



DOI: <http://dx.doi.org/10.23857/dc.v8i41.2531>

Ciencias Técnicas y Aplicadas
Artículo de Investigación

Técnicas de minería de datos para el análisis de la plusvalía inmobiliaria

Data mining techniques for real estate appreciation analysis

Técnicas de mineração de dados para análise de ganhos de capital imobiliário

Elba Bodero-Poveda ^I

ebodero@unach.edu.ec

<https://orcid.org/0000-0003-3807-5203>

Cristian Morales-Alarcón ^{II}

cmorales@unach.edu.ec

<https://orcid.org/0000-0002-0197-0581>

Ana Elizabeth Congacha-Aushay ^{III}

acongacha@unach.edu.ec

<https://orcid.org/0000-0002-2488-9110>

Cristina Estefanía Ramos-Araujo ^{III}

cristina.ramos@epoch.edu.ec

<https://orcid.org/0000-0002-8644-5814>

Correspondencia: ebodero@unach.edu.ec

***Recibido:** 30 de noviembre de 2021 ***Aceptado:** 30 de diciembre de 2021 * **Publicado:** 28 de enero de 2022

- I. Magister en Tecnología de la Información y Multimedia Educativa. Docente en Ciencias Informáticas, Universidad Nacional de Chimborazo, Riobamba, Ecuador. Doctoranda, Doctorado en Ciencias Informáticas, Universidad Nacional de La Plata, La Plata, Argentina.
- II. Magister en Gestión de Sistemas de Información e Inteligencia de Negocios, Gerente Empresa Esprint. Riobamba, Ecuador.
- III. Magister en Gerencia Informática, Universidad Nacional de Chimborazo, Riobamba, Ecuador.
- IV. Ingeniera en Estadística Informática, Máster de Ciencias en Matemáticas Aplicadas. Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.

Resumen

La minería de datos permite la extracción de información, la clasificación de datos y el procesamiento de los mismos, para conocer su comportamiento, realizar predicciones y para la toma efectiva de decisiones. Esta investigación tiene como objetivo determinar cuáles son las técnicas de minería de datos más utilizadas para el análisis de la plusvalía inmobiliaria, para ello se hace uso de una revisión bibliográfica, análisis de contenido y la aplicación de un diseño de teoría fundamentada. En las investigaciones analizadas se pudo determinar que las técnicas de minería de datos aplicadas para predecir precios en bienes raíces son: Redes Neuronales Artificiales, Análisis Clúster, Árbol de decisión, Modelos de regresión, Modelos estadísticos, Clasificador Bayesiano Ingenuo, Algoritmo K-Neighbors, Algoritmo AdaBoost, evidenciando que la técnica de Redes Neuronales Artificiales es la que mejores resultados presenta. Con respecto a las variables de datos de cada propiedad se encontraron 27 entre las más importantes se presentan el tipo de propiedad, ubicación, precio, área de terreno, área de construcción, número de pisos, número de habitaciones, número de baños, número de estacionamientos o garaje, entre otras, además se analizaron los datos de las publicaciones en redes sociales y sitios web de inmobiliarias de la ciudad de Riobamba. Los resultados proporcionados por esta investigación son útiles para la aplicación de técnicas de minería de datos en investigaciones sobre inversiones inmobiliarias o predicciones del avalúo inmobiliario.

Palabras Clave: Minería de datos, algoritmos, predicciones, plusvalía, bienes inmuebles.

Abstract

Data mining allows the extraction of information, data classification and data processing, in order to know their behavior, to make predictions and for effective decision making. The objective of this research is to determine which are the most used data mining techniques for the analysis of real estate capital gains. For this purpose, a literature review, content analysis and the application of a grounded theory design are used. In the research analyzed it was possible to determine that the data mining techniques applied to predict real estate prices are: Artificial Neural Networks, Cluster Analysis, Decision Tree, Regression Models, Statistical Models, Naive Bayesian Classifier, K-Neighbors Algorithm, AdaBoost Algorithm, evidencing that the Artificial Neural Networks technique is the one with the best results. With respect to the data variables of each property 27 were found among the most important are presented the type of property, location, price, land area, construction area, number of floors, number of bedrooms, number of bathrooms, number of parking spaces or garage,

among others, in addition, data from publications in social networks and real estate websites of the city of Riobamba were analyzed. The results provided by this research are useful for the application of data mining techniques in real estate investment research or real estate appraisal predictions.

Keywords: Data mining, algorithms, predictions, capital gains, real estate.

