado en el entorno de trabajo con este nombre y la extensión .tif.

Cuando se le provee un objeto espacial **'sf'** con geometría "POLYGON" o "MULTIPOLYGON", la función permite calcular variables independientes que resumen las características del vecindario e indican la proporción del espacio que se encuentra cubierta por los polígonos de interés. Por ejemplo, permite calcular el porcentaje del vecindario que se encuentra cubierto por plazas o parques.

Cuando se le provee un objeto espacial **'sf'** con geometría "POINT", la función permite calcular variables independientes que resumen las características del vecindario e indican la cantidad de puntos de interés que se encuentran dentro del entorno definido por el usuario. Por ejemplo, permite calcular la cantidad de negocios u oficinas que se encuentran dentro de un radio de "x" metros de cada píxel.

Para ejemplificar la utilización de esta función se realiza el cálculo de espacios verdes (plazas o parques) en el entorno dentro de los límites administrativos de Belo Horizonte, en Brasil. Por lo tanto, el primer paso consiste en definir el área de estudio. Al igual que en el caso de la función `calcular_dist()`, se extrae esta información desde OSM, aunque se podría usar una capa específica con el polígono del área de estudio por utilizar.

```
area_de_estudio <- geo_lite_sf(address = "Belo Horizonte, Brasil", points_only = F)
bbox = st_transform(area_de_estudio, 4326) |> st_bbox(bbox)
```

Una vez definida el área de estudio, se procede a la extracción de información. En este caso, los usuarios han mapeado los espacios verdes con polígonos y con líneas, por lo cual se debe unificar la información dentro de una misma capa de polígonos.

```

espacios_verdes <- opq(bbox) |>
  add_osm_feature(key = "leisure", value = "park") |>
  osmdata_sf()
poligonos = espacios_verdes$osm_polygons
lineas = espacios_verdes$osm_lines
lineas = st_collection_extract(st_polygonize(st_union(lineas)))
lineas = st_as_sf(lineas)
lineas = st_cast(lineas, "POLYGON")
poligonos = poligonos[,c("geometry")]
lineas = lineas[,c("x")]
lineas = dplyr::rename(lineas, geometry = x)
espacios_verdes = rbind(poligonos, lineas)

```

A continuación, se procede a calcular el porcentaje del entorno que se encuentra cubierto por espacios verdes. Para ello, se precisa el área de estudio sobre la que se realizará el cálculo, se define la capa de espacios verdes, se establece el tamaño del píxel del ráster resultante en 50 metros y se especifica que el cálculo debe realizarse para un entorno de 500 metros lineales (mapa 4.3).

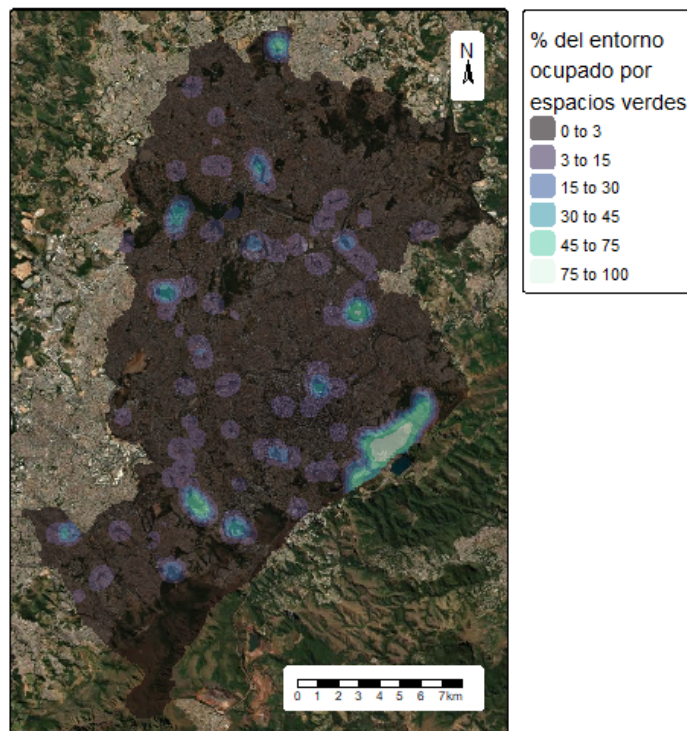
```

calcular_entorno(area_de_estudio, espacios_verdes, 50, 500,
"entorno_espacios_verdes")

```

Mapa 4.3. Resultado de la función `calcular_entorno()` para estimar el porcentaje cubierto por espacios verdes en Belo Horizonte, Brasil

Variable: `entorno_espacios_verdes`



Fuente: Elaboración propia.

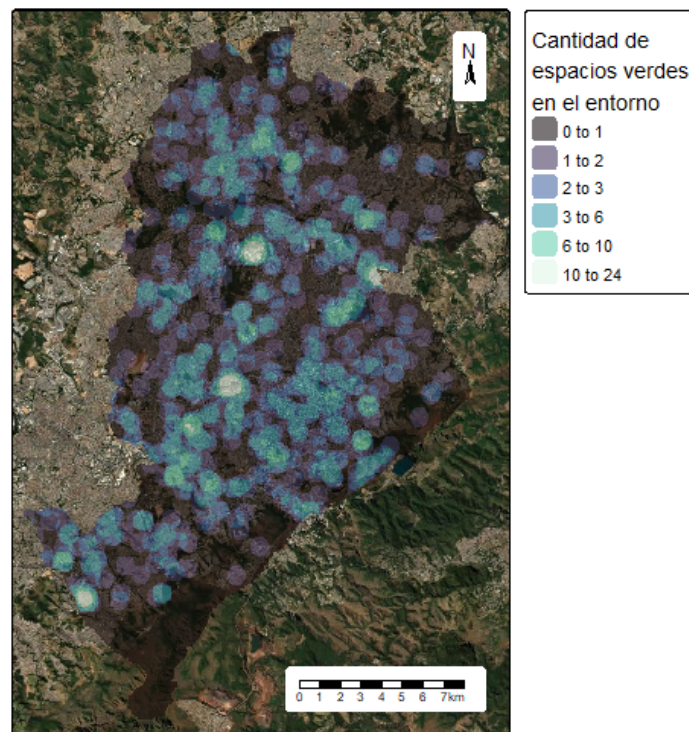
El ráster resultante muestra algunos espacios de la ciudad donde el entorno se encuentra cubierto por espacios verdes en un 80%. Sin embargo, la cobertura promedio para toda el área es igual al 2%.

También se podría calcular la cantidad de espacios verdes en el entorno. Para ello, en lugar de proporcionar a la función `calcular_entorno()` una capa de polígonos sobre los que calcular el porcentaje del espacio que se encuentra cubierto por plazas o parques, se debe proveer una capa de puntos. En este caso, la función contará la cantidad de puntos observados dentro del entorno de cada píxel. Por lo tanto, se procede a ingresar a la función los centroides¹⁶ de cada polígono correspondiente a un espacio verde (mapa 4.4).

```
calcular_entorno(area_de_estudio, st_centroid(espacios_verdes), 50, 500,
"entorno_espacios_verdes")
```

Mapa 4.4. Resultado de la función `calcular_entorno()` para estimar la cantidad de espacios verdes en Belo Horizonte, Brasil

Variable: `entorno_espacios_verdes`



Fuente: Elaboración propia.

Como se aprecia en el ráster resultante, existen sectores donde la cantidad de espacios verdes en el entorno es superior a 20 (con un máximo de 24). La cantidad promedio de plazas o parques en el entorno es igual a 1,47.

16 Esto se logra haciendo uso de la función `st_centroid()` de la librería 'sf'.

4.3. Cálculo de variables ráster: Función `calcular_raster()`

En algunas ocasiones se cuenta con información en formato ráster que, debido a su naturaleza, puede estar expresada en un sistema de coordenadas de referencia diferente del que se está utilizando o tener una resolución espacial distinta (el tamaño de los píxeles puede variar). En estos casos, es necesario readecuar las características del ráster original para hacerlo comparable al resto de la información que se está usando en el proceso de valuación. Esta es la misión de la función `calcular_raster()`, correspondiente al proceso resaltado en rosa en el gráfico 4.1, que tiene los siguientes argumentos:

```
calcular_raster(área, ráster, dim, entorno, nombre)
```

Argumentos:

- *área*: Polígono que indica el área sobre la cual se ejecutará el cálculo. La extensión total de este polígono debe estar comprendida dentro del ráster sobre el que se desea realizar los cálculos, es decir que no puede haber porciones fuera de dicha cobertura.
- *ráster*: Ráster original que se desea readecuar a los criterios aplicados en el proyecto de valuación.
- *dim*: Tamaño de cada píxel del ráster resultante, en metros.
- *entorno*: Distancia en metros lineales en el caso de que se desee transformar el ráster original a uno que resuma el valor promedio en el vecindario. El valor por *default* es cero.
- *nombre*: Nombre de la variable resultante. El ráster será guardado en el entorno de trabajo con este nombre y la extensión `.tif`.

Para ejemplificar la utilidad de esta función, se supone que se cuenta con un polígono que describe el área metropolitana de Medellín, en Colombia. También se dispone de un ráster con información sobre las edificaciones para dicho sector.¹⁷ En este ráster, cada píxel indica la cantidad de metros cuadrados edificados. La resolución espacial del ráster original es de 10 metros (cada píxel tiene 10 metros de alto y 10 metros de ancho) y el sistema de coordenadas de referencia es Mollweide (mapa 4.5). Se considera que en el proyecto de valuación se trabaja con variables resumidas a rásteres con una resolución espacial de 50 metros, en un sistema de coordenadas de referencia Pseudo-Mercator (EPSG:3857).

El primer paso consiste en cargar el polígono del área de estudio y el ráster original dentro del directorio de trabajo. Para el manejo de rásteres se utilizará la librería **'terra'** (Hijmans et al., 2025a).

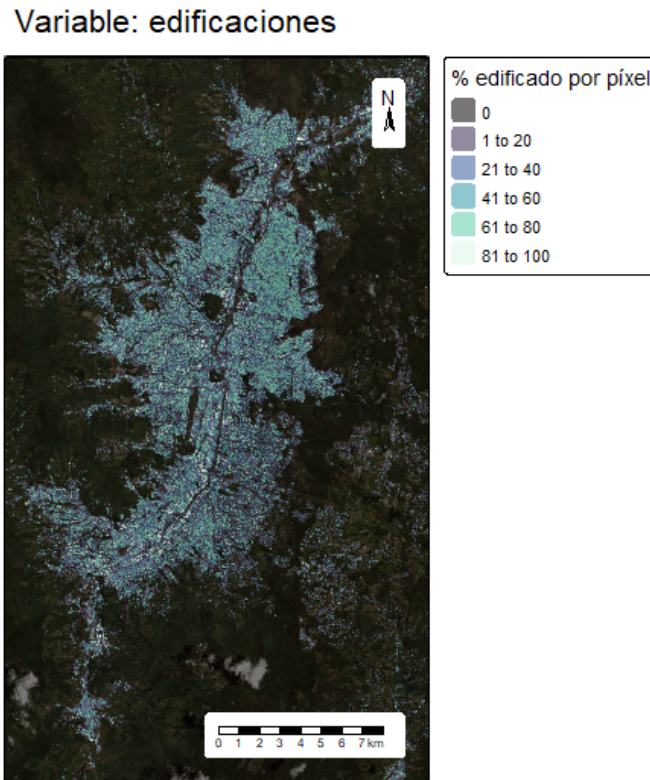
```
library(sf)
library(terra)
edificaciones <- rast(system.file("extdata/edificaciones.tif", package = "valuate"))
área_de_estudio <- st_read(system.file("extdata/area.gpkg", package = "valuate"))
```

¹⁷ Esta información se puede obtener de manera gratuita en el sitio web del proyecto Global Human Settlement Layer, de la Comisión Europea, en el siguiente enlace: <https://human-settlement.emergency.copernicus.eu/>.

Con estos elementos ya se está en condiciones de ejecutar la función.

```
calcular_raster(raster = edificaciones, area = area_de_estudio, entorno = 0, dim = 50, nombre = "edificaciones")
```

Mapa 4.5. Resultado de la función `calcular_raster()` para estimar el porcentaje edificado por píxel en Medellín, Colombia



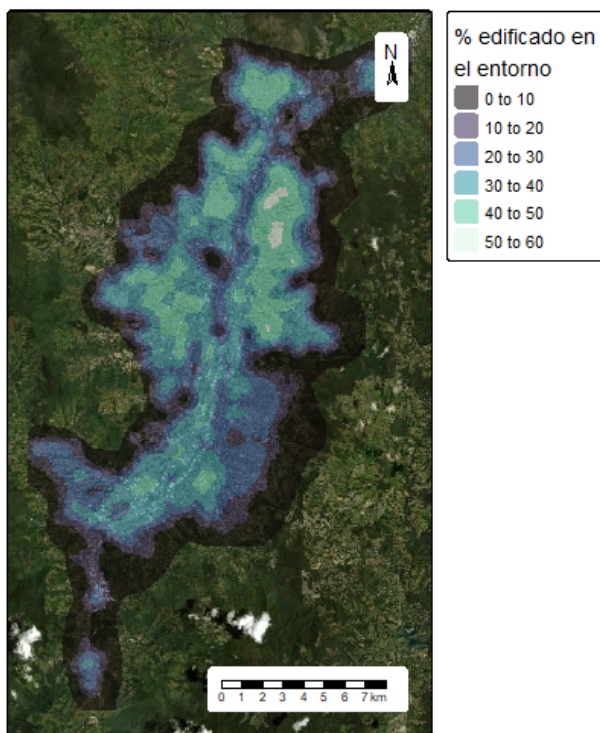
Fuente: Elaboración propia.

Luego se considera que, además, se desea transformar la información del ráster original para dar cuenta del promedio de píxeles edificados en un entorno de 500 metros a la redonda (mapa 4.6). Para lograr esto, se cambia el argumento "entorno" de la función, de la siguiente manera:

```
calcular_raster(raster = edificaciones, area = area_de_estudio, entorno = 500, dim = 50, nombre = "edificaciones_entorno")
```

Mapa 4.6. Resultado de la función `calcular_raster()` para estimar el porcentaje edificado en el entorno en Medellín, Colombia

Variable: `edificaciones_entorno`



Fuente: Elaboración propia.

4.4. Homogeneización de la muestra: Función `calcular_vut()`

La función, que pertenece al proceso resaltado en celeste en el gráfico 4.1, estima el precio por metro cuadrado homogeneizado, de acuerdo con Carranza (2024) y Cerino et al. (2020). Toma como *input* un objeto espacial '**sf**' y ajusta coeficientes de corrección según diferentes variables definidas por el usuario para reexpresar el precio en términos de un "lote típico" para el área de estudio.

La función devuelve un objetivo llamado "resultado", que consiste en el df original al que se agrega una columna correspondiente al valor homogeneizado (*vut*) y otra columna con el coeficiente de homogeneización aplicado (*coef*). Se informa, además, el parámetro de ajuste estimado por el modelo lineal para cada una de las variables independientes involucradas en el proceso de homogeneización.

```
calcular_vut(df, index, dependiente, independientes, dist)
```

Argumentos:

- *df*: Objeto '**sf**' con geometría 'POINT'. Muestra georreferenciada con precios de la tierra cuyo valor se pretende expresar en términos homogéneos.
- *index*: Identificador único para cada variable.
- *independientes*: Vector con los nombres de las columnas que se usarán como variables independientes y el valor de referencia correspondiente a un "lote típico". Si las variables son numéricas se debe indicar si el ajuste a un lote típico debe tomar como referencia la media o la mediana de dicha variable. En el caso de ser variables categóricas, se debe indicar el nombre de la categoría que se tomará como referencia (más adelante se presenta un ejemplo).
- *dependiente*: Nombre de la variable dependiente, aquella que contiene el valor por metro cuadrado.
- *dist*: Distancia que define el vecindario sobre el que se realizará la corrección por "dependencia espacial".

```
data("terrenos", package = "valuate")
head(terrenos)
```

La base de datos es un *dataframe* espacial con geometría de puntos. Cada punto hace referencia a un lote baldío en venta o vendido en la ciudad de Córdoba, Argentina. Las columnas de la base de datos son las siguientes:

	id	valor	tipo_valor	superficie	situacion_juridica	valor_m2
1	9263	250000	Oferta	400	Con Escritura	625.0000
2	8701	45000	Oferta	250	Pre venta	180.0000
3	10556	145000	Oferta	1196	Con Escritura	121.2375
4	5963	250000	Oferta	560	Con Escritura	446.4286
5	7076	80000	Oferta	474	Con Escritura	168.7764
6	2691	140000	Oferta	600	Con Escritura	233.3333

Donde:

- *id*: Es el identificador de cada dato o fila.
- *valor*: Es el monto total de la operación de cada inmueble, ya sea oferta o venta.
- *tipo_valor*: Es una columna que indica, para cada fila, si el dato hace referencia a un inmueble en oferta o a una venta efectivamente realizada.
- *superficie*: Es la superficie del inmueble en metros cuadrados.
- *situación_juridica*: Es una columna que indica, para cada fila, si la operación se realiza con escritura, sin escritura o con escritura en trámite (preventa).
- *valor_m2*: Es el precio por metro cuadrado, igual a $\text{valor} / \text{superficie}$.

