

UNIVERSIDAD INTERNACIONAL SEK

DIGITAL SCHOOL

Trabajo de fin de carrera titulado:

**ANÁLISIS PARA LA PREDICCIÓN DEL VALOR DE UN BIEN INMUEBLE EN LA
CIUDAD DE QUITO POST PANDEMIA**

Realizado por:

WLADIMIR ARSECIO VILCA LINCANGO

Director del proyecto:

Joe Carrión Jumbo, PhD

Como requisito para la obtención del título de:

MAGISTER EN SISTEMAS DE INFORMACIÓN MENCIÓN EN DATA SCIENCE

Quito, marzo 2022

DECLARACIÓN JURAMENTADA

Yo, **WLADIMIR ARSECIO VILCA LINCANGO**, con cédula de identidad 1714478581, declaro bajo juramento que el trabajo aquí desarrollado es de mi autoría, que no ha sido previamente presentado para ningún grado a calificación profesional, y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración, cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la UNIVERSIDAD INTERNACIONAL SEK, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente.

**WLADIMIR ARSECIO
VILCA LINCANGO** Firmado digitalmente por
WLADIMIR ARSECIO VILCA
LINCANGO
Fecha: 2022.04.12 19:59:38 -05'00'

WLADIMIR ARSECIO VILCA LINCANGO

C.C: 1714478581

DECLARATORIA

El presente trabajo de investigación titulado:

“ANÁLISIS PARA LA PREDICCIÓN DEL AVALÚO DE UN BIEN INMUEBLE EN LA CIUDAD DE QUITO POST PANDEMIA”

Realizado por:

WLADIMIR ARSECIO VILCA LINCANGO

Como requisito para la Obtención del Título de:

MASTER EN SISTEMAS DE INFORMACIÓN MENCIÓN EN DATA SCIENCE

Ha sido dirigido por el profesor

Ing. Joe Carrión Jumbo, PhD.

Quien considera que constituye un trabajo original de su autor

**JOE LUIS
CARRION
JUMBO**

Firmado digitalmente
por JOE LUIS
CARRION JUMBO
Fecha: 2022.04.13
15:24:40 -05'00'

Joe Carrión Jumbo, PhD.,

DIRECTOR

PROFESORES INFORMANTES

Después de revisar el trabajo presentado, se lo ha calificado como apto para su defensa oral ante el tribunal examinador.



Firmado electrónicamente por:
**SILVIA DIANA
MARTINEZ
MOSQUERA**

Ing. SILVIA DIANA MARTINEZ MOSQUERA

**CESAR BYRON
GUEVARA
MALDONADO**
Numero de reconocimiento:
SERIAL NUMERO: 040121121003 -
CN=CESAR BYRON GUEVARA
MALDONADO, O=CENTRO DE
CERTIFICACION DE INFORMACION,
O=SECURITY DATA S.A., C=EC
Razon: DEFENSA DE TESIS
Localizacion: QUITO
Fecha: 2022-04-13T16:20:18.217-05:00

Ing. CESAR BYRON GUEVARA MALDONADO

Quito, marzo del 2022

DEDICATORIA

El siguiente trabajo de tesis es un homenaje a una persona esencial en mi vida, mi querida madre, quien desde mi niñez ha sabido guiarme con sus sabios consejos e incondicional apoyo.

A mis amados hermanos que han estado presentes en cada trayecto, compartiendo los buenos y malos momentos, siempre alentándome con sus palabras, hermanos gracias.

Agradezco a Dios por todas las bendiciones recibidas en este largo camino, a Él por otórgame las facultades necesarias para desempeñarme en el transcurso de mi desarrollo profesional.

A Gaby por estar pendiente de mí en todo momento, gracias por tu constante apoyo, amor.

AGRADECIMIENTO

Este Proyecto de Investigación, es la representación de todos mis esfuerzos y la muestra de todas mis capacidades, quiero agradecer a todas aquellas personas que han creído siempre en mí.

A universidad Sek, en cuyo lugar conocí a docentes que contribuyeron tanto con mi formación académica como personal y cuyas enseñanzas dejarán un permanente legado en mi desempeño laboral.

ÍNDICE GENERAL

CAPÍTULO 1

INTRODUCCIÓN	26
1.1 Planteamiento del problema	26
1.2 Justificación	28
1.3 Objetivos	29
1.3.1 Objetivo General	29
1.3.2 Objetivos Específicos	29
1.3.3 Alcances y limitaciones	29
1.4 Marco Conceptual	30
1.4.1 Quito	30
1.4.2 Inmueble	31
1.4.3 Norma técnica de valoración de bienes inmuebles	31
1.4.4 Valor de mercado	31
1.4.5 Avalúo del predio	31
1.4.6 Mercado Inmobiliario	31
1.4.7 Valoración de bienes inmuebles	31
1.4.8 Marco Legal	32
1.4.8.1 Plan metropolitano de Quito para el desarrollo de ordenamiento territorial	32
1.4.8.2 Estructura Urbano Rural	32
1.4.8.3 Clasificación de Suelo	33
1.4.8.4 Uso de suelo	35
1.4.9 Conjunto de datos	38
1.4.10 Inteligencia Artificial	38
1.4.11 Web Scraping	39
1.4.12 Aprendizaje Supervisado	40
1.4.13 Página web	40
1.4.14 Plataformas Inmobiliarias	40
1.4.14.1 Properati	41
1.4.14.2 Plusvalía	42
1.4.14.3 Remax	42

1.4.15	Las Redes Neuronales Artificiales	43
1.4.16	Regresión Lineal	46
1.4.17	Árbol de decisión	48
1.4.18	Random Forest	50
1.4.19	Medida de Desempeño	51
1.4.19.1	Coefficiente de Determinación R2	51
1.4.19.2	Error cuadrático medio RMSE	52
1.4.19.3	Coefficiente de Variación con el RMSE (CV)	52
1.4.20	ROC Y AUC	53
1.4.21	Herramientas de Machine Learning	53
CAPÍTULO II		55
ESTADO DEL ARTE		55
CAPÍTULO III		60
METODOLOGÍA		60
3.1	Metodología	60
3.2	Construcción de los datos	62
3.2.1	Fuente	62
3.2.2	Extracción de datos	62
3.2.3	Transformación de los datos	62
3.2.3.1	Visualización	63
3.2.3.2	Limpieza de columnas y de registros	64
3.2.4	Algoritmos	66
3.2.4.1	Regresión Lineal	66
3.2.4.2	Random Forest	69
3.2.4.3	Árbol decisión	70
3.2.4.4	Redes Neuronales	72
3.2.5	Evaluación de métodos automáticos	73
RESULTADOS		79
4.1	Rendimiento de los datos	79
CONCLUSIONES Y TRABAJOS FUTUROS		81
BIBLIOGRAFÍA		82

ÍNDICE DE FIGURAS

Figura 1. Muestra la distribución zonal de la ciudad de Quito. Fuente: https://www.quito.gob.ec	30
Figura 2. Muestra la clasificación de suelos en DMQ. Fuente: Plan de Uso y Gestión del Suelo	34
Figura 3. Muestra los recursos no renovables (RNNR) del DMQ. Fuente: Plan de Uso y Gestión del Suelo	37
Figura 4. Muestra cómo la Inteligencia Artificial engloba muchas tareas. Fuente: Elaboración propia.....	38
Figura 5. Muestra el proceso de funcionamiento de software web scraping. Fuente: Propia.....	39
Figura 6. Muestra el proceso de aprendizaje que implementa Machine Learning. Fuente: Elaboración propia.....	40
Figura 7. Muestra la interfaz de la página web Properati en Ecuador. Fuente: www.properati.com.ec	41
Figura 8. Muestra la interfaz de la página web de Plusvalía. Fuente: https://www.plusvalia.com/	42
Figura 9. Muestra la interfaz de la página web Remax para Ecuador Fuente: https://www.remax.com.ec/	43
Figura 10. Muestra como una la red neuronal. Fuente: Elaboración propia	44
Figura 11. Muestra métodos de aprendizaje supervisados. Fuente: Elaboración propia	45
Figura 12. Muestra algoritmos de agrupamiento. Fuente: Elaboración propia	46
Figura 14. Muestra que no hay un patrón sistemático. Fuente libro: GUJARATI, Econometría, Quinta Edición, México.....	48
Figura 15. Muestra la formación de un árbol decisión. Fuente: Elaboración Propia.....	49
Figura 16. Muestra el método de Random Forest Fuente: Elaboración Propia.....	51
Figura 17 Muestra el resultado con las palabras claves ingresadas en Scopus. Fuente: Base de datos Scopus	55
Figura 18. Estructura del proceso a utilizar. Fuente: Elaboración propia.....	61
Figura 19. Muestra la correlación de los variables. Fuente: Elaboración propia	65
Figura 20. Muestra el diagrama de dispersión. Fuente: Elaboración propia.....	67
Figura 21. Correlación de los datos con los atributos Price y área. Fuente: Elaboración propia.....	68
Figura 22. Muestra los resultados de la estimación con modelo Regresión Lineal. Fuente: Elaboración propia.	68
Figura 23 Método de Random Forest para el conjunto de datos Fuente: Propia	70
Figura 24. Muestra la formación del árbol decisión. Fuente: Elaboración propia.	71
Figura 25 Muestra la estructura de red neuronal. Fuente: Elaboración propia.....	72

ÍNDICE DE TABLAS

Tabla 1. Detalla el procedimiento del método denominado Random Forest	50
Tabla 2 . Descripción de las variables.	
Fuente: Elaboración propia	63
Tabla 3. Tabla de variables descartadas	
Fuente: Elaboración propia	64
Tabla 4. Tabla de sectores de Quito de acuerdo a los datos descargada.	
Fuente: Elaboración propia	66
Tabla 5. Muestra los resultados obtenidos con el sector norte.	
Fuente: Elaboración propia	73
Tabla 6. Muestra los resultados obtenidos con el sector 2.	
Fuente: Elaboración propia	74
Tabla 7. Muestra los resultados obtenidos con el sector 3	
Fuente: Elaboración propia	74
Tabla 8. Muestra los resultados obtenidos con el sector 4.	
Fuente: Elaboración propia	75
Tabla 9. Muestra los resultados obtenidos con el sector 5.	
Fuente: Elaboración propia	75
Tabla 10. Muestra los resultados obtenidos con el sector 6.	
Fuente: Elaboración propia	76
Tabla 11. Muestra los resultados obtenidos con el sector.	
Fuente: Elaboración propia	76
Tabla 12. Descripción de métodos aplicados a los datos divididos por sectores.	
Fuente: Elaboración propia	79
Tabla 13. Muestra el mejor ajuste al precio de las viviendas de acuerdo RMSE.	
Fuente: Elaboración propia	80
Tabla 14 Análisis ROC y área bajo la curva	
Fuente: Propia.....	80

RESUMEN

El mundo siempre ha atravesado situaciones críticas que han traído grandes devastaciones consigo, es claro el ejemplo de la Segunda Guerra Mundial, evento histórico que alteró la vida y economía de varias naciones. Coyunturalmente, la pandemia de COVID 19 generó muchos efectos negativos para la humanidad, provocó el colapso del mercado global aumentado el desempleo, y condicionó la vida de miles de individuos que tuvieron que asumir una nueva normalidad, en el que el uso de la tecnología constituyó un recurso para no afectar tanto sus actividades laborales como educativas.

En referencia al sector inmobiliario, muchos proyectos quedaron paralizados por falta de presupuesto ya que las ventas decayeron notablemente, así se configuró un panorama desalentador para inversionista y trabajadores, afectados por la falta de empleo. Especialistas en economía a pesar de lo sucedido, pronosticaron que la reactivación del área inmobiliaria se lograría a largo plazo.

Este trabajo sitúa como principal beneficiario al sector inmobiliario, el cual mediante el desarrollo de proyectos genera inversiones importantes, así como la adquisición y venta de bienes inmuebles para el sector público o privado. El área inmobiliaria es idónea para implementar técnicas de aprendizaje automático, una de las razones fundamentales es el alto interés de personas que buscan viviendas para conocer su valor, y así direccionarse a una futura compra o venta.

En correspondencia a lo referido, el presente proyecto de titulación busca realizar un análisis para el establecimiento de un método automático, a partir del uso técnicas de máquinas de aprendizaje para así predecir el valor de un inmueble en la ciudad de Quito.

Debido a que en la tasación de un bien inmueble realizada por un perito calificado, se toman en cuenta distintos aspectos, entre ellos: el área (superficie del terreno) en m^2 , el área de construcción, la ubicación, el número de pisos, el número de habitaciones, el número de baños que posee el inmueble, los acabados, la antigüedad, entre otros.

La tasación se define como: “el proceso de cálculo para la determinación de los avalúos del suelo, construcción y adicionales constructivos, de los bienes inmuebles del Distrito Metropolitano de Quito, de acuerdo a la metodología y técnicas de valuación, en el marco de lo que establece el Código Orgánico de Organización Territorial, Autonomía y Descentralización-COOTAD y Acuerdo Ministerial 029 que expide las "Normas Nacionales para el Catastro de Bienes Norma técnica para la Valoración de los Bienes Inmuebles en el MDMQ” (Norma técnica de Valoración de los Bienes Inmuebles en el MDMQ, 2019, pág. 18)

Los datos recolectados corresponden a las publicaciones de los bienes inmuebles en las diferentes páginas web en Ecuador dedicadas a la venta de inmuebles, los cuales se obtuvieron mediante el uso de *web scraping*, esta herramienta fue aplicado en páginas como son Plusvalía, Properati, Remax. Con los datos obtenidos se procedió a realizar un procedimiento de limpieza de columnas, eliminación de registros.

Con el resultado de los datos extraídos y una vez aplicados el proceso de limpieza se procedió a organizar los datos por sectores y se efectuó la implementación de técnicas de aprendizaje automático como son: regresión lineal, árbol de decisión, redes neuronales, Random Forest para cada conjunto de datos. En consecuencia, al aplicar los diferentes algoritmos, la regresión lineal presentó el mejor desempeño frente a los demás modelos aplicados.

Palabras Clave: aprendizaje automático, datos, inmuebles, pandemia, tasación.

ABSTRACT

The virus named COVID-19 represents the greatest threat to our lives, since the Second World War, this virus has changed from the way people interact to the economies of world powers, since its inception COVID-19 caused global collapse In the market, unemployment increased, radical changes in business models and ways of working, a change that the world had not seen for decades.

The pandemic affects the real estate sector directly, this was evidenced in the lack of investment for projects, stopped projects, drop in sales, among others. Specialists in economics predict that the area of the real estate sector and its reactivation will take place in the long term.

The focus of this work has the real estate sector as its main beneficiary, this sector generates large investments in the creation of real estate projects as well as in the acquisition, sale of real estate for the public or private sector, this area is ideal for implementing learning techniques of machine (Machine Learning), one of the main reasons is high interest of people looking for homes and others their interest to find out the value to plan a future purchase or sale.

In correspondence to the aforementioned, the present titling project seeks to carry out an automatic analysis for the establishment of a method, based on the use of machine learning techniques to predict the value of a property in the city of Quito.

Due to the fact that in the appraisal of real estate carried out by a qualified expert, different aspects are taken into account, among them: the area (surface of the land) in square meters, the construction area, the location, the number of floors, the number of rooms, the number of bathrooms that the property has, the finishes, the age, among others.