Resumo

A mineração de dados permite a extração de informações, a classificação de dados e o processamento dos mesmos, para conhecer seu comportamento, fazer previsões e para tomadas de decisão efetivas. O objetivo desta pesquisa é determinar quais são as técnicas de mineração de dados mais utilizadas para a análise de ganhos de capital imobiliário, para as quais se utiliza uma revisão bibliográfica, análise de conteúdo e a aplicação de um projeto de teoria fundamentada. Nas investigações analisadas foi possível determinar que as técnicas de mineração de dados aplicadas para prever preços de imóveis são: Redes Neurais Artificiais, Análise de Cluster, Árvore de Decisão, Modelos de Regressão, Modelos Estatísticos, Classificador Bayesiano Naive, Algoritmo K-Vizinhos, Algoritmo AdaBoost, mostrando que a técnica de Redes Neurais Artificiais é a que apresenta os melhores resultados. Em relação às variáveis de dados de cada imóvel, foram encontradas 27, dentre as mais importantes estão o tipo de imóvel, localização, preço, área do terreno, área de construção, número de andares, número de cômodos, número de banheiros, número de estacionamentos ou garagens, entre outros, além disso, foram analisados dados de publicações em redes sociais e sites imobiliários da cidade de Riobamba. Os resultados fornecidos por esta pesquisa são úteis para a aplicação de técnicas de mineração de dados em pesquisas de investimentos imobiliários ou previsões de avaliações de imóveis.

Palavras-chave: Mineração de dados, algoritmos, previsões, ganhos de capital, imóveis.

Introducción

En el mundo existe un constante crecimiento del sector inmobiliario, las zonas urbanas se expanden y otras zonas internas se renuevan. Los proyectos inmobiliarios, constructoras y empresas dedicadas a la venta de bienes raíces, presentan importantes inversiones, que son fundamentales para la productividad y desarrollo económico de las ciudades, este desarrollo también involucra una revalorización en el precio de los bienes inmuebles, sea por la construcción, por su espacio de suelo, por sus condiciones de constructibilidad, ventajas de localización, o el potencial que involucra la actividad económica de su ubicación.

La plusvalía según la Real Academia Española-RAE (2021) la define como el incremento del valor de un bien por causas extrínsecas a este. El valor de la plusvalía incrementa, de acuerdo al avalúo del bien fijado por los gobiernos municipales y el valor de la venta del bien inmueble. Así también el “Avalúo es la determinación del valor comercial de un bien, realizado por un profesional competente, de acuerdo a criterios previamente establecidos y desarrollando una metodología adecuada al fin propuesto” (Ministerio de Desarrollo Urbano y Vivienda, 2016, p.3).

Algunos autores acerca de la variación del precio y plusvalía de los bienes inmuebles indican, que está relacionado a las inversiones en infraestructura y al crecimiento económico en la zona donde se encuentra ubicado y sus alrededores. “Los proyectos de obras públicas como la construcción de vías, el abastecimiento de agua y las inversiones en transporte colectivo producen beneficios que se capitalizan inmediatamente en el valor de las tierras circundantes” (Peterson, 2010, p.5).

Según (Merizalde et al., 2020) el incremento del valor del suelo se debe en su mayoría a factores externos a los dueños de la propiedad, como la construcción de obras públicas o privadas, o la inversión comercial en los alrededores del inmueble. También un factor que ha influenciado, es el aumento de la demanda de propiedades, causado por la gran oferta de préstamos bancarios para la adquisición de viviendas y por ende el aumento de proyectos inmobiliarios que intentan satisfacer el requerimiento de los ciudadanos.

Entre los factores que influyen en el valor de las propiedades es la cercanía que tiene a zonas de importancia como: plazas, parques, hospitales, centros comerciales, tiendas departamentales, bancos, hotelería, restaurantes, cafeterías, inmuebles para la cultura, deporte, educación, salud, oficinas, lugares de gestión pública y privada, servicios de transporte y comunicaciones, etc. (Márquez, 2020). Para establecer el valor de una propiedad, el Registro Oficial de la República del Ecuador (2016) considera en forma obligatoria, el valor del suelo que es el precio unitario del mismo, urbano o rural; el valor de las edificaciones, que es el precio de las construcciones que se hayan desarrollado con carácter permanente sobre un solar, calculado sobre el método de reposición; y el valor de reposición que se determina aplicando un proceso que permite la simulación de construcción de la obra que va a ser evaluada, a costos actualizados de construcción, depreciada de forma proporcional al tiempo de vida útil. (p. 36)

Por el importante crecimiento e inversión inmobiliaria existe una necesidad de estudio de teorías y herramientas computacionales que permite extraer información útil (conocimiento) de los volúmenes de datos digitales en rápido crecimiento (Banerji y Saxena, 2012). Por medio de la minería de datos,

la información inmobiliaria puede procesarse y convertirse en predicciones del comportamiento o estrategias importantes para el sector inmobiliario.

La minería de datos es la extracción de información y conocimiento profundo, es un campo de investigación multidisciplinario e interdisciplinario orientado a aplicaciones, el cual combina teorías y tecnologías en campos diferentes, como el aprendizaje automático, las bases de datos y las estadísticas matemáticas. En la actualidad, los algoritmos de minería de datos que más se utilizan son principalmente: árboles de decisión, algoritmos genéticos, redes neuronales artificiales y tecnología difusa (Huang, et al., 2020).