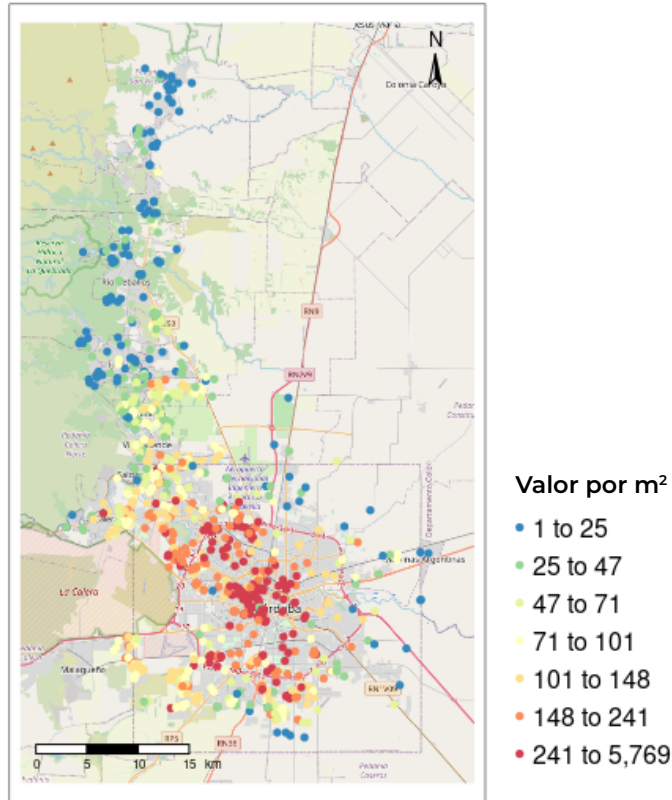
Las estadísticas descriptivas de cada variable son las siguientes:

```
summary(terrenos)
```

id	valor	tipo_valor	superficie	situacion_juridica	valor_m2
Min.: 24	Min.: 3700	Oferta: 975	Min.: 104	Con Escritura : 811	Min.: 1.463
1st Qu.: 3678	1st Qu.: 31250	Venta : 25	1st Qu.: 361	Pre venta: 170	1st Qu.: 40.456
Median: 6188	Median: 70000		Median: 610	Sin título/Posesión: 19	Median: 87.963
Mean: 6017	Mean: 213556		Mean: 2981		Mean: 153.893
3rd Qu.: 8782	3rd Qu.: 130000		3rd Qu.: 1360		3rd Qu.: 165.000
Max 10876	Max.: 49261135		Max.: 300000		Max.: 5769.231

La distribución espacial del valor por metro cuadrado (variable *valor_m2*) se presenta en el mapa 4.7.

Mapa 4.7. Cálculo del valor homogéneo: Distribución espacial del valor por metro cuadrado en la ciudad de Córdoba, Argentina (en US\$ por m²)



Fuente: Elaboración propia.

El objetivo, entonces, es calcular el precio por metro cuadrado en términos de un lote homogéneo. A tal fin se define un lote homogéneo como aquel que tiene una superficie igual a la mediana observada en la base de datos, efectivamente vendido (es decir, que no sea una oferta) y cuya operación se haya realizado con la escritura correspondiente. Para el cálculo de la autocorrelación espacial (interdependencia del precio o de variables no observadas en el entorno) se define una distancia de 500 metros lineales.

```

calcular_vut(df = terrenos,
            index = "id",
            independientes = c("tipo_valor" = "Venta",
                              "superficie" = "mediana",
                              "situacion_juridica" = "Con Escritura"),
            dependiente = "valor_m2",
            dist = 500)

```

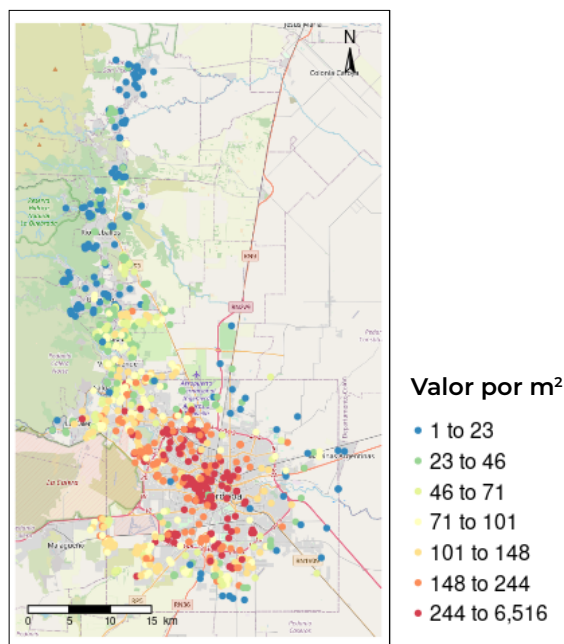
Los efectos estadísticamente significativos fueron los siguientes:

Variable	p-valor	Efecto
superficie	0.000000	-0.3770692
situacion_juridica:Sin título/Posesión	0.0026828	-0.5532533

Los resultados muestran efectos estadísticamente significativos en las variables superficie y situación jurídica (mapa 4.8). Así, por ejemplo, si hay dos lotes baldíos vecinos y muy similares, cuya única diferencia es que uno tiene una superficie igual a la mediana de la muestra y el otro es un 10% más grande, este último tendrá un precio por metro cuadrado un 3,7% más bajo. En el caso de la situación jurídica, el parámetro estimado implica que un lote baldío vendido sin escritura tiene un precio por metro cuadrado un 55,3% más bajo que los lotes baldíos vendidos de manera regular en el vecindario. Los demás efectos incluidos en la muestra no resultaron estadísticamente significativos.

Con estos parámetros, la función realiza los ajustes correspondientes, sugeridos por Carranza (2024) y Cerino et al. (2020), para reexpresar el valor de mercado de cada lote baldío en términos homogéneos y comparables.

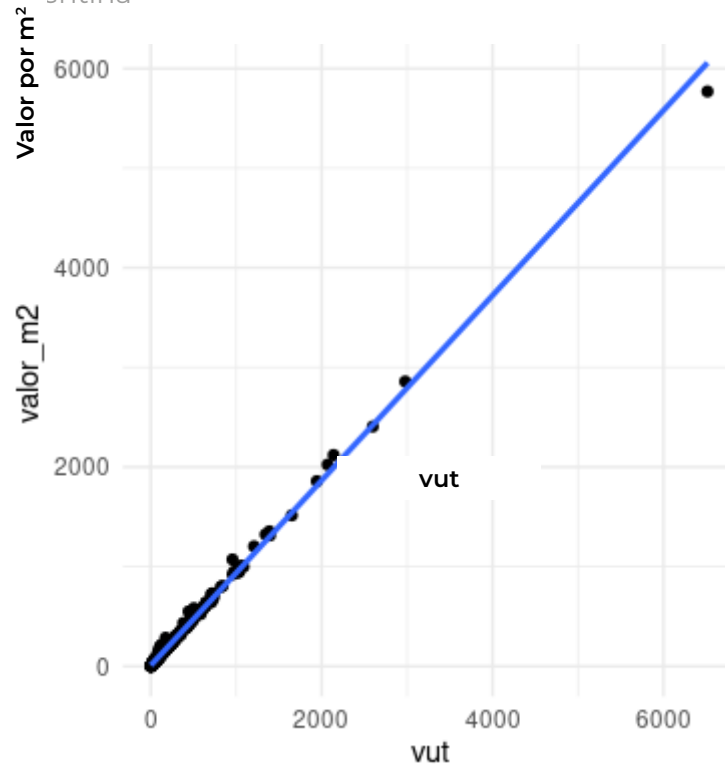
Mapa 4.8. Distribución en función del valor del metro cuadrado homogeneizado en la ciudad de Córdoba, Argentina (en US\$ por m²)



Fuente: Elaboración propia.

Si bien al comparar los mapas de las variables `valor_m2` y `vut` (mapas 4.7 y 4.8, respectivamente) la estructura espacial de ambas parece similar, las correcciones realizadas por la función se pueden apreciar en las escalas de colores calculadas a partir de los cuantiles de cada una de sus distribuciones. El valor máximo es más alto en la variable `vut` que en `valor_m2` (gráfico 4.2). Se observan, además, pequeñas correcciones en todo el conjunto de datos (excepto, claro, en los lotes baldíos que tienen originariamente las características de un lote típico).

Gráfico 4.2. Comparación entre el valor por metro cuadrado original y el homogeneizado en la ciudad de Córdoba, Argentina



Fuente: Elaboración propia.

4.5. Identificación de observaciones atípicas: Función `eliminar_outliers()`

Esta función, que también integra el proceso resaltado en celeste en el gráfico 4.1, permite eliminar observaciones cuyo valor es muy diferente de los observados en el entorno. Es posible identificar *outliers* espaciales mediante el cálculo del índice de Moran local (Anselin, 1995) o el promedio observado en el entorno ponderado por la distancia.

```
eliminar_outliers(df, dist, variable, umbral, moran)
```

Argumentos:

- *df*: *Dataframe* sobre el que se realizará el cálculo. Debe ser un objeto espacial de la librería '**sf**' con geometría 'POINT'.
- *dist*: Distancia euclidiana que define la extensión del vecindario en el que se comparan los valores.
- *variable*: Nombre de la variable sobre la que se realizará el cálculo.
- *umbral*: Umbral para caracterizar una observación como atípica con relación a lo observado en el vecindario. Por ejemplo, un umbral igual a 0,1 implica que se caracterizan como *outliers* todas las observaciones con valor superior o inferior al observado en el vecindario en una magnitud igual a +/- 10%.
- *moran*: Logical. Si asume el valor 'TRUE' se procede a identificar *outliers* mediante el cálculo del índice de Moran local. En este caso, el parámetro 'umbral' queda sin efecto.

La función permite identificar valores que son muy diferentes de los observados en el entorno. Para ello, compara el valor de cada observación con el promedio del entorno (ponderado por la distancia, mediante la construcción de una matriz de pesos espaciales). También permite identificar *outliers* espaciales por medio del cálculo del índice de Moran local, de acuerdo con Anselin (1995).

4.6. Entrenamiento de modelos: Función `entrenar_modelo()`

Esta función es crítica para el proceso de implementación de un AVM y pertenece a la sección resaltada en verde en el gráfico 4.1. Su utilización permite entrenar modelos y cuantificar su error en la predicción del valor de los inmuebles (o cualquier otra variable territorial que se esté analizando).

```
entrenar_modelo(
  df,
  dependiente,
  independientes,
  modelo = "qrf",
  umbral = 0.3,
  eliminar = 0.4)
```

Argumentos:

- *df*: *Dataframe* sobre el que se realizará el cálculo. Debe ser un objeto de la librería '**sf**' con geometría 'POINT'.
- *dependiente*: Nombre de la variable por estimar o variable dependiente.
- *independientes*: Vector con los nombres de los rásteres que contienen información sobre las variables independientes, generadas a partir de las funciones `calcular_dist()`, `calcular_raster()` o `calcular_entorno()`.
- *modelo*: Nombre del modelo para entrenar. Se corresponde con las abreviaciones previstas en la librería '**caret**' (Kuhn, 2008).¹⁸ El valor por *default* ("**qrf**") se corresponde al algoritmo Quantile Regression Forest (Meinshausen, 2006).
- *umbral*: Valor que define el error máximo permitido en la estimación. La función evalúa la precisión fuera de la muestra mediante un proceso de validación cruzada en 10 grupos. Para cada dato en la muestra se registra el error fuera de la muestra. Si el nivel de error agregado es superior al umbral definido (nivel de error igual o superior al establecido por el parámetro 'eliminar'), la función procede a eliminar los datos (gráfico 4.3).
- *eliminar*: Valor que define el nivel de error a partir del cual se depura la muestra, siempre que el error general de estimación se encuentre por encima del definido por el parámetro 'umbral'.

Esta función implementa un flujo completo de entrenamiento, evaluación y ajuste de un modelo supervisado para la estimación de una variable dependiente a partir de un conjunto de variables independientes en formato ráster. Específicamente, está diseñada para trabajar con datos espaciales almacenados en un *dataframe* con geometría de tipo 'POINT', por medio de herramientas de las librerías '**sf**' y '**terra**'. El proceso recorre los siguientes pasos principales:

¹⁸ Para más información, véase el siguiente enlace: <https://topepo.github.io/caret/train-models-by-tag.html>.

- **Preprocesamiento de datos.** Los rásteres de las variables independientes se combinan con el *dataframe* espacial mediante la extracción de valores correspondientes a los puntos de la geometría. Los datos se depuran al eliminar filas con valores faltantes.
- **Configuración del modelo.** Se define un modelo a través de la librería **'caret'** (Kuhn, 2008), con soporte para una amplia variedad de algoritmos. El modelo predeterminado es Quantile Regression Forest ("qrf"), aunque se pueden especificar otros métodos soportados por **'caret'**.
- **Entrenamiento iterativo.** La función utiliza validación cruzada de 10 particiones para evaluar el error relativo promedio en valor absoluto (MAPE, por sus siglas en inglés) fuera de la muestra. Si el MAPE promedio supera el umbral especificado, se eliminan observaciones cuyo MAPE individual es igual o superior al 40% y el modelo se entrena nuevamente con los datos restantes. Este proceso se repite hasta que el MAPE promedio cumple con el umbral definido.
- **Generación de predicción espacial.** Una vez finalizado el entrenamiento, el modelo entrenado se utiliza para interpolar las predicciones sobre los rásteres de entrada, a fin de generar un nuevo ráster de salida que se guarda en el directorio de trabajo con el nombre **'vut.tif'**.
- **Salida de datos.** La función genera dos conjuntos de datos en el entorno global: `datos_utilizados` (datos que cumplieron con el umbral de error) y `datos_eliminados` (datos descartados por errores altos).

La función está optimizada para aplicaciones que requieren análisis espaciales precisos y está diseñada para integrarse en flujos de trabajo geospaciales complejos. Además, emplea paralelización para acelerar el proceso de entrenamiento en máquinas con múltiples núcleos. El diseño de esta función está orientado a contextos donde es crucial garantizar una alta precisión en la estimación, especialmente en aplicaciones que involucran modelización espacial o análisis urbano. Además, facilita la integración con flujos de trabajo basados en datos geospaciales mediante las librerías **'sf'** y **'raster'** (Hijmans et al., 2025b).

