



Facultad de Ciencias Económicas y Empresariales
ICADE

Evaluación y Aplicación de Modelos Automatizados de Valoración (AVMs)

Autor: Miriam Portillo Quesada
Director: Leandro Sergio Escobar Torres

MADRID | Marzo, 2024

RESUMEN

El presente Trabajo de Fin de Grado trata de explicar y analizar el impacto del Big Data y la Inteligencia Artificial en la valoración de inmuebles mediante los Modelos de Valoración Automatizada (AVM). A lo largo del mismo, se realiza una revisión de los métodos tradicionales que sirven como base para entender los fundamentos de estas nuevas metodologías y los agentes involucrados en las valoraciones. Adicionalmente, se realizará una revisión de la literatura para entender en profundidad el funcionamiento de estos algoritmos, tecnologías y técnicas, así como su impacto y aplicación en el sector inmobiliario. Además, se realiza una propuesta de cómo será el futuro proceso de aplicación de estas nuevas tecnologías en el sector, junto con las perspectivas y desafíos que las deparan. Se concluirá el trabajo con un caso práctico del Distrito de Chamartín, siendo uno de los 21 distritos que forman la ciudad de Madrid, caracterizado por su heterogeneidad en la composición social y económica de sus barrios, afectando a la valoración de los bienes raíces. En este caso práctico, se aplicarán las metodologías de precios hedónicos, K-vecinos cercanos y la combinación de ambas, debido a que son los únicos métodos permitidos en España de forma oficial.

PALABRAS CLAVE: Valoración, AVM, metodologías, estimación, precio, KNN, regresión.

ABSTRACT

This Final Degree Project aims to explain and analyse the impact of Big Data and Artificial Intelligence on property valuation through Automated Valuation Models (AVM). Throughout the thesis, a review of traditional methods is conducted as a foundation to understand the principles of these new methodologies and the stakeholders involved in valuations. Additionally, a review of the literature will be carried out to understand in depth the functioning of these algorithms, technologies, and techniques, as well as their impact and application in the real state sector. Moreover, the thesis proposes how the future process of applying these new technologies in the sector will look like, along with the perspectives and challenges they present. The project will conclude with a practical case study of the Chamartín District, one of Madrid's 21 districts, characterized by its heterogeneity in the social and economic composition of its neighborhoods, affecting real estate valuations. In this case study, the methodologies of hedonic pricing, K-nearest neighbors, and the combination of both will be applied, as they are the only officially permitted methods in Spain.

KEY WORDS: Valuation, AVM, methodologies, estimation, price, KNN, regression.

ÍNDICE DE CONTENIDO

LISTA DE ACRÓNIMOS	7
1. INTRODUCCIÓN	8
1.1 Propósito general	8
1.2 Objetivos.....	8
1.3 Metodología.....	9
1.4 Contextualización del tema.....	9
1.5 Justificación del tema.....	11
2. VALORACIÓN INMOBILIARIA TRADICIONAL	13
2.1 Actores involucrados en la tasación.....	13
2.2 Normativa y Regulación en España.....	15
2.3 Proceso de Valoración Tradicional.....	17
2.4 Métodos de Valoración tradicional.....	18
2.4.1 Enfoque de comparación	19
2.4.2 Enfoque basado en el Ingreso.....	19
2.4.3 Enfoque de coste	20
2.4.4 Otras consideraciones.....	20
3. VALORACIÓN INMOBILIARIA MEDIANTE EL BIG DATA Y LA INTELIGENCIA ARTIFICIAL	21
3.1 Concepto y Fundamento	21
3.2 Usos de los AVM.....	22
3.3 Legislación AVMs en España.....	23
3.4 Proceso de Valoración Automatizada.....	25
3.4.1 Negociación previa al proceso de valoración	25
3.4.2 Investigación	25
3.4.3 Tratamiento e interpretación de datos	28
3.4.5 Informes sobre el proceso posterior a la evaluación	31
3.5 Nuevos Métodos de Valoración.....	31
3.5.1 Modelos de precios hedónico	32
3.5.2 Método de comparables.....	35
3.5.3 Inteligencia Artificial (IA): Redes Neuronales.....	37
3.5.4 Otros Modelos de Valoración.....	40
3.6 Perspectiva futuras y desafíos de la valoración inmobiliaria.....	41

4. CASO PRÁCTICO DE VALORACIÓN DE INMUEBLES CON AVMs.....	44
4.1 Selección del área de estudio	44
4.2 Objetivos y Metodología.....	45
4.3 Análisis preliminar de los datos	45
4.4 Aplicación de los AVM.....	47
4.4.1 Modelo Hedónico.....	47
4.4.2 K-Vecinos cercanos.....	53
4.4.3 Modelo Mixto.....	55
5. CONCLUSIÓN.....	58
6. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado.....	60
7. REFERENCIAS BIBLIOGRÁFICAS	61
8. ANEXO	65
8.1 ANEXO I: Variables del Caso Práctico	65
8.2 ANEXO II: Código de R Studio del Caso Práctico.....	67

ÍNDICE DE FIGURAS

Figura 1: Fases del proceso de valoración.....	17
Figura 2: Resumen del proceso de valoración.....	18
Figura 3: Construcción de un modelo AVM.....	28
Figura 4: Representación del método k-vecinos	35
Figura 5 :Elementos de una red neuronal.....	38
Figura 6: Distrito de Chamartín por barrios	44
Figura 7: Diagrama de Caja y Bigotes por Barrio.....	46
Figura 8: Matriz de Correlaciones.....	47
Figura 9: Histograma de errores Stepmodel.....	53
Figura 10: Valores óptimos de k	54
Figura 11: Histograma de los errores en KNN.....	55
Figura 12: Métricas de Precisión Modelo Mixto.....	56

ÍNDICE DE TABLAS

Tabla 1: Usos de los AVM en España.....	23
Tabla 2: Resumen del Modelo 1.....	49
Tabla 3: Resumen del Modelo 2.....	50
Tabla 4: Resumen del Modelo 3.....	51
Tabla 5: Resumen de los modelos hedónicos.....	52
Tabla 6: Modelo por Pasos para la selección de variables.....	52
Tabla 7: Validación Cruzada KNN.....	54
Tabla 8: Resumen del Modelo Mixto.....	55
Tabla 9: Resumen de las variables.....	65

LISTA DE ACRÓNIMOS

AVM	Modelos de Valoración Automática
IA	Inteligencia Artificial
P2P	Peer to Peer
IoT	Internet of Things
FII	Fondos de Inversión Inmobiliario
SOCIMIs	Sociedades Anónimas Cotizadas de Inversión en el Mercado Inmobiliario
EVS	Estándares Europeos de Valoración
IVS	Estándares Internacionales de Valoración
TEGoVa	Asociación de Tasadores Europeos
RICS	Royal Institution of Chartered Surveyors
PIB	Producto Interior Bruto
FCD	Flujo de Caja Descontado
SIG	Sistemas de Información Geográfica
MLS	Multiple Listing Service
AEV	Asociación Española de Análisis del valor
AAIO	International Association of Assessing Officers
RICS	Institution of Chartered Surveyors
KNN	K-Vecinos más cercanos
RNA	Redes Neuronales
MLP	Redes neuronales de propagación o Multicapa
CNN	Redes convolucionales
AG	Algoritmo Genético
LOPD	Ley Orgánica de Protección de Datos
ME	Error medio
RMSE	Raíz Cuadrada Media
MAPE	Error absoluto porcentual Medio
MAE	Error Medio absoluto

1. INTRODUCCIÓN

1.1 Propósito general

El objeto del presente trabajo consiste en el estudio de la valoración de bienes inmuebles, partiendo de una revisión de los métodos tradicionales para introducir las bases de los denominados Modelos de Valoración Automática (AVM). En la actualidad, los fenómenos del Big Data y la Inteligencia Artificial están revolucionando las economías a nivel mundial cobrando una mayor importancia en el sector público y privado. Este trabajo pretende revisar la literatura para comprender qué futuro les depara a estas nuevas tecnologías en el sector inmobiliario, especialmente en la estimación del valor. Además, este estudio estará acompañado de una aplicación práctica de la teoría a través de un caso de valoración empleando los nuevos métodos de valoración impulsados por estos fenómenos.

1.2 Objetivos

El presente Trabajo de fin de Grado pretende abordar un conjunto de objetivos específicos que se enfocan en la correcta valoración de los bienes inmuebles y la comprensión de las diversas metodologías empleadas, tanto tradicionales como automatizadas. Para ello, se establecen cinco objetivos con el propósito de remarcar la relevancia de los nuevos enfoques matemáticos y su futuro proceso de valoración en el sector.

Los objetivos de este trabajo se pueden dividir en los siguientes cinco apartados:

1. Se pretenderá reflejar la importancia de una correcta tasación de un inmueble y los agentes afectados por éstas.
2. Se buscará aportar la base teórica de los métodos tradicionales, para facilitar la comprensión de en qué consiste un proceso de tasación, las metodologías en que se basan y la legislación a la que se rige.
3. Se examinará cómo están empezando a desarrollarse los AVM y sus correspondientes desafíos y perspectivas.
4. Se realizará una revisión de la literatura para conocer de forma precisa el concepto y funcionamiento de los nuevos métodos de valoración.
5. Se mostrará mediante un caso práctico el funcionamiento de los AVM permitidos

en España.

1.3 Metodología

La metodología del trabajo consiste en la revisión de la literatura, aportando una base teórica necesaria para entender tanto el funcionamiento del proceso de valoración como los algoritmos empleados en los AVM. Se examinarán los procesos, metodologías y regulaciones tradicionales empleadas por las Sociedades de Tasación, para poder seguir utilizando la misma estructura en el futuro proceso de tasación con ayuda de los AVM. Asimismo, se explorará el fenómeno del Big Data y la Inteligencia Artificial para presentar las distintas metodologías que han sido instauradas en la valoración de inmuebles. Para la revisión de la literatura, se utilizará mayoritariamente documentos encontrados en *Google Scholar*, *Dialnet* y *ResearchGate* siendo buscadores fiables de publicaciones académicas.

Adicionalmente, se realizará un modelo AVM con la ayuda de R Studio como herramienta principal en el desarrollo del algoritmo y Microsoft Excel para la limpieza y recopilación de los datos. En este caso práctico, la base de datos empleada en el modelo ha sido extraída de Idealista, mostrando los inmuebles en venta más recientes a la fecha actual.

1.4 Contextualización del tema

En la era digital actual, la proliferación de datos generados por el uso de smartphones y redes sociales ha proporcionado un crecimiento exponencial de la cantidad de información disponible (Mckinsey & Company, 2011). Este fenómeno ha convertido al Big Data en un elemento fundamental en diversas áreas económicas a nivel mundial, destacando su importancia en la innovación y eficiencia operativa.

En este contexto, el sector inmobiliario no se ha quedado atrás gracias al surgimiento del *Proptech*, una fusión entre la tecnología y el sector inmobiliario, que se presenta como una solución ante los desafíos existentes de las valoraciones inmobiliarias. Este concepto comenzó a definirse en España a comienzos del siglo XXI, marcado por la aparición de plataformas y sitios web especializados como Idealista, Rentalia y Pisos.com. La importancia del *Proptech* es tan destacada que, en 2021, se reconocieron más de 250 empresas operando bajo esta denominación en el país (Molina, 2021).

Esta tecnología ha transformado radicalmente el sector inmobiliario, introduciendo innovaciones que abarcan desde la comercialización hasta la gestión de propiedades. Se utilizan siete categorías para entender las nuevas integrantes en el mercado inmobiliario (PropTech, s.f.):

1. **Las plataformas de mercado** han alterado el modelo de negocio tradicional de las agencias inmobiliarias, al ofrecer espacios como Fotocasa y Pisos.com, que facilitan la compra, venta y promoción de propiedades. Estas plataformas funcionan como mercados digitales, permitiendo al usuario buscar y anunciar inmuebles para venta o alquiler de forma remota y con interacción directa entre las partes involucradas.
2. **Los modelos Peer to Peer (P2P)** han democratizado el alquiler y venta directa entre particulares sin necesidad de intermediarios. Un ejemplo es Airbnb, donde los propietarios pueden alquilar sus espacios directamente a los huéspedes.
3. **El Big Data** permite el uso de grandes conjuntos de datos para analizar tendencias, valoraciones y otros aspectos relevantes en el sector. Esto incluye el análisis de precios, predicciones de mercado, optimización de inversiones y personalización de la oferta de propiedades según las preferencias del cliente.
4. **Internet of Things (IoT)** consiste en la automatización del hogar mediante el uso de esta tecnología para conectar y automatizar los sistemas y dispositivos de una propiedad, como la iluminación, calefacción y sistemas de seguridad.
5. **La gestión inmobiliaria** se beneficia de un software diseñado para simplificar tareas como el seguimiento de alquileres, el mantenimiento, la comunicación con los inquilinos y la administración financiera.
6. **Las plataformas de Inversión e Hipotecas** permiten a los usuarios acceder a nuevas formas de financiación e inversión relacionadas con el sector inmobiliario como el crowdfunding inmobiliario, inversión en propiedades, financiación de proyectos inmobiliarios y la obtención de hipotecas.
7. **La Realidad Virtual** permite a los usuarios experimentar inmersivamente propiedades a distancia, ofreciendo visitas virtuales que ayudan a la decisión de compra o alquiler sin necesidad de estar físicamente presentes.

En conclusión, la revolución tecnológica se ha introducido en todas las verticales del sector inmobiliario mediante el *Proptech*. El presente trabajo se centra en las herramientas

empleadas en la valoración de inmuebles mediante el Big Data y la Inteligencia Artificial debido a su relevancia en el sector. Estas tecnologías están en continuo desarrollo para abordar el mayor reto presente en el sector inmobiliario, la correcta estimación de precios de los bienes raíces.

1.5 Justificación del tema

En 2013, antes del auge actual del Big data y otras tecnologías, en España se realizaron un total de 661.000 tasaciones (San José, 2020). Fueron realizadas por 52 sociedades de tasación homologadas en ese período, generando un volumen de negocio de 265 millones de euros y abarcando una superficie de 517.000 hectáreas de terreno (San José, 2020). Dentro del sector inmobiliario, resalta el mercado de la vivienda como su principal componente. Históricamente, la vivienda inmobiliaria ha tenido un papel notable en la literatura, mientras que otros segmentos han sido marginales. De las valoraciones mencionadas en 2013, un 59% fueron de viviendas individuales, ya sean parte de edificios o unifamiliares (Núñez Tables, Rey Carmona, & Caridad y Ocerín, 2017)

Cabe destacar que es una industria intensiva de mano de obra, con un registro de 74.000 tasadores solo en Estados Unidos en 2017 (Kok, Koponen, & Martínez-Barbosa, 2017). Esta necesidad de recursos se atribuye a las metodologías utilizadas en la valoración de inmuebles, debido a que requieren un alto nivel de especialización y una dedicación horaria considerable (Guijarro Martínez, 2023). Sin embargo, eso no garantiza que las valoraciones logren un alto grado de precisión en algunas situaciones. Según lo indicado en el estudio de Kok *et al* (2017) los tasadores pueden incurrir en un error relativo del 12%, pudiendo variar dependiendo del grado de madurez y transparencia del país. Además, es una industria que, en países como Reino Unido, Australia y Sudáfrica sus tasadores están caracterizados por estar próximos a la edad de jubilación (San José, 2020).

En el contexto español, una correcta valoración inmobiliaria es imprescindible debido a que la composición de la riqueza de los hogares muestra una marcada preferencia por los activos inmobiliarios. Según el informe sobre la distribución de riqueza del Banco Central Europeo, un 70% de la riqueza de los hogares españoles se concentra en activos no financieros, principalmente inmuebles. Este porcentaje supera significativamente al de países como Francia o Alemania, que presentan una inversión entre el 55 y 60 por ciento en este tipo de activos, situando a España como el país de la zona euro más ligado al sector

inmobiliario (Bondt, Gieseck, & Tujula, 2020).

Esta preferencia ha tenido consecuencias históricas, especialmente visibles durante la crisis del 2008. La sobrevaloración de las viviendas y el *boom* del crédito hipotecario, seguido por el estallido de la burbuja inmobiliaria en 2007, llevó a los hogares españoles a enfrentarse a un nivel de endeudamiento que alcanzó el 130% sobre el total de los activos, en comparación con el promedio del 98% en la zona euro (Bondt, Gieseck, & Tujula, 2020). Tras la crisis financiera, se buscó una mayor transparencia y eficiencia en las valoraciones inmobiliarias para evitar sobrevaloraciones y promover prácticas de concesión de préstamos más conservadoras. Esta situación, impulsó a las sociedades de tasación en España a intensificar el uso de AVMs como medida de refuerzo en la precisión de las valoraciones.

En este contexto, comprender la definición, la regulación y las capacidades de los diferentes métodos de valoración inmobiliaria se convierte en un componente esencial en los conocimientos de un tasador, independientemente de que el proceso se lleve a cabo de forma manual o automática.

2. VALORACIÓN INMOBILIARIA TRADICIONAL

2.1 Actores involucrados en la tasación

Los ciudadanos con frecuencia se encuentran involucrados de una forma u otra en procesos de valoración inmobiliaria. Pueden surgir de situaciones cotidianas, como la compra de una vivienda o el pago de impuestos, así como en el ámbito laboral al realizar inversiones inmobiliarias o gestionar los balances empresariales. En 2023, el sector inmobiliario generó 55.310 millones de euros, de los cuales 35.669 millones correspondían a viviendas (INE, s.f.). La correcta valoración de inmuebles es un proceso complejo y repercute en diversos individuos, implicando la participación de diversos actores, cada uno de los cuales aporta conocimientos especializados y perspectivas al proceso de tasación.