The appraisal is defined as: "the calculation process for the determination of the appraisals of the land, construction and additional construction, of the real estate of the Metropolitan District of Quito, according to the methodology and valuation techniques, within the framework of the that establishes the Organic Code of Territorial Organization, Autonomy and Decentralization-COOTAD and Ministerial Agreement 029 that issues the "National Norms for the Property Registry"

The collected data correspond to the publications of the real estate in the different web pages in Ecuador dedicated to the sale of real estate, which were obtained through the use of web scraping, this tool was applied in pages such as Plusvalía, Properati, Remax. With the data obtained, a column cleaning procedure was carried out, as well as the elimination of records.

With the result of the extracted data and once the cleaning process was applied, the data were organized by sectors and the implementation of automatic learning techniques was carried out, such as: decision tree, neural networks, random forest for each data set. Consequently, applying the different algorithms resulted in the linear regression with the highest performance compared to the other applied models.

Keywords: appraisal, data, machine learning, pandemic, real estate.

CAPÍTULO I

INTRODUCCIÓN

1.1 Planteamiento del problema

Debido a que el área inmobiliaria ha sido afectada por la pandemia, el desempleo se ha convertido en parte del deterioro de la economía de las personas, imposibilitándolas a que puedan adquirir un bien inmueble, pese a ello y con la llegada de la nueva normalidad el mundo inmobiliario ha podido restablecerse poco a poco, esto debido a la aplicabilidad de la tecnología. “La pandemia ha sido un elemento catalizador de los cambios en la percepción de la vivienda y en los hábitos de los compradores que han descubierto que necesitan viviendas en las que confluían confort, digitalización, funcionalidad, flexibilidad, sostenibilidad y espacios más amplios. Por ello, la venta de viviendas unifamiliares se disparó durante los primeros 7 meses de 2021 respecto de años anteriores”, (Higueras, 2021)

En cuanto a ciertas resoluciones usadas para la obtención de un bien inmueble, consta la formación de los precios inmobiliarios, esta responde a la misma lógica que de activos financieros donde la especulación es un comportamiento común. (Zakaria & Fatine, 2021)

Dado que las cifras de la Asociación de Promotores Inmobiliarios de Vivienda del Ecuador (APIVE), las reservas netas del sector entre enero a junio de 2021 crecieron en un 44% respecto a 2020, mientras que las visitas a proyectos crecieron en un 11% durante el mismo periodo en 2021 versus el año anterior. (Ekos, 2021). Lo que indica el gradual crecimiento en el sector inmobiliario.

De acuerdo a la Superintendencia de Bancos en su señalamiento a las normativas legales para calificarse como peritos evaluadores, manifiesta que “todas las personas naturales y jurídicas que tengan en regla su documentación pueden calificarse como peritos valuadores y pueden brindar sus servicios profesionales en las entidades de los sectores financieros público y privado”.

(Superintendencia De Bancos, 2021). De modo que dichos profesionales calificados realizan la tasación de un inmueble.

La experiencia de los profesionales calificados en una tasación es importante, así pues, se debe contemplar la particularidad del sitio en donde está ubicado el inmueble ya que esto puede influir al momento de valorar un bien y por tanto hará que su cotización aumente o disminuya, a modo de ejemplo se puede citar el caso de Quito, allí los valores de inmuebles en el sector norte de la ciudad no son semejantes a aquellos inmuebles del sector sur de Quito.

Entre otros aspectos para la tasación de un bien inmueble contempla un tiempo determinado, tomando en cuenta los recursos del profesional calificado antes mencionado, como movilidad, herramientas de medición, etc., también es importante analizar las circunstancias geográficas que engloban las características de la zona en donde se encuentra el inmueble.

Teniendo en cuenta la descripción del territorio por el municipio de Quito, donde se manifiesta “El DMQ está localizado sobre la región central de la sierra ecuatoriana en la provincia de Pichincha, se despliega sobre un contexto geográfico muy irregular.” (Distrito Metropolitano de Quito, 2014). En lineamiento a lo anterior se puede inferir que, para el caso de Quito, uno de los factores a considerar para realizar el avalúo, es el estrato de una vivienda, tomando como base siempre seguir algunas ordenanzas municipales de esta zona.

Este trabajo plantea aplicar técnicas de aprendizaje automático para pronosticar valores de bienes inmuebles en la ciudad de Quito, se trata de aplicar algoritmos, los cuales mediante una forma sencilla facilitan el acceso y disponibilidad de la información, para en consecuencia obtener resultados de una manera más dinámica.

1.2 Justificación

La industria inmobiliaria es sumamente vasta e involucra otras muchas industrias, que incluyen terrenos, materiales de construcción, construcción de edificios, decoración, electrodomésticos y otras industrias. (Amenga, 2020).

La llegada de la pandemia supuso nuevos retos para el resurgimiento de algunos sectores estratégicos, en la situación del área inmobiliaria la tecnología fue la apuesta segura para que la misma fortalezca su economía, de modo que las personas interesadas en adquirir o vender un inmueble usan medios disponibles como son páginas web, redes sociales. Por consiguiente, surge la necesidad de realizar un análisis sobre este tema.

Entre otros aspectos relevantes consta la financiación que buscan las personas al momento de adquirir un inmueble, ya sea proveniente de alguna institución que pueda proporcionar los denominados préstamos quirografarios o hipotecarios, dependiendo del valor del inmueble adquirir.

El presente proyecto pretende construir un método automático que, a aporte en la adquisición de un inmueble, poniendo énfasis en la predicción del valor del bien, como una solución rápida para este objetivo, se implementará técnicas de aprendizaje automático dinamizando a la gran información existente en Internet junto a un proceso de tratamiento de datos extraídos para realizar la predicción lo más cercanas a los valores reales.

Este trabajo toma como referencia la investigación presentada en Colombia por el ingeniero Juan Carlos Borrero Calixto, quien aplicó métodos automáticos en su trabajo de titulación, con el objetivo de pronosticar el avalúo comercial de viviendas en la ciudad de Bogotá. Por esta razón, el presente proyecto de titulación se implementará las técnicas de aprendizaje automático para predecir el valor de un inmueble en la ciudad de Quito.

1.3 Objetivos

1.3.1 Objetivo General

Aplicar métodos automáticos para pronosticar el valor de un bien inmueble en la ciudad de Quito utilizando técnicas aprendizaje automático.

1.3.2 Objetivos Específicos

Se contemplan los siguientes objetivos para el presente trabajo:

- Recolectar datos usando web-scraping en sitios webs inmobiliarios de Ecuador.
- Determinar el modelo que más se aproxime en la predicción al valor de inmueble.
- Evaluar el rendimiento del método automático mediante las medidas de desempeño como es el coeficiente de variación.

1.3.3 Alcances y limitaciones

- Este proyecto tiene como enfoque principal predecir los valores comerciales de bienes inmuebles particulares en una zona determinada, dejando de lado bienes que pertenecen al sector público.
- El área donde se aplicará este proyecto será la ciudad de Quito por ser una de las ciudades más importantes del país y de mayor territorio, lo que establece un gran mercado de compra y venta en el sector inmobiliario.
- El tipo de inmuebles analizados en este trabajo, serán casas y departamentos.

1.4 Marco Conceptual

En la presente sección se revisará una serie de conceptos abordados en los que se sustentan el presente proyecto. Los conceptos a tener en cuenta son: Quito, inmueble, normas de valoración de bienes inmuebles, valor de mercado, avalúo del predio, mercado inmobiliario, valoración de bienes inmuebles entre los más estacados.

1.4.1 Quito

El Distrito Metropolitano de Quito es el cantón con más población del Ecuador. Quito aparte de ser la capital del Ecuador cuenta con un constante crecimiento del sector inmobiliario, de acuerdo al comportamiento en este sector. “en Quito existe promedio 580 proyectos en marcha en los últimos 4 años, y Guayaquil 89” (marketwatch, 2020.). En la figura 1 se puede observar el mapa de las administraciones zonales, las cuales son 9, estas se dividen en 32 parroquias urbanas, 33 parroquias rurales y suburbanas.



Figura 1. Muestra la distribución zonal de la ciudad de Quito.
Fuente: <https://www.quito.gob.ec>

1.4.2 Inmueble

Un inmueble es una propiedad que no puede movilizarse, se caracteriza por ser un bien que está pegado al suelo, ya sea porque forma parte de él o porque está adherido a él, por ejemplo: una casa, un edificio, un local comercial, un terreno, etc. (Concepto Jurídico, 2021)

1.4.3 Norma técnica de valoración de bienes inmuebles

“Conjunto de conceptos, criterios y reglas que norman y regulan el cálculo para la determinación de la valoración de los bienes inmuebles”. (Norma técnica de Valoración de los Bienes Inmuebles en el MDMQ, 2019)

1.4.4 Valor de mercado

Es el valor que un comprador estaría dispuesto a pagar y que un vendedor estaría dispuesto a recibir por un bien inmueble, considerando las condiciones del mercado. (Norma técnica de Valoración de los Bienes Inmuebles en el MDMQ, 2019)

1.4.5 Avalúo del predio

Es el avalúo de los bienes inmuebles que consta en el sistema catastral predial. (Norma técnica de Valoración de los Bienes Inmuebles en el MDMQ, 2019)

1.4.6 Mercado Inmobiliario

El mercado inmobiliario es el conjunto de las acciones de oferta y demanda de bienes inmuebles. La naturaleza de estos bienes puede ser muy distinta, diferenciándose entre bienes de naturaleza residencial, comercial, industrial, urbano, etc. (Realía, 2022)

1.4.7 Valoración de bienes inmuebles

Es el proceso de cálculo de un valor económico, establecido de acuerdo con unas metodologías de evaluación y a unas técnicas que tratan de construir un valor objetivo, atendiendo a las características propias del bien y de las características de su entorno. (Vozmediano, 2018)

1.4.8 Marco Legal

Esta ley dentro de su objetivo manifiesta: “establecer los principios y reglas generales que rigen el ejercicio de las competencias de ordenamiento territorial, uso y gestión del suelo urbano y rural,”. (Ley Orgánica de Ordenamiento Territorial, Uso y Gestión del Suelo , 2016)

En el Capítulo I, denominado Objeto y Ámbito, (Ley Orgánica de Ordenamiento Territorial, Uso y Gestión del Suelo, 2016) determina: orientar las políticas públicas relativas al ordenamiento territorial, desarrollo urbano, ¡derecho a la vivienda adecuada y digna; promover un uso eficiente, equitativo, racional y equilibrado del suelo urbano y rural a través de la definición de principios, directrices y lineamientos, y generar un hábitat seguro y saludable en todo el territorio. (pág. 4).

1.4.8.1 Plan metropolitano de Quito para el desarrollo de ordenamiento territorial

Según lo determina la página oficial de municipio de Quito, el “13 de septiembre de 2021 el Concejo Metropolitano de Quito aprobó la actualización del Plan Metropolitano de Desarrollo y Ordenamiento Territorial (PMDOT) 2021 – 2033 del Distrito Metropolitano de Quito” (Municipio de Quito, 2022).

1.4.8.2 Estructura Urbano Rural

El Modelo Territorial Deseado (MTD) del Plan Metropolitano de Desarrollo y Ordenamiento Territorial (PMDOT), determina el modelo de ocupación del territorio y sustenta su desarrollo con los siguientes ejes: gobernabilidad e institucionalidad social, territorial, económica y social.

En base a (PLAN METROPOLITANO DE DESARROLLO DE ORDENAMIENTO TERRITORIAL, 2021-2023), el trabajo del territorio está encaminado a planificar, orientar, clasificar, subclasificar y aprovechar el uso del suelo. El Modelo Territorial definido en el PMDOT y los sistemas que lo componen son:

SISTEMA AMBIENTAL Y DE RIESGO NATURAL

- Áreas Naturales Protegidas
- Zonas de susceptibilidad y amenazas a fenómenos naturales.
- Polígonos industriales y zonas agroproductivas.
- Zonas de aprovechamiento extractivo.

- Sistema de Microrregiones Rurales Sostenibles.
- Sistema Policéntrico.
- Sistemas Públicos de Soporte.

- Servicios básicos
- Equipamientos
- Sistema Vial
- Sistema de Movilidad y Conectividad

1.4.8.3 Clasificación de Suelo

La distribución de suelo según el (PLAN METROPOLITANO DE DESARROLLO DE ORDENAMIENTO TERRITORIAL, 2021-2023), en el Capítulo III manifiesta: “la clasificación de suelo se definirá a partir del análisis de las proyecciones de crecimiento y densidad poblacional, la consolidación de la ocupación de los predios, estructura predial y su nivel de fraccionamiento, amanzanamiento y estructura vial regular, capacidad de los sistemas públicos de soporte que garanticen la dotación de servicios como los de agua potable, alcantarillado, luz eléctrica, aseo de calles y otros de naturaleza semejante; suelo urbano existente (consolidado y vacante), áreas industriales, factores de riesgo, productivos y de protección; y a la vez, se determinará la necesidad real para extender, disminuir o mantener el límite.”

Suelo Urbano

Es aquel ocupado por asentamientos humanos concentrados, que están dotados total o parcialmente de infraestructura básica y servicios públicos, y que constituye un sistema continuo e interrelacionado de espacios públicos y privados. (Plan de Uso y Gestión de Suelo del Distrito Metropolitano de Quito, 2021)

Suelo Rural

Es aquel destinado principalmente a actividades agro-productivas, extractivas, o forestales, que, por su condición natural o ambiental, su interés paisajístico, histórico, cultural, o que, por sus especiales características biofísicas o geográficas debe ser protegido o reservado para futuros usos urbanos. (Plan de Uso y Gestión de Suelo del Distrito Metropolitano de Quito, 2021).

En la figura 2 se muestra la clasificación de suelos en la cuida de Quito como lo son suelo rural en color amarillo y el suelo urbano de color azul.

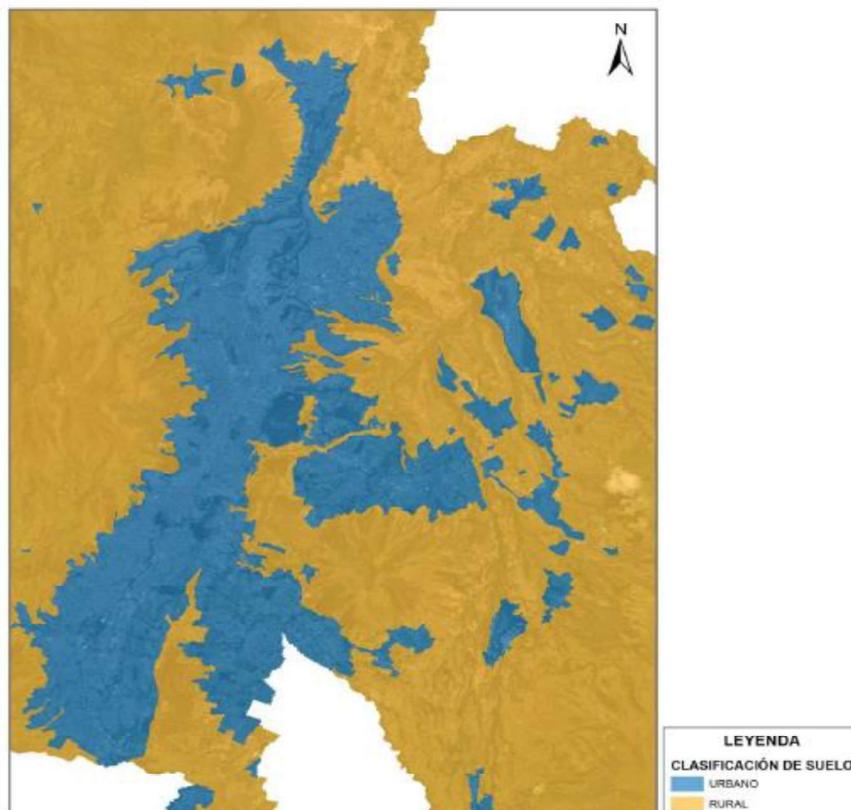


Figura 2. Muestra la clasificación de suelos en DMQ.
Fuente: Plan de Uso y Gestión del Suelo

Aprovechamiento Urbanístico

El capítulo III del (Plan de Uso y Gestión de Suelo del Distrito Metropolitano de Quito, 2021), este documento manifiesta que el aprovechamiento urbanístico acuerda la utilización de suelo, por medio de normativas de uso, la construcción de edificaciones, la planificación y gestión.