El proceso de la función `entrenar_modelo()` se resume en el gráfico 4.3. Por defecto, la función entrena el algoritmo Quantile Regression Forest (Meinshausen, 2006), que ha mostrado un excelente desempeño para predecir fenómenos territoriales vinculados con el mercado inmobiliario (Carranza et al., 2022; Córdoba et al., 2021). Sin embargo, es posible seleccionar varios algoritmos, según las especificaciones de la librería **'caret'**.¹⁹ El modelo se entrena mediante un proceso de validación cruzada, donde la muestra se subdivide aleatoriamente en 10 subconjuntos de similar tamaño, cada uno de estos subconjuntos es retirado recursivamente de la muestra, el modelo es entrenado con los subconjuntos restantes y el nivel de error es estimado en el subconjunto extraído. De esta manera, se cuenta con información sobre el nivel de error con que el modelo estima el precio de los inmuebles en cada dato de la muestra, con esta observación fuera del proceso de entrenamiento. La medida de error aplicada por la función es el MAPE.

$$MAPE = \frac{1}{n} \sum \frac{|y - \hat{y}|}{y}$$

¹⁹ Para una lista detallada de los algoritmos aplicables en esta función, véase el siguiente enlace: <https://topepo.github.io/caret/train-models-by-tag.html>.

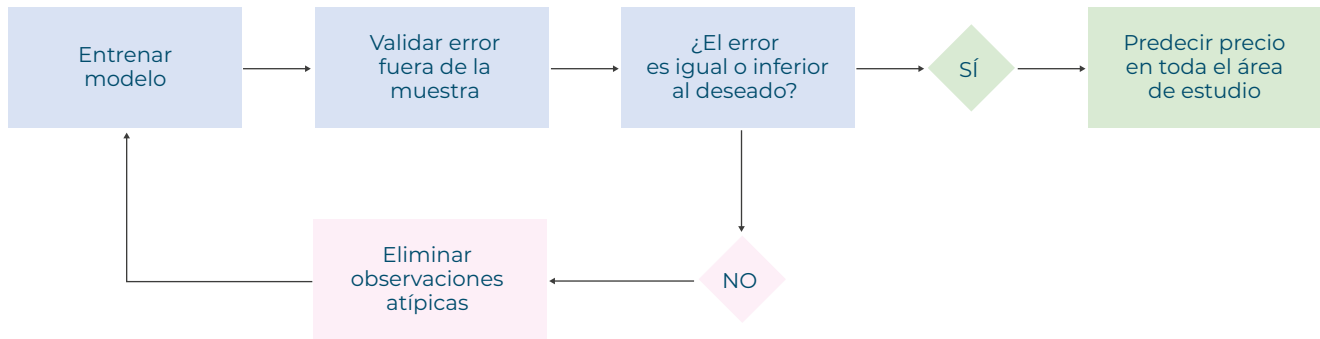
Donde:

y es el valor observado.

\hat{y} es el valor predicho por el modelo fuera de la muestra.

n es el tamaño de la muestra.

Gráfico 4.3. Flujo del proceso de entrenamiento de algoritmos



Fuente: Elaboración propia.

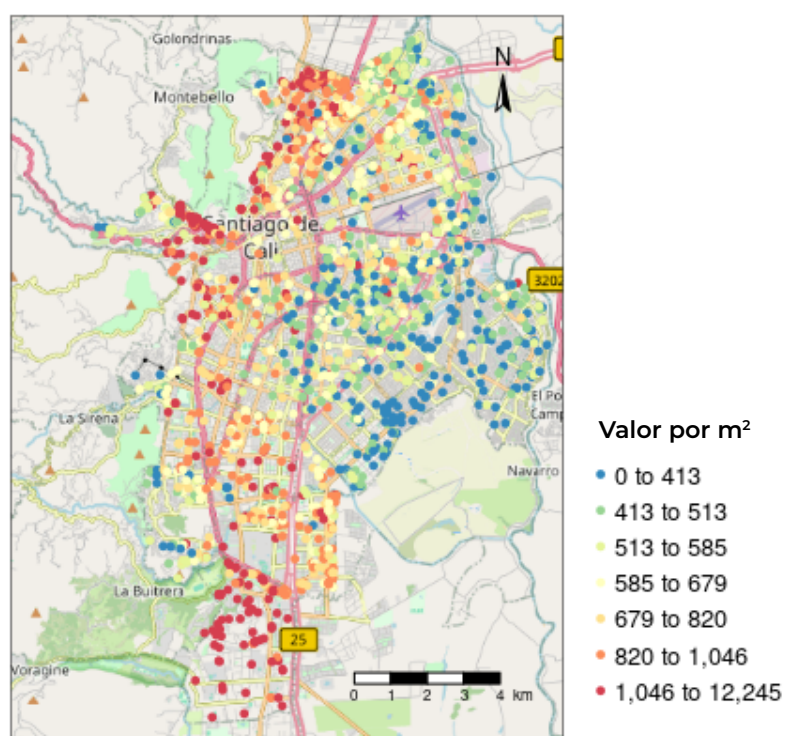
Si el nivel general de error fuera de la muestra estimado por el modelo es inferior al umbral predefinido por el usuario, se procede a calcular el precio de los inmuebles en toda el área de estudio. La función devuelve un ráster con esa información, donde la dimensión de cada píxel es igual a la definida en las funciones utilizadas para el cálculo de las variables independientes. En el caso de que el nivel de error sea superior al umbral predefinido por el usuario, la función procede a remover los datos estimados con un error fuera de la muestra elevado, al asumir que el algoritmo clasifica estas observaciones como atípicas a partir de las relaciones entre el precio de ese inmueble en particular y las variables independientes utilizadas en la estimación. Posteriormente, el proceso se repite de manera iterativa hasta que se alcanza el nivel de error mínimo indicado.²⁰

20 Es importante destacar que, si se fija un nivel de error demasiado bajo, inferior al error estructural de los datos contenidos en la muestra, la función puede entrar en un *loop* que se detendrá solo cuando la muestra quede sin observaciones.

4.6.1. Caso de aplicación: Estimación del valor de los inmuebles en la ciudad de Cali, Colombia

Para explicar el proceso se toma como ejemplo el caso de la ciudad de Cali, en Colombia. La muestra de datos utilizada proviene del Observatorio Inmobiliario de la Alcaldía de Cali y consta de 1.952 observaciones correspondientes a casas o apartamentos en venta en toda el área de estudio.²¹ La muestra contiene, además, información respecto del área construida de cada inmueble, la superficie del terreno, la antigüedad de las edificaciones y su estado de conservación. Su distribución espacial en términos del precio por metro cuadrado se presenta en el mapa 4.9.

Mapa 4.9. Distribución espacial de la muestra de inmuebles según el valor del metro cuadrado en Cali, Colombia (en US\$ por m²)



Fuente: Elaboración propia.

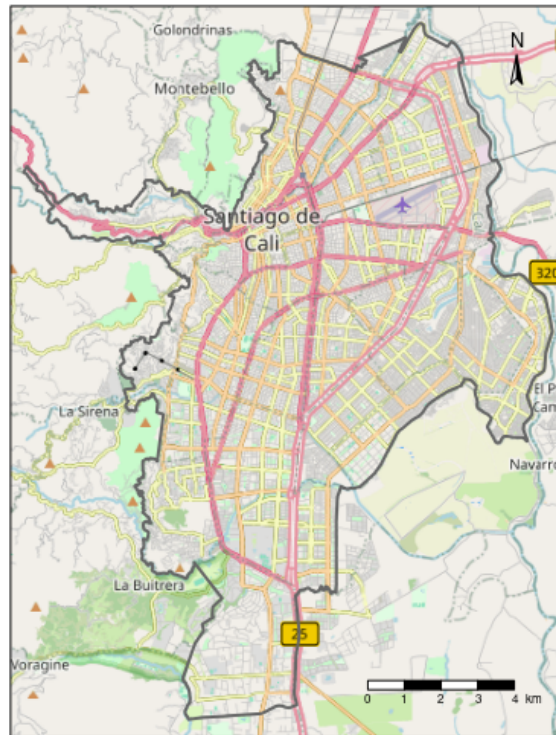
Una vez conformada la muestra, se calculan las variables independientes. Para asegurar la replicabilidad del proceso, solo se utilizan bases de datos libres y abiertas. Si bien esto puede tener algún efecto en la calidad de los resultados, permite avanzar hacia uno de los objetivos principales de esta guía, que es la construcción de capacidades para la implementación de análisis territoriales basados en datos. Por lo tanto, en primer lugar se define de manera precisa el área de estudio, tal como se indica en el mapa 4.10.

21 Para más información, véase el siguiente enlace: <https://www.cali.gov.co/hacienda/publicaciones/147970/observatorio-inmobiliario/>.

En este ejemplo se usa un recorte del área metropolitana de Cali, que está incluido en los archivos de la librería **'valuate'**. Para cargarlo al entorno de trabajo, se debe ejecutar el siguiente comando:

```
load(system.file("extdata/area_cali.rda", package = "valuate"))
```

Mapa 4.10. Definición del área de estudio: Polígono administrativo de Cali, Colombia



Fuente: Elaboración propia.

Sobre esta área de estudio se calculan diferentes variables independientes. Con base en información extraída desde la API de OSM (2023), se estiman variables de distancias a hoteles, vías principales y vías secundarias.

```

bbox = sf::st_transform(area, 4326) |> sf::st_bbox(bbox)
hotel <- osmdata::opq(bbox) |>
  osmdata::add_osm_feature(key = "tourism", value = "hotel") |>
  osmdata::osmdata_sf()
hotel <- hotel$osm_points
calcular_dist(area = area,
              objeto = hotel,
              dim = 20,
              nombre = "dist_hoteles")

vias_prim <- osmdata::opq(bbox) |>
  osmdata::add_osm_feature(key = "highway", value = "primary") |>
  osmdata::osmdata_sf()
vias_prim <- vias_prim$osm_lines
calcular_dist(area = area,
              objeto = vias_prim,
              dim = 20,
              nombre = "dist_vias_primarias")

vias_sec <- osmdata::opq(bbox) |>
  osmdata::add_osm_feature(key = "highway", value = "secondary") |>
  osmdata::osmdata_sf()
vias_sec <- vias_sec$osm_lines
calcular_dist(area = area,
              objeto = vias_sec,
              dim = 20,
              nombre = "dist_vias_secundarias")

```

Asimismo, sobre la misma fuente de datos, se calculan variables de entorno sobre espacios verdes, bancos y restaurantes.

```

bancos <- osmdata::opq(bbox) |>
  osmdata::add_osm_feature(key = "amenity", value = "bank") |>
  osmdata::osmdata_sf()
bancos <- bancos$osm_points
calcular_entorno(area = area,
  ext = 500,
  objeto = bancos,
  dim = 20,
  nombre = "bancos_entorno")

restaurant <- osmdata::opq(bbox) |>
  osmdata::add_osm_feature(key = "amenity", value = "restaurant") |>
  osmdata::osmdata_sf()
restaurant <- restaurant$osm_points
calcular_entorno(area = area,
  ext = 500,
  objeto = restaurant,
  dim = 20,
  nombre = "restaurants_entorno")

plazas <- osmdata::opq(bbox) |>
  osmdata::add_osm_feature(key = "leisure", value = "park") |>
  osmdata::osmdata_sf()
plazas_1 <- plazas$osm_polygons
plazas_2 <- plazas$osm_multipolygons
plazas_1 = plazas_1[,c("geometry")]
plazas_2 = plazas_2[,c("geometry")]
plazas = rbind(plazas_1, plazas_2)
plazas = st_cast(plazas, "POLYGON")
plazas = st_make_valid(plazas)
calcular_entorno(area = area,
  ext = 500,
  objeto = plazas,
  dim = 20,
  nombre = "plazas_entorno")

```

También se calculan variables de entorno a partir de datos abiertos de la Alcaldía de Cali:²² las pérdidas acumuladas por fenómenos sísmicos en el entorno y la cantidad de eventos por ruidos molestos registrados en el entorno.

22 Para más información, véase el siguiente enlace: <https://datos.cali.gov.co/>.

```

load(system.file("extdata/ruido_cali.rda", package = "valuate"))
ruido = st_transform(ruido, 3857)
calcular_entorno(area,
                ruido,
                ext = 500,
                dim = 20,
                "visitas_ruido")

load(system.file("extdata/riesgo_cali.rda", package = "valuate"))
riesgo = st_transform(riesgo, 3857)
r <- raster::raster(st_buffer(area, 100), res = 20)
r <- fasterize::fasterize(area, r)
r <- terra::crop(r, area)
r <- terra::rast(r)
terra::crs(r) <- "epsg:3857"
terra::values(r) <- 1
perdidas <- fasterize::fasterize(riesgo, raster::raster(r),
                                field = "losses", fun = "last")
writeRaster(perdidas, "perdidas_sismos.tif")

```

Además, se utiliza información satelital desarrollada por la Comisión Europea en el marco del proyecto Global Human Settlement Layer (Pesaresi et al., 2024). A partir de la clasificación automática de imágenes satelitales, este proyecto estima los píxeles edificados y la altura promedio de las edificaciones. Esta misma fuente proporciona, además, una estimación de la distribución de la población en el espacio, en función de la interacción entre imágenes satelitales y fuentes oficiales de censos. Por último, se agrega una variable que informa acerca de la elevación en metros sobre el nivel del mar, obtenida de la Shuttle Radar Topography Mission, mediante la librería de R **'geodata'** (Hijmans et al., 2024).

```

edificaciones <- rast(system.file("extdata/edificaciones_cali.tif", package = "valuate"))
calcular_raster(raster = edificaciones, area = area, entorno = 500, dim = 20, nombre = "edificaciones_entorno")

altura <- rast(system.file("extdata/altura_cali.tif", package = "valuate"))
calcular_raster(raster = altura, area = area, entorno = 500, dim = 20, nombre = "altura_entorno")

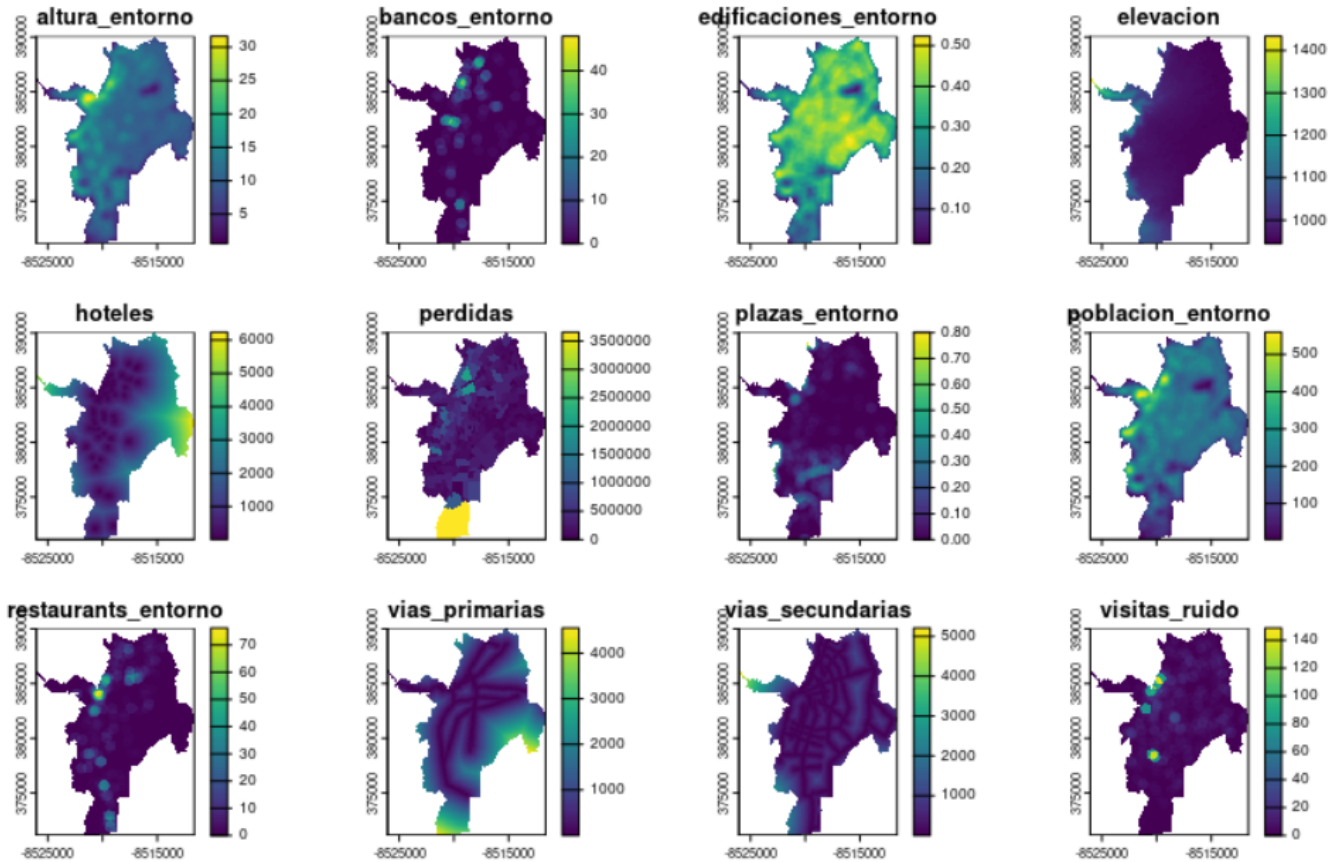
poblacion <- rast(system.file("extdata/poblacion_cali.tif", package = "valuate"))
calcular_raster(raster = poblacion, area = area, entorno = 500, dim = 20, nombre = "poblacion_entorno")

library(geodata)
coordenadas = st_centroid(st_transform(area, 4326))
coordenadas = st_coordinates(coordenadas)
elevacion <- elevation_3s(lon = coordenadas[[1]], lat = coordenadas[[2]], path=tempdir() )
elevacion = elevacion / 1
elevacion <- terra::project(elevacion, "epsg:3857")
elevacion = mask(elevacion, area)
calcular_raster(raster = elevacion, area = area, dim = 20, entorno = 0, nombre = "elevacion")

```

Por lo tanto, mediante la utilización de las funciones `calcular_dist()`, `calcular_entorno()` y `calcular_raster()` se crean y se guardan en el entorno de trabajo los rásteres correspondientes a cada una de las variables independientes. La distribución espacial de cada una de las variables independientes calculadas se puede apreciar en el mapa 4.11.

