



Desarrollo de aplicación web prototipo para estimación de valores de viviendas en México aplicando modelos de aprendizaje automático

Development of a prototype web application for estimating housing values in Mexico using machine learning models

Fabián Espinoza Garza

Universidad Autónoma Indígena de México, Los Mochis, México

fabianespinozagarza@uaim.edu.mx

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



<https://doi.org/10.36825/RITI.13.32.003>

Recibido: Junio 15, 2025

Aceptado: Octubre 08, 2025

Resumen: Se presenta el desarrollo de un prototipo de aplicación web para estimar el valor de viviendas en México, utilizando modelos de aprendizaje automático (AA). Se describe la evolución de la valuación en México y la oportunidad que representa el AA para mejorar la precisión y eficiencia. El artículo se basa en datos abiertos del Sistema Nacional de Información e Indicadores de Vivienda (SNIIV) y emplea modelos *Random Forest* (RF) y *Linear Regression* (LR), destacando la superioridad del primero en rendimiento. Se detalla la metodología, arquitectura y herramientas tecnológicas utilizadas para construir la aplicación y se presentan los resultados de precisión por entidad federativa, mostrando la viabilidad técnica del prototipo a pesar de limitaciones como la variabilidad del desempeño y la falta de normatividad para su aplicación en casos reales.

Palabras clave: Valuación Inmobiliaria, Vivienda, Prototipo, Aprendizaje Automático.

Abstract: The development of a prototype web application for estimating home values in Mexico using machine learning models (ML) is presented. The evolution of valuation in Mexico is described, along with the opportunity that machine learning represents to improve accuracy and efficiency. The article is based on open data from the National Housing Information and Indicators System (SNIIV) and employs Random Forest (RF) and Linear Regression (LR) models, highlighting the superiority of the former in performance. The methodology, architecture, and technological tools used to build the application are detailed, and the accuracy results are presented by state, demonstrating the technical feasibility of the prototype despite limitations such as performance variability and the lack of regulations for its application in real-life cases.

Keywords: *Real Estate Valuation, Housing, Prototype, Machine Learning.*

1. Introducción

La práctica de la valuación en México ha atravesado un proceso de evolución, pasando de un ejercicio empírico basado en el “leal saber y entender” del perito, hacia una disciplina técnica, normativa y profesionalmente consolidada. Esta transformación se vio impulsada por la intervención de instituciones como Banobras y la Comisión de Avalúos de Bienes Nacionales, así como, por la estandarización promovida por la Comisión Nacional Bancaria, lo que derivó en una mayor profesionalización del gremio [1].

La formación de asociaciones especializadas y del Instituto Nacional de Valuadores consolidó un marco que favoreció la organización y normatividad, dando lugar al desarrollo de normas técnicas, generación de bases de datos sistematizadas y creación documentos oficiales que definen los principios y valores que deben guiar la conducta del valuator, como la objetividad, la confidencialidad, la honestidad y la independencia [1].

En este contexto, la normatividad constituye una referencia clave para garantizar la calidad y confiabilidad de los servicios de valuación, al establecer lineamientos metodológicos que protegen los intereses de los usuarios, unifican criterios técnicos, y promueven la homologación de principios a nivel nacional e internacional [2].

Sin embargo, a pesar de estos avances, los métodos convencionales de valuación siguen siendo vulnerables a sesgos al realizar la selección de propiedades comparables, a uso de información imprecisa y a la omisión de variables relevantes, lo que provoca estimaciones poco representativas del mercado y disminuye la precisión y confiabilidad de los resultados [3].

La integración de técnicas de AA representa una oportunidad para optimizar el proceso al realizar valuaciones mediante modelos que mejoran en precisión y eficiencia frente a los métodos convencionales [4].

No obstante, la mayoría de las herramientas disponibles actualmente son de acceso restringido o cerradas al escrutinio académico y profesional. De acuerdo con [5], se tiene una limitada disponibilidad de soluciones de código abierto que permitan una evaluación transparente y replicable de los modelos aplicados.

Por otra parte, el avalúo inmobiliario convencional es costoso, lento y propenso a errores humanos, dado que requiere cálculos manuales y comparaciones por parte de un valuator profesional. Si bien, estudios previos han demostrado que los modelos de AA superan en precisión, velocidad y costos a los métodos tradicionales. Además, algunas otras iniciativas también aplican técnicas de AA a la valuación inmobiliaria, son pocas las que ofrecen soluciones basadas en código abierto [5].

El uso de AA en la valuación inmobiliaria permite mejorar la precisión de las estimaciones, reducir los sesgos humanos, procesar grandes volúmenes de datos y disminuir tiempos y costos operativos. Estos modelos ofrecen una actualización continua de valores en función de nuevas condiciones del mercado y permiten identificar con claridad las variables que más influyen en los precios. En conjunto, su implementación representa un avance hacia valuaciones más objetivas, eficientes y escalables frente a los métodos tradicionales [6].

En consonancia con lo planteado por [5], el desarrollo de herramientas abiertas que permitan generar avalúos aplicando modelos de AA, adaptables a cualquier región, representa una oportunidad clave para México. Este enfoque no solo facilitaría la investigación y el acceso a tecnologías avanzadas, sino que también impulsaría la equidad, la eficiencia y la modernización de las valuaciones en contextos con recursos limitados.

A pesar del notable avance en la profesionalización y normatividad de la valuación inmobiliaria en México, aún no es posible incorporar formalmente técnicas de AA en la elaboración de avalúos, debido a la ausencia de un marco legal y normativo que regule su aplicación. Las normas vigentes, continúan sustentando metodologías tradicionales y no contemplan el uso de modelos automatizados o algoritmos predictivos, lo que limita la integración de enfoques modernos [7].

