



Una revisión sistemática de la literatura sobre la precisión de modelos de aprendizaje automático aplicados a la tasación de bienes raíces

A systematic literature review on the accuracy of machine learning models applied to real estate appraisal

Fabián Espinoza Garza

Universidad Autónoma Indígena de México, Los Mochis, México

fespinozagarza@gmail.com

ORCID: 0009-0005-7145-9542

Yobani Martínez Ramírez

Universidad Autónoma de Sinaloa, Los Mochis, México

yobani@uas.edu.mx

ORCID: 0000-0002-4967-9187

Alan Ramírez-Noriega

Universidad Autónoma de Sinaloa, Los Mochis, México

alandramireznoriega@uas.edu.mx

ORCID: 0000-0002-8634-9988

Iván Noel Álvarez Sánchez

Universidad Autónoma Indígena de México, Los Mochis, México

ivanalvarezsanchez@uais.edu.mx

ORCID: 0000-0002-8819-5310

doi: <https://doi.org/10.36825/RITI.12.28.002>

Recibido: Junio 14, 2024

Aceptado: Agosto 25, 2024

Resumen: El objetivo del presente documento es identificar los modelos de Aprendizaje Automático (AA) más precisos para predecir el valor de una propiedad inmobiliaria, basándose en una Revisión Sistemática de la Literatura (RSL). La cual se realizó de investigaciones publicadas entre 2022 y 2023 que analizaron la precisión de modelos de AA en la valuación inmobiliaria. Se extrajo información sobre los modelos de AA utilizados, las bases de datos empleadas y los modelos destacados por su precisión. Por último, se identificó una variedad de modelos de AA utilizados en la valuación inmobiliaria, incluyendo *Random Forest* (RF), *XGBoost*, *Gradient Boosting Machine* (GBM), Regresión Lineal (LR) y Regresión Lasso. Los modelos RF y LR se destacaron como los más precisos en las investigaciones analizadas, encontrando que la precisión de los modelos de AA varía según la base de datos, las características de las propiedades y el contexto de la valuación. Concluyendo que los modelos de AA, como RF y LR, son herramientas prometedoras para mejorar la precisión de la valuación inmobiliaria. Además, la elección del mejor modelo de AA depende de factores como la base de datos, las características de las

propiedades y los objetivos específicos de la valuación. Es necesario continuar investigando la aplicación de AA en la valuación inmobiliaria, considerando aspectos como la interpretabilidad de los modelos y el contexto del mercado inmobiliario.

Palabras clave: *Valuación Inmobiliaria, Aprendizaje Automático, Precisión, Modelo.*

Abstract: The objective of this document is to identify the most accurate machine learning (ML) models for predicting the value of a real estate property, based on a systematic literature review (SLR). It was conducted on research published between 2022 and 2023 that analyzed the accuracy of ML models in real estate valuation. Information was extracted regarding the ML models used, the databases employed, and the models highlighted for their accuracy. A variety of ML models used in real estate valuation were identified, including Random Forest (RF), XGBoost, Gradient Boosting Machine (GBM), Linear Regression (LR), and Lasso Regression. RF and LR models stood out as the most accurate in the analyzed research, showing that the accuracy of ML models varies depending on the database, property characteristics, and valuation context. It was concluded that ML models like RF and LR are promising tools for improving the accuracy of real estate valuation. Additionally, the choice of the best ML model depends on factors such as the database, property characteristics, and specific valuation objectives. Further research is needed on the application of ML in real estate valuation, considering aspects such as model interpretability and the real estate market context.

Keywords: *Real Estate Valuation, Machine Learning, Accuracy, Model.*

1. Introducción

La precisión de las valuaciones ha sido un tema de debate tanto popular como académico durante varias décadas [1]. En México existen normas como la NMX-R-081-SCFI-2015 que establece las reglas mediante las cuales se define detalladamente los requisitos generales que deben observarse para la prestación del servicio de valuación [2]. Sin embargo, la tecnología actual y la información disponible expone nuevas oportunidades para explorar metodologías alternativas que podrían mejorar la eficiencia y la precisión. Se ha explorado la viabilidad de utilizar aprendizaje automático (ML, por sus siglas en inglés *Machine Learning*), esta tecnología podría ser útil en la valuación de bienes raíces, sosteniendo que con la llegada de modelos más avanzados y el crecimiento de datos disponibles se logrará desarrollar un sistema de valuación automatizado [1]. Si bien, deberá ser probado exhaustivamente, representa un nuevo reto a considerar en el área de la valuación inmobiliaria en México.

Mediante la inteligencia artificial se logra analizar e interpretar datos de forma similar a los seres humanos [3]. El aprendizaje automático (AA) es una rama de la inteligencia artificial que se centra en el desarrollo de sistemas capaces de aprender y mejorar a partir de la experiencia, sin estar explícitamente programados para ello [4]. En esencia, el AA utiliza algoritmos para aprender de los datos y tomar decisiones o realizar predicciones.

Estos algoritmos se pueden clasificar en tres categorías principales: 1) Aprendizaje supervisado: empleado cuando se cuenta con un conjunto de datos con características bien identificadas; 2) Aprendizaje no supervisado: empleado cuando se cuenta con datos sin identificar, este modelo identifica patrones y relaciones entre estos datos; 3) Por último, Aprendizaje por refuerzo: se emplea una retroalimentación al modelo para guiar el aprendizaje [5].