1.4.8.4 Uso de suelo

De acuerdo a (Plan de Uso y Gestión de Suelo del Distrito Metropolitano de Quito, 2021): “el uso es la destinación asignada al suelo conforme con su clasificación y subclasificación en relación con las potenciales densidades edificatorias y a las actividades a ser desarrolladas”.

En correspondencia al (Plan de Uso y Gestión de Suelo del Distrito Metropolitano de Quito, 2021), el uso en general se caracteriza un límite espacial, en la cual contienen las actividades a desarrollarse en su mayoría, los usos generales son:

RESIDENCIAL (R): Corresponde al uso asignado a los polígonos de intervención territorial urbanos o rurales, que tienen como destino mayoritario la vivienda permanente.

SUELO DE EXPANSIÓN (SE): “Corresponde al uso asignado a los polígonos de intervención territorial que únicamente pertenecen a la subclasificación de suelo rural de expansión urbana, en los cuales su destino dependerá de la situación territorial circundante y las decisiones de planeamiento para su incorporación a suelo urbano. Podrán permitir actividades residenciales, de protección natural y cultural, múltiples e industriales según corresponda, de acuerdo con su planificación detallada mediante plan parcial.”

MÚLTIPLE (M): “Corresponde al uso asignado a los polígonos de intervención territorial en los que pueden coexistir residencia, comercio, industrias de bajo impacto, servicios y equipamientos compatibles de acuerdo con las disposiciones del PUGS”.

PATRIMONIO CULTURAL (PC): “Corresponde al uso asignado a los polígonos de intervención territorial urbanos y rurales, que mayoritariamente contengan o constituyan bienes patrimoniales de orden natural, espacial o cultural, tangibles e intangibles, que forman parte del proceso de desarrollo de los asentamientos humanos y que han adquirido tal significado social, que los hace representativos de su tiempo y de la creatividad humana.”

INDUSTRIAL (I): “Corresponde al uso asignado a los polígonos de intervención territorial urbanos, cuyo destino mayoritario es la elaboración, transformación, tratamiento y manipulación de materias primas para producir bienes o productos materiales, así como su almacenamiento, mediante la utilización de maquinaria especializada y recursos humanos organizados, de acuerdo con las disposiciones del PUGS.”

EQUIPAMIENTO (E): “Corresponde al uso asignado a los polígonos de intervención territorial urbanos y rurales; los cuales, están destinados a actividades e instalaciones que generen bienes y servicios para satisfacer las necesidades de la población, garantizar el esparcimiento y mejorar la calidad de vida en el distrito, independientemente de su carácter público o privado.”

PROTECCIÓN ECOLÓGICA (PE): Corresponde al uso asignado a los polígonos de intervención territorial urbanos o rurales en los que su uso mayoritario se destina a la conservación del patrimonio natural. Permite las actividades de gestión ambiental y ecológica con el objetivo principal de proteger y restaurar la diversidad ecológica. Incluye a las áreas naturales protegidas que forman parte del Sistema Nacional de Áreas Protegidas (SNAP) y Bosque Protectores dentro del DMQ, al Patrimonio Natural Distrital constituido por el Subsistema Metropolitano de Áreas Naturales Protegidas (SMANP), y demás áreas de protección y conservación de acuerdo con las disposiciones de este plan.

RECURSO NATURAL RENOVABLE (RNR): Corresponde al uso asignado a los polígonos de intervención territorial en suelo rural, en los cuales su uso mayoritario es destinado al manejo, producción sostenible, uso racional y eficiente de los recursos naturales renovables, debido a sus condiciones biofísicas y agroclimáticas y a la aptitud del suelo. Incluye a la producción agrícola,

pecuaria, forestal, silvícola o acuícola, actividades recreativas, ecoturísticas, de protección agraria y otras actividades productivas.

La figura 3 muestra la zona de uso de suelo de los recursos no renovables del DMQ, dicho suelo posee una serie de normativas que respaldan su preservación y explotación sostenible.

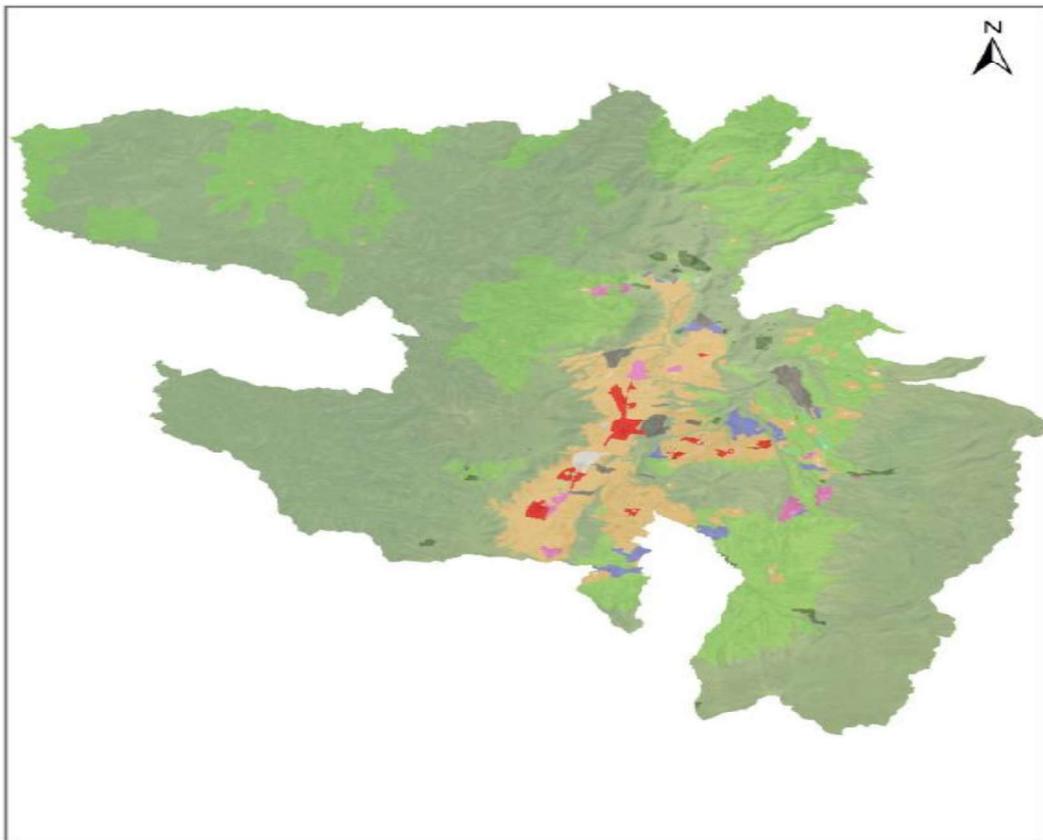


Figura 3. Muestra los recursos no renovables (RNNR) del DMQ.
Fuente: Plan de Uso y Gestión del Suelo

1.4.9 Conjunto de datos

Se define como un grupo de números, o bytes, que a menudo se muestran en una tabla con las columnas que clasifican los datos en subconjuntos. (Spiegato, 2022)

1.4.10 Inteligencia Artificial

“La Inteligencia Artificial (IA) es un subcampo de la informática que se creó en la década de 1960, y que trata de solucionar tareas que son sencillas para los seres humanos, pero difíciles para las computadoras. (García, 2022)

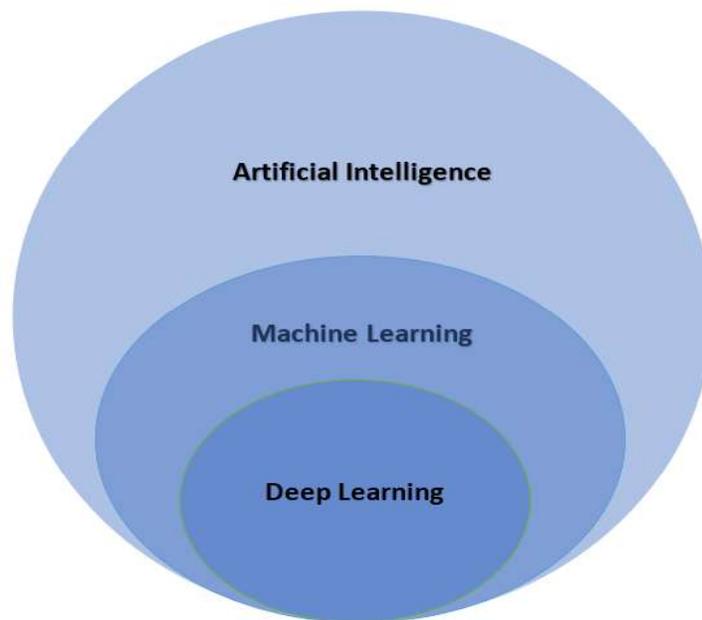


Figura 4. Muestra cómo la Inteligencia Artificial engloba muchas tareas.
Fuente: Elaboración propia

La figura 4 muestra los tipos de inteligencia artificial existentes, se puede adiestrar cosas mediante aprendizaje automático, sin embargo, pueden también instruirse por sí mismo como es el caso del aprendizaje profundo.

1.4.11 Web Scraping

El web scraping es una técnica que se utiliza para la extracción y almacenamiento de información de cualquier página web a través de un programa de software, que suele ser un crawler. (Koller, 2022)

La figura 5 muestra el proceso de recopilación de datos web de forma estructurada y modo automático, de esta manera se puede aprovechar la gran cantidad de datos disponibles en la web.

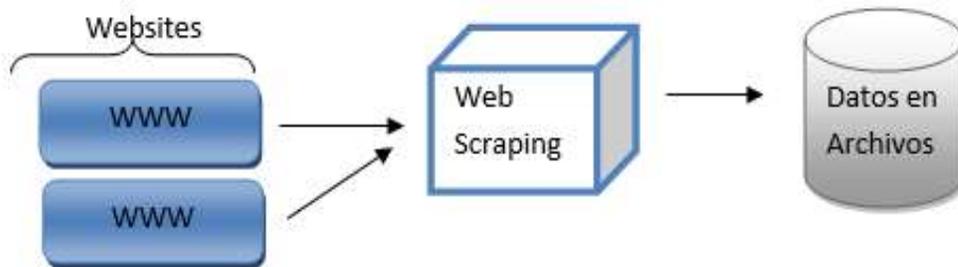


Figura 5. Muestra el proceso de funcionamiento de software web scraping.
Fuente: Propia

Machine Learning

El Machine Learning en su uso más básico es la práctica de usar algoritmos para analizar datos, aprender de ellos y luego ser capaces de hacer una predicción o sugerencia sobre algo. Los programadores deben perfeccionar algoritmos que especifiquen un conjunto de variables para ser lo más precisos posibles en una tarea en concreto. (Rodríguez, 2017)

En la figura 6 se puede ver el desarrollo de aprendizaje basados en la experiencia como entrada, su capacidad descubrir patrones y producir predicciones como salida.



Figura 6. Muestra el proceso de aprendizaje que implementa Machine Learning.
Fuente: Elaboración propia

1.4.12 **Aprendizaje Supervisado**

El aprendizaje supervisado son un conjunto de técnicas que permite realizar predicciones futuras basadas en comportamientos o características analizadas en datos históricos etiquetados. Una etiqueta no es más que la salida que ha mostrado el conjunto de datos para datos históricos, ya conocidos. (Calvo, 2019)

1.4.13 **Página web**

Se define una página web como un documento disponible en Internet, o World Wide Web (www), codificado según sus estándares y con un lenguaje específico conocido como HTML. (Oliver, 2019)

1.4.14 **Plataformas Inmobiliarias**

A medida que las empresas se expanden y publican cada vez más información sobre sus actividades comerciales en sus sitios web, los datos de estos sitios prometen ser una fuente valiosa para investigar la innovación. (Abdullah, Waterworth, & Shapira, 2015)

1.4.14.1 Properati

Sitio web digital de bienes que se enfoca en propiedades para la venta y arriendo. “Quienes busquen un nuevo hogar o quieran invertir en propiedades encontrarán en Properati, además de las ofertas disponibles, valiosa información para tomar las mejores decisiones: promedios de precios, características de los barrios, comparaciones, etc.” (Properati, 2002).

En la figura 7 se muestra el entorno del sitio web, el cual consta de propiedades con fechas más recientes de publicación, un menú de búsquedas y un espacio para encontrar otras ofertas.

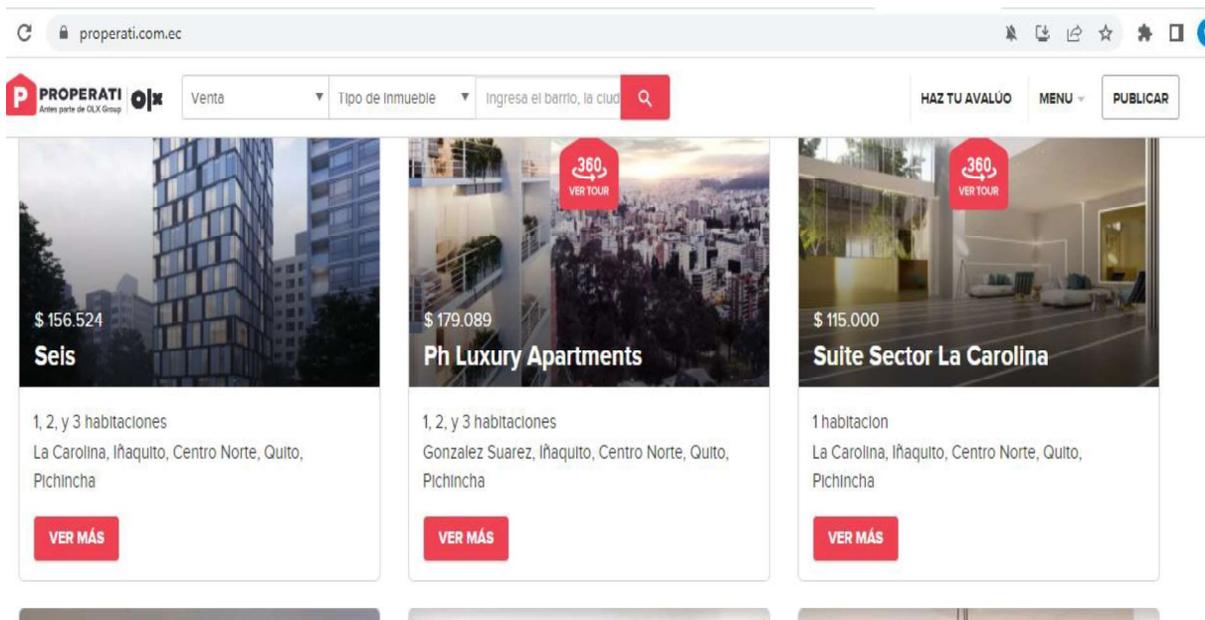


Figura 7. Muestra la interfaz de la página web Properati en Ecuador.
Fuente: www.properati.com.ec

1.4.14.2 Plusvalía

Esta plataforma utiliza un buscador para filtrar las opciones por localidad y definir el rango de precios de interés para el usuario. (Plusvalía, 2022). Al igual que PROPERATI se encuentra presente en varios países de Latinoamérica, su objetivo principal es ayudar a encontrar la propiedad de interés del cliente a través de su portal web, en donde se ofertan inmuebles como casas, departamentos, oficinas, terrenos y otros inmuebles.

