

Artículo completo

Ciencias Ambientales, Agroindustrias y de la Tierra

Aplicación de técnicas de *Machine Learning* para Valuación Masiva de Inmuebles en ciudades intermedias de la provincia de Mendoza

Machine Learning techniques application for Mass Appraisal in middle cities of Mendoza province

Polo, Renzo Enrique¹; Alberdi, Ramiro²; Erba, Diego Alfonso¹¹ Consultor independiente² Fac. de Ing. y Cs. Hídricas de la Univ. Nac. del Litoral

DOI: 10.59872/icu.v7i9.428

Correo de correspondencia: renzoepolo@gmail.com

Recepción: 16/06/2023; Aceptación: 07/12/2023;

Publicación: 19/04/2024

Palabras claves: Valuaciones

Keywords: Machine learning

Resumen

La necesidad de valorar masivamente los inmuebles se ha comprobado en el desarrollo de las políticas públicas en general. Los métodos tradicionalmente aplicados para la valuación masiva de inmuebles (VMI) han involucrado el uso de fórmulas e índices complejos, difíciles de obtener, procesar y mantener, elegidos en parte por la falta de conocimiento en tecnologías de automatización de procesos. Los procedimientos de VMI no se realizan con una frecuencia apropiada para acompañar la dinámica del mercado inmobiliario y por ese motivo nunca se llega a una valuación precisa de los inmuebles. En búsqueda de soluciones a dicha problemática, se desarrolló este trabajo siguiendo tres acciones: 1) generar una base de datos de entrenamiento capturando, sistematizando y procesando datos de oferta de inmuebles en el mercado inmobiliario de tres ciudades intermedias de la provincia de Mendoza; 2) caracterizar las bases de entrenamiento y predicción mediante variables geográficas; 3) asignar valores a todas las parcelas urbanas a través de técnicas de *Machine Learning*, más concretamente a través del algoritmo XGBoost para el modelado de valores. Aun cuando este recurso ha sido poco explorado en este ámbito de aplicación, los resultados y métricas obtenidas muestran que su utilización deriva en parámetros de calidad aceptables para los objetivos planteados, y que su implementación permite diseñar estrategias eficientes y eficaces para la construcción de VMI, a menor costo en dinero y en tiempo que los métodos tradicionales. La clave del éxito, en parte, tiene fuerte dependencia con la estrategia de recolección de datos.

Abstract

The need to massively valuation of land has been verified in the development of public policies in general. The methods traditionally applied for land mass appraisal (LMA) have involved the use of complex formulas and indices, difficult to obtain, process and maintain, chosen in part due to a lack of knowledge in process automation technologies. LMA procedures are not carried out with an appropriate frequency to follow the dynamics of the real estate market and for this reason an accurate valuation of the properties is never reached. In search of solutions to this problem, this work will be carried out following three actions: 1) generate a training database capturing, systematizing and processing real estate supply data in the real estate market of three middle size cities in the province of Mendoza; 2) characterize the training and prediction bases through geographic variables; 3) assign values to all urban plots through Machine Learning techniques, more specifically, the XGBoost algorithm for value modeling. Even though this resource has been little explored in this field of application, the results and metrics obtained show that its use leads to acceptable quality parameters for the proposed objectives, and that its implementation allows designing strategies for the construction of efficient and effective LMAs. at a lower cost in money and time than traditional methods. The key to success, in part, is strongly dependent on the data collection strategy.



Introducción

Las políticas públicas han manifestado sus intenciones de gravar más a los bienes inmuebles de mayor valor, tanto en la normativa que las materializa cuando en la verbalización de quienes las ejecutan. En general este principio no se discute, las dudas permanecen en el «cómo», es decir, con qué métodos se pueden llevar a la práctica. Inexorablemente el análisis recae sobre dos líneas: por un lado, la estrictamente tributaria que refiere al cálculo del impuesto de base territorial, y por otro, la determinación de la base imponible que, particularmente para los bienes inmuebles, está estrechamente vinculada a la información que registran y generan los catastros al aplicar diferentes métodos para valuar dichos inmuebles. Este trabajo se enfoca en el último aspecto por la filiación temática con el catastro, área con la que los autores tienen un vínculo estrecho, pero además por la importancia que en los últimos años ha tomado la actualización tecnológica y conceptual del proceso de la valuación masiva de inmuebles (VMI en adelante) en América Latina. Los métodos de VMI más modernos, apoyados en estadística e inteligencia artificial, se popularizan, se usan cada vez más y consecuentemente han tenido impacto en un gran número de políticas territoriales y tributarias dado que como resultado de su aplicación se pueden realizar evaluaciones de impactos económicos de grandes obras urbanas, definir estrategias de recupero de plusvalía, acelerar el desarrollo de planes de ordenamiento territorial, entre muchas otras.

En ese marco, se evidencia que la incorporación de tecnologías de información geográfica (TIG) abre para la Agrimensura oportunidades de proveer profesionales e investigadores capacitados para llevar adelante estos procesos de modernización.

El gran salto conceptual y técnico apuntado está dado por el cambio de paradigma que implica evolucionar de los sistemas tradicionales de valuación basados en tabulaciones, coeficientes, parámetros estáticos y mediciones en cartografía analógica, hacia los métodos modernos que se basan en la consideración de los valores de mercado de los inmuebles para construir un mapa continuo del territorio que muestre los matices económicos, apoyados en TIG y en la potencia de cálculo que ofrecen las técnicas estadísticas actuales. Este nuevo paradigma, en el contexto del libre mercado de inmuebles que funciona en la región con pocas restricciones, tiene la ventaja de poder registrar la riqueza real que representan las propiedades inmuebles, dado que, en términos ideales, se tiene en cuenta el valor al que el inmueble efectivamente se comercializa. Frente a los modelos tradicionales, este método incluye muchas variables que en general son difíciles de medir o representar pero que de una u otra manera están impresos en el valor final de transferencia (preferencias del comprador, urgencias del vendedor, expectativas, conocimiento territorial, etc.).

Trabajar con valores de mercado implica usar muestras compuestas por precios de oferta que en general corresponden a un número reducido del conjunto de inmuebles existentes. Por esta razón, los métodos de VMI basados en precios de mercado deben recurrir a técnicas estadísticas que permitan asignar un valor a cada uno de los inmuebles dentro del territorio, a través de casos testigo o tipo, y esto siempre conlleva un grado o nivel de precisión que es necesario conocer.