Mapa 4.11. Variables independientes utilizadas en el entrenamiento del modelo para estimar el valor de los inmuebles en la ciudad de Cali, Colombia



Fuente: Elaboración propia.

A continuación, se carga la muestra de datos de mercado al entorno de trabajo de la siguiente forma:

```
load(system.file("extdata/datos_entrenar_modelo.rda", package = "valuate"))
```

Se procede, entonces, a entrenar el modelo Quantile Regression Forest (Meinshausen, 2006) sobre la muestra de precios de inmuebles edificados en Cali, con las variables independientes construidas antes. Se define un umbral de error mínimo igual a +/- 20%.

```

entrenar_modelo(df = datos,
  dependiente = "vut",
  independientes = c("edificaciones_entorno.tif",
                    "altura_entorno.tif",
                    "poblacion_entorno.tif",
                    "bancos_entorno.tif",
                    "elevacion.tif",
                    "dist_hoteles.tif",
                    "perdidas_sismos.tif",
                    "plazas_entorno.tif",
                    "restaurants_entorno.tif",
                    "dist_vias_primarias.tif",
                    "dist_vias_secundarias.tif",
                    "visitas_ruido.tif"),

  modelo = "qrf",
  umbral = 0.2)

```

...entrenando modelo...

El proceso de entrenamiento del modelo ha finalizado. El MAPE resultó igual a +/- 16.48%.

...realizando interpolación...

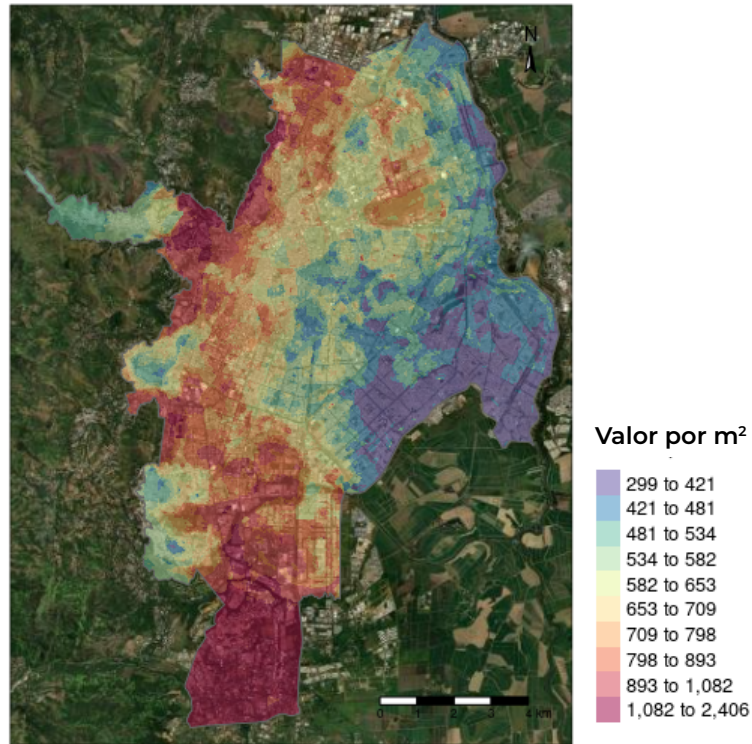
En el directorio de trabajo se ha guardado un ráster con la predicción, con el nombre 'vut.tif'.

Como se observa en el resultado de la función, el modelo giró una vez con la totalidad de la muestra y registró un error igual a +/- 16,48%. Debido a que este nivel de error es inferior al umbral definido de +/- 20%, el proceso se detiene en este punto. La distribución espacial del precio de los inmuebles resultante para toda el área de estudio se muestra en el mapa 4.12.

```

library(tmap)
library(tmaptools)
tm_shape(read_osm(bb(st_bbox(area)), ext = 1.05, type = "osm-german")) +
  tm_rgb() +
  tm_shape(val) +
  tm_raster(alpha = 0.5, palette = rev(RColorBrewer::brewer.pal(10, "Spectral")),
    n = 10, style = "quantile", title = "Precio por m2") +
  tm_shape(area) +
  tm_borders(lwd = 1) +
  tm_legend(position = c("left", "bottom"),
    bg.color = "white",
    bg.alpha=1, text.size = 0.8,
    width = 1, title.size = 1.2, outside = T) +
  tm_compass(type = "arrow", position=c("right", "top"), size = 1) + tm_scale_bar()

```

Mapa 4.12. Distribución espacial del valor de los inmuebles en Cali, Colombia (en US\$ por m²)

Fuente: Elaboración propia.

Para fines catastrales y para determinar la base imponible del impuesto a la propiedad, esta información se puede agregar a nivel parcelario, con objeto de proporcionar un nivel máximo de detalle a la información. También se puede agregar en otras unidades espaciales, con el propósito de adecuarla a la normativa tributaria específica de cada ciudad. Por ejemplo, si se deseara agregar la información del precio por metro cuadrado de los inmuebles en Cali a nivel de barrios, corresponde realizar lo siguiente:

```
# Cargar previamente la capa de polígonos correspondientes a barrios, o a la unidad espacial sobre la cual se quiere agregar la predicción del precio por metro cuadrado (parcelas, manzanas, sectores, etc.)
```

```
load(system.file("extdata/barríos_cali.rda", package = "valuate"))
```

```
barríos$pm2 = extract(vut, barríos, fun = "mean", na.rm = TRUE)[,2]
```

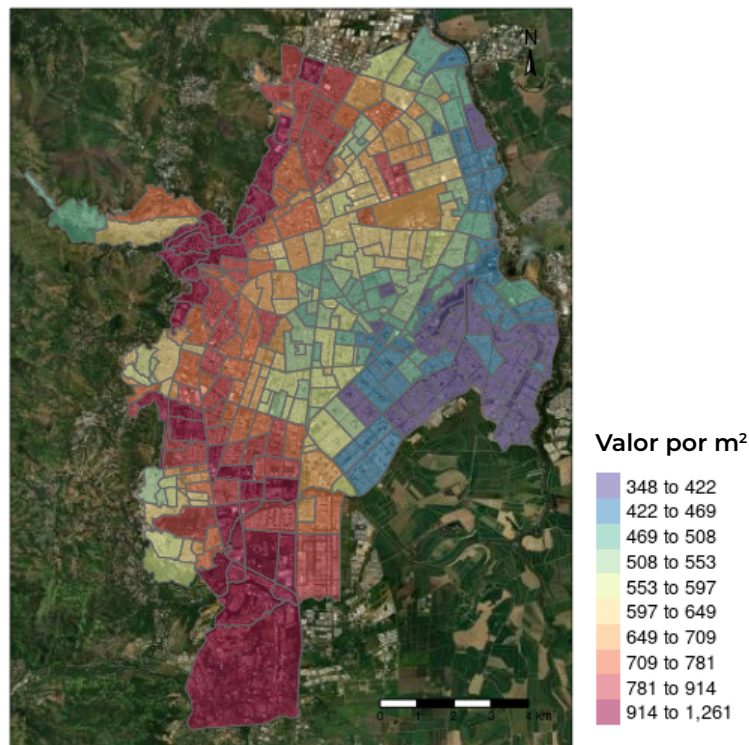
Para visualizar el resultado en un mapa (mapa 4.13), es preciso implementar el proceso que se describe a continuación:

```

tm_shape(tmaptools::read_osm(bb(st_bbox(area)), ext = 1.05, type = "osm-german")) +
tm_rgb() +
tm_shape(val) +
tm_fill("pm2", alpha = 0.5, palette = rev(RColorBrewer::brewer.pal(10, "Spectral")),
      n = 10, style = "quantile", title = "Precio por m2") +
tm_legend(position = c("left", "bottom"),
      bg.color = "white",
      bg.alpha = 1, text.size = 0.8,
      width = 1, title.size = 1.2, outside = T) +
tm_compass(type = "arrow", position=c("right", "top"), size = 1) + tm_scale_bar()

```

Mapa 4.13. Distribución espacial del valor por metro cuadrado de los inmuebles a nivel de barrios en Cali, Colombia (en US\$ por m²)



Fuente: Elaboración propia.

A partir de esta información es posible calcular, por ejemplo, intervalos de confianza para el precio por metro cuadrado promedio dentro de cada barrio (o polígono de interés, sean parcelas, manzanas u otros sectores administrativos). También se puede transformar esta información geográfica en alfanumérica, para conocer, por ejemplo, los 10 barrios con el valor del metro cuadrado más alto, en promedio:

```
library(kableExtra)
barrios |>
  st_drop_geometry() |>
  arrange(desc(pm2)) |>
  select(barrio, pm2) |>
  head(, 10) |>
  kable()
```

barrio	pm2
Parcelaciones Pance	1261.024
Urbanización Ciudad Jardín	1241.199
Normandía	1206.153
Arboledas	1200.292
Santa Teresita	1182.748
La Playa	1120.842
Ciudad Campestre	1104.976
Unicentro Cali	1099.266
Santa Rita	1059.174
Juanambú	1039.614

De manera similar, para saber cuáles son los 10 barrios con el menor precio promedio por metro cuadrado, solo se debe modificar el criterio de ordenamiento de la tabla y eliminar la función `desc()` para que los datos se ordenen de manera ascendente.

```
barrios |>
  st_drop_geometry() |>
  arrange(pm2) |>
  select(barrio, pm2) |>
  head(, 10) |>
  kableExtra::kable()
```

barrio	pm2
Villamercedes I - Villa Luz - Las Garzas	347.7855
El Vallado	350.4579
Pizamos II	352.3539
Pizamos III - Las Dalias	353.5421
El Retiro	355.8474
Las Orquídeas	356.0155
Potrero Grande	357.0737
Villamercedes I - Villa Luz - Las Garzas	357.1838
Pizamos I	365.4277
Los Comuneros I	368.2900

4.7. Valuación de intervenciones urbanas: Función `simular_escenario()`

Esta función constituye la principal innovación de la librería propuesta y forma parte de la sección resaltada en morado del proceso detallado en el gráfico 4.1. Su utilización permite simular de manera objetiva el valor de las externalidades generadas por potenciales modificaciones normativas, obras de infraestructura o eventos climáticos, entre otras, susceptibles de ser modeladas mediante la inclusión de diferentes variables independientes en la estimación del modelo de valuación original. La función tiene la siguiente estructura:

```
simular_escenario(modelo, independientes, original)
```

Argumentos:

- *modelo*: Es el modelo predictivo creado mediante la función `entrenar_modelo()` y guardado automáticamente en la carpeta seteada como directorio de trabajo.
- *independientes*: Son las mismas variables utilizadas para entrenar el modelo con `entrenar_modelo()`, a excepción de las variables recalculadas para la simulación del escenario. Las variables generadas para la simulación deben tener una estructura exactamente igual a la de las variables originales (la misma resolución espacial, los mismos nombres, etc.).
- *original*: Ruta a la estimación original del valor de los inmuebles. Por defecto, la función `entrenar_modelo()` se guarda en el directorio de trabajo seteado.

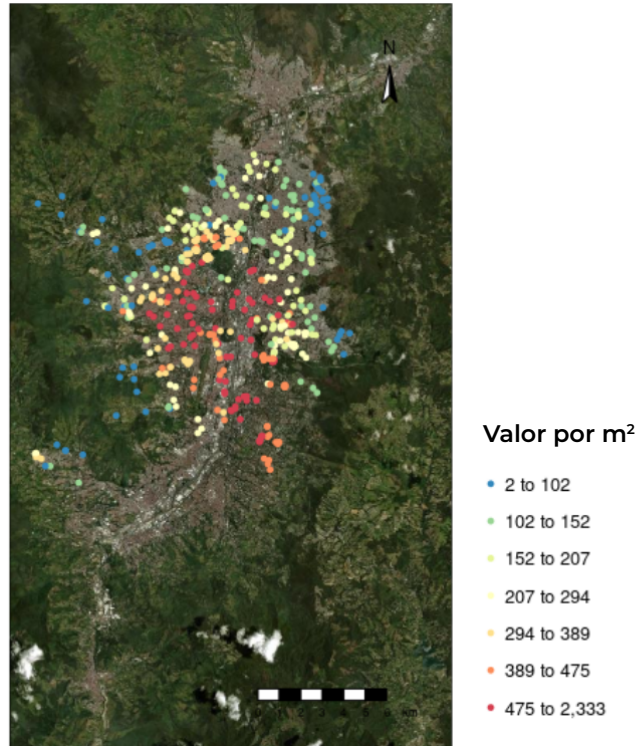
4.7.1. Caso de aplicación: Valuación del impacto de una nueva vía principal en Medellín, Colombia

Para ejemplificar la aplicación de esta función, se considera simular el impacto sobre el precio de la tierra de la construcción de una nueva vía primaria en la ciudad de Medellín (Colombia), a partir del modelo estimado mediante la función `entrenar_modelo()`.

Por lo tanto, se debe estimar un modelo que permita predecir el precio de la tierra en el área de estudio. De manera análoga a la que se usó en el ejemplo de la ciudad de Cali, el primer paso es obtener una muestra georreferenciada de los lotes en venta. En este caso, se utilizan los datos de ejemplo que integran la documentación de ayuda de la librería `'valuate'`, los cuales fueron extraídos como datos abiertos del Mapa de Valores del Suelo en América Latina.²³ Se trata de una submuestra de 500 observaciones de lotes en venta, cuya distribución espacial se describe en el mapa 4.14.

23 Para más información, véase el siguiente enlace: <https://valorsueloamericalatina.org/>.

Mapa 4.14. Distribución espacial de la muestra de valores de suelo en la ciudad de Medellín, Colombia (en US\$ por m²)



Fuente: Elaboración propia.

La primera medida que se debe tomar para simular escenarios es estimar un modelo predictivo que, a partir de diferentes variables independientes, pueda determinar el precio de los inmuebles (en este caso, el precio de la tierra) en toda el área de estudio.

Por lo tanto, mediante las funciones `calcular_dist()`, `calcular_entorno()` y `calcular_raster()` se construyen diferentes variables independientes y se procede a la utilización de la función `entrenar_modelo()`, presentada en la sección anterior.