La base fundamental para la aplicación de los algoritmos de AA es la información, por lo que contar con bases de datos con información relevante y suficiente determina la eficiencia que pueden alcanzar. En México, se encuentra disponible en el Sistema Nacional de Información e Indicadores de Vivienda una base de datos en la cual se registran los datos históricos de financiamientos a la vivienda, esta contiene información desde el 2013 y se encuentra actualizada con datos hasta septiembre de 2023 [6]. Por lo que, se considera que existe la oportunidad de emplear estas bases de datos en el AA y así, estimar el valor de las propiedades en todo el territorio mexicano.

La revisión sistemática de literatura (RSL, del inglés *Review Systematic Literature*) es la base para conocer el panorama general, esta ayuda a comprender de forma general los temas a investigar, evitando sesgos y conformando los límites desde la búsqueda y revisión. De esta manera, se intenta identificar información relevante con representación similar. Además, permite evaluar de manera crítica los resultados obtenidos, lo que contribuye significativamente al desarrollo y mejora de futuras investigaciones. Por lo que, la RSL no solo proporciona una plataforma sólida de conocimientos existentes, también conforma una base sólida para el desarrollo de nuevas investigaciones [7].

El objetivo de este trabajo es conocer cuáles modelos de AA son más precisos para estimar el valor de un bien inmueble, mediante el análisis de diversos trabajos empleando la RSL. El presente documento se encuentra estructurado en siete apartados: el primero presenta la introducción; el segundo, aborda la metodología aplicada; el tercero, presenta los resultados obtenidos; el cuarto, discute los resultados encontrados; el quinto, expone las conclusiones; el sexto, describe el trabajo futuro, y finalmente, en el séptimo apartado se muestran las referencias en las que se apoya la presente investigación.

2. Metodología

En el presente realiza una RSL y se lleva a cabo tomando en cuenta un enfoque mixto, basándose en un análisis de la precisión de diferentes modelos de AA, así como, los modelos considerados, el propósito es analizar los modelos más viables y precisos al aplicarse en la valuación inmobiliaria. Este trabajo se basa en la metodología propuesta por [7]. Los criterios de inclusión y exclusión se implementaron como se muestra en la Tabla 1.

Tabla 1. Criterios de inclusión y exclusión.

Criterio	Inclusión	Exclusión
1	Publicación del 2022 al 2024.	Estudio duplicado.
2	Analiza o compara la precisión de diferentes modelos de AA.	Describe o realiza revisiones de modelos de AA teóricamente
3	Documentos en español e inglés.	Otros idiomas.
4	Publicaciones de conferencias y artículos de revistas de investigaciones académica.	Publicaciones no académicas
5	Bases de datos con información de las características de las propiedades y sus precios.	Bases de datos con información espaciales (coordenadas), imágenes, información 3d.

Fuente: Elaboración propia.

La búsqueda se realizó en las bases de datos: *IEEE Xplore*, *Mendeley*, *ScienceDirect* y *Scopus*, empleando las palabras clave en los siguientes idiomas: (a) Inglés: *real AND estate AND (valuation OR appraisal) AND machine AND learning AND model AND accuracy*; (b) Español: *precisión AND aprendizaje AND automático AND (valuación OR tasación) AND inmobiliaria*. Así también, se consideraron los resultados del 01 enero del 2022 hasta el 09 de febrero del 2024.

Primero, se localizaron 86 documentos en las diferentes bases de datos aplicando los criterios antes mencionados: *IEEE Xplore* (18), *Mendeley* (14), *ScienceDirect* (23) y *Scopus* (31). Posteriormente, considerando el título y el resumen de cada documento, se seleccionaron 31 artículos. Luego, se realizó una revisión más minuciosa basado en su contenido, para ello se procedió a realizar un análisis detallado teniendo en cuenta los criterios previamente establecidos en la Tabla 1. De este conjunto de 31 artículos, se seleccionaron 9 para ser incluidos en el presente documento por estar altamente relacionado con el enfoque de la RSL.

En la Tabla 2, se enlistan las abreviaturas que se emplearán más adelante para hacer referencia a cada modelo de AA y otras utilizadas en el documento.

Tabla 2. Abreviaturas.

RF	<i>Random Forest</i> (Bosques Aleatorios)	DT	<i>Decision Trees</i> (Arboles de decisión)	SVM	<i>Support Vector Machine</i> (Máquina de Soporte de Vectores)
LR	<i>Linear Regression</i> (Regresión lineal)	VR	<i>Voting Regressor</i> (Regresor de Votación)	DNN	<i>Deep Neural Networks</i>

				(Red Neuronal Profunda)
NN	<i>Neural Network</i> (Red Neuronal)	Ridge	<i>Ridge Regression</i> (Regression Ridge)	Boosting Modelo de Conjuntos (Impulsado)
MLR	<i>Multiple Linear Regression</i> (Regresión Lineal Multiple)	Lasso	<i>Lasso Regression</i> (Regresión Lasso)	GBM/Gradient Boosting Máquina de Aumento de Gradiente
SVR	<i>Support Vector Regressor</i>	PR	<i>Regression Polynomial</i> (Regresión polinomial)	XGBoost Aumento de Gradiente Extremo
AVM	<i>Automated Valuation Models</i> (Modelos de Valoración Automatizados)	IML	<i>Interpretable Machine Learning</i> (Aprendizaje Automático Interpretable)	ML/AA <i>Machine Learning</i> (Aprendizaje Automático)

Fuente: Elaboración propia.

En la Tabla 3 se muestran los 9 artículos seleccionados para ser incluidos en esta investigación. El análisis realizado indica los datos extraídos de cada uno de los trabajos: año, título, país, base de datos, modelos analizados y modelo destacado.