Actualmente, las técnicas estadísticas se han visto notablemente beneficiadas por el avance tecnológico en lo que a software libre y hardware accesible se refiere, reduciendo tiempos de cálculo, incorporando estrategias de aprendizaje autónomo o *machine learning*, en sus diferentes variantes. A su vez, la posibilidad que brindan las TIG de obtener representaciones tanto variadas como detalladas del territorio, en diferentes aspectos y a través de numerosas fuentes, hizo posible que las aplicaciones estadísticas adquieran más insumos y controles para la generación masiva de valores.

El talón de Aquiles de los métodos actuales suele ser, justamente, la captura de datos de mercado desde diferentes fuentes. Los requisitos que las muestras deben cumplir para que resulten adecuadas en la corrida de modelos informáticos no son pocas y, frecuentemente, los datos desperdiciados son inevitables. Más aún, considerando que una de las estrategias de obtención de información es la recorrida en campo, puede advertirse que los principales desafíos se ubican en esta etapa del proceso.

En ese contexto, este trabajo muestra los resultados obtenidos con la generación, aplicación y evaluación de un método de VMI de suelo urbano basada en valores de mercado, la cual pone de manifiesto todas las ventajas y desafíos antes mencionadas, junto a estrategias para potenciarlas y superarlas, respectivamente. La experiencia relatada en este artículo se llevó a cabo en tres localidades intermedias de la provincia de Mendoza durante el año 2022 enfocándose en lotes baldíos como inmueble tipo y los resultados presentados son alentadores para ampliar el alcance de este método a escalas mayores y más complejas.

Bases conceptuales

En América Latina y el Caribe la valuación catastral es mayoritariamente dissociativa, es decir, los valores del suelo y de las construcciones se determinan separadamente, para luego sumarlos y conformar el valor del inmueble. El valor del suelo en particular es generalmente agrupado, no necesariamente se determina «lote a lote», en sectores o zonas considerados homogéneos dentro de los cuales no necesariamente hay homogeneidad ni de uso del suelo, ni de distribución de redes de infraestructura, ni de normativa urbanística. Estos procesos subjetivos influyen negativamente sobre la equidad en la determinación del valor con relación a los practicados en el mercado, no obstante, es posible dar objetividad al método a través de la identificación de submercados mediante análisis de componentes principales y de agrupamientos (Hayles, 2006; Napoli *et al.*, 2017), técnicas de análisis *cluster* y análisis exploratorio de datos espaciales, análisis de cointegración, agrupamiento difuso, funciones de suavizado, entre otras (Royuela & Duque, 2013), las cuales pueden ser aplicadas con ventajas en ambiente SIG.

La Regresión Lineal Múltiple (RLM) es una técnica comúnmente utilizada para la VMI que permite determinar las relaciones lineales existentes entre una variable dependiente y variables independientes: la variable dependiente normalmente es el precio, mientras que las variables independientes pueden ser distancias a diferentes puntos de valorización, inclusión en zonas con altos índices de aprovechamiento edilicio o con afectaciones ambientales, entre otras. La RLM ha sido utilizada por varios autores como base para comparar su rendimiento frente a otras técnicas emergentes (Bencure *et al.*, 2019; Cohen *et al.*, 2020; Doumpos *et al.*, 2020; Wang *et al.*, 2020; Yilmazer & Kocaman, 2020), y si bien es de fácil comprensión, su aplicación para generar modelos de VMI para toda una ciudad no siempre se torna apropiada puesto que, eventualmente, no tiene la capacidad de identificar los elementos determinantes para la concretización de un negocio inmobiliario, generando frecuentemente problemas relacionados con autocorrelación espacial de los residuos y heterocedasticidad (Kauko & d’Amato, 2017).

La geoestadística puede ser utilizada para analizar el mercado inmobiliario a través de un conjunto de técnicas y permite predecir valores de inmuebles distribuidos a lo largo del espacio y/o del tiempo. En este ámbito los valores no se consideran independientes, sino que se correlacionan unos con otros, es decir, existe una dependencia espacial. Intuitivamente esto implica que mientras más cercanos estén situados dos inmuebles, más correlacionados están sus valores y mientras más separados, menos relación habrá entre ellos (Díaz Viera, 2002).

El Aprendizaje de Máquina (*Machine Learning*) es una rama de la IA que se aplica en la VMI, pues se encarga de estudiar métodos automatizados para generar modelos matemáticos, es decir, algoritmos y estadísticas que permitan resolver tareas de forma tal que el *software* descifra por sí mismo cómo resolverlo. En el área de VMI los modelos de aprendizaje de máquina suelen ser criticados por ser de difícil comprensión y asemejarse a los modelos de tipo «caja negra», especialmente cuando se comparan con los modelos econométricos de RLM donde el coeficiente de cada estimador es interpretado como la contribución marginal de la variable independiente al precio observado.

Desde otro ángulo puede afirmarse que los algoritmos de aprendizaje de máquina son estocásticos, por lo que su comportamiento puede variar de acuerdo al conjunto de datos. En ese sentido, las técnicas de remuestreo, como la validación cruzada o *leave-one-out*, sirven para evaluar la varianza de los resultados obtenidos, alejándose de la idea de «caja negra». Los métodos de conjunto o ensamblajes utilizan múltiples predicciones de estimadores para conseguir un mejor rendimiento predictivo que el que se podría obtener aisladamente (Polikar, 2006). En términos de varianza, los ensamblajes ayudan a reducirla y a generar modelos que generalizan el problema. Ejemplos de ellos son las técnicas de bosques aleatorios (*Random Forests* - Breiman, 2001), los cuales en regresión se refieren a múltiples árboles de decisión (Figura 1) contruidos independientemente, para luego calcular la estimación final como el promedio de las estimaciones para cada uno de ellos. Sumado a ello, existen las técnicas de *Boosting* (Freund & Schapire, 1996), que consisten en que cada árbol construido reduzca los residuos de la predicción del árbol anterior minimizando una función de pérdida. Las implementaciones XGBoost, CATBoost y LightGBM son optimizaciones de esta técnica¹.