Entidades Financieras

El sector financiero se ve influenciado por la valoración inmobiliaria principalmente en las áreas de préstamos hipotecarios, gestión de riesgos, seguros e inversiones inmobiliarias.

Respecto a los créditos hipotecarios, las entidades crediticias requieren una tasación por una empresa especializada siendo un requisito impuesto por el Banco de España para garantizar la adecuada concesión y devolución de este tipo de préstamos. Esta valoración determina la cantidad de capital que se prestará, debido a que la propiedad actúa como colateral, convirtiéndose en un elemento para la gestión de riesgos de la entidad. En caso de incumplimiento, el inmueble se emplea para la mitigación de las pérdidas ocasionadas.

En el ámbito de los seguros, la valoración inmobiliaria influye en la fijación de primas y en la evaluación de los riesgos asociados a la propiedad (Fabra Garcés & Sevillano Rubio, 2000). Esta tasación establece el límite máximo de cobertura de la póliza, asegurando una protección adecuada al valor del inmueble. Un ejemplo de ello son aquellos inmuebles en áreas propensas a desastres naturales, los cuales requieren primas más altas debido a un mayor riesgo de pérdida.

Desde la perspectiva de la inversión, obtener una tasación precisa del valor de mercado

es un componente esencial. Esta valoración ayuda a determinar el precio justo de compra y si el valor subyacente de la propiedad es consistente con las expectativas de rendimiento de dicha inversión. Las opciones de inversión varían desde transacciones directas en propiedades hasta la participación en fondos de inversión inmobiliario (FII) y Sociedades Anónimas Cotizadas de Inversión en el Mercado Inmobiliario (SOCIMIs).

Administración Pública

Las entidades gubernamentales tienen la responsabilidad de supervisar, regular y utilizar las valoraciones inmobiliarias como base de toma de decisiones para lograr los objetivos fiscales, legales y de planificación urbana.

El sistema tributario, específicamente el área destinada a la recaudación se basa en la imposición de ciertos impuestos sobre los bienes inmuebles. Estos pueden ser directos, como el IVA inmobiliario, el impuesto sobre transmisiones patrimoniales y el impuesto sobre bienes inmuebles, o indirectos, como el impuesto sobre patrimonio, IRPF y el impuesto de sucesiones (Fabra Garcés & Sevillano Rubio, 2000).

Además, las tasaciones son requeridas en situaciones legales para resolver disputas sobre las propiedades. Desde la antigüedad, el derecho a la propiedad y a la posesión han sido fuente de innumerables litigios debido a discrepancias y disconformidades en el valor del inmueble (Antón-Merino & Garrido-Villén, 2016). Existen diversos casos de disputas sobre la propiedad, siendo las más comunes la división de bienes en un divorcio, herencias y sucesiones, daños a la propiedad y límites territoriales.

En la planificación urbana, las tasaciones se realizan antes de la aprobación de grandes proyectos, como la construcción de viviendas residenciales, infraestructura o instalaciones comerciales, con el propósito de comprender cómo estos proyectos en desarrollo afectarán en el valor de las propiedades circulares (Becerra Lois, Pons García, & Ruíz Domínguez, 2020). Este uso multifacético de las tasaciones subraya la importancia para las políticas gubernamentales, la legislación y el desarrollo urbano estratégico.

Ámbito Particular

Los propietarios son los encargados iniciar el proceso de valoración, aportando

información precisa del inmueble y colaborando estrechamente con el tasador. Los informes resultantes tienen diferentes finalidades, tales como fijar un precio de venta justo, utilizarlo como base de negociaciones en divorcios o divisiones de herencia, facilitar la obtención de financiamiento y asegurar una correcta tributación sobre la propiedad.

Ámbito Empresarial

En el contexto empresarial, la importancia de la valoración inmobiliaria es innegable. Las empresas frecuentemente recurren a estas valoraciones para optimizar sus operaciones, ya sea para integrar propiedades mediante fusiones o adquisiciones, o para respaldar la expansión y consolidación de la compañía (Fabra Garcés & Sevillano Rubio, 2000). Estos activos no solo pueden ser fundamentales para el crecimiento empresarial, sino también influir en el balance general y en las obligaciones fiscales. Además, la correcta valoración de inmuebles contribuye significativamente a la planificación estratégica, gestión de riesgos, y evaluación de inversiones, destacando su importancia transversal en diversas áreas de negocio.

2.2 Normativa y Regulación en España

La normativa que rige la valoración de un inmueble depende de su finalidad específica. Si tiene fines urbanísticos, se aplica la Ley del Suelo, concretamente el artículo 34 del Real Decreto Legislativo 7/2015. En este artículo se establece que las valoraciones del suelo, instalaciones, construcciones y derechos relacionados se rige por esta Ley en situaciones de verificación de operaciones de reparto de beneficios y cargas, la fijación del justiprecio en expropiación, la determinación del precio en ventas o sustituciones forzosas, y la evaluación de la responsabilidad patrimonial de la Administración Pública.

Para valoraciones de un inmueble registrado en el catastro, se utiliza el Real Decreto Legislativo 1/2004, que aprueba el texto refundido de la Ley de Catastro Inmobiliario. Este marco legal establece las pautas y normativas para llevar a cabo la valoración de inmuebles en el contexto catastral.

En el ámbito financiero, la regulación se basa en la Orden ECO/805/2003 para la valoración de inmuebles involucrados en la concesión de préstamos hipotecarios. Esta Orden establece las normas de valoración del inmueble y los derechos de compradores y

vendedores financieros. Estas normas se fundamentan en diez principios que se deben aplicar en las valoraciones y cuatro métodos de valoración inmobiliaria, además de definir cómo deben estructurarse los informes y certificados de tasación. Según el artículo 2, estas directrices se aplican en contextos como garantías hipotecarias, cobertura de provisiones técnicas de entidades aseguradoras, determinación del patrimonio de los FII, y los Fondo de Pensiones.

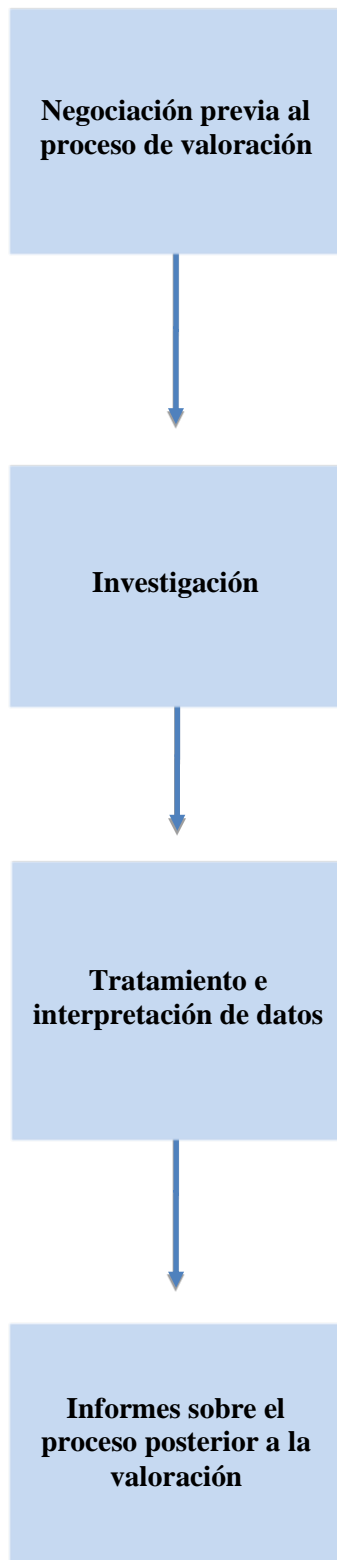
Las instituciones autorizadas para llevar a cabo préstamos y emitir títulos hipotecarios son los bancos, las cajas de ahorro, la Confederación Española de Cajas de ahorros, las entidades oficiales de crédito, las cooperativas de crédito y los establecimientos dedicados a la financiación de crédito. Estos préstamos deben contar con una garantía, siendo obligatorio que esa garantía sea una hipoteca inmobiliaria con rango de primera sobre la propiedad completa. El Real Decreto 716/2009 regula las condiciones de estos préstamos, limitando la financiación al 60% del valor de tasación, o hasta el 80% para propósitos la construcción, rehabilitación o adquisición de viviendas. Además, exige a las entidades financieras mantener registros contables precisos de los préstamos y créditos asociados a bonos hipotecarios.

En la Orden del Ministerio de Economía y Hacienda de 1994, designa a las Sociedades Tasadoras como responsables de las valoraciones inmobiliarias para créditos hipotecarios, provisiones técnicas de las entidades aseguradoras y para la determinación de patrimonio en FII, y Fondo de Pensiones. Además, si la valoración tiene una finalidad diferente a las mencionadas anteriormente, se permite la aplicación de métodos de valoración de cualquiera de ellas, debidamente justificado. Sin embargo, la normativa más utilizada para otras valoraciones generalmente es la Orden ECO/805/2003.

Por último, a nivel europeo e internacional existen Los Estándares Europeos de Valoración (EVS) y Los Estándares Internacionales de Valoración (IVS) que son recomendaciones diseñadas para proporcionar una guía uniforme sobre la valoración de activos, con especial atención a la valoración de bienes inmuebles. El EVS fue desarrollado por la Asociación de Tasadores Europeos (TEGoVa) para crear un marco común de valoración de propiedades en Europa y promover la coherencia y transparencia en las prácticas de valoración. Por otro lado, El IVS fue creado por el International Valuation Standards Council para establecer principios comunes y estandarizar las prácticas de valoración a nivel global.

2.3 Proceso de Valoración Tradicional

Figura 1: Fases del proceso de valoración



El tasador y el cliente acuerdan los **términos del contrato** de forma escrita, antes de empezar la valoración inmobiliaria. El objetivo es que ambas partes conozcan las obligaciones y derechos que asumen.

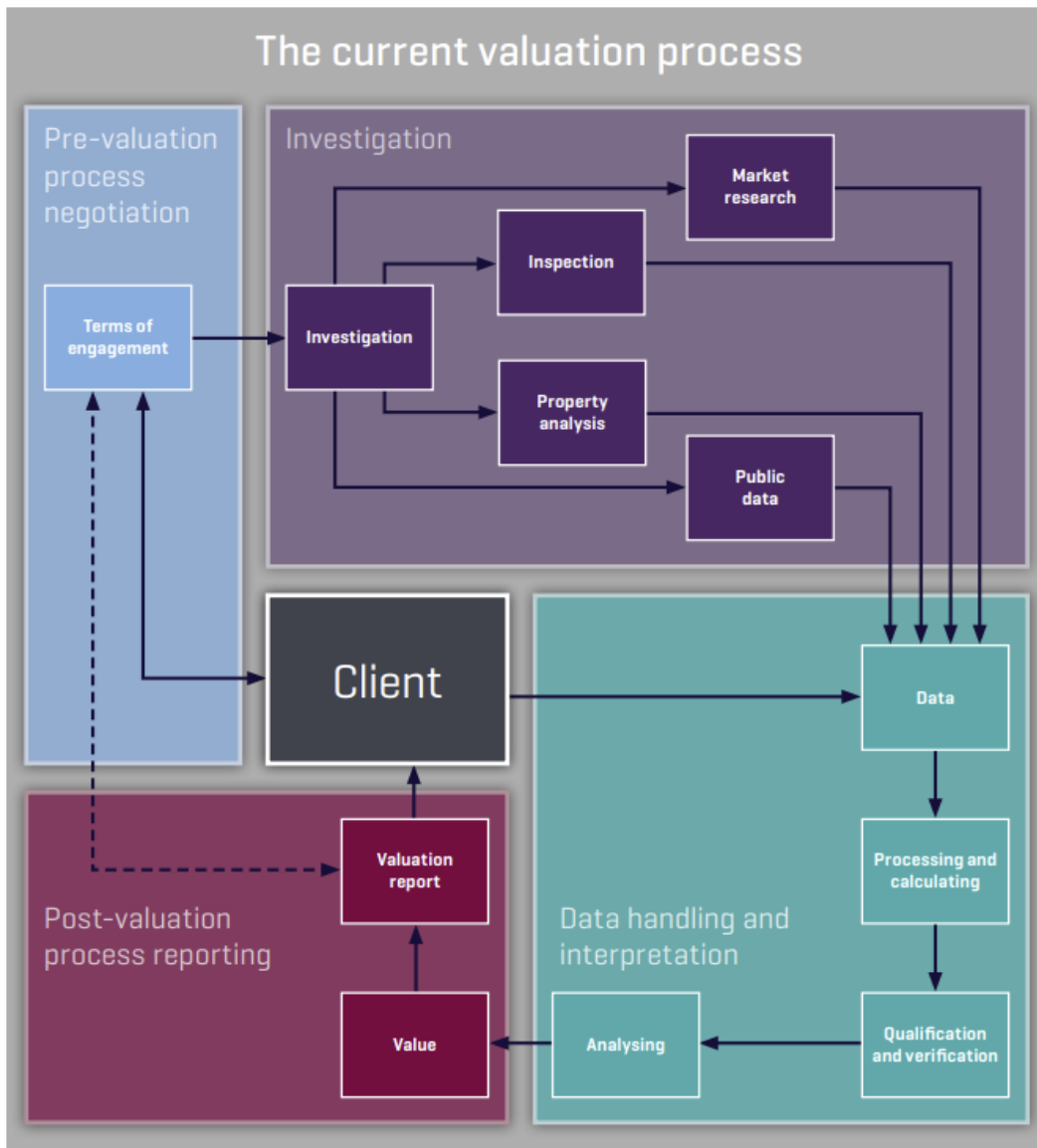
Este paso, consiste en una **investigación** de manera formal o sistemática de la propiedad. Hay cuatro aspectos a considerar en la examinación: inspección externa o interna de la propiedad, análisis de la propiedad, estudio de mercado y uso de bases de datos públicas.

Las etapas para el **tratamiento e interpretación de los datos** incluyen la verificación, procesamiento y análisis de los datos. En la verificación, se identifican y aseguran la calidad de los datos relevantes; en el procesamiento, se realiza la transformación y limpieza de los datos para su posterior aplicación; y en el análisis se interpretan los resultados obtenidos, extrayendo conclusiones que se comunicarán posteriormente al cliente.

Finalmente, en la valoración inmobiliaria, el resultado debe ser un valor único que deberá representarse de forma clara y argumentada en el **informe**. Este informe ofrece una confirmación escrita del valor de la propiedad siendo respaldado por los distintos enfoques y metodologías aplicados.

Fuente: Elaboración propia

Figura 2: Resumen del proceso de valoración



Fuente: (Scheurwater, 2017)

2.4 Métodos de Valoración tradicional

En la valoración se deben tener en cuenta los métodos de valoración correspondientes y apropiados a cada situación. Se pueden usar uno o más enfoques de valoración con el objetivo de alcanzar el valor justo de un inmueble, especialmente en situaciones donde la disponibilidad de datos es limitada. Además, en casos en los que los enfoques ofrecen estimaciones divergentes, el tasador debe comprender las razones subyacentes, y reconsiderar los métodos empleados para obtener la estimación más fiable. Sin embargo, los valuadores no están obligados a emplear más de un método de valoración si confían

en la precisión de uno en concreto.

Según el Consejo de normas internacionales de valoración existen tres enfoques de valoración principales, basándose en principios económicos de equilibrio del precio, anticipación de beneficios o sustitución (2022). A continuación, se explicará en detalle los distintos enfoques y métodos según la norma vigente en la IVS 105 (IVSC, 2022)

2.4.1 Enfoque de comparación

El enfoque de comparación es de los métodos más utilizados para determinar el valor de una propiedad. Este método consiste en una estimación del valor al comparar un activo con otros idénticos o comparables, para los cuales exista información de precios. La aplicación de este enfoque se recomienda en contextos donde el inmueble ha experimentado una reciente transacción, está activo en el mercado público o cuando se observan recientes y frecuentes transacciones de activos similares.

Dentro del enfoque comparativo, el tasador dispone de dos métodos: la comparación directa de transacciones y la comparación con empresas cotizadas. El primero se basa en datos de venta de propiedades similares, mientras que el segundo utiliza indicadores financieros de empresas cotizadas con activos comparables. Aunque ambos métodos se centran en la comparabilidad, la comparación con empresas cotizadas se beneficia de la disponibilidad de métricas en el momento de la tasación e información pública conforme a las normas contables aceptadas.

2.4.2 Enfoque basado en el Ingreso

El enfoque basado en el Ingreso calcula el valor de un inmueble convirtiendo futuros flujos de caja en un valor presente. Este método valora la propiedad basándose en los ingresos futuros, los flujos de efectivo, o el ahorro en costes que el activo puede generar, ofreciendo así una perspectiva financiera del valor del inmueble.

Aunque existen diversas formas de implementar el enfoque basado en ingresos, se destacan métodos como el Flujo de Caja Descontado (FCD) que calcula el valor presente descontando los flujos de caja futuros a la fecha de valoración. Este método en algunos casos se calcula utilizando únicamente el valor terminal sin un periodo determinado, conociéndose como método de capitalización de rentas. Alternativamente, en activos de largo plazo o de vida indefinida, el FCD puede incluir un valor residual que representa el

valor del activo al final del periodo explícito proyectado.

2.4.3 Enfoque de coste

El enfoque de coste se basa en la idea de que nadie pagaría por un activo más de lo que costaría adquirir o construir otro de igual utilidad. Este enfoque proporciona una estimación de valor al calcular el coste de reposición o reproducción de un inmueble, realizando las posibles deducciones debido a deterioro físico u otras obsolescencias relacionadas.

Dentro del enfoque del coste, existen tres métodos distintos: el método de coste de reproducción, el método de coste de reposición y el método aditivo. El primer método estima el valor basado en el coste de un activo similar con utilidad equivalente. El segundo método, se centra en calcular el coste de recrear una réplica exacta del activo. Finalmente, el método aditivo determina el valor sumando los valores de sus componentes, siendo especialmente útil en empresas de inversión donde el valor de la inversión es determinado por el valor de sus participaciones.