El análisis predictivo es una rama de la minería de datos la cual se utiliza para predecir o pronosticar tendencias o comportamientos. La predicción del avalúo inmobiliario puede convertirse en una herramienta estratégica para inversionistas o compradores de bienes inmuebles. Por esta razón es necesario definir ¿qué técnicas de minería de datos son útiles para analizar la plusvalía inmobiliaria?, con la finalidad de en futuras investigaciones aplicar los mismos y así determinar patrones en la variación de los precios de los bienes inmuebles de la ciudad de Riobamba, lo cual a su vez permitirá la toma de decisiones por parte de inmobiliarias, agentes de bienes raíces, municipios y organismos de control.

Esta investigación es parte del proyecto de investigación “Minería de datos para la predicción de la plusvalía inmobiliaria en el cantón Riobamba” de la Universidad Nacional de Chimborazo y tiene como objetivo definir los métodos de minería de datos útiles para el análisis de la plusvalía inmobiliaria y así responder a las siguientes preguntas ¿Cuáles son las técnicas de minería de datos que se utilizan en el sector inmobiliario? ¿Cuáles son las características, ventajas y desventajas de las técnicas más utilizadas? ¿Cuáles son las variables entorno a la plusvalía inmobiliaria que pueden utilizarse para la aplicación de minería de datos?

Trabajos relacionados

Los estudios relacionados con la presente investigación se muestran en la tabla 1, estos han sido tomados de revistas indexadas en bases de datos científicas de impacto mundial y regional.

Técnicas de minería de datos para el análisis de la plusvalía inmobiliaria

Tabla 1. Investigaciones analizadas, resumen, objetivos, aspectos importantes

ID	Investigación	Resumen
A-1	Research and application of data mining algorithm (Huang, et al., 2020)	Mencionan los conceptos y principios básicos de los algoritmos de la minería de datos. Proponen un método de fusión entre la regresión de crestas y el bosque aleatorio, para la creación de un modelo preciso y estable para ser utilizado en la predicción de precios inmobiliarios, con el objetivo de orientar en la toma de decisiones y mejorar las ventas en el sector de bienes raíces. Utiliza la tecnología de remuestreo bootstrap para recopilar un número fijo de datos. “Cada árbol de decisión se combina y la variable numérica se toma como una variable de predicción de regresión para formar un modelo de predicción de regresión forestal aleatoria no lineal multivariante” (Huang, et al., 2020, p.3)
A-2	Survey on predicting property price (Shinde y Gawande, 2018)	Realiza un análisis bibliográfico de técnicas predictivas utilizadas para pronosticar los precios de las propiedades inmuebles. Describe los métodos y procesos de la red neuronal y el árbol de decisiones aplicados para predecir los precios de las viviendas. Concluye que la red neuronal artificial funciona mejor y brinda la mayor precisión entre todas las técnicas analizadas, además presenta el error más bajo.
A-3	Housing price variations using spatio-temporal data mining techniques (Soltani et al., 2021)	Realizan un análisis de cómo los factores relacionados con la ubicación de una propiedad afectan su precio a lo largo del tiempo, aplicando la combinación de técnicas de regresión, técnicas de agrupamiento al análisis espacio-temporal, para identificar los conglomerados espaciales óptimos del precio de la vivienda en base a los datos analizados. Concluyen que la variable de ubicación, como las zonas metropolitanas o cercanía a zonas comerciales, tienen gran influencia en el precio de las propiedades.
A-4	Estimating the market values of houses in Tirana using data mining (Asilkan et al., 2012)	Examina la aplicabilidad de los modelos de Regresión hedónica (HR) y Red neuronal artificial (ANN) en el mercado inmobiliario. Según el modelo HR, la estimación de los precios está relacionado a los atributos de la vivienda y este determina el precio de la vivienda completa. Por lo general, cuantos más atributos deseados contiene la casa, mayor es su precio. El modelo ANN produce un precio de la vivienda basado en elementos particulares. Como resultado concluye que el método de ANN Red neuronal artificial supera al modelo de RRHH, por presentar menos errores.
A-5	Housing price prediction using neural networks (Teng et al., 2016)	El autor realiza una comparación de técnicas de minería de datos aplicadas en el análisis predictivo de precios de condominios en Singapur. Aplica la técnica de redes neuronales artificiales (ANN) con el promedio móvil integrado autorregresivo (ARIMA) y el análisis de regresión múltiple (MRA) para los precios de los condominios en Singapur. Como resultado aplicando la técnica del menor error cuadrático medio (MSE) se demostró la superioridad de ANN sobre otras herramientas predictivas. Utilizando este modelo se podría lograr la mayor precisión en la estimación.
A-6	House Resale Price Prediction Using Classification Algorithms	Analizan y aplican varias metodologías de predicción de datos y diferentes algoritmos de clasificación para la predicción de los valores de bienes inmuebles. Para el procesamiento de los datos aplican algoritmos de clasificación como árbol de decisión, bosque aleatorio, AdaBoost, Naïve Bayes, regresión logística.