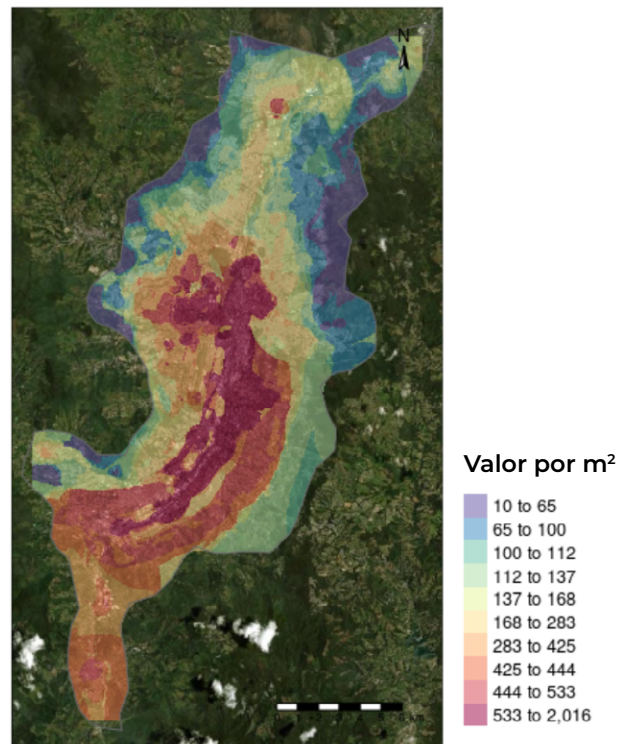
Las variables independientes calculadas para este ejercicio son:

- Distancia a predios de enterramiento de basura (con base en OSM, 2024).
- Distancia a predios industriales (con base en OSM, 2024).
- Distancia a plazas o parques (con base en OSM, 2024).
- Distancia a vías de tren (con base en OSM, 2024).
- Distancia a autopistas (con base en OSM, 2024).
- Distancia a vías primarias (con base en OSM, 2024).
- Distancia a vías secundarias (con base en OSM, 2024).
- Cantidad de hoteles en el entorno (con base en OSM, 2024).
- Proporción del espacio cubierto por plazas o parques en el entorno (con base en OSM, 2024).
- Cantidad de comercios en el entorno (con base en OSM, 2024).

- Altura de las edificaciones en el entorno (con base en Pesaresi et al., 2024).
- Proporción del espacio cubierto por edificaciones en el entorno (con base en Pesaresi et al., 2024).

Con estos insumos se hace uso de la función `entrenar_modelo()` a fin de entrenar el algoritmo Quantile Regression Forest para predecir el precio de la tierra en toda el área urbana de Medellín. Los resultados se muestran en el mapa 4.15. El modelo resultante arrojó un MAPE igual al 16%, lo cual es indicativo de un desempeño muy aceptable en función de la escasa cantidad de datos (solo 500 observaciones) y la realización del cálculo de las variables independientes utilizando únicamente información libre y abierta.

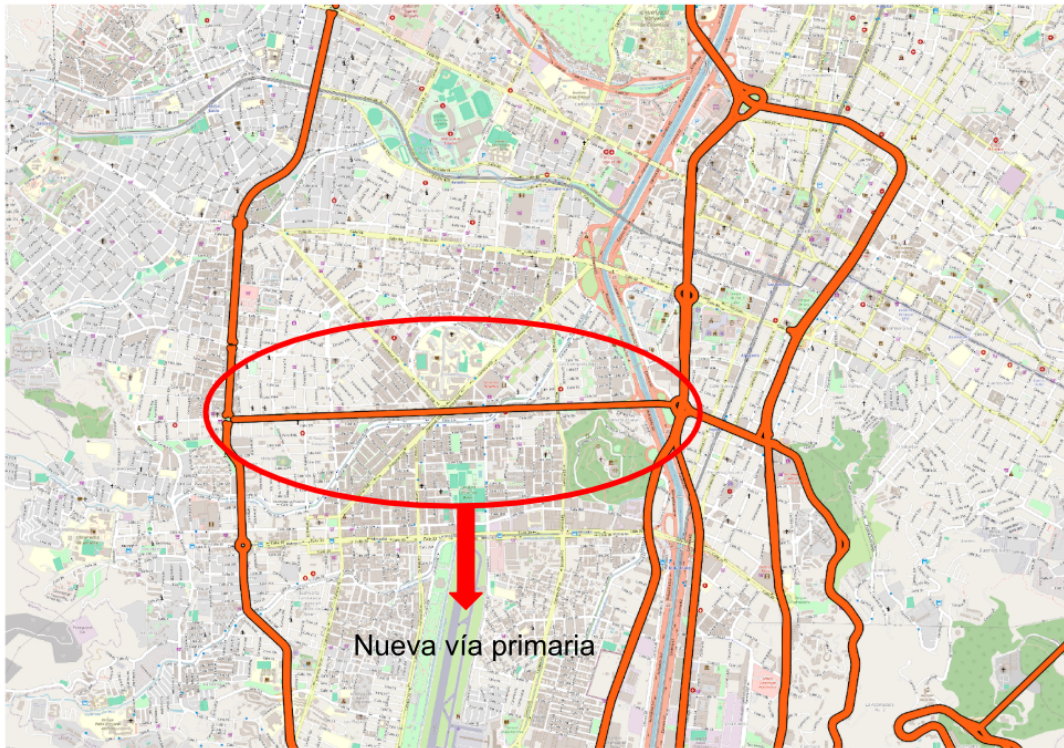
Mapa 4.15. Estimación del valor de la tierra en el área urbana de Medellín, Colombia (en US\$ por m²)



Fuente: Elaboración propia.

Una vez que se cuenta con el modelo predictivo se pueden simular escenarios urbanos. El proceso consiste en redefinir una o más variables independientes para informar una intervención pública en el espacio y volver a realizar la predicción con el modelo original, que ya ha incorporado las relaciones matemáticas entre cada una de estas variables independientes con la variable objetivo (el precio por metro cuadrado de la tierra, en este caso).

La intervención simulada en este ejercicio se basa en la construcción hipotética de una nueva vía primaria en la calle 33, actualmente catalogada como vía secundaria, en el entorno urbano detallado por el mapa 4.16. La nueva vía primaria es la que corta el centro del mapa de este a oeste.

Mapa 4.16. Simulación de una nueva vía primaria en Medellín, Colombia

Fuente: Elaboración propia.

En primer lugar, para dar cuenta de la situación simulada, se debe recalculer la variable “`dist_vias_prim`” mediante la utilización de la función `calcular_dist()`. A tal fin, en cualquier herramienta de gestión de información geográfica se agrega la nueva línea correspondiente a esta vía hipotética a la capa de vías primarias descargada originalmente desde OSM (2024). Esta nueva capa se guarda en una nueva carpeta llamada “simulación”, dentro del entorno de trabajo del proyecto. Con este insumo listo, se está en condiciones de ejecutar la función `simular_escenario()`.

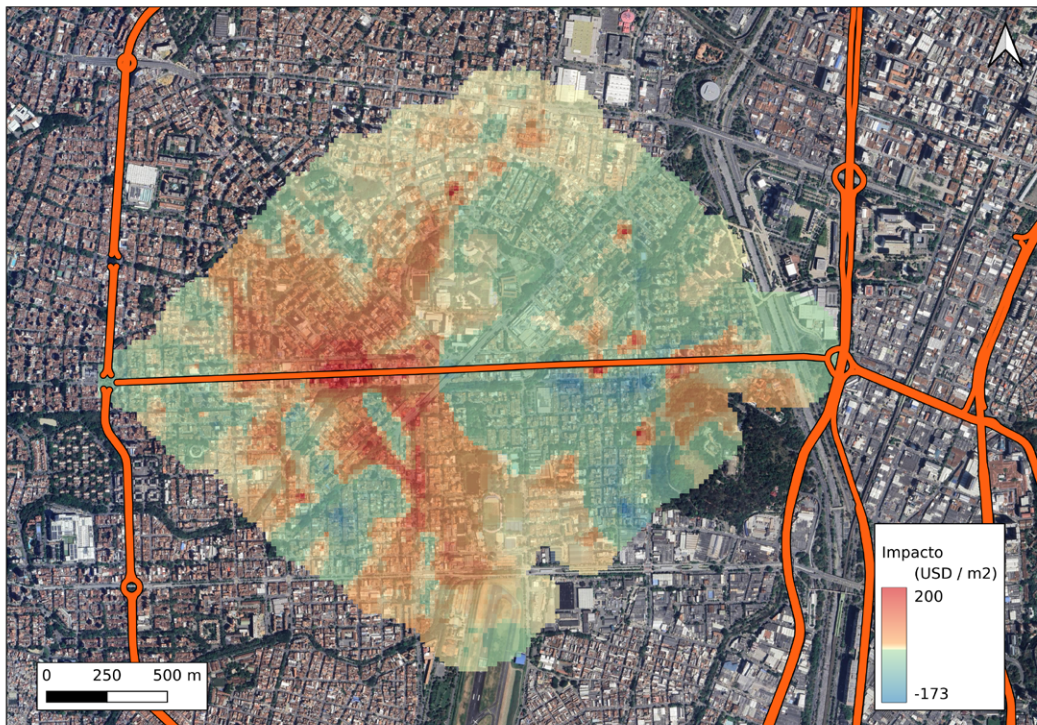
En el argumento “modelo” se debe indicar simplemente “modelo.rda”, ya que la función `entrenar_modelo()` guarda automáticamente el archivo con ese nombre en el directorio de trabajo. En la lista de variables independientes se debe proceder de manera análoga a lo hecho con la función `entrenar_modelo()`, con la salvedad de que es necesario proveer al ráster la ruta de la nueva variable independiente (o nuevas variables independientes) que simula la situación hipotética. En este caso, se ha recalculado la variable “`dist_vias_prim`” y se ha guardado el nuevo ráster resultante en la carpeta “simulación”. Las demás variables deben ser las mismas que se utilizaron en la estimación del modelo original. Finalmente, el último argumento hace referencia a la localización de la estimación original.

```

simular_escenario(modelo = "modelo.rda",
  independientes = c("simulacion/dist_vias_prim.tif",
    "dist_basura.tif",
    "dist_industria.tif",
    "dist_espacios_verdes.tif",
    "dist_tren.tif",
    "dist_vias_autopista.tif",
    "dist_vias_sec.tif",
    "entorno_altura.tif",
    "entorno_espacios_verdes.tif",
    "entorno_comercios.tif",
    "entorno_edificaciones.tif",
    "entorno_hoteles.tif"),
  original = "vut.tif")

```

Mapa 4.17. Efecto esperado de una nueva vía primaria en Medellín, Colombia



Fuente: Elaboración propia.

Como se puede apreciar en el mapa 4.17, el principal impacto se concentra en torno de la nueva vía simulada. No obstante, en sectores que se encuentran hacia el cuadrante sudeste, próximos a un espacio verde (Cerro Nutibara), la intervención se traduce en una reducción del precio de la tierra. La posibilidad de obtener impactos positivos en algunas partes del espacio y negativos en otras es una característica superadora de los modelos de aprendizaje automático (ML), con relación a las técnicas de regresión lineal, ya que estas últimas ofrecen parámetros constantes para toda el área de estudio, lo que constituye una limitación al intentar valorar las externalidades provocadas por diferentes intervenciones urbanas que pueden ofrecer soluciones en algunos sectores de la ciudad y externalidades negativas en otros, como congestión vehicular, inseguridad o ruidos molestos.

En total, **la intervención tiene un valor externo neto, capitalizado en el valor de mercado del suelo urbano, de US\$37.182.410**. Este cálculo surge de considerar que cada píxel de 20 metros por 20 metros representa una superficie de 400 metros cuadrados y que el impacto neto de la intervención parte de un mínimo de -US\$173 por metro cuadrado, con el primer cuartil igual a -US\$5,11, el segundo cuartil igual a US\$1,65, el tercer cuartil igual a US\$14,03 y el máximo igual a US\$521,49. La media del impacto de la intervención es igual a US\$8,50 por metro cuadrado.

El gobierno local dispone ahora de un elemento adicional de crucial importancia para evaluar el costo total de la intervención contra el beneficio neto externo generado por la misma.

4.7.2. Caso de aplicación: Valuación del impacto de un nuevo espacio verde en Belo Horizonte, Brasil

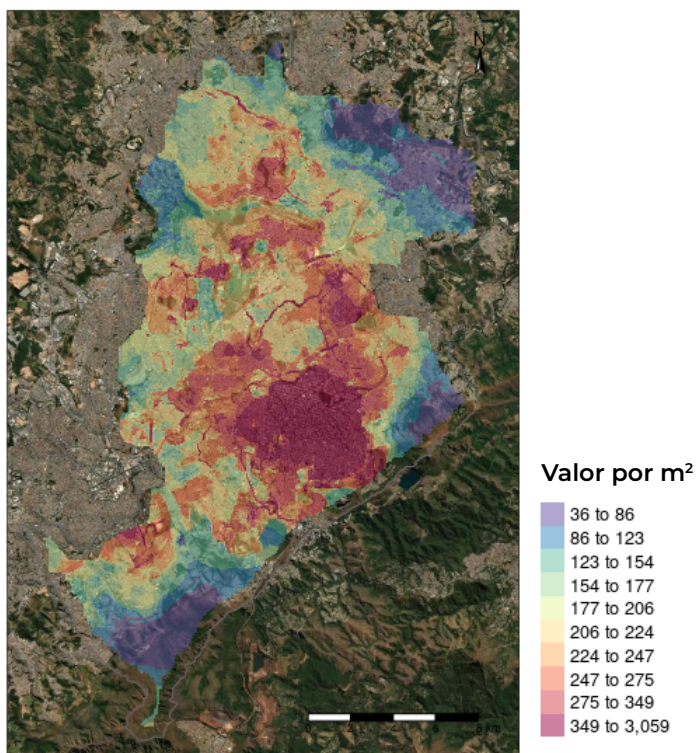
Se procede a estimar el impacto generado por la construcción de una nueva plaza en un sector de Belo Horizonte, en Brasil. De manera análoga al ejemplo anterior, el primer paso es estimar un modelo predictivo del valor de la tierra. Para ello, se utiliza una muestra de 7.041 lotes en venta extraída del Mapa de Valores de América Latina.²⁴ Como se trata de un ejercicio cuyo objetivo es ejemplificar la aplicación de las funciones contenidas en la librería **'valuate'**, solo se calculan nueve variables independientes a partir de fuentes de información libre y abierta, mediante las funciones `calcular_dist()`, `calcular_entorno()` y `calcular_raster()`:

- Distancia a vías primarias (con base en OSM, 2024).
- Distancia a vías secundarias (con base en OSM, 2024).
- Cantidad de hoteles en el entorno (con base en OSM, 2024).
- Cantidad de bancos en el entorno (con base en OSM, 2024).
- Cantidad de restaurantes en el entorno (con base en OSM, 2024).
- Proporción del espacio cubierto por plazas o parques en el entorno (con base en OSM, 2024).
- Altura de las edificaciones en el entorno (con base en Pesaresi et al., 2024).
- Población en el entorno (con base en Pesaresi et al., 2024).
- Proporción del espacio cubierto por edificaciones en el entorno (con base en Pesaresi et al., 2024).

Como se explicó anteriormente, se introduce esta información en la función `entrenar_modelo()` para estimar un modelo predictivo del precio de la tierra urbana en toda el área de estudio. En función de los resultados de esta función, el mapa 4.18 muestra la distribución del precio de la tierra por metro cuadrado en la ciudad de Belo Horizonte, con un MAPE igual a +/- 12,4%.

²⁴ Para más información, véase el siguiente enlace: <https://valorsueloamericalatina.org/>.