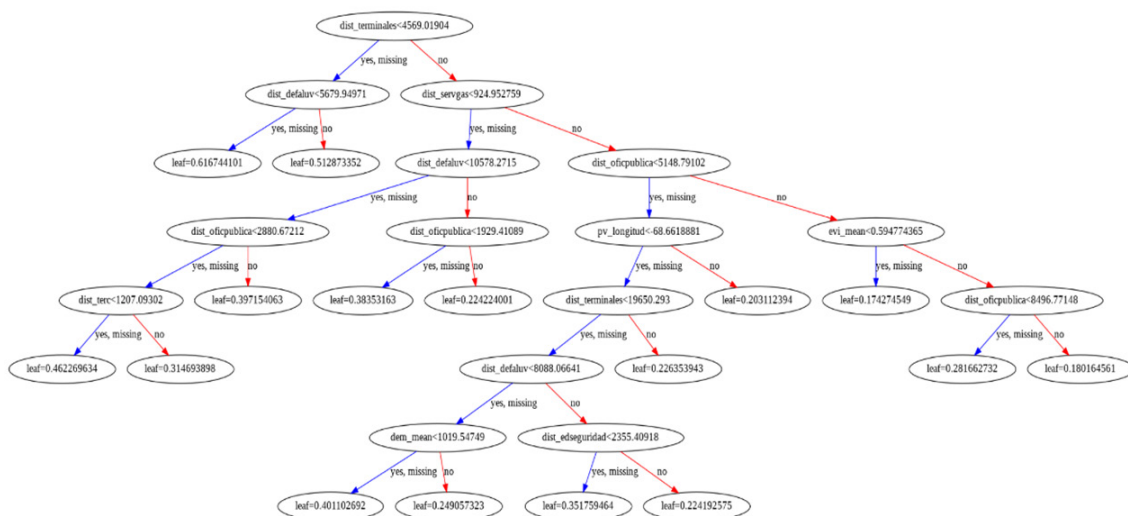


Figura 1. Ejemplo esquemático de árbol de decisión. Fuente: elaboración propia.

¹Más detalles sobre esas implementaciones pueden ser consultados en: XGBoost: Chen & Guestrin (2016), Brownlee (2016); CATBoost: Prokhorenkova et al. (2018) y LightGBM: Ke et al. (2017).

Una estrategia alternativa en la construcción de ensamblajes es la generación de meta-modelos, es decir, arreglos que sirvan para combinar las predicciones efectuadas por otros algoritmos de aprendizaje. Esta técnica llamada *Stacking* fue propuesta por Wolpert (1992), y consiste en la construcción de múltiples modelos de predicción para que sus resultados sean considerados como variables independientes en un modelo final, que las combinará para obtener un valor de síntesis.

Todo modelo de *Machine Learning* requiere configurar diferentes bases de datos para su implementación, las cuales pueden clasificarse como:

1. **Bases de entrenamiento** que contienen la nube de puntos con los datos depurados del Observatorio, complementados con las variables geográficas y ordenados por áreas de procesamiento según características territoriales. Estas bases de entrenamiento son el insumo esencial para la generación de modelos de predicción, y deben subdividirse en 3 partes:

- **Entrenamiento:** Observaciones utilizadas para encontrar relaciones entre la variable dependiente y las independientes.
- **Validación:** Observaciones utilizadas en el proceso de entrenamiento para calcular la función de pérdida a minimizar durante la optimización de hiperparámetros.
- **Testeo:** Observaciones que son utilizadas al finalizar el proceso de entrenamiento para evaluación de modelos y cálculo de métricas de desempeño.

2. **Bases de predicción**, constituidas por las unidades territoriales a las cuales se determina el valor del suelo calculado a través de los modelos entrenados ad hoc, y deben tener incorporadas las mismas variables geográficas que las bases de entrenamiento. Para el ámbito urbano generalmente el parcelario se utiliza como referencia para el cálculo del valor del suelo libre de mejoras.

Estado del arte

Todo proceso de revalúo tradicional inicia con el análisis de los datos catastrales correspondientes a valores de inmuebles y su relación con la política tributaria bajo la normativa vigente. En Argentina en general, y en la provincia de Mendoza en particular, las normas vigentes proponen valuar a los inmuebles y sus mejoras a través de modelos predeterminados que, si bien explícitamente aspiran a relacionar los valores catastrales con los precios practicados en el mercado («... Los valores unitarios de la tierra... serán asignados ... en función del estado del Mercado Inmobiliario local...»), la forma en que se componen evidencia que es alta la probabilidad de que no haya una correlación uniforme. Las fórmulas de cálculo del valor de todos los tipos de inmuebles contemplados en la norma mendocina son de fácil comprensión, no obstante, las variables que representan los valores del m² de terreno (VUT) y de construcción (VUC) derivan de la ubicación de la propiedad, de la jerarquía de la zona, de los servicios y, en definitiva, de un puntaje de categorización todo lo cuales no necesariamente atiende a las expectativas del mercado. De hecho, respecto a los métodos basados en coeficientes de forma, Monzani *et al.* (2020) comprobaron que pueden mostrar diferencias significativas comparados con las expectativas del mercado, a pesar de ser un método que se emplea desde hace muchas décadas.

En Argentina las experiencias desarrolladas con perspectivas del mercado no son numerosas y, si bien no se han implementado políticas centralizadas o coordinadas a escala nacional, en 2018 se promulgó el decreto PEN 938/2018, modificado por el PEN 450/22, que crea el Organismo Federal de Valuaciones de Inmuebles. En el art. 1. se establece que «tendrá por objeto determinar los procedimientos y las metodologías uniformes de aplicación para las valuaciones fiscales de los inmuebles ubicados en el territorio de la Ciudad Autónoma de Buenos Aires y en las provincias, con el objeto de lograr [...] que tiendan a reflejar la realidad del mercado inmobiliario² y la dinámica territorial [...]», lo cual brinda un marco e iniciativa para abordar la VMI bajo esas premisas. En ese contexto e incluso desde antes de dicha promulgación, algunas provincias y localidades han implementado observatorios a los fines de relevar los datos necesarios, pero pocas han avanzado en el procesamiento de ellos y muchas menos han aplicado métodos actuales de VMI. Al respecto, como ejemplo más avanzado de este tipo de políticas puede mencionarse al Gobierno de la Provincia de Córdoba que viene llevando adelante desde hace más de un lustro revalúos masivos a través de métodos de IA, como puede verse en sus reportes de publicación periódica³. En esa provincia generaron el mapa de valores del suelo a través de la implementación de un observatorio y de la aplicación de técnicas de machine learning que incluyeron algoritmos como *Random Forest* (RF) (Carranza *et al.*, 2018, 2019;), *Quantile Random Forest* (QRF) (Cerino *et al.*, 2021) y *Gradient Boosting Model* (GBM) (Piumetto *et al.*, 2019). De manera más incipiente y en la misma línea, en la provincia de Buenos Aires se ha puesto en marcha en 2021 el «Observatorio de valores del suelo e instrumentos de financiamiento del desarrollo urbano»⁴, el cual ya presenta sus primeros resultados en Dioguardi *et al.* (2022).