2.4.4 Otras consideraciones

En España cabe destacar que la Orden ECO/805/2003 establece cuatro métodos para la valoración inmobiliaria: el método de comparación, el método de actualización de rentas, el método residual y el método de coste. Sin embargo, ningún método es aplicable de manera universal, aunque el método de comparación en un mercado activo se considera generalmente como la evidencia más sólida para la valoración. No obstante, si la información proviene de mercados inactivos, pueden estar sujetas a ajustes subjetivos realizados por el valuador. En situaciones concretas, el tasador y cliente pueden acordar este tipo de procedimientos, aunque esta valoración podría calificarse como limitada o no cumplir con los estándares de las IVS.

3. VALORACIÓN INMOBILIARIA MEDIANTE EL BIG DATA Y LA INTELIGENCIA ARTIFICIAL

La digitalización y automatización de procesos se han posicionado como una de las principales prioridades en el sector inmobiliario durante varios años. La pandemia aceleró la adopción de servicios digitales, permitiendo mantener las operaciones normales durante el estado de alarma y confinamiento de 2020. Estos servicios permitieron llevar a cabo actividades cotidianas de forma remota, como visitas virtuales a propiedades, valoraciones automatizadas de inmuebles y el cierre de transacciones mediante contratos en línea.

En este contexto, los AVM en 2020 y principios de 2021 experimentaron un incremento y expansión en la industria. En 2020, se realizaron más de 2,5 millones de valoraciones inmobiliarias, de las que el 60 % se realizaron con modelos automatizados (AltoDirectivo, 2021). Sin embargo, a pesar del auge que tuvo durante la pandemia, ha mantenido su crecimiento en los años siguientes. En 2021, entre octubre y diciembre se produjo un incremento del 400% respecto al tercer trimestre de ese mismo año, alcanzando casi 1,2 millones de valoraciones inmobiliarias (AltoDirectivo, 2021).

3.1 Concepto y Fundamento

Según AEV (2019), “se consideran AVM aquellas metodologías de valoración basadas en programas informáticos que permiten obtener, con un grado de confianza determinado, valores de tasación de inmuebles a partir de un conjunto de datos y parámetros relativos a los mismos, empleando para ello prácticas de valoración generalmente aceptadas”. Estos modelos se diferencian de las tasaciones tradicionales, de utilizar fórmulas matemáticas para calcular las valoraciones, sin necesidad de una visita física ni de la intervención directa de un tasador profesional.

Los primeros intentos de desarrollar los AVM surgieron en Norteamérica durante la década de los 80s, impulsados por innovaciones tecnológicas como los Sistemas de Información Geográfica (SIG) y plataformas digitales que mostraban las propiedades disponibles. En 1981, el desarrollo de los SIG destacó por el lanzamiento comercial del software ARC/INFO, permitiendo a los profesionales del sector inmobiliario visualizar datos geospaciales y facilitando el análisis sobre las ubicaciones de las propiedades

(ESRI, s.f.). Simultáneamente, en la mitad de los años 80, la creación de plataformas como el *Multiple Listing Service* (MLS) permitió a los agentes inmobiliarios compartir y acceder a datos de forma más eficiente, centralizada y en tiempo real. Todo esto, unido a las mejoras computacionales y el desarrollo de modelos estadísticos establecieron las bases de desarrollo de esta nueva tecnología.

La relevancia de los AVM se intensificó en la década de los 2000 con la llegada de internet y la expansión de plataformas inmobiliarias en línea como Zillow, Redfin y Realtor. Este avance permitió a los usuarios obtener estimaciones de sus propiedades, sin necesidad de intermediarios y de forma remota. El auge de estas plataformas impulsó la necesidad de mejorar la precisión de los modelos mediante la integración de análisis masivos de datos y modelos de aprendizaje supervisado. Cabe destacar que la precisión de cada AVM varía según sus fórmulas específicas, la tecnología usada y la calidad de los datos. Un ejemplo ilustrativo es la comparación entre los modelos de Redfin y Zillow antes y después de establecer el precio de venta. Antes de la publicación del precio de venta, Redfin muestra una tasa de error mayor (9.1%) que Zillow (7.8%). Sin embargo, tras comunicar el precio, el error de Redfin disminuye notablemente a un 2.7%, superando en precisión a Zillow, que registra un 4.1% (Thind, 2017). Esta variación evidencia que la precisión de las estimaciones está directamente vinculada a las estrategias, fórmulas adaptadas en cada modelo y a la habilidad del profesional responsable en desarrollar el modelo (IAAO, 2018).

3.2 Usos de los AVM

Los modelos de valoración automática son ampliamente utilizados en todo el mundo, siendo Estados Unidos y Canadá pioneros en su implementación y posterior transferencia de conocimiento a otros países de habla inglesa, tales como Reino Unido, Australia, Nueva Zelanda y Sudáfrica (Downie & Robson, 2008). En España, los AVM son ampliamente utilizados por entidades financieras, incluyendo bancos y fondos de inversión inmobiliaria, así como entidades gubernamentales y compañías de tasación (AEV, 2019). En la siguiente tabla, se aprecian los principales usos de los AVM por cada agente involucrado en la valoración.

Tabla 1: Usos de los AVM en España

Fuente: Adaptado de AEV (2019)

Entidades Financieras	Entidades Bancarias	<ul style="list-style-type: none"> – Herramienta de control de calidad para la actualización de bienes adjudicados – Valoración del portfolio de activos hipotecados y efectos contables de inmuebles
	Estados financieros	<ul style="list-style-type: none"> – Revisión y actualización periódica del valor de carteras de colaterales – Gestión de riesgos
	Fondos de Inversión inmobiliario	<ul style="list-style-type: none"> – Apoyo en la fijación de valores de venta o determinación del valor razonable – Soporte de revisión de valoraciones propias y gestión inmobiliaria – Obtención de orientación sobre precios en mercados locales
Entidades Gubernamentales	Administración Pública	<ul style="list-style-type: none"> – Cálculo de las bases imponibles para impuestos o tasas sobre inmuebles – Valoraciones catastrales y revisiones
Ámbito empresarial	Empresas de Valoración	<ul style="list-style-type: none"> – Elaboración de índices en mercados locales – Control de valoraciones, calibración y contraste de base de datos de precios – Valoración online de viviendas individuales

3.3 Legislación AVMs en España

Los AVMs tienen la responsabilidad de adherirse a las prácticas establecidas en la normativa internacional, europea y española para la valoración de inmuebles, como se detalló anteriormente (IVS, EVS, ECO 805/2003). Los AVMs deben seguir los principios, métodos y normas de los enfoques de valoración tradicionales al ser empleados. Por ejemplo, según la Orden ECO 805/2003, es importante que los valuadores tengan una extensa base de datos con transacciones realizadas y ofertas ajustadas. Esta información permite identificar los parámetros en la homogeneización de inmuebles comparables.

En el contexto español, los AVMs carecen de regulación o normativa específica (AEV,

2019). No obstante, se reguló en 2017 el uso de los AVM para atender encargos procedentes de entidades crediticias, con la finalidad de realizar una valoración contable de inmuebles utilizados como garantías y adjudicados. Esta regulación se establece en el Anexo 9 de la Circular 4/2017, de 27 noviembre, que aborda las normas de información financiera pública y privada y los modelos de estados financieros. Esta directriz está dirigida a las empresas de tasación inscritas en los registros del Banco de España con el objetivo de identificar las mejores prácticas para establecer y aplicar procedimientos sólidos utilizando AVM.

En el Anexo 9, se especifica que los AVM solo son aptos para valorar propiedades inmobiliarias que presenten características homogéneas y que estén presentes en un mercado activo de transacciones. Además, según el punto 78.b.ii, los AVM deben seguir prácticas de valoración aceptadas, entre las que se incluyen: el modelo de regresión múltiple hedónico, de comparables y el modelo mixto (la combinación de los dos anteriores).

Cabe destacar que en la Circular 3/2008 del Banco de España, que transpone la Directiva europea 2006/48/CE a la normativa española, establece en su artículo 4.2 que la aplicación de modelos AVM es para la actualización del valor de inmuebles residenciales y locales comerciales, pero no para la tasación principal del bien hipotecado. Además, especifica la necesidad de valoraciones adicionales por un tasador independiente, si las condiciones del mercado u otras circunstancias están afectando a los precios provocando una reducción del riesgo de crédito.

En el ámbito internacional, existen varios estándares relacionados con los modelos AVM. En 2018, la *International Association of Assessing Officers (AAIO)* aprobó *Standard on Automated Valuation Models (AVMs)*, mientras que, en 2019, la Asociación Española de Análisis de Valor estableció el estándar sobre valoración de inmuebles mediante modelos automatizados (AVM). Por último, cabe mencionar el glosario de términos relacionados con los modelos AVM presente en los *European Standards for Statistical Valuation Methods for Residential Properties* elaborado por la *European AVM Alliance*, el documento técnico informativo contenido en el EVIP 6, realizado por el EVS y una nota informativa publicada por el *Royal Institution of Chartered Surveyors (RICS)* llamada *Automated Valuation Model*.

3.4 Proceso de Valoración Automatizada

3.4.1 Negociación previa al proceso de valoración

En el escenario actual de transformación digital, la automatización de procesos emerge con el objetivo de optimizar la eficiencia y transparencia en diversas áreas, incluidos los acuerdos contractuales. En este contexto, los *smart contracts* surgen como solución innovadora para simplificar el proceso de negociación, reducir riesgos, eliminar intermediarios y acelerar la ejecución de obligaciones contractuales marcando un cambio significativo en cómo se realizarán las transacciones y los contratos.

Los *smart contracts* o contratos inteligentes, son programas almacenados en la *blockchain* que se ejecutan de manera automática y autónoma cuando se cumplen unos parámetros programados. Generalmente se utilizan para automatizar la ejecución de un acuerdo, de modo que todos los participantes estén seguros del resultado, sin la intervención de ningún intermediario. Esto es debido al uso de la tecnología *blockchain* que su principal función es reforzar la seguridad, transparencia y confianza entre los participantes, evitando malentendidos, falsificaciones o modificaciones (Iberdrola, s.f.). Estos contratos toman relevancia especialmente en la automatización de pagos y gestión de garantías en el ámbito de las hipotecas, facilitan la transparencia de propiedad y administración de contratos de arrendamiento y agilizan la tramitación de reclamaciones de aseguradoras. No obstante, aún se considera una tecnología incipiente en el sector inmobiliario, y si alguien deseara adquirir una propiedad mediante contratos inteligentes, el procedimiento que debería seguir no se encuentra totalmente definido.

3.4.2 Investigación

La investigación sobre Modelos de Valoración Automatizada consiste en inspeccionar y analizar las propiedades sin necesidad de la intervención de un tasador. Se centra en el desarrollo y perfeccionamiento de tecnologías como sensores remotos, Internet of Things y recopilación de imágenes.

Sensores remotos

Los sensores remotos son dispositivos que capturan la información sobre el entorno o la superficie terrestre desde una distancia considerable, sin necesidad de intervención humana. Su principal objetivo es obtener datos geoespaciales con precisión mediante la utilización de sensores ópticos, radar o térmicos, pudiendo estar implementados en

satélites, aviones o drones. Esta tecnología se ha convertido en una de las pocas herramientas capaces de captar las condiciones ambientales exteriores de una vivienda, facilitando la valoración al ofrecer una vista aérea desde lo alto de la propiedad. Los sensores remotos destacan por recibir y procesar tres tipos de fuentes:

En primer lugar, **las imágenes multiespectrales** son un tipo de fuente que se centra en representar el espacio verde urbano mediante el cálculo del índice de vegetación de diferencia normalizada, evaluando así las condiciones ambientales en términos de vegetación y superficie impermeable (Huerta & Yépez, 2019). Esta tecnología simplifica la valoración del entorno cercano a la propiedad ofreciendo atributos más sólidos relacionados con los vecinos al proporcionar información ambiental más detallada que las imágenes estándares (Wei, Fu, Yang, Tang, & Xiong, 2022).

En segundo lugar, la **luz nocturna** surge como indicador en la valoración inmobiliaria, debido a que existe una fuerte correlación espacial entre el Producto Interior Bruto (PIB) y la densidad poblacional. Algunos investigadores la utilizan exitosamente en la modelización de los precios de las viviendas al ofrecer una representación eficaz del desarrollo urbano (Wei, Fu, Yang, Tang, & Xiong, 2022)

Por último, los **radares láser** resaltan por ofrecer nuevos parámetros tridimensionales en el modelo de valoración. Se emplean para medir el área y volumen de las viviendas, y para evaluar en qué manera el ángulo de visión al mar influye en el valor de viviendas situadas en la playa (Wei, Fu, Yang, Tang, & Xiong, 2022).

En términos generales, la valoración de inmuebles está estrechamente vinculada a su entorno, demostrándose que el uso de sensores remotos y otros datos geoespaciales mejoran considerablemente la precisión de las valoraciones inmobiliarias. Sin embargo, la incorporación de esta información conlleva desafíos debido a su complejo proceso de extracción.

Internet of Things

El Internet de las Cosas se refiere a la interconexión de dispositivos mediante el internet, permitiendo la comunicación entre ellos y la nube. Esta tecnología se ha implementado en variados sectores, incluidos los hogares inteligentes, la logística y la seguridad alimentaria, gracias a su capacidad de intercambio a tiempo real de la información. Esta

característica permite el seguimiento de la ubicación y estado de los dispositivos en el IoT, siendo particularmente útil en la valoración inmobiliaria a través del análisis de datos como tarjetas inteligentes, posicionamiento GPS y monitoreo ambiental.

Los **datos de tarjetas inteligentes**, que representan datos de viajes anónimos, se emplean para evaluar la accesibilidad a recursos públicos en áreas residenciales. Estos datos no solo influyen en la valoración al calcular la relación entre la accesibilidad y los precios inmobiliarios, sino que también permiten caracterizar aspectos socioeconómicos de los usuarios. Además, ofrecen información valiosa sobre el comportamiento turístico y las condiciones económicas de posibles compradores (Wei, Fu, Yang, Tang, & Xiong, 2022).

Además, los **datos de posicionamiento GPS** cumplen una función análoga a las tarjetas inteligentes, reflejando trayectorias de viajes. Sin embargo, esta tecnología se aplica tanto en los GPS de vehículos como en el de los teléfonos móviles, con el propósito de recopilar datos sobre la actividad del individuo y analizar las preferencias de las personas desde una perspectiva de demanda inmobiliaria (Wei, Fu, Yang, Tang, & Xiong, 2022).

Por último, los **datos de monitoreo ambiental** se obtienen mayormente a través de organismos gubernamentales y de sensores ambientales. Son esenciales para la valoración de inmuebles debido a que abarcan distintos aspectos como el ruido, la calidad del aire y la contaminación del agua (Wei, Fu, Yang, Tang, & Xiong, 2022). Aunque su inclusión en la valoración se ve limitada por la dificultad en obtener datos detallados, si se incorpora de forma apropiada puede potenciar la capacidad explicativa de los modelos, especialmente en regiones con riesgos ambientales significativos.

Recopilación de Imágenes

Aunque no es una tecnología nueva, los dispositivos de grabación se presentan como herramientas útiles durante la inspección interna de una propiedad, ahorrando tiempo al tasador (Scheurwater, 2017).

La cuestión fundamental radica en si cualquier individuo puede llevar a cabo estas grabaciones o si se requiere de una formación específica para asegurar la calidad y precisión de las imágenes. Otra alternativa valiosa sería la transmisión en vivo de la grabación, permitiendo al tasador guiar de forma remota al vendedor. Por ejemplo, Zoho Lens ofrece una solución de asistencia remota que permite a los expertos guiar a los

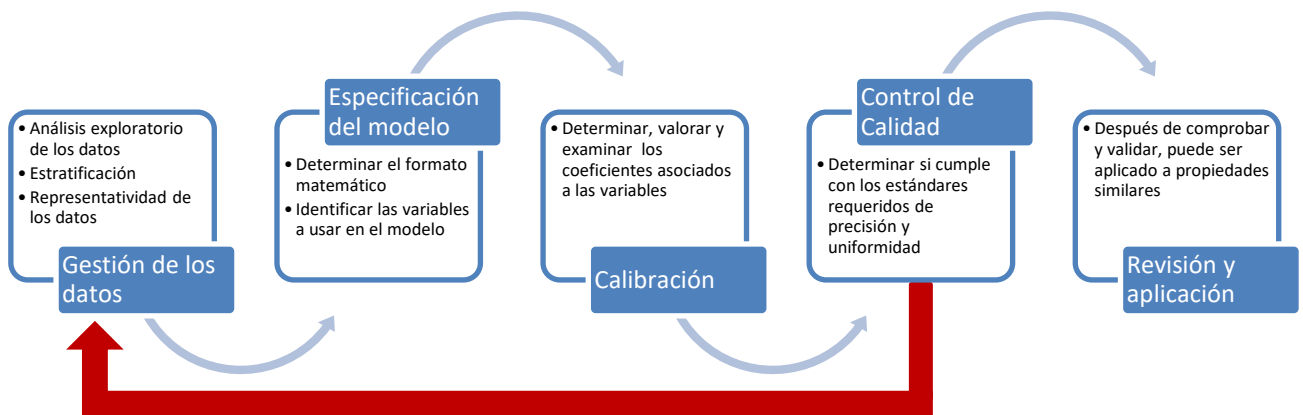
usuarios mediante la inspección visual de la propiedad con su dispositivo móvil. Estas imágenes resultan valiosas al poder ser revisadas posteriormente por el tasador, y, además, pudiendo ser analizadas directamente mediante métodos de Inteligencia Artificial.