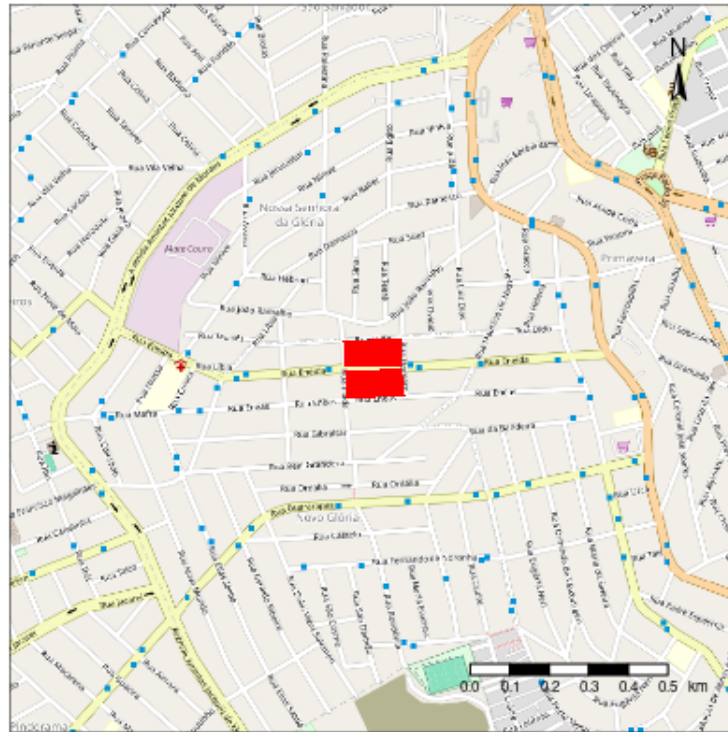
Mapa 4.18. Estimación del valor del suelo en Belo Horizonte, Brasil (en US\$ por m²)



Fuente: Elaboración propia.

De manera análoga al ejemplo anterior, se recalcula la variable que informa de la intervención en el espacio público. En este caso, la variable que captura la proporción del espacio cubierto por plazas o parques en el entorno. Se simula, entonces, la construcción de un nuevo espacio verde en la región noroeste de la ciudad, sobre Rua Eneida, entre Rua Ilíada y Rua Homero. Se trata de un área de dos manzanas, con un espacio total de 1,8 hectáreas, cuya localización se puede apreciar en el mapa 4.19.

Mapa 4.19. Simulación de un nuevo espacio verde en Belo Horizonte, Brasil

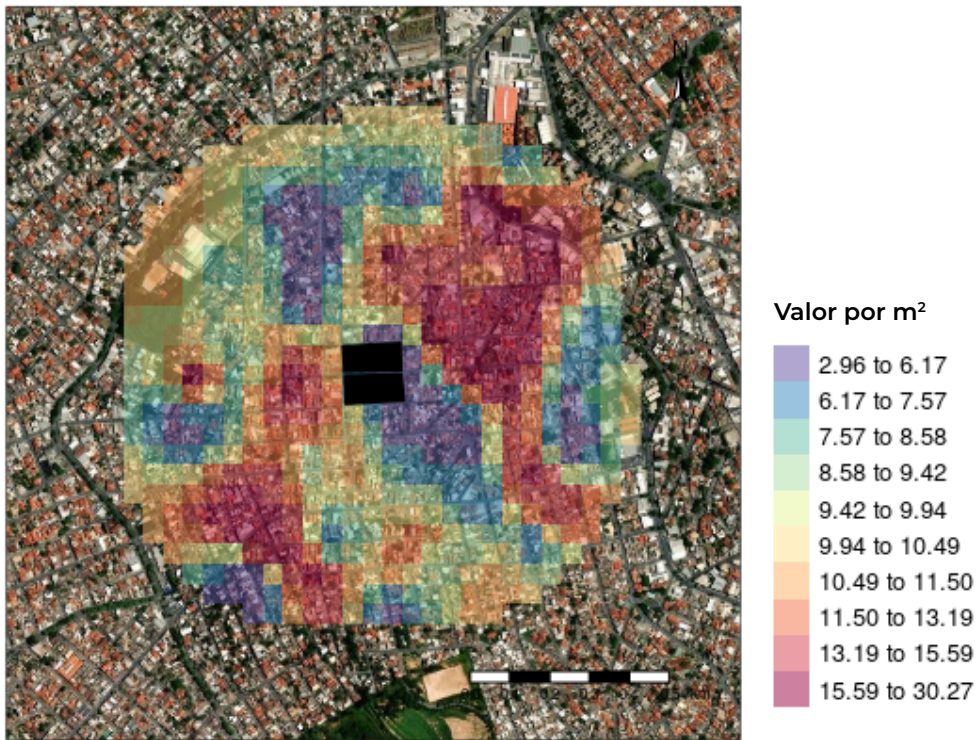


Fuente: Elaboración propia.

Luego de incorporar este nuevo polígono a las plazas y los parques descargados de OSM, se recalcula la variable correspondiente mediante la función `calcular_entorno()` y el ráster resultante se guarda en una nueva carpeta llamada "simulación". Al igual que en el ejemplo anterior, se procede a cargar todos los argumentos necesarios para ejecutar la función `simular_escenario()`:

```
simular_escenario(modelo = "modelo.rda",
                  independientes = c("simulacion/entorno_espacios_verdes.tif",
                                   "dist_vias_prim.tif",
                                   "dist_vias_sec.tif",
                                   "entorno_bancos.tif",
                                   "entorno_hoteles.tif",
                                   "entorno_restaurantes.tif",
                                   "entorno_altura.tif",
                                   "entorno_poblacion.tif",
                                   "entorno_edificaciones.tif"),
                  original = "vut.tif")
```

Mapa 4.20. Efecto esperado de un nuevo espacio verde en Belo Horizonte, Brasil (en US\$ por m²)



Fuente: Elaboración propia.

Como se puede apreciar en el mapa 4.20, la creación del nuevo espacio verde tiene un impacto positivo en el precio de la tierra del entorno, que se concentra hacia el noreste, contra la Avenida Brigadeiro Gomes. En total, si se considera el precio por metro cuadrado de cada píxel y la superficie de cada uno de ellos (50 metros por lado, es decir 2.500 metros cuadrados), **la intervención tiene un valor neto, capitalizado en el valor de mercado del suelo urbano, de US\$14.915.896.** El impacto neto de la intervención parte de un mínimo de US\$2,96 por metro cuadrado, con el primer cuartil igual a US\$8,06, el segundo cuartil igual a US\$9,94, el tercer cuartil igual a US\$12,08 y el máximo igual a US\$30,27. La media del impacto de la intervención es igual a US\$10,54 por metro cuadrado. Es importante resaltar que, a diferencia de la simulación anterior sobre una nueva vía primaria de circulación que mostraba sectores donde el impacto se traducían en un menor valor, el impacto del nuevo espacio verde es positivo en todos los sectores. Además, el conjunto de herramientas propuestas proporciona una estimación objetiva del valor de las propiedades existentes en el sector donde se propone la intervención hipotética, por si es necesario realizar expropiaciones.

4.7.3. Caso de aplicación: Estimación de las externalidades generadas por la disminución del riesgo de incendio en Córdoba, Argentina

El objetivo de este ejemplo es identificar el impacto de la eliminación de factores de riesgo de incendio en el área metropolitana de la ciudad de Córdoba, Argentina. Para la definición de riesgo, se toma como base el trabajo realizado por IDECOR (2023), que define el riesgo de incendio de la siguiente manera: “El mapa de riesgo muestra niveles de bajo riesgo en zonas de alta compacidad urbana y, en menor medida, en áreas agrícolas, donde, debido a la baja continuidad y distribución de la cobertura vegetal no favorece la propagación de los incendios forestales. Niveles de riesgo medios se distribuyen en zonas de interfase clásica y en áreas con cobertura de vegetación del tipo matorral/arbustal y pastizales. En cuanto a los niveles altos de riesgo se encuentran en zonas de interfase mixta, sobre todo donde existen viviendas aisladas inmersas en una matriz de vegetación”. Se trabaja, entonces, sobre la categoría de nivel alto de riesgo de incendio, conforme a la definición anterior y a la geolocalización de los polígonos que cumplen con estas características, según los datos de Infraestructura de Datos Espaciales de la Provincia de Córdoba (IDECOR).²⁵

Como en los ejemplos anteriores, la variable dependiente es el valor de mercado por metro cuadrado de la tierra urbana en dólares (mapa 4.21). La muestra utilizada contiene 1.139 lotes baldíos en venta o vendidos, disponibles como datos abiertos en IDECOR.²⁶

Además del riesgo de incendio mencionado, se calculan las siguientes variables independientes:

- Riesgo (0 = sin riesgo elevado, 1 = riesgo elevado) (con base en IDECOR, 2024).
- Distancia al banco más cercano (con base en OSM, 2024).
- Distancia al barrio cerrado más cercano (con base en OSM, 2024).
- Distancia al barrio popular o asentamiento informal más cercano (con base en el Registro Nacional de Barrios Populares [RENABAP]).²⁷
- Distancia a la industria más cercana (con base en OSM, 2024).
- Distancia a la vía primaria más cercana (con base en OSM, 2024).
- Distancia a la vía secundaria más cercana (con base en OSM, 2024).
- Cantidad de unidades habitacionales en un entorno de 500 metros (con base en IDECOR, 2024).
- Porcentaje del territorio que se encuentra edificado en un entorno de 500 metros (con base en Pesaresi et al., 2024).
- Promedio del cociente entre la cantidad de metros cuadrados edificados en cada parcela y su superficie, en un entorno de 500 metros (con base en IDECOR, 2024).
- Cantidad de parcelas en un entorno de 500 metros (con base en IDECOR, 2024).
- Porcentaje del territorio que se encuentra ocupado por plazas o parques en un entorno de 500 metros (con base en OSM, 2024).

25 Para más información, véase el siguiente enlace: <https://mapascordoba.gob.ar/viewer/mapa/353>.

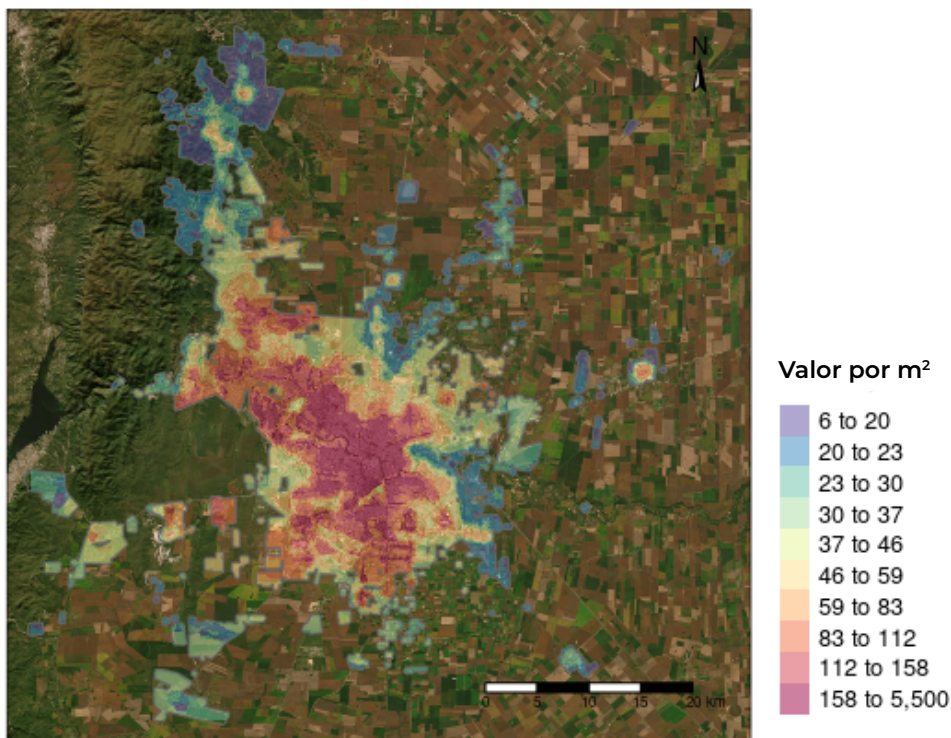
26 Para más información, véase el siguiente enlace: <https://mapascordoba.gob.ar/#/descargas>.

27 Para más información, véase el siguiente enlace: <https://www.argentina.gob.ar/obras-publicas/sisu/renabap/mapa>.

Se observa que un 8,5% de la muestra corresponde a lotes ubicados dentro de áreas con riesgo de incendio elevado. Como se ha dicho, el primer paso es entrenar un modelo de ML para que prediga el precio de la tierra en toda el área de estudio, mediante la utilización de la función `entrenar_modelo()`.

A fin de establecer el impacto de la simulación de una intervención pública, que consiste en eliminar un polígono sujeto a riesgo elevado, se entrena el algoritmo Quantile Regression Forest. Mediante un proceso de validación cruzada en 10 grupos, el error fuera de la muestra se estima en +/- 15,11%.

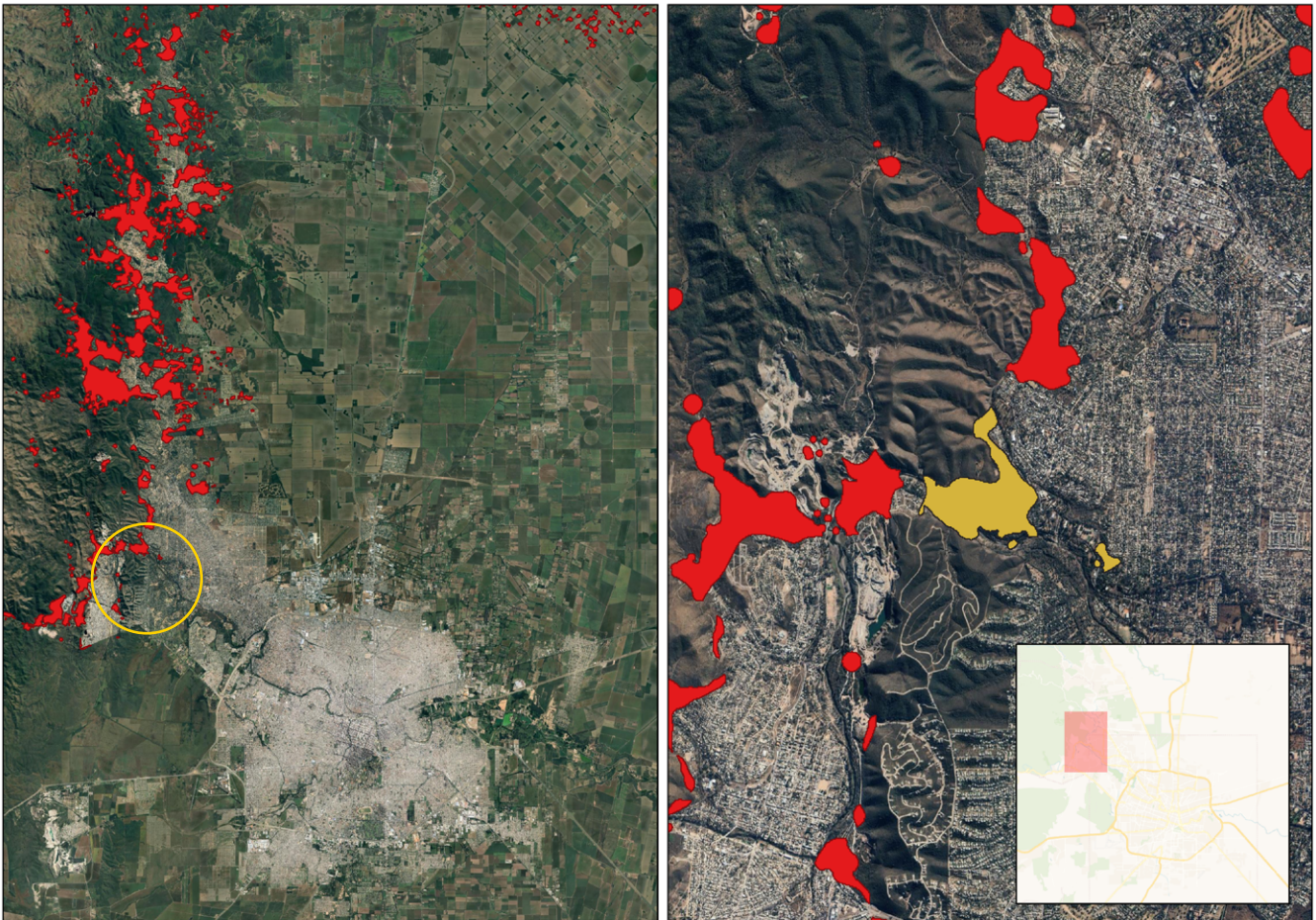
Mapa 4.21. Estimación del valor de la tierra en el área metropolitana de la ciudad de Córdoba, Argentina (en US\$ por m²)



Fuente: Elaboración propia.