Con relación a lo anterior, si bien el término observatorio tiene amplios usos y sentidos, en este trabajo se los define como una estructura administrativa que monitorea el territorio a través de imágenes y/o censos, vinculada a instituciones catastrales, empresas públicas y/o privadas a través de alianzas con intereses comunes (Eguino & Erba, 2020). Concretamente, a lo que se refieren los objetivos de este trabajo es a la creación de un Observatorio del Mercado Inmobiliario de Mendoza (OMIM).

² El destacado en negrita es propio.

³ Ver, por ejemplo: Nuevos Valores Unitarios de la Tierra Urbana 2022 en Córdoba: <https://www.idecor.gob.ar/wp-content/uploads/2022/12/Informe-Valores-Tierra-Urbana-2022.pdf>

⁴ https://www.gba.gob.ar/habitat/noticias/se_present%C3%B3_el_%E2%80%99Observatorio_de_valores_del_suelo_e_instrumentos_de_financiamiento

Área de estudio

La provincia de Mendoza presenta un territorio heterogéneo donde las tierras aprovechables en los términos actuales se concentran en el 4,8% del total de superficie⁵ (los denominados oasis). Estas áreas concentran el 95% de la población con densidades máximas que llegan a los 300 habitantes/km² en el ámbito urbano. A su vez, la organización urbana de la provincia concentra el 62% de la población en el Gran Mendoza en primer lugar, y como segundo núcleo urbano se ubica la ciudad de San Rafael y alrededores en el oasis centro. Si bien resulta complejo definir en forma precisa el concepto de ciudades medias para la provincia de Mendoza, en este trabajo se analizaron las ciudades de San Rafael, Malargüe y Tunuyán (Figura 2), cuyo tamaño y población no se ubican en los extremos, a diferencia de Gran Mendoza u otras localidades menores. Esto último no implica que sean representativas de toda la realidad territorial, pero dan un buen panorama sobre las posibilidades de extender las metodologías a zonas semejantes (caso Tunuyán y Malargüe) y/o áreas urbanas mayores (caso San Rafael). En la tabla 1 se muestran algunos datos territoriales de referencia de las localidades.

Tabla 1. Datos básicos de las localidades estudiadas

| Localidad | Superficie del departamento ¹ (km ²) | Superficie total de inmuebles urbanos y suburbanos ² (km ²) | Cantidad de inmuebles urbanos ¹ | Población ³ |
|------------|---|--|--|------------------------|
| San Rafael | 31.435 | 56,0 | 54.092 | 210.468 |
| Tunuyán | 3.362 | 15,2 | 13.862 | 58.534 |
| Malargüe | 40.525 | 3,6 | 5.575 | 32.977 |

¹Según datos de la Dirección de Catastro de la Administración Tributaria de Mendoza (DCATM).

²Obtenida mediante sumatoria de superficies de parcelas registradas en la base de datos de la DCTAM.

³Según datos del Censo Nacional 2022.



Figura 2. Localidades estudiadas en el contexto provincial. Fuente: elaboración propia.

⁵Según datos del Departamento General de Irrigación de Mendoza.

Materiales y métodos

La estrategia metodológica se inspiró en experiencias similares como la de la provincia de Córdoba (Piumetto *et al.*, 2019), y consistió en organizar el proceso en tres etapas bien definidas y sucesivas:

1) **Observatorio**, con la responsabilidad de obtener, sistematizar y depurar los datos de mercado provenientes de diferentes fuentes.

Los recursos se generaron en forma secuencial atendiendo a la demanda de representatividad y diversidad geográfica en los entornos urbanos. De este modo, el primer relevamiento se hizo a través de técnicas de *web scraping*, mediante las cuales se «rasparon» sitios especializados como por ejemplo www.inmoclick.com, y se obtuvieron muestras de lotes baldíos en las localidades de estudio. Una vez reconocidas las áreas de vacancia se visitaron puntualmente otros sitios web y/o avisos en línea para completar la nube de puntos, entre los cuales se destaca www.properati.com.

En paralelo, se realizaron relevamientos en terreno con recorridas intensivas a las localidades para capturar avisos de particulares y/o de inmobiliarias que no tuvieran precio publicado en sitios web. El mecanismo consistió en capturar la posición del inmueble mediante fotografía de fachada con el cartel de venta visible (Figura 3) y posterior contacto con el comercializador para obtener los datos de interés para el trabajo.



Figura 3. Avisos e inmuebles relevados en el terreno. Fuente: elaboración propia.

Seguidamente se procedió con la sistematización y depuración de esos datos, fundamentalmente, a través del análisis estadístico, buscando identificar y descartar valores fuera de contexto (*outliers*) y evaluar la exactitud en la posición a través de la constatación de las características de cada inmueble con la base parcelaria de la DCATM. Una vez depuradas las nubes de puntos se constituyeron las bases de entrenamiento para cada localidad.

2) **Cartografía y SIG**: con la función de generar variables geográficas para la caracterización del territorio.

Los métodos automatizados de VMI se basan en la correspondencia geográfica de los valores de muestra y de población. En otras palabras, la generación de valores para todos los inmuebles a través de una muestra representativa requiere que se identifiquen parámetros correlacionados con las diferencias de valor. De este modo, fue confeccionada una amplia lista de variables geográficas que en conjunto potencialmente componen el valor del suelo e influyen en las variaciones entre muestras. Dichas variables se calcularon tanto para la base de entrenamiento como para la base de predicción tomando como base el parcelario urbano y suburbano. Las fuentes de información para esta tarea fueron muy diversas pero tomaron como base, por un lado, las capas SIG disponibles en geoservicios oficiales tales como la Infraestructura de Datos Espaciales de Mendoza⁶ (IDEM), la DCATM y el Sistema de Información Territorial Ambiental⁷ (SIAT) en combinación con la generación de algunas capas nuevas y extracción de fuentes libres (OpenStreetMap⁸, por ejemplo) y, por otro, servicios web de imágenes satelitales libres. Ambos conjuntos fueron recopilados, procesados y presentados combinando herramientas de uso libre como QGIS⁹, Google Earth Engine¹⁰ y PostGIS¹¹. Cabe destacar que toda la información se organizó y almacenó en servidores en la nube, en una base de datos PostgreSQL¹², administrada en combinación con QGIS y pgAdmin¹³.