No obstante, el estudio de estas tecnologías es relevante y necesario, ya que las tendencias globales muestran una adopción progresiva en diversos ámbitos profesionales. La revisión de artículos científicos recientes confirma un creciente interés y validación académica del uso de modelos de AA en la valuación inmobiliaria, lo que anticipa su futura incorporación normativa. Ejemplo de ello son las investigaciones de [8], [9] y [10], que aplican modelos de AA a la valuación de viviendas utilizando múltiples bases y conjuntos de datos, evaluando su precisión mediante métricas como R^2 y MAPE.

Adicionalmente, en [5] el autor propone un prototipo funcional de código abierto que automatiza la valuación inmobiliaria mediante AA. Desarrollado con el marco de trabajo Django, el sistema genera reportes con el valor estimado, incluyendo las propiedades empleadas como comparables, usando como modelo predeterminado RF, seleccionado por su alta precisión, R^2 superior a 0.9. La herramienta es adaptable a cualquier base de datos y región, y representa una valiosa aportación práctica para la investigación y automatización en el ámbito del avalúo inmobiliario.

En el presente trabajo se desarrolla un prototipo, inspirado en el prototipo planteados por [5], orientado a la valuación automatizada de viviendas en México. El objetivo es sentar las bases para futuras investigaciones, y el desarrollo de esta investigación se organiza de la siguiente manera:

En el primer lugar se describen los enfoques de valuación tradicionales, se propone el uso de modelos de AA como alternativa metodológica y se mencionan los modelos de AA más viables y la fuente de datos.

En el segundo lugar se establece la metodología aplicada, detallando el modelo de proceso evolutivo y la filosofía "Agile" utilizada para el desarrollo del prototipo, las características funcionales que incluirá, las métricas de evaluación, las variables independientes a considerar y el proceso de preprocesamiento de los datos. También se especifican las herramientas tecnológicas seleccionadas.

En el tercer lugar se describe el desarrollo del prototipo, explicando la arquitectura de la aplicación web, el entrenamiento de los modelos y la interfaz gráfica para la captura de parámetros y visualización de resultados, incluyendo las métricas de evaluación.

En el cuarto lugar se muestran los resultados, presentando el desempeño de los modelos en distintas entidades federativas, mostrando tablas y gráficos con las métricas RMSE (*Root Mean Squared Error* - Raíz del Error Cuadrático Medio), MAPE (*Mean Absolute Percentage Error* - Error Porcentual Absoluto Medio) y R^2 (*Coefficient of Determination* - Coeficiente de Determinación), y ejemplos de la interfaz del prototipo con una estimación de valor.

En el quinto lugar se desarrolla la discusión, analizando las fortalezas del prototipo, como su diseño modular y el uso de datos abiertos, así como sus limitaciones, incluyendo el desempeño variable por entidad, la interpretabilidad limitada de los modelos de AA y la ausencia de un marco normativo.

Por último, se llega a las conclusiones, resumiendo el potencial del prototipo como una alternativa viable y eficaz para la valuación automatizada de viviendas en México, destacando la precisión y replicabilidad de los modelos y la transparencia de la arquitectura tecnológica, al tiempo que reitera los desafíos persistentes.

2. Antecedentes

Para llevar a cabo una valuación, se inicia con la recopilación de la información necesaria y definición del propósito del avalúo para determinar los enfoques de valuación aplicables (mercado, costo e ingresos), seguido por la delimitación del alcance del trabajo, en función del propósito y tipo de bien a valorar. Posteriormente, se realiza la implementación de la valuación mediante inspección física, análisis documental y estudio del mercado. Finalmente, se elabora un informe técnico claro y sustentado, que incluye la descripción del inmueble, la metodología aplicada y los resultados obtenidos, cumpliendo con los requisitos técnicos, legales y éticos [11].

Se reconoce que, debido a la naturaleza de la metodología aplicada en el presente estudio, no es posible aplicar de manera estricta ninguno de los enfoques reconocidos y aceptados. No obstante, se propone la implementación de modelos de AA como una alternativa metodológica que pueda apoyar como una herramienta adicional para la validación de resultados. Esta estrategia busca aprovechar la capacidad de dichas técnicas para identificar patrones complejos en grandes volúmenes de datos y generar estimaciones de valor con niveles adecuados de precisión, eficiencia y replicabilidad [6].

Con base en la revisión sistemática de literatura realizada, los modelos de AA más viables para la estimación del valor de propiedades inmobiliarias son RF y LR, destacando por su alta precisión en diversos estudios internacionales. RF fue identificado como el modelo con mejor desempeño, mientras que LR también se encontró

entre los mejores, además, de su reiterada inclusión en los diferentes estudios. Ambos modelos demostraron ser eficaces al aplicarse sobre bases de datos estructuradas que contienen características numéricas y categóricas de los inmuebles, lo que sugiere que son herramientas prometedoras para mejorar la precisión y objetividad en la valuación inmobiliaria [12].

En la actualidad, los datos se han convertido en un recurso fundamental, ya que el AA permite analizarlos y extraer patrones que impulsan decisiones más precisas, automatizadas y eficientes en diversos sectores [13].

En México se cuenta con Datos Abiertos del SNIIV, administrada por la Secretaría de Desarrollo Agrario, Territorial y Urbano (SEDATU), que ofrece acceso público a datos estructurados sobre el sector vivienda en México, incluyendo información sobre subsidios, financiamientos, inventarios, oferta y demanda de vivienda, así como datos geográficos. Esta plataforma permite descargar los datos en formatos analizables, facilitando su uso en investigación. Además, cuenta con un repositorio histórico que permite realizar análisis comparativos y estudiar tendencias en el ámbito habitacional del país [14].

Se considera que el desarrollo de un prototipo para valuación inmobiliaria es viable, ya que los modelos de AA como RF y LR permiten generar estimaciones precisas y replicables. Además, el acceso a datos abiertos del SNIIV proporciona una base sólida para entrenar estos modelos con información real del mercado mexicano.