A continuación, se procede a simular la eliminación de un polígono de 34 hectáreas sujeto a riesgo de incendio elevado. Las zonas de riesgo de incendio elevado se encuentran al noroeste de la ciudad, un sector de sierras bajas, con áreas de bosque nativo, que está sujeto hace algo más de dos décadas a un proceso de presión urbanística muy importante. Se elimina el polígono marcado en amarillo en el mapa 4.22, para simular una intervención que reduce el riesgo de incendio en la zona.

Mapa 4.22. Simulación de la eliminación de un polígono para reducir el riesgo de incendio en la ciudad de Córdoba, Argentina



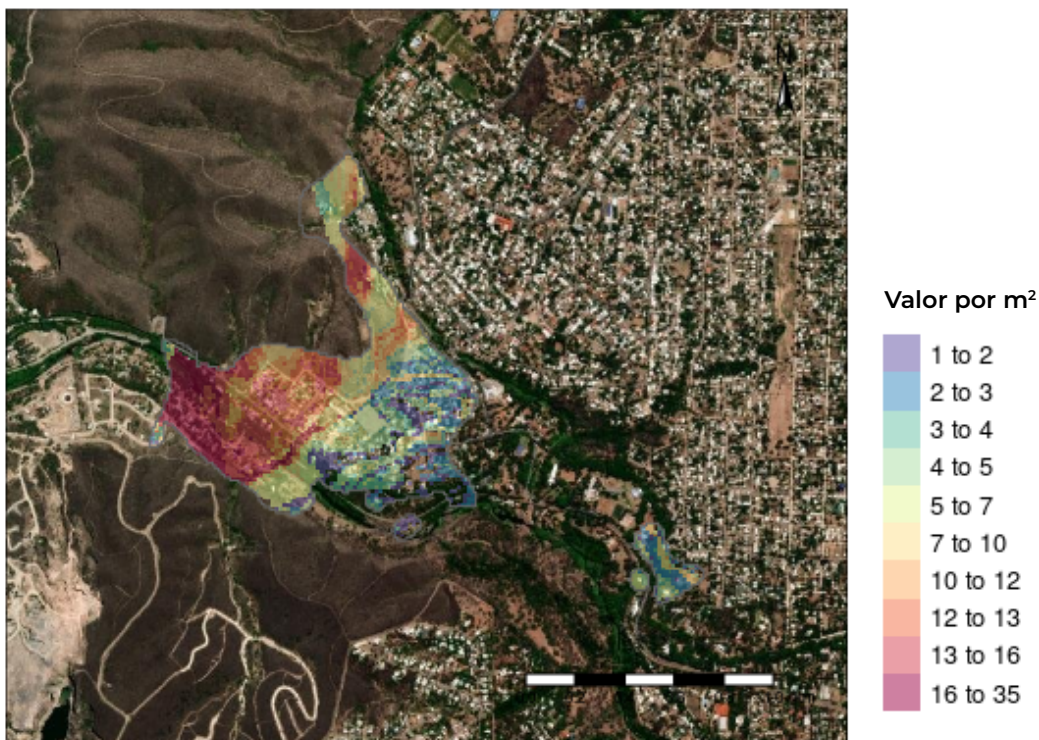
Fuente: Elaboración propia.

Luego de remover el polígono señalado y recalcular la variable "riesgo", se procede a cargar los argumentos correspondientes a la función `simular_escenario()`:

```
simular_escenario(modelo = "modelo.rda",
                  independientes = c("simulacion/riesgo.tif",
                                   "dist_bancos.tif",
                                   "dist_barrio_cerrado.tif",
                                   "dist_barrios_populares.tif",
                                   "dist_industrias.tif",
                                   "dist_vias_primarias.tif",
                                   "dist_vias_secundarias.tif",
                                   "entorno_cuentas.tif",
                                   "entorno_edificaciones.tif",
                                   "entorno_fot.tif",
                                   "entorno_parcelas.tif",
                                   "entorno_plazas.tif"),
                  original = "vut.tif")
```

El uso de la función `simular_escenario()` indica que la intervención pública para eliminar el riesgo elevado de incendio en el sector analizado **genera un aumento total del precio de la tierra igual a US\$6.772.725**. Esto implica un **incremento del precio de la tierra del 13,31%** para toda el área sujeta a la intervención. La distribución espacial del impacto de la intervención se puede observar en el mapa 4.23.

Mapa 4.23. Efecto esperado de la reducción del riesgo de incendio en Córdoba, Argentina (en US\$ por m²)



Fuente: Elaboración propia.



CAPÍTULO 5. Lineamientos de acción y consideraciones finales

HUÁSCAR EGUINO Y JUAN PABLO CARRANZA



La presente guía tiene tres propósitos: En primer lugar, proporcionar un marco de referencia teórico sobre las técnicas de valuación masiva automatizada (AVM); en segundo lugar, contribuir a mejorar las capacidades técnicas de los especialistas interesados en usar las técnicas de valuación inmobiliaria en las finanzas públicas y el desarrollo urbano; y en tercer término, extender el campo de aplicación de los AVM a aspectos que van más allá de las finanzas públicas locales, para abarcar campos de suma importancia para la gestión territorial como la priorización de inversiones públicas, la valoración de externalidades y la gestión de riesgos climáticos y ambientales. A modo de conclusión, se destacan los siguientes conceptos:

Importancia de la valuación inmobiliaria en la gestión pública subnacional. A lo largo del documento se demuestra que la valuación o tasación inmobiliaria es mucho más que una herramienta técnica: se trata de un elemento estratégico para la gestión de los gobiernos subnacionales. Lejos de ser un instrumento meramente técnico, la correcta estimación del valor de los bienes inmuebles permite a los gobiernos subnacionales fortalecer su capacidad de generar ingresos, implementar políticas urbanas y mejorar su capacidad de respuesta frente a riesgos climáticos, entre otros.

La modernización de los catastros y los sistemas de valuación no solo mejora la administración tributaria, en particular respecto del impuesto a la propiedad, sino que también facilita la implementación de instrumentos financieros basados en la captura de la valorización inmobiliaria, como las contribuciones por mejoras y los esquemas de financiamiento por incremento de impuestos. Además, el hecho de contar con información actualizada, desagregada objetiva y territorialmente del valor de los inmuebles, permite que los gobiernos locales tomen decisiones más objetivas, transparentes, eficientes y oportunas en materia de ordenamiento territorial y planificación de las inversiones en infraestructuras y servicios.

Avances tecnológicos y su impacto en las prácticas de valuación. El documento describe en detalle la evolución metodológica de las técnicas de valuación masiva, desde los enfoques tradicionales, basados en regresiones lineales y zonas homogéneas, hasta la incorporación de elementos más complejos y precisos como los algoritmos de aprendizaje automático (ML). Tanto en el ámbito científico como en el área de implementación de políticas de valuación inmobiliaria se destacan algoritmos como Random Forest, Quantile Regression Forest y XGBoost, que permiten procesar grandes volúmenes de datos con un alto grado de precisión, incluso en contextos con información incompleta o dispersa.

Estos avances tecnológicos, combinados con el creciente acceso a fuentes de datos abiertos, ha generado nuevas oportunidades para diseñar e implementar políticas basadas en inteligencia artificial (IA). No obstante, también se reconocen limitaciones, como la falta de capacidades técnicas locales, el riesgo de sesgos y opacidad en los modelos utilizados (“cajas negras”), y los desafíos éticos asociados al tipo de información que se usa en las decisiones de tipo fiscal o urbanístico. Por ello, se destaca la necesidad de acompañar la aplicación de avances tecnológicos con procesos de formación, institucionalización y construcción de confianza en la ciudadanía, así como implementar políticas de transparencia y datos abiertos que proporcionen legitimidad a este tipo de enfoques.

Usos y aplicaciones prácticas. Uno de los mayores aportes de la guía radica en la inclusión de casos de estudio de ciudades latinoamericanas como Córdoba (Argentina), Fortaleza (Brasil), Cali, Bogotá y Medellín (Colombia). Estas experiencias muestran que los modelos de valuación masiva basados en IA se pueden aplicar en contextos urbanos diversos y contribuyen a mejorar varios ámbitos de la gestión subnacional, como la recaudación del impuesto a la propiedad, el acceso a financiamiento con base en la valorización inmobiliaria, la gestión de riesgos físicos y la gestión de inversiones públicas.

En particular, merecen destacarse los casos de Medellín y Bogotá. En Medellín, el uso de modelos de valuación masiva sirvió de base para implementar un esquema de financiamiento urbano a través de un distrito con financiamiento por incremento de impuestos (distrito TIF). En Bogotá, en tanto, estos modelos fueron de utilidad para aplicar de forma efectiva el financiamiento de obras de infraestructura mediante las contribuciones por valorización.

Instrumentos de gestión urbana basados en el valor del suelo. La posibilidad de simular escenarios futuros, como cambios normativos, desarrollos inmobiliarios, riesgos climáticos o intervenciones en infraestructura, es una de las principales virtudes de los modelos de valuación masiva. De esta forma es posible anticipar impactos, planificar inversiones y diseñar políticas públicas más ajustadas a la realidad y a las necesidades identificadas en el territorio. Además, otorga a los gobiernos subnacionales una herramienta poderosa para monetizar el valor generado por sus intervenciones, lo cual permite una gestión financiera más sostenible.

El estudio destaca que las reformas urbanísticas y de construcción, como los cambios en los coeficientes de construcción en Curitiba (Brasil), pueden ser más efectivos y transparentes si se basan en modelos de valuación bien diseñados. También revela que el uso conjunto de los mapas de riesgo y los modelos de valuación (como en el caso de Córdoba, Argentina) puede ser una herramienta de enorme utilidad para simular el impacto de eventos climáticos y ambientales, y establecer medidas de mitigación, como las inversiones en resiliencia urbana.

Código abierto para implementar valuaciones masivas. Uno de los principales aportes de esta guía es el diseño y la apertura de una librería en el lenguaje de programación R, que facilita la realización de múltiples cálculos complejos en las diferentes etapas del proceso de implementación de una política de actualización masiva de las valuaciones inmobiliarias.

La librería **'valuate'** está orientada a los cuadros técnicos de la administración pública y permite realizar, de manera relativamente sencilla para un usuario con formación inicial en estadística y manejo de lenguajes de programación, una serie de cálculos que son esencialmente complejos e implican la manipulación de información geográfica. El alcance de la librería comprende el cálculo de variables territoriales independientes, el tratamiento y homogeneización de la muestra y el entrenamiento de algoritmos de ML para predecir el valor de los inmuebles.

Quizás la innovación más relevante es la incorporación de una función (**simular_escenario()**) que permite valorar de forma objetiva inversiones públicas, eventos climáticos o cambios de normativa edilicia, lo cual contribuye a la implementación de políticas públicas basadas en evidencia.

Responsabilidades y capacidades de los gobiernos subnacionales. Finalmente, el documento pone en evidencia la importancia de apoyar a los gobiernos subnacionales interesados en transitar hacia una gestión más moderna y eficiente del valor inmobiliario. Al respecto, se sugieren las siguientes acciones:

- *Invertir en datos e infraestructura tecnológica.* Se requiere fortalecer los observatorios inmobiliarios, los sistemas de información geográfica y los catastros digitales, que son la base para el uso de los modelos de valuación masiva.
- *Mejorar la capacitación del personal técnico.* Los gobiernos subnacionales deben invertir en la formación de sus equipos técnicos en temas como estadística, programación, IA, gestión de datos espaciales, sistemas catastrales y financiamiento subnacional.
- *Establecer alianzas con universidades y el sector privado.* La colaboración con el mundo académico y con el sector privado (en particular, las empresas de seguros y las entidades financieras que proporcionan financiamiento hipotecario) puede suplir las carencias técnicas y fomentar la innovación. Por este motivo, es crucial que los gobiernos subnacionales trabajen de manera coordinada con las entidades académicas locales y con el sector privado en los proyectos de seguimiento de la dinámica urbana y de monitoreo de los valores de los bienes inmuebles.
- *Revisar el marco normativo y establecer las necesidades de reforma.* Las reformas legales deben permitir la adopción de nuevas tecnologías, sin que ello implique una pérdida de transparencia o de derecho a la privacidad. En especial, es importante que las normas tributarias, catastrales y urbanísticas faciliten el uso de las técnicas de AMV como parte de las herramientas usadas por los encargados de la toma de decisiones.
- *Estrategias de comunicación.* Es fundamental explicar de forma clara cómo se valoran las propiedades y cómo se usan estos valores para diseñar políticas públicas. Esto contribuye a reducir la resistencia a la aplicación de las técnicas de AMV y de IA en la gestión financiera y la planificación subnacional.

REFERENCIAS

- Alonso, W. 1964. *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press. Disponible en: <https://doi.org/10.4159/harvard.9780674730854>.
- Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht, Países Bajos: Springer Dordrecht. Disponible en: <https://doi.org/10.1007/978-94-015-7799-1>.
- 1995. Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2):93-115. Disponible en: <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- Barreix, A., M. Bès, U. González, J. Roca y F. Velayos. 2023. El estado actual del impuesto al patrimonio en América Latina. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0004791>.
- BID (Banco Interamericano de Desarrollo). 2024. Perú: Mejoramiento de la gestión de las finanzas públicas subnacionales para la sostenibilidad fiscal (PE-L1278). Propuesta de préstamo. Washington, D.C.: BID. Disponible en: <https://www.iadb.org/es/proyecto/PE-L1278>.
- Bidanset, P. y R. Rakow. 2022. Survey On the Use of Automated Valuation Models (AVMs) in Government Assessment Offices: An Analysis of AVM Use, Acceptance, and Barriers to More Widespread Implementation. *Journal of Property Tax Assessment & Administration*, 19(2). Disponible en: <https://researchexchange.iaao.org/jptaa/vol19/iss2/3>.
- Blanco Blanco, A., V. Fretes Cibils y A. Muñoz Miranda (eds.). 2016. Expandiendo el uso de la valorización del suelo: La captura de plusvalías en América Latina y el Caribe. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0012751>.
- Blanco Blanco, A., N. Moreno, D. Vetter y M. Vetter. 2016. El potencial de la captura de plusvalías para la financiación de proyectos urbanos: Consideraciones metodológicas y casos prácticos. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0000597>.
- Bonet, J., A. Muñoz Miranda, C. Pineda Mannheim, L. Díaz Frers, L. Castro, C. De Cesare, R. Alves Dantas, et al. 2014. El potencial oculto: Factores determinantes y oportunidades del impuesto a la propiedad inmobiliaria en América Latina. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0012692>.
- Bottou, L., F. Curtis y J. Nocedal. 2018. Optimization Methods for Large-Scale Machine Learning. *SIAM Review*, 60(2):223-311. Disponible en: <https://doi.org/10.1137/16M1080173>.
- Breiman, L. 1996. Bagging predictors. *Machine Learning*, 24(2):123-140. Disponible en: <https://doi.org/10.1007/BF00058655>.
- 2001. Random Forests. *Machine Learning*, 45(1):5-32. Disponible en: <https://doi.org/10.1023/A:1010933404324>.
- Breiman, L., J. Friedman, R.A. Olshen y C. Stone. 1984. *Classification and Regression Trees*. Nueva York: Chapman and Hall/CRC. Disponible en: <https://doi.org/10.1201/9781315139470>.
- Bueno Villegas, P., J. Alcántara Obregón, V. Leyva Picazo y S. Hernández López. 2022. Áreas de oportunidad para la actualización de valores catastrales de suelo habitacional en zonas urbanas. *Perspectivas de la Ciencia y la Tecnología*, 5(8):70-81. Disponible en: <https://revistas.uaq.mx/index.php/perspectivas/article/view/680>.