⁶ <https://www.mendoza.gov.ar/idem/>

⁷ <http://siat.mendoza.gov.ar/>

⁸ <https://www.openstreetmap.org/>

⁹ <https://qgis.org/es/site/>

¹⁰ <https://earthengine.google.com/>

¹¹ <https://postgis.net/>

¹² <https://www.postgresql.org/>

¹³ <https://www.pgadmin.org/>

El resultado final de este procedimiento es una lista de variables que fue adosada a cada registro de la base de datos.

3) **Revalúo masivo:** con el cometido de aplicar algoritmos que, alimentados por la base de datos de entrenamiento, generaron los valores para la base de predicción de valores de mercado. De la amplia lista de algoritmos disponibles mencionados en apartados anteriores, en este trabajo se optó por el empleo del XGBoost, propuesto por Chen & Guestrin (2016), el cual permite evaluar, a través de diferentes iteraciones, la pertinencia de las variables geográficas en un gran número de modelos que se sintetizan en un óptimo resultante y arroja los siguientes parámetros:

- *Valor por m² o Valor Unitario del Suelo (VUT):* es el resultado directo de los modelos sobre las bases de predicción, considerando características de las parcelas.
- *Valor del suelo homogeneizado o Valor Unitario del Suelo Homogeneizado (VUTH):* presenta los resultados del cálculo del VUT considerando un lote tipo de tamaño y forma regular para todo el territorio. Funciona como indicador del valor independiente de las características de las parcelas.
- *Valor total de la parcela:* es el producto del cálculo entre el VUT por el área total de la parcela. Es el valor completo del terreno libre de mejoras.

Por su carácter incipiente, en este trabajo se decidió emplear el VUT como elemento representante de la aplicación de los modelos.

Los resultados de la modelación se evaluaron a través del Error Porcentual Medio Absoluto (*Mean Absolute Percent Error - MAPE*), que mide el error expresado en porcentaje a través del promedio de los errores absolutos.

Esta etapa fue desarrollada íntegramente en lenguaje Python¹⁴, utilizando Pandas¹⁵ y Scikit-Learn¹⁶ como biblioteca de manipulación y análisis de datos; Hyperopt¹⁷ para optimización de hiperparámetros en árboles de regresión; y la biblioteca XGBoost¹⁸ como implementación de este algoritmo.

Resultados

Los resultados de la etapa 1 se resumen en la tabla 2, donde se detalla la cantidad de observaciones sistematizadas en total y las que efectivamente fueron empleadas luego de la depuración; y en las figuras 4, 5 y 6, en las que se muestran las nubes de puntos resultantes por localidad, incluyendo puntos totales relevados y puntos descartados por diferentes factores.

Tabla 2. Total de muestras sistematizadas y procesadas en el proceso de revalúo.

| Localidad | Scraping automatizado | Otros sitios web | Relevamiento en campo | Total observaciones | Total depuradas |
|------------|-----------------------|------------------|-----------------------|---------------------|-----------------|
| San Rafael | 31 | 9 | 67 | 107 | 70 |
| Tunuyán | 30 | 21 | 0 | 51 | 23 |
| Malargüe | 4 | 4 | 15 | 23 | 12 |
| Totales | 65 | 34 | 82 | 181 | 105 |

¹⁴ <https://www.python.org/>

¹⁵ <https://pandas.pydata.org/>

¹⁶ <https://scikit-learn.org/stable/>

¹⁷ <http://hyperopt.github.io/hyperopt/>

¹⁸ <https://xgboost.readthedocs.io/en/stable/>

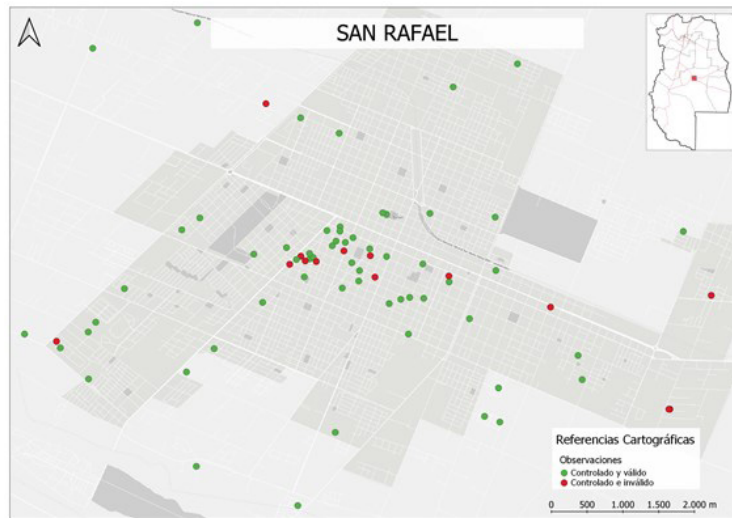


Figura 4. Nube de puntos resultantes del observatorio para San Rafael. Fuente: elaboración propia.

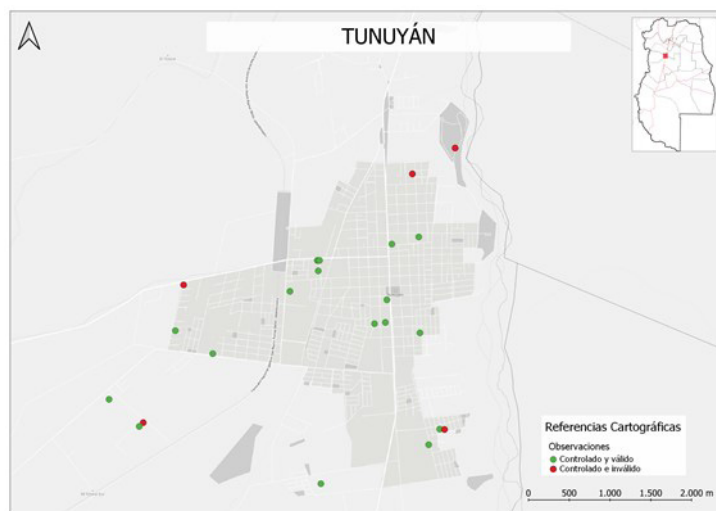


Figura 5. Nubes de puntos resultantes para Tunuyán. Fuente: Elaboración propia.