La búsqueda a través de filtros en la página web de Plusvalía mejora la experiencia al momento de localizar el bien de interés para el usuario, así como se muestra en la figura 8.



Figura 8. Muestra la interfaz de la página web de Plusvalía.

Fuente: <https://www.plusvalia.com/>

1.4.14.3 Remax

El sitio web Remax, fundado en 1973, es uno de los sitios digitales que se encuentra presente en Latinoamérica, destacándose por contar con agentes de bienes raíces de RE/MAX, quienes apoyan al comprador durante todo el proceso de venta, ofreciéndole todos los servicios que necesita para asegurar la venta exitosa de propiedades.

En su sitio web se puede apreciar un gran catálogo digital de propiedades de interés para el usuario, allí se ofertan bienes tanto para la compra y alquiler en distintos lugares, ofreciendo una lista variada de opciones.

En la figura 9 se puede observar el sitio web de Remax para Ecuador con la particularidad de que esta página ofrece el servicio de compra, venta de inmuebles mediante una agente.

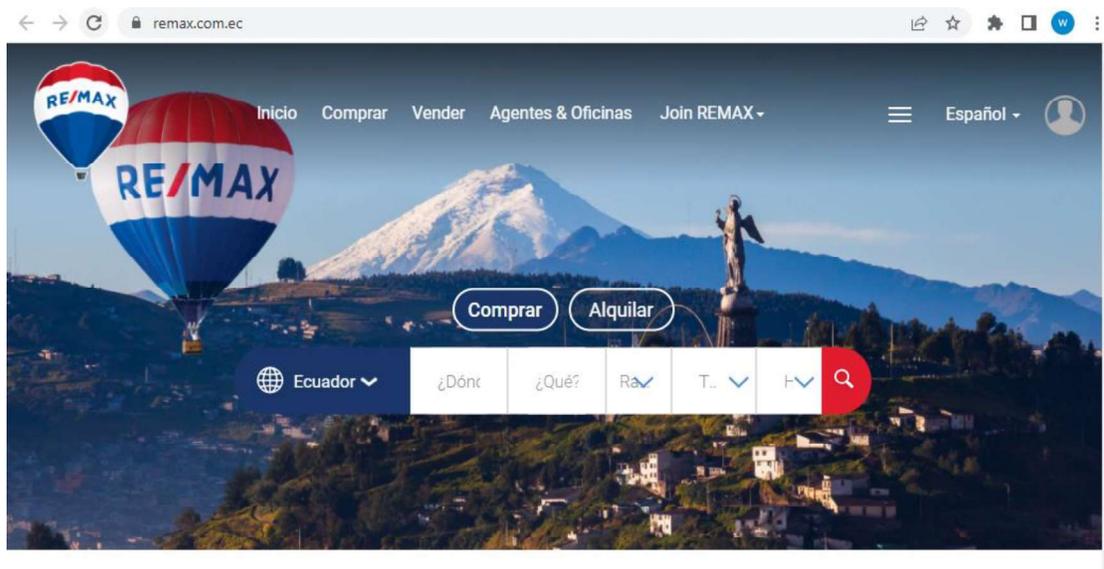


Figura 9. Muestra la interfaz de la página web Remax para Ecuador
Fuente: <https://www.remax.com.ec/>

1.4.15 Las Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA) son capaces de aprender, generalizar resultados y responder adecuadamente a datos incompletos o previamente desconocidos. (Chiarazzo, Caggiana, & Marinellia, 2014).

Una red neuronal (NN) es un proceso informático inspirado libremente en la estructura del cerebro humano, proporciona un marco para que múltiples algoritmos de aprendizaje automático trabajen juntos para procesar datos complejos.

En la figura 10 se muestran las capas de una red neuronal, la cual está compuesta “n” neuronas de entrada, “m” neuronas en la capa oculta y una neurona de salida.

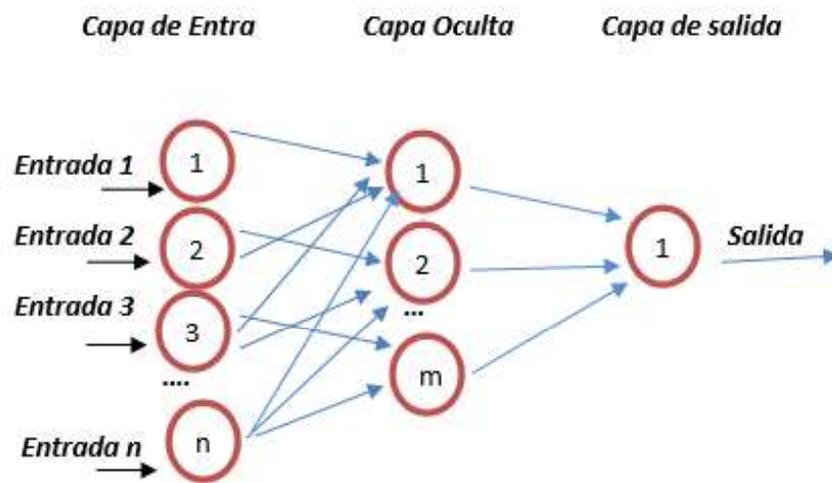


Figura 10. Muestra como una la red neuronal.

Fuente: Elaboración propia

Paradigmas del aprendizaje

Aprendizaje Supervisado

El proceso de aprendizaje supervisado consiste en entrenar la red neuronal a partir de un conjunto de datos de entrada y sus respectivas salidas, el algoritmo de aprendizaje ajusta los parámetros de la red de tal manera que la salida generada por la ANN se ajuste a los datos de salida dada una cierta entrada. Se dice que es supervisado ya que se conoce desde el inicio cual es el patrón de salida. (Jarrín & Riofrío, 2019)

En la figura 11 se pueden visualizar diferentes modelos de algoritmos, supervisados con aquellos con los que puede trabajar para resolver varios problemas.

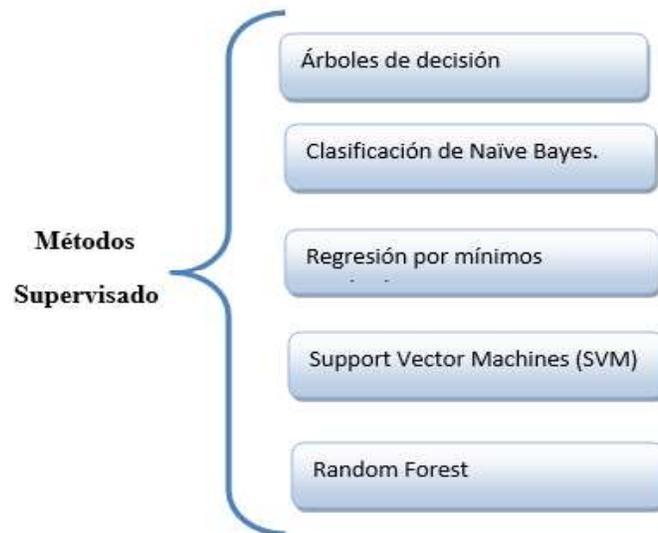


Figura 11. Muestra métodos de aprendizaje supervisados.
Fuente: Elaboración propia

Auto-organización (No supervisado)

En este tipo de aprendizaje no se concibe la figura “supervisor” y su principal objetivo es implementar una salida, es en base a la redundancia en las entradas, el aprendizaje lo extrae de los patrones que consigue por la entrada de datos.

“Las redes neuronales emplean su capacidad de aprendizaje adaptativo para autoorganizar la información que reciben durante el aprendizaje y/o la operación. Mientras que el aprendizaje es la modificación de cada elemento procesal, la autoorganización consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico” (Matich, 2001).

Las aplicaciones primordiales del aprendizaje no supervisado están vinculadas con el agrupamiento de los algoritmos más utilizados, así como lo podemos ver en la figura 12.

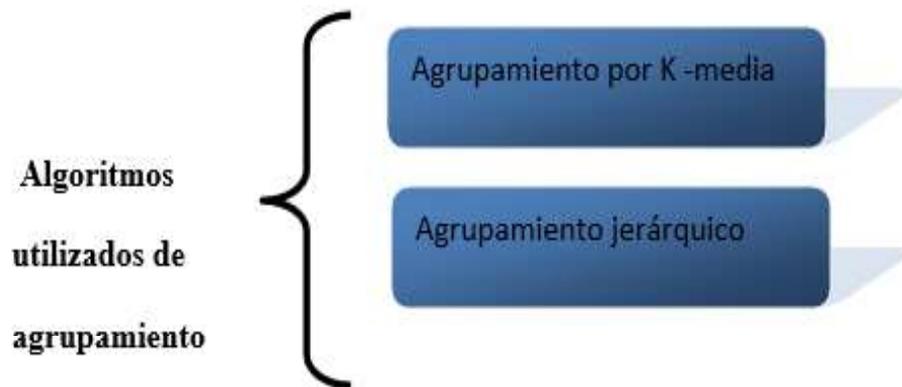


Figura 12. Muestra algoritmos de agrupamiento.
Fuente: Elaboración propia

1.4.16 Regresión Lineal

Los modelos de Regresión Lineal son muy populares y, por lo tanto, ampliamente utilizados en muchas áreas diferentes de la ciencia (Yalçınkaya, Balay, & Enoğlu, 2021).

“El método de MCO (mínimos cuadrados ordinarios) es el más común en el análisis de regresión, sobre todo por ser mucho más intuitivo y matemáticamente más sencillo que el método de máxima verosimilitud.” (Gujarati, 2009), según el autor este modelo plantea siete supuestos:

Supuesto 1.-Modelo de Regresión Lineal

El supuesto manifiesta que es lineal en los parámetros, aunque puede o no ser lineal en las variables.

Supuesto 2.- Valores fijos de X

Para este supuesto determina que los valores que toma la x pueden considerarse fijos en muestras repetidas.

Supuesto 3.- El valor medio de la perturbación

El supuesto establece que el valor de la media de u_i , que depende de las X_i dadas es cero, geoméricamente este supuesto se representa mediante la figura 13, demostrando algunos valores de la variable X y las poblaciones Y, asociadas a cada uno de ellos. (Gujarati, 2009)

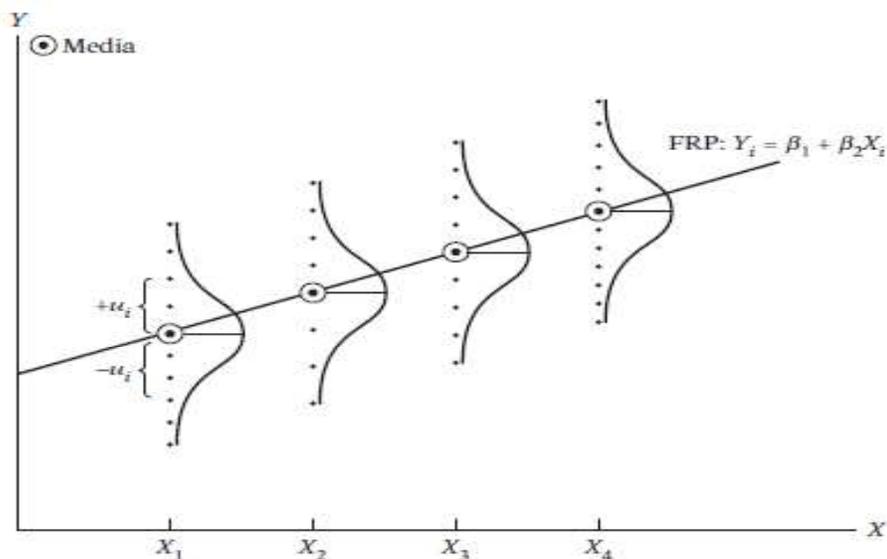


Figura 13. Distribución de la condicional.
Fuente libro: GUJARATI, Econometría, Quinta Edición, México

Supuesto 4.- Varianza constante

La varianza del término de error o de perturbación, es la misma sin importar el valor de X.

Supuesto 5.-No hay autocorrelación entre las perturbaciones

El supuesto indica que, dados dos valores que pueden ser X , X_i y X_j , la correlación entre dos u_i y u_j , podrían sacar como resultado cero, es decir a la correlación están ausentes. Las desviaciones siguen patrones sistemáticos para indicar la existencia de la correlación.

En la figura 14 se muestra que no existe un patrón sistemático de u , esto señala que la correlación es cero.

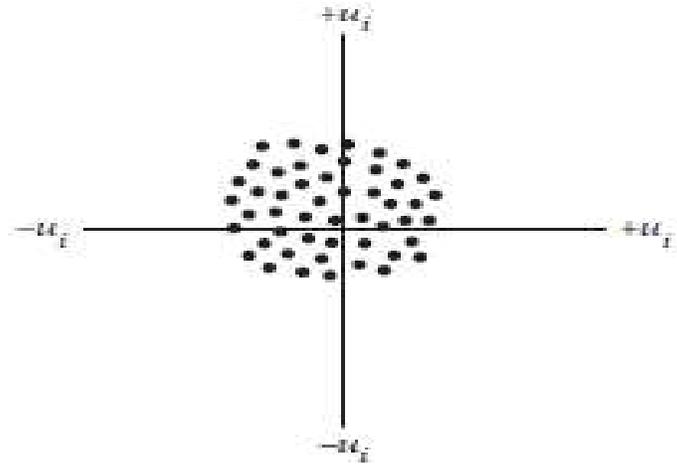


Figura 14. Muestra que no hay un patrón sistemático.
Fuente libro: GUJARATI, Econometría, Quinta Edición, México

Supuesto 6.- El número de observaciones n debe ser mayor que el número de parámetros por estimar

El supuesto indica que el número de observaciones debe ser mayor que el número de variables explicativas.

Supuesto 7.- La naturaleza de las variables X

El supuesto indica que no todos los valores X en una muestra determinada deben ser iguales, al igual muestra que no puede haber valores atípicos de la variable X , es decir, valores muy grandes en relación con el resto de las observaciones.

1.4.17 Árbol de decisión

El método del Árbol de decisión es un algoritmo de clasificación y regresión universal muy bien conocido, en donde los probadores pueden verificar las construcciones internas o el rendimiento de la aplicación (Tahraoui, Amrane, Belhadj, & Zhang, 2021)

Es un método de aprendizaje supervisado, ya que su desarrollo precisa de una variante dependiente en el grupo de entrenamiento, favorece a la realización de modelos predictivos de analítica de datos que comparten grandes volúmenes de información. Su clasificación se fundamenta según diferentes características o propiedades, mediante la relación entre distintas variables para predecir el valor de otra.

Estructura

Los árboles de decisión están formados por nodos, ramas y su sentido para leer se realiza de arriba hacia abajo.

La figura 15 muestra de manera gráfica los acontecimientos que pueden resultar de las variables que contengan sus datos y, a su vez, de todos los sucesos que finalmente deriven en la toma de una decisión.