3. Metodología

Se lleva a cabo el desarrollo del prototipo mediante un modelo de proceso evolutivo, para ello se construyeron versiones cada vez más completas, este método es útil en situaciones en las que se define un conjunto de objetivos generales para el software, pero no se identifican requisitos detallados para las funciones y característica, sin embargo, se planea su desarrollo implementando una arquitectura que permita integrar nuevos módulos, realizar ajustes estructurales y escalar funcionalidades sin comprometer la estabilidad del sistema [15].

3.1. Metodología Agile

Para el desarrollo del proyecto se considera la filosofía “Agile” donde se implementa el proceso *Extreme Programming* (XP), tal como se ilustra en la Figura 1. Este proceso se estructura en torno a cuatro actividades principales: planificación, diseño, codificación y pruebas [15].

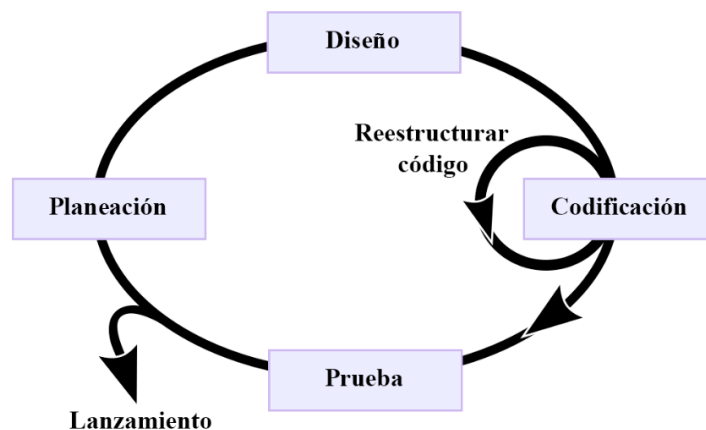


Figura 1. Proceso *Extreme Programming*. Fuente: Adaptado de [15].

En la Figura 1 se muestran las etapas de a) Planeación: Incluye elementos como historias del usuario, valores, criterios de las pruebas de iteración, plan de iteración; b) Diseño: Involucra un diseño simple, soluciones provisionales y prototipos; c) Codificación: Contempla programación en pares, integración continua y refactorización o reestructuración de código; d) Prueba: Considera pruebas de unidad (comprobar el correcto funcionamiento de las unidades más pequeñas e independientes del código, como funciones, métodos o clases) y pruebas de aceptación; y por último, e) lanzamiento: incremento de software (parte del software que se construye

y entrega en un ciclo o iteración) y velocidad calculada del proyecto (medida utilizada para estimar el ritmo de avance y planificar futuras iteraciones) [15].

3.2. Prototipo de aplicación web

El prototipo propuesto para la estimación del valor de viviendas se desarrolla considerando las principales características funcionales. Para esto, se contempla el entrenamiento de al menos un modelo de AA utilizando la información del SNIIV (Sistema Nacional de Información e Indicadores de Vivienda), con el objetivo de predecir el valor de una propiedad a partir de parámetros relevantes. Se incluye una interfaz gráfica accesible desde una página web que permita al usuario capturar datos como año, entidad federativa, municipio, modalidad de vivienda, rangos de ingreso y superficie construida, y visualizar tanto el valor estimado como las métricas R^2 , MAPE y RMSE.

3.3. Métricas

La métrica RMSE por sus siglas en inglés, *Root Mean Squared Error* (raíz del error cuadrático medio), es la desviación típica de los errores en la estimación, es decir, muestra qué tanto se equivocan en promedio las predicciones del modelo. Si usamos $\hat{y}_i$ para representar el precio estimado de la vivienda número i y y_i como el precio real de esta misma vivienda, y tenemos un total de n viviendas en el conjunto de datos [6]. Se calcula con la Ecuación (1).

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (1)$$

Otra métrica muy utilizada es MAPE por sus siglas en inglés, *Mean Absolute Percentage Error* (Error Porcentual Absoluto Medio), es una medida parecida al RMSE, pero en lugar de usar los errores al cuadrado, usa el valor absoluto de los errores y los divide entre el precio real de la vivienda. Así, nos muestra en promedio qué tanto se equivoca el modelo en forma de porcentaje [6]. Se calcula con la Ecuación (2).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (2)$$

Por último [16], indica que R^2 representa el Coeficiente de Determinación, mide qué tanto del cambio en la variable que queremos predecir puede explicarse usando las otras variables del modelo. Su medición es de 0 a 1, y un valor más alto de R^2 significa que el modelo hace mejores predicciones. Se calcula con la Ecuación (3).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

Por otra parte, para combinar el resultado de dos modelos de AA es posible emplear AVG, por sus siglas en inglés, *Average* (Promedio), “promedio ingenuo” de dos modelos de AA, esto se refiere a una forma básica de combinar sus predicciones para mejorar la estimación. Se calcula promediando, punto a punto, las salidas de ambos modelos. Es una técnica sencilla de ensamble que no requiere entrenamiento adicional [17]. Se calcula con la Ecuación (4).

$$AVG_j = \frac{1}{n} \sum_{i=1}^n z_{i,j} \quad (4)$$

Una vez explicado brevemente como se calculan estas métricas que ayudarán a conocer la precisión de los modelos entrenados, abordaremos el diseño del prototipo, para lo cual se propone integrar un panel administrativo para la gestión y actualización de las bases de datos, si bien el prototipo en su fase inicial solo presentará una página de bienvenida, formulario de datos de entrada con selectores y un reporte final, no obstante, se ha dejado listo para integrar (a corto plazo) un panel administrativo para la gestión y actualización de las bases de datos. Con esta idea

en mente, en sus próximas etapas se planea implementar un proceso de preprocesamiento de datos para actualizar la información de las bases de datos empleadas en el entrenamiento y actualización de modelos. Además de almacenar las consultas realizadas por los usuarios, lo cual permitirá su consulta y análisis posteriores.