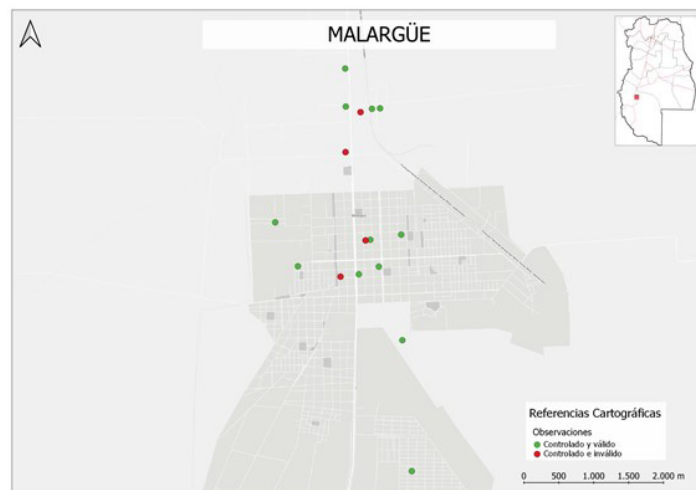


Figura 6. Nubes de puntos resultantes para Malargüe. Fuente: elaboración propia.

En la etapa 2 se generaron en total 79 variables geográficas organizadas en 3 categorías:

- *Catastrales*: derivadas de atributos propios de las parcelas tales como superficie construida, tamaño, etc.
- *De localización*: representan distancias a sitios o zonas de interés (hospitales, escuelas, etc.), su densidad (cantidad de paradas de bus en un radio determinado) o indican pertenencia a áreas con características particulares (áreas de alto valor, zonas homogéneas, etc.).
- *Satelitales*: obtenidas a través de la generación de diferentes índices espectrales o estadísticos calculados con productos satelitales (imágenes y modelos digitales de terreno).

Los procedimientos empleados se resumen en la Figura 7, y fueron aplicados tanto para la base de entrenamiento como para la de predicción, completando los conjuntos de datos de la etapa anterior para ejecución de la tercera etapa.

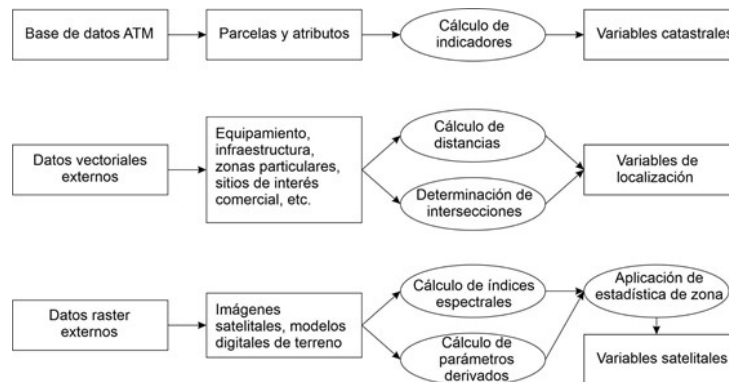


Figura 7. Flujo de trabajo para la generación de variables geográficas. Fuente: elaboración propia.

La etapa 3 consistió en la aplicación del algoritmo XGBoost. La implementación radicó en la variación de los datos de entrenamiento y validación a partir del cambio de semilla de inicialización. Los hiper parámetros de los árboles se obtuvieron a través de una búsqueda bayesiana (Bergstra *et al.*, 2013), previa configuración del espacio de búsqueda de estos parámetros y minimización de la función de pérdida. Este procedimiento se reiteró obteniendo 25 predicciones que luego fueron ensambladas con un meta-modelo del cual se obtuvo la predicción final.

La evaluación de calidad realizada mediante la obtención del índice MAPE arrojó un valor de 29%, que bajo estas condiciones de aplicación resulta aceptable y acorde a valores obtenidos en trabajos similares (Reyes-Bueno *et al.*, 2018; Carranza *et al.*, 2022; Oliveira, 2020; Oliveira *et al.*, 2022), aunque todavía por encima del rango establecido por la IAAO (2018) y procesos con varios años de experiencia que alcanzan actualmente cifras cercanas al 10% (IDECOR, 2022), aunque con diferencias metodológicas de obtención del MAPE. Posteriormente, se procedió a la asignación de valores de VUT sobre la base de datos de predicción, logrando los mapas que se presentan a continuación en las figuras 8 a 10.

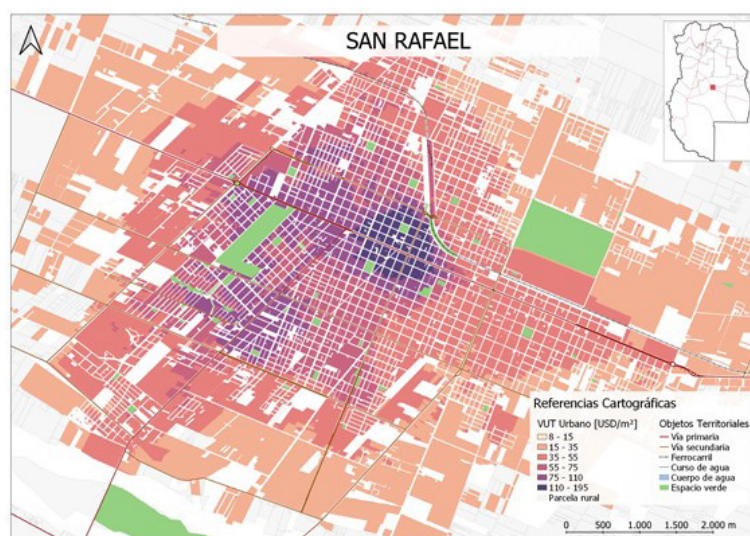


Figura 8. Mapa de valores del suelo urbano de la localidad de San Rafael. Fuente: elaboración propia.

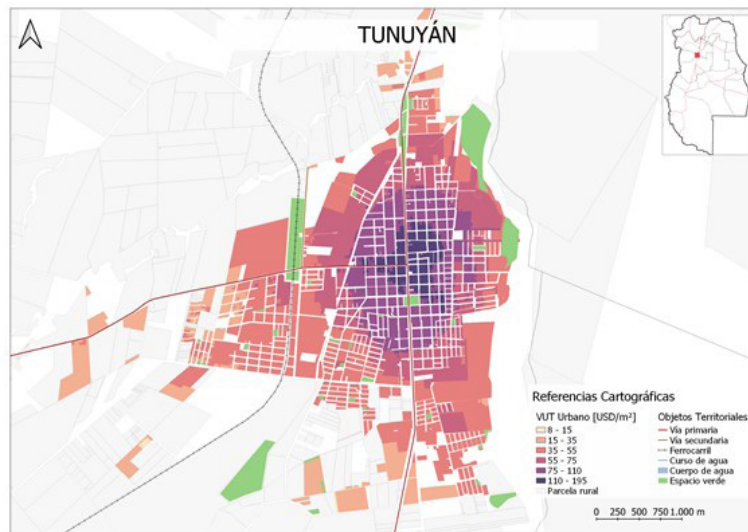


Figura 9. Mapa de valores del suelo urbano de la localidad de Tunuyán. Fuente: elaboración propia.