Tabla 3. Modelo destacado y bases de datos consideradas por artículo.

Año y Autor	Título	País	BD	Modelo destacado
2023 Chiasson, E., et al. [8]	<i>REALM: Automating Real Estate Appraisal with Machine Learning Models</i>	Estados Unidos	3000 casas vendidas en Ames, Iowa, entre 2006 y 2010.	RF
2023 Deppner, J., et al. [9]	<i>Boosting the Accuracy of Commercial Real Estate Appraisals: An Interpretable Machine Learning Approach</i>	Estados Unidos	Propiedades NCREIF (NPI) entre 1997 y 2021	Boosting
2023 Mody, P., et al. [10]	<i>Enhancing Real Estate Market Insights through Machine Learning: Predicting Property Prices with Advanced Data Analytics</i>	India	Scraping de diversas fuentes WEB	RF
2023 Nazarov, M, y Yarmatov, S. [11]	<i>Optimization of Prediction Results Based on Ensemble Methods of Machine Learning</i>	Uzbekistán	16.000 registros de propiedades Samarcanda	VR
2023 Stang, M., et al. [12]	<i>From human business to machine learning—methods for automating real estate appraisals and their practical implications</i>	Alemania	1,2 millones de propiedades residenciales en Alemania entre 2014 y 2020	XGBoost

<i>The Comparison Study of Regression Models (Multiple Linear Regression, Ridge, Lasso, Random Forest, and Polynomial Regression) for House Price Prediction in West Nusa Tenggara</i>				
2023 Putri, M. R., et al. [13]		Indonesia	web scraping del sitio web Lamundi, Nusa Tenggara	LR/Lasso /RF
<i>Model agnostic interpretable machine learning for residential property valuation</i>				
2023 Gunes, T. [14]		Turquía	Ankara	RF
<i>Does machine learning prediction dampen the information asymmetry for non-local investors</i>				
2022 Jung, J., et al. [15]		Estados Unidos	19,640 transacciones oficinas en áreas metropolitanas de EU	RF/GBM
<i>Real Estate Price Prediction using Supervised Learning</i>				
2022 Matey, V., et al. [16]		India	Sitios web, incluidos kaggle, magicbricks, 99acres y ready rates	LR

Fuente: Elaboración propia.

Así también, se puede apreciar en la Tabla 3 el interés por aplicar AA para analizar una variedad de fuentes de datos, incluyendo bases de datos públicas, transacciones de rentas y sitios web de promoción de bienes raíces. Además, estos análisis se llevan a cabo en diversas regiones del mundo, lo que destaca tanto la relevancia como el interés en examinar esta vasta fuente de información digital disponible públicamente. Así mismo, se observa una búsqueda constante de nuevas metodologías que podrían beneficiar significativamente la valuación inmobiliaria en el futuro.

En la Figura 1 se identifican gráficamente los países en los cuales se realizaron las investigaciones incluidas en el presente documento.



Figura 1. Países en los cuales se realizaron las investigaciones [17].

En la Tabla 4 se presenta un resumen de los modelos de AA analizados en las diferentes publicaciones de investigación sobre valuación inmobiliaria durante el período 2022-2023.

Como se puede apreciar, en la Tabla 4 se representa por cada fila una publicación, identificada por el año y los autores, mientras que las columnas indican los diferentes modelos de aprendizaje automático considerados. Una “X” en la intersección de una fila y una columna significa que la publicación correspondiente analizó ese modelo específico. Esta ofrece una visión general de los modelos de AA analizados en cada documento incluido en la presente investigación.

Tabla 4. Modelos analizados por publicación.

Año y Autor	Modelo														
	RF	LR	NN	Boosting	GBM	XGBoost	MLR	SVR	DT	VR	Ridge	Lasso	SVM	DNN	PR
2023 Chiasson, E., <i>et al.</i> [8]	X	X	X												
2023 Deppner, J., <i>et al.</i> [9]				X											
2023 Mody, P., <i>et al.</i> [10]	X					X	X	X	X						
2023 Nazarov, M, y Yarmatov, S. [11]		X			X				X	X	X	X			
2023 Stang, M., <i>et al.</i> [12]					X										
2023 Putri, M. R., <i>et al.</i> [13]	X	X									X	X			X
2023 Gunes, T. [14]	X					X		X							
2022 Jung, J., <i>et al.</i> [15]	X				X								X	X	
2022 Matey, V., <i>et al.</i> [16]		X							X			X			

Fuente: Elaboración propia.

3. Resultados

Esta sección debe proporcionar una descripción concisa y precisa de los resultados experimentales y su interpretación. La Tabla 5 presenta un breve un panorama de las contribuciones recientes en la aplicación de AA a la valuación inmobiliaria.

Tabla 5. Resumen de aportaciones.