3.4. Preprocesamiento de datos

A continuación, en la Tabla 1 se enlistan las variables independientes a considerar para la estimación del valor de viviendas en México.

Tabla 1. Variables independientes [14].

Nombre del campo	Alias
año	Año
entidad	Entidad federativa
municipio	Municipio
rango de ingresos	Rangos de ingresos mensuales del acreditado
valor de la vivienda	Valor de la vivienda clasificado en: Económica, Popular, Tradicional, Media, Residencial y Residencial plus.
modalidad	Modalidad de financiamiento (Vivienda: nueva o usada)
monto	Monto del financiamiento

Fuente: Elaboración propia con información de base de datos SNIIV [14].

Las variables consideradas permiten capturar diversos factores que influyen en la determinación del valor de las viviendas en México, integrando dimensiones temporales, geográficas, socioeconómicas y características del financiamiento.

La variable “año” permite analizar la evolución del mercado a lo largo del tiempo, considerando factores como la inflación y la apreciación del suelo. Las variables “entidad federativa” y “municipio” introducen una dimensión espacial que refleja variaciones regionales y locales en los precios inmobiliarios, influenciadas por diferencias en desarrollo urbano, infraestructura y normativas. El “rango de ingresos mensuales del acreditado” funciona como un indicador del poder adquisitivo de los compradores, lo cual guarda una estrecha relación con el tipo de vivienda demandada en la zona. La “modalidad de financiamiento” permite diferenciar entre propiedades con características físicas distintas, lo que incide en su valuación. La variable “valor de la vivienda”, proporciona información cualitativa útil para el análisis de segmentos del mercado.

En este estudio, la variable dependiente es el “monto del financiamiento”, ya que representa una estimación cuantitativa directamente relacionada con el valor de la propiedad, y puede modelarse en función del resto de las variables mencionadas. En conjunto, estos atributos conforman la base de datos para el desarrollo de modelos predictivos enfocados en la estimación del valor de las viviendas en México.

En el análisis tecnológico se identificó que, para la aplicación de modelos de AA, resulta esencial disponer de una fuente de datos con la calidad y cantidad necesarias para el entrenamiento adecuado de los modelos. En el presente estudio, se utilizaron los datos abiertos del SNIIV, los cuales como se mencionó anteriormente proporcionaron la información detallada necesaria sobre financiamientos para realizar el proceso de entrenamiento de modelos de AA que llevarán a estimar los valores de las viviendas.

El proceso de preprocesamiento de la base de datos consistió en una depuración sistemática de los registros provenientes de la base de datos. Inicialmente, se eliminaron atributos redundantes o irrelevantes para la estimación del valor de las viviendas, quedando un conjunto reducido de variables clave. Además, se excluyeron registros correspondientes a modalidades distintas a vivienda nueva o usada, así como datos nulos y valores atípicos, eliminando los extremos del 5 % de la distribución. Los datos fueron agrupados por combinaciones de variables socioeconómicas y geográficas, promediando valores por año.

Por otra parte, se delimitó el rango de análisis a registros posteriores a 2016, año en que entró en vigor la Unidad de Medida y Actualización (UMA) dato fundamental para estimar los valores de las viviendas a través del tiempo, resultando en una base consolidada y coherente. La base de datos conformada inicialmente por 4.09 millones de registros de créditos para vivienda del 2013 al 2025. Finalmente, tras depuración y filtrado, quedó conformada por 7 variables clave como se muestra en la Tabla 1 y 300,619 registros consolidados.

3.5. Selección de modelos

Se consideraron diferentes modelos de AA, destacando los modelos RF (*Random Forest*) y LR (*Linear Regression*) con base en su capacidad para manejar datos estructurados y su reconocida aplicabilidad en problemas de regresión, como es el caso de la estimación de valores inmobiliarios. El modelo RF fue priorizado por su precisión superior, al combinar múltiples árboles de decisión y reducir el sobreajuste, lo cual se reflejó en un mayor coeficiente de determinación R^2 y menor error porcentual absoluto MAPE. Por otro lado, aunque LR presenta un menor desempeño estadístico, su utilidad práctica radica en su capacidad para identificar tendencias generales y realizar proyecciones. Esta selección responde al análisis previo de estudios similares y a los resultados obtenidos sobre la base de datos del SNIIV, asegurando una evaluación robusta y representativa en el contexto mexicano [12].

3.6. Arquitectura tecnológica

Para el desarrollo del sistema prototipo, se seleccionaron diversas herramientas tecnológicas cuya elección se fundamenta en sus características técnicas y su idoneidad para proyectos colaborativos y de análisis de datos.

Se eligió Python como lenguaje base por su sintaxis clara, amplia comunidad y compatibilidad multiplataforma, además de ofrecer una extensa gama de bibliotecas útiles para el AA y el desarrollo web [18]. Adicionalmente la librería *Scikit-learn* complementa este entorno al proporcionar algoritmos eficientes para clasificación, regresión y preprocesamiento de datos [19]. *Google colab* se incorpora como entorno interactivo que facilita la documentación y ejecución de código en tiempo real, permitiendo integrar visualizaciones y análisis [20]. *Pandas* se emplea como herramienta esencial para la manipulación y análisis de datos estructurados [21].

Django se utiliza como marco de trabajo web robusto para construir la interfaz del sistema, ya que cuenta con una arquitectura que permite fácilmente integración de nuevos programadores, destacando por su sistema de autenticación, panel administrativo y su ORM para la gestión eficiente de bases de datos [22].

La arquitectura establecida por el marco de trabajo de Django aporta lo necesario para llevar a cabo el prototipo, ya que sigue el patrón de diseño *Model-View-Template* (MVT), o Modelo Vista Plantilla (MVP) como se muestra en la Figura 2, y sigue varios principios de diseño que complementan su arquitectura, una de ellas es la arquitectura en capas [22]. Esta arquitectura según [15], consiste en la separación clara entre la capa de presentación, la lógica de negocio y la persistencia de datos para separar las diferentes capas de una aplicación web.