Técnicas de minería de datos para el análisis de la plusvalía inmobiliaria

	(Durganjali y Pujitha, 2019)	Concluye que el modelo AdaBoost y árbol de decisión están más concentrados en generar reglas y presentan mejores resultados para ese conjunto de datos.
A-7	Analysis of Data Mining Techniques on Real Estate (Banerji y Saxena, 2012)	Analizan técnicas de minería de datos, para la obtención de información de grandes volúmenes de datos para la predicción de precios de bienes raíces. Aplican la técnica de perceptrón multicapa con regresión de ritmo, sugieren que este método es un patrón eficiente para el descubrimiento de conocimientos. El análisis se realiza utilizando un conjunto de datos inmobiliarios que buscan la identificación del comportamiento de los precios inmobiliarios para la inversión. Concluyen que la regresión del ritmo es la mejor técnica entre otras técnicas de regresión, sin embargo, el método de redes de perceptrones multicapa la supera en resultados, precisión y margen de error.
A-8	Real Estate Cost Estimation Through Data Mining Techniques. (Khare et al., 2021)	En esta investigación se realizan dos conjuntos de predicciones, uno con todas las características disponibles necesarias para comprar la casa y el otro con algunas características seleccionadas. Se aplican diversas técnicas de regresión de aprendizaje automático como la regresión lineal, regresión polinomial, árbol de decisiones y bosque aleatorio. Concluye que de las técnicas analizadas la regresión polinomial es la que da los mejores resultados.
A-9	Research on real estate pricing methods based on data mining and machine learning (Yu et al., 2021)	Analizan los modelos de precios inmobiliarios basados en la minería de datos y el aprendizaje automático, presentando métodos para la estimación de tendencias, fijación de precios, que presenten ventajas tales como dinámica, precisión y flexibilidad. Propone un modelo de precios de bienes raíces basado en la estimación de tendencia, variable en el tiempo de suavizado exponencial cuadrático. Combinando la tecnología de minería de datos y tecnología de aprendizaje automático. Se basa principalmente en la construcción de: un árbol de precios de bienes raíces, un árbol de VPN (decisión) y un árbol de opciones.
A-10	Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. (Grybauskas et al., 2021)	Aplicaron 15 modelos de aprendizaje automático para la predicción de precios de bienes inmobiliarios, posterior a la pandemia. En la primera etapa, el conjunto de datos se dividió en conjuntos de datos de prueba y entrenamiento, en la segunda etapa se pusieron a prueba los algoritmos de aprendizaje automático. De los 15 modelos diferentes probados, el algoritmo de gradiente de árboles reforzados (XGB) fue el más preciso y se concluyó que es el mejor modelo para aplicarse.
A-11	Un modelo econométrico para determinar el valor de venta de proyectos inmobiliarios en la ciudad de Santiago. (López, 2020)	Aplicó un análisis descriptivo de datos y técnicas de estimación, aplicando modelos de regresión, con una base de datos robusta de proyectos inmobiliarios, con el objetivo de determinar qué variables influyen en el precio de un proyecto inmobiliario. Concluye que el modelo de regresión lineal puede presentar errores y no ser exacto. Con respecto a las variables que más influyen en el valor final de un proyecto inmobiliario describe que son: el precio final, costo de construcción y otros gastos inmobiliarios.
A-12	Análisis clúster de precios de vivienda de 2017 en la ciudad de Valencia.	Realizan la aplicación de técnicas multivariantes para clasificación y uso del análisis clúster con el método jerárquico aglomerativo o ascendente. para la creación de modelos de estimación de valor de viviendas. El análisis de conglomerados lo realizaron a partir de la