Este enfoque permitiría agilizar el proceso de tasación, pero plantea interrogantes sobre su capacidad para proporcionar una valoración integral de la propiedad. La pregunta se centra en si una simple imagen puede captar y reflejar todos los aspectos relevantes para la tasación de una vivienda. Por ello, la ausencia física del tasador podría dar lugar a la omisión de datos significativos o a la interpretación sesgada de ciertos aspectos para la valoración.

3.4.3 Tratamiento e interpretación de datos

El proceso tratar e interpretar datos en la construcción de un modelo AVM incluye cinco etapas: gestión de los datos, especificación del modelo, calibración, control de calidad y finalmente, revisión y aplicación. La figura siguiente detalla estos pasos con mayor profundidad, ofreciendo una visión global de cada etapa:

Figura 3: Construcción de un modelo AVM



Fuente: Elaboración propia

Gestión de los datos

La primera fase de la construcción de un AVM consiste en la recopilación, verificación, tratamiento y mantenimiento de los datos. Esta etapa asegura que el modelo se base en

datos precisos y relevantes, permitiendo así estimaciones fiables.

El proceso comienza con la recopilación de datos, que pueden provenir de la investigación de la propiedad o ser suministrados por el cliente. Estos datos se clasifican en tres categorías: información de la vivienda, datos de localización y datos relevantes del mercado (IAAO, 2018). Posteriormente, se realiza una verificación para asegurar que los datos recopilados contengan información relevante para el análisis y se determinen las propiedades que tiene que estar excluidas del modelo (AEV, 2019). La verificación se realiza mediante un análisis exploratorio de los datos, el cual abarca una evaluación exhaustiva de la calidad de estos. En este punto, se identifican y eliminan errores, se analizan distribuciones en los datos, se investigan tendencias del mercado y fluctuaciones temporales, asegurando así la integridad y relevancia de los datos utilizados (IAAO, 2018).

Adicionalmente, se aplican los índices de los precios locales, se efectúa una segmentación detallada y se registra la fecha de valoración. Por último, es fundamental garantizar que la muestra utilizada represente de manera fiel y exhaustiva las características más importantes de la población total para una estimación del valor más precisa.

Especificación del modelo

La especificación del modelo consiste en seleccionar formalmente una metodología y su posterior desarrollo en una ecuación, apoyándose en el análisis de datos y en los principios de la valoración. (IAAO, 2018).

Las diferentes metodologías emplean algoritmos específicos y se apoyan en un conjunto de variables determinadas para calcular el valor. Por lo tanto, el primer paso consiste en la selección de variables consistiendo en la “selección del subconjunto de atributos que pueden resultar de máxima utilidad para las tareas de clasificación o regresión posteriores” (Roque López, 2021). Identificar las variables más relevantes es uno de los desafíos más notables en el análisis masivo de datos, debido a que la correcta selección permite reducir la posibilidad de incurrir en *overfitting*, mejorar la precisión del modelo, eliminar variables redundantes, simplificar el modelo y reducir el tiempo de proceso (Roque López, 2021). Existen dos técnicas para elegir las variables del modelo, la técnica de eliminación de variables y técnicas de selección de variables (Gallego Mora-

Esperanza, 2008).

Por último, la Sociedad de Tasación ajustará la metodología en función de las particularidades del encargo y las características de los inmuebles valorados, siguiendo las mejores prácticas estadísticas y evitando metodologías que no se ajusten a los datos disponibles (AEV, 2019). Más adelante, se profundizará en las nuevas metodologías utilizadas en los Modelos de Valoración Automatizada.

Calibración

La calibración del modelo consiste en establecer los coeficientes asociados a cada variable, evaluando su relevancia y coherencia en un modelo AVM (IAAO, 2018). Estos coeficientes se calculan mediante un análisis del mercado de las variables involucradas, utilizando métodos estadísticos para asegurar la calidad de la calibración. Los datos se dividen en una muestra de entrenamiento (70%) y una muestra de control (30%) para la validación del modelo. Posteriormente, el analista puede realizar ajustes en la especificación del modelo, como transformaciones y otras técnicas, hasta optimizar los coeficientes. Las técnicas de calibración se aplican en las metodologías de regresión lineal como no lineal y Redes Neuronales.

Los modelos basados en regresión utilizan una amplia gama de estadísticas diagnósticas para evaluar, incluidas las medidas de bondad de ajuste (R^2 , R^2 ajustado, coeficiente de variación y error estándar de la estimación) y significancia de las variables (R , t , F , etc.) (Roque López, 2021). Por otro lado, los modelos creados mediante redes neuronales, el analista introduce cada variable asignándole un peso (coeficiente) y el algoritmo calibra los pesos en una capa oculta de tal manera que reduce el error cuadrático (Núñez Tables, Rey Carmona, & Caridad y Ocerín, 2017).

Control de calidad

En España, las valoraciones automatizadas están reguladas en la Circular del Banco de España 4/2017 que obliga a la realización de un *backtesting* con la finalidad de asegurar la consistencia y calidad de los resultados del modelo.

El *backtesting* es una técnica que evalúa la efectividad del modelo analizando la muestra de control. Al examinar este conjunto, se realizan inferencias de las características

generales de todos los inmuebles valorados. El objetivo principal es de verificar si el modelo alcanza los estándares de precisión y coherencia requeridos. Además, la aceptabilidad de los resultados dependerá de las expectativas del cliente y las políticas de gestión de riesgos establecidas.

Revisión y aplicación

Una vez es aprobado y validado, el modelo AVM puede ser empleado en propiedades similares dentro de áreas o regiones específicas para las cuáles fue diseñado. Además, para mantener su validez a largo plazo, es necesario realizar revisiones de los valores que genera, asegurando que las valoraciones sigan siendo razonables y coherentes con las dinámicas futuras (IAAO, 2018).

3.4.5 Informes sobre el proceso posterior a la evaluación

La mayoría de los informes de valoración se presentan en formato papel, siendo poco atractivos para la lectura y limitar el interés del cliente a solo dos aspectos: el valor final de la propiedad y la firma del tasador (Scheurwater, 2017). No obstante, la adopción de tecnologías avanzadas, como visualización 3D, mapas de calor y realidad virtual podría mejorar la presentación de estos informes. Estas herramientas enriquecen la experiencia del usuario al ofrecer una interfaz más interactiva y entendible, facilitando una mejor visualización y comprensión de la valoración de su propiedad.

3.5 Nuevos Métodos de Valoración

Los nuevos métodos de valoración surgen como respuesta a la búsqueda de precisión en la estimación del precio, a su vez que, ayudar a los tasadores en amenizar el proceso de valoración. Existen diversas metodologías, que al igual que los métodos de valoración tradicionales, el encargado del modelo tendrá que valorar cuál se adapta más a las características del inmueble y de las especificaciones del cliente. En la actualidad, los principalmente empleados son los modelos hedónicos, los sistemas basados en la Inteligencia Artificial y los clasificadores de comparables (Flores Guillén, 2016). Sin embargo, se realizará un análisis adicional sobre otras metodologías que están siendo incipientes en el mercado.

3.5.1 Modelos de precios hedónico

Concepto y fundamento

El modelo hedónico consiste en utilizar un modelo de regresión, con el objetivo de estimar el valor de un bien o servicio descomponiendo sus características específicas y evaluando cómo contribuye al precio total (Gila García & Novás Filguera, 2012). Se basa en la idea de que el precio está determinado por sus características individuales, tanto factores internos como externos. Esta metodología se ha empleado en una amplia gama de productos con características homogéneas, como teléfonos, ordenadores y automóviles, entre otros. Por ello, en la valoración de inmuebles es dónde sin ninguna duda existe un mayor número de contribuciones en la literatura científica (Núñez Tabales & Rey Carmona, 2017).

La AEV (2019) define el modelo hedónico como:

“Método de valoración estadística que estructura y descompone cada inmueble en parámetros por razón de su localización, entornos, atributos, características, etc., y que, mediante algoritmos matemáticos, explica y atribuye el valor para un inmueble relacionado con una muestra parametrizada de testigos de su entorno próximo”.

El origen de esta metodología se remonta a 1922, cuando Hass emplea este enfoque para determinar los precios de la tierra cultivable en función de la fertilidad del suelo, la distancia a la ciudad de mercado, el valor de construcción, entre otras variables (Caridad, Ceular Villmandos, & Núñez Tabales, 2008). Además, en 1929 hay evidencias de que se utilizó esta metodología para el análisis de la calidad de las legumbres. Sin embargo, la mayoría de los autores sitúan el origen de esta metodología a la investigación realizada por Court (1939) para la asociación de fabricantes del sector automotriz, cuando el gobierno estadounidense mostró preocupación por posibles oligopolios en Detroit. El análisis de Court proporcionó el fundamento del método de precios hedónicos al demostrar que el aumento de los precios se debía a mejoras significativas en la calidad y las características de los productos.

La primera aplicación de este enfoque en el sector inmobiliario, se encuentran en las investigaciones de Ridker y Henning (1967), quienes proporcionaron pruebas empíricas de que la contaminación tenía un impacto en el precio de las viviendas. Por lo tanto, las

propiedades de una vivienda pueden clasificarse en mínimo dos categorías distintas: las categorías estructurales de la vivienda y las características vinculadas a la ubicación y entorno físico (Caridad, Ceular Villmandos, & Núñez Tabales, 2008). Además, en 1974, Shervin Rosen fue el pionero al desarrollar un enfoque unificado para analizar cómo los precios reflejan diferentes características de los productos o servicios, mediante la metodología de precios hedónicos. Desde ese momento, el modelo concebido por Rosen ha alcanzado consenso general como paradigma central de este enfoque, especialmente en países anglosajones, tanto para calcular el precio de la vivienda como obtener el índice de precios (Gila García & Novás Filguera, 2012).

Selección de variables

En la literatura científica existen diversos estudios sobre los factores que determinan el precio del inmueble, destacando la investigación de Sirmans et al. (2005), quienes realizaron 125 estudios sobre viviendas entre 1994 y 2004, llegaron a dos conclusiones:

En primer lugar, cada vivienda dispone de un conjunto único de atributos que influyen en su valor, y estas características son apreciadas de distinta manera en cada zona geográfica. Por ejemplo, un sistema de calefacción central podría ser más apreciado en una región con inviernos extremadamente fríos, o un sistema de aire acondicionado en una zona con veranos calurosos. Estas diferencias hacen que los resultados obtenidos en una región específica sean complicados de generalizar a otras ubicaciones debido a variaciones en condiciones climáticas, aspectos sociales y culturales.

Por último, cada potencial comprador puede valorar los atributos de la vivienda de manera distinta. Por ejemplo, en Finlandia, los compradores están dispuestos a pagar un 4.8% más por una vivienda si tiene vistas a un bosque (Tyrväinen & Mettinen, 2000), mientras que, en Suecia, el precio del inmueble disminuye un 30% en vecindarios afectados por ruido del tráfico (Wilhelmsson, 2000). De esta manera, varios compradores pueden valorar de forma las características de una propiedad, lo que dificulta la tarea de valoración.

Por todo lo mencionado anteriormente, en cada tasación se definen y calculan las variables de manera distinta, dificultando la comparación entre estudios. Sin embargo, según Núñez-Tabales y Rey-Carmona (2017), las variables más utilizadas se ordenaban

de la siguiente manera:

1. Edad del inmueble
2. Superficie
3. Garaje, tamaño del terreno y chimenea
4. Número de dormitorios, número de baños, piscina y sótano
5. Ambiente natural, como vistas a un lago, mar o simplemente “buenas vistas”
6. Ubicación, el índice de criminalidad, cercanía a un campo de golf y presencia de árboles.
7. Servicios públicos, como distrito escolar, presencia de población minoritaria creciente en escuela y acceso a alcantarillado
8. Ocupación y factores de venta como estado de conservación, impuestos sobre la propiedad, condiciones de habitabilidad, entre otras

Expresión del modelo

Un modelo hedónico se expresa a través de una función matemática que relaciona el precio del bien inmueble con sus atributos, es decir, utilizando un modelo de regresión. La expresión general del modelo es la siguiente:

$$P=\beta_0+\beta_1X_1+\beta_2X_2+\dots+\beta_nX_n+\varepsilon$$

Donde:

- P es el precio del inmueble
- X_1, X_2, \dots, X_n representan las diferentes características del inmueble
- $\beta_0, \beta_1, \dots, \beta_n$ son los coeficientes que indican la contribución relativa de cada atributo al precio, es decir, la variación de la X en la variable Y
- ε corresponde a la perturbación aleatoria que influye en la Y, que captura cualquier variación no explicada por las variables incluidas en el modelo.

Sin embargo, la forma funcional no necesariamente corresponde a un modelo de regresión lineal, debido a que la relación entre el precio y algunas de sus variables explicativas pueden no serlo. La expresión del modelo resultaría en un modelo de regresión no lineal entre la variable independiente y dependientes mediante funciones cuadráticas, logarítmicas, exponenciales o de interacción. Algunos ejemplos de que puede existir

regresión no lineal en este tipo de modelos son los siguientes:

- El modelo tiende a seguir formas logarítmicas en relación con la superficie, ya que el precio tiende a decaer a medida que la superficie aumenta de manera significativa.
- El modelo puede incorporar interacciones significativas, como la relación entre el aumento del precio debido a las diferencias de tamaño y si es nueva o de segunda mano. Esto implicaría la introducción de parámetros adicionales al modelo, evidenciando la interacción entre estas dos variables.

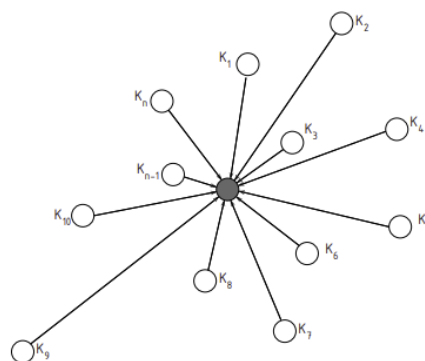
3.5.2 Método de comparables

Concepto y fundamento

El modelo de comparables, también conocido como *sales comparison model*, consiste en la automatización de los métodos tradicionales de comparación y de actualización de rentas mencionados anteriormente (Agrasar González, 2019). En la actualidad, el clasificador de comparables más utilizado es el método KNN, o K-vecinos más cercanos (Flores Guillén, 2016).

Concretamente, el KNN es una técnica de aprendizaje supervisado que clasifica o predice la categoría de un punto de datos basándose en las características de los ‘k’ puntos más cercanos a él. Simplificadamente, si un punto de datos necesita ser clasificado, el KNN busca las ‘k’ propiedades más parecidas, y según estas, asigna el punto a un grupo o un clúster específico.

Figura 4: Representación del método k-vecinos



Fuente: (Jaureguizar Oriol, 2017)

Funcionamiento del método K-vecinos

El algoritmo KNN se caracteriza por ser no paramétrico, es decir, no genera ninguna fórmula. Utiliza reglas lógicas aplicadas a un grupo de características similares para estimar el precio (Jaureguizar Oriol, 2017). En la tasación inmobiliaria, se identifican los inmuebles más parecidos al evaluado, entre sus vecinos más cercanos, es decir, aquellos inmuebles que conocemos su precio.

Se parte de una base de datos de inmuebles de los que se conoce el precio, y al que se han aplicado la calificación, tratamiento, selección de variables, etc. (Gallego Mora-Esperanza, Modelos de valoración Automática, 2008). En esta primera etapa es importante estandarizar las variables, para que en los pasos siguientes no influya mucho más una variable sobre otra.

Para calcular el valor un bien raíz, el primer paso es identificar cuáles son sus vecinos más cercanos en la base de datos. Para determinar la proximidad, se calcula la distancia entre la propiedad que deseamos valorar y aquellas que conocemos su precio. Esta distancia no es física, sino lógica, debido a que incorpora todas las variables, no solo las relacionadas con la ubicación (Gallego Mora-Esperanza, Modelos de valoración Automática, 2008). Se emplean varios métodos para calcular la distancia, como la euclídea, Manhattan, Mahalanobis, Minkowski, Chevishev, entre otros, siendo la distancia euclídea la más comúnmente utilizada. Después de calcular la distancia entre la propiedad a valorar y todas las propiedades con precios conocidos, se seleccionan aquellas con menos distancia.

Posteriormente, se eligen los vecinos más cercanos o el número de clústeres que participarán en el cálculo. Es la incógnita más relevante de aplicar el método de KNN debido a que no existe ningún criterio objetivo para la elección del número correcto, pudiendo afectar significativamente en el resultado. Para facilitar la elección, se utilizan el método del codo y el método de la silueta, ambos proporcionando en una visualización el número adecuado de clústeres. Finalmente, para calcular el valor con este método, hay tres opciones entre las cuales el usuario puede elegir: valor medio, valor ponderado por distancia y método global.

3.5.3 Inteligencia Artificial (IA): Redes Neuronales

Concepto y fundamento

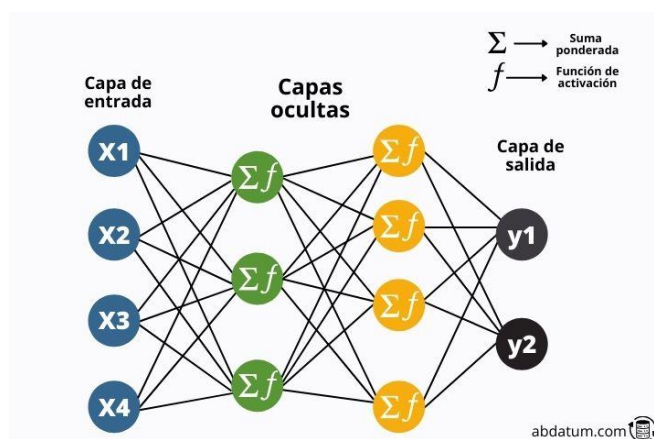
Las redes neuronales artificiales (RNA) son estructuras computacionales que se inspiran en ciertas características de los humanos, como la capacidad de memorizar y establecer patrones. Desde una perspectiva más técnica, y sin adentrarnos en los detalles del funcionamiento que se abordarán posteriormente, las redes neuronales son sistemas complejos compuestos por elementos simples conocidos como nodos o neuronas. Estos elementos están dispuestos en capas y establecen conexiones entre sí mediante enlaces de transmisión de la información. Así, las redes neuronales pueden aprender de experiencias pasadas, permitiéndoles ajustar su comportamiento o respuesta ante decisiones futuras. Por lo tanto, las redes neuronales son simplemente una representación artificial y simplificada de la neurona biológica.