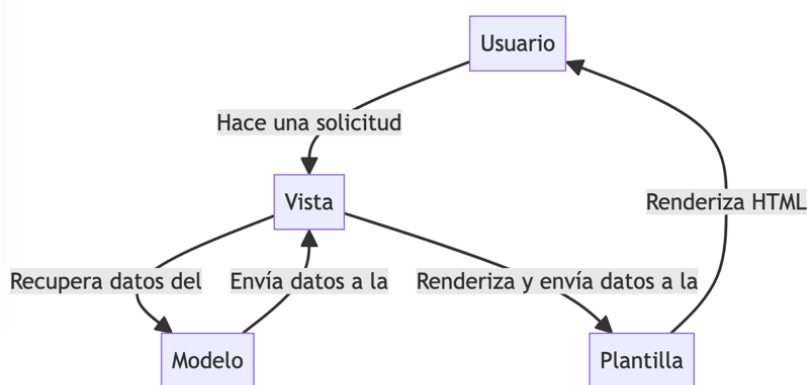


Figura 2. Diseño de la arquitectura del prototipo MVP. Fuente: Adaptado de [22].

Finalmente, Docker se integra como plataforma de virtualización ligera, permitiendo encapsular la aplicación junto con todas sus dependencias, configuraciones y sistema base en un contenedor. Esto asegura que, sin importar el sistema operativo o la configuración del equipo donde se ejecute, el entorno será idéntico, evitando errores causados por diferencias en versiones o configuraciones, lo cual garantiza la reproducibilidad y estabilidad del entorno de desarrollo entre distintos colaboradores [23].

4. Desarrollo del prototipo

El esquema presentado en la Figura 3 describe el funcionamiento general del prototipo para la estimación del valor de viviendas mediante modelos de AA, integrando componentes de procesamiento de datos e interfaz de usuario. El proceso inicia con la captura de datos desde la fuente de información SNIIV, los cuales son posteriormente actualizados y sometidos a un preprocesamiento de datos. Esta información procesada se utiliza para entrenar y ajustar modelos de AA que, una vez evaluados y validados, se almacenan como modelos entrenados los cuales se emplean para la estimación de los valores de las viviendas.

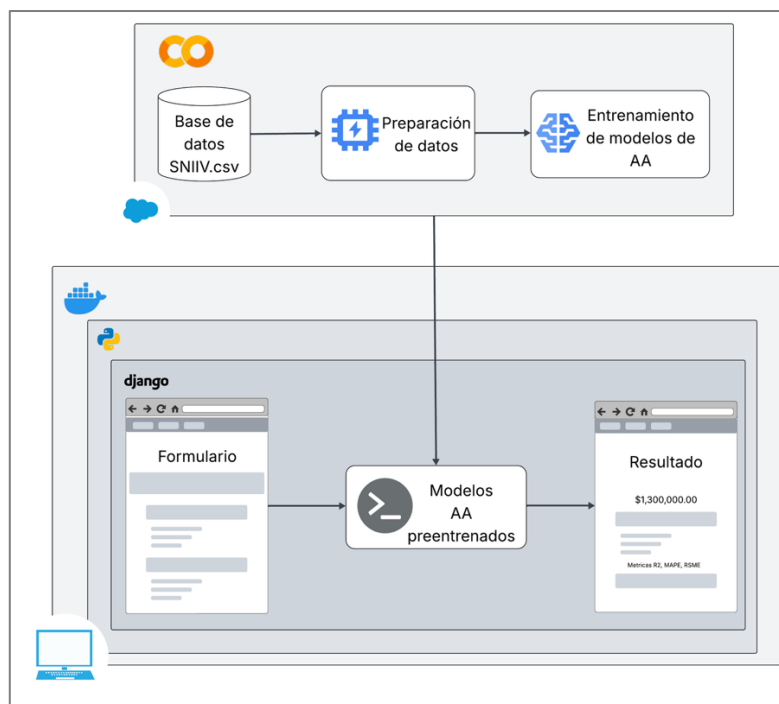


Figura 3. Esquema de aplicación web. Fuente: Elaboración propia.

Es importante resaltar que se emplea el panel de administración de Django para establecer el rumbo de este desarrollo y evitar perder el avance generado hasta este momento, así mismo, se emplea Docker para realizar un contenedor con un prototipo que contenga todas las dependencias. Esto permitirá facilitar su despliegue, escalabilidad y replicabilidad en distintos entornos de desarrollo y producción [23] [24].

Los modelos predictivos son preentrenados en *Google Colab*, empleando como fuente la base de datos del SNIIV. Se utilizan dos algoritmos supervisados de regresión comúnmente aplicados en valuación inmobiliaria: RF y LR, con el objetivo estimar el valor final de viviendas en México.

La interfaz gráfica del sistema permite capturar parámetros clave que influyen en la valuación de viviendas. Empleando datos como el año, clave de la entidad federativa, clave del municipio, la modalidad de la vivienda (nueva o usada), el rango de ingresos promedio de la zona, el rango de superficie construida en metros cuadrados y otras características relevantes de las propiedades. Con base en esta información, el sistema procesará los datos y generará una estimación del valor de la vivienda.

Además del valor estimado, el sistema presenta tres métricas resultado de la evaluación del modelo con el fin de validar la precisión de las predicciones: el coeficiente de determinación (R^2), el error porcentual absoluto medio (MAPE) y la raíz del error cuadrático medio (RMSE), permitiendo así una interpretación integral del nivel de ajuste y del margen de error asociado a la estimación obtenida.