Técnicas de minería de datos para el análisis de la plusvalía inmobiliaria

(Fernández et al., 2017)		herramienta estadística SSPS. Concluyen que por medio del análisis de conglomerados se pueden identificar pautas en la distribución del precio de la vivienda, la variable ubicación específicamente la cercanía a la zona metropolitana o comercial influye de gran manera en el valor de la vivienda.
A-13	Una metodología de minería de datos para la agrupación de series temporales: aplicación al sector de la construcción residencial. (Ferrán, 2011)	Presenta una metodología de minería de datos denominada Metodología del haz de rectas, aplicada al Sector de la Construcción inmobiliaria mediante un Análisis Cluster, Series temporales, utilizando como criterio de agrupación el Algoritmo de las k-medias, con el objetivo de realizar un análisis comparativo y predictivo de los valores de las viviendas.
A-14	Mercado inmobiliario de España: Una herramienta para el análisis de la oferta. (Casas et al., 2019)	Proponen el desarrollo de un software desarrollado en Java y R, que, a partir de la información obtenida mediante la minería de datos en un portal inmobiliario, permite la realización de un completo análisis estadístico de los datos, modelos de estimación de variables, como el precio de oferta, con distintas técnicas como modelos de regresión lineales y no lineales o estimación a través de la creación de redes neuronales de tipo perceptrón. Además de la generación de mapas con la localización geográfica de las propiedades.
A-15	Método automático para la predicción del avalúo comercial de un inmueble en la ciudad de Bogotá. (Martínez-Sánchez y Téllez-Buitrago, 2021)	En este proyecto se realizó el diseño, desarrollo e implementación de un método automático mediante técnicas de Machine Learning con el cual se puede predecir el valor de inmuebles de vivienda. Se calcularon las medidas de cada técnicas pro medio del desempeño de Coeficiente de determinación (R ²), Error cuadrático medio (RMSE) y Coeficiente de Variación (CV) respecto a la varianza, para así seleccionar el mejor método. Se concluyó que la técnica con mejores resultados fue el bosque aleatorio.

Fuente: elaboración propia

Metodología

En la investigación se aplicó un método cualitativo de recolección y análisis de los datos. Este proceso está basado en el diseño de teoría fundamentada Trinidad et al. (2006), menciona que esta metodología de análisis es una alternativa para el análisis cualitativo en el que se propone crear teoría de una manera sistemática, a partir de los datos en los procesos sociales, su objetivo es crear propuestas teóricas fundamentadas en datos empíricos. Esta investigación se centra en el análisis de datos empíricos, constituyendo la formulación de hipótesis teórica.

En la primera etapa se realizó una revisión bibliográfica y análisis de contenido a investigaciones relacionadas al tema o que presenten datos relevantes, posteriormente se seleccionó la información que puede responder a las preguntas de investigación, se estructuró y sintetizó el contenido con los

datos: autores de la investigación, resumen del objetivo y métodos aplicados. Para analizar cuáles son las técnicas más utilizadas y variables utilizadas en esta temática, se realizó un diagrama asociativo entre las investigaciones analizadas y las técnicas de minería de datos que las mismas mencionan. Para concluir se describen los aspectos importantes las ventajas y desventajas de las técnicas de minería de datos que más se utilizaron y de estas cuál es la que mejores resultados presentó en las investigaciones analizadas.

Discusión y Resultados

En esta sección se presentan los resultados de la investigación. En la figura 1 mediante un diagrama de asociación se establecen las técnicas de minería de datos que han sido aplicadas en los estudios relacionados a la plusvalía inmobiliaria, de esta manera se puede establecer cuáles son las más utilizadas. Posteriormente en la tabla 1, identificadas las técnicas que más se utilizaron se definieron los aspectos importantes, ventajas y desventajas para el análisis y conclusión de cuál es la técnica más adecuada para la aplicación en esta temática.

Técnicas de minería de datos utilizadas

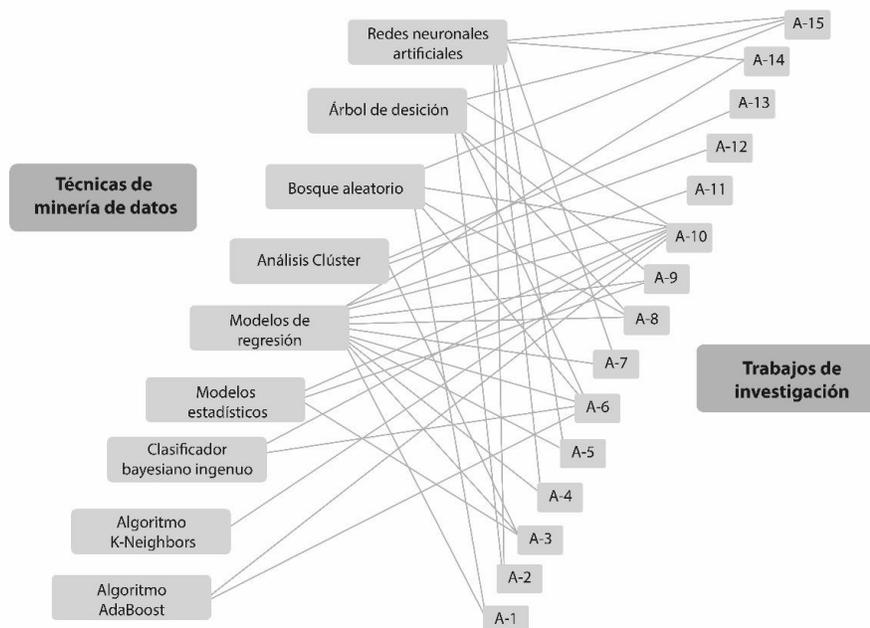


Figura 1. Diagrama asociativo de técnicas de minería de datos aplicadas en la predicción de la plusvalía inmobiliaria
Fuente: elaboración propia