Todo comenzó en 1943, cuando el neurofisiólogo Water Pitts y el matemático Warren McCulloch publicaron el artículo *A Logical Calculus of the ideas immanent in nervous activity* en el Boletín de Biofísica. Fueron los primeros en concebir una teoría basándose en el funcionamiento del cerebro con propósitos computacionales. Esta teoría se basaba en el funcionamiento de las neuronas, los cuales modelaron una red neuronal simple mediante circuitos eléctricos (Matich, 2001). A partir de este artículo, emergieron diversas líneas de investigación en el ámbito de las redes neuronales: una corriente dirigida a los procesos biológicos y otra enfocada en el desarrollo de la IA. Las investigaciones más influyentes originadas de estas nuevas corrientes se remontan a 1953 con la introducción de la neurona artificial o perceptrón, seguido por la propuesta de la retropropagación de Paul Werbos en 1974 y, hacia finales de los años 80, el desarrollo del entrenamiento de algoritmos liderado por Yann LeCun.

Elementos básicos

Una neurona artificial es un conjunto de neuronas conectadas y organizadas en capas. Estas capas se dividen en tres tipos: la **capa de entrada** que recibe la información externa, las **capas ocultas**, denominadas así por no ser visibles, con un número variable de capas dependiendo el objetivo a conseguir y, por último, **la capa de salida** que, según el tipo de escenario (regresión o clasificación), tendrá una o más neuronas. A continuación, se puede ver la estructura de una red neuronal:

Figura 5 :Elementos de una red neuronal



Fuente: (Canadas, 2022)

Funcionamiento de la RNA

En primer lugar, el entrenamiento de una red neuronal es de los primeros factores a considerar para implementar este tipo de metodología. Al igual que los seres humanos, las RNA tienen la capacidad de abordar situaciones mediante la experiencia acumulada. El proceso de aprendizaje consiste en que las RNA analizan la información, predicen el resultado y se autocorrijen en caso de error, perfeccionando su conocimiento en cada corrección. Este proceso se repite hasta que la metodología pueda predecir con exactitud la salida de nuevos datos.

En cuanto al funcionamiento de una red neuronal, es más sencillo si se imagina cada nodo como un propio modelo de regresión lineal formado por datos de entrada, ponderaciones, un sesgo y una salida.

$$\sum w_i x_i + \text{sesgo} = w_1 x_1 + w_2 x_2 + w_3 x_3 + \text{sesgo}$$

Primero, se asignan de manera aleatoria los pesos iniciales a las conexiones entre neuronas, influenciando una neurona sobre otra. Una vez establecidos estos pesos, los datos de entrada son introducidos a la red y da lugar a dos actividades dentro de la neurona: la combinación lineal de las entradas ponderadas y la función de activación. Esta función evalúa el estado de actividad de la neurona, transformando la entrada en un valor en el rango de 0 a 1. En este intervalo, una neurona puede estar inactiva (0 o -1) o activa (1) (Matich, 2001). Si la salida supera cierto umbral, activa el nodo y transfiere los datos

a la capa siguiente la red. En consecuencia, la salida de un nodo se convierte en la entrada del nodo siguiente, convirtiéndose en una red de propagación hacia delante. Posteriormente, la información calculada alcanza la capa de salida y se compara con resultados reales, creándose una función de coste que determina el grado de error de la predicción.

En este caso el umbral a superar es 3 o valor de sesgo de -3:

$$\text{Error} = (5*1) + (2*0) + (4*1) - 3 = 6$$

Finalmente, los pesos se ajustan utilizando algoritmos de optimización y el ciclo se repite hasta llegar a un error aceptable del 5% en la valoración de inmuebles (Gallego Mora-Esperanza, 2004). En este momento, finaliza el entrenamiento de RNA y se puede utilizar para calcular el valor del inmueble siendo lo más preciso posible al valor del mercado.

Tipos de RNA y aplicación en el sector inmobiliario

Desde su creación, han surgido distintos tipos de RNA adaptadas a funciones específicas. En el ámbito de las valoraciones inmobiliarias, se destacan principalmente dos tipos de RNA, cada una optimizada para abordar diferentes aspectos y desafíos del análisis y tasación de propiedades:

- **Redes neuronales de propagación o Multicapa (MLP):** son las redes en las que nos hemos centrado con anterioridad. Este tipo de redes neuronales se utilizan para calibrar los modelos de valoración inmobiliaria, al estar enfocados en encontrar una combinación de parámetros que reduzca el error cuadrático (IAAO, 2018)
- **Redes convolucionales (CNN):** son redes especializadas en procesar datos con estructura cuadrícula, como imágenes o videos. Tiene varias capas, como las convolucionales, agrupamiento y capas conectadas para reconocer patrones en las imágenes. Los valuadores utilizan este tipo introduciendo imágenes de exteriores o interiores del inmueble que reflejen sus características (antigüedad, decoración, etc.) y las RNA valoran dichas características.

Por último, estos tipos de sistemas destacan por su precisión en los bienes raíces con características atípicas en comparación al mercado (Núñez Tables, Rey Carmona, & Caridad y Ocerín, 2017). Sin embargo, las redes neuronales se enfrentan al desafío de la

“caja negra”, es decir, que la interpretación de sus resultados es especialmente compleja, lo cual ha generado limitaciones en su aceptación en el mercado (Flores Guillén, 2016).

3.5.4 Otros Modelos de Valoración

En los subapartados anteriores, se han detallado las principales metodologías empleadas en los modelos automatizados de valoración. Sin embargo, existe una amplia variedad de métodos y tecnologías que ayudan en el proceso de selección de variables y valoración de inmuebles. A continuación, se presentan algunos ejemplos:

Técnica de la lógica difusa

La técnica de la lógica difusa es un enfoque matemático que permite tratar la información imprecisa, siendo relevante en procesos complejos sin modelos matemáticos precisos. Esta metodología se aplica especialmente en procesos no lineales y cuando intervienen conocimientos que no están estrictamente definidos, es decir, que son imprecisos o subjetivos y están asociados al lenguaje natural (Fernández Agüero, Fernández Cuesta, & Moya, 2008). En la valoración de inmuebles se utiliza como complemento al proceso convencional de tasación, para interpretar datos como “casi seguro”, “bien ubicado”, “piso céntrico”, etc. Además, algunas inmobiliarias en línea, como Zillow, han integrado estos algoritmos para optimizar la búsqueda de viviendas en su plataforma.

Sistemas de Expertos

Sistemas de Expertos son programas informáticos diseñados para abordar problemas específicos, empleando la IA para emular el razonamiento humano. Se les atribuye la denominación de sistemas expertos debido a su habilidad para imitar la toma de decisiones de un especialista. (UNIR, 2022). Actualmente, existen múltiples portales inmobiliarios que permiten a los usuarios ingresar datos sobre la propiedad para obtener una tasación de forma rápida.

Algoritmo Genético (AG)

El Algoritmo Genético es una técnica de búsqueda y optimización inspirada en la Teoría de la evolución de Darwin (Núñez Tabales & Rey Carmona, 2017). Esta teoría postula que, en una población dada, los individuos que logran sobrevivir son aquellos que mejor se adaptan al entorno. Estos algoritmos tienen múltiples aplicaciones, siendo relevante en

la selección de características determinantes del precio de una propiedad (Gallego Mora-Esperanza, 2008).

Análisis Espacial

El Análisis Espacial es un procedimiento que aborda problemas complejos relacionados con la ubicación, facilitando a las empresas el análisis de datos geográficos. Esta metodología permite establecer conexiones, identificar y medir configuraciones, analizar tendencias, así como realizar pronósticos y tomar decisiones (ESRI, s.f.). Además, facilita la integración de datos geográficos y descriptivos de varias fuentes, enriqueciendo los modelos y visualizaciones. Por último, con esta técnica permite crear modelos predictivos basados en datos de una región específica, pudiendo determinar el precio del inmueble en otras áreas.

Teoría multicriterio

Teoría Multicriterio es un conjunto de métodos, técnicas, herramientas y aproximaciones destinados a “mejorar la calidad integral de los procesos de decisión seguidos por los individuos y sistemas, esto a mejorar la efectividad y eficiencia de los procesos de decisión y a incrementar el conocimiento de los mismos” (Moreno Jiménez, 2002). Este conjunto de métodos basados en la Teoría de multicriterio permite abordar valoraciones de manera más integral. Esto implica que el valuador puede considerar aspectos intangibles del inmueble, al mismo tiempo que permite la aplicación de la experiencia y el conocimiento del tasador durante el proceso (Aznar Bellver & Guijarro Martínez, 2020).

3.6 Perspectiva futuras y desafíos de la valoración inmobiliaria

En los inicios, la adopción de los Modelos de Valoración Automatizada en comparación con los métodos tradicionales generó resistencia entre los profesionales, quienes mostraban preocupación por cómo estas tecnologías iban a afectar en su trabajo. Esta nueva preocupación condujo a dos corrientes del pensamiento: una que sostiene que los AVM podrían reemplazar a los tasadores y otra viéndolos como una herramienta complementaria (Scheurwater, 2017).

Actualmente, existe una mayor evidencia de que los AVM funcionan como herramientas

complementarias para los tasadores, transformando su forma de trabajar y el disminuyendo tiempo destinado a diversas fases del proceso. Como ya hemos destacado, la aplicación de esta tecnología se limita a casos específicos dentro de la valoración inmobiliaria y siempre se realiza bajo la supervisión de un profesional. Estos factores, combinados con las notables ventajas relacionadas con la eficiencia y el ahorro de tiempo y costes para los tasadores, han impulsado la progresiva adopción de los AVM tanto en el sector público como en las empresas de tasación. Además, esta corriente propone que la IA puede eliminar la parte ‘científica’ de la valoración, pero no reemplazará el ‘arte’ de valorar (Scheurwater, 2017). Esta perspectiva se vuelve especialmente relevante al abordar los gustos y preferencias individuales, que, al ser subjetivos, introducen un nivel adicional de complejidad en las valoraciones. El desafío radica en que el tasador debe abordar con precisión las tendencias del mercado, cuantificarlas objetivamente y generar tasaciones que reflejen de manera precisa las expectativas de los compradores.

Respecto a la corriente que plantea el posible reemplazo de los tasadores por los AVM, surge un escenario en el que las nuevas tendencias, especialmente la generación de los millenials¹, podrían respaldar esta transformación. Esta teoría se fundamenta en su inclinación hacia la tecnología, viviendo completamente conectados al internet, y a su demanda de valoraciones más rápidas, con variables sostenibles y una perspectiva futura (Scheurwater, 2017). Todo esto unido a su búsqueda de reducir el papeleo para ajustarse a los requisitos de negocio y evitar incurrir en coste de oportunidad.

Sin embargo, ambas corrientes se encuentran ante diversos desafíos en el ámbito de la valoración inmobiliaria, resaltando los siguientes:

- La **necesidad de formar a los empleados** para que puedan analizar la gran cantidad de datos generados y utilizar las nuevas metodologías. Esta formación resulta exigente y complicada, ya que requiere habilidades específicas para comprender y manejar eficazmente la información disponible.
- Existe una **falta de estandarización** en los métodos de recopilación de datos, lo que dificulta la comparabilidad y la consistencia en los resultados.
- Los AVM presentan una **interpretación compleja** y a menudo son percibidos como "cajas negras" debido a la dificultad para interpretarlos.
- **Consideraciones legales y éticas** relacionados con la privacidad y leyes de protección de datos. En España, la Ley Orgánica de Protección de Datos (LOPD)

exige que las entidades obtengan el consentimiento explícito para el tratamiento de datos, asegurando así la privacidad y respeto de los derechos individuales. Además, impone requerimientos para proteger la información y establece mecanismos para que las personas puedan ejercer sus derechos de acceso, rectificación y cancelación de datos.

A pesar de todos los desafíos previamente expuestos,

“El mayor reto para el sector inmobiliario europeo es planificar y construir cosas en las que la gente realmente quiera estar, no nosotros, sino la generación que nos sigue”

(PWC; ULI, 2017)

4. CASO PRÁCTICO DE VALORACIÓN DE INMUEBLES CON AVMs

4.1 Selección del área de estudio

El caso de estudio será el distrito de Chamartín, siendo uno de los 21 distritos que forman la ciudad de Madrid en España. Este distrito está ubicado al norte del centro de la ciudad, y está compuesto por los barrios de Castilla, Nueva España, Bernabéu, El Viso, Ciudad Jardín y Prosperidad.

Figura 6: Distrito de Chamartín por barrios



Fuente: (Idealista, 2024)

Este distrito se caracteriza por albergar una de las principales estaciones de tren y un intercambiador que integra diversas líneas de autobuses y metro, ofreciendo una excelente conectividad con otros puntos de Madrid y España. Además, se encuentra ubicado estratégicamente por su proximidad al complejo de las Cuatro Torres, un destacado centro de negocios, y el Estadio Bernabéu, prestigioso hogar del Real Madrid. Todo esto unido a la combinación de viviendas de alta calidad, accesibilidad, servicios, colegios de prestigio y zonas verdes contribuyen a hacer de Chamartín una de las áreas más deseables para vivir en Madrid.

Adicionalmente, el distrito de Chamartín se caracteriza por su heterogeneidad entre sus

distintos barrios, debido a diferencias demográficas, sociales, culturales y económicas. Desde sus orígenes ya presentan diferencias significativas estos barrios, siendo Castilla el más atípico por sus estratos sociales más variados (Flores Guillén, 2016). Por ello, los barrios que conforman Chamartín son una fuente de estudio interesante para la aplicación de AVMs.

4.2 Objetivos y Metodología

El objetivo del presente estudio es realizar una valoración del mercado inmobiliario del distrito de Chamartín aplicando AVMs. Más concretamente, se utilizarán las metodologías de precios hedónicos, KNN y la combinación de ambos, por ser los únicos métodos permitidos de forma oficial en España. Además, este estudio tiene como propósito mostrar de forma práctica los conceptos mencionados anteriormente, aportando un análisis detallado de la precisión de los distintos modelos.

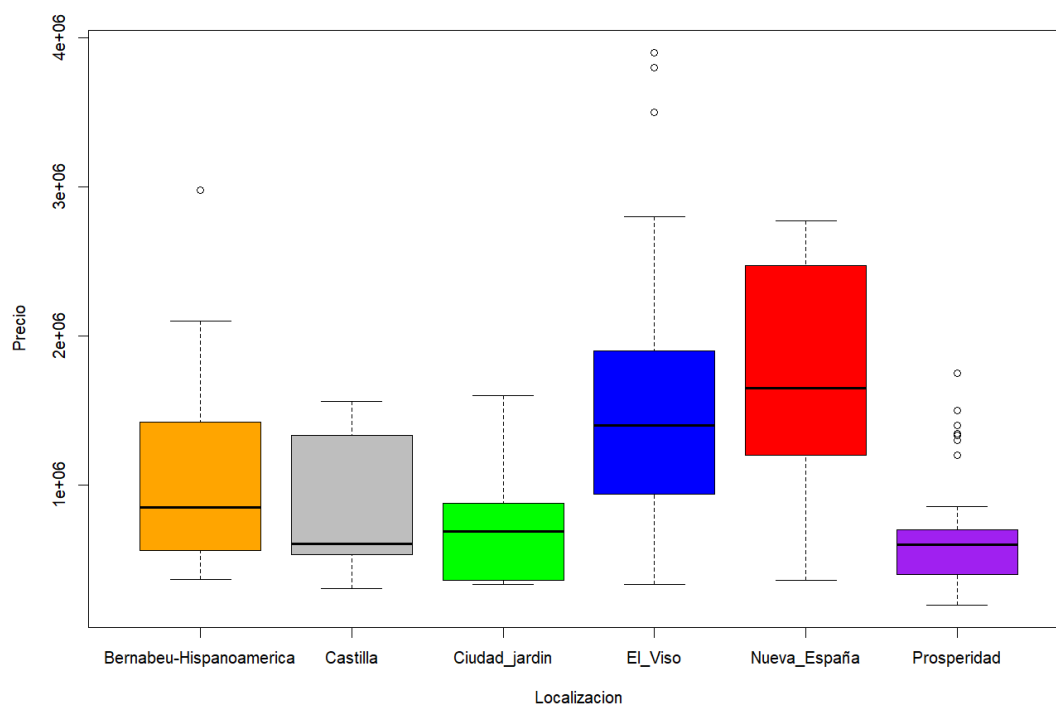
Para ello, se han obtenido los datos de Idealista, portal inmobiliario líder en España. Estos datos se han seleccionado de forma manual, debido a las 'reCAPTCHA'. Estos, son sistemas clásicos que diferencian entre humanos y ordenadores, impidiendo extraer los datos mediante la técnica de *Web scraping*. Por ello, se ha limitado el alcance del caso práctico a 284 observaciones y 15 variables (Anexo A).

Respecto a las herramientas empleadas en el modelo, se utilizó Microsoft Excel para la recopilación y preprocesamiento inicial de los datos. Posteriormente, la creación de gráficos, la identificación y eliminación de valores atípicos, la selección de variables, así como el desarrollo y análisis de los modelos se llevaron a cabo con R Studio.