5. Resultados

A continuación, se presentan los resultados de una prueba funcional de los modelos del sistema. Estos resultados se relacionan con las métricas obtenidas durante el entrenamiento de los modelos RF y del modelo combinado RF con LR como se muestra en la Figura 4 y Figura 5. La selección de estos dos enfoques se fundamenta en sus propiedades complementarias: por un lado, el modelo RF demostró una alta precisión al predecir valores dentro

del rango de fechas contenidas en las bases de datos; por otro lado, el modelo LR ofrece la capacidad de proyección a fechas futuras. La combinación de ambos modelos permite aprovechar simultáneamente la precisión del RF en el periodo de entrenamiento y la capacidad predictiva del LR en horizontes temporales futuros, lo cual resulta en un desempeño robusto y generalizable.

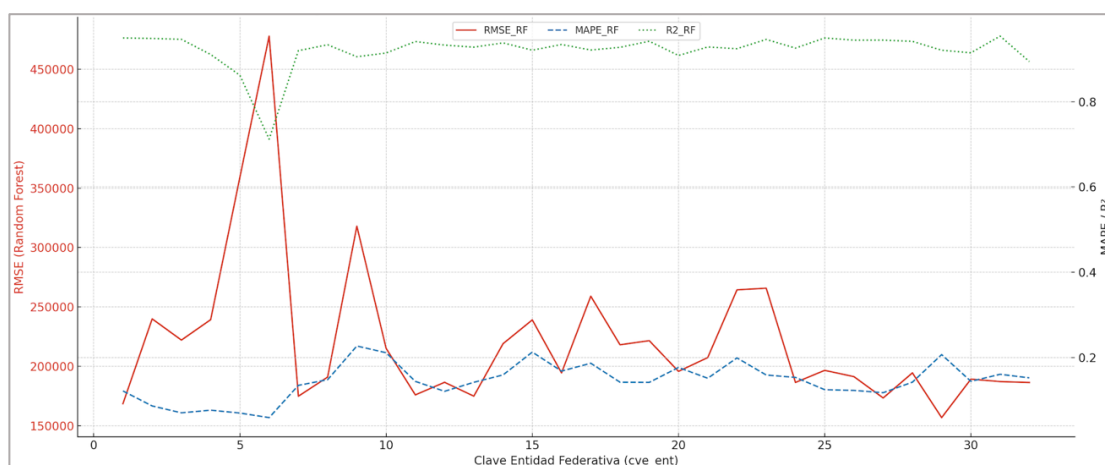


Figura 4. Comparación de métricas del modelo RF por entidad federativa.

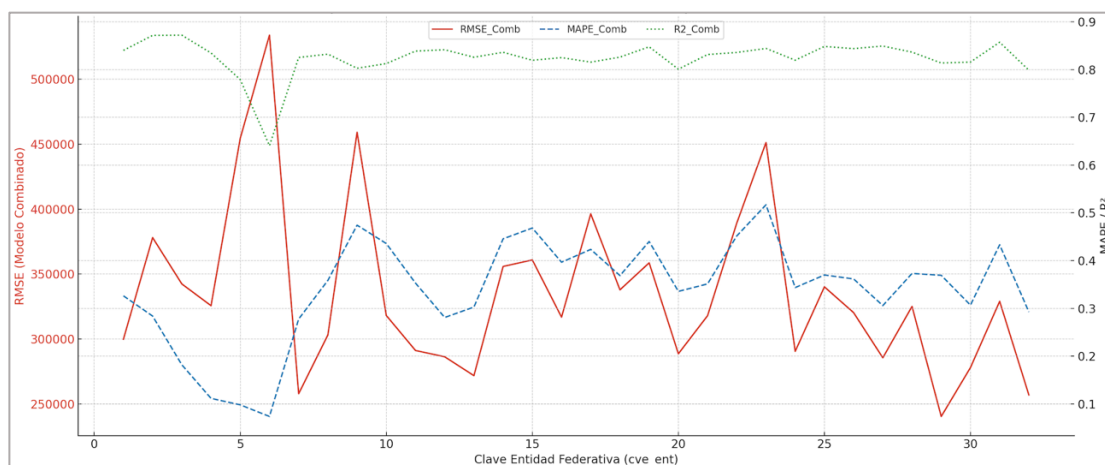


Figura 5. Comparación de métricas del modelo combinando RF y LR por entidad federativa.

Las gráficas generadas presentan una visualización comparativa de tres métricas clave del modelo RF y modelos combinados RF y LR a nivel estatal. En el eje vertical izquierdo se representa RMSE mediante una línea continua de color rojo, mientras que en el eje derecho se muestran de forma simultánea el MAPE es trazado en azul con línea discontinua, y el R^2 indicado mediante una línea punteada de color verde. Esta configuración permite analizar de manera conjunta y diferenciada la variabilidad del desempeño del modelo en función de cada métrica, facilitando la evaluación comparativa entre entidades federativas.

En el análisis de desempeño, las entidades federativas con mejores resultados fueron aquellas como cve_ent 1, 2, 3, 25, 27 y 31, caracterizadas por valores altos de R^2 igual o mayor a 0.94 y bajos de MAPE, lo que indica que los modelos explican gran parte de la variabilidad de los precios y cometen pocos errores relativos. En contraste, las entidades con peor desempeño como cve_ent 6, 29 y 32 presentaron R^2 bajos igual o menor a 0.81 y MAPE elevados, evidenciando menor capacidad predictiva y mayor error en las estimaciones. Por último, el parámetro RMSE determina el desempeño, midiendo el error promedio en las predicciones en las mismas unidades de la variable objetivo.

El desempeño desigual entre entidades federativas puede atribuirse a diversos factores. Entre ellos destacan la heterogeneidad del mercado inmobiliario, que dificulta la identificación de patrones consistentes, la limitada representatividad o sesgo en los datos disponibles, la presencia de valores atípicos que distorsionan las estimaciones y la ausencia de variables relevantes, como pueden ser: infraestructura, actividad turística, niveles de seguridad o políticas locales que influyen significativamente en el valor de los inmuebles.