Técnicas de minería de datos para el análisis de la plusvalía inmobiliaria

Tabla 2. Técnicas más utilizadas de minería de datos aplicadas en el sector inmobiliario

Métodos	Aspectos importantes	Ventajas	Desventajas
Redes neuronales	Las redes neuronales son modelos de cálculo caracterizados por algoritmos eficientes que operan de forma masivamente paralela y permiten desarrollar tareas cognitivas como el aprendizaje de patrones, predicciones, la clasificación o la optimización. -Es un tipo de aprendizaje no supervisado	-Se puede crear para descubrimiento y extracción de conocimiento. -Clasifican conjuntos de datos -Son tolerantes a fallas. -Presentan márgenes bajos de error. -Clasifican patrones desconocidos con patrones conocidos que comparten las mismas. Características distintivas. -Sigue funcionando, aunque presente una falla significativa de sus neuronas e interconexiones.	-Son complejas para el aprendizaje de grandes tareas. -No tienen la capacidad de interpretar resultados. -Pueden presentar extensos tiempos para el aprendizaje. (Rivas y Mazón, 2018)
Análisis Clúster	Esta técnica consiste en la construcción y clasificación de grupos de datos similares a partir de conjuntos de datos. Presenta varios métodos para distintos tipos de datos y objetivos como el Agrupamiento jerárquico y de K-Means. -Es un tipo de aprendizaje no supervisado	-Agrupa datos similares -Puede reducir complejas cantidades de información en pequeños grupos con características similares. -Es fácil de entender, aplicar y adaptar. -Es eficiente, tiene buen desempeño y resultados. - Se puede aplicar a datos estáticos como a datos dinámicos	-Algunas técnicas de clustering no son convenientes para grandes grupos de datos y necesitan que el usuario defina el número de clústeres. -Puede llegar a ser computacionalmente costoso.
Árbol de decisión	Es un método de análisis que mediante una representación esquemática facilita la toma de decisiones en el uso de resultados y probabilidades asociadas. búsquedas binarias -Es un tipo de aprendizaje supervisado	-Son fáciles de representar e interpretar. -Permiten llegar a la toma de decisiones o acciones específicas. - Concentrado en generar reglas, da resultados más precisos. - Reduce el número de variables independientes	-Seleccionar al atributo adecuado para la raíz del árbol. -No es funcional para pequeños conjuntos de datos. -Propenso a sobreajuste
Modelos de regresión	-Es un tipo de aprendizaje supervisado Los modelos de regresión encontrados son: -Regresión lineal -Regresión múltiple -Regresión hedónica -Regresión ponderada geográfica -Regresión ponderada temporalmente -Regresión de crestas - Regresión polinomial	- Se pueden realizar relaciones complejas. -Existen diversas técnicas de regresión que pueden aplicarse a distintos procesos u objetivos. -Fáciles de desarrollar y entender. -Útiles para la comprensión de datos y toma de decisiones. -Útiles para la clasificación de texto e imágenes.	-Se pueden presentar valores atípicos. -Deficiente para realizar modelaciones complejas.

Técnicas de minería de datos para el análisis de la plusvalía inmobiliaria

<p>Bosque aleatorio</p>	<p>Se construye con múltiples árboles de decisión que combina las predicciones o clasificaciones de cada árbol, para obtener predicciones más precisas y estables, reduciendo las altas varianzas de los árboles individuales. -Es un tipo de aprendizaje supervisado</p>	<p>-Es simple de utilizar y se puede aplicar para tareas de clasificación o regresión. -Procesamiento de variables cualitativas y cuantitativas sin la necesidad de crear variables indicadoras. - Realiza validaciones al mismo tiempo que construye los árboles, generando estimaciones de errores.</p>	<p>-Ausencia de interpretabilidad. -Propenso a sobreajuste. -Por presentar múltiples árboles de decisión, su proceso puede ser lento.</p>
--------------------------------	--	---	---

Fuente: elaboración propia

Variables entorno a la plusvalía inmobiliaria

Según las investigaciones analizadas, en la figura 2 se presentan los datos que influyen en la variación del avalúo inmobiliario y se utilizan como variables en las técnicas de minería de datos. En la figura 3 se presentan las variables que pueden utilizarse en las predicciones de avalúo inmobiliario en la ciudad de Riobamba, para obtener esta información se realizó un análisis de los datos que se presentan con mayor frecuencia en los principales sitios web inmobiliarios y publicaciones en redes sociales de agencias de bienes raíces o constructoras.