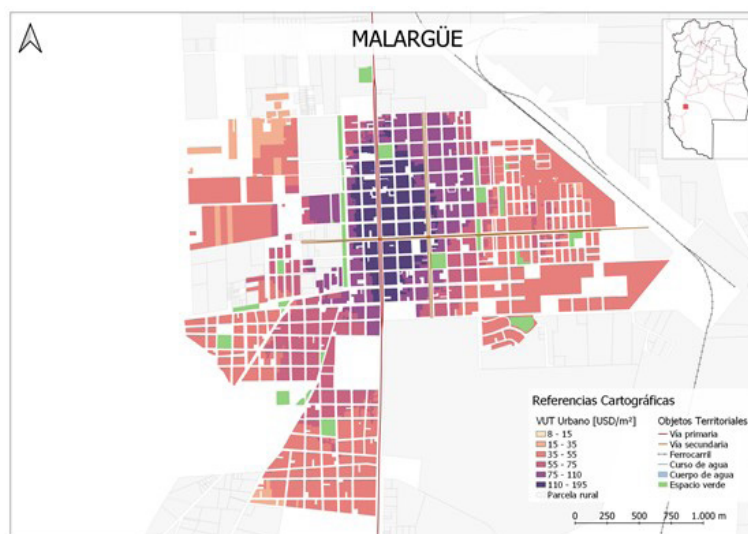


Figura 10. Mapa de valores del suelo urbano de la localidad de Malargüe. Fuente: elaboración propia.

Discusiones

Los métodos tradicionales aplicados para determinar valores catastrales en general, y el referido en la norma de Avalúos de Mendoza en particular, se basan en la obtención de valores a través del procesamiento de variables territoriales y parcelarias. La complejidad que pueden alcanzar estos modelos es tan amplia como detallado se requiera que sea el método, sin embargo, no existe alguno que concentre todos los elementos que intervienen en la formación de valores, por la sencilla razón de que hay parámetros «escondidos» o no determinables (gustos, preferencias, urgencias, entre otros). No obstante, al tratarse de una cantidad limitada y bien conocida de variables, los métodos pueden aplicarse sin problemas o limitaciones a todo un territorio, aunque los resultados finales pueden no estar relacionados con la riqueza real existente ni con el impacto efectivo de las obras públicas, permisos de edificación, etc. Por el contrario, en los valores de mercado que se pueden relevar en forma directa sí se incluyen esos factores escondidos, aunque sin conocer cuál es su influencia y/o especificidad, razón por la cual deben emplearse métodos estadísticos para arribar al mapa completo a partir de muestras. Si bien estos mecanismos no están exentos de errores, la precisión es cuantificable, puede determinarse y mitigarse por arreglos informáticos o bien reducirse mejorando la calidad del proceso (introduciendo más datos, por ejemplo). En resumen, la posibilidad de determinar la riqueza real del territorio es un horizonte al que se pueden dirigir las políticas de VMI, tomando los mejores aspectos de los métodos tradicionales, pero sin perder de vista el funcionamiento real del mercado. Los resultados presentados en este trabajo para San Rafael, Tunuyán y Malargüe muestran que la refinación de los métodos a través de la densificación de datos,

ampliación y/o jerarquización de variables geográficas, junto a la actualización o testeo de nuevos parámetros de calibración del modelo XGBoost, permiten proponer nuevos métodos de VMI a escala provincial, a relativo bajo coste y menores tiempos de ejecución, pero con mayor fidelidad y captación de los efectos de políticas territoriales que generan riqueza.

Conclusiones

Referir los valores al mercado inmobiliario está en el espíritu de la normativa catastral en gran parte de las jurisdicciones latinoamericanas, no obstante, esto raramente acontece en la práctica.

Existe vasta literatura sobre valuación masiva de inmuebles urbanos, desde el enfoque de precios hedónicos con modelos de regresión lineal múltiple, regresión espacial, con y sin el uso de técnicas geoestadísticas, hasta la aplicación de técnicas de machine learning; lo que lo que falta es difundir estos métodos y técnicas en Argentina y poner en práctica particularmente el algoritmo XGBoost además del Random Forest utilizado en los revalúos masivos desarrollados en la provincia de Córdoba (Carranza, J.P.; 2018). Tal como lo indica Muzaffer Can Iban (2022), si bien el rendimiento predictivo de los algoritmos basados en árboles puede superar al del método de mínimos cuadrados clásico y al XGBoost, el regresor que utiliza este último puede ser ligeramente mejor que los métodos competitivos basados en árboles

Este artículo describe un flujo de trabajo aplicado en localidades de la provincia de Mendoza, cuyo fin fue la obtención del mapa de valores de mercado a escala de parcela en zonas urbanas. Más allá de la consecución de dicho mapa, se ha demostrado que es posible aplicar métodos rápidos y económicos para referir la riqueza inmobiliaria a valores reales del mercado. No obstante, es importante destacar que la robustez del método radica fuertemente en las muestras reales del observatorio, de modo que su densidad debe establecerse como un parámetro de calidad al inicio del proceso. Aun así, los resultados estadísticos son promisorios para una implementación a mayor escala y sostenida en el tiempo.

El área de valuaciones masivas está en la pauta de los administradores públicos, se abre para la Agrimensura un nuevo espacio de trabajo que requerirá, no obstante, actualizaciones por parte de los profesionales actuantes y renovación de los programas de las disciplinas afines.