Como ejercicio final, se realizó una prueba funcional de la interfaz, como se muestra en la Figura 6 y Figura 7 para probar los resultados finales. Por otra parte, en la Tabla 2 se describe el aporte de cada uno de los parámetros considerados en el prototipo [24].

Tabla 2. Significado de parámetro empleados y su aplicación.

Parámetro	Significado	Aplicación
Año	Año en que se realizó la venta o se recopiló el dato	Se usa como variable temporal para capturar tendencias y ciclos del mercado
Entidad federativa	Estado donde se ubica el inmueble	Se codifica como variable categórica para segmentar precios según condiciones macroeconómicas y geográficas
Municipio	División administrativa local	Se usa como variable categórica o geoespacial para capturar mercados específicos
Tipo de vivienda	Condición física	Variable categórica que influye en el precio esperado
Rango de ingresos en la zona	Nivel socioeconómico del comprador típico	Se usa como rango socioeconómico para predecir precios y segmentar datos
Valor de la vivienda	Precio de venta o valor estimado de mercado	Es la variable objetivo (variable dependiente)

Fuente: Elaboración propia con información de base de datos SNIIV [14].

Para realizar la prueba práctica se asignaron los siguientes parámetros: Año=2025, CVE_ENT=25 (Sinaloa), CVE_MUN=1 (Ahome), Modalidad = 2 (Vivienda usada), Rango de ingresos= Medio (de 20,671 a 30,955 pesos mensuales), Vivienda Valor=Media (Superficie construida en promedio: 102m², con 2 baños, cocina, sala, comedor, de 2 a 3 recámaras, cuarto de servicio y 1 a 2 cajones de estacionamiento), obteniéndose un valor de \$1,379,728.16 pesos.

Figura 6. Prueba funcional de la Interfaz gráfica para realizar pruebas prácticas.

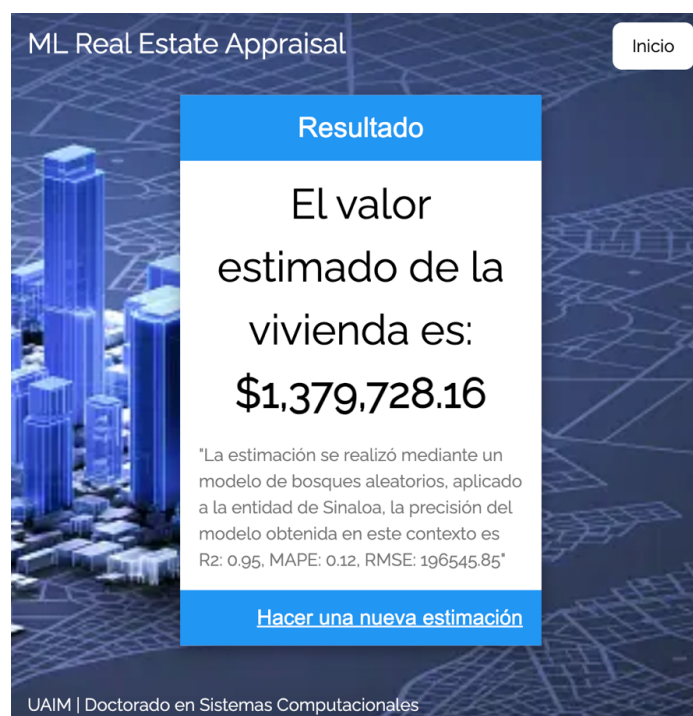


Figura 7. Resultado del prototipo.

Como puede apreciarse en la Figura 7 se informa que "La estimación se realizó mediante un modelo de RF (bosques aleatorios), aplicado a la entidad de Sinaloa, la precisión obtenida del modelo en este contexto es R2: 0.95, MAPE: 0.12, RMSE: 196,545.85", esto indica cual modelo fue seleccionado por la aplicación, así como, la información acerca de la precisión de los resultados.

6. Discusión

Entre las principales fortalezas del prototipo destaca su diseño modular y abierto, el cual integra herramientas ampliamente utilizadas en el ámbito académico y profesional, como Python, Scikit-learn, Django y Docker. Esta arquitectura permite no solo su fácil despliegue y escalabilidad, sino también su adaptación a distintos contextos territoriales o cambios en la estructura de datos. Además, el uso de datos abiertos provenientes del Sistema SNIIV que garantiza transparencia, trazabilidad y congruencia en las predicciones generadas.

No obstante, se identificaron algunas limitaciones relevantes. En primer lugar, si bien los modelos empleados ofrecen alta precisión en promedio, se detectó que existen entidades federativas con bajo ajuste, lo cual puede estar asociado a deficiencias en la representatividad de los datos disponibles, ya que la diversidad de las variables determina en gran medida la capacidad de los modelos de AA, por lo que incluir más variables incrementa la capacidad de capturar relaciones complejas. En segundo lugar, el sistema presenta dificultades inherentes a los algoritmos utilizados, como la limitada interpretabilidad del modelo RF, que impide una comprensión detallada de la lógica oculta en cada predicción.

Finalmente, la falta de un marco normativo vigente que respalde el uso de estas herramientas limita su adopción en procesos oficiales de valuación, a pesar de la creciente evidencia científica sobre su efectividad. En conjunto, los resultados de las pruebas funcionales respaldan la viabilidad técnica del prototipo y abren una vía prometedora hacia la automatización y acceso libre a algoritmos de valuación inmobiliaria en México aplicando AA, que permitan avanzar en la investigación de nuevos métodos de valuación, y en un futuro su implementación, siempre y cuando se superen las limitaciones identificadas y se promueva el desarrollo normativo correspondiente.