- Nodo raíz: es la primera segmentación en función de la característica más importante.
- Ramas: representan la decisión en función de una determinada condición
- Hojas: se ubican en la parte inferior e indican la clasificación final

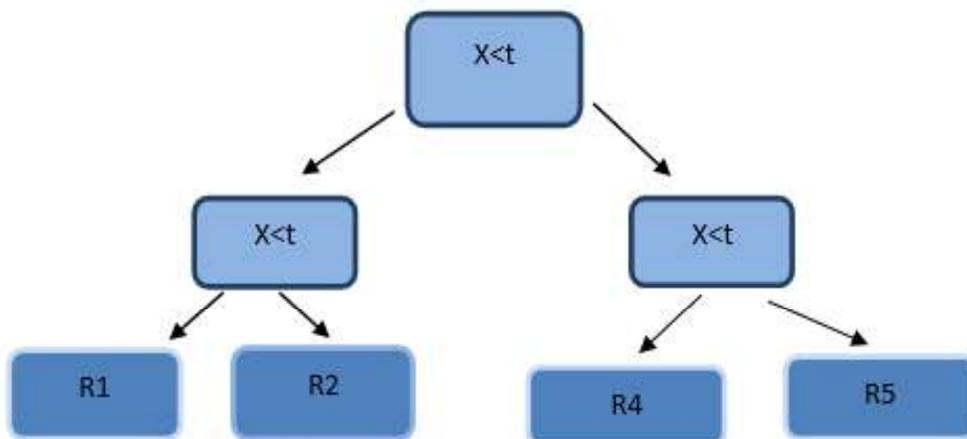


Figura 15. Muestra la formación de un árbol de decisión.
Fuente: Elaboración Propia

1.4.18 Random Forest

Es un algoritmo de predicción que contiene múltiples árboles de decisión. En el proceso de modelado, no existe relación entre cada árbol de decisión; por lo tanto, cada árbol se modela de forma independiente. (Linqi, Zhou, Liu, & Kong, 2021)

¿Cómo funciona Random Forest?

Consiste en la elaboración de una gran cantidad de árboles de decisión individuales que actúan como un grupo, cada árbol en el bosque aleatorio pronostica una clase y la clase con más votos se transforma en la predicción de nuestro modelo, este proceso se detalla de mejor manera en la Tabla 1.

Tabla 1. Detalla el procedimiento del método denominado Random Forest

Pasos	Descripción
1	En el bosque aleatorio, se toman n números de registros aleatorios del conjunto de datos que tiene k números de registros.
2	Se construyen árboles de decisión individuales para cada muestra.
3	Cada árbol de decisión generará una salida.
4	El resultado final se considera en función de la votación mayoritaria o el promedio para la clasificación y la regresión, respectivamente.

Clasificación en Random Forest

La clasificación en bosques aleatorios emplea una metodología de conjunto para alcanzar el objetivo, presenta la utilidad de manejar grandes volúmenes de datos, así como manipular una extensa cantidad de variables de entrada e identificar las más relevantes.

Como se muestra en la figura 16 el modelo de Random Forest, crea muchos árboles por lo que hace predicciones muchos más ventajosas y satisfactorias al tomar la clase con más votos.

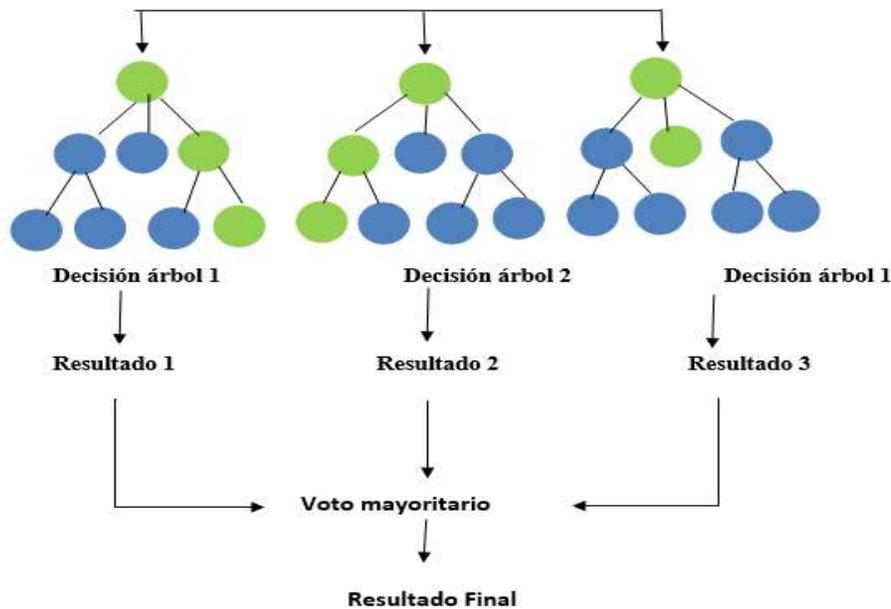


Figura 16. Muestra el método de Random Forest
Fuente: Elaboración Propia

1.4.19 Medida de Desempeño

1.4.19.1 Coeficiente de Determinación R^2

El R^2 tiene como objetivo medir el desempeño del modelo indicador que ayuda a definir o determinar si el mismo se ajusta al precio de la vivienda, para este caso el coeficiente de determinación tiene como propósito valorar que tan eficiente es un modelo. Este factor permite decidir que tanto se adaptan los modelos al valor del inmueble.

$$\frac{\sum_{t=1}^T (\hat{Y}_i - Y_t)^2}{\sum_{t=1}^T (Y_i - Y_t)^2} \quad (1)$$

Y_i = Valor del inmueble conocido.

\hat{Y}_i = Valor del inmueble predicho por el modelo.

1.4.19.2 Error cuadrático medio RMSE.

Como se puede ver en la ecuación 1, el coeficiente de determinación se obtiene con la sumatoria del valor del inmueble conocido, menos el valor originado por cada uno de los modelos, ya que esto se divide para la sumatoria de los valores conocidos menos el valor resultante de los modelos.

El error cuadrático medio es una medida que indica la equivocación entre un conjunto de datos, se equipara un valor que se pronostica con un valor real, se halla representado por las siglas RMSE; esta predicción cuantifica qué tan desigual es un conjunto de resultados o rendimientos, cuanto más pequeño es el valor demuestra que se aproxima al valor real

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (2)$$

P_i = Precio de vivienda predicho por el modelo

O_i = Precio de vivienda

n = número total de predicciones

Como se puede apreciar en la ecuación 2, el RMSE se obtiene con el valor de cada algoritmo que ha predicho en referencia al inmueble, menos el valor conocido en conjunto de pruebas, esto permite tazar que tan semejante o apartado está al valor real.

1.4.19.3 Coeficiente de Variación con el RMSE (CV)

Es una media estadística que es conocida como coeficiente de variación de Pearson, la cual indica que tan dispersas se encuentran estas cifras del conjunto de datos.

(3)

$$CV = \sqrt{\frac{RMSE}{|\bar{x}|}}$$

RMSE: Raíz del error cuadrático medio.

$|\bar{x}|$: es la media de la variable Y , es decir, el conjunto de precios de todas las viviendas.

En la ecuación 3, se muestra como el coeficiente de variación es obtenido mediante el valor *RMSE*, alcanzado en la ecuación 2; anteriormente se calculaba cada uno de los modelos sobre el valor de la media la variable Y , el cual pertenece al grupo de valores de los inmuebles en nuestros datos.

1.4.20 ROC Y AUC

Cuando necesitamos comprobar o visualizar el rendimiento de problemas de clasificación se utiliza la gráfica ROC y el área bajo la curva AUC, están compuestas por el eje y, en donde se encuentra la tasa de verdaderos positivos y el eje x conformado por falsos positivos, AUC representa el área baja de la curva, mostrando que una situación ideal se produce cuando las curvas no se superponen $AUC=1$.

Cuando AUC es igual 0.70 significa que hay un 70% de probabilidad y que el modelo puede distinguir entre clase positiva y clase negativa, cuando la AUC es igual a 0,5 el modelo no tiene la capacidad de distinguir entre clase de positiva y clase.

1.4.21 Herramientas de Machine Learning

El aprendizaje automático contiene una gran variedad de herramientas, plataformas y software de aprendizaje automático, a continuación, se detalla algunas de estas herramientas utilizadas en el presente proyecto.

Weka

Es un software de código abierto y fácil instalación, tiene una interfaz gráfica para el usuario, la aplicación contiene varios modelos aprendizaje automático.

Orange

Es un software de minería de datos, permite la manipulación de datos con diferente extensión, el uso de métodos es más visuales e intuitivos.

R-Studio

R studio es un software dedicado a la computación estadística y gráfico, contiene variedad de librerías para la implantación de modelos de aprendizaje automático.

CAPÍTULO II

ESTADO DEL ARTE

La siguiente sección evidencia diferentes casos de estudio que se han recolectado, los cuales muestran mayores coincidencias con el presente proyecto, donde se aplicaron técnicas de aprendizaje automático para la predicción de valores de viviendas.

Los casos de estudio presentados fueron alcanzados a través de la base de datos “Scopus”, bajo determinados criterios de búsqueda se utilizaron las siguientes palabras: “real estate + neural network” y “real estate + deep learning”, con el objetivo de obtener mejores resultados y no limitar la búsqueda. el acceso la base datos Scopus se dio gracias a las credenciales otorgadas por la Universidad Internacional SEK del Ecuador.

La razón técnica de usar “Scopus” es debido a que cuenta con una gran cantidad de artículos científicos de gran calidad además de tener variedad de contenidos en diferentes disciplinas. Los artículos buscados en Scopus, arrojaron algunos resultados en idioma inglés, recolectados desde 2018 hasta 2021, como datos totales se obtuvieron 114 documentos al ingresar las palabras claves “real estate + neural network”. En la figura 17 se visualizan los resultados que arrojó la búsqueda en la revista “Scopus” con criterios y palabras claves como: “real estate + neural network” .



Figura 17 Muestra el resultado con las palabras claves ingresadas en Scopus.
Fuente: Base de datos Scopus

Ante todos los resultados obtenidos cabe señalar que el proyecto tiene como objetivo predecir el precio del inmueble por sectores en la ciudad de Quito al ser por sectores se tiene un valor más aproximado al real de cada uno, esto sumado a las medidas de desempeño implementadas en el presente trabajo, aportan a este tipo de investigaciones, así como a futuros trabajos.

En su totalidad fueron tres las investigaciones más sobresalientes que coincidieron con el presente trabajo, dado que se usan métodos de aprendizaje automático, y muestran el uso de variables como: área, precio, entre otros. Las cuales son:

- 1.- Análisis predictivo utilizando Big Data para el mercado inmobiliario durante la pandemia del COVID-19. (Grybauskas, Pilinkienė, & Stundžienė, 2021)
- 2.- Impulsar las predicciones de precios de la vivienda mediante la integración de redes geoespaciales. (Snigdha, Li, & Sellis, 2020)
- 3.- Inversión inmobiliaria en Dubái: un análisis predictivo y de series temporales (Krishna, 2021)

A continuación, se realiza una síntesis de los trabajos encontrados para el siguiente análisis.

- **Análisis predictivo utilizando Big Data para el mercado inmobiliario durante la pandemia del COVID-19.**

El estudio proporciona información sobre la comprensión de los modelos de Machine Learning aplicados, así también como cuáles fueron los más precisos para la predicción de bienes raíces, al igual aporta una visión a los hogares, empresarios e inversores que están relacionados con el sector inmobiliario, sobre las variables que se deben considerar para anticipar caídas de precios en este sector.

El trabajo describe también la recopilación de datos con el software web-scraping y los pasos metodológicos tomados en la construcción de los modelos de máquinas de aprendizaje, considerando variables como: tipo de edificación, años de antigüedad, tamaño en metros cuadrados, número de habitación, número de pisos, garajes y otras características de la vivienda,

además la cercanía a universidades, escuelas, centros comerciales, estaciones de tren como las características más significativas que afectan en el precio.

Entre los recursos empleados se tuvo el uso de Machine Learning, proceso que tuvo dos etapas distintas, divididas en conjuntos de datos, de prueba 70% y entrenamiento 30%, los métodos aplicados fueron los siguientes:

1. CatBoost Classifier
2. Extreme Gradient Boosting (XGB)
3. Light Gradient
4. BoostingMachine
5. Random Forest Classifier
6. Extra Trees Classifier
7. Gradient Boosting
8. Classifier Linear Discriminant Analysis
9. Logistic Regression
10. Ridge Classifier
11. Naive Bayes
12. Ada Boost Classifier
13. K-Neighbors Classifier
14. Decision Tree Classifier Quadratic Discriminant Analysis and SVM.

El estudio finaliza refiriéndose a los bienes raíces en épocas de pandemia, demostrando que la caída en los precios no fue tan catastrófica como se pensaba, una baja de valores promedio de cuatro meses solo alcanzó -7.20% y -4.2% para alquiler y venta operaciones, respectivamente. Comparado al modelo de pronóstico de llamado Giudice, que predijo una caída del $4,8\%$ en el corto plazo.

- **Impulsar las predicciones de precios de la vivienda mediante la integración de redes geospaciales.**

El objetivo de esta investigación fue predecir el valor de una vivienda, considerando que el precio es uno de los factores más críticos en el proceso de la compra de una casa y también resaltando la importancia de la ubicación, localización geoespacial, la cual se basa en la instalación del vecindario, además de la revisión de algunos aspectos como número de habitación, número de baños, área de construcción, etc.

El estudio considera como un factor para el alza de servicios, la cercanía del inmueble a escuelas, estaciones de tren, supermercados, contando que sea un área en una ubicación residencial cuyo ambiente vaya en armonía con el estilo de vida de los futuros compradores.

- **Inversión inmobiliaria en Dubái: un análisis predictivo y de series temporales**

El objetivo de este estudio fue proporcionar a los inversores inmobiliarios una cartera óptima para invertir en la ciudad de Dubái mediante un análisis predictivo, aplicando Machine Learning, para esto el conjunto de datos se basó en palabras claves como tierras de Dubái y departamento.

Para el análisis del estudio se seleccionaron los siguientes modelos: Regresión Lineal, Random Forest, y Árbol de decisión. En este trabajo los datos utilizados contemplan 21 columnas de las cuales 9 fueron variables potenciales para el análisis, entre ellas destacan el área, valor del metro cuadrado, valor de la renta.

El valor de la renta y su precio por metro cuadrado constituyó información crucial para el progreso de este trabajo, el cual demostró que las diez principales áreas de propiedad absoluta de Dubái elegidas por los inversores estaban en el centro de la ciudad, además se evidenció que la gente invierte principalmente en propiedades de una y dos habitaciones en comparación con propiedades tipo estudio o de tres o más habitaciones.

Se utilizaron modelos analíticos de datos de análisis predictivo para predecir el precio de venta de la propiedad en función de tres variables: tamaño de la propiedad en pies cuadrados, número de dormitorios (estudio, uno, dos, tres, cuatro, cinco y seis) y calidad de construcción (bajo, medio y alto).

El estudio contempla la predicción de precios inmobiliarios, la cual hace notar el grado de dificultad debido a muchos factores directos e indirectos que influyen en la precisión de los costos, siendo estos cuantitativos o cualitativos. Entre otras técnicas, se manejó la valoración de atributos de bienes raíces, por ejemplo, la inclusión de precios de ventas pasadas, área de terreno, años de construcción, espacio de piso, superficie, número de plantas y estado del edificio; se utilizaron cuatro modelos de aprendizaje automático para emprender el pronóstico de los bienes inmuebles.

CAPÍTULO III

METODOLOGÍA

3.1 Metodología

En la creación de una solución automática de un avalúo a un bien inmueble, es necesario comprender el proceso que un perito realiza y aplica al momento de realizar el avalúo, uno de los métodos que utilizan es el denominado “comparación”, el cual consiste en determinar el valor de suelo del área de intervención valorativa que no presenta determinados servicios e infraestructura, en comparación con otro de características homogéneas en cuanto a zonificación, uso y potencialidad” (Norma técnica de Valoración de los Bienes Inmuebles en el MDMQ, 2019).