4.3 Análisis preliminar de los datos

Como ya hemos mencionado anteriormente, el distrito de Chamartín es caracterizado por su heterogeneidad. Esta característica se refleja de forma notoria en la composición social y económica de sus barrios, afectando en la valoración de la propiedad. Se ha realizado un Diagrama de Cajas y Bigotes, para analizar estas tendencias en el precio e identificar y eliminar los valores atípicos de nuestra base de datos.

Figura 7: Diagrama de Caja y Bigotes por Barrio



Fuente: Elaboración propia

En primer lugar, se aprecia que las propiedades de mayor valor se localizan en El Viso y Nueva España, distinguiéndose como las zonas más caras y exclusivas de Madrid. Estos barrios son predominantemente habitados por las clases altas y se caracterizan por sus grandes chalés y apartamentos de lujo. No obstante, es importante destacar que el espectro de precios en estas áreas es el más amplio, lo que permite hallar propiedades cuyos valores se asemejan a los encontrados en los otros barrios.

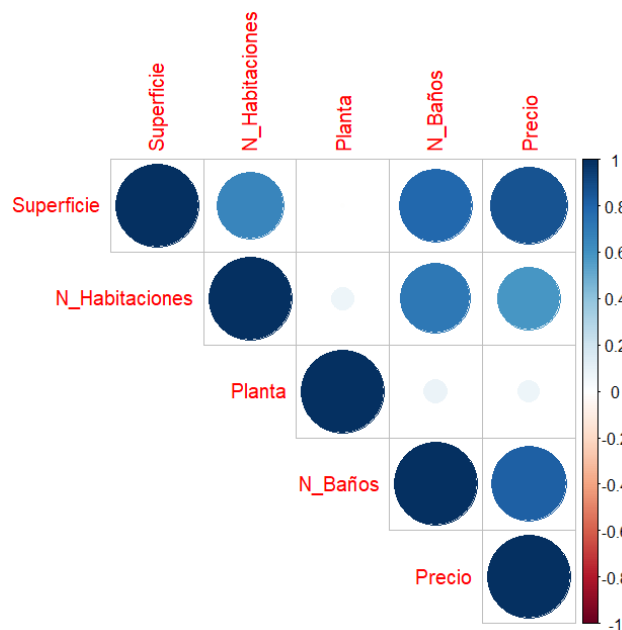
En segundo lugar, el barrio de Bernabéu cuenta con propiedades de alto valor por su ubicación cercana al estadio Santiago Bernabéu y a sus numerosos restaurantes, bares y comercios. Además, la población se caracteriza por su diversidad, acogiendo tanto a familias como a jóvenes, y ofreciendo una amplia variedad de tipos de viviendas.

En tercer lugar, los barrios de Ciudad Jardín y Castilla tienen a atraer individuos de clase media y media-alta. En estas áreas existe una notable variabilidad en su demografía, traducándose en un amplio rango de precios en el mercado inmobiliario adaptándose tanto a opciones más accesibles como a propiedades de mayor valor.

Por último, el barrio de Prosperidad destaca por ser el más asequible del distrito, ofreciendo una mezcla de viviendas antiguas y modernas. Esta área alberga una diversa comunidad en cuanto a estratos sociales, abarcado tanto familias trabajadoras como a profesionales jóvenes.

Adicionalmente, se ha realizado un análisis bivalente básico, representado en la Figura 7 la matriz de correlaciones entre las variables numéricas. Este análisis permite estudiar cuáles son las variables más correlacionadas con el precio del inmueble, que es la variable que se intenta predecir. La figura muestra que la superficie tiene la mayor correlación con el precio (87%), seguido por el número de baños (82%), el número de dormitorios (59%) y el número de planta (7%). Cabe destacar que en este análisis no aparecen las variables categóricas, pudiendo tener una posible relación con el precio.

Figura 8: Matriz de Correlaciones



Fuente: Elaboración propia

4.4 Aplicación de los AVM

4.4.1 Modelo Hedónico

La primera metodología empleada va a consistir en la aplicación de modelos hedónicos por su capacidad de interpretación y selección de las variables en el modelo. Esta metodología permite descomponer las variables y conocer cómo contribuye en el precio

final, pudiendo identificar las variables más significativas que afectan al valor de las propiedades en el distrito. Además, se han propuesto las siguientes hipótesis a la hora de crear los distintos modelos:

1. Las propiedades con mayor superficie, número de habitaciones y baños, y servicios adicionales como piscina, trastero y garaje, se valoran significativamente más alto que las propiedades que carecen de estas características.
2. La superficie y el precio exhiben un patrón logarítmico, de tal manera que el precio experimenta una disminución conforme la superficie de la propiedad se incrementa de manera significativa.
3. Las propiedades ubicadas en pisos superiores y no equipados con ascensor presentan precios más bajos en comparación con aquellos que poseen esta comodidad. Se hipotetiza que existe una interacción entre el número de planta y la presencia de ascensor, la cual impacta en el valor del inmueble.

Adicionalmente, para evitar la colinealidad perfecta se han convertido las variables categóricas (Localización, Tipo, Exterior, Ascensor, Garaje, Terraza, Trastero, Zonas Verdes y Piscina) en dicotómicas, introduciendo en el modelo todas excepto la categoría de referencia. Respecto a los valores atípicos, se han identificado 11 de ellos y han sido eliminados de la base de datos. En los tres modelos propuestos a continuación, se han empleado todas las variables de la base de datos, para tener una visión explicativa de cómo las variables independientes afectan a la variable Precio en los distintos casos.

En los siguientes modelos, las variables van a tener asociados unos coeficientes, como se mencionó en la teoría, que miden el cambio esperado en la variable dependiente por cada unidad de cambio en la variable independiente correspondiente. Estos coeficientes tienen asociados un error estándar que muestra la precisión de la estimación de los coeficientes y un p-valor que indica la probabilidad de obtener estos resultados si su beta real fuese cero, es decir, se busca que las variables tengan un p-valor inferior al 10% para que sean consideradas significativas con un 90% de confianza. Además, los modelos ofrecerán unas métricas denominadas bondad de ajuste, que están formadas por: R^2 que especifica la proporción de la variabilidad en la variable precio que puede ser explicada por las variables independientes y R^2 ajustado que penaliza el número de predictores del modelo, proporcionando una mejor estimación.

Modelo 1: Regresión lineal

El primer modelo presenta la siguiente estructura:

$$P = \beta_0 + \beta_1 X_{\text{Superficie}} + \beta_2 X_{\text{N_Habitaciones}} + \dots + \beta_{17} X_{\text{Piscina_Si}} + \varepsilon$$

La construcción del primer modelo se fundamenta en la primera hipótesis, la cual propone que un aumento en la superficie, el número de habitaciones y baños, así como la disponibilidad de servicios adicionales, conlleva a un incremento en el valor de la propiedad. Estas premisas indican que los valores de los coeficientes asociados a las variables indicadas tienen que ser positivos.

Tabla 2: Resumen del Modelo 1

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-92031.8	88025.3	-1.046	0.296703
Superficie	3563.9	219.3	16.248	< 2e-16 ***
N_Habitaciones	-15027.5	16752.8	-0.897	0.370496
Planta	22728.0	5091.4	4.464	1.17e-05 ***
N_Baños	166652.2	22128.5	7.531	7.26e-13 ***
Localizacion_Bernabeu-Hispanoamerica`	198174.5	42668.3	4.645	5.29e-06 ***
Localizacion_Castilla	-122778.4	63735.5	-1.926	0.055088 .
Localizacion_Ciudad_jardin	131310.2	67257.3	1.952	0.051910 .
Localizacion_El_Viso	416645.0	44609.4	9.340	< 2e-16 ***
Localizacion_Nueva_España	289046.8	49829.6	5.801	1.81e-08 ***
Tipo_Duplex	198084.6	76533.7	2.588	0.010161 *
Tipo_Piso	8220.5	57853.6	0.142	0.887112
Exterior_Si	-107792.3	48976.5	-2.201	0.028574 *
Ascensor_Si	-65002.3	60823.3	-1.069	0.286139
Garaje_Si	120071.8	31886.6	3.766	0.00203 ***
Terraza_Si	35286.7	27776.0	1.270	0.205015
Trastero_Si	-14872.9	30676.5	-0.485	0.628182
Zonas_Verde_Si	-7826.5	33386.8	-0.234	0.814835
Piscina_Si	90474.8	42132.6	2.147	0.032637 *

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 216900 on 275 degrees of freedom
 Multiple R-squared: 0.9051, Adjusted R-squared: 0.8988
 F-statistic: 145.6 on 18 and 275 DF, p-value: < 2.2e-16

Según se muestra en la Tabla 2, existe evidencia de que, a mayor superficie, el número de baños y la disponibilidad de una piscina y garaje, contribuyen al incremento del valor de la propiedad. Sin embargo, no existe evidencia de que el número de habitaciones y la disponibilidad de ascensor, terraza, trastero y zonas verdes influyan en la variable dependiente.

Modelo 2: Regresión no lineal con efectos logarítmicos

El segundo modelo presenta la siguiente estructura:

$$\ln(P) = \beta_0 + \beta_1 \ln(X_{\text{Superficie}}) + \beta_2 X_{\text{N_Habitaciones}} + \dots + \beta_{17} X_{\text{Piscina_Si}} + \varepsilon$$

El segundo modelo se fundamenta en la premisa de que, no existe una relación lineal entre el precio y la superficie, sino que es de carácter logarítmico. Por lo tanto, en la construcción del modelo se han incluido los logaritmos neperianos en las variables influenciadas, indicando que el precio del inmueble tiende a disminuir a medida que la variable superficie aumenta de forma significativa.

Tabla 3: Resumen del Modelo 2

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.816433	0.161400	60.820	< 2e-16	***
log(Superficie)	0.682006	0.038904	17.530	< 2e-16	***
N_Habitaciones	-0.009732	0.015001	-0.649	0.51705	
Planta	0.020531	0.004164	4.931	1.42e-06	***
N_Baños	0.094348	0.018151	5.198	3.94e-07	***
Localizacion_Bernabeu-Hispanoamerica`	0.277521	0.035678	7.778	1.48e-13	***
Localizacion_Castilla	-0.070469	0.052595	-1.340	0.18140	
Localizacion_Ciudad_jardin	0.156867	0.055331	2.835	0.00492	**
Localizacion_EL_Viso	0.429268	0.037487	11.451	< 2e-16	***
Localizacion_Nueva_Espana	0.312697	0.041961	7.452	1.20e-12	***
Tipo_Piso	-0.068295	0.047247	-1.445	0.14947	
Tipo_Duplex	0.012325	0.062598	0.197	0.84406	
Exterior_Si	-0.088516	0.040203	-2.202	0.02851	*
Ascensor_Si	0.079556	0.049937	1.593	0.11228	
Garaje_Si	0.118439	0.026251	4.512	9.53e-06	***
Terraza_Si	0.035717	0.022802	1.566	0.11840	
Trastero_Si	-0.012282	0.025421	-0.483	0.62938	
Zonas_Verde_Si	-0.007893	0.027314	-0.289	0.77281	
Piscina_Si	0.052065	0.033811	1.540	0.12473	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.1781 on 275 degrees of freedom					
Multiple R-squared: 0.9234, Adjusted R-squared: 0.9184					
F-statistic: 184.1 on 18 and 275 DF, p-value: < 2.2e-16					

En la Tabla 3 se confirma la segunda hipótesis, evidenciando que la variable superficie influenciada por el logaritmo neperiano, es significativa, junto con una mejora notable en la precisión del modelo debido a un aumento en R^2 y R^2 ajustado.

Modelo 3: Regresión no lineal con interacción

El tercer modelo presenta la siguiente estructura:

$$\ln(P) = \beta_0 + \beta_1 \ln(X_{\text{Superficie}}) + \beta_2 X_{\text{N_Habitaciones}} + \dots + \beta_{17} X_{\text{Piscina_Si}} + \beta_{18} X_{\text{Ascensor_Si}} * X_{\text{Planta}} + \epsilon$$

El tercer modelo se ha basado en la segunda y tercera hipótesis, que establece adicionalmente al modelo 2 una relación entre el número de la planta del inmueble y si existe un ascensor. En otras palabras, las propiedades situadas en plantas altas que carecen de ascensor tienden a tener precios inferiores en comparación con aquellas que disponen de este servicio.

Tabla 4: Resumen del Modelo 3

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.881651	0.165576	59.681	< 2e-16 ***
Ascensor_Si	0.015539	0.062883	0.247	0.80501
Planta	-0.051470	0.043416	-1.185	0.23685
log(Superficie)	0.680877	0.038785	17.555	< 2e-16 ***
N_Habitaciones	-0.007005	0.015043	-0.466	0.64183
N_Baños	0.091952	0.018150	5.066	7.47e-07 ***
Localizacion_Bernabeu-Hispanoamerica`	0.274951	0.035597	7.724	2.13e-13 ***
Localizacion_Castilla	-0.075141	0.052501	-1.431	0.15350
Localizacion_Ciudad_jardin	0.154198	0.055176	2.795	0.00556 **
Localizacion_El_Viso	0.425788	0.037425	11.377	< 2e-16 ***
Localizacion_Nueva_Espana	0.308940	0.041887	7.376	1.95e-12 ***
Tipo_Piso	-0.071672	0.047139	-1.520	0.12955
Tipo_Duplex	0.011120	0.062401	0.178	0.85869
Exterior_Si	-0.082832	0.040219	-2.060	0.04039 *
Garaje_Si	0.114960	0.026250	4.379	1.69e-05 ***
Terraza_Si	0.037003	0.022742	1.627	0.10487
Trastero_Si	-0.007573	0.025497	-0.297	0.76669
Zonas_Verde_Si	-0.010397	0.027267	-0.381	0.70327
Piscina_Si	0.053942	0.033721	1.600	0.11083
Ascensor_Si:Planta	0.072515	0.043527	1.666	0.09686 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1775 on 274 degrees of freedom
Multiple R-squared: 0.9242, Adjusted R-squared: 0.9189
F-statistic: 175.7 on 19 and 274 DF, p-value: < 2.2e-16

En última instancia, la Tabla 4 verifica la tercera hipótesis, mostrando evidencia de una correlación positiva entre la presencia de ascensor y el número de planta del inmueble. Sin embargo, esta nueva variable de interacción muestra ser significativa solo un 9%, lo cual implica que, dependiendo del nivel de confianza deseado para el modelo, podría ser excluida o mantenida dentro del mismo.

Comparación de modelos hedónicos

En la Tabla 5 se ofrece un resumen comparativo de los tres modelos propuestos, destacando sus respectivas bondades de ajuste, el error estándar asociado y las variables que resultaron ser estadísticamente significativas. Se destaca que el Modelo 3 sobresale como el más explicativo entre los propuestos, evidenciado por sus valores superiores en los coeficientes de determinación R^2 y R^2 ajustado, junto con el error estándar más bajo.

Tabla 5: Resumen de los modelos hedónicos

<i>Modelos</i>	<i>R²</i>	<i>R² ajustado</i>	<i>Error estándar</i>	<i>Variables significativas</i>
<i>Modelo 1</i>	0,9051	0,8988	216900	Superficie, Planta, N_Baños, las disntitas localizaciones, Tipo_Dúplex, Exterior_Si, Garaje_si y Piscina_Si
<i>Modelo 2</i>	0,9234	0,9184	178100	Log(superficie), Planta, N_Baños, las localizaciones excepto Castilla, Exterior_Si y Garaje_ si
<i>Modelo 3</i>	0,9242	0,9189	177500	Log(superficie), Planta, N_Baños, las localizaciones excepto Castilla, Exterior_Si y Garaje_ si

Selección de variables

La metodología de precios hedónicos tiene la capacidad de seleccionar de forma automática las variables relevantes mediante la Regresión *Stepwise*. Este procedimiento consiste en agregar y eliminar iterativamente las variables del modelo, buscando el conjunto óptimo de predictores, que mejor se ajuste a los datos con el objetivo de minimizar la estimación del error. Además, este procedimiento se tiene que llevar a cabo en el conjunto de entrenamiento del Modelo 1 debido a que incorpora todas las variables en su forma original. Se ha realizado la partición del conjunto de datos (70/30), y se han concluido que las variables significativas son las siguientes: superficie, número de baños y planta, las distintas localizaciones, si es un dúplex, si es exterior y si contiene garaje y piscina.