- Carranza, J.P. 2024. *Inteligencia artificial aplicada a políticas de valuación masiva de la tierra urbana*. [Tesis de doctorado, Universidad Nacional de Córdoba]. Disponible en: <https://rdu.unc.edu.ar/handle/11086/554984>.
- . 2025. *r-valuate*: Caja de herramientas para la realización de estudios territoriales. R package version 0.1.0. Disponible en: <https://github.com/r-valuate/valuate>.
- Carranza, J.P., M. Piumetto, M. Salomón, F. Monzani, M. Montenegro y M. Córdoba. 2019. Valuación masiva de la tierra urbana mediante inteligencia artificial. El caso de la ciudad de San Francisco, Córdoba, Argentina. *Vivienda y Ciudad*, 6. Disponible en: <https://revistas.unc.edu.ar/index.php/ReViyCi/article/view/27090>.
- Carranza, J.P., M. Piumetto, C. Lucca y E. Da Silva. 2022. Mass Appraisal as Affordable Public Policy: Open Data and Machine Learning for Mapping Urban Land Values. *Land Use Policy*, 119:106211. Disponible en: <https://doi.org/10.1016/j.landusepol.2022.106211>.
- Castro, L., L. Frers, A. Alfieri y A. Bovino. 2014. Desafíos y potencial del impuesto inmobiliario en la Argentina. Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento. Documento de trabajo No. 120. Buenos Aires, Argentina: CIPPEC. Disponible en: <https://www.cippec.org/wp-content/uploads/2017/03/1359.pdf>.
- Cerino, R., J.P. Carranza, M. Piumetto, M. Bullano, F. Monzani y M. Córdoba. 2020. Homogeneización del valor de la tierra urbana mediante técnicas de econometría espacial en valuaciones masivas automatizadas. [Sesión de congreso] 14° Congresso de Cadastro Multifinalitário e Gestão Territorial. 2° Encontro de Professores de Cadastro Territorial. Modalidad virtual. Disponible en: <https://rdu.unc.edu.ar/handle/11086/28446>.
- Cerino, R., J.P. Carranza, M. Piumetto, M. Bullano, V. Donalizio y F. Monzani. 2021. Propuesta metodológica para la valuación masiva del suelo urbano. Aplicación espacial del algoritmo Quantile Regression Forest. *Vivienda y Ciudad*, 8. Disponible en: <https://revistas.unc.edu.ar/index.php/ReViyCi/article/view/35166>.
- Córdoba, M., J.P. Carranza, M. Piumetto, F. Monzani y M. Balzarini. 2021. A Spatially Based Quantile Regression Forest Model for Mapping Rural Land Values. *Journal of Environmental Management*, 289:112509. Disponible en: <https://doi.org/10.1016/j.jenvman.2021.112509>.
- De Cesare, C. (ed.). 2016. *Sistemas del impuesto predial en América Latina y el Caribe*. Cambridge, MA: Lincoln Institute of Land Policy. Disponible en: <https://www.lincolninst.edu/publications/books/sistemas-del-impuesto-predial-en-america-latina-el-caribe/>.
- Derdouri, A. e Y. Murayama. 2020. A comparative study of land price estimation and mapping using regression kriging and machine learning algorithms across Fukushima Prefecture, Japan. *Journal of Geographical Sciences*, 30(5):794-822. Disponible en: <https://doi.org/10.1007/s11442-020-1756-1>.
- Eguino, H. y D. Erba (eds.). 2020. *Catastro, valoración inmobiliaria y tributación municipal: Experiencias para mejorar su articulación y efectividad*. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0002437>.
- . 2024. Mercado inmobiliario e impuesto predial: Aplicaciones de técnicas de valuación masiva. Washington, D.C.: BID. Disponible en: <https://doi.org/10.18235/0005488>.
- El Jaouhari, A., A. Samadhiya, A. Kumar, A. Šešplaukis y S. Raslanas. 2024. Mapping the Landscape: A Systematic Literature Review on Automated Valuation Models and Strategic Applications in Real Estate. *International Journal of Strategic Property Management*, 28(5):286-301. Disponible en: <https://doi.org/10.3846/ijspm.2024.22251>.
- Ferreira de Oliveira, A., F. Reyes-Bueno, M. Stumpf Gonzalez y E. da Silva. 2024. Comparing Traditional and Machine Learning Techniques in Apartments Mass Appraisal in Fortaleza, Brazil. *Aestimum*, 84:21-38. Disponible en: <https://oaj.fupress.net/index.php/ceset/article/view/15344>.

- Freund, Y. 1990. Boosting a Weak Learning Algorithm by Majority. En: M.A. Fulk, J. Case (eds.). *Proceedings of the Third Annual Workshop on Computational Learning Theory. University of Rochester, Rochester, New York, August 6-8, 1990*, Volumen 3. Nueva York. Elsevier Science & Technology Books. Disponible en: <https://dl.acm.org/doi/10.5555/92571.92640>.
- Friedman, J. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5):1189-1232. Disponible en: <https://www.jstor.org/stable/2699986>.
- Gao, Q., V. Shi, C. Pettit y H. Han. 2022. Property Valuation Using Machine Learning Algorithms on Statistical Areas in Greater Sydney, Australia. *Land Use Policy*, 123:106409. Disponible en: <https://doi.org/10.1016/j.landusepol.2022.106409>.
- García, E., G. Suárez, M. Esquivel, A. Ruiz, D. Zuloaga y O. Chevalier. 2019. Bases generales para el desarrollo de estudios de reducción de riesgos hidrológicos en ciudades. Lecciones aprendidas de la iniciativa Ciudades Emergentes y Sostenibles ante el reto del cambio climático en Latinoamérica y el Caribe. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0002128>.
- Hastie, T., R. Tibshirani y J. Friedman. 2009. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. (2ª ed.) Nueva York: Springer. Disponible en: <https://doi.org/10.1007/978-0-387-84858-7>.
- Hernangómez, D., J. Lacko y A. White. 2025. `_nominatimlite`: Interface with 'Nominatim' API Service. s.l.: CRAN.R-project.org. Disponible en: <https://cran.r-project.org/web/packages/nominatimlite/index.html>.
- Hijmans, R., M. Barbosa, A. Ghosh y A. Mandel. 2024. `_geodata`: Download Geographic Data_. R package version 0.6-2. s.l.: CRAN.R-project.org. Disponible en: <https://CRAN.R-project.org/package=geodata>.
- Hijmans, R., M. Barbosa; R. Bivand, A. Brown, M. Chirico, E. Cordano et al. 2025a. `_terra`: Spatial Data Analysis_. R package version 1.8-42. s.l.: CRAN.R-project.org. Disponible en: <https://doi.org/10.32614/CRAN.package.terra>.
- Hijmans, R., J. van Etten, M. Sumner, J. Cheng, D. Baston, A. Bevan et al. 2025b. `_raster`: Geographic Data Analysis and Modeling_. R package version 3.6-32. s.l.: CRAN.R-project.org. Disponible en: <https://doi.org/10.32614/CRAN.package.raster>.
- Hong, J., H. Choi y W. Kim. 2020. A House Price Valuation Based on the Random Forest Approach: The Mass Appraisal of Residential Property in South Korea. *International Journal of Strategic Property Management*, 24(3):140-152. Disponible en: <https://doi.org/10.3846/ijspm.2020.11544>.
- IAAO (International Association of Assessing Officers). 2022. Standard on Automated Valuation Models (AVMS). Kansas City, MO: IAAO. Disponible en: https://www.iaao.org/wp-content/uploads/Standard_on_Automated_Valuation_Models.pdf.
- IDECOR (Infraestructura de Datos Espaciales de la Provincia de Córdoba). 2023. Mapas de riesgo local e interfase para incendios forestales. Corredores de Sierras Chicas, Traslasierra, Punilla, Calamuchita y sector norte de la Provincia de Córdoba. Córdoba, Argentina: IDECOR. Disponible en: https://obs-idecor-mapas-docs.obs.sa-argentina-1.myhuaweicloud.com/m353/Mapa_Riesgo_Local_2023.pdf.
- . 2024. Mapas Córdoba. Mapas temáticos y datos libres en forma de geoservicios en línea o a través de descarga directa. Córdoba, Argentina: IDECOR. Disponible en: <https://mapascordoba.gob.ar/#/mapas>.
- Jafary, P., D. Shojaei, A. Rajabifard y T. Ngo. 2024. Automated Land Valuation Models: A Comparative Study of Four Machine Learning and Deep Learning Methods Based on a Comprehensive Range of Influential Factors. *Cities*, 151:105115. Disponible en: <https://doi.org/10.1016/j.cities.2024.105115>.

- Kim, J., J. Won, H. Kim y J. Heo. 2021. Machine-Learning-Based Prediction of Land Prices in Seoul, South Korea. *Sustainability*, 13(23):13088. Disponible en: <https://doi.org/10.3390/su132313088>.
- Kleinberg, J., J. Ludwig, S. Mullainathan y Z. Obermeyer. 2015. Prediction Policy Problems. *American Economic Review*, 105(5):491-495. Disponible en: <https://doi.org/10.1257/aer.p20151023>.
- Krige, D.G. 1951. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119-139. Disponible en: https://hdl.handle.net/10520/AJA0038223X_4792.
- Kuhn, M. 2008. Building Predictive Models in R Using the Caret Package. *Journal of Statistical Software*, 28(5):1-26. Disponible en: <https://doi.org/10.18637/jss.v028.i05>.
- López-Ghio, R., C. Salazar Echavarría y L. Schloeter. 2020. Tax Increment Financing for Urban Projects: An Alternative to Fund Infrastructure in Latin American Cities? The Case of Medellín's Carrera 80 Tram Project. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0002574>.
- Ma, J., J.C. Cheng, F. Jiang, W. Chen y J. Zhang. 2020. Analyzing Driving Factors of Land Values in Urban Scale Based on Big Data and Non-Linear Machine Learning Techniques. *Land Use Policy*, 94:104537. Disponible en: <https://doi.org/10.1016/j.landusepol.2020.104537>.
- Maloney, W., J. Zambrano, G. Vuletin, G. Beylis y P. Garriga. 2024. Impuestos a la riqueza para la equidad y el crecimiento. Informe Económico América Latina y el Caribe. Washington, D.C.: Banco Mundial. Disponible en: <http://hdl.handle.net/10986/42001>.
- MBA (Mortgage Bankers Association). 2019. The State of Automated Valuation Models in the Age of Big Data. Washington, D.C.: MBA. Disponible en: https://www.mba.org/docs/default-source/uploadedfiles/member-white-papers/stateofautomatedvaluationmodels-final.pdf?sfvrsn=6cd37b1_0.
- Meinshausen, N. 2006. Quantile Regression Forests. *Journal of Machine Learning Research*, 7(35):983-999. Disponible en: <https://jmlr.org/papers/v7/meinshausen06a.html>.
- OCDE (Organización para la Cooperación y el Desarrollo Económicos). 2022. Informe de Síntesis 2022. Observatorio Mundial sobre las Finanzas y la Inversión de los Gobiernos Subnacionales. París: OCDE. Disponible en: https://www.sng-wofi.org/SNG-WOFI%20Informe%20de%20Sintesis%202022_Spanish.pdf.
- Odeh, I., A. McBratney y D. Chittleborough. 1995. Further Results on Prediction of Soil Properties from Terrain Attributes: Heterotopic Cokriging and Regression-Kriging. *Geoderma*, 67(3-4):215-226. Disponible en: [https://doi.org/10.1016/0016-7061\(95\)00007-B](https://doi.org/10.1016/0016-7061(95)00007-B).
- OSM (OpenStreetMap). 2023. [Datos obtenidos en 2023]. Cambridge: OSMF. Disponible en: <https://www.openstreetmap.org>.
- . 2024. [Datos obtenidos en 2024]. Cambridge: OSMF. Disponible en: <https://www.openstreetmap.org>.
- Padgham, M., B. Rudis, R. Lovelace y M. Salmon. 2017. osmdata. *Journal of Open Source Software*, 2(14):305. Disponible en: <https://doi.org/10.21105/joss.00305>.
- Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439-446. Disponible en: <https://doi.org/10.32614/RJ-2018-009>.
- Perdomo-Calvo J., C. Mendoza-Álvarez, J. Mendieta-López y A. Baquero-Ruiz. 2007. Study of the Effect of the TransMilenio Mass Transit Project on the Value of Properties in Bogotá, Colombia. Lincoln Institute of Land Policy. Documento de trabajo No. WP07CA1. Cambridge: Lincoln Institute of Land Policy. Disponible en: <https://dx.doi.org/10.2139/ssrn.1586381>.

- Pesaresi, M., M. Schiavina, P. Politis, S. Freire, K. Krasnodębska, J. H. Uhl, A. Carioli et al. 2024. Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, 17(1), 2390454. Disponible en: <https://doi.org/10.1080/17538947.2024.2390454>.
- Piometto, M., G. García, V. Monayar, J.P. Carranza y H. Morales. 2019. Modernización de la valuación masiva de la tierra en la provincia de Córdoba a través de técnicas de aprendizaje computacional. *Revista de la Facultad de Ciencias Exactas, Físicas y Naturales*, 6(2). Disponible en: <https://ri.conicet.gov.ar/handle/11336/183406>.
- Piometto, M., V. Monayar, A. Sammartino y V. Fernández. 2022. Impacto territorial y fiscal de una valuación masiva inmobiliaria a precio de mercado (Córdoba, Argentina). *Economía, Sociedad y Territorio*, 22(70):767-800. Disponible en: <https://doi.org/10.22136/est20221757>.
- R Core Team. 2025. *_R: A Language and Environment for Statistical Computing_*. Viena: R Foundation for Statistical Computing. Disponible en: <https://www.R-project.org/>.
- Radics, A., F. Vázquez, N. Pérez Benítez e I. Ruelas. 2022. Panorama de las relaciones fiscales entre niveles de gobierno de países de América Latina y el Caribe. Washington, D.C.: BID. Disponible en: <http://dx.doi.org/10.18235/0004201>.
- Schapiro, R. 1990. The strength of weak learnability. *Machine Learning*, 5(2):197-227. Disponible en: <https://doi.org/10.1007/BF00116037>.
- Tennekes, M. 2018. tmap: Thematic Maps in R. *Journal of Statistical Software*, 84(6):1-39. Disponible en: <https://doi.org/10.18637/jss.v084.i06>.
- . 2025. *_tmaptools: Thematic Map Tools_*. R package version 3.2. s.l.: CRAN.R-project.org. Disponible en: <https://doi.org/10.32614/CRAN.package.tmaptools>.
- Wang, D. y V. Li. 2019. Mass Appraisal Models of Real Estate in the 21st Century: A Systematic Literature Review. *Sustainability*, 11(24):7006. Disponible en: <https://doi.org/10.3390/su11247006>.
- Worachairungreung, M., K. Thanakunwutthirot y S. Ninsawat. 2021. A Study on Estimating Land Value Distribution for the Talingchan District, Bangkok Using Points-of-Interest Data and Machine Learning Classification. *Applied Sciences*, 11(22):11029. Disponible en: <https://doi.org/10.3390/app112211029>.
- Yilmazer, S. y S. Kocaman. 2020. A Mass Appraisal Assessment Study Using Machine Learning Based on Multiple Regression and Random Forest. *Land Use Policy*, 99:104889. Disponible en: <https://doi.org/10.1016/j.landusepol.2020.104889>.
- Zhang, P., S. Hu, W. Li, C. Zhang, S. Yang y S. Qu. 2021. Modeling Fine-Scale Residential Land Price Distribution: An Experimental Study Using Open Data and Machine Learning. *Applied Geography*, 129:102442. Disponible en: <https://doi.org/10.1016/j.apgeog.2021.102442>.