Para este proyecto se utilizará la metodología CRISP-DM (CRoss Industrial Standard Process for Data Mining) porque describe los procesos clave para llevar a cabo un proyecto de minería de datos.

- 1.-Comprensión empresarial: al principio, el objetivo es obtener una visión general del sector inmobiliario.
- 2.-Preparación de datos: una fase que llevó mucho tiempo, ya que los conjuntos de datos seleccionados y sus campos tuvieron que limpiarse antes de que pudieran usarse.
- 3.-Modelado: en esta fase, el análisis se realizó utilizando las técnicas de modelado seleccionadas.
- 4.- Evaluación: una vez que se construyeron los modelos y en base a las métricas implantadas se procedió a la evaluación

En la figura 18 se muestra la estructura del proceso a utilizar para pronosticar el valor de un bien inmueble con el conjunto de datos obtenidos para la ciudad de Quito en el presente proyecto.

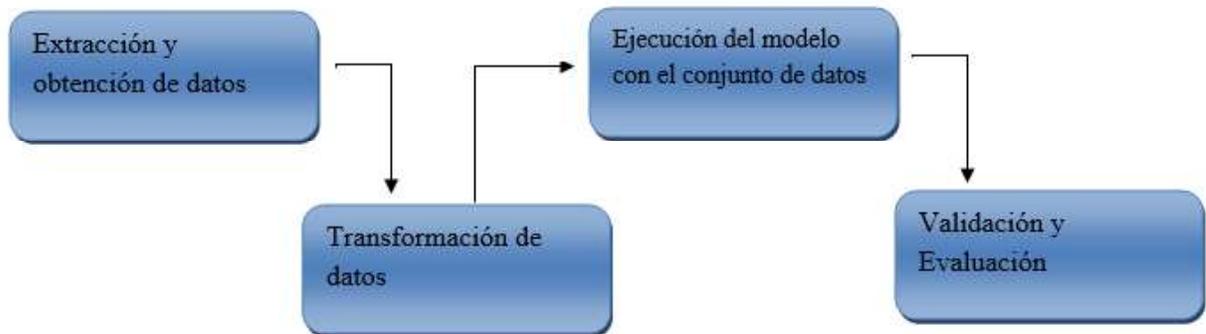


Figura 18. Estructura del proceso a utilizar.
Fuente: Elaboración propia

Extracción y obtención de los datos

Para obtener la recolección de los datos se buscó ayuda en las páginas webs disponibles, en conformidad a lo que ellas mostraban en sus portales, en el caso de algunas páginas web los datos se obtuvieron directamente en formato csv y en otros usando *web scraping*.

Transformación de los datos

Examinar el archivo con los datos obtenidos, realizar la limpieza de los datos, así como eliminar registros en blanco.

Ejecución de modelos a los datos.

Con los datos procesados, aplicamos los diferentes métodos de Machine Learning.

Validación y Evaluación

Luego de la ejecución del modelo, se realiza el análisis y evaluación con las medidas de desempeño de cada uno de los métodos aplicados para seleccionar el óptimo.

3.2 Construcción de los datos

En el estudio de la Analítica de datos, existe una amplia gama de métodos, los cuales, se componen de tres aspectos: adquisición y almacenamiento de los datos, limpieza y depuración de los datos y, la preparación para su análisis. (Cumba & Riofrío, 2018)

3.2.1 Fuente

Las páginas web visitadas para la elaboración de esta investigación fueron: PROPERATI, PLUSVALÍA, REMAX, en las cuales se aplicó la herramienta *web scraping*, de manera que se recolectó la información para el presente proyecto.

3.2.2 Extracción de datos

El levantamiento de los datos, se realizó mediante el acceso a la URL de las páginas inmobiliarias antes mencionadas, en donde se encuentran información publicadas de casas y posterior guardado en formato csv. entre todas ellas destaca PROPERATI debido a que en su portal permite la descarga archivos csv, esto más los datos descargados con la herramienta *web scraping* de las otras páginas web constituyeron el conjunto datos para el presente proyecto, es necesario resaltar que los datos obtenidos corresponden al año 2020.

3.2.3 Transformación de los datos

En la transformación del conjunto de datos descargados, inicialmente se analizó con la herramienta Excel los datos de los inmuebles recolectados, esto con el propósito de visualizar el contenido de

la información, de modo que fue el punto de partida para continuar con el proceso de transformación de los datos.

3.2.3.1 Visualización

Una vez descargados el conjunto de datos y para el análisis del presente trabajo se observó que sus columnas estaban en inglés: "id", "ad_type", "start_date", "end_date", "created_on", "lat", "lon", "11", "12", "13", "14", "15", "16", "rooms", "bedrooms", "bathrooms", "surface_total", "surface_covered", "price"; esta información consta de 288984 observaciones y de 24 columnas o variables.

Se analizó el contenido de cada columna y se verificó su aporte para esta investigación, para un mejor entendimiento, todo ello se describe en forma detallada en la Tabla 2.

Tabla 2 . Descripción de las variables.

Fuente: Elaboración propia

Variable	Descripción
ad_type	Tipo de propiedad
Id	Código generado automáticamente numérico
start_date	Fecha de publicación
created_on	Fecha de creación de la publicación
end_date	Fecha de finalización
Lat	Latitud
Lon	Longitud
11	País
12	Provincia
13	Cantón
14	Nombre de la ubicación del sector
15	Nombre del sector
16	Nombre del barrio
Rooms	Número de cuartos
bedrooms	Número de baños
Área	Área total en m ² del inmueble
Construcción	Área de construcción
Price	Precio en \$
Surface	Superficie en m ²
Currency	Divisa USD
Description	Descripción de que el usuario ingresa
Title	Título de la publicación
Price_period	Periodos de los precios

3.2.3.2 Limpieza de columnas y de registros

Para la limpieza se procedió con la verificación de las características de los datos, se seleccionó y eliminó columnas, esto debido al contenido de las mismas ya que muchas de estas columnas no aportaban valor significativo para el presente trabajo. En la tabla 3 se indican el nombre de la variable y una breve descripción de su contenido.

Tabla 3. Tabla de variables descartadas
Fuente: Elaboración propia

Variable	Descripción
id	Código generado automáticamente numérico y no aporta con información.
created_on	Fecha de creación de la publicación.
end_date	Fecha de finalización.
Surface	Superficie en m^2
Currency	Divisa USD
Description	Descripción de que el usuario ingresa.
Title	Título de la publicación
Price_period	Periodos de los precios

Para selección y limpieza de registros dentro de los datos se procedió a utilizar comando de R-studio, se eliminaron indicios que contenían = "9999-12-31" y vacíos, debido que no aportan a los datos.

De acuerdo con la investigación “Modelo de Clasificación de Riesgo Crediticio utilizando Random Forest en financiera del Ecuador”, se debe revisar si existen valores faltantes que pueden generar errores al momento de entrenar al algoritmo”. (Freire & Guevara, 2021)

Conjunto de datos resultantes

El resultado de la transformación de los datos mediante la visualización, análisis y limpieza de registros permitió obtener un conjunto de datos listos para trabajar en el presente proyecto.

Aplicación de métodos automáticos

En el presente trabajo se aplicaron algunos algoritmos de aprendizaje automático como son regresión lineal, árboles decisión, Random Forest, redes neuronales, con el fin de responder al objetivo planteado y posterior medir el desempeño de los modelos.

En la implantación de los métodos automáticos se hizo uso del software de código abierto R Studio, así como Orange y Weka antes descritos de la sección marco teórico.

Análisis de los datos

Se utiliza una gráfica para analizar la correlación que presentan los datos, los círculos azules grandes representan la mejor correlación entre las variables.

En la figura 19 se muestra la correlación de los variables que posee la fuente de datos, siendo los círculos más grandes los de mejor correlación hasta llegar a los círculos más pequeños con escasa correlación entre las variables, obteniendo que las variables de mayor correlación son el área, la construcción y el precio.

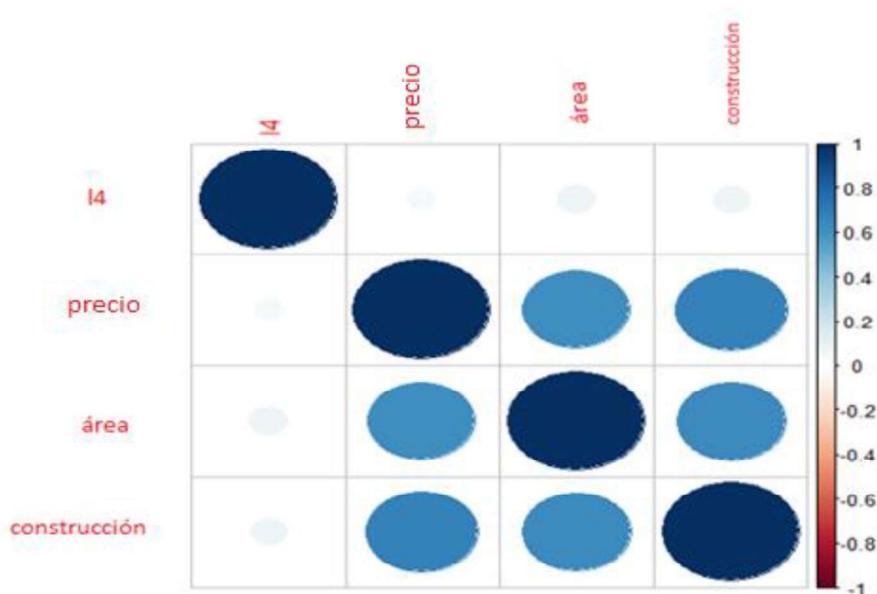


Figura 19. Muestra la correlación de los variables.
Fuente: Elaboración propia.

En la tabla 4 se observa que para el gran conjunto de datos se procedió a segmentar en sectores a algunas zonas de la ciudad de Quito, estas secciones están enumeradas del 1 al 7, esto debido a que en los datos descargados constaban dichos sectores, posterior a ello se realizó su respectivo análisis.

Tabla 4. Tabla de sectores de Quito de acuerdo a los datos descargada.
Fuente: Elaboración propia

Sector	Número
Norte	1
Sur De Quito	2
Valle de Tumbaco	3
Centro Colonial	4
Valle Los Chillos	5
Centro Norte	6
Otros	7

3.2.4 Algoritmos

Un primer aspecto a considerar es que existen varias técnicas de aprendizaje automático, las cuales pueden clasificar, agrupar o predecir datos. Por consiguiente, se utilizarán los modelos antes descritos en la sección de marco teórico, dichos modelos se aplicaron a los datos obtenidos. Los modelos utilizados fueron:

3.2.4.1 Regresión Lineal

La técnica de Regresión Lineal permite predecir el valor de una variable con valores de otra, en este caso se usaron las variables *price y área*, con las dos se puede ver la relación entre el área y el precio.

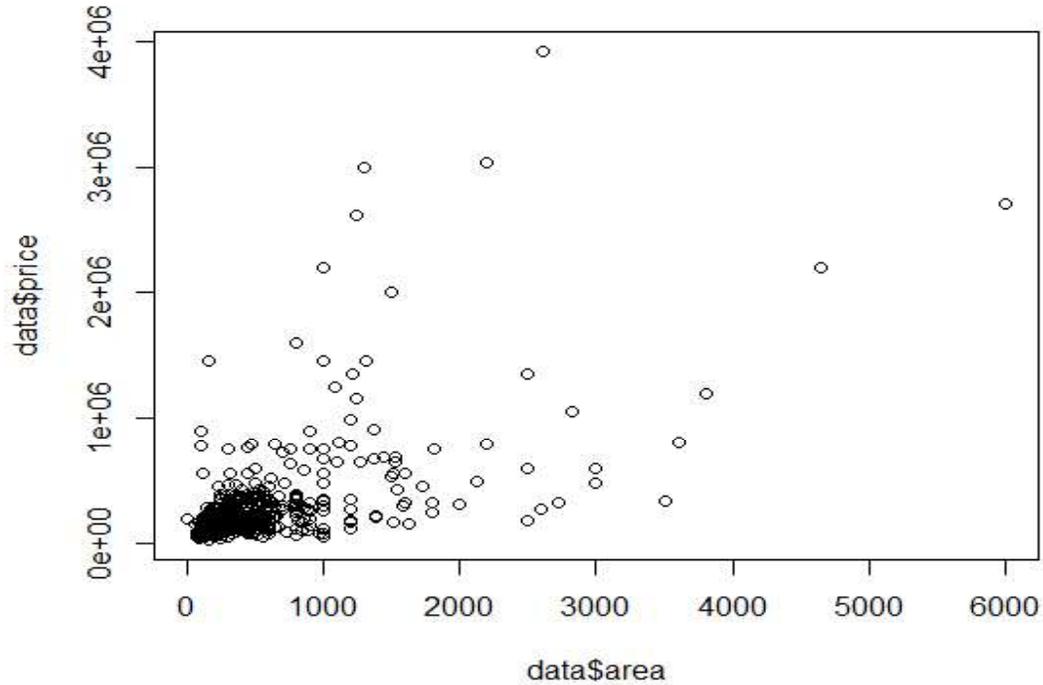


Figura 20. Muestra el diagrama de dispersión.
Fuente: Elaboración propia

La figura 20 muestra la dispersión de los datos en el eje x tenemos el área en m^2 y el eje y el precio en USD, podemos observar el comportamiento de los datos y sus variaciones, los cuales indican que a medida que el área del inmueble aumenta el precio también lo hace.

De modo similar se puede analizar la correlación por medio la correlación de *Pearson*. En la figura 21 se muestra la correlación de *Person* tomando como parámetros las variables *price* y *área*. Al realizar la correlación de las variables precio y área se aprecia que existe una fuerte correlación de las *price* y *área*, y dado que valor de $p - value < 2.2e^{-16}$ existe una correlación lineal.

```

> cor.test(data$price,data$area)

Pearson's product-moment correlation

data: data$price and data$area
t = 21.149, df = 741, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5666028 0.6564812
sample estimates:
 cor

```

Figura 21. Correlación de los datos con los atributos Price y área.
Fuente: Elaboración propia

De esta manera se puede plantear un modelo de regresión lineal llamado “Modelo_RL” atreves de la función *lm* y posterior se puede visualizar los resultados con *summary(Modelo_RL)*, de modo que se pueda interpretar sus resultados.

```

> summary(Modelo_RL)

Call:
lm(formula = data$price ~ data$area)

Residuals:
    Min       1Q   Median       3Q      Max
-1077584  -67573  -30817   21545  2852068

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 78644.45  12543.01    6.27 6.13e-10 ***
data$area    379.70    17.95   21.15 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 266100 on 741 degrees of freedom
Multiple R-squared:  0.3764,    Adjusted R-squared:  0.3756
F-statistic: 447.3 on 1 and 741 DF,  p-value: < 2.2e-16

```

Figura 22. Muestra los resultados de la estimación con modelo Regresión Lineal.
Fuente: Elaboración propia.

Al observar la figura 22, una vez aplicado el modelo de Regresión Lineal se muestra el valor *pi value* es menor 0.05, trabajando el modelo a un 95% de confianza, al tener un valor de significancia menor a 0.05. podemos asegurar que el modelo es válido.

Para formar la ecuación tomamos los valores de los coeficientes 78644.45 y 379.70 de la figura 22.

$$y = 78644.45 + 379.70 x \quad (4)$$

La ecuación 4 de la recta nos permite predecir el valor de un bien inmueble con cierta área, según la ecuación de la Regresión Lineal, basta desplazar x por el área en m^2 del bien inmueble.

3.2.4.2 Random Forest

El algoritmo de bosque aleatorio funciona mediante la agregación de las predicciones realizadas por múltiples árboles de decisión de profundidad variable, este método permite clasificar correctamente cada variable en función de un objetivo determinado.