Tabla 6: Modelo por Pasos para la selección de variables

```

Residuals:
    Min       1Q   Median       3Q      Max
-696494 -129901  14063  124967  720664

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -149292.1   64217.5  -2.325  0.021129 *
Superficie      3351.6     244.1   13.731  < 2e-16 ***
Planta         27624.2     6339.3    4.358  2.14e-05 ***
N_Baños       158156.3    23984.9    6.594  4.04e-10 ***
`Localizacion_Bernabeu-Hispanoamerica`  191607.7    51084.8    3.751  0.000233 ***
Localizacion_Castilla -157458.3    77972.4   -2.019  0.044835 *
Localizacion_Ciudad_jardin  130692.8    78408.6    1.667  0.097181 .
Localizacion_El_Viso    395316.5    51965.0    7.607  1.22e-12 ***
Localizacion_Nueva_Espana  299014.1    58721.0    5.092  8.42e-07 ***
Tipo_Duplex    199495.6    69641.0    2.865  0.004640 **
Exterior_Si   -89810.4    60431.5   -1.486  0.138879
Garaje_Si     110128.1    37666.7    2.924  0.003874 **
Piscina_Si    115292.6    48427.4    2.381  0.018257 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 221600 on 192 degrees of freedom
Multiple R-squared:  0.9054,    Adjusted R-squared:  0.8995
F-statistic: 153.1 on 12 and 192 DF,  p-value: < 2.2e-16

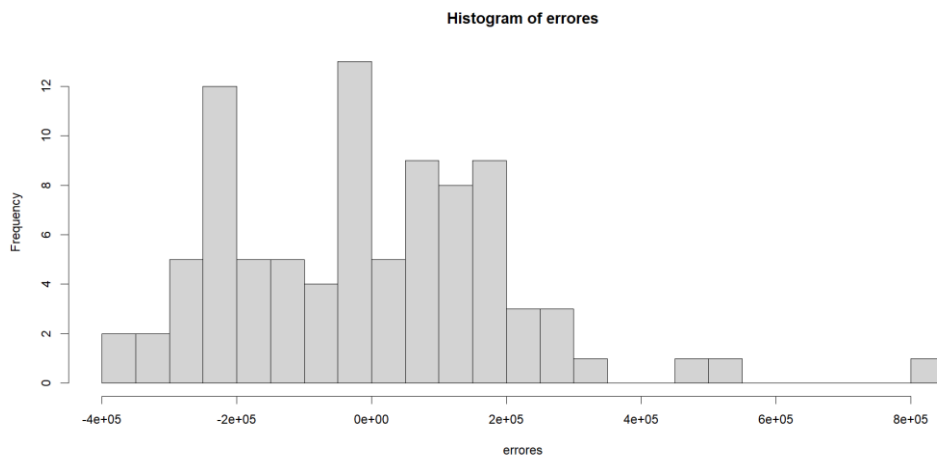
```

Modelo por pasos o ‘StepModel’

Pese a que se han discutido otros modelos hedónicos a lo largo del caso práctico, el modelo que se utilizará para el análisis del rendimiento del algoritmo será el Stepmodel por su relevancia en la selección de variables y simpleza. Además, para medir el rendimiento del algoritmo se van a estimar los errores producidos en la muestra de control (30%), existiendo diferentes métricas para calcularlo: Error Medio (ME), Raíz de error cuadrada media (RMSE), Error absoluto porcentual medio (MAPE) y Error Medio absoluto (MAE). El ME indica si el modelo tiende a sobreestimar o infraestimar los valores. En este caso, con un ME de -5.508,272 euros, se observa que el modelo tiene una tendencia a sobreestimar las predicciones. Además, el modelo en cuestión presenta un RMSE de 207.419,8 euros, un MAE de 161.885,7 euros y un MAPE de 18,12214 euros.

Adicionalmente, se ha representado en un histograma los errores de este modelo pudiéndose comprobar que el modelo tiende a infravalorar en algunas situaciones. La mayor frecuencia se encuentra en 0 y se aprecia un valor atípico de 800.000 euros, indicando la necesidad de un mayor estudio de la posible causa de esa predicción.

Figura 9: Histograma de errores Stepmodel



Fuente: Elaboración propia

4.4.2 K-Vecinos cercanos

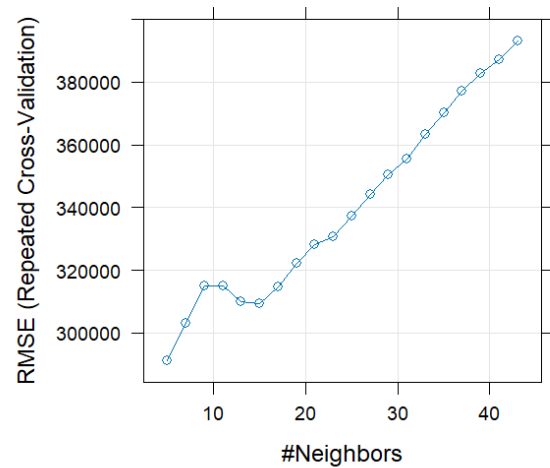
En primer lugar, se seleccionaron las variables previamente identificadas en el Stepmodel como significativas y se procedió a su estandarización. Además, el hiperparámetro K puede ser escogido a priori por el investigador, sin embargo, se ha utilizado un proceso de validación cruzada con el propósito de reducir el error de generalización. Para ello, se

ha realizado la partición de los datos (70/30) y se han empleado 10-folds de validación, es decir, el modelo se entrena 10 veces y se repite 3. Tras ejecutar el código en R Studio, el resultado del valor óptimo de k es 5.

Tabla 7: Validación Cruzada KNN

k	RMSE	Rsquared	MAE
5	291277.0	0.8259116	210569.1
7	303129.9	0.8154448	221969.8
9	315025.4	0.8044598	238212.7
11	315027.7	0.8098856	241177.0
13	309930.2	0.8260694	239045.9
15	309462.1	0.8333470	241028.6
17	314572.9	0.8372873	247064.5
19	322385.7	0.8397323	253255.2
21	328169.8	0.8432918	257653.0
23	330740.3	0.8488411	257948.9
25	337222.6	0.8495560	261442.3
27	344152.2	0.8469655	264145.7
29	350617.9	0.8481368	267452.3
31	355607.1	0.8509163	269974.0
33	363293.8	0.8487794	276624.7
35	370317.6	0.8477926	281671.8
37	377169.0	0.8476574	284903.3
39	382701.9	0.8451775	288646.1
41	387040.0	0.8475131	291045.5
43	393049.2	0.8447362	295180.2

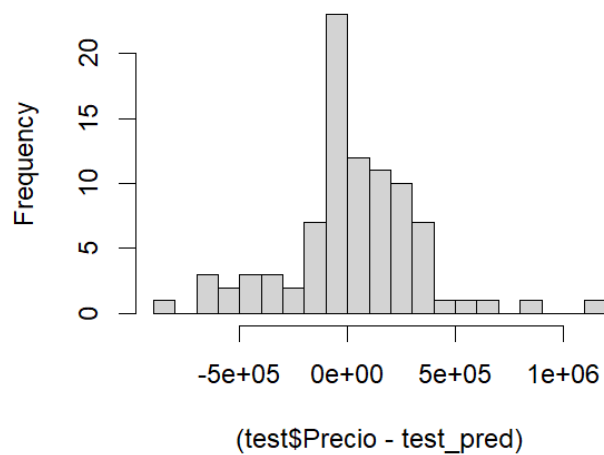
Figura 10: Valores óptimos de k



Fuente: Elaboración propia

Para verificar el modelo, se ha calculado las estimaciones con el conjunto de control (30%). Además, se ha realizado un histograma de los errores en la predicción para entender cómo se distribuyen en el modelo. Se destaca que, la mayor frecuencia se encuentra en 0 euros y se aprecian dos valores atípicos por infravaloración. Sin embargo, este modelo, sobrestima e infraestima siendo superior en la infravaloración, indicando un problema de sesgo que puede afectar a la utilidad del modelo. También es relevante mencionar que el RMSE experimentó un incremento al valor previsto durante la validación cruzada, lo cual se atribuye a la selección aleatoria de los datos del conjunto de control, resultando en un valor de 310.210 euros.

Figura 11: Histograma de los errores en KNN
Histogram of (test\$Precio - test_pred)



Fuente: Elaboración propia

4.4.3 Modelo Mixto

Este modelo mixto se va a realizar mediante la técnica avanzada de *Ensemble Stacking* o apilamiento de conjuntos que combina múltiples modelos para mejorar la precisión de las predicciones. En otras palabras, en lugar de utilizar un único modelo predictivo, el stacking utiliza varios modelos base, en este caso el modelo de pasos y KNN. Esta metodología se emplea para mejorar las predicciones reduciendo el sesgo y la varianza simultáneamente, a pesar de incurrir en un mayor coste computacional y mayor complejidad. En la tabla a continuación se muestra el mínimo, el primer cuartil, la mediana, la media, el tercer cuartil, el máximo y si existen valores perdidos en el modelo en las variables de precisión.

Tabla 8: Resumen del Modelo Mixto

```

Models: glm, knn
Number of resamples: 20

MAE
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
glm 107610.5 151565.8 187191.7 180706.4 211874.2 227478.5    0
knn 143046.9 186330.8 215546.0 213163.1 239409.3 300404.9    0

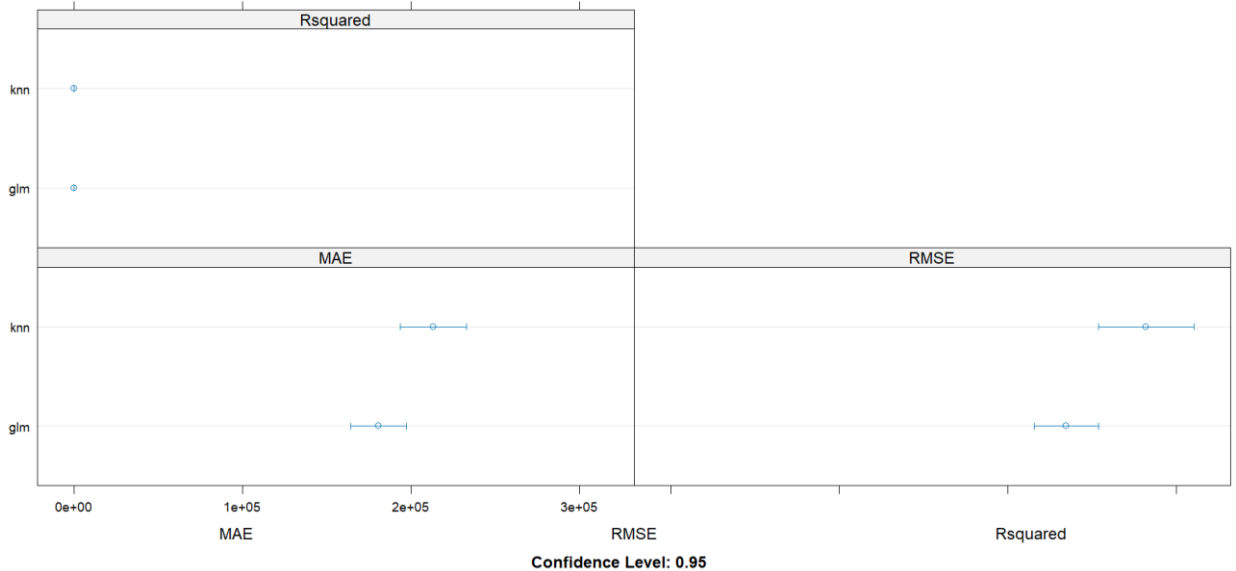
RMSE
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
glm 150477.6 202306.1 245589.4 234767.7 270967.6 288835.8    0
knn 178345.5 249554.4 279385.5 282090.5 311172.1 415875.0    0

Rsquared
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
glm 0.8413624 0.8788852 0.9015070 0.8991822 0.9228001 0.9598369    0
knn 0.6571972 0.8237096 0.8535728 0.8471205 0.8837906 0.9459973    0

```

Para facilitar la visualización de los datos anteriores, en la figura 8 se representan las medias del R^2 , MAE y RMSE obtenidas mediante el Ensemble Stacking:

Figura 12: Métricas de Precisión Modelo Mixto



Fuente: Elaboración propia

Por último, este modelo permite combinar las distintas predicciones en un único conjunto, obteniendo los siguientes resultados: un RMSE de 224.170,2, un R^2 de 0,8972258 y un MAE de 167.347 euros.

4.5 Resultados

El presente caso práctico tenía la finalidad de aplicar la teoría mostrada en un caso real del Distrito de Chamartín. Se han empleado las metodologías de precios hedónicos, KNN y la combinación de ambas, obteniendo los resultados explicados a continuación.

En primer lugar, se ha empleado la metodología de precios hedónicos para establecer tres posibles hipótesis que influyen en el valor del inmueble y determinar las variables significativas del modelo. Respecto a las tres hipótesis planteadas, la primera hipótesis se confirma parcialmente al no existir evidencia de que el número de habitaciones y la disponibilidad de ascensor, terraza, trastero y zonas verdes influyan en el modelo. Además, se confirman la segunda y tercera hipótesis al evidenciar que existe una relación

logarítmica entre el precio y la superficie, y una interacción entre el número de planta del inmueble y si dispone de un ascensor. Finalmente, se ha realizado una selección de variables mediante la metodología de regresión *Stepwise* concluyendo que las variables significativas del modelo son superficie, número de baños y planta, las distintas localizaciones, si es tipo dúplex y dispone de garaje y piscina. Este Modelo de Pasos se caracteriza por sobreestimar e infraestimar mostrando una tendencia superior a la sobreestimación. Además, el modelo en cuestión presenta un RMSE de 207.419,8 euros, un MAE de 161.885,7 euros y un MAPE de 18,12214 euros.

En segundo lugar, se ha realizado la metodología de KNN utilizando las variables obtenidas en la metodología anterior. Además, mediante una validación cruzada se ha seleccionado el número óptimo de k , concluyendo que era 5. Este modelo se caracteriza por sobreestimar e infraestimar, teniendo una inclinación a infraestimar. En esta metodología se muestra también un problema de sesgo que puede afectar a la utilidad del modelo. Además, en la validación del modelo influenciada por la muestra de control seleccionada se obtuvo un RMSE de 310.210 euros.

En el último análisis, se realizó una combinación de ambas metodologías mediante *Ensemble Stacking* obteniendo un RMSE de 224.170,2, un R^2 de 0,8972258 y un MAE de 167.347 euros.

En base a los resultados expuestos, la metodología de precios hedónicos muestra unas medidas de precisión superiores al resto. Sin embargo, cabe destacar que, la combinación de ambas debería aportar una mayor precisión, pero la metodología KNN influye negativamente al poseer un RMSE superior en 102.790,2 euros que la primera metodología. Pese a ello, la combinación de ambos ha alcanzado un RMSE de 224.170 euros y un R^2 ligeramente superior al Stepmodel.

5. CONCLUSIÓN

El fenómeno del Big Data y la Inteligencia Artificial han transformado radicalmente las economías mundiales, introduciendo innovaciones en el sector inmobiliario a través del *PropTech*. Sin embargo, la mayor incógnita sigue siendo cómo conseguir la máxima precisión en la valoración de bienes raíces, tanto de forma tradicional como automatizada. Tras la crisis del 2008, las Sociedades de Tasación establecieron estas nuevas tecnologías como medidas de refuerzo para ofrecer objetividad, estandarización y precisión en las valoraciones. En la actualidad, los AVM no han conseguido ser estandarizados, y por consecuencia ser precisos en la fijación de los precios.

A lo largo de este Trabajo de Fin de Grado, se ha analizado cómo debería ser el proceso de valoración automatizada existiendo varias áreas de mejora durante las distintas etapas. En primer lugar, los *smarts contracts* siguen siendo una tecnología incipiente en el sector, sin tener un proceso definido de aplicación. En segundo lugar, la investigación sin intervención de un tasador despierta dudas sobre si existen las tecnologías necesarias para capturar todos los datos relevantes del inmueble y si no es necesaria la presencia de una profesional para la validación de la información. Por último, existe una falta de estandarización durante el proceso, especialmente en la recolección de los datos. Estos modelos están muy influenciados por la calidad de los datos, pudiendo divergir las estimaciones en gran medida por ellos. En el caso práctico es muestra de ello, existiendo valores atípicos en la base de datos, y en caso de haber escogido otra base distinta, las estimaciones resultantes podrían haber variado. Esto destaca la relevancia de la investigación, selección y el tratamiento de los datos en el proceso de análisis. Además, las variables del modelo explican al 90% la variabilidad del modelo, reflejando que las características principales del inmueble influyen al valor. Los modelos planteados se caracterizan por sobreestimar e infraestimar, existiendo valores atípicos en las estimaciones, dando lugar a un amplio rango de mejora en esta área.

Por último, tras la realización de este trabajo surgen propuestas de estudio para futuros proyectos en relación con la evaluación y aplicación de los AVM, destacando los siguientes:

- Estudio de las tecnologías de recolección masiva de datos sobre inmuebles, y la

influencia de LOPD sobre estas.

- Estudio de la viabilidad del uso de las nuevas tecnologías de investigación sin la intervención de un tasador.
- Definición del proceso de valoración automatizada y su posible aplicación en el sector.
- Aplicación de las distintas metodologías del Big Data y la Inteligencia Artificial en un caso práctico.

6. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Miriam Portillo Quesada, estudiante de E2+ Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Evaluación y Aplicación de Modelos Automatizados de Valoración (AVM)", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
3. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
4. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
5. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 18/03/2024

Firma: Miriam Portillo Quesada

7. REFERENCIAS BIBLIOGRÁFICAS

- AEV. (2019). *Estándar sobre valoración de inmuebles mediante modelos automatizados (AVM)*. Asociación Española de Análisis de Valor .
- Agrasar González, B. (2019). *Valoración Inmobiliaria a través de automated valuation models (AVMs)*. Universidad Pontificia de Comillas.
- AltoDirectivo. (2021). Los Modelos de Valoración Automatizada experimenta un crecimiento exponencial y se consolidan en el sector inmobiliario. *Alto Directivo*. Retrieved marzo 1, 2024, from <http://www.altodirectivo.com/secciones/27534/los-modelos-de-valoracion-automatizada-experimentan-un-crecimiento-exponencial-y-se-consolidan-en-el-sector-inmobiliario>
- Antón-Merino, A., & Garrido-Villén, N. (2016). La valoración judicial de la prueba pericial en la delimitación de la propiedad inmobiliaria. *Revista del Ilustre Colegio Oficial de Ingenieros Técnicos en Topografía*, 33(171), 123-130.
- Aznar Bellver, J., & Guijarro Martínez, F. (2020). Nuevos métodos de valoración. Modelos multicriterio. *Universitat Politècnica de València*.
- Becerra Lois, F., Pons García, R., & Ruíz Domínguez, R. (2020). Procedimiento para evaluar el impacto de proyectos de desarrollo local de tipo económico en un territorio. *Universidad y Sociedad*, 12(3), 312-318.
- Bondt, G., Gieseck, A., & Tujula, M. (2020). *Household wealth and consumption in the euro area*. European Central Bank.
- Canadas, R. (2022, julio 15). *Redes Artificiales/ Qué son y cómo funcionan*. Retrieved febrero 28, 2024, from Abdatum: <https://abdatum.com/tecnologia/redes-neuronales-artificiales>
- Caridad, J. M., Ceular Villmandos, N., & Núñez Tabales, J. M. (2008). *Metodología de precios hedónicos vs Redes Neuronales Artificiales como alternativa a la valoración de inmuebles. Un caso Real*. Universidad de Córdoba.
- Court, A. (1939). *Hedonic price indexes with automotive examples*.
- Downie, M.-L., & Robson, G. (2008). Automated Valuations Models: an international perspective. *RICS Automated Valuations Models Conference: AVMs Today and Tomorrow*. Londres: RICS.
- ESRI. (n.d.). *Análisis espacial en ArcGis Pro/ Documentación*. Retrieved marzo 5, 2024, from ESRI: <https://pro.arcgis.com/es/pro-app/latest/help/analysis/introduction/spatial-analysis-in-arcgis-pro.htm>
- ESRI. (n.d.). *Historia de los SIG*. Retrieved febrero 20, 2024, from esri.com:

<https://www.esri.com/es-es/what-is-gis/history-of-gis>

- Fabra Garcés, L. A., & Sevillano Rubio, A. (2000). La valoración inmobiliaria [ambitos de aplicación]. *Revista de relaciones laborales*, 8, 11-30.
- Fernández Agüero, C., Fernández Cuesta, I., & Moya, D. (2008). *Valoración de inmuebles mediante técnicas de lógica difusa*. Universidad Complutense de Madrid.
- Flores Guillén, A. I. (2016). *La aplicación de métodos estadísticos en la valoración masiva de inmuebles: una propuesta de regulación en el marco normativo del Banco de España*. Tésisi Doctoral, Universidad Politécnica de Madrid.
- Gallego Mora-Esperanza, J. (2004). La Inteligencia Artificial aplicada a la valoración de inmuebles: un ejemplo para valorar Madrid. *Catastro*(50), 51-69.
- Gallego Mora-Esperanza, J. (2008). Modelos de valoración Automática. *Catastro*(62), 7-26.
- Gila García, A., & Novás Filguera, M. (2012). El uso del método hedónico para ajustar los cambios de calidad: experiencia del IPV. *Estadística Española*(179), 299-310.
- Guijarro Martínez, F. (2023). Valoración automática de inmuebles residenciales mediante modelos de machine learning. *Revista de Estudios Empresariales, Segunda época*, 2(2023), pp. 27-39. doi:<https://doi.org/10.17561/ree.n2.2023.7823>
- Huerta, R., & Yépez, F. (2019). Análisis de imágenes multiespectrales para el cálculo de la calidad de áreas verdes urbanas. Caso de estudio: Zona Metropolitana de Monterrey. *Reunión Anual de la Unión Geofísica Mexicana*. Puerto Vallarta: CICESE.
- IAAO. (2018). *Automated Valuations Models (AVMs). A criterion for measuring fairness, quality, equity and accuracy*. International Association of Assessing Officers.
- Iberdrola. (n.d.). *Smart Contracts: contratos inteligentes para formalizar acuerdos en la era digital*. Retrieved febrero 10, 2024, from Iberdrola: <https://www.iberdrola.com/innovacion/smart-contracts>
- Idealista. (2024, Marzo 13). *Idealista Mapa Chamartín*. Retrieved Marzo 13, 2024, from Idealista: <https://www.idealista.com/venta-viviendas/madrid/chamartin/mapa>
- INE. (n.d.). *PIB pm Oferta (Precios corrientes) (30678)*. Retrieved marzo 3, 2024, from Instituto Nacional de Estadística: <https://www.ine.es/jaxiT3/Datos.htm?t=30678>
- IVSC. (2022). *International Valuations Standards*. London: International Valuation Standards Council.
- Jaureguizar Oriol, I. (2017). *Aplicación de valoración hedónica. Caso de los corrales de Buelna*. Trabajo de Grado, Universidad Politécnica de Madrid.
- Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017, September). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio*

- Management*, 43, pp. 202-211. doi:10.3905/jpm.2017.43.6.202
- Matich, D. J. (2001). *Redes neuronales: conceptos básicos y aplicaciones*. Universidad Tecnológica de Rosario.
- Mckinsey & Company. (2011). *Big Data: The next frontier for innovation, competition and productivity*. Mckinsey Global Institute.
- Molina, D. (2021, Octubre 8). *Qué es el proptech y cómo ha cambiado el sector inmobiliario*. Retrieved marzo 9, 2024, from IEBS: <https://www.iebschool.com/blog/que-es-el-proptech-y-como-ha-cambiado-el-sector-inmobiliario-innovacion/>
- Moreno Jiménez, J. M. (2002). El proceso analítico jerárquico (AHP). Fundamentos, metodologías y aplicaciones. *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 1, 28-77.
- Núñez Tabales, J. M., & Rey Carmona, F. J. (2017). Alternativas para la valoración de inmuebles urbanos. *Revista Publicando*, 1(11), 3-19.
- Núñez Tabales, J. M., Caridad y Ocerin, J. M., & Nuria, C. (2009, enero 9). Estimación del precio de la vivienda mediante redes neuronales artificiales (RNA) en diferentes marcos temporales. *Metodologías de Encuestas*, pp. 79-101.
- Núñez Tables, J. M., Rey Carmona, F. J., & Caridad y Ocerín, J. M. (2017). Redes Neuronales (RN) aplicadas a la valoración de locales comerciales. *Informes de construcción*, 69. doi:10.3989/ic.15.053
- PropTech. (n.d.). *¿Qué es PropTech?* Retrieved marzo 9, 2024, from PROPTech: <https://proptech.es/que-es-proptech/>
- PWC; ULI. (2017). *Emerging Trends in Real State, new markets reality Europe 2017*. PWC & Urban Land Institute.
- Ridker, R., & Henning, A. (1967). The determinants housing prices and the demand for clean air. *Journal Environmental Economy Management*, 81-102.
- Roque López, J. (2021). *Técnicas de selección de variables en regresión lineal múltiple*. Trabajo de Fin de Máster, Universidad de Huelva.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Political Economy*, 82(1), 34-55.
- San José, A. (2020). *El futuro de la valoración inmobiliaria: El Big data y modelos AVM*. Escuela Superior de Arquitectura, Universidad Politécnica de Madrid.
- Scheurwater, S. (2017). *The Future of Valuations. The relevance of real estate valuations for institutional investors and banks- views from a European expert group*. London: The Royal Institution of Chartered Surveyors (RICS).

- Sirmans, G., & Zietz, E. (2005). The composition of hedonic pricing models. *Journal of Real Estate Literature*, 3-43.
- Thind, J. (2017, Mayo 24). *Home Value Estimates: Understanding Their Purpose and Evaluating Their Results*. Retrieved Febrero 18, 2024, from Zillow: <https://www.zillow.com/tech/home-value-estimates/>
- Tyrväinen, L., & Mettinen, A. (2000). Property prices and urban forest amenities. *Journal of environmental economics and management*, 205-223.
- UNIR. (2022). *¿Qué es un sistema experto? Usos y aplicaciones en Inteligencia Artificial*. Retrieved marzo 4, 2024, from UNIR: <https://www.unir.net/ingenieria/revista/sistema-experto/>
- Wei, C., Fu, M., Yang, H., Tang, F., & Xiong, Y. (2022). The Research Development of Hedonic Price Model-Based Real Estate Appraisal in the Era of the Big Data. *Land*, 11(3), 2-30. doi:<https://doi.org/10.3390/land11030334>
- Wilhelmsson, M. (2000). El impacto del ruido del tráfico en los valores de casas unifamiliares. *Revista de planificación y gestión ambiental*, 43(6), 799-815. doi: <https://doi.org/10.1080/09640560020001692>

8. ANEXO

8.1 ANEXO I: Variables del Caso Práctico

Tabla 9: Resumen de las variables

Nombre	Descripción	Tipo	Rango valores	Comentarios
ID_Inmueble	Es la clave única de cada inmueble	Numérica	1-305	Esta variable sirve para identificar a los distintos inmuebles si fuera necesario
Localización	Los distintos barrios del distrito	Factor	Castilla, Nueva España, El Viso, Ciudad Jardín, Bernabéu y Prosperidad	Esta variable se introducirá en el modelo como un set de dicotómicas, surgiendo 6 nuevas variables
Tipo	Los distintos tipos de inmuebles	Factor	Piso, Ático y Dúplex	Esta variable se introducirá en el modelo como un set de dicotómicas, surgiendo 3 nuevas variables
Superficie	Los m ² correspondientes a cada vivienda	Numérica	35-593 m ²	
N_Habitaciones	El número de habitaciones en cada vivienda	Numérica	1-8 habitaciones	
Planta	El número de planta en el que está ubicado el apartamento	Numérica	0-13 planta	La planta 0 es considerada el bajo de los apartamentos, posteriormente será planta 1, etc.

Exterior	Si el inmueble está situado al exterior o no	Factor	Si/No	Esta variable se introducirá en el modelo como una variable dicotómica (1/0)
Ascensor	Si el inmueble cuenta con un ascensor	Factor	Si/No	Esta variable se introducirá en el modelo como una variable dicotómica (1/0)
Garaje	Si el inmueble cuenta con garaje	Factor	Si/No	Esta variable se introducirá en el modelo como una variable dicotómica (1/0)
N_Baños	Número de baños	Numérica	1-6 baños	
Terraza	Si el inmueble cuenta con terraza	Factor	Si/No	Esta variable se introducirá en el modelo como una variable dicotómica (1/0)
Trastero	Si el inmueble cuenta con trastero	Factor	Si/No	Esta variable se introducirá en el modelo como una variable dicotómica (1/0)
Zonas_Verde	Si el inmueble cuenta con zonas verdes	Factor	Si/No	Esta variable se introducirá en el modelo como una variable dicotómica (1/0)
Piscina	Si el inmueble cuenta con piscina comunitaria	Factor	Si/No	Esta variable se introducirá en el modelo como una variable dicotómica (1/0)
Precio	Tasación del inmueble	Numérica	190000-3900000 €	En el modelo será la variable dependiente, es decir, la que buscamos predecir

8.2 ANEXO II: Código de R Studio del Caso Práctico

```
#-----Cambiar directorio e instalación de paquetes-----#

rm(list = ls())
library(corrplot)
library(dplyr)
library(plotly)
library(fastDummies)
library(MASS)
library(ggplot2)
library(lattice)
library(caret)
library(e1071)
library(tidyverse)
library(caretEnsemble)

#-----Visualización y limpieza de los datos-----#

#Subir el archivo excel-->en file e importar dataset en archivo excel

summary(InmueblesChamartin)

# Se selecciona únicamente las variables numéricas para realizar un análisis de correlaciones

Inmuebles_num=InmueblesChamartin[,c(4,5,6,10,15)]
View(Inmuebles_num)
cor(Inmuebles_num)

#Nos muestra un Scatter plot
names(InmueblesChamartin)[2] <- "Localizacion" #Cada color corresponde a un equipo
plot_ly(data = InmueblesChamartin, x = ~Precio, y = ~Superficie, color = ~Localizacion, #Se
mide la correlaci?n entre puntos y salario
      text = ~paste("Id_Inmueble: ", ID_inmueble)) #Permite reconocer cada punto con el
individuo correspondiente

#gráfico para las correlaciones en forma de triángulo

corrplot(cor(Inmuebles_num %>%
```

```

    dplyr::select(Superficie:Precio),
    use = "complete.obs"),
    method = "circle",type = "upper")

#gráfico de caja y bigotes para saber si existen outliers en la muestra

boxplot(InmueblesChamartin$Precio)
boxplot(Precio~Localizacion, InmueblesChamartin,
        col = c("orange", "grey", "green", "blue", "red", "purple"))

# Se eliminan los outliers

InmueblesChamartin=InmueblesChamartin[c(-180,-198,-292,-79,-86,-293,-234,-258,-20,-81,-
11),] #dataset con 156 observaciones

#Se comprueba que ya no existen outliers

boxplot(InmueblesChamartin$Precio)
boxplot(Precio~Localizacion, InmueblesChamartin,
        col = c("orange", "grey", "green", "blue", "red", "purple"))

#-----MODELO EXPLICATIVO-----#

hist(InmueblesChamartin$Precio)

# Se convierten en dicotómicas las variables categóricas

datos= dummy_cols(InmueblesChamartin, select_columns = c('Localizacion', 'Tipo',
'Exterior','Ascensor', 'Garaje', 'Terraza', 'Trastero', 'Zonas_Verde', 'Piscina'))
datos=datos[,c(-1,-2,-3,-7,-8,-9,-11,-12,-13,-14)]

#Para evitar colinealidad perfecta, se elimina la categoría de referencia
datos=datos[,c(-11,-12,-15,-17,-19,-21,-23,-25,-27)]

# MODELO: REGRESIÓN LINEAL

modelo1=lm(Precio~.,data = datos)
summary(modelo1)

```

```

modelo2=lm(log(Precio)~log(Superficie)+N_Habitaciones+Planta+N_Baños+`Localizacion_Bernabeu-
Hispanoamerica`+Localizacion_Castilla+Localizacion_Ciudad_jardin+Localizacion_El_Viso+Local
izacion_Nueva_España+Tipo_Piso+Tipo_Duplex+Exterior_Si+Ascensor_Si+Garaje_Si+Terraza_Si
+Trastero_Si+Trastero_Si+Zonas_Verde_Si+Piscina_Si, data=datos)
summary(modelo2)

```

```

modelo3=lm(log(Precio)~Ascensor_Si*Planta+log(Superficie)+N_Habitaciones+Planta+N_Baños
+`Localizacion_Bernabeu-
Hispanoamerica`+Localizacion_Castilla+Localizacion_Ciudad_jardin+Localizacion_El_Viso+Local
izacion_Nueva_España+Tipo_Piso+Tipo_Duplex+Exterior_Si+Ascensor_Si+Garaje_Si+Terraza_Si
+Trastero_Si+Trastero_Si+Zonas_Verde_Si+Piscina_Si, data=datos)
summary(modelo3)

```

```

#-----Partición de los datos-----#

```

```

set.seed(12)
n<-nrow(datos)
trainIndex<-sample(1:n, 0.7*n, replace=FALSE)
train<-datos[trainIndex,]
test<-datos[-trainIndex,]

```

```

#-----PREDICCIÓN DEL MODELO-----#

```

```

# Se estima el modelo con todas las variables

```

```

fullmodel<-lm(Precio~., data=train)

```

```

# Se aplica el procedimiento stepwise

```

```

stepmodel<-stepAIC(fullmodel, direction="both", trace=1)

```

```

summary(stepmodel)

```

```

# Se obtienen las predicciones en el Test Set

```

```

predicciones=predict(stepmodel, newdata = test)

```

```

# Se calculan los errores de predicción

```

```

errores<-test$Precio-predicciones

```

```

boxplot(errores) # Muestra la gráfica de los errores de predicción

```

```

hist(errores, breaks=20)

```

```

h<-nrow(test) # numero de observaciones en el test

```

```

# cálculo de las medidas de Predictive performance en el test
ME<-sum(errores)/h
RMSE_modelo1<-sqrt(sum(errores^2)/h)
MAE<-sum(abs(errores))/h
MAPE<-sum (abs(errores)/test$Precio)/h*100

print(ME)
print(RMSE_modelo1)
print(MAE)
print(MAPE)

calc_rmse = function(actual, predicted) {
  sqrt(mean((actual - predicted) ^ 2))
}

#-----KNN-----#

# Se establecen controles en el conjunto de entrenamiento
repeats = 3
numbers = 10
tunel = 20
x = trainControl(method = "repeatedcv",
                 number = numbers, # n?mero de folds
                 repeats = repeats)

# Modelo de KNN
set.seed(12)
knn1 <- train(Precio~Superficie+Planta+N_Baños+`Localizacion_Bernabeu-
Hispanoamerica`+Localizacion_Castilla+Localizacion_Ciudad_jardin+Localizacion_El_Viso+Local
izacion_Nueva_España+Garaje_Si+Piscina_Si+Tipo_Duplex, data = train,
             method = "knn",
             preProcess = c("center","scale"),
             trControl
             = x,
             tuneLength = tunel)

```

```

# Resumen del modelo
knn1
plot(knn1)

# Predicciones
test_pred <- predict(knn1, newdata=test, type = "raw")
test_pred

RMSE_knn = calc_rmse(test_pred, test$Precio)

RMSE_knn
hist((test$Precio-test_pred), breaks=20)

#-----ENSEMBLE-----#
#primer nivel
controlensemble <- trainControl(method="repeatedcv", number=10, repeats=3)
algorithmList<- c("glm", "knn")

set.seed(12)
ensemble1 <- caretList(Precio~Superficie+Planta+N_Baños+`Localizacion_Bernabeu-
Hispanoamerica`+Localizacion_Castilla+Localizacion_Ciudad_jardin+Localizacion_El_Viso+Local
izacion_Nueva_Espana+Garaje_Si+Piscina_Si, data=train, trControl=controlensemble,
                      methodList=algorithmList, tuneLength=10)

#Se muestra los resultados del modelo
results <- resamples(ensemble1)
summary(results)

#Gráfica de los dos modelos
dotplot(results)

#segundo nivel

stackControl <- trainControl(method="repeatedcv", number=10,
                             repeats=3)
set.seed(1)
stack.glm <- caretStack(ensemble1, method="glm", metric="RMSE", trControl=stackControl)
print(stack.glm)

```

```
#RMSE de los modelos individuales  
print(RMSE_modelo1)  
print(RMSE_knn)
```