Año y Autor	Contribución
2023 Chiasson, E., <i>et al.</i> [8]	Proporciona una herramienta de código libre realizada en Python, empleando aprendizaje automático, obteniendo un

	formato final con información acerca de la valuación realizada.
2023 Deppner, J., <i>et al.</i> [9]	Evalúa el potencial de los algoritmos de aprendizaje automático para reducir las desviaciones entre los valores de mercado y los precios de transacción posteriores, buscando hacerlas más eficientes y objetivas.
2023 Mody, P., <i>et al.</i> [10]	Implementación y evaluación de diferentes modelos de aprendizaje automático, proporcionando nuevas opciones valiosas para el ámbito de las propiedades inmobiliarias, como son el aprovechar el aprendizaje automático y los datos obtenidos de la web por medio de <i>scraping</i>
2023 Nazarov, M, y Yarmatov, S. [11]	Prueba y análisis de aplicación de diferentes modelos de aprendizaje automático en la valuación inmobiliaria, se sugiere el uso de métodos de conjunto para mejorar la precisión, debido a que los modelos individuales pueden ser propensos a sobreajuste o falta de robustez
2023 Stang, M., <i>et al.</i> [12]	Se prueban diferentes modelos de aprendizaje automático y se mide su desempeño, concluyendo que el uso de métodos de aprendizaje automático para la valoración de inmuebles es beneficioso en muchas situaciones y que su aprobación debería ser discutida por las autoridades reguladoras.
2023 Putri, M. R., <i>et al.</i> [13]	Se propone analizar mediante diferentes modelos aprendizaje automático la información disponible en la web, utilizando métodos recopilación de información automatizado, se realizan pruebas mediante Júpiter IDE.
2023 Gunes, T. [14]	Pruebas y análisis de predicción del precio de la vivienda usando el aprendizaje supervisado, propone el algoritmo LIME y los valores de Shapley, que pueden explicar las predicciones de cualquier clasificador o regresor de una manera fiel.
2022 Jung, J., <i>et al.</i> [15]	Se prueban diferentes modelos de aprendizaje automático y se mide su desempeño, además se considera que los tasadores que reciben la retroalimentación del precio de transacción tienden a ajustar sus valoraciones hacia el precio observado, sugiriendo un sesgo de anclaje persistente, que puede ser evitado empleando aprendizaje automático en las valuaciones inmobiliarias.
2022 Matey, V., <i>et al.</i> [16]	Se prueban diferentes modelos de aprendizaje automático y se mide su desempeño, afirmando que los métodos aprendizaje automático pueden aportar transparencia a la valuación inmobiliaria.
Fuente: Elaboración propia.	

Como se observa en la Tabla 5, se presenta un panorama dinámico de investigación y desarrollo en el campo de la valuación inmobiliaria con AA, sintetiza las aportaciones de distintos autores e investigadores durante el periodo 2022-2023 en el ámbito de la valuación inmobiliaria con técnicas de aprendizaje automático. Se aprecia una tendencia evidente hacia la exploración y aplicación de estas técnicas para mejorar la exactitud, eficiencia y objetividad de las valuaciones.

En la Figura 2 se representan los modelos contrastados gráficamente con los modelos destacados por su precisión, considerando dentro de la categoría “analizado”, la cantidad de modelos de AA que se consideraron para su análisis de precisión dentro de los trabajos considerados en esta revisión y, por otra parte, los “destacados”

como los modelos de AA que se clasifican como modelos con una alta precisión según los resultados de cada documento.

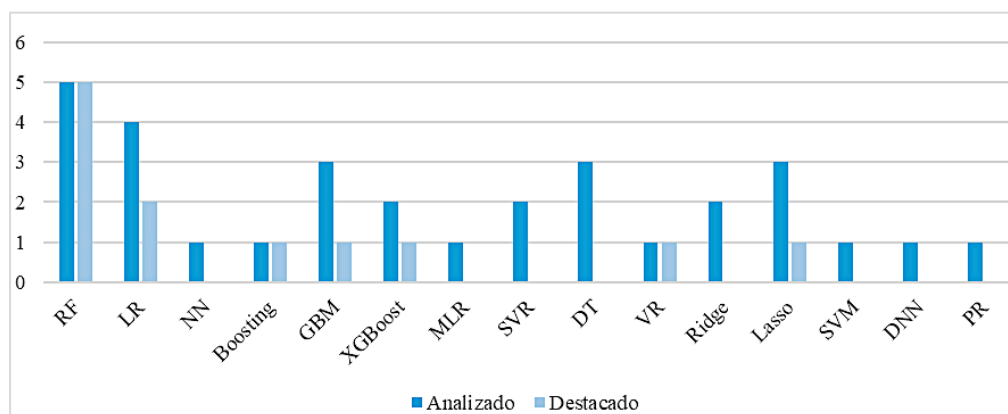


Figura 2. Modelos analizados / destacados. Fuente: Elaboración propia.

En la Figura 2 se realizó una comparación de modelos de AA, contrastando los modelos analizados con los que se destacaron por su precisión, se incluyeron 15 modelos diferentes conforme los artículos analizados, destacando 7 que podrían ser considerados para un análisis más profundo o para su implementación.

4. Discusión

A continuación, se discute información relevante de los artículos seleccionados, tomando como propósito de este estudio analizar y discutir la literatura de investigación publicada del 2022 al 2023, referente a la implementación de AA en la valuación inmobiliaria, todas las publicaciones seleccionadas profundizan en la aplicación de AA en la valuación inmobiliaria y presentan un análisis de la precisión resultante del entrenamiento de diferentes modelos, empleando información de inmuebles relacionados con su valor comercial. Si bien, el enfoque principalmente fue revisar el desempeño de los diferentes modelos de AA, se incluyen comparaciones con otros métodos como son el método hedónico y manual, obteniendo los modelos de AA un mejor desempeño [8]. De forma general los estudios utilizan conjuntos de datos específicos (una ciudad, región o tipo de propiedad), revelando la importancia en mejorar la interpretabilidad de los modelos e integrar datos contextuales adicionales [16].