En el conjunto de datos que se tiene al utilizar el Random Forest las posibilidades de obtener un valor más razonable para las propiedades por sectores son altas, ya que la decisión se basa en los resultados de múltiples árboles de decisión.

La figura 23 muestra la aplicación de método Random Forest, se creó una partición de los datos con un 0.70 de los datos originales, para luego se construye un *Modelo* con Random Forest, donde x representa las variables independientes, y represente la dependiente, la variable *n_{tree}* representa el número de árboles a generar. Posterior realizamos la predicción con el restante de los datos. Dando como resultado la clasificación de los bienes inmuebles en cada sector.

```

set.seed(123)
entrenamiento <- createDataPartition(dataQuito$14, p=0.70,list=FALSE)
Modelo<- randomForest(x= dataQuito[entrenamiento,2:4],
                      y= dataQuito[entrenamiento,1],
                      ntree=500,
                      keep.forest=TRUE)

prediccion <- predict(Modelo,dataQuito[-entrenamiento,])

```

Actual	Predicho						
	1	2	3	4	5	6	7
1	203	3	9	0	4	0	0
2	1	16	0	0	1	0	0
3	5	1	65	0	3	1	2
4	3	0	10	60	0	0	0
5	6	0	0	0	20	0	0
6	2	0	2	0	1	8	0
7	2	0	1	0	0	0	11

Figura 23 Método de Random Forest para el conjunto de datos Fuente: Propia

3.2.4.3 Árbol decisión

El Árbol de decisión es un conjunto de algoritmos estadísticos que a través de modelos predictivos, clasifican datos según ciertas características o propiedades obteniendo como resultado modelos precisos y confiables. En un inicio se utilizan el *área*, *construcción* del bien inmueble, lo cual se puede verificar que influye más para el *price* como en este caso.

En el estudio de resoluciones, es posible implementar un Árbol de decisiones para mostrar de forma visual el árbol y así en base a un conjunto de variables tomar una decisión acertada.

En la figura 24 el algoritmo “n” permite visualizar la construcción del árbol de acuerdo con los datos ingresados y sus variables, en el nodo raíz creado por el árbol, se toma como atributo principal el *Price* (valor de bien inmueble), en un primer nivel podemos observar el 100% de los datos la primera distinción que realiza en base a precio, es decir los inmuebles con valores menores al tomado como referencia es un 66% y los inmuebles mayores a dicho valor corresponde al 34%.

La segunda distinción realizada corresponde al 34% y toma al precio para distinguir un 25% menores al precio referido y un 8% mayores al este. Para la tercera distinción toma a la construcción, de la construcción menores a 254 m^2 corresponde al 13% y las construcciones mayores a este valor está 13%.

La cuarta distinción corresponde 13% y toma al área de inmueble mostrando como resultado que el área mayor 318 m^2 corresponde al 3% y menores a este valor está en 9%.

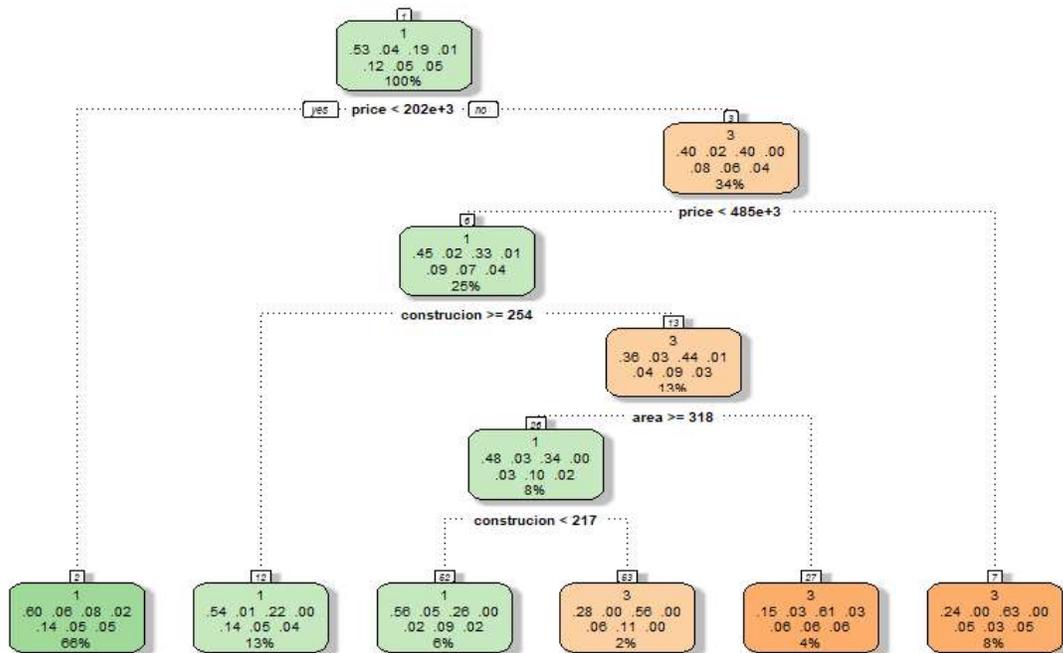


Figura 24. Muestra la formación del árbol decisión.
Fuente: Elaboración propia.

3.2.4.4 Redes Neuronales

La red neuronal es un algoritmo de aprendizaje automático, puede resolver problemas mediante aprendizaje automático, mismo que entrena redes neuronales con diferentes estructuras, lo que permite extraer con mayor precisión mejores resultados. (Shia & Zhangb, 2022)

Las redes artificiales están elaboradas por nodos o “neuronas” que reciben, transmiten y envían información, suelen estar conformadas por capas. Estas redes son entrenadas lo cual les permite crear, modificar o eliminar alguna conexión entre los nodos con el objetivo de dar una respuesta.

Se aplica el método de red neural, para la data presentada en la presente red neuronal, como entrada se tienen todas las variables independientes que se correlacionan con la variable *price*.

La figura 25 muestra la entrada de la red neuronal y su salida que es el objetivo de la red, su objetivo es el *Price* (precio), la red neuronal presentada se realizó en 193 pasos, la cual trabaja con 9 neuronas en la primera capa y para la segunda capa 7 neuronas.

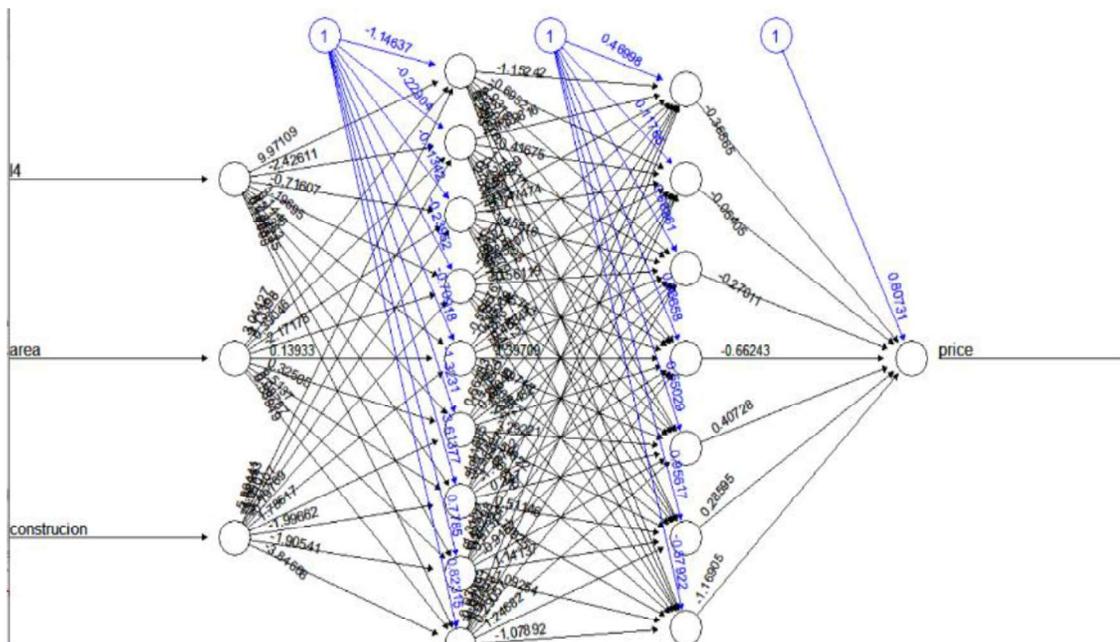


Figura 25 Muestra la estructura de red neuronal.
Fuente: Elaboración propia

3.2.5 Evaluación de métodos automáticos

En este apartado se muestra el resultado de las medidas de desempeño tras la ejecución de los modelos en los siete conjuntos de datos. El procedimiento para todos los modelos fue igual en cada caso tanto para el entrenamiento y prueba.

1.- Resultado sector norte

A continuación, la Tabla 5 muestra el resultado de las medidas de desempeño generados por los modelos aplicados a los datos en el sector norte de Quito.

Tabla 5. Muestra los resultados obtenidos con el sector norte.
Fuente: Elaboración propia

	<u>Regresión Lineal</u>	<u>Árbol decisión</u>	<u>Random Forest</u>	<u>Red Neuronal</u>
R^2	0.36	0.087	0.188	-0.57
RMSE	205148.504	245482.743	231397.99	322566.067
CV	1.0	1.25	1.18	1.65

El resultado de la aplicación de los modelos concluye que el mejor coeficiente de determinación es la Regresión Lineal con un valor de 0.36, igualmente, se puede apreciar que el menor error RMSE es la Regresión Lineal, esto permite verificar que el valor de predicción del inmueble se acerca al real en el sector, y por último se mira como el coeficiente de variación también es menor en la Regresión Lineal.

2.- Resultado sector sur de Quito

Tabla 6. Muestra los resultados obtenidos con el sector 2.
Fuente: Elaboración propia

	<u>Regresión Lineal</u>	<u>Árbol decisión</u>	<u>Random Forest</u>	<u>Red Neuronal</u>
R2	0.24	-0.40	0.125	-3.23
RMSE	44852.982	61047.422	48207.507	106028.419
CV	0.48	0.65	0.24	0.54

En la Tabla 6 se pueden determinar que el mejor coeficiente de determinación es la Regresión Lineal con un valor de 0.24, seguido de Random Forest, esto permite verificar que el valor de predicción del inmueble se acerca al real en el sector; se puede evidenciar que el coeficiente de variación también es menor en Random Forest, el método coincide con el sector norte.

3.- Resultado sector valle de Tumbaco

Tabla 7. Muestra los resultados obtenidos con el sector 3
Fuente: Elaboración propia

	<u>Regresión Lineal</u>	<u>Árbol decisión</u>	<u>Random Forest</u>	<u>Red Neuronal</u>
R2	0.62	0.48	0.54	-0.78
RMSE	340741.160	396205.408	374709.079	490164.994
CV	0.69	0.80	0.21	0.99

La Tabla 7 muestra el rendimiento del sector del valle de Tumbaco, en sus resultados indica que el mejor coeficiente de determinación es la Regresión Lineal con un 0.62, seguido de Random

Forest con 0.54, para este sector la red neuronal es el de menor desempeño, se muestra que el valor de los inmuebles con el sector sur de Quito es más económico versus el sector valle de Chillos.

4.- Resultado sector centro colonial

Tabla 8. Muestra los resultados obtenidos con el sector 4.
Fuente: Elaboración propia

	<u>Regresión Lineal</u>	<u>Árbol decisión</u>	<u>Random Forest</u>	<u>Red Neuronal</u>
R2	-0.11	-0.19	-0.33	-12.97
RMSE	161446.168	49453.644	55661.92	163246.749
CV	0.98	0.30	0.34	0.33

Como se puede ver en la tabla 8 se muestra como los modelos aplicados no contribuyen al análisis, ello se debe a que los datos para este sector son reducidos y en consecuencia no aportan resultados significativos, por lo que se excluyen de este diagnóstico.

5.- Resultado sector valle de los Chillos

Tabla 9. Muestra los resultados obtenidos con el sector 5.
Fuente: Elaboración propia

	<u>Regresión Lineal</u>	<u>Árbol decisión</u>	<u>Random Forest</u>	<u>Red Neuronal</u>
R2	0.67	0.18	0.49	-1.2
RMSE	89926.373	143180.899	113115.389	240211.485

CV	0.49	0.79	0.62	1.3
----	------	------	------	-----

Con los resultados de la Tabla 5 se pueden evidenciar que el método de Regresión Lineal tiene un mayor coeficiente de determinación, seguido de Random Forest; revisando los anteriores resultados para el sector norte ,2,3, se deduce que la Regresión Lineal tiene un mejor impacto en los sectores.

6.- Resultado sector centro norte

En la Tabla 10 se verifica que el mejor coeficiente de determinación es el método de Regresión Lineal, se puede evidenciar que los bajos resultados se deben al escaso conjuntos de datos del sector centro norte.

Tabla 10. Muestra los resultados obtenidos con el sector 6.
Fuente: Elaboración propia

	<u>Regresión Lineal</u>	<u>Árbol decisión</u>	<u>Random Forest</u>	<u>Red Neuronal</u>
R2	0.07	-0.9	-0.4	-1.8
RMSE	147133.416	211231.11	183767.651	258027.961
CV	0,70	0.98	0.88	0.99

7.- Resultado sector otros

Tabla 11. Muestra los resultados obtenidos con el sector.
Fuente: Elaboración propia

<u>Regresión Lineal</u>	<u>Árbol decisión</u>	<u>Random Forest</u>	<u>Red Neuronal</u>
-------------------------	-----------------------	----------------------	---------------------

R2	-0.13	-0.23	0.11	-1.17
RMSE	202503.134	211365.420	178830.719	280151.443
CV	0.98	0.99	0.86	1.36

La Tabla 11 muestra como resultado para el sector otros el coeficiente de determinación con mayor importancia es Random Forest con 0.11 y su coeficiente de variación de 0.86 es menor comparado con los otros métodos implementados.

8.- Análisis de resultados con ROC Y AUC

Del análisis realizado y los resultados obtenidos con los métodos aplicados, se puede realizar el llamado ROC análisis y AUC, que permiten verificar el rendimiento de clasificación de los métodos aplicados.