Bibliografía

- Bencure, J. C., Tripathi, N. K., Miyazaki, H., Ninsawat, S., & Kim, S. M. (2019).** Development of an Innovative Land Valuation Model (iLVM) for Mass Appraisal Application in Sub-Urban Areas Using AHP: An Integration of Theoretical and Practical Approaches. En *Sustainability* (Vol. 11, Número 13). <https://doi.org/10.3390/su11133731>.
- Bergstra, J., Yamins, D. & Cox, D. (2013).** «Hyperopt: A Python library for optimizing the hyperparameters of Machine Learning algorithms». Proceedings of the 12th Python In Science Conference.
- Breiman, L. (2001).** «Random Forests». *Machine Learning*, 45. pp. 5-32. Elsevier. <https://doi.org/10.1023/A:1010933404324>.
- Brownlee, J. (2016).** A gentle introduction to XGBoost for applied Machine Learning. *Machine Learning Mastery*, consultado por última vez el 29/05/2023: <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>.
- Carranza, J. P., Salomón, M. J., Piumetto, M. A., Monzani, F., Montenegro, M. G. & Córdoba, M. A. (2018).** Random Forest como técnica de valuación masiva del valor del suelo urbano: una aplicación para la ciudad de Río Cuarto, Córdoba, Argentina. Congreso de Cadastro Multifinalitário e Gestao Territorial, COBRAC. Florianópolis.
- Carranza, J. P., Piumetto, M. A., Salomón, M. J., Monzani, F., Montenegro, M. G. & Córdoba, M. A. (2019).** Valuación masiva de la tierra urbana mediante inteligencia artificial. EL caso de la ciudad de San Francisco, Córdoba, Argentina. *Revista Vivienda & Ciudad*, Vol. 6, ISSN 2422-670X.
- Carranza, J. P.; Piumetto, M. A.; Lucca, C. M. & Da Silva, E. (2022).** «Mass appraisal as affordable public policy: Open data and machine learning for mapping urban land values». *Land Use Policy*, Vol. 119. Elsevier.
- Cerino, R. M., Carranza, J. P., Piumetto, M. A., Bullano, M. E., Caffaratti Donalisio, V. & Monzani, F. (2021).** Propuesta para la valuación masiva del suelo urbano. Aplicación espacial del algoritmo Quantile Regression Forest. *Revista Vivienda y Ciudad*, Vol. 8, ISSN 2422-670X.
- Chen, T. & Guestrin, C. (2016).** «XGBoost: A scalable tree boosting system». KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp 785-794.
- Cohen, J. P., Coughlin, C. C., & Zabel, J. (2020).** Time-Geographically Weighted Regressions and Residential Property Value Assessment. *The Journal of Real Estate Finance and Economics*, 60(1), 134-154. <https://doi.org/10.1007/s11146-019-09718-8>.
- Díaz Viera, M. A. (2002).** *Geoestadística Aplicada*.
- Dioguardi, F., Torres, D., Antonelli, R. L. & del Río, J. P. (2022).** Construcción de un grafo de conocimiento para un observatorio inmobiliario. Libro de actas del XXVIII Congreso Argentino de Ciencias de la Computación - CACIC 2022.
- Doumpos, M., Papastamos, D., Andritsos, D., & Zopounidis, C. (2020).** Developing automated valuation models for estimating property values: a comparison of global and locally weighted approaches. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-020-03556-1>.
- Eguino, H. & Erba, D.; (Eds.) (2020).** Catastro, valoración inmobiliaria y tributación municipal: Experiencias para mejorar su articulación y efectividad. Banco Interamericano de Desarrollo. <https://doi.org/http://dx.doi.org/10.18235/0002437>.
- Freund, Y. & Schapire, R. E. (1996).** «Experiments with a new boosting algorithm». En Saitta (Ed.): *Machine Learning*, Proceedings of the 13th International Conference. pp. 148-156. Morgan Kaufmann.
- Hayles, K. (2006).** A Property Valuation Model for Rural Victoria [School of Mathematical and Geospatial Science. RMIT University]. <http://researchbank.rmit.edu.au/eserv/rmit:6265/Hayles.pdf>
- IAAO (2018).** Standard on Automated Valuation Models (AVMs). International Association of Assessing Officers. Disponible en: https://www.iaao.org/media/standards/Standard_on_Automated_Valuation_Models.pdf
- IDECOR (2022).** Informe: Estudio del Mercado de Suelo Urbano de la provincia de Córdoba. Disponible en: <https://www.idecor.gob.ar/wp-content/uploads/2022/12/Informe-Valores-Tierra-Urbana-2022.pdf>
- Kauko, T. J., & d'Amato, M. (Eds.). (2017).** *Advances in Automated Valuation Modeling*. En *AVM after the non-agency mortgage crisis*. Springer. <https://doi.org/10.1007/978-3-319-49746-4>.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. & Liu, T. (2017).** «LightGBM: A Highly Efficient Gradient Boosting Decision Tree». 31st Conference on Neural Information Processing Systems (NIPS).
- Monzani, F., Carranza, J. P., Piumetto, M. A., Salomón, M. J., Polo, R., ECórdoba, M. A. & Monayar, M. V. (2020).** Análisis estadístico del impacto de la superficie, el frente y el fondo sobre el valor unitario de la tierra urbana. *Revista de la Facultad de Ciencias Exactas, Físicas y Naturales*, Vol. 7 Nº 1.
- Napoli, G., Giuffrida, S., Trovato, M. R., & Valenti, A. (2017).** Cap Rate as the Interpretative Variable of the Urban Real Estate Capital Asset: A Comparison of Different Sub-Market Definitions in Palermo, Italy. *Buildings*, 7(3).
- Oliveira, A. A. F. de. (2020).** «Avaliação em massa com modelos de aprendizado de máquina aplicados aos terrenos urbanos do Município de Fortaleza». Dissertação (Mestrado em Economia do Setor Público). Universidade Federal do Ceará.
- Oliveira, A. A. F. de; Bandeira, S. R. V.; Silva, C. V. A. (2022).** «Estimativa de desempenho de métodos de aprendizado de máquina baseados em árvores de decisão na valoração do solo no município de Fortaleza, Brasil». *Revista Valorem*, [S. l.], v. 1, n. 1, p. 1-15, 2022.
- Polikar, R. (2006).** The Wavelet Tutorial. Rowan University.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. & Gulin, A. (2018).** «CatBoost: unbiased boosting with categorical features». *Advances in Neural Information Processing Systems* 31.
- Reyes-Bueno, F., García-Samaniego, J. M. & Sánchez-Rodríguez, A. (2018).** «Large-scale simultaneous market segment definition and mass appraisal using decision tree learning for fiscal purposes». *Land Use Policy*, Vol. 79, pp. 116-122. Elsevier.
- Royuela, V., & Duque, J. C. (2013).** HouSI: Heuristic for delimitation of housing submarkets and price homogeneous areas. *Computers, Environment and Urban Systems*, 37(1), 59-69.

Wang, D., Li, V. J., & Yu, H. (2020). Mass Appraisal Modeling of Real Estate in Urban Centers by Geographically and Temporally Weighted Regression: A Case Study of Beijing's Core Area. *Land*, 9(5), 143. <https://doi.org/10.3390/land9050143>.

Wolpert, D. H. (1992). «Stacked generalization». *Neural Networks*, 5 (2). pp. 241-259. Elsevier. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).

Yilmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy*, 99, 104889. <https://doi.org/https://doi.org/10.1016/j.landusepol.2020.104889>.