Chiasson *et al.* [8] propone el sistema nombrado REALM, una solución de código abierto para realizar valuaciones inmobiliarias empleando aprendizaje automático. El problema se aborda debido a la falta de soluciones para realizar tasaciones automáticas que pueda ser una vía diferente a los métodos hedónicos automatizados y métodos manuales que tradicionalmente se emplean en la valuación inmobiliaria. REALM automatiza la valuación inmobiliaria, utiliza un modelo de predicción de precios con bosques aleatorios RF y extrae propiedades comparables para generar informes de la valuación en PDF. Se evalúa la herramienta con el conjunto de precio de viviendas en Ames, Iowa, publicado por “*The Journal of Statistics Education*”, que contiene propiedades información y precios de venta de aproximadamente 3,000 casas vendidas en Ames, Iowa, entre 2006 y 2010, comparando la precisión de los modelos RF, LR y NN, así como, modelos de valuación hedónica automatizados y métodos manuales, se eligió el modelo de RF por defecto para la plataforma sobre los modelos NN y LR debido a la mayor precisión presentada en las pruebas realizadas utilizando el conjunto de datos de muestra. La contribución principal es la automatización eficaz de la valuación inmobiliaria con informes detallados. Sin embargo, REALM no considera factores externos como ubicación específica de las propiedades o servicios cercanos, entre otros factores importantes en la estimación del valor de una propiedad, por lo que se considera que los resultados podrían mejorarse aún más integrando datos adicionales.

Jung *et al.* [15] trata sobre el uso de métodos de aprendizaje automático para estimar los precios de transacción de los bienes raíces comerciales. Los autores comparan la precisión predictiva de varios métodos de aprendizaje automático, como RF, GBM, SVM y DNN, con el enfoque hedónico, emplean una base de datos de 19,640 propiedades de oficinas en 10 áreas metropolitanas de EE. UU. entre 2004 y 2017. Se explican los diferentes aspectos que influyen al tasar un inmueble, como pueden ser las desviaciones en la información de los

métodos de capitalización derivado de flujos de pagos en los que no se reflejan las transacciones realizadas en efectivo o como los costos de mantenimiento muy variables influyen directamente en la utilidad. Además, expone como los inversores que no son de la localidad al adquirir una propiedad pagan un 13.5% más, y venden un 7% por debajo de su valor en relación con los precios locales. Por lo que se consideran los modelos de aprendizaje automático como una posible respuesta, ya que las desviaciones que se presentan en el valor final al implementar métodos tradicionales se atribuyen a características imprecisas, impactando directamente en cálculos predefinidos que modifican los valores de las propiedades en cuestión. Sin embargo, al implementar aprendizaje automático apunta a una presión mayor, ya que en principio la precisión puede ajustarse para mejorar el comportamiento de los modelos. Se aplican los cuatro métodos de aprendizaje automático antes mencionados para predecir los precios de las transacciones y comparar su precisión con el modelo hedónico. Concluyendo que RF y el GBM tienen un mejor rendimiento que los otros métodos. Por otra parte, los resultados muestran que los tasadores que reciben la retroalimentación del precio de transacción tienden a ajustar sus valoraciones hacia el precio observado. Esto sugiere que el sesgo de anclaje es persistente y puede afectar la calidad y la credibilidad de las valoraciones comerciales.

Nazarov y Yarmatov [11] analiza el uso de métodos de conjunto en el aprendizaje automático para mejorar la precisión de las predicciones en la valoración de bienes inmuebles. Derivado de las limitaciones de los modelos de aprendizaje automático independientes debido a que los modelos individuales pueden ser propensos a sobreajuste o falta de robustez, se proponen métodos de conjunto para mejorar la precisión. Los autores proponen VR y *Gradient Boosting*, que combinan las predicciones de varios modelos de aprendizaje automático, como RL, DT y RF. Se evalúan el rendimiento de los métodos propuestos utilizando un conjunto de datos de 16,000 registros de propiedades inmobiliarias en la región de Samarcanda, Uzbekistán, y comparan sus resultados con los de los modelos individuales y otros métodos de conjunto. Se concluye que VR es el más efectivo para predecir el precio de las propiedades inmobiliarias, con una precisión del 90%, seguido de xgBoost, con una precisión del 86%. Los autores sugieren que estos métodos pueden ser útiles para los inversores que quieran financiar proyectos inmobiliarios en el futuro.

Mody *et al.* [10] presenta un estudio sobre la predicción de precios de propiedades inmobiliarias en Mumbai usando técnicas de regresión de aprendizaje automático. El conjunto de datos utilizado se compone de propiedades en ciudades metropolitanas, obtenidas mediante *scraping*; una técnica utilizada en el campo de la informática, en particular en el aprendizaje automático y la minería de datos se emplea para extraer información de sitios web. Se realizaron análisis estadísticos y visualizaciones para comprender las relaciones entre las variables e identificar patrones relevantes. Por último, se implementaron y evaluaron diversas técnicas de regresión de aprendizaje automático, como MLR, SVR, DT, RF, XGBoost, los resultados muestran que el modelo de RF logra el mayor nivel de precisión y el mejor ajuste a los datos. El estudio aporta el aprovechar el aprendizaje automático y los datos obtenidos de la web, proporcionando percepciones valiosas para los interesados en el ámbito de las propiedades inmobiliarias. Estas percepciones pueden ayudar a la formulación de decisiones bien informadas sobre inversiones inmobiliarias, metodologías de valoración y tendencias del mercado.