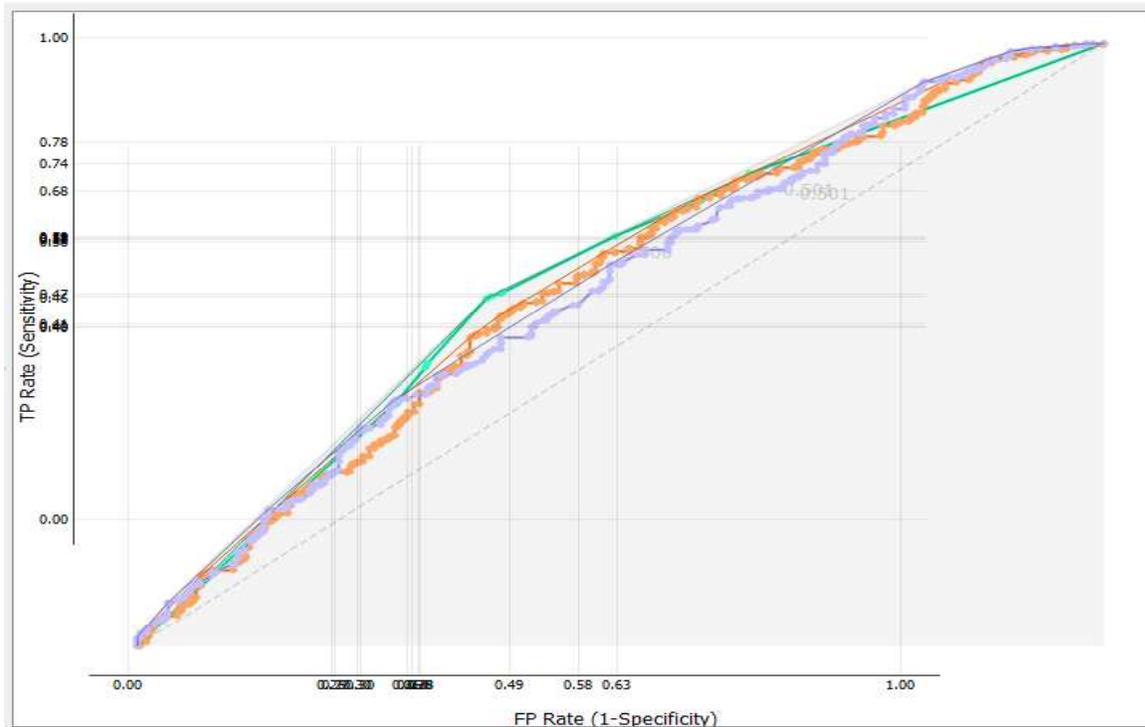


Figura 26. Muestra análisis ROC y AUC
Fuente:Propia

La Figura 25 muestra la evaluación de los métodos, se observa que, por cada método aplicado, sobresale el área bajo de la curva AUC con su respectivo valor.

En la figura de la curva ROC, el eje Y corresponde a la proporción de verdaderos positivos sobre el total de los inmuebles en los sectores establecidos, y el eje X corresponde a la proporción de falsos positivos sobre el total de inmuebles en los sectores establecidos; examinando la Figura de curva ROC se muestra la "proporción de verdaderos positivos" (eje Y) versus la "proporción de falsos positivos" (eje X)

CAPÍTULO IV

RESULTADOS

En esta sección se detallan las técnicas aplicadas en la elaboración del presente trabajo, así como su interpretación, en total se trabajaron con 1263 registros, los cuales, en la aplicación de los métodos escogidos, contaron con una partición de 70% para entrenamiento y 30% para pruebas, de esta manera se pudieron analizar los resultados obtenidos y su desempeño.

4.1 Rendimiento de los datos

Se elaboró la Tabla 12 como explicación y resumen de las técnicas aplicadas para cada conjunto de datos, mostrando los porcentajes antes mencionados

Tabla 12. Descripción de métodos aplicados a los datos divididos por sectores.
Fuente: Elaboración propia.

Clase	Desempeño	Datos	Partición	Variables Utilizadas
Modelo de Regresión Lineal Árboles de decisión Random Forest	Correlación entre variables	Consta 1263 registros	Contempla la siguiente partición: 70% Entrenamiento 30% prueba	14 sectores de Quito, Price, Área m2, Construcción m2

Luego de obtener los resultados de los métodos aplicados, se puede observar que la Regresión Lineal predomina sobre los sectores 2,3,5,6 a diferencia del sector 7 en el que el método de Random Forest fue el mejor.

En mención al RMSE, se verifica que los sectores 1, 3 obtuvieron un valor alto en las viviendas, así se puede decir que los sectores con mayor valor referente a las viviendas son esas zonas, el modelo que más predominó fue la regresión lineal.

Con los resultados arrojados por RMSE distribuidos para cada sector, se obtuvo un valor de los inmuebles por sectores.

Tabla 13. Muestra el mejor ajuste al precio de las viviendas de acuerdo RMSE.
Fuente: Elaboración propia

Sector	RMSE
Norte	\$ 251148,83
Sur De Quito	\$ 65034,082
Valle Tumbaco	\$ 400455,160
Valle Los Chillos	\$ 146608,53
Centro Norte	\$ 200040,03
Otros	\$ 218212,679

De acuerdo con los valores pronosticados por los métodos, se muestra en la Tabla 13 el promedio de los posibles valores alcanzados por los métodos de aprendizaje automático; se evidencia que el sector con alta plusvalía de acuerdo con el costo de las viviendas es el Valle de Tumbaco y el sector con menor plusvalía es el sur de Quito.

Aplicando el análisis de ROC y AUC a los métodos aplicados, se muestran los valores de AUC y precisión en la Tabla 14.

Tabla 14 Análisis ROC y área bajo la curva
Fuente: Propia

MODELO	AUC	Precisión
Árbol decisión	0.59	0.45
Random Forest	0.64	0.40
Red Neuronal	0.65	0.39

Mediante el análisis ROC y el AUC arrojados por los modelos aplicados a los datos de Quito, la Regresión Lineal presenta el mayor porcentaje de precisión al momento de predecir el valor del inmueble en el conjunto de datos.

CAPÍTULO IV

CONCLUSIONES Y TRABAJOS FUTUROS

- En este estudio se propuso un proceso de obtención de datos mediante el uso de web-scraping, esta herramienta proporciono datos de los diferentes portales web disponibles los cuales están involucrados en la rama de bienes inmuebles, estos datos fueron de gran utilidad al momento de entrenar los modelos implementados.
- El uso de diferentes técnicas de aprendizaje automático, así como el uso de diferentes librerías para aplicar dichos modelos en los datos, se obtuvieron resultados satisfactorios, destacando la utilidad de la aplicación de regresión lineal, random forest, redes neuronales y árboles de decisión, siendo el método de regresión lineal el más destacado para determinar el valor de un inmueble, en vista de las medidas de desempeño R^2 , $RMSE$ y CV implantadas al conjunto de datos.
- La limpieza de la información y procesamiento de los datos contempló la reducción de los registros repetidos, datos vacíos, selección de características relevantes de los datos con la finalidad de obtener un conjunto de datos óptimo y aplicar los métodos descritos en este trabajo.
- Se recomienda para trabajos futuros tomar en cuenta más variables de las consideradas en el presente trabajo como es el caso de años de antigüedad de la vivienda, ubicación geográfica, costo de remodelación entre otros. Así como también el uso de otros modelos de aprendizaje automático.

BIBLIOGRAFÍA

1. Abdullah, Waterworth, A., & Shapira, P. (2015). Use of web mining in studying innovation. 1.
2. Amenga, X. (2020). Analysis of the impact of business administration informatization. *ScienceDirect*, 3.
3. AntriaNovation. (22 de 10 de 2019). *Inteligencia Artificial*. Obtenido de Inteligencia Artificial: <https://www.atriainnovation.com/que-son-las-redes-neuronales-y-sus-funciones/>
4. Calvo, D. (23 de 3 de 2019). *Definición de aprendizaje supervisado*. Obtenido de <https://www.diegocalvo.es/aprendizaje-supervisado/>
5. Chiarazzo, V., Caggiana, L., & Marinellia, M. (2014). Modelo basado en redes neuronales para la estimación de precios de bienes raíces.. *ScienceDirect*, 811.
6. Concepto Jurídico. (2021). *Bienes Inmuebles*. Obtenido de conceptojuridico.com
7. Cumba, D., & Riofrío, D. (1 de 11 de 2018). *repositorio.uisek*. Obtenido de [repositorio.uisek: https://repositorio.uisek.edu.ec/bitstream/123456789/3224/2/TESIS%20MTI%20PAUL%20CUMBA.pdf](https://repositorio.uisek.edu.ec/bitstream/123456789/3224/2/TESIS%20MTI%20PAUL%20CUMBA.pdf)
8. Davis, J. L., & Rudolph, A. (2002). *Neurotechnology for biomimetic robots*. Londres: Massachusetts Institute of technology.
9. *deusto formacion*. (13 de 11 de 2018). Obtenido de [deusto formacion: https://www.deustoformacion.com/blog/programacion-tic/que-son-datasets-dataframes-big-data](https://www.deustoformacion.com/blog/programacion-tic/que-son-datasets-dataframes-big-data)
10. Distrito Metropolitano de Quito. (2014). *Diagnóstico del Territorio del DMQ*. Quito: desc.
11. Dr. Swapna Borde, A. R. (2017). *Asesoramiento en Inversiones Inmobiliarias Mediante Machine Learning*. Obtenido de *Asesoramiento en Inversiones Inmobiliarias Utilizando Machine Learning*:: <https://www.semanticscholar.org/paper/Real-Estate-Investment-Advising-Using-Machine-Borde-Rane/627e72cfcf666922d8b59e1e014e6aea51acf69d>
12. Ekos. (27 de 10 de 2021). Obtenido de <https://ekosnegocios.com/articulo/el-sector-inmobiliario-se-dinamiza>
13. EMOTIV Limited. (2010). *EMOTIV Software Development Kit, release 1.0.0.5*.
14. Etecé, E. (05 de 08 de 2021). *concepto*. Obtenido de [concepto: https://concepto.de/](https://concepto.de/)
15. Freire, J., & Guevara, B. (10 de Agosto de 2021). *Modelo de Clasificación de Riesgo Crediticio Utilizando Random Forest en financiera del Ecuador*. Obtenido de

<https://repositorio.uisek.edu.ec/handle/123456789/4256>:
<https://repositorio.uisek.edu.ec/handle/123456789/4256>

16. García, C. (02 de 01 de 2022). *indrscompany*. Obtenido de <https://www.indracompany.com/>
17. Gujarati. (2009). *Econometria*. Mexico: McGRAW-HILL.
18. healthdataminer. (16 de 12 de 2021). *healthdataminer*. Obtenido de healthdataminer: <https://healthdataminer.com/data-mining/aprendizaje-supervisado-y-no-supervisado/>
19. Higuera, J. C. (12 de 12 de 2021). El sector inmobiliario crece luego del rezago ocasionado por la pandemia. *eltelegrafo*.
20. <https://www.remax.com.ec/>. (s.f.). <https://www.remax.com.ec/>. Obtenido de remax.
21. Jarrín, D., & Riofrío, D. (6 de 09 de 2019). *Repositorio Digital SEK*. Obtenido de Repositorio Digital SEK: https://repositorio.uisek.edu.ec/bitstream/123456789/3532/1/Tesis_DAVID_JARRIN_VERSION_FINAL.pdf
22. Koller, S. (26 de 3 de 2022). *¿Qué es el web scraping?* Obtenido de <https://seranking.com/es/blog/web-scraping/>
23. Ley Orgánica de Ordenamiento Territorial, Uso y Gestión del Suelo . (2016). Quito: Asamblea Nacional.
24. Linqi, Zhou, X., Liu, W., & Kong, Z. (2021). Predicción de registro de contenido de carbono orgánico total basada en aprendizaje automático. 10.
25. López, J. F. (10 de 08 de 2021). *repositorio.uisek.edu.ec*. Obtenido de repositorio.uisek.edu.ec: <https://repositorio.uisek.edu.ec/bitstream/123456789/4256/1/Freire%20L%20c%3%b3pez%20c%20Juan.pdf>
26. Matich, D. J. (2001). Redes Neuronales . En D. J. Matich, *Redes Neuronales* (pág. 40). Rosario, Argentina.
27. Moreno, C. G. (03 de 02 de 20022). *indracompany*. Obtenido de [indracompany](https://www.indracompany.com/): <https://www.indracompany.com/>
28. Moyano, E. (2008). Cartografía de la psicología contemporánea: pluralismo y modernidad. Santiago de Chile: LOM.
29. Municipio de Quito. (3 de 01 de 2022). Obtenido de <https://www.quito.gob.ec/>
30. Norma técnica de Valoración de los Bienes Inmuebles en el MDMQ. (2019). *Normas Técnica para la valoración de bienes inmuebles urbanoss y rurales del Distrito Metropolitano de Quito*. Quito.

31. Oliver, B. (1 de 11 de 2019). *¿Qué es una página web?* Obtenido de <https://www.aboutespanol.com/que-es-una-pagina-web-3202308>
32. Peiró, R. (08 de 07 de 2019). *economipedia*. Obtenido de [economipedia: https://economipedia.com/](https://economipedia.com/)
33. Peiro, R. (08 de 07 de 2019). *economipedia.com*. Obtenido de [economipedia.com: https://economipedia.com/definiciones/pagina-web.html](https://economipedia.com/definiciones/pagina-web.html)
34. Plan de Uso y Gestión de Suelo del Distrito Metropolitano de Quito. (2021). Quito.
35. PLAN METROPOLITANO DE DESARROLLO DE ORDENAMIENTO TERRITORIAL. (2021-2023). *Gobierno Municipal de Quito*. Obtenido de <https://gobiernoabierto.quito.gob.ec/plan-pmdot/>
36. Plusvalía. (15 de 01 de 2022). Obtenido de www.plusvalia.com
37. Properati. (14 de 01 de 2002). *Definición de Properati*. Obtenido de <https://www.properati.com.ec/>
38. realia. (01 de 01 de 2019). Obtenido de www.realia.es: <https://www.realia.es/>
39. Realia. (26 de 3 de 2022). *¿Qué es el mercado inmobiliario?* Obtenido de <https://www.realia.es/que-es-mercado-inmobiliario>
40. remax. (s.f.). *remax*. Obtenido de [remax: https://www.remax.com.ec/](https://www.remax.com.ec/)
41. Rodríguez, T. (27 de 3 de 2017). *Discutiendo entre el Machine Learning y el Deep Learning ¿A qué nos referimos con cada uno?* Obtenido de <https://www.xataka.com/robotica-e-ia/machine-learning-y-deep-learning-como-entender-las-claves-del-presente-y-futuro-de-la-inteligencia-artificial>
42. Santos, P. R. (02 de 12 de 2021). *thinkbig*. Obtenido de [thinkbig: https://empresas.blogthinkbig.com/](https://empresas.blogthinkbig.com/)
43. scopus. (s.f.). *scopus*. Obtenido de [scopus: https://www.scopus.com/results/results.uri?sort=plf-f&src=s&st1=real+state+and+predicton&sid=0ce392c556a8dd41a6964a7d3ed083cb&sot=b&sdt=b&sl=43&s=TITLE-ABS-KEY+%28+real+state+and+prediction+%29&origin=resultslist](https://www.scopus.com/results/results.uri?sort=plf-f&src=s&st1=real+state+and+predicton&sid=0ce392c556a8dd41a6964a7d3ed083cb&sot=b&sdt=b&sl=43&s=TITLE-ABS-KEY+%28+real+state+and+prediction+%29&origin=resultslist)
44. Shia, Y., & Zhangb, Y. (2022). The neural network methods for solving Traveling Salesman. *ELSEVIER*, 2.
45. Spiegato. (26 de 3 de 2022). Obtenido de <https://spiegato.com/es/que-es-un-conjunto-de-datos>
46. Superintendencia De Bancos. (15 de 12 de 2021). *Calificación Peritos Valuadores*. Obtenido de <https://www.superbancos.gob.ec/bancos/ufaq/calificacion-peritos-valuadores/>

47. Tahraoui, H., Amrane, A., Belhadj, A.-E., & Zhang, J. (2021). Modeling the organic matter of water using the decision tree. *ELSER*, 4.
48. Vozmediano, J. R. (19 de 10 de 2018). *CONCEPTO DE VALORACIÓN INMOBILIARIA Y NORMATIVA APLICABLE*. Obtenido de <https://negocioinmo.com/concepto-de-valoracion-inmobiliaria-y/>
49. wikipedia. (10 de 01 de 2022). *wikipedia*. Obtenido de wikipedia: 2022
50. Yalçınkaya, A., Balay, I. G., & Enoğlu, B. S. (2021). A new approach using the genetic algorithm for parameter estimation in. *Elsesr*, 1.
51. Zakaria, F., & Fatine, F. A. (2021). Towards the hedonic modelling and determinants of real estates price. *ELSERVER*, 1.