7. Conclusiones

El prototipo mostró mejoras en la eficiencia y la precisión de las valuaciones inmobiliarias. Por una parte, la facilidad y rapidez con la que se logran obtener resultados en cualquier entidad de México se supera en relación con el tiempo que se tomaría aplicando métodos convencionales ya que se cuenta con los modelos preentrenados

con información confiable y actualizada de la base de datos del SNIIV que dan respuesta inmediata tras indicar los valores de los parámetros. Por otra parte, la precisión de los modelos de AA presenta un promedio de RMSE: 221,502, MAPE: 0.1444 y R^2 : 0.9213, lo que indica que el modelo presenta un ajuste robusto y una capacidad predictiva elevada.

El desarrollo del prototipo para la estimación automatizada de valores de viviendas en México, mediante la aplicación de modelos de AA, demuestra ser una alternativa técnicamente viable y potencialmente eficaz frente a los métodos tradicionales de valuación inmobiliaria. Así mismo, la implementación de una arquitectura tecnológica basada en herramientas de libre acceso, código abierto y contenedores virtuales favorece la escalabilidad, portabilidad y transparencia del sistema, atributos fundamentales para su adopción en contextos institucionales y académicos.

Finalmente, las aportaciones para los usuarios finales, se traduce en estimaciones confiables y accesibles en segundos, optimizando la toma de decisiones. En el caso de las instituciones financieras y organismos de vivienda, la solución brinda eficiencia operativa, estandarización replicable y transparencia, favoreciendo la integración en procesos crediticios. Para la academia, representa un caso de aplicación práctica de inteligencia artificial con potencial para el desarrollo de investigaciones metodológicas y comparativas.

La plataforma podría implementarse a gran escala mediante infraestructura en la nube, una API pública para integración con sistemas institucionales, actualización automática de bases oficiales e interfaces adaptadas a diferentes perfiles de usuario. Además, la incorporación de mecanismos de retroalimentación fortalecería la capacidad de mejora continua del sistema.

Por último, se considera viable impulsar líneas de investigación orientadas al desarrollo de modelos regionalizados y a la integración de datos no estructurados, como imágenes y textos de anuncios publicados en portales inmobiliarios. Así mismo, resulta relevante analizar el impacto de la aplicación de modelos de AA en el mercado y estudiar el marco normativo que respalde su reconocimiento legal.

8. Referencias

- [1] Sánchez-Juárez, R. (1986). *Historia moderna de la valuación en la República Mexicana*. https://www.gob.mx/cms/uploads/attachment/file/640528/HISTORIA_MODERNA_SHCP.pdf
- [2] Secretaría de Economía. (2015). *NMX-R-081-SCFI-2015. Servicios-Servicios de Valuación-Metodología*. <http://economia-nmx.gob.mx/normas/nmx/2010/nmx-r-081-scfi-2015.pdf>
- [3] 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>
- [4] Verma, P. K., Arya, S., Asbe, C. (2023). Predicting future housing prices: a machine learning approach. *Multidisciplinary Science Journal*, 5, 1-7. <https://doi.org/10.31893/multiscience.2023ss0206>
- [5] 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>
- [6] Guijarro Martínez, F. (2023). Valoración automática de inmuebles residenciales mediante modelos de Machine Learning. *Revista de Estudios Empresariales. Segunda Época*, (2), 27-40. <https://doi.org/10.17561/ree.n2.2023.7823>
- [7] Secretaría de Hacienda y Crédito Público. (2017). *Metodologías de los servicios valuatorios regulados por INDAABIN*. <https://www.gob.mx/indaabin/documentos/metodologias-de-caracter-tecnico-24208>
- [8] 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>
- [9] Mody, P., 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>
- [10] 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>

- [11]Hernández Ruiz, E. A. (2023). *Valuación inmobiliaria*. Trillas.
- [12]Espinoza Garza, F., Martínez Ramírez, Y., Ramírez-Noriega, A., Álvarez Sánchez, I. N. (2024). Una revisión sistemática de la literatura sobre la precisión de modelos de aprendizaje automático aplicados a la tasación de bienes raíces. *Revista de Investigación en Tecnologías de la Información (RITI)*, 12 (28), 4-16. <https://doi.org/10.36825/RITI.12.28.002>
- [13]Deppner, J., von Ahlefeldt-Dehn, B., Beracha, E., Schaefer, W. (2023). Boosting the accuracy of commercial real estate appraisals: an interpretable machine learning approach. *The journal of real estate finance and economics*, 71, 314-351. <https://doi.org/10.1007/s11146-023-09944-1>
- [14]Secretaría de Desarrollo Agrario, Territorial y Urbano. (2025). *Financiamientos a la vivienda. Sistema Nacional de Información e Indicadores de Vivienda (SNIIV)*. https://sniiv.sedatu.gob.mx/Reporte/Datos_abiertos
- [15]Pressman, R. S. (2005). *Software engineering: a practitioner's approach*. Palgrave Macmillan.
- [16]Gunes, T. (2024). Model agnostic interpretable machine learning for residential property valuation. *Survey Review*, 56 (399), 525-540. <https://doi.org/10.1080/00396265.2023.2293366>
- [17]Hong, J., Kim, W. S. (2022). Combination of machine learning-based automatic valuation models for residential properties in South Korea. *International Journal of Strategic Property Management*, 26 (5), 362-384. <https://doi.org/10.3846/ijspm.2022.17909>
- [18]Python Software Foundation. (2025). *Lenguaje de programación Python*. <https://www.python.org/>
- [19]Scikit-learn developers. (2025). *Scikit-learn: Machine Learning in Python*. <https://scikit-learn.org/>
- [20]Google. (2025). *Google Colaboratory* (Versión 1.0). <https://colab.research.google.com/>
- [21]Pandas Development Team. (2025). *Pandas*. <https://pandas.pydata.org/>
- [22]Django Software Foundation. (2025). *Django* (Versión 5.0). <https://www.djangoproject.com/>
- [23]Docker, Inc. (2025). *Docker Desktop* (Versión 4.29.0). <https://www.docker.com/>
- [24]Parsons, D. (2009). *Desarrollo de aplicaciones Web dinámicas con XML y Java*. Anaya multimedia.