Matey *et al.* [16] explora el uso de aprendizaje supervisado para la predicción del precio de la vivienda en la ciudad de Pune, India. Se menciona que el mercado inmobiliario en India es opaco y los precios de las viviendas a menudo están inflados por lo que existe una necesidad de un sistema que prediga los precios de las viviendas de manera precisa y transparente, el proyecto busca ayudar a los compradores de viviendas a encontrar propiedades que se ajusten a sus presupuestos. Se considera Varios sitios web, incluidos *kaggle*, *magicbricks*, *99acres* y *ready rates*, un sitio web del gobierno que proporciona valores inmobiliarios actualizados. Se analizaron las características más relevantes para la predicción de precios, empleando tres técnicas de regresión: LR, Lasso y DT, basándose en varios factores como los atributos físicos, el diseño y la ubicación de las propiedades. Se obtiene un buen desempeño de los modelos analizados, sin embargo, se concluye que LR tiene más precisión entre los conjuntos de entrenamiento y prueba. El análisis de características reveló que el número de habitaciones, el área y la ubicación son factores importantes en la predicción de precios. El aprendizaje supervisado puede ser utilizado para predecir los precios de las viviendas con una precisión razonable lo que se considera una herramienta útil para los compradores de viviendas en Pune. En trabajos futuros se propone explorar otros algoritmos de aprendizaje automático, como DT, *Naive Bayes* y SVM. Además de utilizar el algoritmo *K*-vecinos más cercanos KNN para mejorar la precisión. Por otra parte, considerar factores adicionales que influyen en los precios de las viviendas, como la proximidad a servicios esenciales, que pueden ser considerados en trabajos futuros.

Stang *et al.* [12] compara diferentes enfoques para construir Modelos de Valoración Automatizados AVM a nivel nacional en Alemania, con el objetivo de proporcionar evidencia empírica para el debate sobre el uso futuro de las valoraciones automatizadas en el proceso de préstamos. Se comparan cuatro métodos: Función Experta EXF: Una automatización del método de comparación de ventas utilizando filtros y funciones de similitud; Regresión de Mínimos Cuadrados Ordinarios OLS: Un modelo hedónico tradicional que asume una relación lineal entre las variables; Modelo Aditivo Generalizado GAM: Un modelo hedónico más flexible que puede capturar relaciones no lineales; Por último, se compara con el modelo XGBoost basado en DT. Se utiliza un conjunto de datos único de 1.2 millones de propiedades residenciales en Alemania, valoradas por tasadores profesionales entre 2014 y 2020. Se encuentra que el método de aprendizaje automático XGBoost ofrece el mejor rendimiento en términos de precisión de las estimaciones, flexibilidad y facilidad de implementación, también muestran que diferentes tipos de métodos funcionan mejor en diferentes regiones y que la disponibilidad de datos influye en el rendimiento de los métodos. Los algoritmos de aprendizaje automático como XGBoost son prometedores para su uso en AVMs debido a su alta precisión y flexibilidad. Por último, este estudio proporciona evidencia empírica que respalda la inclusión de los algoritmos de aprendizaje automático en la discusión sobre el futuro de las valoraciones automatizadas.

Putri *et al.* [13] resalta la importancia de la vivienda como necesidad básica de las personas, así como, la importancia que tiene la valuación para los inversionistas y compradores para conocer el valor que logre un beneficio para ambos. Basado en estudios anteriores revisados por los autores se determina que ML es viable para realizar predicciones del precio de las viviendas, además, destaca que los modelos de regresión proporcionan la relación que existe en una variable continua con uno o más variables independientes. Por lo tanto, se comparan cinco modelos de regresión: MLR, Ridge, Lasso, RF y PR, para predecir precios de las viviendas en la provincia Nusa Tenggara, la base de datos se obtuvo mediante web *scraping* del sitio web Lamundi, limitando los resultados a la provincia de Nusa Tenggara occidental, contando con 600 datos de viviendas con diversas especificaciones, estos datos se analizaron removiendo datos duplicados, rectificando información errónea y revisando inconsistencias. Posteriormente, se analizaron los datos considerando las características de las propiedades. Para el análisis y entrenamiento de los modelos se empleó Júpiter IDE, una aplicación basada en web que ayuda a crear y compartir documentos de dichos procesos. Según la evaluación realizada se muestra que el mejor modelo es MRL y regresión Lasso con el mismo R^2 y RMSE, mientras tanto RMSE (validación cruzada) muestra que el mejor modelo es RF. Los resultados presentan implicaciones significativas para todas las personas involucradas en transacciones de bienes raíces, particularmente viviendas. Este estudio ofrece una aplicación profunda del desempeño y comparación de modelos de regresión utilizados para la predicción de precios de viviendas.

Gunes [14] trata sobre la opacidad que puede existir en la aplicación de ML, a pesar de ello, se reconoce el mejor desempeño en la valuación de propiedades. Por lo que, revisa el uso de métodos de aprendizaje automático ML y de interpretabilidad de modelos IML para la valoración de propiedades residenciales. El autor aplica tres algoritmos: XGBoost, RF y SVR, para predecir los precios de las propiedades y compara sus rendimientos. Afirma que los métodos de ensamble mejoran de forma general el desempeño de las predicciones combinando múltiples modelos de ML. En este caso se selecciona el modelo de RF como el mejor y utiliza métodos de IML para explicar su comportamiento tanto a nivel global como local. Se utilizan herramientas como la importancia de las características, los efectos de las características, las interacciones de las características, el LIME y los valores de Shapley para revelar las relaciones internas entre las características de las propiedades y los precios. LIME (*Local Interpretable Model-Agnostic Explanations*) es un algoritmo que puede explicar las predicciones de cualquier clasificador o regresor de una manera fiel, aproximándolo localmente con un modelo interpretable. Los Valores de Shapley son un método de la teoría de juegos cooperativos que se utiliza para distribuir equitativamente las características evitando tendencias que puedan afectar los resultados. El autor afirma que los métodos de IML pueden aportar transparencia a los modelos de ML opacos y visualizar los efectos e interacciones de las características en los modelos de valoración de propiedades.