Tipo de propiedad <small>Casa, departamento, terreno, local, estacionamiento.</small>	Año de construcción	Topografía del suelo	Amoblado
Ubicación	Precio	Calidad y materiales de construcción	Características adicionales <small>Piscina, ascensor, calefón.</small>
Nombre del barrio o sector	Número de pisos	Acabados	Seguridad
Nombre del conjunto o urbanización	Número de habitaciones	Zonas externas <small>Patio, jardín, terraza.</small>	Precio alicuotas
Ubicación en el mapa	Número de baños	Tiempo en venta	
Área de terreno	Número de estacionamientos	Servicios básicos <small>Agua, luz, alcantarillado, internet.</small>	
Área de construcción	Distancia a zonas comerciales <small>Centros educativos, hospital, parques, centros comerciales, restaurantes, tiendas, etc.</small>	Características del entorno	

Figura 2. Variables para el análisis inmobiliario en investigaciones realizadas

Fuente: elaboración propia

Técnicas de minería de datos para el análisis de la plusvalía inmobiliaria

Tipo de propiedad	Esquinera	Área de construcción	Número de estacionamientos
Ubicación Calles, barrio, sector.	Tipo de calle	Número de pisos	Cercanía a zonas comerciales
Estado de la propiedad	Servicios básicos Agua, luz, alcantarillado.	Número de habitaciones	
Precio	Área de terreno	Número de baños	

Figura 3. Variables para el análisis inmobiliario, basadas en datos de plataformas de venta de bienes raíces
Fuente: elaboración propia

Conclusiones

La minería de datos puede convertirse en una herramienta estratégica para los inversionistas o compradores de bienes inmuebles. Por medio del análisis y procesamiento de datos se puede predecir tendencias o comportamientos en el precio de las propiedades. En las investigaciones analizadas se pudo determinar que las técnicas de minería de datos aplicadas en el sector inmobiliario son: Redes neuronales artificiales, Análisis Clúster, Árbol de decisión, Modelos de regresión (Regresión lineal Regresión múltiple Regresión hedónica, Regresión ponderada geográfica Regresión ponderada temporalmente Regresión de crestas, Regresión polinomial), Modelos estadísticos, Clasificador bayesiano ingenuo, Algoritmo K-Neighbors, Algoritmo AdaBoost.

Posteriormente se realizó un análisis de las ventajas y desventajas de las técnicas de minería de datos aplicadas a los estudios inmobiliarios, se pudo evidenciar que la técnica más utilizada y que mejores resultados proporciona, es la técnica de Redes neuronales artificiales, por medio de esta técnica se puede extraer conocimiento, clasificar conjuntos de datos, además de que son tolerantes a fallas y presentan los más bajos márgenes de error.

Con respecto a los datos de las propiedades que se aplican como variables en las técnicas de minería de datos se encontraron 25, de las cuales las principales variables para el análisis inmobiliario en la ciudad de Riobamba, basado en las investigaciones y que concuerdan con los datos de las publicaciones en redes sociales y sitios web son: tipo de propiedad, ubicación (calles, barrio, sector), estado de la propiedad, precio, esquinera, tipo de calle, servicios básicos (agua, luz, alcantarillado), área de terreno, área de construcción, número de pisos, número de habitaciones, número de baños,

número de estacionamientos o garaje y cercanía a zonas comerciales. Los datos proporcionados por esta investigación son de gran utilidad para la aplicación de técnicas de minería de datos en futuras investigaciones sobre inversiones inmobiliarias o predicciones del avalúo inmobiliario en la ciudad de Riobamba.

Bibliografía

1. Asilkan, O., Faqolli, A., Gerdeci, A., & Cico, B. (2012). Estimating the market values of houses in Tirana using data mining. *Global Journal on Technology*, 1. <http://archives.unpub.eu/index.php/P-ITCS/article/view/857/1082>
2. Banerji, G., & Saxena, K. (2012). Analysis of Data Mining Techniques on Real Estate. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(3), 223-230. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.649.7414&rep=rep1&type=pdf>
3. Casas, J., Casas, D., Caridad, J., & Núñez, J. (2019). Mercado inmobiliario de España: Una herramienta para el análisis de la oferta. *Cuadernos de Economía*, 42(120). <http://www.cude.info/index.php/CUDE/article/view/69>
4. Durganjali, P., & Pujitha, M. V. (2019). House Resale Price Prediction Using Classification Algorithms. 2019 International Conference on Smart Structures and Systems (ICSSS) (pp. 1-4). <https://ieeexplore.ieee.org/abstract/document/8882842>
5. Fernández, L., Cózar, A., & Llorca, A. (2017). Análisis clúster de precios de vivienda de 2017 en la ciudad de Valencia. XII Congreso de ciencia regional de Andalucía. Comercio Regional y empleo: Una perspectiva Regional. https://old.reunionesdeestudiosregionales.org/sevilla2017/media/uploads/2017/10/05/Paper_Analisis_cluster.pdf
6. Ferrán, M. (2011). Una metodología de minería de datos para la agrupación de series temporales: aplicación al sector de la construcción residencial. [Tesis Doctoral, Universidad Complutense de Madrid]. <https://core.ac.uk/download/pdf/19715384.pdf>
7. Grybauskas, A., Pilinkienė, V., & Stundžienė, A. (2021). Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. *Journal of big data*, 8(1), 1-20. <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00476-0>