Deppner *et al.* [9] investiga la precisión de las valoraciones de inmuebles comerciales en Estados Unidos y el potencial del aprendizaje automático para mejorarla. Se analizan datos de transacciones de propiedades del NCREIF *Property Index* entre 1997 y 2021, se utiliza una muestra de 7,133 transacciones de propiedades comerciales de cuatro tipos: apartamentos, industriales, oficinas y comerciales. De este modo, se encuentran que las valoraciones tradicionales suelen ser inexactas y presentar sesgos estructurales. El estudio aplica un algoritmo de *boosting* para explicar y reducir las desviaciones entre los valores de tasación y los precios de venta reales, logrando mejoras significativas en la precisión y eliminando el sesgo en los cuatro tipos de propiedades

comerciales principales: apartamentos, industriales, oficinas y minoristas. Los resultados sugieren que el aprendizaje automático puede proporcionar información valiosa para mejorar las prácticas de valoración actuales, especialmente al considerar factores espaciales y estructurales que los métodos tradicionales podrían pasar por alto. Si bien el aprendizaje automático presenta limitaciones, ofrece un gran potencial para aumentar la eficiencia, la objetividad y la precisión en la valoración de inmuebles comerciales, beneficiando a una amplia gama de actores del sector. Se concluye que los métodos de aprendizaje automático pueden ofrecer una mejora sustancial a las prácticas actuales de tasación, al hacerlas más eficientes y objetivas, pero también advierten de las limitaciones y desafíos de estos métodos, como son la disponibilidad de datos, falta de justificación ya que los modelos de aprendizaje automático se basan en la identificación de patrones en los datos, pero no consideran las leyes económicas ni establecen relaciones de causalidad. Esto dificulta la comprensión de las razones detrás de las predicciones y la toma de decisiones, además, los modelos de ML son difíciles de interpretar, lo que genera dudas sobre su confiabilidad y transparencia, entre otros. Por lo que, la industria de la valoración inmobiliaria puede ser reacia a adoptar nuevas tecnologías y métodos, especialmente si no se comprenden completamente sus fundamentos y limitaciones. Se propone encontrar un equilibrio entre la precisión de los modelos de aprendizaje automático y la interpretabilidad de los métodos tradicionales para lograr una valoración eficiente y confiable. Se discute la viabilidad de emplear modelos de aprendizaje automático para la valuación inmobiliaria. A pesar de implementar y evaluar diferentes modelos de aprendizaje automático, arrojando resultados positivos, el funcionamiento de estos no es totalmente transparente. Sin embargo, su utilidad para inversores inmobiliarios puede ser una herramienta más que ayude a comparar los métodos convencionales, y así, aportar transparencia a la valuación inmobiliaria y evitar sesgos de anclaje en las valoraciones. Además, es necesario continuar analizando el uso de algoritmos para explicar las predicciones de cualquier clasificador o regresor de manera fiel, como el algoritmo LIME y los valores de Shapley, los cuales podrían atenuar la suspicacia acerca de su funcionamiento.

5. Conclusiones

El principal objetivo del estudio se logró identificando los modelos de AA con mejor desempeño. Así como, los más implementados en la valuación inmobiliaria, considerando las características de las bases de datos de la Tabla. Se englobaron diferentes modelos de AA y contextos de valuación. Finalmente, la RSL realizada proporciona una base para futuras investigaciones sobre la aplicación de AA en la valuación inmobiliaria en México.

El propósito de esta RSL se alcanzó al conocer cuáles modelos de aprendizaje automático son más eficientes y precisos para predecir el valor de una propiedad. La información empleada en los diferentes artículos para realizar los entrenamientos de los modelos de aprendizaje automático fue variada y específica para cada estudio. Principalmente, se utilizaron conjuntos de datos que incluyeron características numéricas y categóricas, en el presente análisis no se consideraron bases de datos que incluyeran imágenes, ubicaciones o modelos gráficos. Se considero información de transacciones realizadas o bien ofertas publicadas mediante internet, estos conjuntos de datos usualmente se obtuvieron de fuentes públicas o bases de datos especializadas, y según lo observado a lo largo de la revisión de los artículos incluidos, esta información fue cuidadosamente seleccionada y preprocesada para asegurar su idoneidad para el modelo. Además, se prestó especial atención a la calidad de los datos, asegurándose de que fueran representativos, completos y libres de sesgos, lo cual es esencial para garantizar la precisión y la generalización de los modelos entrenados.

Se concluye que con base a los artículos analizados el modelo de *Random Forest* (RF) o Bosques Aleatorios, presentaron una precisión mayor, decantándose por él en 5 de 9 artículos, por otra parte, el modelo de *Linear Regression* (LR) o Regresión Lineal, se analiza en 4 de los 9 artículos, destacando en dos de ellos como el más preciso, por lo que se infiere que estos 2 modelos se destacan, por lo que estos son viables para el análisis de bases de datos con información de inmuebles que incluyan sus características de los inmuebles, así como, sus precios.

6. Trabajos futuros

Las investigaciones incluidas en esta revisión buscan llegar a encontrar cuáles modelos de aprendizaje automático son los más destacados, conforme a la base de datos con características similares a la disponible en el Sistema Nacional de Información e Indicadores de Vivienda, mencionada al inicio de este documento, se pretende de llevar a cabo el análisis de esta base de datos empleando los modelos reconocidos como los de mejor rendimiento.

Además, es importante destacar que el proceso de valuación inmobiliaria en México privilegia el enfoque comparativo de mercado [18]. En ese sentido, es importante considerar aún más información disponible en páginas web públicas mediante las cuales se promueven bienes raíces como son: vivanuncios.com, inmuebles24.com, lamudi.com.mx, portalterreno.com, etc. Esta información también podría ser analizada mediante aprendizaje automático y así enriquecer la información e intentar incrementar la precisión de los resultados [4].

7. Agradecimientos

Se agradece a la Universidad Autónoma de Sinaloa por brindar acceso a su biblioteca y a sus bases de datos científicas, recursos indispensables para la obtención de la información necesaria en la realización de la revisión sistemática de literatura. Asimismo, se reconoce y valora el apoyo brindado por los docentes de la institución, cuyo conocimiento y orientación fueron fundamentales para el desarrollo de este trabajo.

8. Referencias

- [1] Kok, N., Koponen, E. L., Martínez-Barbosa, C. A. (2017). Big data in real estate? from manual appraisal to automated valuation. *Journal of Portfolio Management*, 43 (6), 202–211. <https://doi.org/10.3905/jpm.2017.43.6.202>
- [2] Secretaría de Economía and Gobierno de México. (2023). *NORMA NMX-R-081-SCFI-2015*. <http://economia-nmx.gob.mx/normas/nmx/2010/nmx-r-081-scfi-2015.pdf>
- [3] Shabbir, J., Anwer, T. (2018). *Artificial Intelligence and its Role in Near Future*. arXiv. <http://arxiv.org/abs/1804.01396>
- [4] Shinde, P., Shah, S. (2018). *A Review of Machine Learning and Deep Learning Applications*. Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India. <http://dx.doi.org/10.1109/ICCUBEA.2018.8697857>
- [5] Mahesh, B. (2018). Machine Learning Algorithms-A Review. *International Journal of Science and Research*, 9 (1), 381-386. <https://doi.org/10.21275/ART20203995>
- [6] SNIIV. (2024). *Datos abiertos de financiamiento en México*. https://sniiv.sedatu.gob.mx/Reporte/Datos_abiertos
- [7] Gough, D., Oliver, S., Thomas, J. (2017). *An introduction to systematic reviews* (2nd Ed.). SAGE Publications Ltd.
- [8] Chiasson, E., Kaniecki, M., Koechling, J., Uppal, N., Hammad, I. (2023). *REALM: Automating Real Estate Appraisal with Machine Learning Models*. IEEE World AI IoT Congress (AIIoT). Seattle, WA, USA. <https://doi.org/10.1109/AIIoT58121.2023.10174323>
- [9] Deppner, J., von Ahlefeldt-Dehn, B., Beracha, E., Schaefer, W. (2023). Boosting the Accuracy of Commercial Real Estate Appraisals: An Interpretable Machine Learning Approach. *Journal of Real Estate Finance and Economics*. <https://doi.org/10.1007/s11146-023-09944-1>
- [10] Mody, M., Motiramani, M., Singh, A. (2023). *Enhancing Real Estate Market Insights through Machine Learning: Predicting Property Prices with Advanced Data Analytics*. 4th IEEE Global Conference for Advancement in Technology (GCAT). Bangalore, India. <https://doi.org/10.1109/GCAT59970.2023.10353243>
- [11] Nazarov, F. M., Yarmatov, S. (2023). *Optimization of Prediction Results Based on Ensemble Methods of Machine Learning*. International Russian Smart Industry Conference (SmartIndustryCon). Sochi, Russian Federation. <https://doi.org/10.1109/SmartIndustryCon57312.2023.10110726>
- [12] Stang, M., Krämer, B., Nagl, C., Schäfers, W. (2023). From human business to machine learning—methods for automating real estate appraisals and their practical implications. *Zeitschrift für Immobilienökonomie*, 9 (2), 81–108, <https://doi.org/10.1365/s41056-022-00063-1>
- [13] Putri, M. R., Wijaya, I. G. P. S., Praja, F. P. A., Hadi, A., Hamami, F. (2023). *The Comparison Study of Regression Models (Multiple Linear Regression, Ridge, Lasso, Random Forest, and Polynomial Regression) for House Price Prediction in West Nusa Tenggara*. International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS). Bali, Indonesia. <https://doi.org/10.1109/ICADEIS58666.2023.10270916>

- [14]Gunes, T. (2023). Model agnostic interpretable machine learning for residential property valuation. *Survey Review*, 1-16. <https://doi.org/10.1080/00396265.2023.2293366>
- [15]Jung, J., Kim, J., Jin, C. (2022). Does machine learning prediction dampen the information asymmetry for non-local investors? *International Journal of Strategic Property Management*, 26 (5), 345–361. <https://doi.org/10.3846/ijspm.2022.17590>
- [16]Matey, V. Chauhan, N., Mahale, A., Bhistannavar, V., Shitole, A. (2022). *Real Estate Price Prediction using Supervised Learning*. IEEE Pune Section International Conference (PuneCon). Pune, India. <https://doi.org/10.1109/PuneCon55413.2022.10014818>
- [17]Matplotlib (2024). *Using Matplotlib. Documentación oficial*. <https://matplotlib.org/stable/users/index.html>
- [18]Salas Tafoya, J. M. (2015). El Modelo de Valuación Inmobiliaria en México. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 5 (10), 31-54. <https://www.ride.org.mx/index.php/RIDE/article/view/196>