8. Huang, C., Yang, Y., Wang, H., Zhang, X., Zhao, J., & Wan, J. (2020). Research and application of data mining algorithm. In *Journal of Physics: Conference Series* (Vol. 1748, No. 3, p. 032043). <https://iopscience.iop.org/article/10.1088/1742-6596/1748/3/032043/pdf>
9. Khare, S., Gourisaria, M. K., Harshvardhan, G. M., Joardar, S., & Singh, V. (2021). Real Estate Cost Estimation Through Data Mining Techniques. In *IOP Conference Series: Materials Science and Engineering*, 1099, (1), p. 012053). <https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012053/pdf>
10. López, F. (2020). Un modelo econométrico para determinar el valor de venta de proyectos inmobiliarios en la ciudad de Santiago. *Horizontes Empresariales*, 19(1), 28-50. <http://revistas.ubiobio.cl/index.php/HHEE/article/view/4158/3856>
11. Martínez-Sánchez, D. N., & Téllez-Buitrago, V. J. (2021). Método automático para la predicción del avalúo comercial de un inmueble en la ciudad de Bogotá. [Tesis pregrado, Universidad católica de Colombia]. <https://repository.ucatolica.edu.co/bitstream/10983/26289/1/trabajo%20de%20grado%20Completo.pdf>
12. Márquez, L. (2020). El capital inmobiliario-financiero y la producción de la ciudad latinoamericana hoy. *Cadernos Metrópole*, 22(49), 665-682. <https://doi.org/10.1590/2236-9996.2020-4901>
13. Merizalde, M., Caicedo, F., & Salame, M. (2020). Ley de la Plusvalía en el Ecuador; un breve análisis desde su aprobación hasta su inmediata derogación. *Dilemas Contemporáneos: Educación, Política y Valores*, 7. <https://doi.org/10.46377/dilemas.v33i1.2119>
14. Ministerio de Desarrollo Urbano y Vivienda. (2016, 3 de octubre). Normas técnicas para catastro de bienes inmuebles urbanos, rurales. Registro Oficial Suplemento 853. <https://www.habitatyvivienda.gob.ec/wp-content/uploads/downloads/2016/10/Acuerdo-Ministerial-29-16.pdf>
15. Peterson, G. E. (2010). *La plusvalía de la tierra como opción para el financiamiento de la infraestructura urbana*. The World Bank. <https://doi.org/10.1596/978-0-8213-8329-2>
16. Real Academia Española. (2021). Diccionario de la lengua española (23a ed.). <https://dle.rae.es/plusval%C3%ADa>

17. Registro Oficial Órgano de la República del Ecuador. (2016, 3 de octubre). Valor de la propiedad. Registro oficial no. 853 <https://www.registroficial.gob.ec/index.php/registro-oficial-web/publicaciones/registro-oficial/item/8389-registro-oficial-no-853.html>
18. Rivas, W., & Mazón, B. (2018). *Redes neuronales artificiales aplicadas al reconocimiento de patrones*. Editorial UTMACH.
19. Shinde, N., & Gawande, K. (2018). Survey on predicting property price. In 2018 International Conference on Automation and Computational Engineering (ICACE) (pp. 1-7). <https://doi.org/10.1109/ICACE.2018.8687080>
20. Soltani, A., Pettit, C. J., Heydari, M., & Aghaei, F. (2021). Housing price variations using spatio-temporal data mining techniques. *Journal of Housing and the Built Environment*, 1-29. <https://link.springer.com/article/10.1007/s10901-020-09811-y>
21. Teng, W. T., Wang, L., Wang, Y., & Chang, Q. (2016). Housing price prediction using neural networks. In 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 518-522). IEEE. <https://ieeexplore.ieee.org/abstract/document/7603227>
22. Torrente, Á. (2021). Sistema de Web Scraping orientado a portales del ámbito inmobiliario [Tesis Doctoral, Universidade da Coruña]. http://castor.det.uvigo.es:8080/xmlui/bitstream/handle/123456789/575/TorrentePatino_Alvaro_TFM_2021.pdf?sequence=1
23. Trinidad, A., Carrero, V., & Soriano, R. (2006). *Teoría fundamentada “Grounded theory”: la construcción de la teoría a través del análisis interpretacional* (Centro de Interpretaciones Sociológicas (ed.)). <https://n9.cl/r8bg8>
24. Yu, Y., Lu, J., Shen, D., & Chen, B. (2021). Research on real estate pricing methods based on data mining and machine learning. *Neural Computing and Applications*, 33(9), 3925-3937. <https://sci-hub.se/https://doi.org/10.1007/s00521-020-05469-3>